Advancements towards sequencing the bread wheat genome: An update of the projects of the International Wheat Genome Sequencing Consortium.

Kellye Eversole, IWGSC, Eversole Associates, 5207 Wyoming Road, Bethesda, MD 20816, USA.

Genome sequences hold the key for understanding the molecular basis of phenotypic traits and variation. Although the sequencing of model plant genomes, such as *Arabidopsis thaliana* and rice, has revolutionized our understanding of plant biology over the past 10 years, it has not translated efficiently into crop improvement for maize, wheat, or barley. At the same time, comparative genomic studies have revealed the limits of conservation between rice and the other cereal genomes; thereby necessitating the development of genomic resources and programs for maize, sorghum, wheat, and barley to serve as the foundation for future genome sequencing and the acceleration of genomic-based improvement of these essential crops. Despite the recognition that genome sequencing is critical for crop improvement, the size and complexity of the Triticeae genomes has been perceived as an obstacle for the efficient development of genome sequencing projects. Thus, genomics and its application to the production of wheat has lagged behind advances in other cereal crops, such as rice, sorghum, and maize. Today, wheat is the last major crop for which no genome sequencing effort is underway. Recently, however, technological advances offer the prospects of tractable large-scale programs that can deliver much-needed genomic resources for wheat.

In November 2003, a USDA–NSF-funded international workshop of wheat geneticists and sequencing specialists identified the first objectives towards sequencing the hexaploid wheat genome, i.e., physical mapping and assessment of sequencing strategies (Gill et al. 2004). To capitalize on the momentum of this workshop, the International Wheat Genome Sequencing Consortium (IWGSC, http://www.wheatgenome.org) was established in January 2005 with the goal of coördinating the international effort to build the foundation for and leading the effort to sequence the bread wheat genome.

As an international industry, academic, and governmental agency collaboration, the IWGSC is committed to providing wheat breeders and industry state-of-the-art tools and technologies that enable profitability throughout the wheat industry. The consortium is governed by a coördinating committee, comprised of scientific and financial contributors, who support sequencing the bread wheat genome, and an executive director (K. Eversole) supported by six co-chairs from Europe (C. Feuillet, France, and B. Keller, Switzerland), Australia (R. Appels), the USA (J. Dvorak and B. Gill), and Japan (Y. Ogihara). General membership in the consortium is open to anyone and business meetings are open to the public. Business meetings and workshops or coördinating committee meetings are held in conjunction with most major international plant genomics meetings.

To ensure the rapid delivery of tools to breeders, the IWGSC identifies short-term and long-term strategic goals, advocates for sequencing the wheat genome, coördinates international scientific efforts to build resources for wheat, develops and assists in the development of project proposals, and secures funding for collaborative efforts aimed at meeting identified goals. By implementing a milestone-based strategy, the consortium delivers products and tools while working towards the ultimate goal of a sequenced bread wheat genome. This overall strategy ensures the immediate availability of significant outputs for wheat breeders and the wheat industry at large in parallel to continued advancements in basic research on the wheat genome.

Projects coördinated and endorsed by the IWGSC fall within two broad categories: physical mapping (construction of physical maps for the D genome of *Ae. tauschii* and for the hexaploid wheat genome) and sequencing (the development of the resources necessary for sequencing and the testing of technologies to determine the best method for sequencing). The following provides an update of the IWGSC projects.

Physical mapping.

To provide the greatest resources to enhance wheat production and also advancing our basic understanding of the hexaploid wheat genome, the first priorities for the consortium are to establish a physical map of the 21 hexaploid wheat chromosomes and to complete the physical map of *Ae. tauschii*. This will facilitate the map-based isolation of the hundreds of genes and QTL for traits of agronomic importance as well as delivering ‘perfect’ markers for wheat breeding.

At the same time, the physical map will provide the substrate for sequencing the wheat genome regardless of the ultimate sequencing strategy selected.

In August 2006, the IWGSC road map for physical mapping projects was established and agreed upon by the IWGSC coördinating committee at the ITMI meeting in Victor Harbour, Australia. This road map includes the completion of a physical map of the D genome of the wild diploid *Ae. tauschii,* as a framework for the construction of the physical maps for the seven chromosomes of the D genome of hexaploid wheat. The IWGSC also aims to complete the maps of the homoeologous A and B chromosomes. The D-genome project was initiated six years ago (http://wheat.pw.usda.gov/PhysicalMapping/) and has established efficient protocols and software to perform BAC fingerprinting and contig assembly (Luo et al. 2003). These protocols are being used now for the hexaploid wheat genome project. Funding from the US National Science Foundation has been provided recently to complete the D-genome physical mapping project (PI, J. Dvorak, University of Californai, Davis, CA, USA).

The construction of physical maps in hexaploid wheat is performed with a chromosome-specific strategy that has been pioneered in Europe by the Institute of Experimental Botany in the Czech Republic and the INRA in France. This approach relies on the recent improvement of chromosome sorting and BAC library construction technologies that allowed the construction of chromosome-specific BAC libraries (Dolezel et al. 2007). The first BAC library already has been used successfully in a pilot project to establish a chromosome landing ready physical map of chromosome 3B, the largest wheat chromosome (3X the rice genome) (PI, C. Feuillet; INRA, France) and was published in 2008 (Paux et al. 2008).

The successful pilot project to develop a physical map on chromosome 3B of Chinese Spring by a single laboratory has opened up the route for the international collaborative effort on the 20 remaining chromosomes of hexaploid wheat. During the past two years, physical mapping projects for additional chromosomes have been initiated and physical mapping and sequencing project leaders have been secured for all of the bread wheat chromosomes. A summary of the status of the physical mapping projects follows and an illustration of the current status of the Chinese Spring physical mapping projects is provided in Fig. 1.

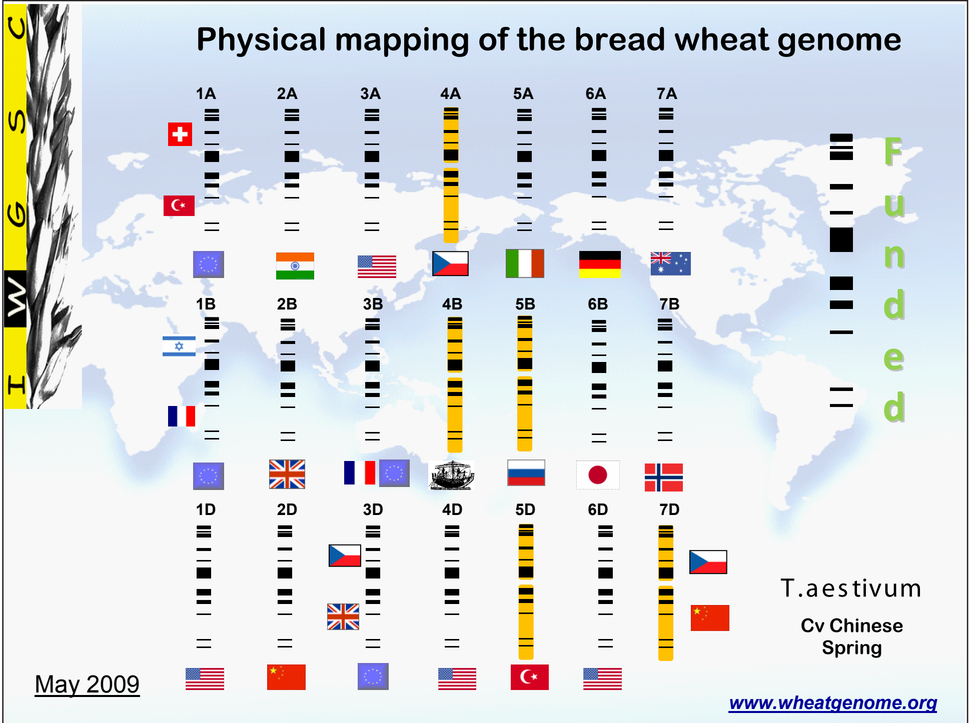

Fig. 1. Project leadership status for chromosome-based physical maps of bread wheat.

IWGSC physical mapping projects.

1. Completed projects

- 3B — Led by C. Feuillet (INRA, France), this has served as the pilot project for developing the first physical map of a flow-sorted chromosome. After a first 9.6X chromosome landing ready physical map (Paux et al. 2008), the project was completed in the framework of the EU FP7 project TriticeaeGenome to an 18X sequence-ready physical map of chromosome 3B of Chinese Spring. The 3B physical map can be found at http://urgi.versailles.inra.fr/gbrowse/cgi-bin/gbrowse/Wheat_FPC/.

2. Funded projects

- 1A, 1B, 3D, (3Bv2) — TriticeaeGenome project funded by the EU Commission under FP7 and coördinated by C. Feuillet (INRA, France) with 17 EU partners. This European project will complete the physical maps of group 1 and 3 chromosomes in wheat (and barley). Map-based cloning of targeted QTL, molecular breeding, and bioinformatics platforms will be developed within the framework of this project as well.
- 1, 4, and 6D of hexaploid wheat and all of *Ae. tauschii* — D-genome project led by J. Dvorak (UC, Davis), B. Gill (KSU), and O. Anderson (USDA–ARS). This project, funded by the US. National Science Foundation, will complete the *Ae. tauschii* physical map and physical maps of chromosomes 1, 4, and 6 of the D genome of Chinese Spring.
- 2AL — Led by K. Singh (Punjab Agricultural University, India). This project for the construction of the physical map of the long arm of chromosome 2A is funded by the Department of Biotechnology of the Indian Ministry of Science and Technology.
- 2D — Led by J. Jia (KL-CGB, CAAS, China) and funded by the CAAS (Chinese Academy of Agricultural Sciences). This project will develop a physical map of chromosome 2D.
- 3AS/3AL — Led by B. Gill (KSU, USA) and funded by the USDA (CSREE–NRI). These projects are developing anchored physical maps of the short and long arms of chromosome 3A.
- 4A – Led by J. Dolezel (Institute of Experimental Botany, Czech Republic). This project resulted in the construction of a BAC library and funding is being sought for physical mapping.
- 5A — Led by L. Cattivelli (Experimental Institute for Cereal Research, Italy) and funded by the Agricultural Research Council of Italy. This project will develop a physical map of chromosome 5A.
- 5B — Led by E. Salina (Institute of Cytology & Genetics, Russia). This project, developed in collaboration with European and U.S. partners, will construct the physical map of chromosome 5B of Chinese Spring. A 2- year grant was approved by the INRA–RFBR (Russian Foundation for Basic Research) for developing markers and genetic maps that will be needed for anchoring. Other grants will be submitted to funding agencies in 2009-2010 for the construction of the 5B BAC library and for fingerprinting.
- 7DS — Led by J. Dolezel (IEB, Olomouc, Czech Republic), this project to construct the physical map of chromosome 7DS has received funding for the construction of the BAC library and fingerprinting has been achieved. A proposal is pending for funds for the anchoring of the map.
- 7DL — This project is led by S. Weining (Northwest A&F University, Yangling, Shaanxi, China), R. Appels (Murdoch University, Perth, Australia), and J. Dolezel (IEB, Olomouc, Czech Republic). Funding for the BAC library was provided by the Czech Science Foundation. Northwest A&F University has provided funding for fingerprinting and anchoring for the completion of the 7DL physical map.
- Radiation hybrid mapping — Led by S. Kianian (North Dakota State University, Fargo, ND, USA) and funded by the U.S. National Science Foundation, this project will develop high-resolution, RH physical maps that will be used to anchor existing bacterial artificial chromosome (BAC) contigs and clones for the D-genome of hexaploid wheat. Other panels are under development for the other chromosomes such as for 3B (Paux et al, 2008).

3. Projects under development.

- 2B — Led by M. Bevan (JIC, UK), this project will develop the physical map of chromosome 2B of Chinese spring and will be submitted to funding agencies in late 2009 or in 2010.
- 4B — Led by M. Nachit (ICARDA, Syria) and D. Habash (Rothamstead Research, UK), the development of the physical map of chromosome 4B will be included in the proposed *4Phoenicia* project that will build scientific capabilities for the Mediterranean region to underpin wheat genomics for sustainable agriculture.
- 5D — Led by H. Budak (Sabanci University, Turkey), this project to establish a physical map of chromosome 5D of Chinese spring is under development and will be submitted to funding agencies in 2009.
- 6A — Led by T. Schnurbush (IPK Gatersleben, Germany), this project will construct the physical map of chromosome 6A.
- 6B — Led by Y. Ogihara (Kyoto University, Japan), this project will establish the physical map of chromosome 6A.
- 7A — Led by R. Appels (Murdoch University, Perth, Australia), this project will establish the physical map of chromosome 7A of Chinese Spring.
- 7B — Led by O. Olsen (Norwegian University of Life Sciences, Aas, Norway and Hedmark University College in Hamar, Norway). A proposal to construct the physical map of chromosome 7B is being finalized for submission to regional and national public/private entities.

Sequencing.

In the past few years, there has been an exponential growth in the number of sequencing strategies and the rhythm at which the technologies are evolving. For the first time, the possibility to sequence the hexaploid wheat genome at a reasonable cost appears to be within reach. To begin to assess how these new technologies can be applied efficiently to the wheat genome in a manner that will deliver high quality reference genome sequences that can be readily exploited by breeders to accelerate crop improvement, the IWGSC was joined by the International Barley Sequencing Consortium in hosting a workshop in September 2008 in Evry, France.

The primary workshop goal was to develop strategic road maps for sequencing the wheat and barley genomes in the next few years. The workshop brought together international experts in the human and agriculturally-important genome sequencing projects, developers of genome sequencing technologies, scientists with extensive knowledge of the structure and organization of the wheat and barley genomes, representatives of international genome sequencing centers interested in *de novo* sequencing of plant species, as well as representatives of governmental funding agencies. At the workshop, a consensus was reached on a two-phase sequencing strategy:

Phase 1: Obtaining a good quality sequence that can be used as soon as possible to develop tools for breeding and that represents a platform for phase 2. Establish pilot sequencing projects on chromosomes 3B of wheat to establish the most cost effective approaches for the wheat genome. Roche 454 Titanium and Illumina Solexa technologies should be tested separately and in combination on sorted chromosomes and on the minimal tiling paths. Furthermore, the potential utility of WGS paired-end datasets on the diverse next generation sequencing platforms in pilot projects should be explored. Such datasets are needed to train algorithms for Triticeae genome characteristics and advance the approach. At the same time, such data will deliver 'gene-catalog' sequence datasets that complement EST resources for marker and breeding-tool development.

Phase 2: Achieving high quality 'gold standard' sequences that will enable all functional and structural analyses of the two genomes.

The full workshop report can be found on the IWGSC website at (http://www.wheatgenome.org/documents).

A number of projects are currently underway to evaluate the feasibility of using new sequencing technologies to accelerate marker development and reduce MTP sequencing cost while maintaining quality and without losing access to the non genic but yet relevant sequences. The following provides a summary of specific IWGSC sequencing projects.

IWGSC sequencing projects.

1. *Chromosome sequencing. 3BSEQ* — Led by C. Feuillet (INRA, Clermont-Ferrand, France), this project entitled 'Sequencing, Annotation, and Characterization of the Bread Wheat Chromosome 3B' (3BSEQ) has been submitted to the French ANR Plant Biotech Flagship Project Call 2009. The 3BSeq project aims at obtaining an annotated sequence of the largest bread wheat chromosome, chromosome 3B, and at exploiting this knowledge to develop tiling arrays of the 3B gene space for further functional and structural characterizations. The project will take advantage of the potential offered by the next generation sequencing and array technologies to develop an original strategy and deliver a high quality draft sequence of the chromosome.

2. *Sequencing of megabase-sized contigs on chromosome 3B.*

- C. Feuillet, (INRA France): Two projects (supported by the ANR-Genoplante and the Genoscope) to sequence more than 20 Mb of BAC contigs distributed in different regions of chromosome 3B have been completed and a publication is forthcoming.
- R. Appels (Murdoch University, Australia): Sequencing of two Mb-sized contigs located on chromosome 3BS (supported by GRDC) has been completed, and a publication is forthcoming.

3. *3AS sample sequencing project* — Led by B. Gill (KSU, USA) and funded by the USDA (CSREES-NRI), this project will generate 18.4 Mb of sequence from the chromosome 3AS BAC libraries, sequence 48 targeted BAC clones, and BAC end sequence 10,000 random clones. The second sequencing component of this project is to compare the 3AS BAC sequences with sequences from homoeologous chromosome arm 3BS.

4. Mining the allohexaploid wheat genome for useful sequence polymorphisms — Led by K. Edwards (University of Bristol, UK) and funded by the Biotechnology and Biological Sciences Research Council (BBSRC), this project will use next generation sequencing to identify sequence differences in the genomes of five key cultivars and explore ways to sequence the wheat genome.

Bioinformatics.

Broad bioinformatics capabilities will be necessary to annotate the sequence and to ensure the greatest utility of data for breeders. To coördinate the manual and automated annotation of the wheat genome, the IWGSC established an annotation working group in 2006 that was expanded subsequently to include all of the Triticeae. The Group is chaired by P. Leroy (INRA, Clermont-Ferrand, France) and T. Wicker (University of Zurich, Switzerland).

During 2006 and 2007, the Triticeae Annotation Working Group developed annotation guidelines (http://www.wheatgenome.org/tools). In 2007, a semi-automated annotation pipeline, TriAnnot (http://urgi.versailles.inra.fr/projects/TriAnnot/), was developed and continues to be improved. The IWGSC coördinated a Bioinformatics Tool Workshop that was held in January 2009 in conjunction with the XVII Plant and Animal Genome conference in San Diego, California. The agenda and links to speaker abstracts can be found on the PAG website (http://www.intl-pag.org/17/17-iwgsc2.html). Under the coördination of D. Matthews (USDA–ARS, Cornell University, Ithaca, NY, USA), GrainGenes has established an IWGSC GrainGenes BAC repository for annotated genome sequences (http://www.tritgenome.org/tawg/). New sequencing technologies, full-length cDNA sequences, and deep transcriptome sequencing will be essential resources for annotation and the IWGSC supports projects to develop these resources.

The major challenge for all genome-sequencing projects is developing bioinformatics capabilities for exploitation of the sequence. During the IWGSC–IBSC workshop on sequencing technologies, it became clear that bioinformatics capacities needed to be developed well in advance of the availability of the sequence. The workshop participants agreed that we need to develop a centralized database that will extend the Ensembl platform to plants and that we had to develop a mechanism for obtaining ongoing feedback from the sequence users. The IWGSC goals for bioinformatics, thus, include the development of a publicly available, centralized, comprehensive database that integrates annotation and other available biological data, including comparative genomics, variation, and regulatory data.

Next steps and future meetings.

During the next year, the IWGSC will continue to seek funding for the development of physical maps and sequencing of all of the chromosomes of hexaploid wheat to ensure that the entire wheat industry can begin to exploit rapidly genomic information while efforts are underway to obtain a complete sequence of bread wheat. Concurrently, the IWGSC will continue to explore the feasibility of using the new sequencing technologies for the construction of physical maps and for sequencing the bread wheat genome and may organize another sequencing technologies workshop in 2010. The IWGSC will expand its efforts in bioinformatics to ensure the development of a centralized database and to coordinate the various bioinformatic tools and resources that will be established through the individual physical mapping and sequencing projects. The IWGSC is exploring the possibility of organizing a workshop to develop a road map for the establishment of bioinformatics capabilities in late 2009 or 2010.

The IWGSC will hold a coördinating committee meeting in conjunction with the ITMI meeting in Clermont-Ferrand, France, in August 2009. In January 2010, the IWGSC will host a workshop and an open business meeting at the Plant and Animal Genome Conference in San Diego, California and may organize another annotation workshop. Meeting information as well as general information will be available at the IWGSC website (http://www.wheatgenome.org).

Finally, as more and more resources for breeders are becoming available through the IWGSC projects, a concerted effort is underway now to increase significantly the active industry involvement in the consortium. Six breeding companies now sponsor the consortium in addition to the public and grower organizations (Fig. 2, p. 16). The IWGSC is working with industry to identify chromosomal regions of importance for early, targeted sequencing and deep sequencing. Additionally, the IWGSC encourages all physical mapping and sequencing projects to include isolation of genes and QTL underlying industry-identified key traits in wheat.

With the completion of the first physical map of a bread wheat chromosome, funding in place for physical maps of 10 other chromosomes and the entire *Ae. tauschii* genome, the submission of the first proposal to sequence a wheat chromosome, and the rapid development of new sequencing technologies, the IWGSC goal of obtaining a first draft sequence of the wheat genome is within reach.

Fig. 2. Sponsors of the IWGSC.

References.

Doležel J, Kubaláková M, Paux E, Bartoš J, and Feuillet C. 2007. Chromosome-based genomics in the cereals. Chromosome Res 15:51-66.

Gill BS, Appels R, Botha-Oberholster A-M, Buell CR, Bennetzen JL, Chalhoub B, Chumley F, Dvorák J, Iwanaga M, Keller B, Li W, McCombie WR, Ogihara Y, Quetier F, and Sasaki T. 2004. A workshop report on wheat genome sequencing: International genome research on wheat consortium. Genetics 168:1087-1096.

Luo M C, Thomas C, You F, Hsiao J, Ouyang S, Buell CR, Malandro M, McGuire PE, Anderson OD, and Dvorak J. 2003. High-throughput fingerprinting of bacterial artificial chromosomes using the snapshot labeling kit and sizing of restriction fragments by capillary electrophoresis. Genomics 82:378-389.

Paux E, Sourdille P, Salse J, Saintenac C, Choulet F, Leroy P, Korol A, Michalak M, Kianian S, Spielmeyer W, Lagudah E, Somers D, Kilian A, Alaux M, Vautrin S, Bergès H, Eversole K, Appels R, Safar J, Simkova H, Dolezel J, Bernard M, and Feuillet C. 2008. A physical map of the 1-gigabase bread wheat chromosome 3b. Science 322:101-104.