IWGSC: PHYSICAL MAPPING STANDARD PROTOCOLS WORKSHOP

Plant and Animal Genome Meeting, San Diego, CA, USA
Tuesday, 12 January, 2010.

Workshop report.

Rudi Appels, Eduard Akhunov, Michael Alaux, Mario Caccamo, Federica Cattonaro, Jaroslav Dolezel, David Edwards, Ming-Cheng Luo, Dave Matthews, Nicolas Guilhot, Etienne Paux, Thomas Wicker, Kellye Eversole, and Catherine Feuillet.

On 12 January, 2010, the International Wheat Genome Sequencing Consortium (IWGSC) organized a workshop to develop and discuss protocols and standards for the physical mapping of the hexaploid wheat genome and develop a consensus. In addition, the workshop surveyed the sequencing efforts undertaken within the consortium to coördinate the studies carried out in member laboratories. The goal was to ensure homogeneity in the procedures used for constructing the wheat physical maps by providing guidelines developed in expert laboratories and distributing these to the groups participating in the physical mapping and sequencing of bread wheat chromosomes under the auspice of the IWGSC.

The road map for achieving a high-quality reference sequence of the bread wheat genome established by the IWGSC includes, as a first step, the construction of physical maps in hexaploid wheat using a chromosome-specific strategy. This approach relies on recent improvements in chromosome sorting and BAC library construction technologies that have allowed the construction of chromosome-specific BAC libraries (Dolezel et al. 2007). The first physical map has been achieved for the largest wheat chromosome, 3B (1 Gb) (Paux et al. 2008; http://urgi.versailles.inra.fr/projects/Triticum/index.php) and its sequencing has been initiated this year in the framework of an ANR flagship project (3BSEQ). Physical mapping and sequencing project leaders have been secured for all of the bread wheat chromosomes (Fig. 1), and a number of projects are in the initial phases of fingerprinting and contig assembly (see http://wheat.pw.usda.gov/ggpages/awn/55/TEXTFILES/IWGSC_REPORT.pdf).

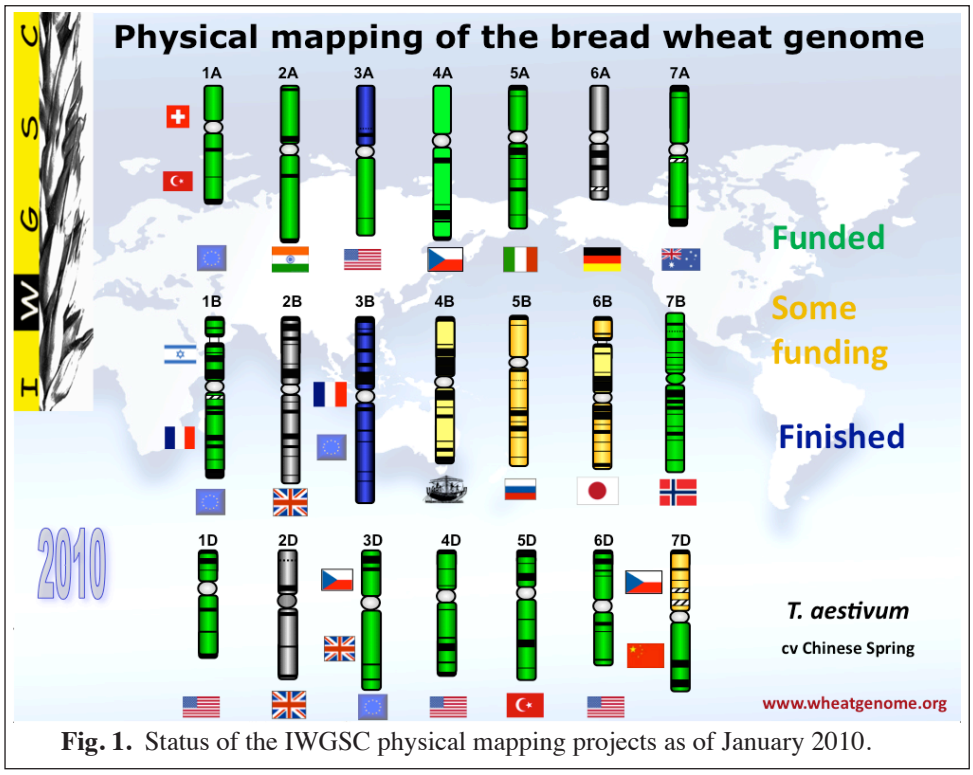

Fig. 1. Status of the IWGSC physical mapping projects as of January 2010.

The workshop was organized in five sessions of 45 minutes each that covered the following topics:
• Fingerprinting (Ming-Cheng Luo and Federica Cattonaro)

- Contig assembly (Etienne Paux and Ming-Cheng Luo)
- Anchoring (Eduard Akhunov and Etienne Paux)
- Sequencing with next generation technologies for marker development (Thomas Wicker and David Edwards)
- Databases and online tools for displaying the anchored physical maps (Michael Alaux and Dave Mathews)

In an introduction to the workshop, Jaroslav Dolezel (Institute of Experimental Botany (IEB), Czech Republic), who pioneered the chromosome sorting approach, presented the methods used to construct chromosome-specific BAC libraries and the effort required to produce high-quality DNA from sorted chromosomes. About 5 μg DNA, which corresponds to 5–10 million chromosomes, sorted during about 6–8 weeks of work by the team at IEB is needed to construct one library. Except for chromosome 3B that can be sorted directly from Chinese Spring, specific cytogenetic stocks (ditelosomic, telosomic lines) are used for sorting the different chromosome arms. About 10,000–20,000 seeds of good quality are needed to construct each BAC library. The contamination of sorted fractions by other chromosomes is generally low (5–15%) and does not interfere with the subsequent analyses as the contamination represents a mixture of unrelated chromosomes. To date, BAC libraries have been constructed and delivered to the lead laboratories for 12 chromosomes (1A, 2A, 3A, 4A, 5A, 1B, 3B, 1D, 3D, 4D, 6D, and 7D; see Table 1 for details) and another six will be completed this year. Recently, constructed libraries are cloned in phage-resistant bacteria, and their average insert size is about 120 Kb. All BAC libraries are constructed with the *Hin*dIII restriction site. Two low-coverage libraries (1, 4, 6D, and 3B) were constructed with the *Eco*RI cloning site as well. Finally, the demand is increasing for chromosomal DNA to perform whole-chromosome (arm), shotgun sequencing by next-generation technologies for chromosome composition surveys and marker development. This is performed now routinely in different projects with DNA amplified from flow-sorted chromosomes and chromosome arms. Typically, 50 seeds are needed to isolate sufficient numbers of chromosome arms to produce 3–5 μg DNA after whole-genome amplification with the GenomiPhi V2 DNA Amplification Kit (GE Healthcare). The available data indicate representative amplification with only several-fold quantitative differences in the rate of amplification of various genomic loci.

Table 1. List of chromosome-specific BAC libraries currently available.

Chromosome	Country sponsor/lead	Library name	Insert size	Coverage
1D, 4D, 6D	USA	TaaCsp146eA	110 kb	1.3 x
1D, 4D, 6D	CZ, FR	TaaCsp146hA	85 kb	3.4 x
1D, 4D, 6D	USA	TaaCsp146hB	102 kb	6.9 x
1D, 4D, 6D	USA	TaaCsp146hC	116 kb	7.4 x
1AL	EU/TR	TaaCsp1ALhA	103 kb	8.0 x
1AL	EU/TR	TaaCsp1ALhB	109 kb	7.7 x
1AS	EU/CH	TaaCsp1AShA	111 kb	11.8 x
1BL	EU/FR	TaaCsp1BLhA	114 kb	15.4 x
1BS	EU/IL	TaaCsp1BShA	113 kb	15.7 x
2AS	IND	TaaCsp2AShA	123 kb	15.4 x
3AL	USA	TaaCsp3ALhA	106 kb	10.2 x
3AL	USA	TaaCsp3ALhB	114 kb	5.2 x
3AS	USA	TaaCsp3AShA	80 kb	10.9 x
3AS	USA	TaaCsp3AShB	115 kb	15.9 x
3B	EU/FR	TaaCsp3BFeA	107 kb	1.9 x
3B	CZ/FR	TaaCsp3BFhA	103 kb	6.2 x
3B	EU/FR	TaaCsp3BFhB	126 kb	9.1 x
3DL	EU/UK	TaaCsp3DLhA	105 kb	12.2 x
3DS	EU/CZ	TaaCsp3DShA	110 kb	11.0 x
4AL	CZ	TaaCsp4ALhA	126 kb	17.4 x
4AS	CZ	TaaCsp4AShA	131 kb	16.6 x
5AL	I	TaaCsp5ALhA	123 kb	18.3 x
5AS	I	TaaCsp5AShA	120 kb	16.5 x
7DL	CZ/PRC	TaaCsp7DLhA	115 kb	14.8 x
7DS	CZ	TaaCsp7DShA	114 kb	12.2 x
1BS	USA	TaaPav1BShA	82 kb	14.5 x
3B	CZ/AUS	TaaHop3BFhA	78 kb	6.0 x

The first session, led by Ming-Cheng Luo (University of California, Davis, USA) and Frederica Cattonaro (IGA, Italy), was dedicated to presenting and discussing 'Fingerprinting protocols'. The labs of Ming-Cheng Luo and Jan Dvorak have pioneered the use of the SNaPshot kit to perform BAC fingerprinting (Luo et al., 2003) and have used it to develop a physical map of *Ae. tauschii* (http://wheatdb.ucdavis.edu), the ancestral D-genome donor of hexaploid wheat. This D-genome progenitor map will serve as a framework to support the assembly of the D-genome physical map of hexaploid wheat cultivar Chinese Spring. The SNaPshot protocol is the basis for all physical mapping projects in wheat as well as in other species, including *Brachypodium distachyon*, soybean, cowpea, cassava, walnut, banana,

and citrus. The laboratory of Federica Cattonaro has been involved in physical mapping several plant species, including grape wine, as a basis for genome sequencing. To date, the Cattonaro laboratory has fingerprinted all of the wheat BAC libraries (1AS, 1AL, 1BS, 1BL, 3DS, 3DL, and 3Bv2) of the European project TriticeaeGenome (www.trticeae-geome.eu). The Luo and Cattonaro laboratories presented their working protocols and latest improvements in terms of DNA isolation, SNaPshot reaction, and electrophoresis on capillary sequencers. The main differences between the two protocols concerned the type of size marker that was used (GS500Liz vs. GS1200Liz), and a consensus was reached on the fact that the use of the GS1200Liz is recommended to gain more fingerprint information. The main message from the two specialists was that although fingerprinting a few clones is not so complicated both theoretically and technically, it is one of the most complex tasks when fingerprints have to be generated over months or years and for these fingerprints to remain comparable in a large mapping and sequencing project.

A number of steps are critical for producing high quality fingerprints:

- the bacterial growth conditions need to be carefully determined before starting the production phase;
- it is recommended that the DNA preparations should not be done all at the same time before starting with fingerprinting because the conditions are not always stable and, sometimes, need to be re-optimized during the production phase;
- it is critical to use fresh β-mercaptoethanol and BSA;
- high-fidelity restriction enzymes and a good control of the digestion conditions (temperature, buffer) need to be used to avoid star activity (reaction volume should not be too small); and
- the running conditions for electrophoretic analyses need to be optimized.

Thus, to ensure the construction of high-quality and homogenous physical maps within the IWGSC, the consortium requires that the labs involved in the physical maps project use one of the two protocols developed by Ming-Cheng Luo and Federica Cattonaro. These protocols are available on the IWGSC website at http://www.wheatgenome.org/News-and-Reports/General-reports/Physical-mapping-standard-protocol-workshop.

The second session was led by Etienne Paux (INRA, France) and Ming-Cheng Luo (UC Davis, USA) and concerned the 'assembly of the physical contigs' from the fingerprints. Both experts presented the critical parameters that need to be taken into account to ensure reliable and robust assemblies. Paux presented a detailed protocol for contig assembly using FPC that was developed by the European project TriticeaeGenome. The protocol provides a step-by-step description of the assembly starting from the BAC naming convention to be adopted for the IWGSC projects to the MTP selection. The protocol includes the recently developed FPB software that permits the elimination of spurious background noise (Scalabrin et al. 2009; available at http://www.appliedgenomics.org/tools.php). Some of the critical parameters and processes were discussed, and there was agreement on the fact that the assembly needs to be performed first with high stringency, which reduces the risk of spurious assemblies due to the 10–20% of bands that are shared randomly between BACs, reflecting the repetitive nature of the wheat genome composition. Luo mentioned that some of the parameters need to be adjusted accordingly, especially the cut-off value (Sulston Score), which varies with the number of clones in assembly, the average number of fragments, fragment sizing range, and the value of tolerance (instrument sizing precision). A general consensus was established to use the guideline established by the TriticeaeGenome project and implement it with parameter settings that can be applied depending on the size of the project. The guideline is available through the IWGSC website at http://www.wheatgenome.org/News-and-Reports/General-reports/Physical-mapping-standard-protocol-workshop.

Finally, Paux mentioned a new algorithm called LTC (for Linear Topography Contig) that has been developed by A. Korol's group at the Institute of Evolution in Haifa and tested on the 3B physical contigs. The first results are very encouraging and indicate that LTC improves the quality of the assembly by identifying and resolving nonlinear topological structures. LTC enables the construction of highly reliable and longer contigs, the detection of 'weak' connections in contigs, and their 'repair', as well as the elongation of contigs obtained by other assembly methods such as FPC. A publication is underway, and A. Korol is willing to help colleagues use LTC for their physical assembly.

The third session, led by Etienne Paux (INRA, France) and Ed Akhunov (Kansas State University, Manhattan, USA), was dedicated to the different options available for 'anchoring the physical maps to the genetic maps'. Paux presented different strategies for pooling BAC libraries and MTPs that need to be screened with different markers. He then described the three approaches that can be used to anchor the physical contigs to the genetic maps: (1) by using markers (SSR, DArT, EST, and SNPs) already localized on genetics maps (forward approach) to screen the BAC contigs (pools) by PCR (individual PCR or through the hybridization of BAC pools on marker arrays); (2) by developing markers from

the BAC contigs (through BAC end sequencing) and mapping those on genetic maps (reverse approach); or (3) by developing markers from whole-chromosome or genome shotgun sequencing) to map them on genetic maps and screen the BAC pools (hybrid approach). To support these approaches, a number of scripts have been developed by INRA Clermont-Ferrand to define SSR and ISBP markers from 454 sequences obtained from sorted chromosome sequences or BAC end sequences (Paux et al. 2010). These are available publicly upon request to Frederic Choulet (fchoulet@clermont.inra.fr). One of the main conclusions from the experience of the initial wheat projects is that there is little overlap between the different type of markers used so far (EST, SSR, and ISBPs) and that it is better to use a diverse set of markers to ensure that most of the contigs will be anchored by at least one marker and that all regions of the chromosome are covered. The high-throughput platforms such as the Illumina bead express for SNPs or EST arrays that are emerging for marker genotyping in wheat should be used to improve the cost efficiency of the process. New platforms are currently under development (e.g., ISBPs on chip) and will continue to improve the efficiency of the anchoring. Akhunov presented the strategy planned for chromosome 3A that uses the sequences generated by 454 sequencing of BAC pools representing the chromosome 3A minimum tiling path (MTP) to develop molecular markers for anchoring. The major type of markers to be used for anchoring 3A is SNPs due to their amenability to high-throughput detection using various genotyping technologies. Development of SNP markers will be performed by targeted resequencing of low-copy genomic regions in the parents of mapping populations. The products of selective amplification generated using the PCR primers designed on the basis of MTP sequences will be pooled for 454 sequencing and variant discovery. Polymorphic sites and their flanking sequences will be submitted to Illumina for the design of 1536-plex Oligo Pool Assay (OPA). The data generated by genotyping of mapping populations on BeadArray platform using Illumina's GoldenGate assay will be utilized for the anchoring of chromosome 3A MTP.

A discussion followed on the genetic material that can be used for anchoring in wheat. Recombination mapping will be very useful to support map based cloning projects but will remain limited because 50% of the chromosomal regions do not show high levels of recombination. A number of populations are now available for anchoring including a large RIL of 2,000 individuals from a cross between Chinese Spring and Renan that will be publicly available through the TriticeaeGenome project in 2011 and a new large (1,600 RILs) ITMI population that is developed by M. Sorrels and C. Qualset. In addition, deletion mapping with the series of deletion lines that are available for the 21 wheat chromosomes is recommended and these can be ordered by contacting Bikram Gill at Kansas State University (bsgill@k-state.edu). Even if the resolution obtained with these lines is rather low due to the large size of the deletion bins, the lines provide valuable information of the location of centromeric and pericentromeric contigs that cannot be obtained by meiotic mapping. To increase the mapping resolution with a technique that does not rely on recombination and polymorphism, S. Kianian (North Dakota State University, Fargo, USA) has developed radiation hybrids in wheat (Hossain et al. 2004). A first panel has been developed for chromosome 3B (Paux et al. 2008) and the results indicated that a resolution of about 263 kb can be achieved with RH panels. RH populations will be developed now for the D-genome physical mapping project and other chromosomes are planned for the IWGSC. The main limitation at the moment is the greenhouse space required to develop the first generation of RH plants.

Thomas Wicker (University of Zurich, Switzerland) and David Edwards (University of Queensland, Australian Centre for Plant Functional Genomics, Australia) chaired the fourth session on sequencing with next generation technologies for marker development'. They first presented a survey of the ongoing efforts at the international level to produce sequence from wheat using next generation techniques. A number of projects are underway to either produce sequence surveys from whole genomes (A, D, and ABD) or from individual chromosomes and chromosome arms (Table 2). Wicker and Edwards will update the information regularly and request that the IWGSC members keep them informed about new projects to avoid redundancy and increase coordination and collaborations within the consortium. It is also im-

Table 2. Projects underway to produce sequence surveys from whole genomes or individual chromosomes and chromosome arms.

Country	Target	Coverage	Technique
United Kingdom	WGS	5x	454 Titanium
United Kingdom	3DL	75x	GAII PE
TriticeaeGenome	Group 1	1.5x	454 Titanium
France	3B	2x	GAII
Australia	WGS	0.2x	GAII PE
Australia	7DS	16x	GAII PE
United States	3A	2x	454 Titanium
Italy	5A	2x	454 Titanium
Switzerland	A/B/D ancestors and R(ye)	0.1x each	454 Titanium
China	D (*Ae. tauschii)*	40x	GAII PE
China	A (*T. urartu*)	40x	GAII PE

portant to gain more details about the methods that have been used for library production (Native or amplified DNA, Cp DNA contamination, WGS DNA library preparation, PE library features, etc…) as well as the quality controls that were applied on the sequence. This is essential to evaluate potential bias and assess the quality of the data produced. A 'WIKI type' platform will be established by Edwards and Wicker the IWGSC website to store and access all information related to the sequence data produced and the contact persons within the consortium. Edwards also presented the visualization tools he has developed for displaying short reads and using them for SNP discovery (see PowerPoint presentation online). He is happy to distribute them upon request (dave.edwards@uq.edu.au). Edwards also hosts a web tool, TAGdb, for searching paired read information for targeted assembly and cloning in wheat, which is available at http://flora.acpfg.com.au/tagdb.

The last session was dedicated to 'databases and online tools for displaying the anchored physical maps' and it was led by Michael Alaux (INRA URGI, France) and Dave Mathews (Cornell University, Ithaca, NY, USA). Mathews presented the different tools and resources available at GrainGenes (http://graingenes.org), where the physical maps and the contigs are displayed using a CMap interface. Links to the deletion bins are lacking in this representation, but this can be done in collaboration with Alaux and the URGI who have developed a Gbrowse interface that can display the bins along the chromosomes (http://urgi.versailles.inra.fr/projects/Triticum/deletionBin.php). The URGI-Gbrowse interface has a clickable link to the BAC contigs assigned to each bin (http://urgi.versailles.inra.fr/cgi-bin/gbrowse/wheat_FPC_pub/) and the subsequent information about each BAC contig (markers, sequence etc…). Mario Caccamo presented the PGP-viewer, a set of software tools, databases and interfaces developed to assist the work of the pig and zebrafish genome projects at the Sanger Institute (UK). The system integrates, in one platform, data from different sources including whole-genome assemblies, genetic markers, and expression information. The PGP-viewer interface extends the Ensembl browser with customized tracks that use a colour-based schema to distinguish robust versus more unstable sequences. The information is stored in an underlying database that can be used to access the data programmatically and also is attached to other interfaces, such as GBrowse. Caccamo mentioned that BACs excluded as singletons during the FPC assembly could be recovered at the sequence stage by using information provided by the alignment of BAC ends. The same information was used to guide the selection of BACs to close or extend over gaps in the underlying physical maps. One of the most valuable tools was a 'punch list' that provided an interface to the users for getting requests and tracking comments from the community. These comments were taken into account by the curators to improve the versions of the sequence release. It was decided that such a 'punch list' should be put in place for the IWGSC. Finally, Caccamo mentioned the genome reference consortium (http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/) that was established to curate the human and mouse genomes and can be followed as an example for the wheat sequencing effort.

The PowerPoint slides that were presented by the different speakers are available as PDF files on the IWGSC website at http://www.wheatgenome.org/News-and-Reports/General-reports/Physical-mapping-standard-protocol-workshop.

References.

Dolezel J, Kubalakova M, Paux E, Bartos J, and Feuillet C. 2007. Chromosome-based genomics in the cereals. Chromosome Res 15:51-66.

Hossain KG, Riera-Lizarazu O, Kalavacharla V, Vales MI, Maan SS, and Kianian SF. 2004. Radiation hybrid mapping of the species cytoplasm-specific (scs(ae)) gene in wheat. Genetics 168:415-423.

Luo MC, Thomas C, You FM, 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, Faure S, Choulet F, Roger D, Gauthier V, Martinant JP, Sourdille P, Balfourier F, Le Paslier MC, Chauveau A, Cakir M, Gandon B, and Feuillet C. 2010. Insertion site-based polymorphism markers open new perspectives for genome saturation and marker-assisted selection in wheat. Plant Biotechnol J 8:196-210.

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, Berges 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.

Scalabrin S, Morgante M, and Policriti A. 2009. Automated fingerprint background removal: FPB. BMC Bioinformatics 10:127.