DOI: 10.1534/genetics.104.034769

Workshop Report

A Workshop Report on Wheat Genome Sequencing: International Genome Research on Wheat Consortium

Bikram S. Gill,*[,1] Rudi Appels,[†] Anna-Maria Botha-Oberholster,[‡] C. Robin Buell,[§] Jeffrey L. Bennetzen,[¶] Boulos Chalhoub,[&] Forrest Chumley,[**] Jan Dvořák,[††] Masaru Iwanaga,[‡‡] Beat Keller,[§§] Wanlong Li,* W. Richard McCombie,[¶¶] Yasunari Ogihara,[&&] Francis Quetier[***] and Takuji Sasaki[†††]

*Plant Pathology Department, Kansas State University, Manhattan, Kansas 66506-5502, [†]Department of Agriculture, Murdoch University, 6983 Perth, Western Australia, Australia, [‡]Department of Genetics, Forestry and Agricultural Biotechnology Institute, University of Pretoria, ZA-0002 Pretoria, South Africa, [§]The Institute for Genomic Research, Rockville, Maryland 20850, [¶]Department of Genetics, University of Georgia, Athens, Georgia 30602, [&]Unité de Recherches en Génomique Végétale , Evry, 91157 Cedex, France, **Kansas Agricultural Experiment Station, Kansas State University, Manhattan, Kansas 66506, [††]Department of Agronomy and Range Science, University of California, Davis, California 95616-8515, [‡‡]International Maize and Wheat Improvement Center, Lisboa 27, 06600 Mexico, D.F., Mexico, [§§]Institute of Plant Biology, University of Zürich, CH-8008 Zürich, Switzerland, [¶¶]Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, [&&]Kihara Institute for Biological Sciences, Yokohama City University, Yokohama 244-0813, Japan, ***GENOSCOPE, Evry, 91057 Cedex, France and [†††]Genome Research Department, National Institute of Agrobiological Sciences, 305-8602 Tsukuba, Ibaraki, Japan*

ABSTRACT

Sponsored by the National Science Foundation and the U.S. Department of Agriculture, a wheat genome sequencing workshop was held November 10–11, 2003, in Washington, DC. It brought together 63 scientists of diverse research interests and institutions, including 45 from the United States and 18 from a dozen foreign countries (see list of participants at http://www.ksu.edu/igrow). The objectives of the workshop were to discuss the status of wheat genomics, obtain feedback from ongoing genome sequencing projects, and develop strategies for sequencing the wheat genome. The purpose of this report is to convey the information discussed at the workshop and provide the basis for an ongoing dialogue, bringing forth comments and suggestions from the genetics community.

WHEAT AS AN IMPORTANT CROP SPECIES

Wheat was the first domesticated crop and is the youngest polyploid species among the agricultural crops (see Figures 1–3 for background information). Together with rice and maize, wheat provides >60% of the calories and proteins for our daily life. Wheat is best adapted to temperate regions, unlike rice and maize, which prefer tropical environments. Wheat occupies 17% of all crop area (in 2002, 210 million hectares *vs.* 147 million for rice and 139 million for maize). The trade value of wheat exceeds that of any other cereal species, including rice and maize: $31 billion of world trade in 2001 *vs.* $13 and $19 billion for rice and maize (FAOstat database: http://apps.fao.org/default.jsp). To meet human needs by 2050, grain production must increase at an annual rate of 2% on an area of land that will not increase much beyond the present level. Significant advances in the understanding of the wheat plant and grain biology must be achieved to increase absolute yields and protect the crop from an estimated average annual loss of 25% caused by biotic (pests) and abiotic stresses (heat, frost, drought, and salinity). Genome sequencing is a widely accepted mechanism for accelerating achievement of these objectives, because it leverages similar work from other crops and plants and enables more rapid genetic improvement. In addition to food security, wheat genome sequencing will lead to improved human health and nutrition.

CEREAL GENOME STRUCTURE

Rice, maize, and wheat, which coevolved from a common ancestor ~55–75 million years ago (Kellogg 2001;

[1]*Corresponding author:* Plant Pathology Department, Throckmorton Hall, Kansas State University, Manhattan, KS 66506-5502. E-mail: bsgill@ksu.edu

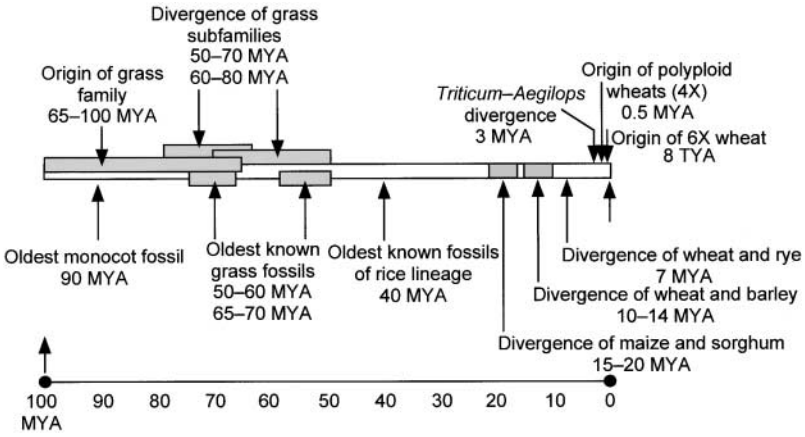

FIGURE 1.—Grasses as a single genetic system and recent coevolutionary history of cereals and humans. Grasses originated 55–75 million years ago and now dominate 20% of the land area. The three major cereals (rice, maize, and wheat), which diverged from a common ancestor ~40 million years ago, provide most of the food for humans. Humans and wheat share a remarkably parallel evolutionary history. About 3 million years ago, humans diverged from apes, and diploid A, B, and D progenitor species of wheat diverged from a common ancestor. About 200,000 years ago, at nearly the same time that modern humans originated in Africa, two diploid grass species hybridized to form polyploid wheat in the Middle East. Humans domesticated wheat ~15,000 years ago in the fertile crescent (modern-day Iraq and parts of Turkey, Syria, and Iran), marking the dawn of modern civilization. Comparative genetic and genomic studies during the last 10 years revealed extensive synteny among major cereals, and the concept of grasses as a single genetic system emerged. The genome sequence of rice (420 Mb) is nearly completed, and it will serve as the anchor genome to promote gene discovery in all cereals. However, recent data suggest that most domestication-driven, agronomic, and end-use traits, as well as those genes involved in landmark speciation events such as polyploidy, are crop and species specific. The emerging view is that DNA sequence information of all key species is essential for investigating grasses as a single genetic system. Maize (2500 Mb) genome sequencing is underway, and wheat (16,000 Mb) genome sequencing was discussed at the workshop. Figure courtesy of W. J. Raupp, based on discussions with P. F. Byrne, Colorado State University, Fort Collins, with additional data from HUANG *et al.* (2002).

Figure 1), differ greatly in genome size. Among agricultural crops, common bread or hexaploid wheat (*Triticum aestivum* L., $2n = 6x = 42$, AABBDD) has the largest genome at 16,000 Mb, ~8-fold larger than that of maize and 40-fold larger than that of rice (ARUMUGANATHAN and EARLE 1991). Amplification of transposable elements (TEs), coupled with duplication of chromosome segments, was a major driving force for cereal genome expansion, although polyploidization also contributed to the large genome size of wheat. The rice genome at 430 Mb consists of at least 22% TEs (MA *et al.* 2004). The maize genome at ~2500 Mb consists of >50% TEs (MEYERS *et al.* 2001, WHITELAW *et al.* 2003). About 90% of the wheat genome consists of repeated sequences and 70% of known TEs (LI *et al.* 2004). Low-copy TEs and miniature inverted repeat TEs are most often associated with active genes, but high-copy TEs mainly insert in the intergenic space (SANMIGUEL *et al.* 1996, 2002). Gene distribution along chromosomes is relatively homogeneous in the small genome of rice, but gene clusters (gene-rich regions) are separated by long stretches of TEs (gene-poor or gene-free regions) in the wheat genome, as demonstrated by deletion mapping (GILL *et al.* 1996a,b; FARIS *et al.* 2000) and BAC-based physical mapping (J. DVOŘÁK, unpublished results). Within some gene-rich regions of the wheat genome, gene density is similar to that of smaller genomes (FEUILLET and KELLER 1999).

Comparative mapping of cereal genomes using a standard set of probes showed extensive conservation in gene content and order at a low-resolution genetic map level. It seems a logical choice to use the small genome of rice as a surrogate for positional cloning of agriculturally important genes from large cereal genomes based on microcolinearity. However, small translocations, deletions, inversions, and duplications often violate microcolinearity and complicate this process (TIKHONOV *et al.*, 1999; KELLER and FEUILLET 2000; DUBCOVSKY *et al.* 2001; LI and GILL 2002; SORRELLS *et al.* 2003). Most importantly, there is little colinearity for disease resistance genes due to their rapid evolution among grasses (LEISTER *et al.* 1998). Sequence comparisons between rice and the wheat genomic regions that harbor the genes *Lr10*, *Pm3* (GUYOT *et al.* 2004), *Q*, and *Tsn1* (J. D. FARIS, J. P. FELLERS, H. LU, K. M. HAEN and B. S. GILL, unpublished results) detected extensive microrearrangements or complete loss of microcolinearity. All this suggests that rice will often not be a good model for positional cloning in wheat. However, sequencing the wheat genome will advance comparative studies of grass genomes and lead to better understanding of the relationship among grass lineages (FREELING 2001).

Comparative analysis of orthologous sequences from two related species has been a powerful tool for *de novo* prediction of genes and identification of noncoding functional elements on the basis of the assumption that sequences conserved among species that have diverged for many millions of years must have functional roles in their genomes. Comparison of human and mouse genome sequences increased the specificity of genome annotation in both species (MOUSE GENOME SEQUENCING CONSORTIUM 2002). Alignment of sequences orthologous to a 1.8-Mb region of human chromosome 7q31, from several animal genomes, detected numerous functional elements such as transcription-factor-binding sites and noncoding RNA transcripts (MARGULIES *et al.*

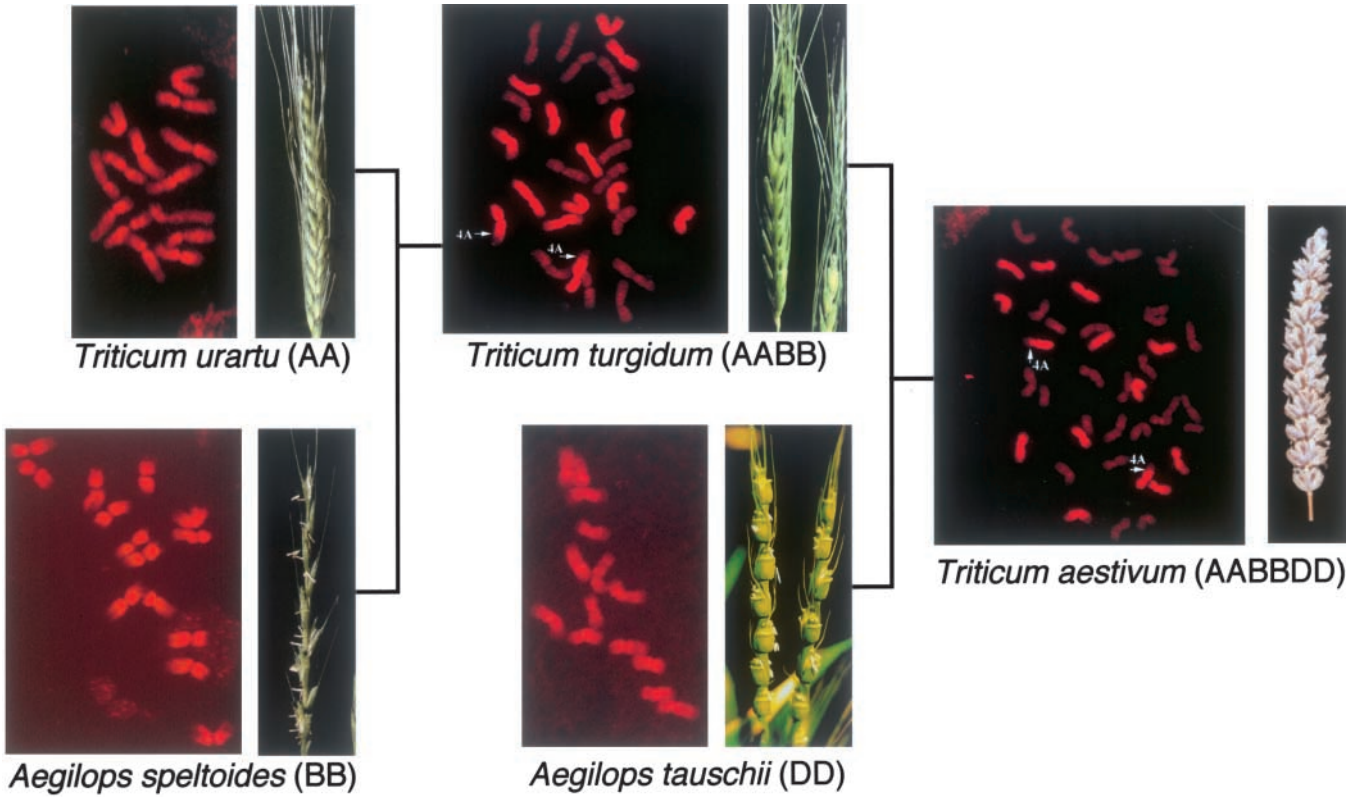

Figure 2.—By 1915, botanists had described three classes of cultivated wheats; the one-seeded monococcum (2*x*), the two-seeded emmer (4*x*), and the dinkel (6*x*). The one-seeded wild relative of monococcum was reported in Greece and Anatolia between 1834 and 1884. The two-seeded wild relative of emmer was discovered by Aaronsohn in 1910 in Lebanon, Syria, Jordan, and Israel. Therefore, it was well accepted, as Candolle had suggested in 1886, that since wild wheats grow in the Euphrates basin, wheat cultivation must have originated there. Between 1918 and 1925, T. Sakamura and his student H. Kihara at Hokkaido University in Japan and K. Sax at Harvard University reported their classic studies on the genetic architecture of the three groups of wheats. They analyzed meiosis in wheat species and hybrids and were the first to establish the basic chromosome number of seven and document polyploidy in the wheat group. This was an exciting observation and established polyploidy as a major macrospeciation process and wheat as a polyploid genetic model. This method of delineating species evolutionary relationships on the basis of chromosome-pairing affinities in interspecific hybrids came to be called the genome-analyzer method. These hybrids of course could also be exploited in plant breeding for interspecific gene transfers as well as in an approach of shuttle genetics where, for example, genetic mapping can be done in diploid wheat and aneuploid mapping in polyploid wheat. Chromosome figures from Zhang (2002).

2003). Similarly, a significant portion of rice genes were identified on the basis of their similarities with Arabidopsis genes at the protein sequence level. Comparison of genomic sequences of 52 orthologous genes of rice and maize indicated that most genes contain conserved noncoding sequences (CNSs) and that upstream regulatory genes tend to be enriched in CNSs (Inada *et al.* 2003). From this point of view, sequencing the wheat genome will facilitate annotation of all plant genomes, especially grass genomes.

CURRENT UNDERSTANDING OF THE WHEAT GENOME AND ITS SEQUENCE

Common wheat is an allohexaploid consisting of seven groups of chromosomes, each group containing a set of three homeologous chromosomes belonging to the A, B, and D genomes, derived from a common ancestor (Figures 2 and 3). Despite their close homology, homeologs are normally prevented from pairing by the *Ph1* gene on the long arm of chromosome 5B. Thus, common wheat functions much like a diploid organism, although it is able to tolerate aneuploidy due to the buffering effect of polyploidy. Sets of viable mono-, tri-, and tetrasomic cytogenetic stocks were developed for all chromosomes, and nullisomics were developed for 11 chromosomes (Sears 1954). Since the loss of a pair of chromosomes can be compensated by two additional doses of a homeolog, 42 compensating nulli-tetrasomics were developed (Sears 1966). The monosomic chromosomes tend to misdivide and this property was exploited to produce a series of chromosome-arm aneuploids: monotelosomics, ditelosomics, tritelosomics, and iso-chromosome lines (Sears and Sears 1978). More recently, taking advantage

of the gametocidal chromosome introduced from the *Aegilops cylindrica* host, Endo and Gill (1996) developed 436 segmental deletion lines in Chinese Spring (CS). All of these genetic stocks, which are in the CS background, have been used to localize genes or markers to a specific chromosome, chromosome arm, or subarm region and play a central role in wheat genetics and genomics.

The 21 wheat chromosomes can be readily identified by heterochromatic banding (Gill *et al.* 1991; see Figure 3) or *in situ* hybridization patterns using repetitive DNA probes (Pedersen and Langridge 1997). A specific chromosome or chromosome arm can be flow sorted at high purity using the genetic stocks (Vrána *et al.* 2000). These sorted chromosomes have been used for construction of chromosome-specific BAC libraries (Safar *et al.* 2004), together with other genetic and molecular resources for wheat genome sequencing (summarized in Table 1).

Genes and recombination events are not randomly distributed along wheat chromosomes. They are clustered in the distal regions, while proximal regions are largely gene poor or gene free (Werner *et al.* 1992; Gill *et al.* 1993, 1996a,b; Kota *et al.* 1993; Delaney *et al.* 1995a,b; Mickelson-Young *et al.* 1995; Faris *et al.* 2000; Weng *et al.* 2000; Qi *et al.* 2003). A detailed study of the short arm of the group 1 chromosomes demonstrated that 70% of the genes and 82% of the total recombination distance were contained within two major gene-rich regions (1S0.8 and 1S0.5) that physically encompass only 14% of the arm (Sandhu *et al.* 2001). This picture of uneven wheat gene distribution has been strongly supported by the recent assignment of nearly 6000 wheat expressed sequence tags (ESTs) to 159 deletion bins across the 21 chromosomes (Qi *et al.* 2004). Associated with recombination, gene duplications also show similar distribution patterns along the wheat chromosomes (Akhunov *et al.* 2003).

Currently, ESTs (~500,000 to date) are the largest sequence resource for wheat. ESTs are cDNA clones, and as such they do not contain promoters, introns, and other functional elements. The gene coverage of human ESTs is 75%; of mouse, 56% (Mouse Genome Sequencing Consortium 2002); of Arabidopsis, 60% (Arabidopsis Genome Initiative 2000); of tomato, 47% (Van der Hoeven *et al.* 2002); and of rice, 36% (Feng *et al.* 2002). It is estimated that the gene coverage of the wheat EST collection is ~60%, close to that of Arabidopsis (Li *et al.* 2004), indicating that ~40% of wheat genes are not represented in EST collections.

In addition to the more than half million ESTs, ~6 Mb of wheat genomic DNA has been sequenced, including ~3 Mb from a random shotgun genomic library and ~3 Mb from large-insert genomic clones. BAC clones selected as hybridizing with specific genes revealed that gene density varies greatly, ranging from 1 (Faris *et al.* 2003) to 16 genes/BAC (Brooks *et al.* 2002), and that genes tend to be clustered into gene islands (Wicker *et al.* 2001; Brooks *et al.* 2002; SanMiguel *et al.* 2002). Feuillet and Keller (1999) reported a gene density

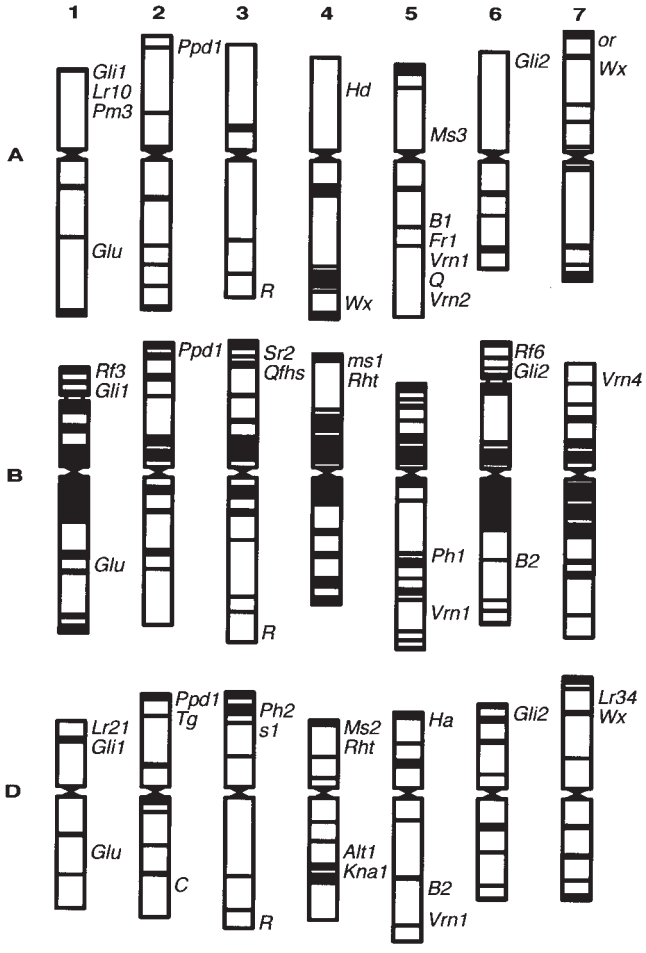

Figure 3.—Chromosome bin mapping of wheat genes that founded human civilization. In the 1950s, E. R. Sears developed aneuploid stocks and identified and mapped many of the unique genes that make wheat humankind's most important crop plant. These included gluten genes (*Gli* and *Glu*) that impart bread-making qualities; grain color (*R*), texture (*Ha*), and starch compostion (*Wx*) genes that greatly impact different market classes; photoperiod (*Ppd*) and vernalization (V*rn*) genes that make wheat the most widely adapted plant; durable disease resistance genes (*Sr2* and *Lr34*) and QTL (Qfhs); domestication genes (*Tg, Q, s1, C, Rht, Hd, B1, B2*); male sterility (*ms, Ms*), and restorer genes (*Rf*); and genes regulating polyploid meiosis (*Ph1* and *Ph2*) and abiotic stress (*Alt1, Fr, Kna, or*; for more details, check out the Catalog of Gene Symbols for Wheat: http:// wheat.pw.usda.gov/ggpages/wgc/98/). In the 1970s, euchromatic and heterochromatic regions were identified in the wheat genome, and a standard karyotype was developed. Deletion stocks were developed in the 1990s. Over 10,000 unigenes, including most of the abovementioned genes and hundreds of QTL, were mapped in chromosome bins (http://wheat.pw.usda.gov/index.shtml). Because of the abundance of genetic resources, several genes unique to wheat such as *Lr21* (Huang *et al.* 2003), *Lr10* (Feuillet *et al.* 2003), *Pm3* (Yahiaoui *et al.* 2004), *Vrn* (Yan *et al.* 2003, 2004), and *Q* (Faris *et al.* 2003) have been isolated by map-based cloning. Wheat karyotype after Gill *et al.* (1991).

TABLE 1

Genetic and molecular resources that can be applied to wheat genome sequencing analysis

Target	Genetic maps	Physical genome resources
Hexaploid wheat, *T. aestivum* L. cv. Chinese Spring	More than 4700 markers have been placed into 20 genetic maps of wheat and its relatives, and all the mapping data are deposited in the GrainGenes database (http://wheat.pw.usda.gov/ggpages/mapsframe.html)	A total of 500,000 ESTs are recorded and ~6000 unigenes are assigned to 159 bins across the 21 wheat chromosomes on the basis of deletion and ditelosomic stocks (http://wheat.pw.usda.gov/nsf)
Genetic stocks in Chinese Spring wheat, their chromosome constitution and number of lines available:	On the basis of these maps, a composite linkage map has been compiled. Markers on the map are linked electronically to the Catalog of Gene Symbols for Wheat (http://wheat.pw.usda.gov/ggpages/wgc)	A total of 200,000 BAC clones are arranged as contigs across the entire D-genome donor of hexaploid wheat. More than 1000 ESTs are assigned to BAC clones (http://wheatDB.ucdavis.edu)
Ditelosomic (20″ + t″), 21	Individual genetic maps are available in cMAP format through an internet interface	276,480 BAC clones of *T. monococcum* (Lijavetzky *et al.* 1999)
Double-ditelosomic (20″ + tS″ + tL″), 20		516,096 BAC clones of durum wheat (Cenci *et al.* 2003)
Ditelo-monotelosomic (20″ + t″ + t'), 41		1,200,000 BAC clones from cv. Chinese Spring (Allouis *et al.* 2003)
Double monosomic (19″ + 1′ + 1′), 20		87,168 BAC clones available from 3B; 87,168 BAC clones from 1D, 4D + 6D combined, and 65,280 BAC clones from 1BS (Safar *et al.* 2004)
Monosomics (20″ + 1″), 21	.	1,000,000 BAC clones from cv. Renan, INRA-Genoplante Resources (B. Chaloub, personal communication)
Monoisosomic (20″ + 1′), 7		
Trisomics (20″ + 1‴), 21		
Nullisomic-tetrasomic (20″ + 1⁗), 38		
Monosomic-tetrasomic (20″ + 1′ + 1⁗), 4		
Nullisomic (20″), 8		
Single chromosome substitution lines, 184 (http://www.ksu.edu/wgrc)		

The listed BAC libraries are public resources and free of IPR.

of 1 gene/4–5 kb within a small segment on chromosome 1A. The 600-kb contig at the *Tsn1* locus on chromosome 5B contained 13 genes at an average gene density of 1 gene/46 kb (J. D. Faris, J. P. Fellers, H. Lu, K. M. Haen and B. S. Gill, unpublished results). However, 9 of these genes were located within a 90-kb segment, resulting in a gene density of 1 gene/10 kb. In contrast, Faris *et al.* (2003) found only three known genes within a 300-kb BAC contig spanning the *Q* locus on chromosome 5A, yielding an estimate of 1 gene/100 kb. Large tracts of repetitive elements with very few intervening low-copy noncoding sequences separated the three genes.

The physical map of the D-genome donor species *Aegilops tauschii* Coss. is under construction. Five BAC libraries have been constructed and fingerprinted using a new, high-resolution method. Briefly, BAC DNA is simultaneously digested with four 6-bp restriction endonucleases (*Bam*HI, *Eco*RI, *Xba*I, and *Xho*I) and a 4-bp restriction endonuclease (*Hae*III). Subsequently, each of the four recessed 3′ ends generated by the four 6-bp restriction enzymes is labeled with a different fluorescent dye and the fragments are sized using a capillary DNA sequencer (Luo *et al.* 2003a). At the same time, wheat RFLP markers and ESTs have been placed onto the physical map to anchor the BAC contigs to genetic maps and deletion bins (Luo *et al.* 2003b).

SELECTION OF A TARGET WHEAT GENOME FOR SEQUENCING

Ancient or modern farmers have grown four wheat species: einkorn (monococcum), emmer (durum), timopheevi, and common (hexaploid, or bread) wheat. However, only durum and common wheat are currently

used for food production, accounting for 4 and 96% of the total wheat acreage, respectively. The diploid relatives of bread wheat have smaller genomes than that of hexaploid wheat (5500 Mb *vs.* 16,000 Mb), and sequencing one of them should require approximately one-third of the time and expense of sequencing hexaploid wheat. However, in addition to its great economic importance, there are several lines of evidence that led many participants to favor sequencing common wheat. The A, B, and D subgenomes of common wheat have undergone dynamic evolution since they came together to form hexaploid wheat (see Figures 1 and 2). Today, they differ significantly from one another and from the genomes of cultivated diploid wheats. Sequencing hexaploid wheat could yield the greatest store of important new information about wheat and crop plant biology and provide the greatest return on investment.

First, it is the wild wheat species *T. urartu* and not einkorn wheat (the cultivated diploid) that is the A-genome donor of polyploid wheat (Dvořák 1998). The A genomes in these two diploid wheat species have diverged for a million years (Huang *et al.* 2002). During this period, the A genomes of diploid and polyploid wheat accumulated many genetic differences. Both B- and D-genome donors exist only as wild species. It would be difficult to find a diploid species that truly reflects the related, but highly diverged, genome found in hexaploid wheat. Therefore, although the physical map of the diploid D genome will be extremely valuable, hexaploid wheat should be the focus of a sequencing effort as all the characterized genes, mapping populations, and cytogenetic stocks exist in hexaploid wheat.

Second, polyploidy is a major force in plant evolution and especially in agriculture as most crop plants are also polyploid. At least 70% of angiosperms are thought to have undergone one or more cycles of polyploidization (Stebbins 1966; Masterson 1994). Rapid genetic and epigenetic changes and restructuring of the genomes occur in synthetic amphiploids where different genomes are forced to share the same nucleus (Comai 2000; Wendel 2000). Sequence elimination (Feldman *et al.* 1997; Liu *et al.* 1998; Ozkan *et al.* 2001), reactivation of transposable elements (Kashkush *et al.* 2003), and changes in methylation (Shaked *et al.* 2001) and gene expression patterns (He *et al.* 2003) were observed in wheat upon amphiploid formation. Actually, common wheat experienced two sequential polyploidization events at different times (Figure 2). Sequencing the genome of common wheat will offer a unique opportunity to elucidate mechanisms of polyploid speciation and evolution.

Third, many agronomically important genes or their alleles are chromosome specific and are not triplicated (see Figure 3). For example, known useful alleles of pest resistance genes such as those against fungal diseases (*e.g.*, rusts and powdery mildew) and insects (*e.g.*, greenbug, Russian wheat aphid, and Hessian fly) are present in only one of the three homeologous chromosomes; the *Ph* pairing-control genes mentioned above are located on chromosomes 5B and 3D; grain hardness gene *Ha* is located only on 5D because the copies on 5A and 5B were eliminated after polyploidization (Gautier *et al.* 2000).

Comparative sequence analyses have demonstrated that plant genomes are more dynamic than animal genomes. Studies on the *bz* (Fu and Dooner 2002) and *z1C* (Song and Messing 2003) regions of maize revealed significant variation in local gene content and colinearity among inbreds. A similar situation was also found in *Ae. tauschii* (S. Brooks and J. P. Fellers, unpublished results) and *T. monococcum* L. (Scherrer *et al.* 2002). Even though diploid wheats and the subgenomes of hexaploid wheat can be expected to differ significantly from one another, it seems unlikely that an orthologous locus would have been deleted from all three homeologous chromosomes in hexaploid wheat. Thus, sequencing the genome of common wheat would be the best way to harvest all the genes present in the diploid ancestors. ESTs and/or filtered genome sequences of diploid ancestors will be helpful in assigning the wheat BAC sequences to a specific subgenome.

APPROACHES TO SEQUENCING LARGE GENOMES

Depending upon the amount of information required and resources available, a genome can be sequenced by three approaches or their combinations: clone by clone (CBC), whole genome shotgun (WGS), or selective gene sequencing. To date, most finished complex genomes, chromosomes, or subchromosome regions have been sequenced by a CBC approach (*C. elegans* Sequencing Consortium 1998; Arabidopsis Genome Initiative 2000; International Human Genome Mapping Consortium 2001; Feng *et al.* 2002; Sasaki *et al.* 2002; Wood *et al.* 2002; Rice Chromosome 10 Sequencing Consortium 2003; Sherer *et al.* 2003). This strategy requires large insert libraries and fine clone-based physical maps for minimal tiling paths (MTPs). Although the most costly approach, CBC produces long, if not complete, pseudomolecules of a genome or a chromosome and provides the most comprehensive information about structure and function of a genome.

In contrast, WGS takes advantage of computing power to produce draft sequences for a genome relatively quickly by sequencing and assembling small insert libraries (Adams *et al.* 2000; Venter *et al.* 2001; Goff *et al.* 2002; Yu *et al.* 2002). The draft sequences can be used to extract important information such as gene content (Venter *et al.* 2001; Yu *et al.* 2002) and compositional gradients of genes (Wong *et al.* 2002) and to develop markers (Goff *et al.* 2002). The genome image inferred from draft sequences is incomplete, particularly with respect to gene context. The high proportions

of repeated sequences in large genomes pose a major difficulty for WGS in computing capacity, sequence assembly, and financial cost. WGS is very useful for small genomes (Adams *et al.* 2000; Galagan *et al.* 2003), especially for labs with limited genomics resources. A WGS variant is the whole chromosome shotgun, in which a separated chromosome rather than the whole genome is used for library construction (Churcher *et al.* 1997; Bowman *et al.* 1999; Glockner *et al.* 2002). For sequencing the large genome of mouse, a mixed strategy was adopted. Assembly of sevenfold WGS sequences generated a draft genome sequence. CBC sequencing of BACs created a hybrid WGS-BAC assembly while BACs were used for finishing (Mouse Genome Sequencing Consortium 2002).

In recent years, two genome filtration strategies, methylation filtration (MF) (Rabinowicz *et al.* 1999) and C_0t-based cloning and sequencing (CBCS; Peterson *et al.* 2002) or high C_0t (HC; Yuan *et al.* 2003) were proposed for selectively sequencing the gene space of large genomes. MF is based on the characteristic of plant genomes in which genes are largely hypomethylated but repeated sequences are highly methylated. Methylated DNA is cleaved when transferred into a *Mcr+ Escherichia coli* strain and only hypomethylated DNA is recovered. CBCS/HC separates single- and low-copy sequences, including most genes, from the repeated sequences on the basis of their differential renaturation characteristics. Both MF and HC have been used for efficient characterization of the maize gene space (Palmer *et al.* 2003; Whitelaw *et al.* 2003) although their ability to discover >90% of maize genes has not yet been proven. Combining CBCS with genome filtration can reduce the cost greatly while retaining high coverage of genic regions. Another alternative may be identification of gene-rich regions on a detailed physical map and sequencing large-insert clones from these regions.

WHEAT GENOME SEQUENCING: WHAT APPROACHES ARE APPROPRIATE?

In the workshop, various approaches to sequencing the wheat genome were considered. These included selected BAC/CBCS, MF, HC, and/or a combination approach. The discussion was focused on the relative efficiency of each strategy in relation to cost and division of labor among the international participants.

The WGS approach was considered too difficult mainly because of the large size and highly repetitive nature of the wheat genome. Several participants proposed a selected BAC approach, in which the gene-containing BACs were isolated by hybridization with ESTs and fingerprinted to construct MTPs and the gene-rich MTPs were sequenced. It was argued that a global physical map should be considered rather than only the gene-rich regions for greater impact on map-based cloning of agriculturally important genes. For gene filtration, preliminary results showed that MF could enrich wheat genes by 2- to 3-fold (Li *et al.* 2004) or even 5-fold (P. Rabinowicz, A. Bedell, M. A. Budiman, N. Lakey, A. O'Shaughnessy, V. Balija, L. Nascimento, W. R. McCombie and R. A. Martienssen, unpublished results). However, MF results in barley, a close relative of wheat, show a higher level of enrichment, suggesting that the effectiveness of MF in wheat may be underestimated (P. Rabinowicz, A. Bedell, M. A. Budiman, N. Lakey, A. O'Shaughnessy, V. Balija, L. Nascimento, W. R. McCombie and R. A. Martienssen, unpublished results). CBCS/HC enriches for genes by ~10-fold (D. Lamoureux and B. S. Gill, unpublished results). It was suggested that a combination BAC/CBC approach and gene filtration would greatly reduce the cost while retaining high coverage of genic regions. Comprehensive genetic maps were considered pivotal for the assembly of BAC contigs.

Considering the large genome size and the possibility of international cooperation, a chromosome-based approach was suggested, in which a specific chromosome would be flow sorted (Vrána *et al.* 2000; Safar *et al.* 2004) and used to construct a BAC library and a gene filtration library. One advantage of this approach lies in its potential for a division of labor. One country or center could concentrate on one chromosome or one homeologous group.

There was general agreement in the meeting and in follow-up communications within the group that wheat genome sequencing should be conducted in three phases: pilot, assessment, and scale up. The pilot phase was recommended to be a 5-year wheat genome project focused mainly on physical and genetic mapping along with sample sequencing of the wheat genome aimed at better understanding wheat genome structure. This would involve generation of more refined physical maps for wheat and anchoring of these to the genetic map. With respect to sequencing, two types of pilot sequencing were recommended. First, more extensive sampling of the wheat genome would be conducted by sequencing a large number of BACs (>100) that represent the gene-rich regions of the wheat genome and another >100 that would provide a random sampling of the genome. This would provide a test of the gene-rich-region model for the structure of the wheat genome. The pilot phase of sequencing should also involve deeper sequencing of enriched libraries (MF and high C_0t) to provide a more statistically significant sampling of the wheat genome. The pilot phase would also measure the appropriateness of chromosome-arm enrichment methods as they have great promise but as of yet have not been thoroughly sampled. Another vital component of the pilot phase is development and validation of wheat genome annotation methods. The current use of nonstandardized annotation methods prevents assessment of gene content across research groups. The assessment phase will involve a collective analysis of all

available wheat sequence and annotation data. This will then allow for a well-developed, scientifically supported, and economically feasible approach to be proposed for sequencing the entire wheat genome. Thus, the scale-up phase will not be initiated until completion of the assessment phase.

WORKSHOP CONCLUSIONS

There was a strong consensus among the workshop participants that sequencing the wheat genome would deliver positive impacts across the spectrum of education, new research technologies, and application to the agricultural industry and provide new insights into the functioning of a polyploid genome. Wheat genome sequencing could also be expected to stimulate national and international collaboration. An international wheat genome project could be established through the following steps:

1. Constructing an accurate, sequence-ready, global physical (BAC-contig) map anchored to the high-resolution genetic and deletion maps of the 21 chromosomes (see 4 below) of the hexaploid wheat genotype Chinese Spring.
2. Exploring the use of flow-sorted chromosome- and arm-specific libraries in the assembly of the global physical map and in preparation for the sequencing of the gene-containing regions of homeologous chromosome groups.
3. Identifying genomic sequence tags using gene-enrichment procedures such as hi-C_0t or methyl filtration, ESTs, and full-length cDNAs of 2*x*, 4*x*, and 6*x* wheat for an accurate estimation of the wheat unigene set.
4. Leveraging rice sequence and wheat-rice gene synteny, comparative genetics, and wheat unigenes toward the development of high-resolution genetic and deletion maps of the 21 chromosomes of Chinese Spring wheat.
5. Identifying a random set of 100 gene-containing BACs from the physical map and another 100 random BACs for sample sequencing. This will provide a test of the gene-rich model and allow refining the technology for assembling sequences with a high repetitive sequence content. Sample sequencing of BACs from different ploidy wheats and genotypes should also be undertaken.
6. Integrating bioinformatics at every step for project management, data analysis, improved methods of sequence annotation, and dissemination of data.
7. Engaging all wheat stakeholders and educational institutions (K–12) globally, especially in developing countries, and locally in all aspects of the research, technology transfer, workforce training, and promotion of science.
8. Maintaining all data, materials, and resources in the public domain and free of intellectual property rights.
9. Organizing an international steering committee to coordinate and execute all aspects of the wheat genome sequencing project.

This workshop was supported by grants from the National Science Foundation (DBI 0344938) and U.S. Department of Agriculture-National Research Initiative (2003-38836-01639) to B.S.G. This is Kansas Agricultural Experiment Station journal article no. 05-109-J.

LITERATURE CITED

Adams, M. D., S. E. Celniker, R. A. Holt, C. A. Evans, J. D. Gocayne *et al.*, 2000 The Genome sequence of *Drosophila melanogaster*. Science **287:** 2185–2195.

Akhunov, E. D., A. W. Goodyear, S. Geng, L. L. Qi, B. Echalier *et al.*, 2003 The organization and rate of evolution of wheat genomes are correlated with recombination rates along chromosome arms. Genome Res. **13:** 753–763.

Allouis, S., G. Moore, A. Bellee, R. Sharp, P. Faivre Rampant *et al.*, 2003 Construction and characterizasation of a hexaploid wheat (*Triticum aestivum* L.) BAC library from the reference germplasm "Chinese Spring." Cereal Res. Commun. **31:** 331–338.

Arabidopsis Genome Initiative, 2000 Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature **408:** 796–815.

Arumuganathan, K., and E. D. Earle, 1991 Nuclear DNA content of some important plant species. Plant Mol. Biol. Rep. **9:** 208–218.

Bowman, S., D. Lawson, D. Basham, D. Brown, T. Chillingworth *et al.*, 1999 The complete nucleotide sequence of chromosome 3 of *Plasmodium falciparum*. Nature **400:** 532–538.

Brooks, S. A., L. Huang, B. S. Gill and J. P. Fellers, 2002 Analysis of 106 kb of contiguous DNA sequence from the D genome of wheat reveals high gene density and a complex arrangement of genes related to disease resistance. Genome **45:** 963–972.

C. elegans Sequencing Consortium, 1998 Genome sequence of the nematode *C. elegans*: a platform for investigating biology. Science **282:** 2012–2018.

Cenci, A., N. Chantret, K. Xy, Y. Gu, O. D. Anderson *et al.*, 2003 Construction and characterization of a half million clone bacterial artificial chromosome (BAC) library of durum wheat. Theor. Appl. Genet. **107:** 931–939.

Churcher, C., S. Bowman, K. Badcock, A. Bankier, D. Brown *et al.*, 1997 The nucleotide sequence of *Saccharomyces cerevisiae* chromosome IX. Nature **387** (6632 Suppl.): 84–87.

Comai, L., 2000 Genetic and epigenetic interactions in allopolyploid plants. Plant Mol. Biol. **43:** 387–399.

Delaney, D. E., S. Nasuda, T. R. Endo, B. S. Gill and S. H. Hulbert, 1995a Cytogenetically based physical maps of the group 2 chromosomes of wheat. Theor. Appl. Genet. **91:** 568–573.

Delaney, D. E., S. Nasuda, T. R. Endo, B. S. Gill and S. H. Hulbert, 1995b Cytogenetically based physical maps of the group 3 chromosomes of wheat. Theor. Appl. Genet. **91:** 780–782.

Dubcovsky, J., W. Ramakrishna, P. J. SanMiguel, C. S. Busso, L. Yan *et al.*, 2001 Comparative sequence analysis of colinear barley and rice bacterial artificial chromosomes. Plant Physiol. **125:** 1342–1353.

Dvořák, J., 1998 Genome analysis in the *Triticum-Aegilops* alliance, pp. 8–11 in *Proceedings of the 9th International Wheat Genetics Symposium*, Vol. 1, Saskatoon, Saskatchewan, Canada, edited by A. E. Slinkard. University Extension Press, University of Saskatchewan, Saskatoon, Saskatchewan.

Endo, T. R., and B. S. Gill, 1996 The deletion stocks of common wheat. J. Hered. **87:** 295–307.

Faris, J. D., K. M. Haen and B. S. Gill, 2000 Saturation mapping of a gene-rich recombinant hot spot region in wheat. Genetics **154:** 823–835.

Faris, J. D., J. P. Fellers, S. A. Brooks and B. S. Gill, 2003 A bacterial artificial chromosome contig spanning the major domestication locus *Q* in wheat and identification of a candidate gene. Genetics **164:** 311–321.

Feldman, M., B. Liu, G. Segal, S. Abbo, A. A. Levy *et al.*, 1997 Rapid

elimination of low-copy DNA sequences in polyploid wheat: a possible mechanism for differentiation of homoeologous chromosomes. Genetics **147:** 1381–1387.

Feng, Q., Y. Zhang, P. Hao, S. Wang, G. Fu *et al.*, 2002 Sequence and analysis of rice chromosome 4. Nature **420:** 316–320.

Feuillet, C., and B. Keller, 1999 High gene density is conserved at syntenic loci of small and large grass genomes. Proc. Natl. Acad. Sci. USA **96:** 8265–8270.

Feuillet, C., S. Travella, N. Stein, L. Albar, A. Nublat *et al.*, 2003 Map-based isolation of the leaf rust disease resistance gene *Lr10* from the hexaploid wheat (*Triticum aestivum* L.) genome. Proc. Natl. Acad. Sci. USA **100:** 15253–15258.

Freeling, M., 2001 Grasses as a single genetic system. Reassessment. Plant Physiol. **125:** 1191–1197.

Fu, H., and H. Dooner, 2002 Intraspecific violation of genetic colinearity and its implications in maize. Proc. Natl. Acad. Sci. USA **99:** 9573–9578.

Galagan, J. E., S. E. Calvo, K. A. Borkovich, E. U. Selker, N. D. Read *et al.*, 2003 The genome sequence of the filamentous fungus *Neurospora crassa.* Nature **422:** 859–868.

Gautier, M., P. Cosson, A. Guirao, R. Alary and P. Joudrier, 2000 Puroindoline genes are highly conserved in diploid ancestor wheats and related species but absent in tetraploid *Triticum* species. Plant Sci. **153:** 81–91.

Gill, B. S., B. Friebe and T. R. Endo, 1991 Standard karyotype and nomenclature system for description of chromosome bands and structural aberrations in wheat (*Triticum aestivum*). Genome **34:** 830–839.

Gill, K. S., B. S. Gill and T. R. Endo, 1993 A chromosome region-specific mapping strategy reveals gene-rich telomeric ends in wheat. Chromosoma **102:** 374–381.

Gill, K. S., B. S. Gill, T. R. Endo and E. V. Boyko, 1996a Identification and high-density mapping of gene-rich regions in chromosomes of group 5 of wheat. Genetics **143:** 1001–1012.

Gill, K. S., B. S. Gill, T. R. Endo and T. Taylor, 1996b Identification and high-density mapping of gene-rich regions in chromosome group 1 of wheat. Genetics **144:** 1883–1891.

Glockner, G., L. Eichinger, K. Szafranski, J. A. Pachebat, A. T. Bankier *et al.*, 2002 Sequence and analysis of chromosome 2 of *Dictyostelium discoideum.* Nature **418:** 79–85.

Goff, S. A., D. Ricke, T.-H. Lan, G. Presting, R. Wang *et al.*, 2002 A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science **296:** 92–100.

Guyot, R., N. Yahiaoui, C. Feuillet and B. Keller, 2004 *In silico* comparative analysis reveals a mosaic conservation of genes within a novel colinear region in wheat chromosome 1AS and rice chromosome 5S. Funct. Integr. Genomics **4:** 47–58.

He, P., B. R. Friebe, B. S. Gill and J. M. Zhou, 2003 Allopolyploidy alters gene expression in the highly stable hexaploid wheat. Plant Mol. Biol. **52:** 401–414.

Huang, L., S. A. Brooks, W. Li, J. P. Fellers, H. N. Trick *et al.*, 2003 Map-based cloning of leaf rust resistance gene *Lr21* from the large and polyploidy genome of bread wheat. Genetics **164:** 655–664.

Huang, S., A. Sirikhachornkit, X. Su, J. Faris, B. Gill, *et al.*, 2002 Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the *Triticum/Aegilops* complex and the evolutionary history of polyploid wheat. Proc. Natl. Acad. Sci. USA **99:** 8133–8138.

Inada, D., C. A. Bashir, C. Lee, B. C. Thomas, C. Ko *et al.*, 2003 Conserved noncoding sequences in the grasses. Genome Res. **13:** 2030–2041.

International Human Genome Mapping Consortium, 2001 A physical map of the human genome. Nature **409:** 934–941.

Kashkush, K., M. Feldman and A. A. Levy, 2003 Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. Nat. Genet. **33:** 102–106.

Keller, B., and C. Feuillet, 2000 Colinearity and gene density in grass genomes. Trends Plant Sci. **5:** 246–251.

Kellogg, E. A., 2001 Evolutionary history of the grasses. Plant Physiol. **125:** 1198–1205.

Kota, R. S., K. S. Gill, B. S. Gill and T. R. Endo, 1993 A cytogenetically based physical map of chromosome 1B in common wheat. Genome **36:** 548–554.

Leister, D., J. Kurth, D. A. Laurie, M. Yano, T. Sasaki *et al.*, 1998 Rapid reorganization of resistance gene homologues in cereal genomes. Proc. Natl. Acad. Sci. USA **95:** 370–375.

Li, W., and B. S. Gill, 2002 The colinearity of the *Sh2/A1* orthologous region in rice sorghum and maize is interrupted and accompanied by genome expansion in the Triticeae. Genetics **160:** 1153–1162.

Li, W., P. Zhang, J. P. Fellers, B. Friebe and B. S. Gill, 2004 Sequence composition, organization and evolution of the core Triticeae genome. Plant J. (in press).

Lijavetzky, D., G. Muzzi, T. Wicker, B. Keller, R. Wing *et al.*, 1999 Construction and characterization of a bacterial artificial chromosome (BAC) library for the A genome of wheat. Genome **42:** 1176–1182.

Liu, B., J. M. Vega, G. Segal, S. Abbo, M. Rodova *et al.*, 1998 Rapid genomic changes in newly synthesized amphiploids of *Triticum* and *Aegilops.* I. Changes in low-copy non-coding DNA sequences. Genome **41:** 272–277.

Luo, M. C., C. Thomas, F. M. You, J. Hsiao, S. Ouyang *et al.*, 2003a High-throughput fingerprinting of bacterial artificial chromosomes using the snapshot labeling kit and sizing of restriction fragments by capillary electrophoresis. Genomics **82:** 378–389.

Luo, M. C., C. Thomas, K. R. Deal, F. M. You, O. D. Anderson *et al.*, 2003b Construction of contigs of *Ae. tauschii* genomic DNA fragments cloned in BAC and BiBAC vectors, pp. 293–296 in Proceedings of the 10th International Wheat Genetics Symposium, edited by N. E. Pogna, M. Romanó, E. A. Pogna and G. Galterio. Institute Sperimentale per la Cerealicoltura, Roma, Italy.

Ma, J., K. M. Devos and J. L. Bennetzen, 2004 Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. Genome Res. **14:** 860–869.

Margulies, E. H., M. Blanchette, D. Haussler, E. D. Green and NISC Comparative Sequencing Program, 2003 Identification and characterization of multi-species conserved sequences. Genome Res. **13:** 2507–2518.

Masterson, J., 1994 Stomatal size in fossil plants: evidence for polyploidy in a majority of angiosperms. Science **264:** 421–423.

Meyers, B. C., S. V. Tingey and M. Morgante, 2001 Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. Genome Res. **11:** 1660–1676.

Mickelson-Young, L., T. R. Endo and B. S. Gill, 1995 A cytogenetic ladder-map of the wheat homoeologous group-4 chromosomes. Theor. Appl. Genet. **90:** 1007–1011.

Mouse Genome Sequencing Consortium, 2002 Initial sequencing and comparative analysis of the mouse genome. Nature **420:** 520–562.

Ozkan, H., A. A. Levy and M. Feldman, 2001 Allopolyploidy-induced rapid genome evolution in the wheat (*Aegilops-Triticum*) group. Plant Cell **13:** 1735–1747.

Palmer, L. E., P. D. Rabinowicz, A. L. O'Shaughnessy, V. S. Balija, L. U. Nascimento *et al.*, 2003 Maize genome sequencing by methylation filtration. Science **302:** 2115–2117.

Pedersen, C., and P. Langridge, 1997 Identification of the entire chromosome complement of bread wheat by two-colour FISH. Genome **40:** 589–593.

Peterson, D. G., S. R. Schulze, E. B. Sciara, S. A. Lee, J. E. Bowers *et al.*, 2002 Integration of Cot analysis, DNA cloning, and high-throughput sequencing facilitates genome characterization and gene discovery. Genome Res. **12:** 795–807.

Qi, L., B. Echalier, B. Friebe and B. S. Gill, 2003 Molecular characterization of a set of wheat deletion stocks for use in chromosome bin mapping of ESTs. Funct. Integr. Genomics **3:** 39–55.

Qi, L., B. Echalier, S. Chao, G. Lazo, O. D. Anderson *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.

Rabinowicz, P. D., K. Schulz, N. Dedhia, C. Yordan, L. D. Parnell *et al.*, 1999 Differential methylation of genes and retrotransposons facilitates shotgun sequencing of maize genome. Nat. Genet. **23:** 305–308.

Rice Chromosome 10 Sequencing Consortium, 2003 In-depth view of structure, activity, and evolution of rice chromosome 10. Science **300:** 1566–1569.

Safar, J., J. Bartos, J. Janda, A. Bellec, M. Kubalakova *et al.*, 2004 Dissecting large and complex genomes: flow sorting and BAC

cloning of individual chromosome from bread wheat. Plant J. **39:** 960–968.

Sandhu, D., J. A. Chanpoux, S. N. Bondareva and K. S. Gill, 2001 Identification and physical localization of useful genes and markers to a major gene-rich region on wheat group *1S* chromosomes. Genetics **157:** 1735–1747.

SanMiguel, P., A. Tikhonov, Y.-K. Jin, N. Motchoulskaia, D. Zakharova *et al.*, 1996 Nested retrotransposon in the intergenic regions of the maize genome. Science **274:** 765–768.

SanMiguel, P. J., W. Ramakrishna, J. L. Bennetzen, C. S. Busso and J. Dubcovsky, 2002 Transposable elements, genes and recombination in a 215-kb contig from wheat chromosome $5A^m$. Funct. Integr. Genomics **2:** 70–80.

Sasaki, T., T. Matsumoto, K. Yamamoto, K. Sakata, T. Baba *et al.*, 2002 The genome sequence and structure of rice chromosome 1. Nature **420:** 312–316.

Scherer, S. W., J. Cheung, J. R. MacDonald, L. R. Osborne, K. Nakabayashi *et al.*, 2003 Human chromosome 7: DNA sequence and biology. Science **300:** 767–772.

Scherrer, B., B. Keller and C. Feuillet, 2002 Two haplotypes of resistance gene analogs have been conserved during evolution at the leaf rust resistance locus *Lr10* in wild and cultivated wheat. Funct. Integr. Genomics **2:** 40–50.

Sears, E. R., 1954 The aneuploids of common wheat. Mo. Agr. Exp. Sta. Res. Bull. **572:** 1–59.

Sears, E. R., 1966 Nullisomic-tetrasomic combinations in hexaploid wheat, pp. 29–45 in *Chromosome Manipulation and Plant Genetics*, edited by R. Riley and K. R. Lewis. Oliver & Boyd, Edinburgh.

Sears, E. R., and L. M. S. Sears, 1978 The telocentric chromosomes of common wheat, pp. 389–407 in Proceedings of the 5th International Wheat Genetics Symposium, edited by S. Ramanujam. Indian Society of Genetics and Plant Breeding, New Delhi.

Shaked, H., K. Kashkush, H. Ozkan, M. Feldman and A. A. Levy, 2001 Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. Plant Cell **13:** 1749–1759.

Song, R., and J. Messing, 2003 Gene expression of a gene family in maize based on noncollinear haplotypes Proc. Natl. Acad. Sci. USA **100:** 9055–9060.

Sorrells, M. E., M. L. Rota, C. E. Bermudez-Kandianis, R. A. Greene, R. Kantety *et al.*, 2003 Comparative DNA sequence analysis of wheat and rice genomes. Genome Res. **13:** 1818–1827.

Stebbins, G. L., 1966 Chromosomal variation and evolution: polyploidy and chromosome size and number shed light on evolutionary process in higher plants. Science **152:** 1463–1469.

Tikhonov, A. P., P. J. SanMiguel, Y. Nakajima, N. M. Gorenstein, J. L. Bennetzen *et al.*, 1999 Colinearity and its exceptions in orthologous *adh* regions of maize and sorghum. Proc. Natl. Acad. Sci. USA **96:** 7409–7414.

Van der Hoeven, R., C. Ronning, J. Giovannoni, G. Martin and S. Tanksley, 2002 Deductions about the number, organization, and evolution of genes in the tomato genome based on analysis of a large expressed sequence tag collection and selective genomic sequencing. Plant Cell **14:** 1441–1456.

Venter, J. C., M D. Adams, E. W. Myers, P. W. Li, R. J. Mural *et al.*, 2001 The sequence of the human genome. Science **291:** 1304–1351.

Vrána, J., M. Kubaláková, H. Simková, J. Cíhalíková, M. A. Lysák *et al.*, 2000 Flow sorting of mitotic chromosomes in common wheat (*Triticum aestivum* L.). Genetics **156:** 2033–2041.

Wendel, J. F, 2000 Genome evolution in polyploids. Plant Mol. Biol. **42:** 225–249.

Weng, Y., N. A. Tuleen and G. E. Hart, 2000 Extended physical maps and a consensus physical map of the homoeologous group-6 chromosomes of wheat (*Triticum aestivum* L. em Thell). Theor. Appl. Genet. **100:** 519–527.

Werner, J. E., T. R. Endo and B. S. Gill, 1992 Toward a cytogenetically based physical map of the wheat genome. Proc. Natl. Acad. Sci. USA **89:** 11307–11311.

Whitelaw, C. A., W. B. Barbazuk, G. Pertea, A. P. Chan, F. Cheung *et al.*, 2003 Enrichment of gene-coding sequences in maize by genome filtration. Science **302:** 2118–2120.

Wicker, T., N. Stein, L. Albar, C. Feuillet, E. Schlagenhauf *et al.*, 2001 Analysis of a contigous 211 kb sequence in diploid wheat (*Triticum monococcum* L.) reveals multiple mechanisms of genome evolution. Plant J. **26:** 307–316.

Wong, G. K.-S., J. Wang, L. Tao, J. Tan, J. Zhang *et al.*, 2002 Compositional gradients in Gramineae genes. Genome Res. **12:** 851–856.

Wood, V., R. G. William, M.-A. Rajandream, M. Lyne, R. Lyne *et al.*, 2002 The genome sequence of *Schizosaccharomyces pombe.* Nature **415:** 871–880.

Yahiaoui, N., P. Srichumpa, R. Dudler and B. Keller, 2004 Genome analysis at different ploidy levels allows cloning of the powdery mildew resistance gene *Pm3b* from hexaploid wheat. Plant J. **37:** 528–538.

Yan, L., A. Loukoianov, G. Tranquilli, M. Helguera, T. Fahima *et al.*, 2003 Positional cloning of the wheat vernalization gene *Vrn1.* Proc. Natl. Acad. Sci. USA **100:** 6263–6268.

Yan, L., A. Loukoianov, A. Blechl, G. Tranquilli, W. Ramakrishna *et al.*, 2004 The wheat *vrn2* gene is a flowering repressor down-regulated by vernalization. Science **303:** 1640–1644.

Yu, J., S. Hu, J. Wang, G. K.-S. Wong, S. Li *et al.*, 2002 A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science **296:** 79–92.

Yuan, Y., P. J. SanMiguel and J. L. Bennetzen, 2003 High-Cot sequence analysis of the maize genome. Plant J. **34:** 249–55 (erratum: Plant J. **36:** 430).

Zhang, P., 2002 Analysis of the wheat genome by BAC-FISH. Ph.D. Thesis, Department of Plant Pathology, Kansas State University, Manhattan, KS.