

Wheat Science Dynamics: Challenges and Opportunities

PROCEEDINGS INTERNATIONAL
WHEAT QUALITY CONFERENCE –IV
Saskatoon, Saskatchewan, Canada, June 2-6, 2009

EDITORS
RAVINDRA N. CHIBBAR
AND
JAMES E. DEXTER



AGROBIOS (INTERNATIONAL)

Wheat Genome and Gene Analysis

Bikram S. Gill, Sunish K. Sehgal, Bernd Friebe, and Eduard Akhunov

Wheat Genetic and Genomic Resources Center, Department of Plant Pathology, Kansas State University, Manhattan, Kansas, USA

ABSTRACT

A history of wheat genome analysis leading to the current excitement in wheat genome sequencing is reviewed. Genome-interrogation methods pioneered in wheat in the 1920's revealed genome composition and the origin of wheat species (*Triticum monococcum*, $2n=2x=14$, genome AA; *T. turgidum*, $2n=4x=28$, AABB; *T. aestivum*, $2n=6x=42$, AABBDD) through interspecific hybridization and polyploidy. Genetic mapping of agronomic and quality traits was facilitated by the isolation of aneuploid stocks in the 1950's. Wheat chromosomes were individually identified in the 1970's, ushering in studies on wheat chromosome structure, function, and manipulation for crop improvement. Deletion stocks isolated in the 1990's, together with advances in molecular biology, permitted targeted mapping and cloning of a large number of genes by 2005. Around the turn of the century, flow cytometry was used to isolate individual chromosomes and arms for BAC library construction and physical mapping of the 21 wheat chromosomes is in progress (<http://www.wheatgenome.org/>). New sequencing technologies leading to cost savings per read length and the amount of data generated per instrument currently are being explored by research groups in the IWGSC for sequencing the wheat gene space as well as the 'gold standard contiguous sequence' constituting the minimum tiling path of wheat chromosome physical maps.

WHEAT GENOME AND GENE ANALYSIS

At 17 Gb, bread wheat has one of the largest genomes and yet it is amenable to precise genetic manipulation for yield and quality. There is a long and fascinating history of genetics research behind it. It is our intention to briefly review this to the current excitement and the prospect of a complete deciphering of the genetic code of this staple of humankind. An exhaustive review is beyond the scope of this paper but certain major breakthroughs will be reviewed (for recent reviews of most topics touched upon in this book, see Feuillet and Muhelbaur 2009).

By early 1900, taxonomists had recognized three types of wheat namely *Triticum monococcum* called einkorn (one-seeded), *T. turgidum* called dicoccum (two-seeded), and *T. aestivum* called dinkel (spelt, primitive wheat with tapering non-threshing spike). As Mendel's laws of inheritance were rediscovered, it was firmly established that chromosomes were the carriers of hereditary factors. In the 1920's, cytological

(chromosome) analysis of hybrids between different types of wheats (Sax 1922; Kihara 1924) revealed that they were interrelated and evolved through polyploidization (interspecific hybridization followed by spontaneous chromosome doubling of the hybrid) (Fig. 1). The chromosome complement (called $2n$) of einkorn wheat consisted of seven pairs of chromosomes ($2n=14$) as seen in meiosis, 14 pairs ($2n=28$) were observed in dicoccum, and 21 pairs ($2n=42$) in dinkel. The chromosome analysis of F_1 hybrids between einkorn and dicoccum revealed $7''$ (paired)+ $7'$ (unpaired) chromosomes at meiosis and those between dicoccum and dinkel revealed $14'' + 7'$ chromosomes at meiosis (Sax 1922). These analyses revealed that wheat species contain multiples of basic x ($1x=7$) set of chromosomes, and this basic unit was called the 'genome' and this method of genome analysis was called the 'genome analyzer' method (Kihara 1924).

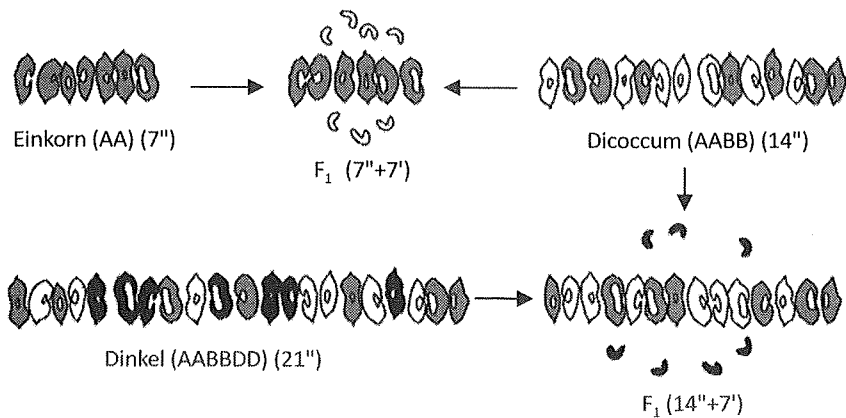


Fig. 1. The meiotic pairing analysis of different wheat species and their hybrids in the 1920's (Sax 1922; Kihara 1924) revealed their evolutionary histories. Einkorn had seven pairs ($7''$) of chromosomes and its hybrid with dicoccum had seven paired and seven unpaired chromosomes indicating that they shared one genome (A) in common. Dicoccum had 14 and dinkel had 21 pairs of chromosomes, and their hybrid showed 14 paired and 7 unpaired chromosomes indicating that dicoccum and dinkel shared two genomes (AB) in common.

The fact that the dicoccum x dinkel hybrid showed 14 paired chromosomes (meiosis only allows pairing between strictly homologous chromosomes) meant that they shared two genomes in common, but the presence of seven unpaired chromosomes indicated that dinkel had one genome that was missing in dicoccum. Similarly, analysis of a einkorn x dicoccum hybrid ($7''$ (paired)+ $7'$ (unpaired)) would lead us to conclude that they shared one genome in common, but dicoccum acquired an additional genome during polyploidization. The *T. monococcum* genome was designated 'A' and the dicoccum genome was designated 'AB'. Between 1944-46, Kihara (1944) and McFadden and Sears (1944, 1946) independently discovered that dinkel, besides having 'AB' genomes derived from dicoccum, obtained its third genome 'D' from *Aegilops tauschii*. (Incidentally, McFadden and Sears logically called it 'C', but this symbol had already been assigned to *Ae. caudata* by Kihara and, thus, the next available symbol 'D' was assigned). Sarkar and Stebbins (1956), based on

brilliant deduction from comparative morphology of einkorn and dicoccum, concluded that the B genome present in dicoccum (and transmitted to dinkel during polyploidization) came from *Ae. speltoides*.

At the time, these were revolutionary findings. The chromosome number and ploidy level were remarkably related to crop traits and grain properties. Einkorn, the diploid wheat, was the first domesticated crop, had a single grain per spikelet, and the grain was suited for speciality foods including porridge. Dicoccum, the tetraploid wheat, had two grains per spikelet and the grain was suitable for macaroni and pasta products. Dinkel wheat had many grains per spikelet and was most suitable for bread making. Breeders began crossing different ploidy wheats to improve the crop and multiple uses of the grain and those distinctions are not as clear-cut in modern day wheats due to extensive mixing of different gene pools by breeders.

Another ploidy change, called aneuploidy, deals with the addition or subtraction of a single chromosome or chromosome arm (a typical chromosome consists of a short arm and a long arm separated by the centromere). Aneuploidy is not tolerated in diploid species, but polyploid species, because of genetic redundancy, can tolerate loss of chromosomes and still be fertile. Sears began aneuploidy research in bread wheat in the 1930's and over the years isolated a complete set of aneuploids where either a whole chromosome or an arm is missing (Sears 1954). Aneuploid stocks simplified chromosome mapping of agronomic and quality traits; any missing trait, protein, or DNA fragment could be assigned to the missing chromosome or the arm. Aneuploids also were used in the production of intercultivar substitution lines and these could be used for genetic mapping of unique agronomic and quality traits of various wheat cultivars. This genetic explosion in wheat research led to the organization of the first International Wheat Genetics Symposium in Winnipeg, Canada, in 1958, and this meeting is held every five years to summarize wheat genetics research.

Aneuploidy unleashed gene and trait mapping of individual chromosomes but very little was known about their structure and chromatin organization. This changed rapidly with the application of new techniques that revealed chromatin states of heterochromatin (gene poor) and euchromatin (gene rich) and these patterns were diagnostic for individual chromosome identification (Gill and Kimber 1974). DNA sequences could be mapped on individual chromosomes in relation to chromatin organization (Rayburn and Gill 1985). A wheat anchor karyotype was developed (Gill et al 1991) and these chromosomes are now the poster children of the International Wheat Genome Sequencing Consortium project (IWGSC; website <http://www.wheatgenome.org/>; see Flags figure).

The ability of rapid chromosome identification and a chromosome breaking genetic element facilitated the isolation of deletion stocks (single break accompanied by loss of acentric fragment) for the 21 chromosomes of wheat (Endo and Gill 1996). Over 500 breakpoints distributed randomly along the length of 21 chromosomes of wheat provide a mapping resolution of 34 Mbp.

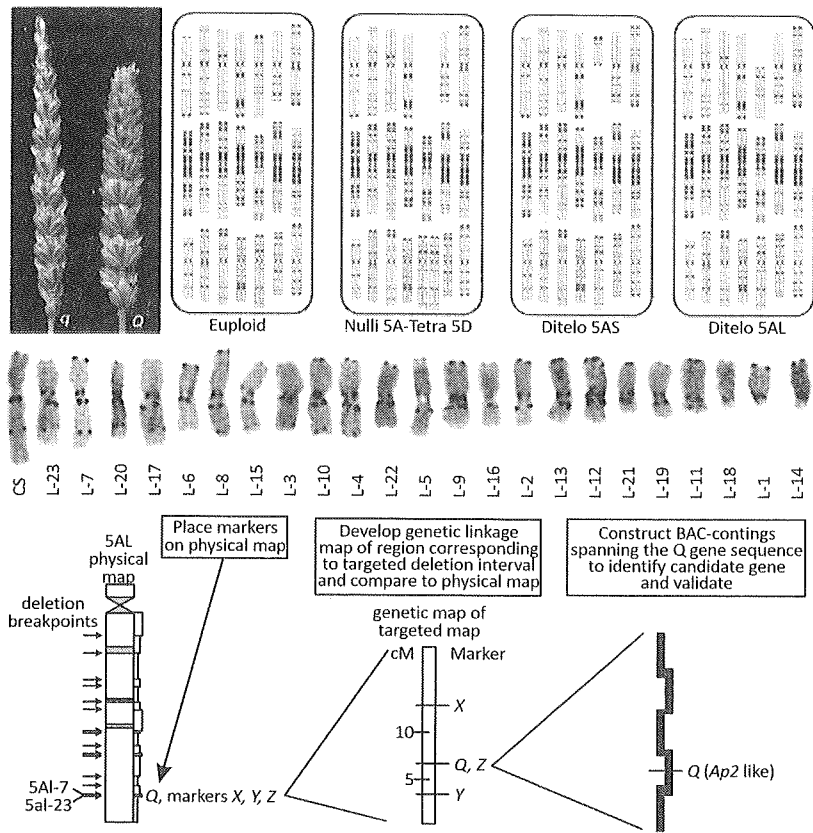


Fig. 2. Targeted mapping and cloning of genes in wheat. Dinkel has primitive speltoid spike due to gene *q*. Mutation at this locus (to *Q*) gave rise to square spike present in modern day wheats (top left panel). The *Q* gene was mapped using Sears (1954) aneuploid stocks (top panel). Phenotype of euploid and ditelosomic (Dt) 5AL (missing arm 5AS) plants was *Q* and those of nullisomic 5A-tetrasomic 5D (missing chromosome 5A) and Dt5AS (missing arm 5AL) was *q*. Therefore, the *Q* gene must be located on 5AL arm. Individual wheat chromosomes are shown having dark- (heterochromatic) and light-staining (euchromatic) regions, and these patterns are diagnostic for individual chromosome identification (Gill and Kimber 1974; Gill et al 1991). The deletion stocks (Endo and Gill 1996), shown for 5A (middle panel) allowed mapping of the *Q* gene to a small interval flanked by deletions 5AL-7 and 5AL-23 (left, lower panel). Targeted molecular and genetic mapping of the *Q* gene region identified a cosegregating DNA marker (middle, lower panel; Faris and Gill 2002) and construction of a BAC-contig map (right, lower panel) led to the cloning of the *Q* gene (Faris et al 2003; Simons et al 2006).

The deletion and Sears' aneuploid stocks allow targeted mapping of genes to specific chromosome segments virtually eliminating complications posed by the massive genome size and polyploidy (Fig. 2). In addition, the wheat transcriptome (called ESTs, expressed sequence tags) was sequenced and mapped into the deletion bins (Qi et al 2004). Most wheat ESTs can be traced to their location in the sequenced genomes of rice and *Brachypodium*, and genetic information from those organisms can be exploited for wheat gene discovery in the targeted deletion bin. The

deletion bin DNA can be mined for additional markers or expressed genes, and cloning of the wheat domestication gene *Q* was an outstanding example of this strategy (Faris and Gill 2002; Faris et al 2003; Simons et al 2006).

Although the above strategy combining targeted and comparative mapping facilitated cloning of a large number of genes in wheat, it was obvious that many genes are unique to wheat, such as gluten proteins that impart bread-making properties to wheat. Wheat genome sequencing provided the best option for accessing all wheat genes and leveraging the extensive knowledge about gene structure and genetic pathways gained from model organisms to wheat improvement (Gill et al 2004).

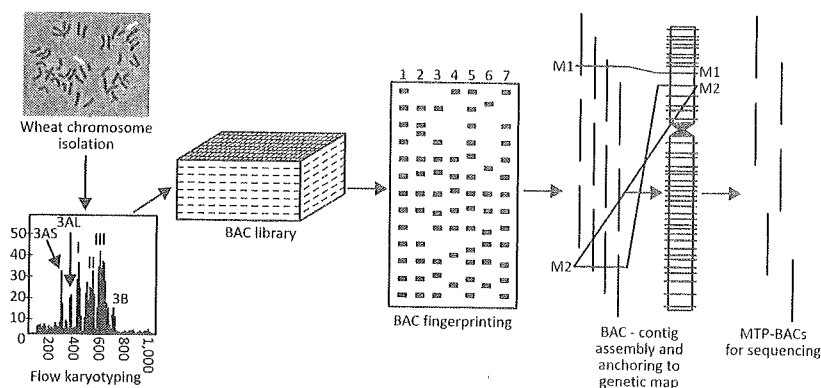


Fig. 3. Flow cytometry permits isolation of wheat chromosomes in high purity for construction of BAC libraries (Dolezel et al 2009) and sequence ready BAC-contig-based physical maps of wheat chromosomes (Stein et al 2009). Most wheat chromosomes are of similar size (top left) and resolve into four fractions. Fractions I, II, and III consist of mixtures of wheat chromosomes except 3B, which forms a separate peak, and its physical map has been published (Paux et al 2008). However, all wheat chromosomes also have been isolated as telosomics (arrowed, top left) that are half the size of a normal wheat chromosome, form a separate peak in a flow karyotype (bottom, left), and will allow physical mapping of all chromosomes except for t5BL, which has the same length as a typical wheat chromosome. The BAC libraries (middle, top) are fingerprinted (middle, bottom) and BAC fingerprints are assembled into contigs based on overlaps and anchored to a genetic map (right panel). A set of unique BAC clones forming the minimum tiling path (MTP) is processed for sequencing (see Fig. 4).

By this time, Dolezel and his group (see review by Dolezel et al 2009) had shown that individual wheat chromosomes and arms (recovered as telocentric chromosomes by Sears and Sears 1978) could be isolated in purity by flow cytometry, processed for high molecular wheat DNA isolation, BAC library preparation, and BAC fingerprinting for the construction of BAC-contig based sequence-ready physical maps (Fig. 3; also see review by Stein 2009). First, a physical map of chromosome 3B was constructed using this strategy (Paux et al 2008) and projects are ongoing in various countries for constructing the physical maps of the remaining 20 wheat chromosomes (see IWGSC website for more details on the ongoing projects).

The ultimate goal of deciphering the wheat genome sequence appears to be within reach, but it will be a challenging task. The size and complexity of the wheat

genome makes it cost-prohibitive to sequence using the standard Sanger sequencing technology. Moreover, the complexity of wheat genome, which consists of over 90% repetitive DNA and only 3% genes (Li et al 2004), poses problems for its assembly and analysis. However, new sequencing technologies, also referred to as next-generation sequencing technologies (NGST) hold a great promise to accelerate analysis of complex genomes.

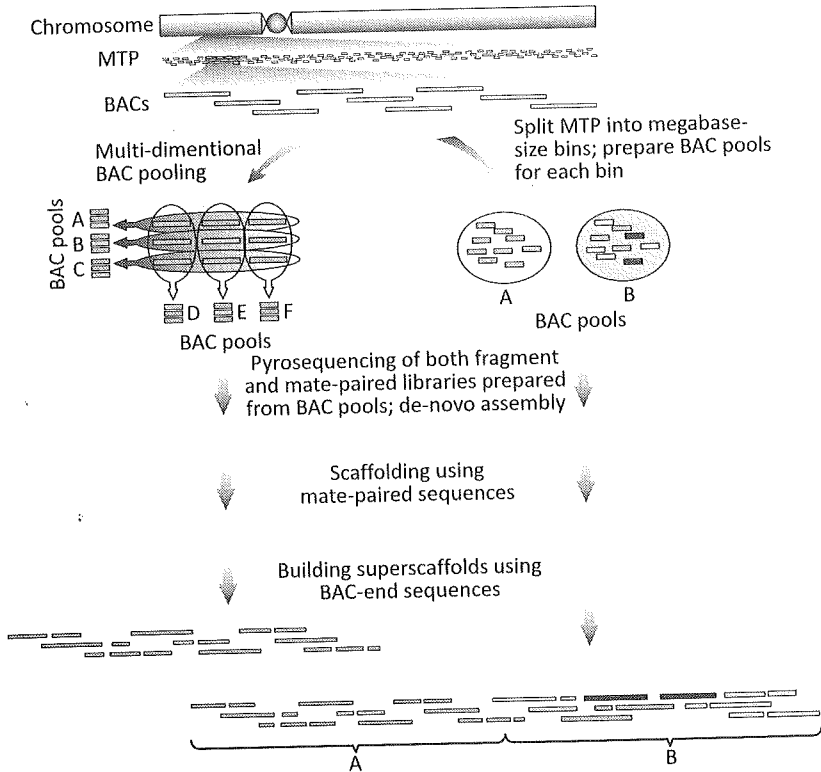


Fig. 4. Possible strategies for generating scaffolds from the MTP of the wheat genome by pooling BAC clones and combining pyrosequencing technology with BAC-end sequencing. Minimum tiling BACs identified from BAC-contigs can be pooled in two or three dimensions to develop multidimensional pools. Simultaneously, the MTP can be split into megabase-size bins to develop bin-based BAC pools. The pools can be pyrosequenced as single fragments or paired-end fragments with 15–20X coverage for *de novo* assembly and scaffolding. The limitation of having one-enzyme based, chromosome-specific libraries hinders complete coverage of the chromosome in a BAC library; hence, random sequencing of sorted chromosomes by NGST may be used to augment sequence assembly and filling of gaps. Further BAC-end sequencing of MTP BACs and random sequencing of sorted chromosomes would help in building superscaffolds.

The NGST dramatically reduced the cost (SOLiD: 30¢/Mb; Helicos: 60¢/Mb; GA: 90¢/Mb; 454 FLX: \$20/Mb of sequence) of production of large volumes of sequence data compared to standard Sanger approach. The three short-read NGST platforms (30–100 bp reads) Illumina/GA, ABI/SOLiD, and Helicos are capable of generating from 18–50 Gb of sequence data per instrument run. Pyrosequencing

technology implemented in Roche's GS FLX 454 instrument can produce 500-bp reads generating 0.5 Gb of data per instrument run. Recently announced by Pacific Bioscience, the Single Molecule Real Time (SMRT™) DNA sequencing technology is capable of generating reads up to 1000 bp. Earlier versions of NGST, due to the short length of the sequence reads, mostly have been used for sequencing organisms with small genomes having low content of repetitive DNA. However, the increased read length and development of mate-pair sequencing strategy significantly improved the capacities of these technologies making them suitable for *de novo* sequencing of more complex repetitive genomes (Imelfort and Edwards 2009; Rounsley et al 2009). The feasibility of sequencing the gene space in large genomes was demonstrated for mungbean, wheat, and barley where NGST were used for assembling the gene space from BAC contigs (Wicker et al 2006) and whole genomes (Tangphatsornruang et al 2009; Wicker et al 2008). Low-coverage, 454 sequencing was used for assessing the repetitive DNA content in barley (Wicker et al 2008). The 454 sequencing of pooled and sequence-tagged BAC clones was shown to be efficient approach for sequencing the gene space in barley genome (Steuernagel et al 2009). The strategy for sequencing the minimum tiling path using pyrosequencing technology was suggested for plant genomes (Rounsley et al 2009). This strategy took the advantage of long reads generated by 454 sequencing technology and long paired-end sequence data to assemble the long arm of the rice *Oryza barthii* (Rounsley et al 2009). Although the repetitive DNA content in the rice genome is significantly lower than that of the wheat genome, this study provided the evidence that NGST can be used for generating the complete sequences of large plant genomes. With the future improvements in NGST, including the increase in read length and the amount of data generated per instrument run, we foresee that it will become feasible to apply similar strategies (Fig. 4) for sequencing the minimum tiling path of the wheat physical map that are currently being generated by research groups in IWGSC.

ACKNOWLEDGEMENTS

The authors are grateful to John Raupp for artwork and editing of the manuscript. Research supported in part by grants from USDA-AFRI (contract nos. 6-00386 and 6-01075), Kansas Wheat Commission and a special USDA grant to Wheat Genetic and Genomic Resources Center. This is contribution #10-241-J from the Kansas Agricultural Experiment Station, Manhattan, KS.

LITERATURE CITED

- Dolezel, J., Simkova, H., Kubalaková, M., Safar, J., Suchankova, P., Cihalikova, J., Bartos, J., and Valarik, M. 2009. Pages 285-316 in: Genetics and Genomics of the Triticeae, Plant Genetics and Genomics: Crops and Models 7. C. Feuillet and G. J. Muehlbauer, eds. Springer Science+Business Media: Dordrecht.
- Endo, T. R. and Gill, B. S. 1996. The deletion stocks of common wheat. *J. Hered.* 87:295-307.
- Faris, J. D., and Gill, B. S. 2002. Genomic targeting and high-resolution mapping of the domestication gene *Q* in wheat. *Genome* 45:706-718.

- Faris, J. D., Fellers, J. P., Brooks, S. A., and Gill, B. S. 2003. A bacterial artificial chromosome contig spanning the major domestication locus *Q* in wheat and identification of a candidate gene. *Genetics* 164:311-321.
- Feuillet, C., and Muehlbauer, G. J., Eds. 2009. *Genetics and Genomics of the Triticeae*. Plant Genetics and Genomics: Crops and Models, Volume 7. Springer Science+Business Media: Dordrecht.
- Gill, B. S., and Kimber, G. 1974. Giemsa C-banding and the evolution of wheat. *Proc. Natl. Acad. Sci. USA*. 10:4086-4090.
- Gill, B. S., Friebe, B., and Endo, T. R. 1991. Standard karyotype and nomenclature system for description of chromosome bands and structural aberrations in wheat (*Triticum aestivum*). *Genome* 34:830-839.
- Gill, B. S., Appels, R., Botha-Oberholster, A-M., Buell, C. R., Bennetzen, J. L., Chalhoub, B., Chumley, F., Dvorak, J., Iwanaga, M., Keller, B., Li, W., McCombie, W. R., Ogihara, Y., Quetier, F., and Sasaki, T. 2004. A workshop report on wheat genome sequencing. The International Genome Research on Wheat Consortium. *Genetics* 168:1087-1096.
- Imelfort, M., and Edwards, D. 2009. De novo sequencing of plant genomes using second-generation technologies. *Brief Bioinform* 10:609-18.
- Kihara, H. 1924. Cytologische und genetische Studien bei wichtigen Getreidearten mit besonderer Rücksicht auf das Verhalten der Chromosomen und die Sterilität in den Bastarden. *Mem. Coll. Sci. Kyoto Imp. Univ. B* 1:1-200.
- Kihara, H. 1944. Discovery of the DD-analyser, one of the ancestors of *Triticum vulgare* (Japanese). *Agric. Hort. (Tokyo)* 19:889-890.
- Li, W., Zhang, P., Fellers, J. P., Friebe, B., and Gill, B. S. 2004. Sequence composition, organization and evolution of the core Triticeae genome. *The Plant J.* 40:500-511.
- McFadden, E. S., and Sears, E. R. 1944. The artificial synthesis of *Triticum spelta*. *Rec. Soc. Genet. Am.* 13:26-27.
- McFadden, E. S., and Sears, E. R. 1946. The origin of *Triticum spelta* and its free-threshing hexaploid relatives. *J. Hered.* 37:81-89, 107-116.
- Paux, E., Sourdille, P., Salse, J., Sautenac, C., Choulet, F., Leroy, P., Korol, A., Michalak, M., Kianian, S., Spielmeier, 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.
- Qi, L. L., Echaliier, B., Chao, S., Lazo, G. R., Butler, G. E., Anderson, O. D., Akhunov, E. D., Dvorak, J., Linkiewicz, A. M., Ratnasiri, A., Dubcovsky, J., Bermudez-Kandianis, C. E., Greene, R. A., Kantety, R., La Rota, C. M., Munkvold, J. D., Sorrells, S. F., Sorrells, M. E., Dilbirligi, M., Sidhu, D., Erayman, M., Randhawa, H. S., Sandhu, D., Bondareva, S. N., Gill, K. S., Mahmoud, A. A., Ma, X-F., Miftahudin, Gustafson, J. P., Wennerlind, E. J., Nduati, V., Gonzalez-Hernandez, J. L., Anderson, J. A., Peng, J. H., Lapitan, N. L. V., Hossain, K. G., Kalavacharla, V., Kianian, S. F., Pathan, M. S., Zhang, D. S., Nguyen, H. T., Choi, D-W., Close, T. J., McGuire, P. E., Qualset, C. O., and Gill, B. S. 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.
- Rayburn, A. L. and Gill, B. S. 1985. Use of biotin-labeled probes to map specific DNA sequences on wheat chromosomes. *J. Hered.* 76:78-81.
- Rounsley, S., Marri, P. R., Yeisoo, Yu. Y., He, R., Sisneros, N., Goicoechea, J. L., Lee, S. J., Angelova, A., Kudrna, D., Luo, M., Affourtit, J., Desany, B., Knight, J., Niazi, F., Egholm, M., and Wing, R. A. 2009. De novo next generation sequencing of plant genomes. *Rice* 2:35-43.

- Sarkar, P., and Stebbins, G.L. 1956. Morphological evidence concerning the origin of the B genome in wheat. *Am. J. Bot.* 43:297-304.
- Sax, K. 1922. Sterility in wheat hybrids. II. Chromosome behavior in partially sterile hybrids. *Genetics* 7:513-552.
- Sears, E. R. 1954. The aneuploids of common wheat. *Mo. Agric. Exp. Sta. Res. Bull.* 572:1-59.
- Sears, E. R., and Sears, L. M. 1978. The telocentric chromosomes of common wheat. Pages 390-407 in: *Proc. 5th. Inter. Wheat Genet. Symp.* New Delhi, India. Ind. Agric. Res. Inst., New Delhi.
- Simons, K. J., Fellers, J. P., Trick, H. N., Zhang, Z., Tai, Y-S., Gill, B. S., and Faris, J. D. 2006. Molecular characterization of the major wheat domestication gene *Q*. *Genetics* 172:547-555.
- Stein, N. 2009. Physical mapping in the Triticeae. Pages 317-335 in: *Genetics and Genomics of the Triticeae, Plant Genetics and Genomics: Crops and Models 7*. C. Feuillet and G. J. Muehlbauer, eds. Springer Science+Business Media, Dordrecht.
- Steuernagel, B., Taudien, S., Gundlach, H., Seidel, M., Ariyadasa, R., Schulte, D., Petzold, A., Felder, M., Graner, A., Scholz, U., Mayer, K. F., Platzer, M., and Stein, N. 2009. De novo 454 sequencing of barcoded BAC pools for comprehensive gene survey and genome analysis in the complex genome of barley. *BMC Genomics* 10:547.
- Tangphatsornruang, S., Somta, P., Uthapaisanwong, P., Chanprasert, J., Sangsrakru, D., Seehalak, W., Sommanas, W., Tragoonrung, S., and Srinives, P. 2009. Characterization of microsatellites and gene contents from genome shotgun sequences of mungbean (*Vigna radiata* (L.) Wilczek). *BMC Plant Biol* 9:137.
- Wicker, T., Schlagenhauf, E., Graner, A., Close, T. J., Keller, B., and Stein, N. 2006. 454 sequencing put to the test using the complex genome of barley. *BMC Genomics* 7:275.
- Wicker, T., Narechania, A., Sabot, F., Stein, J., Vu, G. T., Graner, A., Ware, D., and Stein, N. 2008. Low-pass shotgun sequencing of the barley genome facilitates rapid identification of genes, conserved non-coding sequences and novel repeats. *BMC Genomics* 9:518.