B73 Maize Genome Sequencing (www.maizesequence.org)

Washington University, St Louis, MO (lead institution); Cold Spring Harbor Laboratories, Cold Spring Harbor, NY; University of Arizona, Tucson, AZ; and Iowa State University, Ames, IA

As we approach the last year of the Maize Genome Sequencing Project, we are nearing completion of the initial genome sequence and analysis of maize B73. This includes a primary annotation set (working set genes based on repeat masked DNA), along with a secondary annotation set (protein level). These sets include 16,007 Phase 1 BACs of a total of 16, 625 chosen. At present, we may not necessarily know the order and orientation of all contigs within each BAC, but we do know order and orientation of each BAC in the tiling path. Fusheng Wei at the Arizona Genomics Institute has produced 10 pseudomolecules (chromosomes) along with pseudomolecule 0 (unassigned scaffolds). We also have generated compara gene trees with orthologue calls.

The maize browser, available at http://www.maizesequence.org provides public access to maize BACs and their underlying annotations. The website is tightly integrated with Gramene (http://www.gramene.org) and provides cross-linkage for comparative analysis with other cereal genomes. Mature (improved) BACs have been analyzed using an effective evidence-based gene build strategy in collaboration with Gramene that provides higher-quality gene models. Improved sequences have undergone peptide-based analysis, such as InterPro/GO, to provide greater context for gene models. The maize BAC sequence maps have been integrated with the FPC map. This provides a unified view of the physical and sequence map. Other data sets, such as the maize optical map (see below note), generated by the David Schwartz lab, and full-length cDNAs, provided by the Yeisoo Yu lab (http://www.maizecdna.org/) and Ceres (Alexandrov et al, 2009), have been integrated into the browser, as they became available.

There is still work to be done. Several BACs that only recently were chosen to fill gaps in the tiling path have to be finished and added to the annotation pipeline. As of May 8, 2009, 15, 818 BACs of the total of 16, 625 chosen are finished. This work is ongoing at the Genome Center at Washington University School of Medicine and at The Arizona Genome Institute. These data will undergo analysis and will be integrated into the genomic annotations in the browser.

A manuscript describing the preliminary analysis of the maize genome is in progress and, along with several companion papers, will be published in major journals soon. This will give the first comprehensive look from the genomic level at maize B73.

Alexandrov NN, Brover VV, Freidin S, Troukhan ME, Tatarinova TV, Zhang H, Swaller TJ, Lu YP, Bouck J, Flavell RB, Feldmann KA. Insights into corn genes derived from large-scale cDNA sequencing. Plant Mol Biol. 2009 Jan;69(1-2):179-94. Epub 2008 Oct 21.

Submitted by Sandy Clifton
Washington University, St Louis, MO

B73 Optical Map: A single molecule map of the maize genome

Shiguo Zhou[1], John Nugyen[2], Fusheng Wei[3], Shiran Pasternak[4], Doreen Ware[4], Rod Wing[3], Michael Waterman[2], and David C. Schwartz[1]
[1]University of Wisconsin-Madison, Madison, WI 53706; [2]University of Southern California, Los Angeles, CA 90089; [3]Unversity of Arizona,Tucson, AZ 85721; [4]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724

Maize is one of the most important crops in the U. S., while also being a venerable plant model system for genetic and cytogenetic investigation. Although a substantial body of genetic and cytogenetic studies have provided a global view of maize genome organization, there remains a vast treasure of fine-scaled genomic features awaiting discovery that will surely emerge from analysis of a fully sequenced genome. As part of this effort, we have constructed a genome-wide restriction map for maize using the well-established whole genome shotgun single molecule optical mapping approach [1-13]. Briefly, optical mapping constructs individual restriction maps from millions genomic DNA molecule (300 kb- 2 Mb) that have been stretched on charged surfaces. Automated fluorescence microscopy, coupled to machine vision converts images into data sets comprising high resolution restriction maps. These maps are then assembled in to contigs spanning an entire genome. Although the maize genome is notorious for harboring a complex and extensive panoply of repeats, complicating sequence assembly, such genomic structures are readily characterized by optical mapping because ~500 kb molecules are analyzed. Our optical map of maize genome is facilitating sequence finishing by providing dense restriction marker scaffolds for the ordering and orienting of nascent sequence contigs, the characterization of gaps, and the validation of sequence assemblies.

We constructed a genome-wide optical map of the maize inbred line B73 using SwaI, a methylation insensitive restriction enzyme. SwaI mapped DNA molecules were *de novo* assembled into 68 contigs, each larger than 3 Mb (30.94 Mb, average size) and with a total length of 2103.86 Mb. These optical map contigs span 91.47% of the maize genome (~2300 Mb)[14], with the largest contig spanning more than 100 Mb. A new algorithm was developed in order to utilize the unfinished BAC sequences (http://www.maizesequence.org/index.html release 3a.50) for integration of the maize optical maps with the iMap (FPC physical and Genetic map) [14-21]. This alignment

algorithm includes four steps: *i.* fragment matching – between optical map and sequence contigs; *ii.* BAC alignments – match graph for alignment to optical map contig; *iii.* FPC alignment – dynamic programming for alignment to optical map contig; *iv.* filtering the alignments – based on the colinearity between the alignment order of BACs on optical maps and their order on the FPC map. The algorithm placed 65 of the 68 optical map contigs onto the maize iMap (2082.28 Mb). This result indicates that our optical map is largely congruent with the maize iMap, which in turn cross-validates the BAC physical map resource for maize genome sequencing efforts. Furthermore, as all the anchored optical maps span multiple FPC contigs, gaps between FPC contigs are now estimated. However, we do see multiple regions with conflicting optical map/iMap alignments indicating that some FPC contigs are not correctly placed, or belong to different chromosomes. Further analysis is needed for solving these issues. As most of the BAC sequences in this release comprise of multiple unordered sequence contigs, and some of the BAC sequence contigs can be anchored on optical maps, this also may help finish the BAC sequence assembly. In conclusion, our maize whole genome optical map will be an important resource for finishing maize genome sequence and closing gaps. The finished B73 map/sequence will become a reference for other studies aimed at finding structural differences in other lines, cultivars or varieties of maize.

1. Lin, J. et al., *Whole-genome shotgun optical mapping of Deinococcus radiodurans.* Science, 1999. **285**(5433): p. 1558-62.
2. Lai, Z. et al., *A shotgun optical map of the entire Plasmodium falciparum genome.* Nat Genet, 1999. **23**: p. 309 - 313.
3. Lim, A. et al., *Shotgun optical maps of the whole Escherichia coli O157:H7 genome.* Genome Res, 2001. **11**(9): p. 1584-93.
4. Zhou, S. et al., *A whole-genome shotgun optical map of Yersinia pestis strain KIM.* Appl Environ Microbiol, 2002. **68**(12): p. 6321-31.
5. Zhou, S. et al., *Whole-genome shotgun optical mapping of Rhodobacter sphaeroides strain 2.4.1 and its use for whole-genome shotgun sequence assembly.* Genome Res, 2003. **13**(9): p. 2142-51.
6. Zhou, S. et al., *Shotgun optical mapping of the entire Leishmania major Friedlin genome.* Mol Biochem Parasitol, 2004. **138**(1): p. 97-106.
7. Valouev, A. et al., *Alignment of optical maps.* J Comput Biol, 2006. **13**(2): p. 442-62.
8. Valouev, A. et al., *Refinement of optical map assemblies (original paper).* Bioinformatics, 2006: p. 1217-1224.
9. Valouev, A. et al., *An algorithm for assembly of ordered restriction maps from single DNA molecules.* Proc Natl Acad Sci U S A, 2006. **103**(43): p. 15770-5.
10. Zhou, S. et al., *Validation of rice genome sequence by optical mapping.* BMC Genomics, 2007. **8**: p. 278.
11. Zhou, S., J. Herschleb, and D.C. Schwartz, *A single molecule system for whole genome analysis*, in *New Methods for DNA Sequencing*. 2007, K.R. Mitchelson,Editor, Elsevier B. V.:Amsterdam. p. 269-304.
12. Ananiev, G.E. et al., *Optical mapping discerns genome wide DNA methylation profiles.* BMC Mol Biol, 2008. **9**: p. 68.
13. Kidd, J.M. et al., *Mapping and sequencing of structural variation from eight human genomes.* Nature, 2008. **453**(7191): p. 56-64.
14. Wei, F. et al., *Physical and Genetic Structure of the Maize Genome Reflects Its Complex Evolutionary History.* PLoS Genet, 2007. **3**(7): p. e123.
15. Fang, Z. et al., *cMap: the comparative genetic map viewer.* Bioinformatics, 2003. **19**(3): p. 416-7.
16. Fang, Z. et al., *iMap: a database-driven utility to integrate and access the genetic and physical maps of maize.* Bioinformatics, 2003. **19**(16): p. 2105-11.
17. Cone, K.C. et al., *Genetic, physical, and informatics resources for maize. On the road to an integrated map.* Plant Physiol, 2002. **130**(4): p. 1598-605.
18. Coe, E. et al., *Access to the maize genome: an integrated physical and genetic map.* Plant Physiol, 2002. **128**(1): p. 9-12.
19. Gardiner, J. et al., *Anchoring 9,371 maize expressed sequence tagged unigenes to the bacterial artificial chromosome contig map by two-dimensional overgo hybridization.* Plant Physiol, 2004. **134**(4): p. 1317-26.
20. Nelson, W.M. et al., *Whole-genome validation of high-information-content fingerprinting.* Plant Physiol, 2005. **139**(1): p. 27-38.
21. Yim, Y.S. et al., *Characterization of three maize bacterial artificial chromosome libraries toward anchoring of the physical map to the genetic map using high-density bacterial artificial chromosome filter hybridization.* Plant Physiol, 2002. **130**(4): p. 1686-96.

Mo17 Genome Sequencing (Editors' Note)

Mo17 genome shotgun sequence reads by 454 sequencing have been aligned to the B73 BAC sequence, and are freely available at www.phytozome.org. This project was carried out by the DOE Joint Genome Institute, UC Berkeley. These alignments have been incorporated into the MaizeGDB Genome Browser, representing SNP, insertions and deletions.