

Meeting report

Maize genome in motion

Virginia Walbot

Address: Department of Biology, Serra Mall, Stanford University, Stanford, CA 94305-5020, USA. Email: Walbot@stanford.edu

Published: 7 April 2008

Genome Biology 2008, **9**:303 (doi:10.1186/gb-2008-9-4-303)

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2008/9/4/303>

© 2008 BioMed Central Ltd

A report on the Maize Genetics Conference held in Washington DC, USA, 27 February-1 March, 2008.

Why sequence the genome of maize? Landmark discoveries in *Zea mays* L. (maize or corn) started with Edward East's 1908 report of inbreeding depression and hybrid vigor and were followed by cytogenetic breakthroughs by Barbara McClintock, George Beadle, Marcus Rhoades and other pioneers, who unraveled the relationship between chromosome physical structure and fundamental phenomena such as recombination, nuclear-cytoplasmic interaction, centromere function and origin of chromosome ends. The 'big four' epigenetic phenomena were also discovered in maize: transposable elements by McClintock in the 1940s; paramutation (non-Mendelian heritable silencing of specific alleles) independently by Alexander Brink and Ed Coe in the mid-1950s; Jerry Kermicle's 1970 report of parent-of-origin imprinting impacting on gene expression during early post-fertilization development; and the link between DNA methylation and transposon silencing by Vicki Chandler and myself in 1986. The desirability of a genome sequence for this much-studied and valuable crop plant seems indisputable.

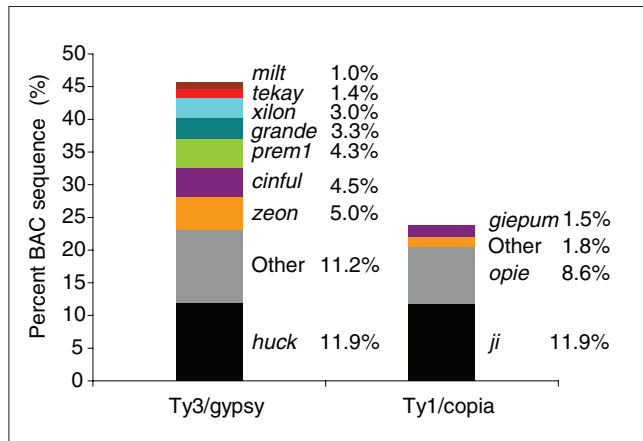
Recent genetic and genomics assessments of diverse maize inbred lines and landraces have uncovered astonishing new information about genome structure and fluidity. Ed Buckler (US Department of Agriculture-Agricultural Research Service (USDA-ARS), Cornell University, Ithaca, USA) pointed out the phenomenal diversity in single nucleotide polymorphisms (SNPs) and indels (insertions/deletions) in maize: the genomes of humans and chimps, separated by 3.5 million years of evolution, are 1.34% divergent, whereas two inbred lines of maize separated for only a few thousand years average 1.42% divergence. Second, as noted by Jean-Philippe Vielle-Calzada (Langebio Cinvestav, Irapuato, Mexico), intra-specific genome size varies from 1,700 million to 3,300 million bases (Mb). This variation reflects two processes: maize intergenic regions are graveyards accumulating

retro-transposon insertions; and DNA transposons catalyze rapid reshuffling of gene order and gene content.

Building the reference for a fluid genome

The draft genome of the maize inbred line B73 is the template for analyzing genome fluidity and epigenetic regulation on a global scale. Rick Wilson (Washington University School of Medicine, St Louis, USA), leader of the Maize Genome Sequencing Consortium (MGSC), reported that maize could be the last genome sequenced using a complete bacterial artificial chromosome (BAC) tiling path. He commented that successful sequencing may have required this method, however, as the maize genome is 78% repetitive DNA, mostly composed of 11 types of retro-transposons (Figure 1). A typical 150-kB BAC contains one or a few compact genes (similar in size to those of other plants but 10-fold shorter than in mammals) and many nested retroelements. Given funding of less than US\$30 million, about 1% the funding for the Human Genome Project, the strategy is to finish genes to very high quality but leave the retrotransposon mess partially unassembled. As transcription, recombination and DNA transposon insertion are all centered on genes, the four or so gaps per BAC in the retrotransposon graveyard between genes should be almost invisible during genetic and molecular analyses.

The B73 draft has around 6,600 completed and 9,000 BACs in the finishing stage to boost gene quality, and Wilson reported that a final set of around 200 BACs will be launched to fill gaps between BACs on the 10 chromosomes. Doreen Ware (Cold Spring Harbor Laboratories, New York, USA) described an automated pipeline, developed with MGSC colleagues, to define exons and introns, using an evidence-based approach employing the approximately 1.4 million publicly available expressed sequence tags (ESTs). The soon to be finished 30,000 B73 full-length cDNAs, reported by Dave Kudrna (University of Arizona, Tucson, USA), who, with Yeisoo Yu, manages the sequencing, and John Fernandes (Stanford University, Stanford, USA), who designs all the primers, will clarify transcription start sites

**Figure 1**

Retrotransposons collectively comprise 76% of the maize genome. In this preliminary analysis of the draft genome sequence by Josh Stein (MGSC, Cold Spring Harbor Laboratory, Cold Spring Harbor, USA) it is clear that just the gypsy family *huck* element plus the copia relative *ji* make up nearly one-quarter of the genome - about 600 Mb. Image courtesy of Richard Wilson and Josh Stein.

and confirm exon-intron boundaries. Currently sitting in sixth place overall in EST count (http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html), maize will soon jump to third place when 2 million additional ESTs determined by commercial companies are released.

Particularly exciting for the entire cereal community was the description by Richard McCombie (Cold Spring Harbor Laboratory, New York, USA) and Fusheng Wei (University of Arizona, Tucson, USA) of Ensembl Compara, a whole-genome alignment tool. McCombie reported that segments of diploid rice (which is separated from maize by about 50 million years evolution) and sorghum align with two chromosomes of maize, reflecting the recent (around 4.8 million years ago) allotetraploidization of maize - when a wide cross between two diploid species led to a double chromosome number in the ancestor of maize. The main story is gene loss (<http://www.maizesequence.org>). Within a duplicated 22-Mb contig analyzed in detail by Wei and colleagues, maize has typically retained only one gene copy: interspecific synteny is retained with short gene gaps in each duplicated maize region and with entirely distinct retrotransposons in each species. For some retained duplicated genes, genetic analysis has already shown subfunctionalization: *R* and *B* are genes for basic helix-loop-helix transcription factors that control where and when anthocyanin pigments accumulate. Most *R* alleles regulate floral and seed color whereas typical *B* alleles regulate leaf/stem color.

In addition to the diverse inbred lines used by geneticists and the hybrid seed industry, even greater diversity is represented in various landraces. Within the past 10,000 years corn was domesticated near Mexico City from the wild

plant teosinte. Since then, migrating people planted this crop from Canada to Patagonia in deserts, savannahs, forests and the tropics. Without a doubt, human selection for a rapidly adaptable plant is the best explanation for retention of around 70% of the allelic diversity of teosinte, despite the bottleneck of domestication, and for the current activity of DNA transposons that quickly generate new alleles and genes. It was thus very welcome news to hear from Vielle-Calzada that a Mexican consortium has completed a draft genome of the popcorn landrace Palomero Toluqueño. This variety was chosen for its small genome size after a survey of 230 Mexican landraces and because archeological evidence points to popcorns as the first domesticated maize. After construction of libraries biased against highly repetitive and methylated DNA, short-read 454 sequencing (ultra-high-throughput pyrosequencing) and a smaller amount of Sanger sequencing yielded 10% of the assembly in contigs greater than 1 kb, as reported by Vielle-Calzada. The current predicted gene number is around 58,000 for Palomero and around 50,000 for B73. The discrepancy will be partly resolved by better annotation, but from previous work the community expects some 5% difference in gene content between lines, in the light, for example, of the classic work by Hugo Dooner showing gene movements to other chromosomes from the *bz1* region of chromosome 9S.

Mihai Miclaus (Rutgers University, Piscataway, USA) reported both duplications and movements of zein storage protein genes. By comparing B73 to other maize lines, he and colleagues have found that helitron transposons (which utilize rolling-circle replication and readily pick up parts of chromosomes), can move intact genes, often generating pseudogenes, and that partial gene copies are often dispersed around the genome by other DNA transposons (that is, packmules of the Mu family). The result is that novel combinations of protein domains generate new genes and pre-existing genes acquire new regulatory regions.

Artificial chromosomes for commerce and science

Maize has supernumerary B chromosomes that contain few if any functional genes. Single Bs undergo non-disjunction during pollen mitosis followed by preferential (70%) fertilization of the egg by the sperm that carries two Bs; 30% of progeny lack the B. Jim Birchler and his team of chromosome engineers (University of Missouri, Columbia, USA) have produced truncated miniBs that contain only a centromere and telomeres, but also carry transgenes and site-specific recombination cassettes. Birchler reported that these miniBs lack non-disjunction, which nevertheless can be restored by the addition of full-length B chromosomes. This introduction provides the means of removing transgenes from the genotype at the next meiosis, thus selecting the 30% of progeny lacking miniBs. These 'top down' natural, but engineered, chromosomes promise to become standard tools for basic and applied research.

The 'bottom up' construction of chromosomes is also advancing rapidly, as reported by two groups from industry. Shawn Carlson (Chromatin Inc, Chicago, USA) described five generations of tests for meiotic stability of the 'bottom-up' chromosomes by design. Sergei Svitashv (Pioneer Hi-Bred International, Johnston, USA) focused on mitotic stability of such chromosomes. The conclusion of each presentation was that currently both mitotic and meiotic stability are less than with the B chromosomes, but stochastic loss could be exploited to select individuals lacking the engineered chromosome.

Phenotyping on the grand scale

Given the incredible diversity of maize, fine-scale mapping has been easy. Selective sweeps are readily pinpointed as mono-allelism. For example, John Doebley (University of Wisconsin, Madison, USA) reported cloning the 'domestication' genes that eliminated the tough fruit case surrounding the seed and remodeled a highly branched wild plant into the single stalk of modern corn, and pinpointing the mutations in corn compared with teosinte. The first step in the analysis is to search for maize loci that lack the expected allelic diversity: a complete selective sweep is presumptive evidence for recent human selection.

To address the more difficult problems of traits controlled by genes with small effects and by genotype-environment interaction, the nested association mapping (NAM) stocks were developed. Mike McMullen (USDA-ARS, University of Missouri, Columbia, USA) representing the Maize Diversity Project (MDP [<http://www.panzea.org>]), described how 25 different lines were crossed to B73 and then 200 recombinant inbred lines derived from each initial cross. These lines have been genotyped for 1,200 SNPs, giving around 1 centiMorgan (cM) resolution across most of the genome. The immortalized NAM lines provide the best resource yet developed for analysis of complex traits in a higher eukaryote. Stock maintenance is cheap, because seeds can be stored for a decade or more.

James Holland (USDA-ARS, North Carolina State University, Raleigh, USA) described how the 5,000 NAM lines, a second mapping population called IBM, association panels and controls have been grown as 6,028 blocks per location in 11 distinct environments to develop a catalog of phenotypic data for around 30 agronomic traits. Because linkage disequilibrium disappears about 2 kb from any gene, careful phenotyping combined with genotyping is placing quantitative trait loci (QTLs) into narrow bins within the maize genome, which is telling us a great deal about additive traits, even when 20 or more genes contribute. This enormous effort is part of the MDP. An illustration of the power of the structured populations was provided by analyzing flowering time. The 26 NAM starting lines each flower at a discrete time, but these dates spread out over a 45-day period and for

93% of the alleles underlying this variation, each allele contributes only a 1.5-day impact on the date on which flowering initiates. Mapping of SNPs within the QTL interval leads to identification of the genes underlying the QTL. The enormous dataset generated by the MDP permits deep and detailed inquiry into the robustness of fitness as well as defining the contributing loci. This is an appropriate goal for the genome age, as it was lobbying by the National Corn Growers Association about a decade ago that launched the Plant Genome Research Program at the National Science Foundation (NSF) and fostered interagency cooperation between NSF and the US Departments of Energy and Agriculture to support funding of the corn genome sequencing. This and many other projects that benefit plant geneticists and ultimately the US public and the world have received an investment of nearly \$1 billion so far.

Historically, many very smart people have worked with corn, exploiting the high resolving power of visible markers and large populations: 10^7 meiotic events per plant in the pollen and easy scoring of 250,000 epidermal cells per kernel on 30 ears of corn gives 10^{-9} resolution for phenomena such as a change in transposon excision frequency. With genomes in hand, the advantages of corn for in-depth study of genome fluidity, epigenetic gene regulation, and genotype-environment interactions should accelerate and attract a new generation of geneticists to share in future discoveries.