Maydica 50 (2005): 285-303

GENETIC, PHYSICAL, MAPS, AND DATABASE RESOURCES FOR MAIZE[1]

E.H. Coe*, M.L. Schaeffer

Plant Genetics Research Unit, Agricultural, Research Service, U. S. Department of Agriculture, and University of Missouri, Division of Plant Sciences, Columbia, MO 65211, USA

Received May 2, 2005

ABSTRACT - Resources for maize genetics and genomics exist in great depth and breadth. They can be brought to bear on its productivity, on selected properties, and on studies of genetic functions, mechanisms of inheritance, phylogeny, and processes of change during domestication. Genetic materials available include the trait variations in tens of thousands of diverse germplasm collections, and a readily accessible, 75-year collection of mutant variations. Descriptions of germplasms as well as of mutant variations are available online. Maps of genes in diagrammatic and in tabulated form, accompanied by linked supportive data, have been updated recently. High-resolution genetic maps with molecular markers undergird physical mapping, anchoring and orienting contigs to the chromosomes. Physical mapping has progressed to the point that minimum tiling paths can be defined for the genome, the gene space is largely encompassed, and sequencing of the genome is ready to proceed.

KEY WORDS: Mutants; Variations; Genetic maps; Physical maps; Genome sequencing; Databases; Informatics.

ACRONYMS

AGCoL	Arizona Genomics Computational Laboratory
AGI	Arizona Genomics Institute
BAC	Bacterial Artificial Chromosome
CHORI	Children's Hospital Oakland Research Institute
CIMDE	Community IBM Map Data Entry
CIMMYT	International Center for Maize and Wheat Improvement
CUGI	Clemson University Genomics Institute
FPC	Fingerprinted Contigs
FSD	FPC Simulated Digest
GRIN	Germplasm Resources Information Network, USDA
IBM	Intermated B73 x Mo17 population
iMAP	Integrated map tool
InDel	Insertion-deletion polymorphism
IWG	Interagency Working Group on Plant Genomes
LIMS	Laboratory Information Management System
MaizeGDB	Maize Genome Database
MMP	Maize Mapping Project
NCGA	National Corn Growers Association
NCGI	National Corn Genome Initiative
PCR	Polymerase Chain Reaction
PGI	Plant Genome Initiative
PGIR	Rutgers Plant Genome Initiative
RFLP	Restriction Fragment Length Polymorphism
SNP	Single-Nucleotide Polymorphism
SSR	Simple Sequence Repeat
USDA	U.S. Department of Agriculture

[1] Support from National Science Foundation 9872655 is acknowledged.

* For correspondence (fax + 1 573 884 7850; email: coee@missouri.edu).

INTRODUCTION

Maize is a focal plant for biological analysis as well as for high grain productivity. Biological advances and productivity advances are made possible by genetic materials; by physical materials that advance genome sequencing; by genetic and physical maps; and by comprehensive databases, including not only content in depth but also efficient informatics access to that content. These are available for maize as the result of nearly a century of cooperative sharing of data and materials among scientists conducting research on this plant. Knowledge and data have expanded at a striking rate during recent years, however, through the stimulus of genomic resources and research.

Accomplishing maize genome analysis required a concerted effort. Recognizing this need, in 1990-1994 members of the Maize Genetics Cooperation brought recommendations to U.S. Government agencies and the National Corn Growers Association (NCGA) on maize research and crop advancement. A National Corn Genome Initiative (NCGI) developed (Table 1; Coe, 1998). This initiative broadened into the U.S. Plant Genome Initiative (PGI) with widespread support by NCGA and other

TABLE 1 - *Events in the development of the genome initiative for maize.*

Date	Key Event	Venue; Organizers	Contacts with
Jan-May 1990	Workshop: Recommendations for Development of Genomics	Madison, WI: Maize Genetics Conference (MGC); Maize Genome Coordinating Committee	Plant Genome Initiative (PGI), USDA/ARS
1990 - 1992	Initiatives and coordinations on plant genomics	Meetings and grants	USDA/ARS and CSRS, NSF, DOE
Jun 1994	Corn Genome Mapping Discussion Group	St. Louis, MO Critical Technologies Partnership; National Corn Growers Association (NCGA)	NCGA; St. Louis Economic Council; Washington Univ; industries
Aug 1994	Forum on Maize Genome Mapping	St. Louis, MO; NCGA	NCGA, PGI
Oct 1994	Strategy planning meeting	St. Louis, MO; NCGA	
Mar 1995	Workshop: National Corn Genome Initiative (NCGI), Grass Synteny	Asilomar, CA; MGC	Cooperators; NCGA
Apr 1996	Workshop: Needs, strategy, components of value of NCGI	St.. Louis, MO; NCGA	Stakeholders, researchers
Mar 1996	Workshop: Maize Genetics Cooperation, NCGI	Pheasant Run, IL; MGC	Cooperators
Aug 1996	Workshop: Needs, strategy, components of value of NCGI	Des Moines, IA; NCGA	Stakeholders, researchers
Dec 1996	Business Plan in Support of Mapping the Corn Genome	NCGA and consortium	
Apr 1997	Meeting: Designing an Agricultural Genome Program	Washington, DC; National Academy of Sciences	Invited Participants
Jun 1997	Colloquium: Protecting Our Food Supply: The Value of Plant Genome Initiatives	Beckman Center, Irvine, CA; National Academy of Sciences	Participants
Sep 1997	Meeting on role of private sector	Interagency Working Group on Plant Genomes (IWG)	Industries
Oct 1997	Meeting on science components of a plant genome initiative	IWG	Invited Participants
Dec 1997	Meeting with commodities groups	IWG	Commodities groups
Dec 1997	NSF announces Plant Genome Program		
Mar 1999	Planning Meeting: Maize Genomics	Allerton Park, IL; Ad Hoc Vision Committee	Cooperators
Mar 1999	Workshop: Genomic Approaches and Projects	Lake Geneva, WI; MGC	Cooperators
Jan 2000	Forum: Advancement of Maize Research	San Diego, CA; Plant and Animal Genome Conference (PAG)	Participants
Mar 2000	Workshop: Genomics Resources	Coeur D'Alene, ID; MGC	Participants
May 2000	Maize Genetics Executive Committee (MGEC) Elected		
Mar 2001	Workshop: Maize Genomics and Future Directions	Lake Geneva, WI; MGC	Cooperators
Jul 2001	Workshop: Sequencing the Maize Genome	St. Louis, MO; MGEC	NCGA, IWG; see BENNETZEN *et al.* (2001)
Mar 2002	Genomics Sessions; White Paper on Sequencing	Kissimmee, FL MGC;MGEC	Cooperators; IWG
Feb 2004	Brief on Sequencing Plan of Work	MGEC	Industry, NSF and IWG, NCGA

crop commodity groups, as well as by scientific societies including the American Society of Plant Biologists. An Interagency Working Group on Plant Genomes (IWG) was appointed in 1997 by the Assistant to the President for Science and Technology, in response to a request from the Senate VA, HUD and Independent Agencies Appropriations Subcommittee (INTERAGENCY WORKING GROUP ON PLANT GENOMES, 1998). The IWG defines the needs, the goals, and the operating principles that guide the coordination of programs with the U.S. National Science Foundation, the U.S. Department of Agriculture, and the U.S. Department of Energy. The brief summary that follows describes the still-advancing outcomes of the NCGI. Internationally, national and regional initiatives and collaborations participate in advancing knowledge of maize genetics and genomics.

Genetic materials

A rich resource of genetic materials is available for maize among germplasm collections and among spontaneous and induced visible and biochemical mutations. Accompanying these materials are extensively documented genes and maps.

The germplasm banks for maize contain tens of thousands of diverse strains, which are carefully maintained, are distributed on request, and are steadily screened by public and private scientists for traits that will contribute to worldwide maize breeding efforts. The banks include those of the USDA, ARS, NATIONAL GENETIC RESOURCES PROGRAM GERMPLASM RESOURCES INFORMATION NETWORK - (GRIN) (2005) and those of the INTERNATIONAL CENTER FOR MAIZE AND WHEAT IMPROVEMENT IN MEXICO (CIMMYT SEED INSPECTION AND DISTRIBUTION UNIT, 2005), in particular.

The maize mutant gene collection, consisting of inherited variations shared by working maize scientists for more than 70 years, is well established, is widely known, and is of great value for research and for teaching. Descriptions and select photographs have been presented in Mutants of Maize (NEUFFER *et al.*, 1997). Extensive descriptions, mapping data, images, gene products, and references are presented in the MAIZE GENOME DATABASE, MAIZEGDB (2005). MAIZEGDB presents a full list of the genes that have been identified in maize. Over 4000 stocks are maintained by the MAIZE GENETICS COOPERATION STOCK CENTER (2005) and are made available in response to requests. This collection continues to grow with new deposits of shared materials.

The first list of genes, with descriptions of phenotypes and a baseline genetic map, was published 70 years ago (EMERSON *et al.,* 1935). Of the 300 genes, 165 were placed to linkage groups and 60 could be mapped in order based on extensively tabulated data. These were the first genetic maps. Subsequent genetic maps were largely accreted onto the 1935 map framework until the maps in Mutants of Maize, which merged genetic data onto a map framework based on molecular markers (NEUFFER *et al.*, 1997). Mutants of Maize provides a list of genes with descriptions and photographs, all of which are included in MAIZEGDB. In 2005, the known genes in MAIZEGDB exceed 5,300.

A comprehensive update of mapped genes (COE, 2005) has been released in tables in MAIZEGDB, including 720 genes with map orders and distances. The tables represent a newly re-evaluated assembly and construct, incorporating retrospective data and data from high-resolution mapping and molecular markers whenever these are available. Supporting data are provided as comments for each mapped locus. Map drawings are in the present paper (Fig. 1). Given the rapid pace at which genes are being cloned and are being mapped, these maps can be expected to be updated with additions and modifications as further data are analyzed and new data become available.

Molecular markers, maps, and physical genomic materials

Because of interest in the biochemical basis of phenotypes, of traits, of diversity, and of heterosis, every new advance in tools has been applied toward comprehension of the maize genome and the genes. Isozymes, i.e., protein polymorphisms, when they first became ready subjects of study, were mapped extensively and were tested for linkage with quantitative traits or for relationship to heterosis (STUBER and MOLL, 1972; GOODMAN *et al.*, 1982; EDWARDS *et al.,* 1987). Isozyme studies in fact set the stage for application of DNA probes in mapping, by identifying mapping parents and mapping populations that were usefully diverse. Prior to 1998 a few thousand anonymous sequences, less than a hundred ESTs, and a few dozen clones from known visible genes, had been derived and mapped at low resolution.

The MAIZE MAPPING PROJECT (MMP) of the PGI/NCGI, which was funded by the National Science Foundation in 1998-2003, produced components and data for an integrated genetic and physi-

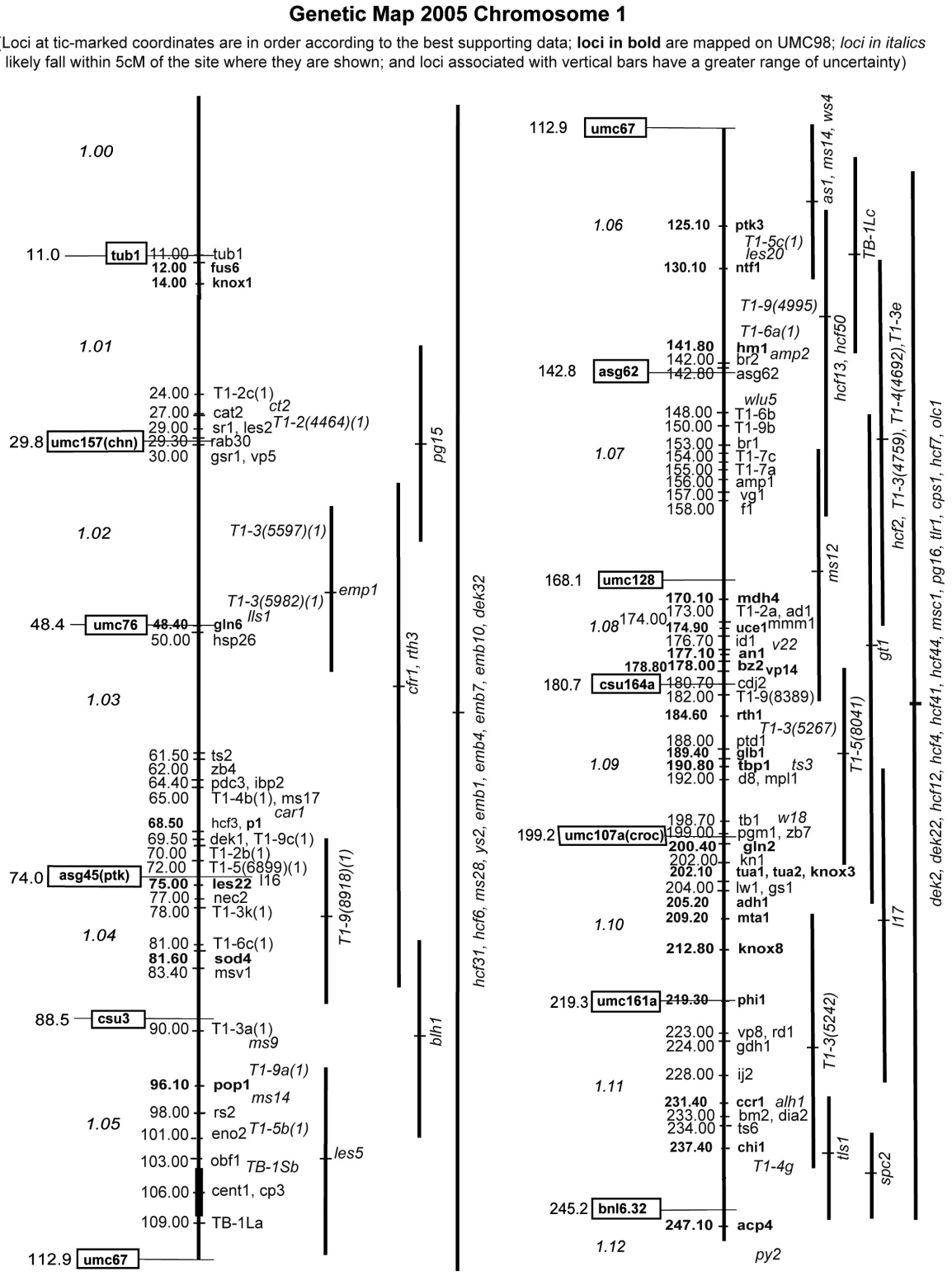
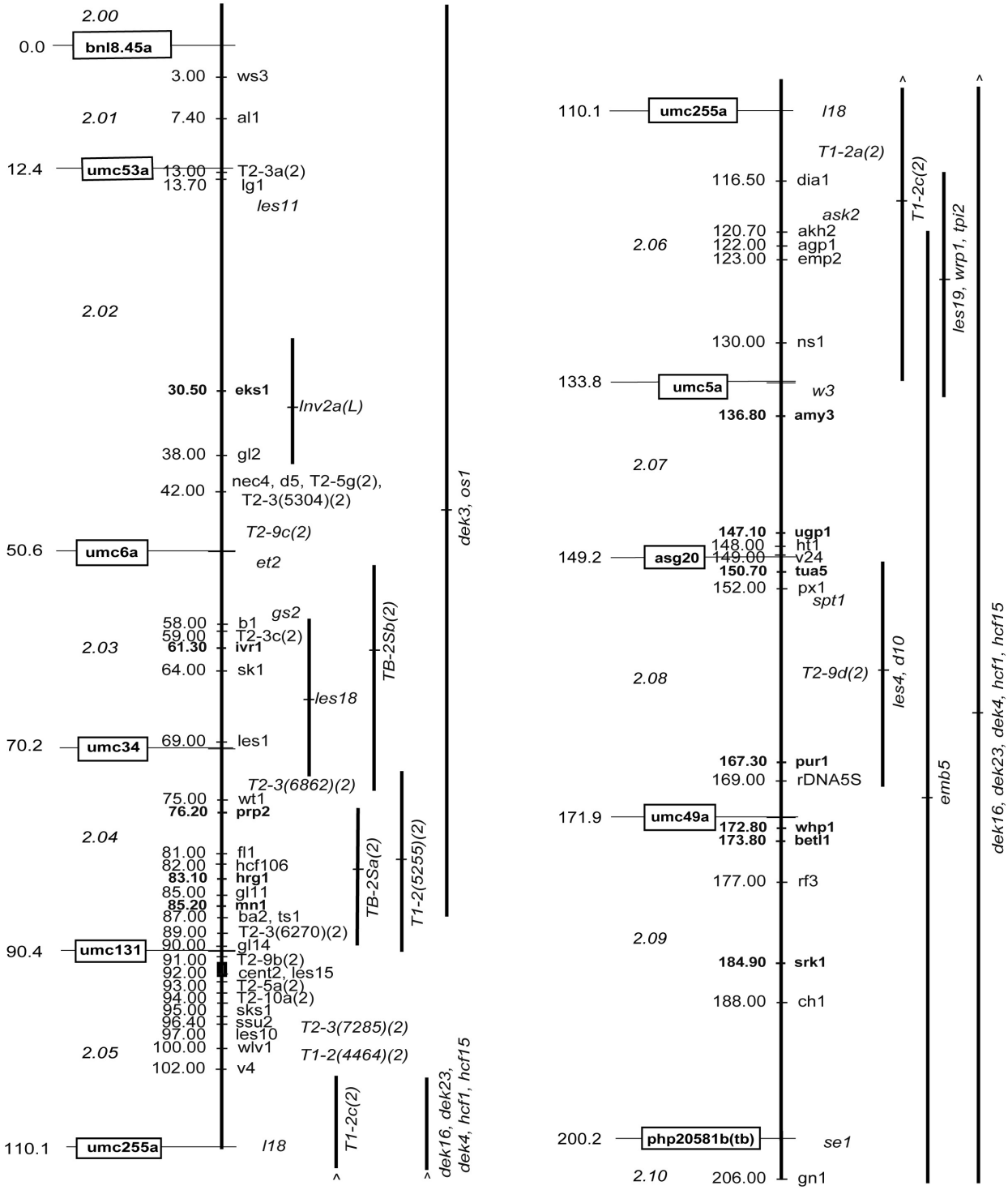


FIGURE 1 - Maps representing the locations of genes on the 10 linkage groups of maize, prepared from retrospective data and from mapping with molecular markers. Tabulations of the map locations and degrees of precision of placement are presented in MaizeGDB (http://www.maizegdb.org), where supporting data are provided as comments for each mapped locus. Version: May 1, 2005.

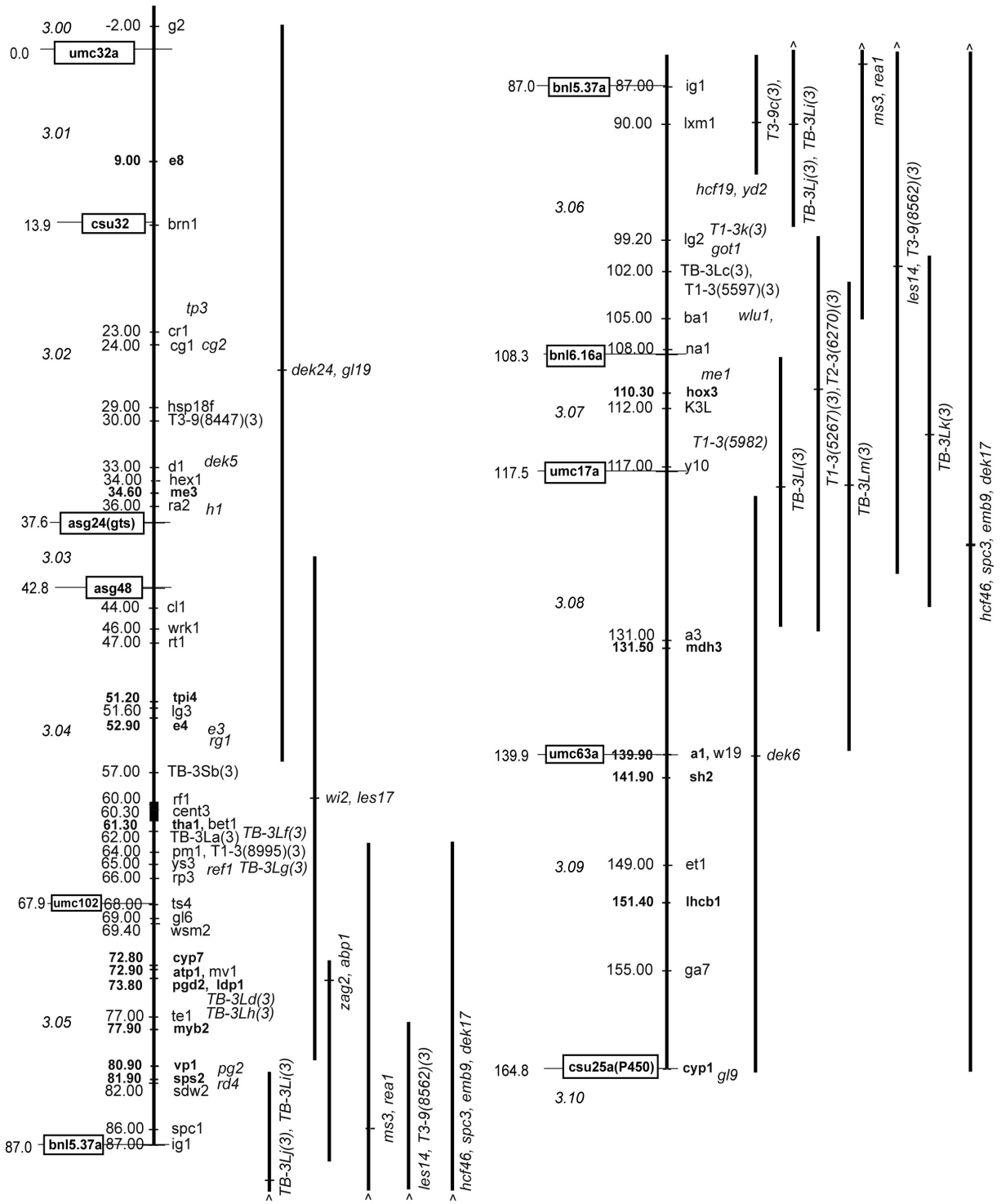
Genetic Map 2005 Chromosome 2

(Loci at tic-marked coordinates are in order according to the best supporting data; **loci in bold** are mapped on UMC98; *loci in italics* likely fall within 5cM of the site where they are shown; and loci associated with vertical bars have a greater range of uncertainty)



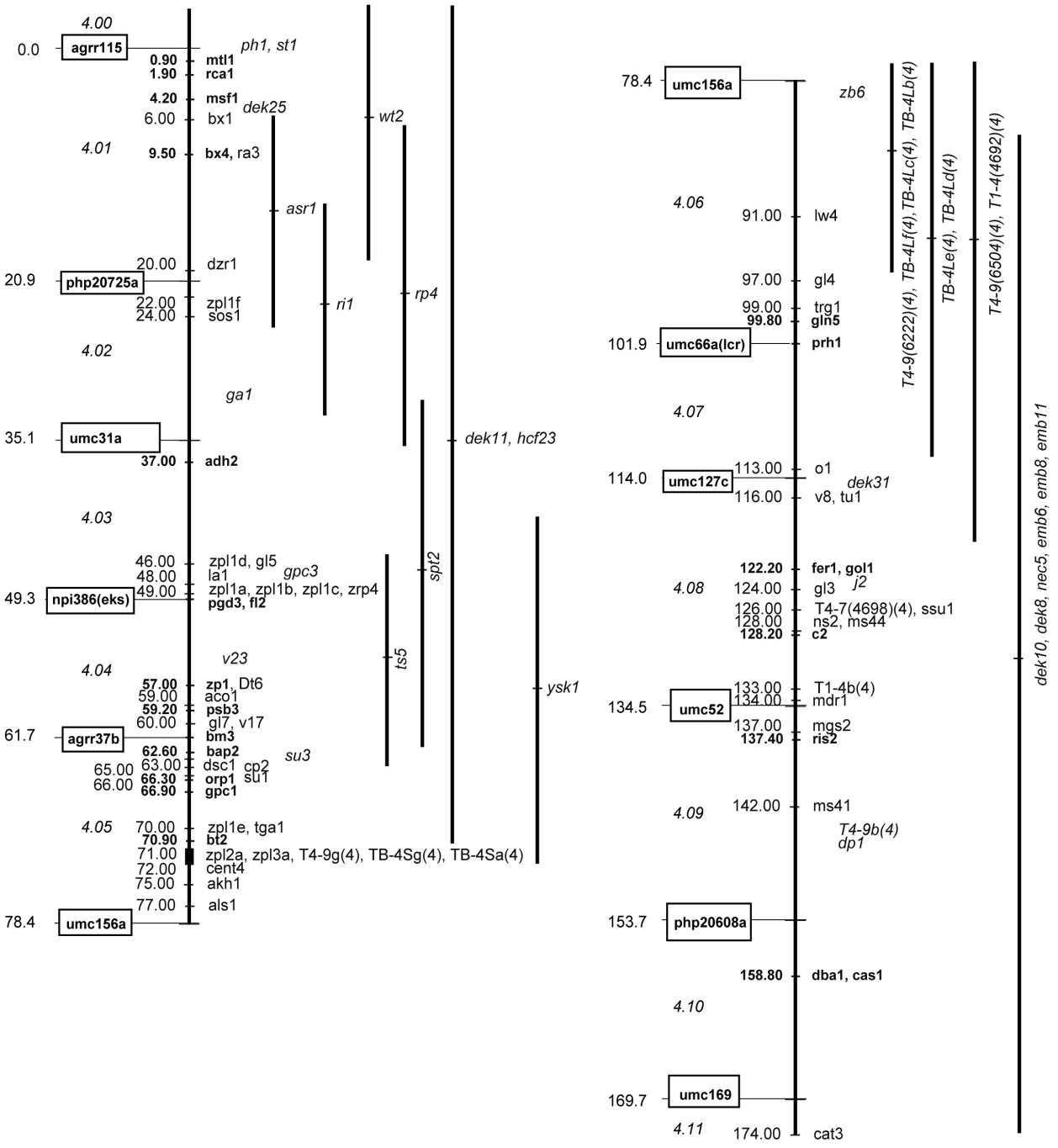
Genetic Map 2005 Chromosome 3

(Loci at tic-marked coordinates are in order according to the best supporting data; **loci in bold** are mapped on UMC98; *loci in italics* likely fall within 5cM of the site where they are shown; and loci associated with vertical bars have a greater range of uncertainty)



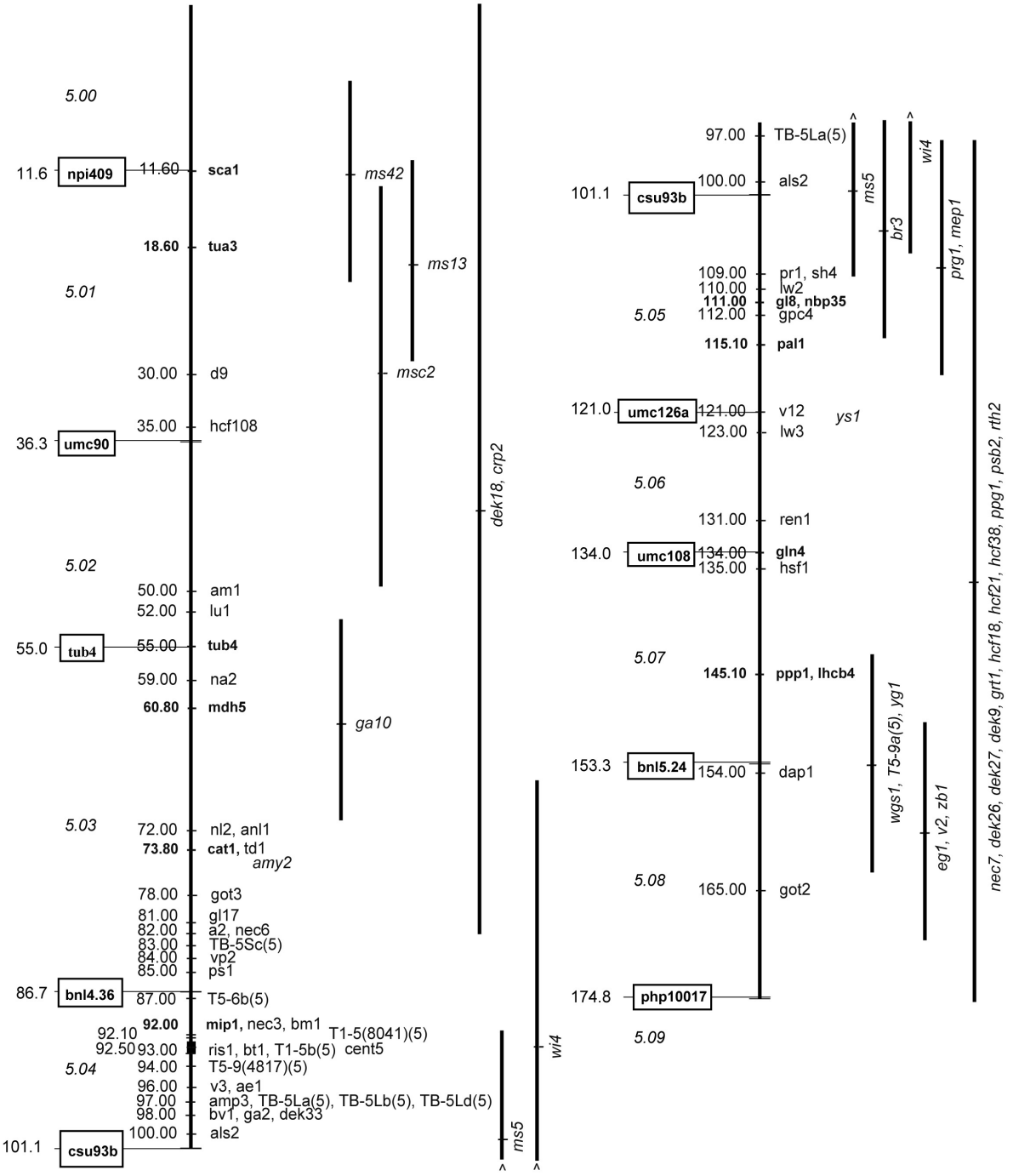
Genetic Map 2005 Chromosome 4

(Loci at tic-marked coordinates are in order according to the best supporting data; **loci in bold** are mapped on UMC98; *loci in italics* likely fall within 5cM of the site where they are shown; and loci associated with vertical bars have a greater range of uncertainty)



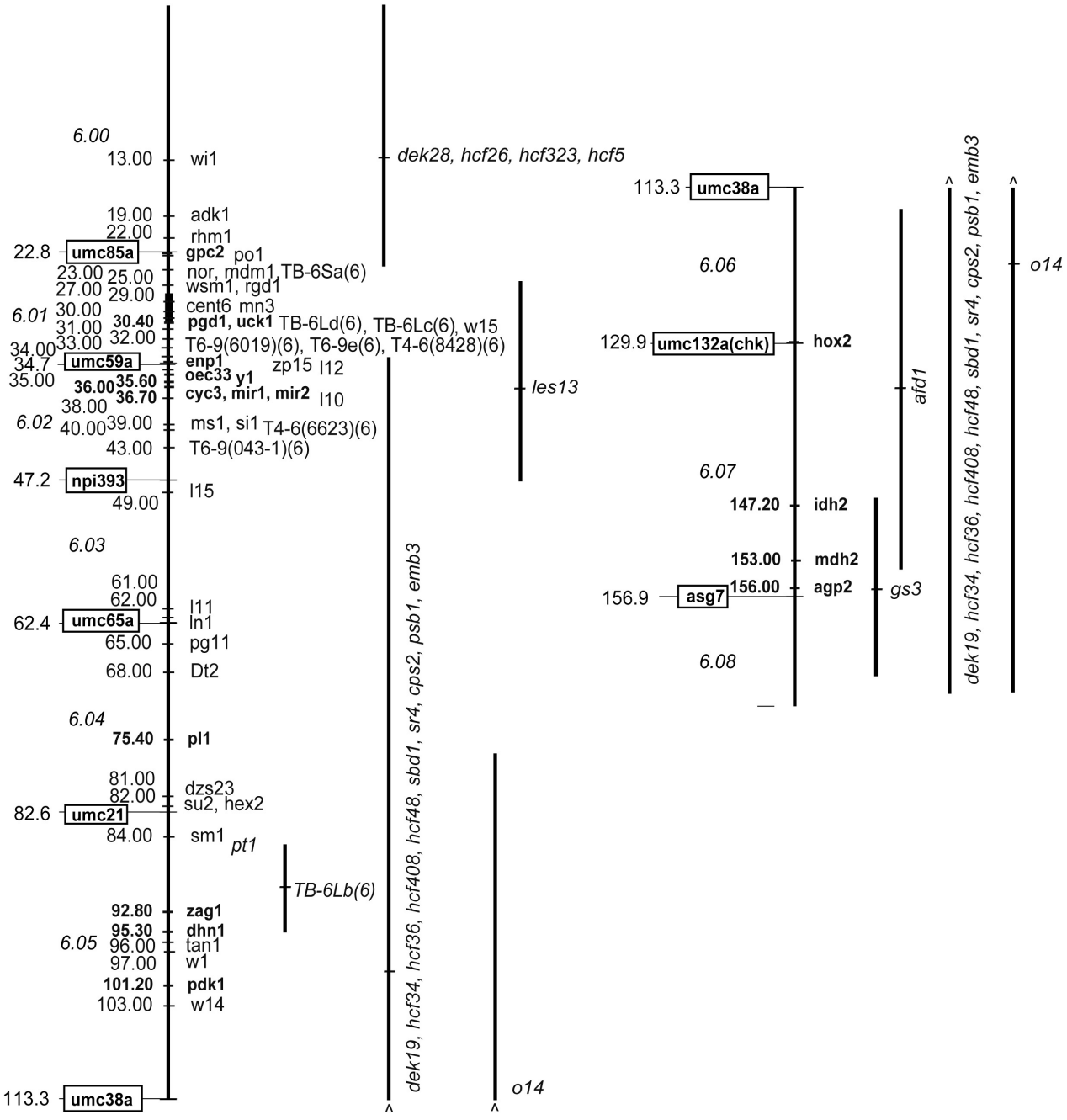
Genetic Map 2005 Chromosome 5

(Loci at tic-marked coordinates are in order according to the best supporting data; **loci in bold** are mapped on UMC98; *loci in italics* likely fall within 5cM of the site where they are shown; and loci associated with vertical bars have a greater range of uncertainty)



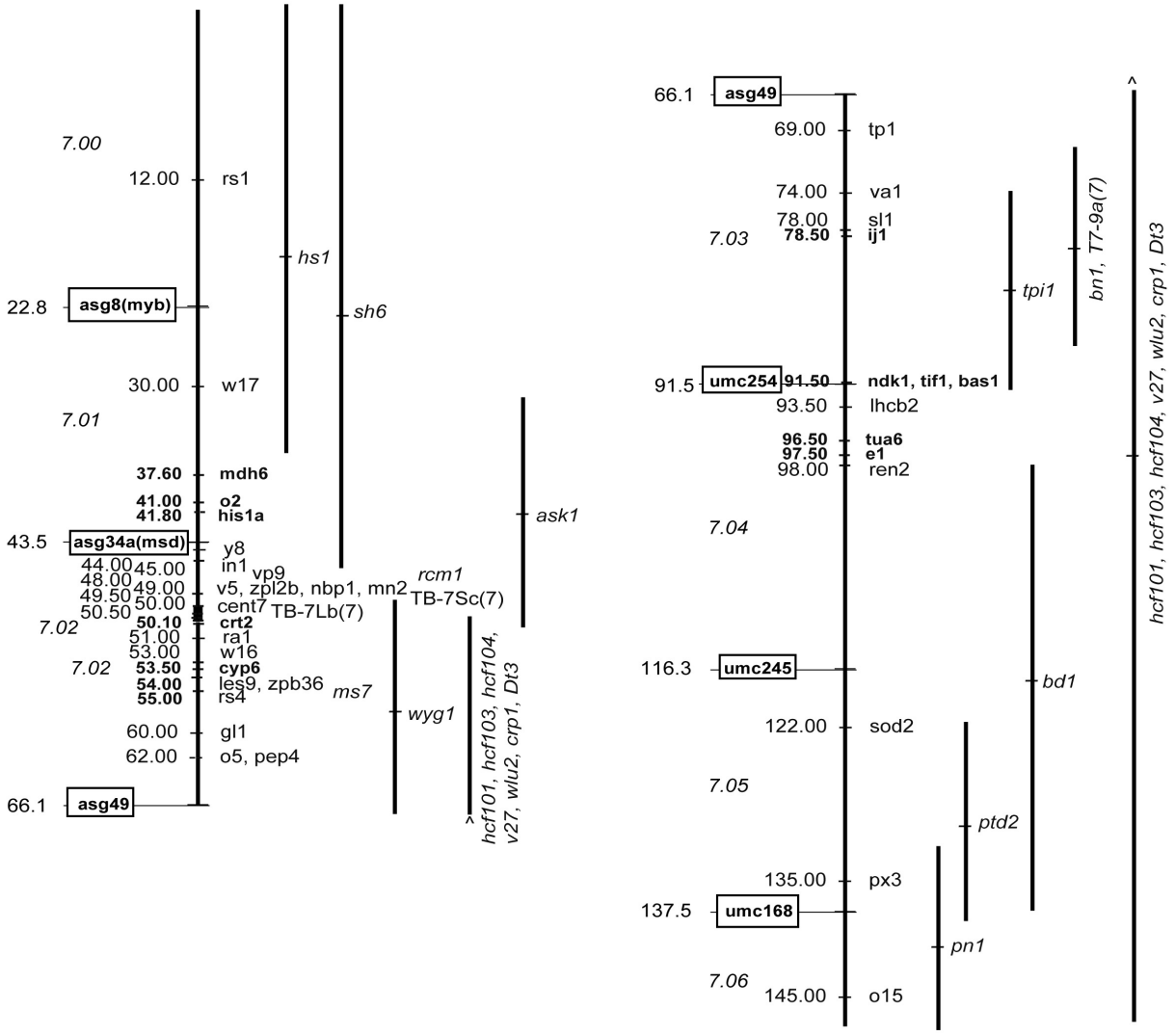
Genetic Map 2005 Chromosome 6

(Loci at tic-marked coordinates are in order according to the best supporting data; **loci in bold** are mapped on UMC98; *loci in italics* likely fall within 5cM of the site where they are shown; and loci associated with vertical bars have a greater range of uncertainty)



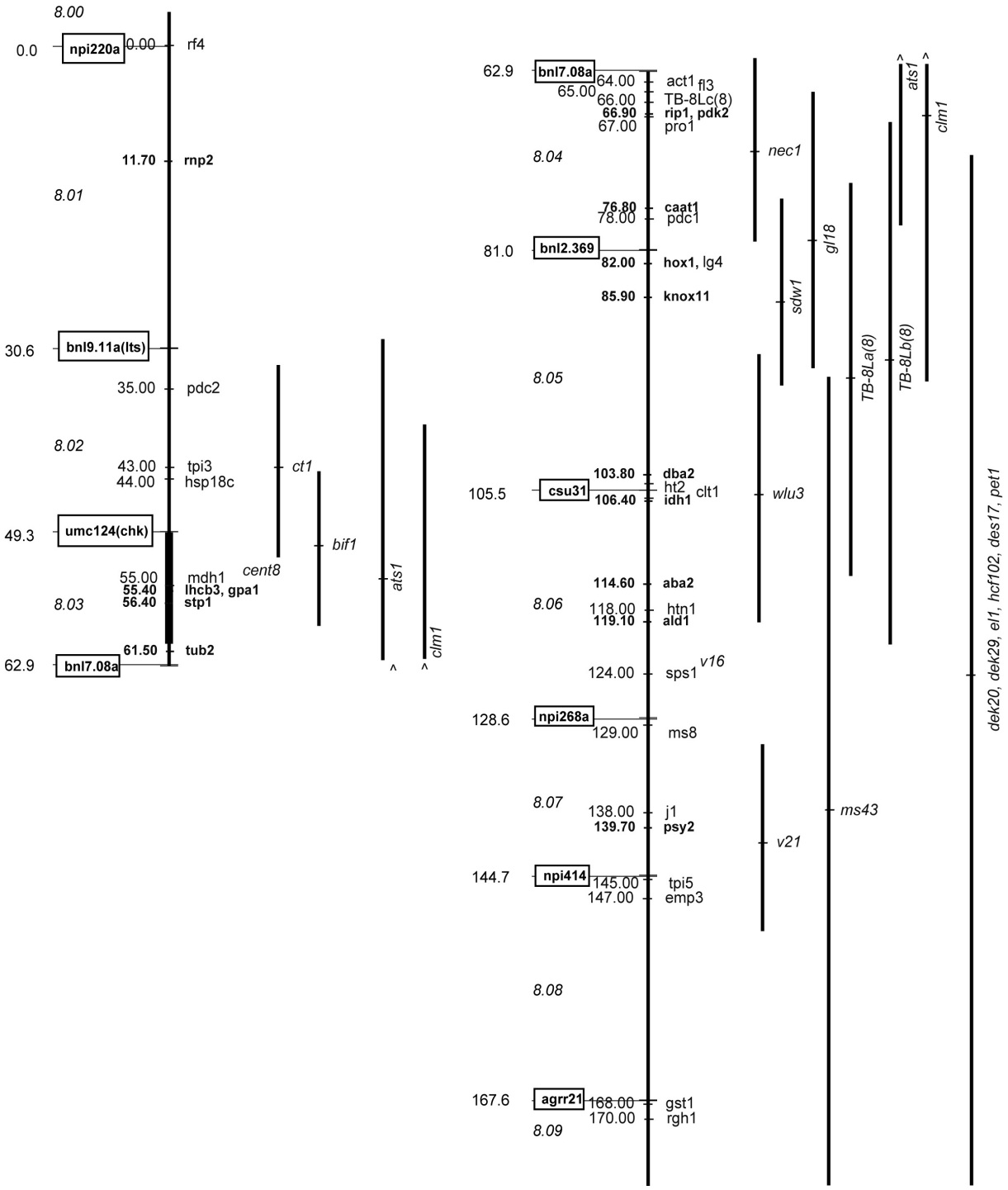
Genetic Map 2005 Chromosome 7

(Loci at tic-marked coordinates are in order according to the best supporting data; loci in bold are mapped on UMC98; loci in italics likely fall within 5cM of the site where they are shown; and loci associated with vertical bars have a greater range of uncertainty)



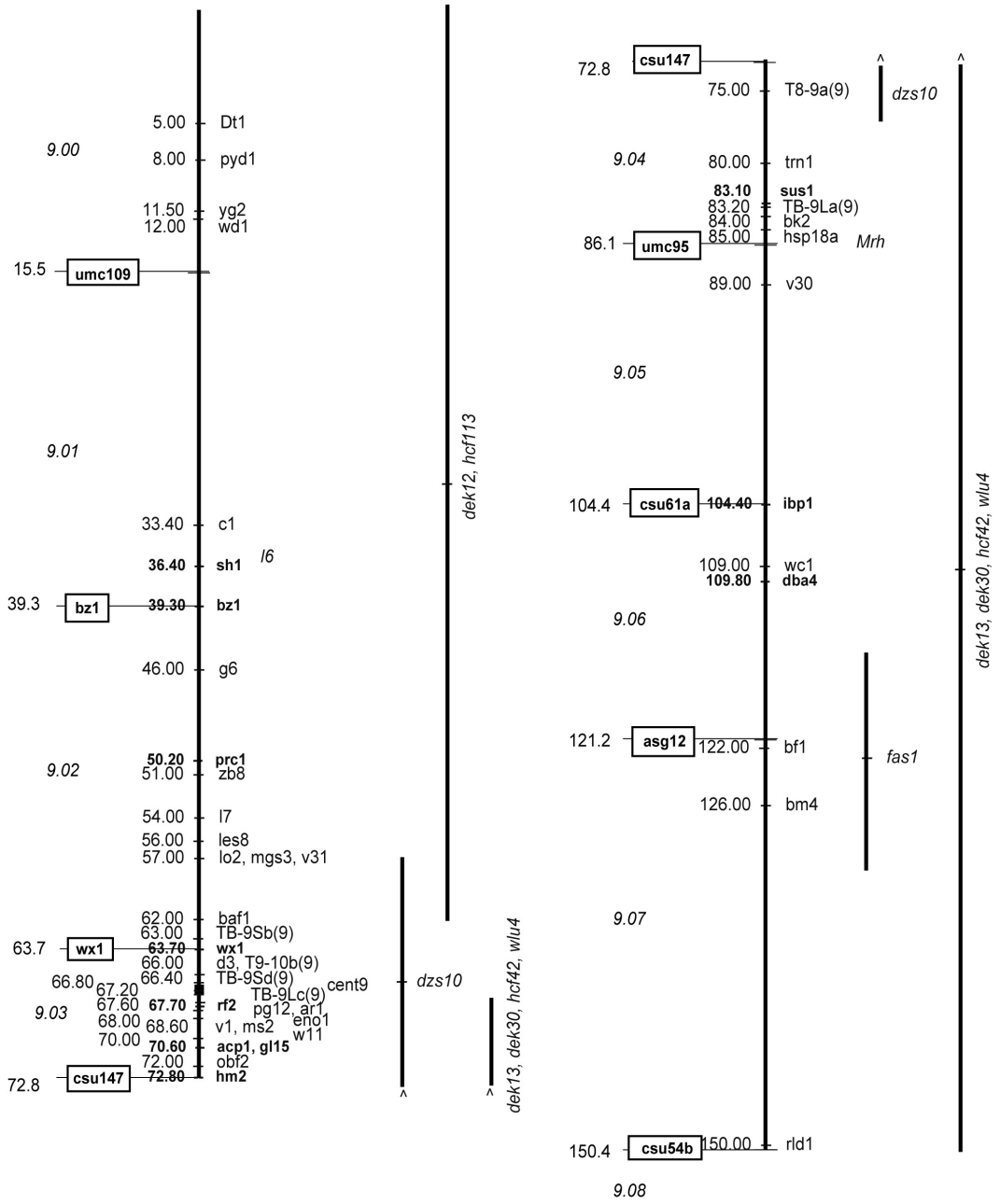
Genetic Map 2005 Chromosome 8

(Loci at tic-marked coordinates are in order according to the best supporting data; **loci in bold** are mapped on UMC98; *loci in italics* likely fall within 5cM of the site where they are shown; and loci associated with vertical bars have a greater range of uncertainty)



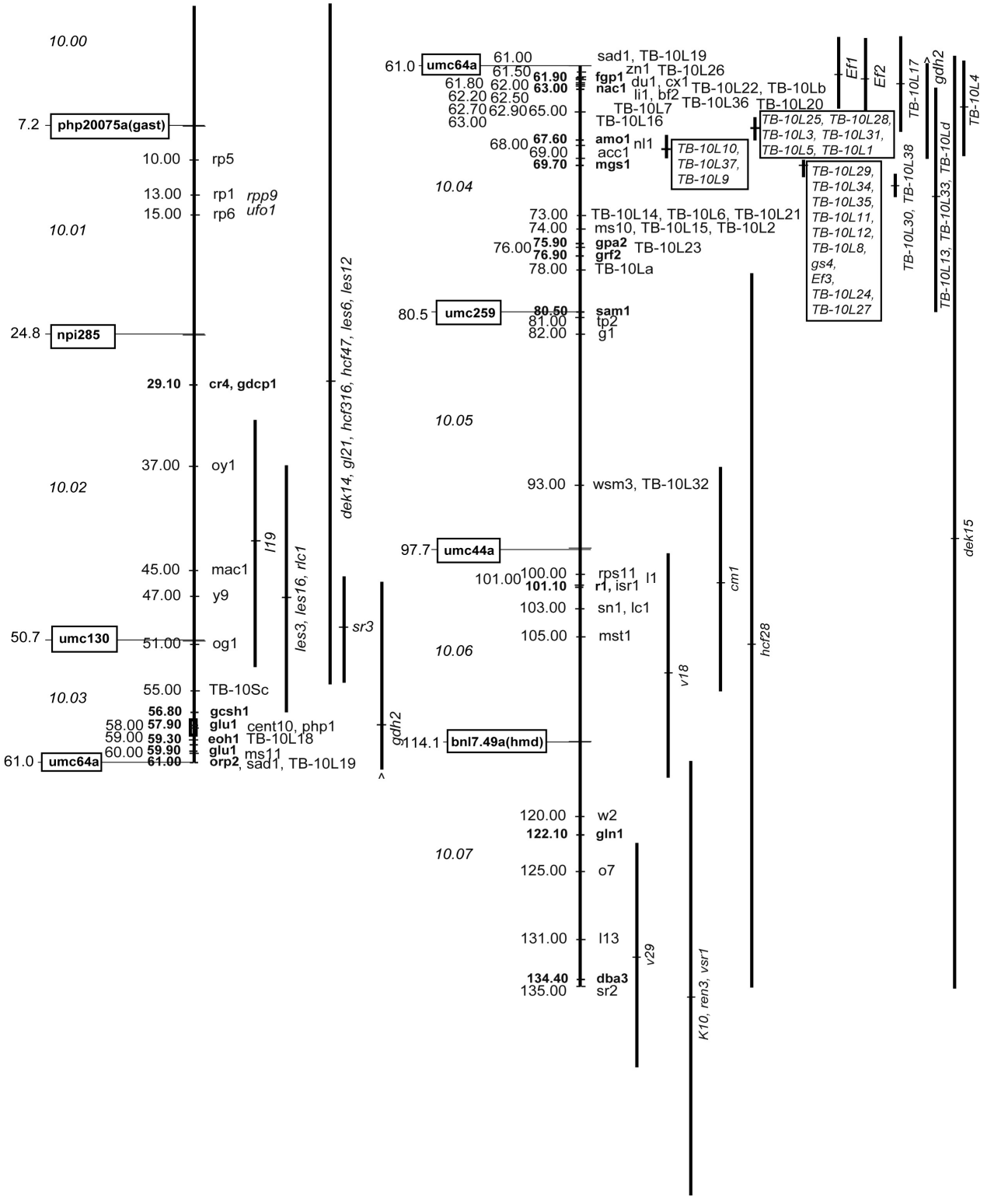
Genetic Map 2005 Chromosome 9

(Loci at tic-marked coordinates are in order according to the best supporting data; **loci in bold** are mapped on UMC98; *loci in italics* likely fall within 5cM of the site where they are shown; and loci associated with vertical bars have a greater range of uncertainty)



Genetic Map 2005 Chromosome 10

(Loci at tic-marked coordinates are in order according to the best supporting data; **loci in bold** are mapped on UMC98; *loci in italics* likely fall within 5cM of the site where they are shown; and loci associated with vertical bars have a greater range of uncertainty)



cal map, i.e., *genomic segments coalesced physically with each other and anchored to the genetic map*. Planning phases and progress of the MMP have been reported (COE *et al.*, 2002; CONE *et al.*, 2002), and have been communicated in a number of meeting abstracts, posters, and presentations. Advancement required integration of three components: (i) informatics infrastructure, (ii) high-resolution genetic mapping, and (iii) physical assembly. The human genome assembly and sequencing effort exemplified what might be accomplished with such a large, complex genome (INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM, 2001), but potential obstacles inherent in the maize genome were expected due to ancient genomic duplication and large numbers of nested retrotransposons (BENNETZEN *et al.*, 2001). At the outset of the MMP the feasibility of the goal of assembling physical segments, i.e., bacterial artificial chromosomes; the scale of each component; and the criteria and standards could be weighed only approximately. Experience gained progressively, with advice from an Advisory Committee, from other colleagues, and from members of the Maize Genetics Community, contributed to the success of the effort, described below. The products and data of the MMP are now being used toward sequencing the genome.

Mapping with molecular markers for genetic studies and for trait analyses

The first genetic maps involving molecular markers were carried out with small populations of F2 or recombinant-inbred individuals. The Brookhaven maps (BURR *et al.,* 1988, 1991) were constructed with 48 and 41 lines combined for analysis from two pairs of parents; and the University of Missouri maps (GARDINER *et al.,* 1993; DAVIS *et al.*, 1999) were constructed with 54 F2 individuals from one of the same pairs of parents. Internationally, studies were conducted on trait variations in the Brookhaven populations (e.g., OTTAVIANO *et al.,* 1991). In the private sector, some early maps were based on the Brookhaven populations and on others (e.g., MURRAY *et al.,* 1989).

As increasingly reliable molecular markers have become available, extensive experiments are conducted with a diverse selection of populations based on germplasms targeted on mapping of quantitative trait loci (QTL) or of simply inherited pest resistances. Application of markers has opened the way to impressive and effective collaborations internationally in the study of pest and stress responses. As a result of QTL studies, marker assisted selection (MAS) in crop improvement has become the norm for effective advancement. The number of distinct maps exceeds 100 in MAIZEGDB, and more studies are published regularly.

Genetic maps for genomics

To serve as the foundation for an integrated genetic and physical map, high resolution and statistical precision is required in order to separate loci by recombination at the scale of a few BAC lengths. The Intermated B73 x Mo17 (IBM) population (LEE *et al.*, 2002) contains over 300 recombinant inbred lines derived from random intermating before selfing to produce recombinant inbreds. The procedure approaches 4-fold expansion per line (WINKLER *et al.*, 2003), and relative to prior sets has a resolution of 0.05 cM, nearly 18 times the resolving power of the map derived from immortal F2 lines. The map was constructed in the MMP with data from several public and private-sector collaborators, with additional low-copy number restriction fragment length polymorphism (RFLP) markers, simple sequence repeat (SSR) markers, and single-nucleotide polymorphisms (SNPs) and insertion-deletion markers (InDels). Its resolution is a dependable gold standard to anchor contigs and to orient them on the genetic map. The resolution places most loci within a contig in the same order as their physical position in the contig. The most recent version of the map, IBM2, with over 2000 markers and map scores for them, is publicly available through MAIZEGDB. It is the primary foundation on which the contigs of the physical map, described below, have been anchored.

Markers from multiple additional mapping populations have been intercalated into a composite "neighbors" map based on shared markers. The precision of placement of intercalated markers is limited by the statistical precision of the added data and by gaps in shared markers, so this map has less-strong statistical support for locus order. The map continues to be revised as new genetic data are accumulated and as information on marker order is garnered from the growing physical map resource. Additional SSR markers developed and mapped on other populations (e.g., SHAROPOVA *et al.*, 2002) have been intercalated, as have large numbers of RFLP loci. The composite map, IBM2 2004 Neighbors, includes nearly 5900 markers. Current versions, IBM2 2004 Neighbors, and the map supported with physical data, IBM2 FPC Neighbors, are available through

MAIZEGDB and iMap at the MAIZE MAPPING PROJECT website (2005).

The community mapping service CIMDE, developed in the MMP, provides a research tool and resource. A subset of 94 lines from the IBM population can be probed by any user. Genotyping data are entered by the user and are submitted for map analysis via the Community IBM Map Data Entry Tool (CIMDE, at the MAIZE MAPPING PROJECT website). With a framework of 250 marker loci, maps for submitted probe data are calculated and returned promptly to individual users. With permission of the users, new data are added to the cumulative public database of the cIBM map posted in MAIZEGDB. The cIBM map has increased to 541 loci by the addition of volunteered, community data, and it is a significant contributor to IBM Neighbors. The resolution, comparable to an RI population of 300-350 individuals, is sufficient to support placement on the physical map. Seed for the IBM-94 population can be ordered through the Maize Genetics-Cooperation Stock Center, or kits containing DNA from the 94 lines in 96-well format, sufficient for 100 PCR reactions, can be ordered from the MMP website.

In all, nearly 8,000 probed loci have been mapped genetically on one or more maps, of which nearly 5,900 are included in IBM Neighbors maps. Incorporation of additional probed sites that have been placed on anchored BAC contigs is in progress, and can be expected to raise the number of markers in IBM Neighbors above 20,000.

Physical maps for genomics

For the U.S. public genomic project, BAC libraries were initiated in 1998-1999. Three BAC libraries were constructed from the B73 inbred line, one of the two IBM mapping parents, for use in generating a physical map. The specific germplasm source for the B73 inbred line, PI 550473, is continuously and reliably available from the USDA North Central Regional PI Station at Ames, Iowa, and can be ordered through the Stock link at MAIZEGDB.

Three different restriction enzymes were used to construct the libraries, to ensure comprehensive genome coverage. The *Hin*dIII library of 247,680 clones, ZMMBBb, constructed at CLEMSON UNIVERSITY GENOMICS INSTITUTE (CUGI) (2005), has an average insert size of 137 kB and a genome coverage of ~17X. Characterizations showed this library to be representative and to be virtually free of chloroplast sequences (TOMKINS *et al.*, 2002). The *Hin*dIII library is available from the ARIZONA GENOMICS INSTITUTE (AGI) (2005) or from CUGI. An *Eco*RI BAC library, ZMMBBc (CHORI 201 segment 1), constructed at CHILDREN'S HOSPITAL OAKLAND RESEARCH INSTITUTE (CHORI), has an average insert size of 163 kB and a genome coverage of ~6.9X. An *Mbo*I library, ZMMBBc (CHORI 201 segment 2), constructed at CHORI in collaboration with Jo Messing, has an average insert size of 167 kB and a genome coverage of ~7X. The *Eco*RI and *Mbo*I libraries are described by and can be ordered from CHORI. Hybridization filters for all three libraries are available from CUGI, AGI, and CHORI.

BACs from all three libraries were fingerprinted using a standard agarose-based method (MARRA *et al.*, 1997). Fingerprints were acquired for BACs representing 17X genome coverage. FPC (SODERLUND *et al.*, 1997, 2000) was used for auto-assembly of BACS into contigs, using standard criteria to judge overlaps between clones. Contigs with questionable (Q) clones were re-evaluated using the DQer function, which uses higher stringency criteria to split these chimeric contigs.

The BAC filters were hybridized with four sets of probes. To place previously mapped standard RFLP markers, 6X-filters of the *Hin*dIII library were screened with 96 RFLP probes representing one to 10 copies in the genome (YIM *et al.*, 2002). The RFLP probes showed a strong correlation and linear fit between copy number and number of BACs hit. Selected complex probes were used to screen 6X of *Hin*dIII, 7X of *Eco*RI, and 7X of *Mbo*I. The probes included four centromere- and two telomere-associated repeats, ribosomal DNA, the 185-bp heterochromatic-knob repeat, the mitochondrial genome, a multifragment chloroplast DNA probe, and lambda phage. Low organellar and lambda contamination was found for all three libraries. The complex nuclear probes characterize the libraries and provide potential addresses for resolving BAC contigs to the targeted regions of the genome (YIM *et al.*, 2002). In a collaboration with DuPont, over 10,600 probes derived from EST sequences were screened on 13X BAC filters with 40-bp overgo probes (GARDINER *et al.*, 2004), which yielded 165,254 BAC addresses from 9,371 probes (18 per probe). In addition, 2,034 overgos derived from sequences that have been genetically mapped in maize, rice, sugarcane, and sorghum (e.g., DRAYE *et al.*, 2001), applied to 13X BAC filters, yielded 85,000 BAC addresses (42 per probe).

A highly efficient BAC screening strategy used

pools of BAC DNA probed with PCR primer pairs from single-copy genetically mapped sequences. BACS from a portion of the *Hin*dIII library (representing 5X genome equivalents) were arrayed in a six-dimensional 48 x 48 x 48 matrix to produce 288 pools according to the strategy employed previously for sorghum (KLEIN *et al.*, 2000). Using the pools, 2,049 markers, including 1,527 mapped SSRs, were addressed. The success rate for addressing BACs in pools with PCR-based markers was 90%; thus, this method proved to be extremely powerful and often provided evidence to disambiguate results obtained with other methods.

Genetically mapped markers, addressed to BACs by hybridization or PCR-based screening of BAC pools, allowed the contigs to be ordered and anchored to chromosomes. The genetic maps used for anchoring are, first, the IBM2 map, supplemented by the IBM2 2004 Neighbors map. FPC can incorporate both framework markers (for which order is well-established, i.e., from IBM2) and approximately placed markers (e.g., from IBM2 2004 Neighbors). In this context, the IBM-mapped markers serve as the framework and the intercalated markers (neighbors) serve as placements. The genetic map with parallel contigs in iMap at the MMP website, and contigs in WebFPC at the AGI website, afford visual representations accompanied by access to underlying data.

In January 2003, after contig auto-assembly was complete, manual editing began to merge previously distinct contigs and to break apart contigs that had been inappropriately assembled. Manual editing uses FPC utilities to judge contig merges. Marker hits to BACs have been an important criterion in making the judgments. Hybridization hits on BACs by genetically mapped markers have served a critical role in assigning chromosomal positions to contigs that did not previously have a map location, in orienting contigs, and in verifying or correcting the positions of automatically anchored contigs. In general, the automatic functions of FPC work well for assigning contigs to chromosomes, but they are conservative. Knowledgeable manual editors must then assign more contigs to chromosomes. For example, for the February 2004 release of the maize FPC assembly, FPC assigned 535 contigs to positions on chromosomes; only 6 of these were later found to be incorrect by manual editing. In contrast, approximately 200 contigs that were not automatically anchored to the genetic map by FPC were newly assigned to chromosomal positions by manual editing.

The major release of October 15, 2004, applied markers, high-information-content fingerprinting, and synteny with the rice pseudochromosomes to edit the assemblies derived by FPC from agarose fingerprinting. Assemblies from 292,201 BACs have merged to 760 contigs, 618 of which include more than 2 BACs (93% of the genome) and 460 include more than 100 BACs; 414 of the contigs have been assigned to chromosomes (80% of the genome). A very large part of the genome of B73 is represented in substantial, contiguous assemblies rather than in dispersed, small islands separated by repetitive element sequences that interfere with assembly. Minimal tiling paths have been chosen for a selection of contigs (accessed through the FPC link at the AGI website), and these BACs are undergoing sequencing in current projects. Composited physical and genetic data integration is also presented at the FPC link of the AGI website. Synteny alignments with the rice genome are presented at the website of the ARIZONA GENOMICS COMPUTATIONAL LABORATORY (AGCoL) (2005).

Through strategies designed to filter the genome and enrich for gene sequences (WHITELAW *et al.,* 2003), an extensive resource of gene-rich sequences has been produced, deriving genome sequences that encompass over 85% of the expressed genes represented in GenBank [THE INSTITUTE FOR GENOMIC RESEARCH (TIGR) 2005]. Information and downloads are also available at the TIGR website and at the website of the DANFORTH CENTER MAIZE GENOME SEQUENCING PROJECTS (2005).

Informatics for genomics

Automated data management is essential to handle the demands of high-throughput data acquisition and analysis, and was developed with the MMP. Tools for internal data management and analysis, pipelines, and presentation were derived and have been made available for other users. The tools include a Laboratory Information Management System (LIMS; SANCHEZ-VILLEDA *et al.*, 2003) to serve multiple purposes: locally, forms for entry of multiple types of data, with self-checks for accuracy; ability to compile and exchange data among the participating laboratories to track and evaluate the pace of the project; and ability to compile and export data to Web-based applications for displaying data for public view. Among the presentation applications is iMap, the integrated map tool at the MMP website, which shows genetic maps in parallel with assembled and anchored BAC contigs (FANG *et al.*, 2003), and accesses associated data.

TABLE 2 - *Resources derived from the Maize Mapping Project.*

Resource	Description	URL
Integrated map	2,000 markers mapped at high resolution plus 3,700 "neighbor" loci, matched to anchored contigs (433 by iMap, 555 by FPC)	http://www.maizemap.org/iMapDB/iMap.html
Markers on BACs	Associations with 3,186 distinct loci	http://www.maizemap.org/MMP_Downloads/
BAC contigs	292,216 BACs, 15X, in 760 contigs as of Oct. 25, 2004	http://genome.arizona.edu/fpc/maize/
BAC libraries and filters	*Hin*dIII (17X), *Eco*RI (7X), and *Mbo*I (7X)	http://genome.arizona.edu/orders/
IBM population	Seed of 300 lines, DNA kits of 94 for mapping	http://www.maizemap.org/resources.htm
SSR primers	1,400 - Sequence, GenBank ID, physical and genetic locations	http://maizegdb.org/
RFLP probes	Sequence, GenBank ID, physical and genetic locations	http://maizegdb.org/
SNP and InDel probes	240 - Sequence, GenBank ID, physical and genetic locations	http://maizegdb.org/
Overgo primers	13,400 - Sequence, GenBank ID, physical and genetic locations	http://maizegdb.org/
EST unigenes ("Cornsensus")	10,600 - Sequence, GenBank ID, physical and genetic locations	http://maizegdb.org/
LIMS	Entry of multiple types of genetic and molecular data with self-checks for accuracy, data exchange, tracking, and compilation and export; open source	http://www.maizemap.org/bioinformatics.htm
iMap	Integrated map utility for development and for presentation	http://www.maizemap.org/bioinformatics.htm
CIMDE	Community mapping utility for data submission and receipt of maps	http://www.maizemap.org/bioinformatics.htm
Data linkages	MaizeGDB, TIGR, Gramene, and Maize WebFPC	http://maizegdb.org/ and others
Gene maps	720 genes with documentation of sources from retrospective and molecular mapping	http://maizegdb.org/

For assembling the BAC contigs, FPC software developed for the human genome project, now centered at AGCoL, automatically accumulates fingerprinted BACs by pattern overlaps and assembles contigs. Sequenced clones are automatically added by performing a simulated digest using FSD (FPC Simulated Digest; ENGLER *et al.*, 2003). Contigs are displayed on the Web using WebFPC and WebChrom with access to associated data.

Data, information, and software are generic and are transferable for application by other users in the future. Source codes for LIMS and iMap are publicly available at the MMP website. The FPC programs are already in wide use for genome assembly in many species.

The present and the future

The products and resources made available in the U.S. Maize Mapping Project are summarized in Table 2. These include genetic maps, the physical map, and integrated views of the two. They also include newly developed markers and sequence information, BAC libraries, software, and information that can be incorporated by other genomics pro-

jects. Additional markers and increasing numbers of genomic sequences, including BAC ends and fully sequenced BACs, continue to be developed in several laboratories.

Although multiple strategies for sequencing the maize genome have been tested and proposed, all agree that the end goal should include not only sequences of all the genes, but also ordering of those genes to the genetic map (CHANDLER and BRENDEL, 2002; JORGENSEN, 2004). Sequences of the genes and ordering of them both require and depend upon an integrated genetic and physical map resource, which is provided in the sequence-ready assembly derived in the MMP. Information on sequencing projects and data are provided at the websites of the RUTGERS PLANT GENOME INITIATIVE, TIGR, the DANFORTH CENTER MAIZE GENOME SEQUENCING PROJECTS, and the Maize Sequencing link at AGI.

NOTE ADDED IN PROOF: November 15, 2005, the U.S. National Science Foundation announced (http://www.nsf.gov/news/news_summ.jsp?cntn_id=104608&org=BIO&from=news) a three-agency, multi-institution award of $32 million to Washington University to sequence the corn genome (http://www.nsf.gov/awardsearch/showAward.do?AwardNumber=0527192).
The major maize FPC release of 19 July 2005 merged 292,519 BACs in 721 contigs. Sequenced BACs are automatically added to FPC when they become available in GenBank. As of 18 November 2005, 3,694 sequenced BACs were in GenBank.

ACKNOWLEDGMENTS - Members of the Advisory Committee for the MMP were Sue Wessler (Chair), Brad Barbazuk, Vicki Chandler, Joe Ecker, Stan Letovsky, and Antoni Rafalski. We are truly grateful for their insights and guidance in advancing and strengthening the project.

REFERENCES

ARIZONA GENOMICS COMPUTATIONAL LABORATORY (AGCoL), 2005 http://www.agcol.arizona.edu/ (accessed 26 April 2005).

ARIZONA GENOMICS INSTITUTE (AGI), 2005 http://genome.arizona.edu/(accessed 26 April 2005).

BENNETZEN J.L., V.L. CHANDLER, P. SCHNABLE, 2001 National Science Foundation-sponsored workshop report. Maize Genome Sequencing. Plant Physiol. **127:** 1572-1578.

BURR B., F. BURR, 1991 Recombinant inbreds for molecular mapping in maize - theoretical and practical considerations. Trends Genet **7:** 55-60.

BURR B., F. BURR, K.H. THOMPSON, M.C. ALBERTSEN, C.W. STUBER, 1988 Gene mapping with recombinant inbreds in maize. Genetics **118:** 519-526.

CHANDLER V., V. BRENDEL, 2002 The maize genome sequencing project. Plant Physiol. **130:** 1594-1597.

CHILDREN'S HOSPITAL OAKLAND RESEARCH INSTITUTE (CHORI), 2005 http://bacpac.chori.org/maize201.htm (accessed 26 April 2005).

CIMMYT SEED INSPECTION AND DISTRIBUTION UNIT, 2005 http://148.223.253.105/ (accessed 26 April 2005).

CLEMSON UNIVERSITY GENOMICS INSTITUTE (CUGI), 2005 https://www.genome.clemson.edu/ (accessed 26 April 2005).

COE E., 1998 Potentials of the National Corn Genome Initiative. Proc. Natl. Acad. Sci. USA. **95:** 2029-2032.

COE E., 2005 Genetic 2005 maps. http://maizegdb.org/cgi-bin/displaymaprecord.cgi?id=940880 through 940889 (accessed 26 April 2005).

COE E., K.C. CONE, M. MCMULLEN, S.S. CHEN, G. DAVIS, J. GARDINER, E. LISCUM, M.L. POLACCO, H. SANCHEZ-VILLEDA, Z. FANG, S. SCHROEDER, S. HAVERMANN, J. E. BOWERS, A.H. PATERSON, C.A. SODERLUND, R. WING, 2002 Access to the maize genome: an integrated physical and genetic map. Plant Physiol. **128:** 9-12.

CONE K.C., M. MCMULLEN, I. VROH BI, G. DAVIS, Y.S. YIM, J. GARDINER, M.L. POLACCO, H. SANCHEZ-VILLEDA, Z. FANG, S. SCHROEDER, S. HAVERMANN, J. E. BOWERS, A.H. PATERSON, C.A. SODERLUND, F. ENGLER, R. WING, E.H. COE, 2002 Genetic, physical, and informatics resources for maize. On the road to an integrated map. Plant Physiol. **130:** 1598-1605.

DANFORTH CENTER MAIZE GENOME SEQUENCING PROJECTS, 2005 http://www.maizegenome.org/ (accessed 26 April 2005).

DAVIS G.L. M. MCMULLEN, C. BAYSDORFER, T. MUSKET, D. GRANT, M.S. STAEBEL, G. XU, M.L. POLACCO, L. KOSTER, S. MELIA-HANCOCK, K. HOUCHINS, S. CHAO, E. COE, 1999 A maize map standard with sequenced core markers, grass genome reference points and 932 expressed sequence tagged sites, ESTs in a 1736-locus map. Genetics **152:** 1137-1172.

DRAYE X., Y.D. LIN, X. QIAN, J.E. BOWERS, G. BUROW, P. MORRELL, D.J. PETERSON, G. PRESTING, S. REN, R. WING, A.H. PATERSON. 2001 Toward integration of comparative genetic, physical, diversity, and cytomolecular maps for grasses and grains, using the sorghum genome as a foundation. Plant Physiol. **125:** 1325-1341.

EDWARDS M., C.W. STUBER, J.F. WENDEL, 1987 Molecular-marker-facilitated investigations of quantitative-trait loci in maize. I. Numbers, genomic distribution, and types of gene action. Genetics **116:** 113-125.

EMERSON R.A., G.W. BEADLE, A.C. FRASER, 1935 A summary of linkage studies in maize. Cornell Univ. Agric. Exp. Stn. Memoir **180:** 1-83.

ENGLER F.W., J. HATFIELD, W. NELSON, C.A. SODERLUND, 2003 Locating sequence on FPC maps and selecting a minimal tiling path. Genome Res. **13:** 2152-2163.

FANG Z., K.C. CONE, H. SANCHEZ-VILLEDA, M.L. POLACCO, M. MCMULLEN, S. SCHROEDER, J. GARDINER, G. DAVIS, S. HAVERMANN, Y.-S. YIM, I. VROH BI, E. COE, 2003 iMap: a database-driven utility to integrate and access the genetic and physical maps of maize. Bioinformatics **19:** 2105-2111.

GARDINER J., E. COE, S. MELIA-HANCOCK, D. HOISINGTON, S. CHAO, 1993 Development of a core RFLP map in maize using an Immortalized-F2 population. Genetics **134:** 917-930.

GARDINER J., S. SCHROEDER, M.L. POLACCO, H. SANCHEZ-VILLEDA, Z. FANG, M. MORGANTE, T. LANDEWE, K. FENGLER, F. USECHE, M. HANAFEY, S.V. TINGEY, H. CHOU, R. WING, C.A. SODERLUND, E. COE, 2004 Anchoring 9,3971 maize expressed sequence tagged unigenes to the bacterial artificial chromosome contig map by two-dimensional overgo hybridization. Plant Physiol. **134:** 1317-1326.

GOODMAN M.M., C.W. STUBER, K. NEWTON, 1982 Isozyme loci in maize. pp. 53-60. *In:* W.F. Sheridan (Ed.), Maize for Biological Research. 1982. Plant Molec. Biol. Assoc., Charlottesville, Virginia.

INTERAGENCY WORKING GROUP ON PLANT GENOMES (IWG), 1998 http://www.ostp.gov/NSTC/html/npgireport.html (accessed 26 April 2005).

INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM, 2001 Initial sequencing and analysis of the human genome. Nature **409:** 860-921.

JORGENSEN R.A., 2004 Sequencing maize: just sample the salsa or go for the whole enchilada? Plant Cell **16:** 787-788.

KLEIN P.E., R.R. KLEIN, S. CARTINHOUR, P. ULANCH, J. DONG, J. OBERT, D.T. MORISHIGE, S. SCHLUETER, K. CHILDS, M. ALE, J.E. MULLET, 2000 A high-throughput AFLP-based method for constructing integrated genetic and physical maps: progress toward a sorghum genome map. Genome Res. **10:** 789-807.

LEE M., N. SHAROPOVA, W.D. BEAVIS, D. GRANT, M. KATT, D.L. BLAIR, A.R. HALLAUER, 2002 Expanding the genetic map of maize with the intermated B73 x Mo17, IBM population. Plant Mol. Biol. **48:** 453-461.

MAIZE GENETICS COOPERATION STOCK CENTER, 2005 http://w3.ag.uiuc.edu/maize-coop/ (accessed 26 April 2005).

MAIZEGENOME.ORG, 2005 http://www.maizegenome.org/ (accessed 26 April 2005).

MAIZE MAPPING PROJECT (MMP), 2005 http://maizemap.org (accessed 26 April 2005).

MAIZE GENOME DATABASE (MAIZEGDB), 2005 http://www.maizegdb.org (accessed 26 April 2005).

MARRA M.A., T.A. KUCABA, N.L. DIETRICH, E.D. GREEN, B. BROWNSTEIN, R.K. WILSON, K.M. MCDONALD, L.W. HILLIER, J.D. MCPHERSON, R.H. WATERSTON, 1997 High throughput fingerprint analysis of large-insert clones. Genome Res. **7:** 1072-1084.

MURRAY M.G., Y. MA, D.P. WEST, J. ROMERO-SEVERSON, J. CRAMER, J.M. PITAS, J. KIRSCHMAN, S. DEMARS, L. VILBRANDT, 1989 General considerations on building an RFLP linkage map with specific reference to maize. pp. 5-9. *In:* T. Helentjaris, B. Burr (Eds.), Development and Application of Molecular Markers to Problems in Plant Genetics. Cold Spring Harbor Laboratory, New York.

NEUFFER M.G, E. COE, S. WESSLER, 1997 Mutants of Maize. Cold Spring Harbor Laboratory, New York. 468 pp.

OTTAVIANO E., M. SARI GORLA, E. PE, C. FROVA, 1991 Molecular markers (RFLPs and HSPs) for the genetic dissection of thermotolerance in maize. Theor. Appl. Genet. **81:** 713-719.

RUTGERS PLANT GENOME INITIATIVE (PGIR), 2005 http://pgir.rutgers.edu/ (accessed 26 April 2005).

SANCHEZ-VILLEDA H., S. SCHROEDER, M.L. POLACCO, M. MCMULLEN, S. HAVERMANN, G. DAVIS, I. VROH BI, K.C. CONE, N. SHAROPOVA, Y.-S. YIM, L. SCHULTZ, N. DURU, T. MUSKET, K. HOUCHINS, Z. FANG, J. GARDINER, E. COE, 2003 Development of an integrated laboratory information management system for the maize mapping project. Bioinformatics **19:** 2022-2030.

SHAROPOVA N., M. MCMULLEN, L. SCHULTZ, S. SCHROEDER, H. SANCHEZ-VILLEDA, J. GARDINER, D. BERGSTROM, K. HOUCHINS, S. MELIA-HANCOCK, T. MUSKET, N. DURU, M.L. POLACCO, K. EDWARDS, T. G. RUFF, J.C. REGISTER, C.R. BROUWER, R.D. THOMPSON, R. VELASCO, E. CHIN, M. LEE, W. WOODMAN-CLIKEMAN, M.J. LONG, E. LISCUM, K.C. CONE, G. DAVIS, E. COE, 2002 Development and mapping of SSR markers for maize. Plant Mol. Biol. **48:** 463-481.

SODERLUND C., I. LONGDEN, R. MOTT, 1997 FPC: a system for building contigs from restriction fingerprinted clones. Comput. Appl. Biosci. **13:** 523-535.

SODERLUND C., S. HUMPHRAY, A. DUNHAM, L. FRENCH, 2000 Contigs built with fingerprints, markers, and FPC V4.7. Genome Res. **10:** 1772-1787. http://www.agcol.arizona.edu/software/fpc/ (accessed 26 April 2005).

STUBER C.W., R.H. MOLL, 1972 Frequency changes of isozyme alleles in a selection experiment for grain yield in maize (Zea mays L.). Crop Sci **12:** 337-341.

THE INSTITUTE FOR GENOMIC RESEARCH (TIGR), 2005 http://maize.tigr.org/ (accessed 2 May 2005).

TOMKINS J.P., G. DAVIS, D. MAIN, Y.-S. YIM, N. DURU, T. MUSKET, J.L. GOICOECHEA, D. FRISCH, E. COE, R. WING, 2002 Construction and characterization of a deep-coverage bacterial artificial chromosome library for maize. Crop Sci. **42:** 928-933.

USDA, ARS, NATIONAL GENETIC RESOURCES PROGRAM. GERMPLASM RESOURCES INFORMATION NETWORK - (GRIN), 2005 http://www.ars-grin.gov/cgi-bin/npgs/html/crop.pl?89 (accessed 26 April 2005).

WHITELAW C.A., W.B. BARBAZUK, G. PERTEA, A. CHAN, F. CHEUNG, Y. LEE, L. ZHENG, S. VAN HEERINGEN, S. KARAMYCHEVA, J. BENNETZEN, P.J. SANMIGUEL, N. LAKEY, J. BEDELL, Y. YUAN, M.A. BUDIMAN, A. RESNICK, S. VAN AKEN, T. UTTERBACK, S. RIEDMULLER, M. WILLIAMS, T. FELDBLYUM, K. SCHUBERT, R.N. BEACHY, C. FRASER, J. QUACKENBUSH, 2003 Enrichment of gene-coding sequences in maize by genome filtration. Science **302:** 2118-2120.

WINKLER C.R., N. JENSEN, M. COOPER, D. PODLICH, O.S. SMITH, 2003 On the determination of recombination rates in intermated recombinant inbred populations. Genetics **164:** 741-745.

YIM Y.S., G. DAVIS, N. DURU, T. MUSKET, E. LINTON, J. MESSING, M. MCMULLEN, C.A. SODERLUND, M.L. POLACCO, J. GARDINER, E. COE, 2002 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. **130:** 1686-1696.