



## Genetic Properties of the Maize Nested Association Mapping Population

Michael D. McMullen *et al.*  
*Science* **325**, 737 (2009);  
DOI: 10.1126/science.1174320

*This copy is for your personal, non-commercial use only.*

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

**The following resources related to this article are available online at [www.sciencemag.org](http://www.sciencemag.org) (this information is current as of June 6, 2013 ):**

**Updated information and services**, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/content/325/5941/737.full.html>

**Supporting Online Material** can be found at:

<http://www.sciencemag.org/content/suppl/2009/08/06/325.5941.737.DC1.html>

A list of selected additional articles on the Science Web sites **related to this article** can be found at:

<http://www.sciencemag.org/content/325/5941/737.full.html#related>

This article **cites 27 articles**, 18 of which can be accessed free:

<http://www.sciencemag.org/content/325/5941/737.full.html#ref-list-1>

This article has been **cited by** 23 article(s) on the ISI Web of Science

This article has been **cited by** 54 articles hosted by HighWire Press; see:

<http://www.sciencemag.org/content/325/5941/737.full.html#related-urls>

This article appears in the following **subject collections**:

Genetics

<http://www.sciencemag.org/cgi/collection/genetics>

17. J. W. Valentine, D. Jablonski, S. M. Kidwell, K. Roy, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 6599 (2006).
18. Materials and Methods are available as supporting material on Science Online.
19. D. Jablonski, K. Roy, J. W. Valentine, *Science* **314**, 102 (2006).
20. J. J. Sepkoski, *Philos. Trans. R. Soc. London Ser. B* **353**, 315 (1998).
21. D. Jablonski, J. A. Finarelli, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 8262 (2009).
22. D. Jablonski, K. Roy, J. W. Valentine, R. M. Price, P. S. Anderson, *Science* **300**, 1133 (2003).
23. J. D. Taylor, S. T. Williams, E. A. Glover, P. Dyal, *Zool. Scr.* **36**, 587 (2007).
24. R. P. Freckleton, P. H. Harvey, M. Pagel, *Am. Nat.* **160**, 712 (2002).
25. S. E. Peters, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 12326 (2005).
26. A. B. Smith, A. J. McGowan, *Palaeontology* **50**, 765 (2007).
27. R. D. Pancost *et al.*, *J. Geol. Soc. London* **161**, 353 (2004).
28. M. L. McKinney, *Paleobiology* **21**, 300 (1995).
29. D. Jablonski, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 5393 (2001).
30. J. B. C. Jackson, in *Extinction Rates*, J. H. Lawton, R. M. May, Eds. (Oxford Univ. Press, Oxford, 1995), pp. 45–54.
31. S. M. Stanley, *Paleobiology* **16**, 401 (1990).
32. J. Zachos, M. Pagani, L. Sloan, E. Thomas, K. Billups, *Science* **292**, 686 (2001).
33. J. A. Crame, *Paleobiology* **26**, 188 (2000).
34. S. M. Stanley, *Paleobiology* **33** (suppl.), 1 (2007).
35. S. Nee, *Annu. Rev. Ecol. Syst.* **37**, 1 (2006).
36. S. Nee, E. C. Holmes, R. M. May, P. H. Harvey, *Philos. Trans. R. Soc. London Ser. B* **344**, 77 (1994).
37. D. L. Rabosky, I. J. Lovette, *Proc. R. Soc. London Ser. B. Biol. Sci.* **275**, 2363 (2008).
38. D. L. Rabosky, I. J. Lovette, *Evolution* **62**, 1866 (2008).
39. J. J. Sepkoski Jr., in *Global Events and Event Stratigraphy*, O. H. Walliser, Ed. (Springer, Berlin, 1996), pp. 35–51.
40. J. W. Valentine, in *Causes of Evolution: A Paleontological Perspective*, R. M. Ross, W. D. Allmon, Eds. (Univ. of Chicago Press, Chicago, 1990), pp. 128–150.
41. J. T. Smith, K. Roy, *Paleobiology* **32**, 408 (2006).
42. We thank S. M. Kidwell and J. W. Valentine for discussions, I. Tëmkin and T. R. Waller for taxonomic advice, and two anonymous reviewers for insightful comments. This work was supported by a grant from NASA.

#### Supporting Online Material

www.sciencemag.org/cgi/content/full/325/5941/733/DC1

Materials and Methods

SOM Text

Figs. S1 to S5

Tables S1 and S2

References

4 March 2009; accepted 19 June 2009

10.1126/science.1173073

# Genetic Properties of the Maize Nested Association Mapping Population

Michael D. McMullen,<sup>1,2</sup> Stephen Kresovich,<sup>3</sup> Hector Sanchez Villeda,<sup>2\*</sup> Peter Bradbury,<sup>1,3</sup> Huihui Li,<sup>4,5,3</sup> Qi Sun,<sup>6</sup> Sherry Flint-Garcia,<sup>1,2</sup> Jeffery Thornsberry,<sup>7</sup> Charlotte Acharya,<sup>3</sup> Christopher Bottoms,<sup>2</sup> Patrick Brown,<sup>3</sup> Chris Browne,<sup>1</sup> Magen Eller,<sup>1</sup> Kate Guill,<sup>1</sup> Carlos Harjes,<sup>3†</sup> Dallas Kroon,<sup>3</sup> Nick Lepak,<sup>1</sup> Sharon E. Mitchell,<sup>3</sup> Brooke Peterson,<sup>1</sup> Gael Pressoir,<sup>3‡</sup> Susan Romero,<sup>1</sup> Marco Oropeza Rosas,<sup>8¶</sup> Stella Salvo,<sup>1</sup> Heather Yates,<sup>3</sup> Mark Hanson,<sup>9</sup> Elizabeth Jones,<sup>10</sup> Stephen Smith,<sup>10</sup> Jeffrey C. Glaubitz,<sup>11</sup> Major Goodman,<sup>8</sup> Doreen Ware,<sup>1,12</sup> James B. Holland,<sup>1,8</sup> Edward S. Buckler<sup>1,3,13||</sup>

Maize genetic diversity has been used to understand the molecular basis of phenotypic variation and to improve agricultural efficiency and sustainability. We crossed 25 diverse inbred maize lines to the B73 reference line, capturing a total of 136,000 recombination events. Variation for recombination frequencies was observed among families, influenced by local (cis) genetic variation. We identified evidence for numerous minor single-locus effects but little two-locus linkage disequilibrium or segregation distortion, which indicated a limited role for genes with large effects and epistatic interactions on fitness. We observed excess residual heterozygosity in pericentromeric regions, which suggested that selection in inbred lines has been less efficient in these regions because of reduced recombination frequency. This implies that pericentromeric regions may contribute disproportionately to heterosis.

The majority of phenotypic variation in natural populations and agricultural plants and animals is determined by quantitative genetic traits (1). Maize (*Zea mays* L.) exhibits extensive molecular and phenotypic variation (2–4). Understanding the genetic basis of quantitative traits in maize is essential to predictive crop improvement. However, only slow progress has been made in identifying the genes controlling quantitative agronomic traits because of limitations in the scope of allelic diversity and resolution in available genetic mapping resources. Linkage mapping generally focuses on the construction and analysis of large families from two inbred lines to detect quantitative trait loci (QTLs) (5). However, resolution of these QTLs can be poor because of the limited number of recombination events that occur during population development. Association analysis takes advantage of historic recombination from deep coalescent history as linkage disequilibrium (LD) generally decays with-

in 2 kb (1, 6). However, because of the number of single-nucleotide polymorphisms (SNPs) required and the confounding effects of population structure, whole-genome association analysis can be difficult in maize (4).

To provide a genetic resource for quantitative trait analysis in maize, we have created the nested association mapping (NAM) population. NAM was constructed to enable high power and high resolution through joint linkage-association analysis, by capturing the best features of previous approaches (7, 8). The genetic structure of the NAM population is a reference design of 25 families of 200 recombinant inbred lines (RILs) per family (fig. S1). The inbred B73 was chosen as the reference inbred line because of its use for the public physical map (9) and for the Maize Sequencing Project (www.maizesequence.org). The other 25 parents [named the 25 diverse lines (25DL)] maximize the genetic diversity of the RIL families (8, 10), independent of any spe-

cific phenotype. The lines were chosen to represent the diversity of maize—more than half are tropical in origin, nine are temperate lines, two are sweet corn lines (representing Northern Flint), and one is a popcorn inbred line (fig. S2).

The NAM genetic map is a composite map created with 4699 RILs combined across the 25 families, representing 1106 loci, with an average marker density of one marker every 1.3 centimorgans (cM) (fig. S3 and table S1). The proportion of SNP loci from the composite map polymorphic in an individual family ranged from 63 to 74%. Among RILs, 48.7% of all marker genotypes were inherited from B73, 47.6% were inherited from the 25DL parent, and 3.6% were heterozygous, which suggests that they were broadly representative of the parents and fall within the expected range for S5 generation RILs. The NAM population captured ~136,000 crossover events, corresponding, on average, to three crossover events per gene. This allows genetic factors to be mapped to very specific regions of the

<sup>1</sup>United States Department of Agriculture–Agriculture Research Service (USDA–ARS). <sup>2</sup>Division of Plant Sciences, University of Missouri, Columbia, MO 65211, USA. <sup>3</sup>Institute for Genomic Diversity, Cornell University, Ithaca, NY 14853, USA. <sup>4</sup>School of Mathematical Science, Beijing Normal University, Beijing 100875, China. <sup>5</sup>Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing 100081, China. <sup>6</sup>Computational Biology Service Unit, Cornell University, Ithaca, NY 14853, USA. <sup>7</sup>Northwest Missouri State University, Maryville, MO 64468, USA. <sup>8</sup>Department of Crop Science, North Carolina State University, Raleigh, NC 27695, USA. <sup>9</sup>Illumina Inc., San Diego, CA 92121, USA. <sup>10</sup>Pioneer Hi-Bred, Johnston, IA 50131, USA. <sup>11</sup>Laboratory of Genetics, University of Wisconsin, Madison, WI 53706, USA. <sup>12</sup>Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA. <sup>13</sup>Department of Plant Breeding and Genetics, Cornell University, Ithaca, NY 14853, USA.

\*Present address: International Maize and Wheat Improvement Center (CIMMYT), kilometer 45, Carretera Mex-Veracruz, El Batán, Texcoco, Mexico.

†Present address: Monsanto, Leesburg, GA 31763, USA.

‡Present address: Fondation CHIBAS, 30 Rue Pacot, Port-au-Prince, Haiti.

¶Present address: Delta Pine/Monsanto, Post Office Box 194, Scott, MS 38772, USA.

||To whom correspondence should be addressed. E-mail: mcmullenm@missouri.edu (M.D.M.); sk20@cornell.edu (S.K.); james\_holland@ncsu.edu (J.B.H.); esb33@cornell.edu (E.S.B.)

genome (11) and leads to a higher-resolution anchoring of the physical map of maize. Recombination varied substantially among the 25 families. The genetic length of the individual family maps relative to the composite map ranged from -104.3 cM (-7.4%) for B73 × Mo18W to +269.4 cM (+19.2%) for B73 × CML228. We attempted to map global recombination controllers by treating the number of crossovers in each individual line as a phenotypic trait (12). However, despite our high statistical power and diversity, we found no shared controllers of recombination; in contrast, shared QTLs for other traits were common (11). By examining individual families, we found evidence for loci controlling whole-genome or individual chromosome recombination at only a 50% or greater false-discovery rate

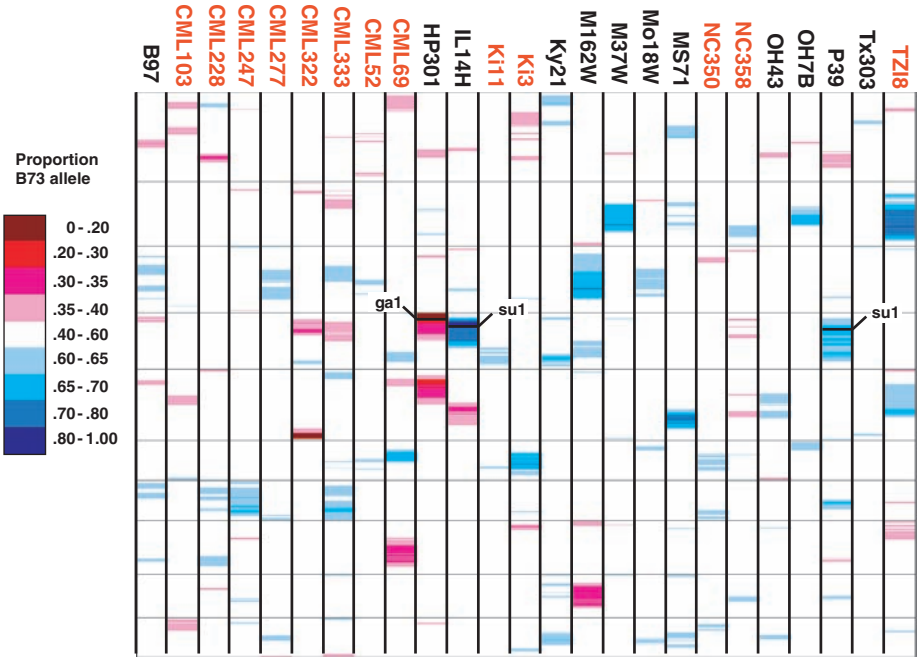
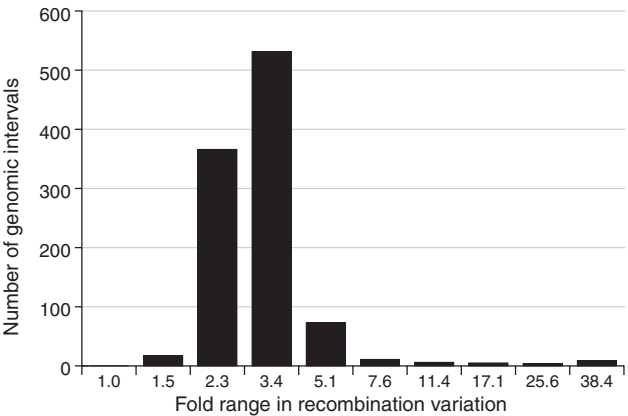
(FDR). Five of the eight loci showing the strongest effects control recombination in cis, which suggested that they are structural variants that modify recombination for specific chromosomes within specific families. The absence of loci with genome-wide effects on recombination suggests that the observed differences in recombination rates were because of numerous, but localized, regions of variation, as suggested by studies of individual families (13–15). We used a sliding window (4 to 6 cM) to examine differential recombination across the 25 families. The average interval had a 2.9-fold difference in recombination rate between the highest and lowest families, however, some intervals exhibited as much as 30-fold differences for recombination rate across families (Fig. 1). Overall, 41% of intervals showed significant dif-

ferences ( $P < 0.05$ ) in the number of recombination events across the 25 families (fig. S4). Across families, we did not observe a normal distribution in differential recombination frequency; the maximum increase relative to the consensus was twofold, whereas recombination was repressed by more than 20-fold in specific regions among specific families. Although recombination does not occur in retrotransposon clusters between genes, differences in the presence or absence of these clusters (which are ubiquitous among maize inbred lines), can result in at least threefold differences in recombination rates within flanking genes (16). Although these differences in non-genic content may explain many of the observed differences in recombination frequency, it seems less likely that they explain the virtual elimination of recombination within 6-cM+ intervals within specific families. Rather, we suspect that previously uncharacterized inversions may be responsible for some of the larger differences observed. For example, for one region on chromosome 5 that represents more than 12 cM of the composite map, we recovered no recombination events in either the B73 × CML322 or B73 × CML52 families, which suggested the presence of a large inversion of this region in CML322 and CML52 relative to B73.

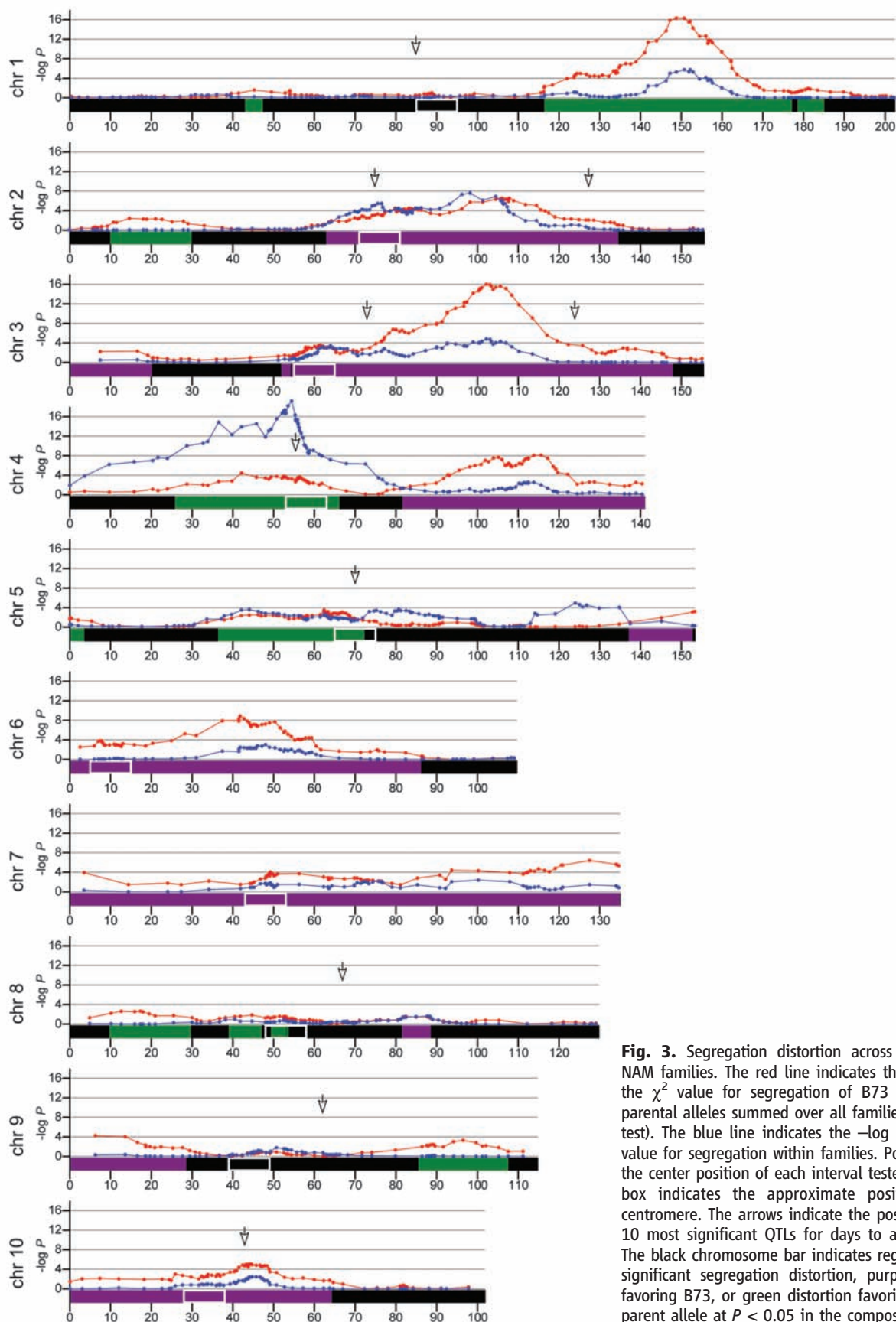
These differences in recombination among families hinder efforts to understand the genetic basis of quantitative traits in maize. All comparisons across mapping populations either by meta-analysis approaches (17) or joint-linkage mapping or joint association-linkage mapping, as with NAM, are confounded by this phenomenon, because these methods assume consistent recombination frequencies across families. The differences in recombination rates among families also indicate that map-based gene cloning projects need to be conducted in genetic backgrounds that demonstrate high recombination rates in the target region.

We developed one-third of the lines for each family under different environmental conditions to minimize inadvertent selection (18). However, because the final RILs resulted from 50,000 meioses and only 22% of the original  $F_2$  plants produced an S5 RIL, the surviving lines were unavoidably subjected to selection for multiple generations. Of all the chromosomal segments, 97% were represented by parental alleles at 45 to 55% frequency, close to the expectation of 50% (fig. S5). Within individual families, 17% of the markers exhibited segregation distortion at  $P < 0.05$ , 8.9% at  $P < 0.01$ , and 4.0% at  $P < 0.001$  (Fig. 2), less than has been reported for individual mapping populations (19–21). Within individual families, only 0.17% of all possible donor-marker combinations had less than 50 donor alleles, which demonstrated that diversity was effectively captured. We saw no bias of selection for temperate alleles, because the 13 families from tropical origin averaged less distorted markers than the total [7.8% for tropical families vs. 8.9% overall, ( $P < 0.01$ )].

**Fig. 1.** A sliding window analysis reveals differential recombination rates among the NAM families for synthetic intervals of ~4 cM (18).



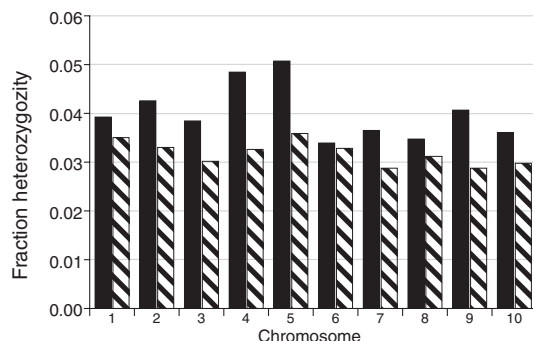
**Fig. 2.** Segregation distortion within individual NAM families. Each column represents one NAM family, the 25DL parent is indicated above the column. Red family names represent those with the 25DL parent from tropical origin. Each family consists of 1116 intervals, with the values of nonpolymorphic or missing marker data inferred on the basis of flanking marker values. Horizontal lines indicate the positions of chromosome boundaries for chromosomes 1 to 10, top to bottom. The proportion of B73 alleles for an interval is indicated by the color scale. The positions of the *ga1* and *su1* genes (22, 24) are indicated.



**Fig. 3.** Segregation distortion across and among NAM families. The red line indicates the  $-\log P$  for the  $\chi^2$  value for segregation of B73 versus 25DL parental alleles summed over all families (composite test). The blue line indicates the  $-\log P$  for the  $\chi^2$  value for segregation within families. Points indicate the center position of each interval tested. The white box indicates the approximate position of the centromere. The arrows indicate the positions of the 10 most significant QTLs for days to anthesis (11). The black chromosome bar indicates regions with no significant segregation distortion, purple distortion favoring B73, or green distortion favoring the 25DL parent allele at  $P < 0.05$  in the composite test.



**Fig. 4.** The proportion of marker genotypes that are heterozygous are shown as calculated for the area within 10 cM on each side of the centromere compared with the remaining chromosome arms. Black bars are within 10 cM on each side of the centromere position; hatched bars represent the rest of the chromosome.



We believe that four of the five most distorted regions within specific families are explained by known genetic factors. *Gametophyte factor 1-strong allele* (*Gal-S*) (22), which excludes *gal* pollen from fertilizing ovules of *Gal-S*, causes the distortion on the short arm of chromosome 4 for the Hp301 family ( $\chi^2$ ,  $P = 1.7 \times 10^{-19}$ ) (Fig. 2). The distortion on chromosome 5 of the Hp301 family probably results from a second gametophyte factor, *ga2* (23). Two sweet corn lines (I114H and P39) show distortion on chromosome 4 against the *sugary1* (*su1*) allele, which causes the sweet corn phenotype but can exhibit reduced germination vigor (24). We also observed substantial distortion in three families: M37W, Oh7B, and Tzi8 on a 65- to 110-cM region of chromosome 2 with a 2:1 bias favoring the B73 allele, perhaps because of QTLs for delayed flowering (*11*). No candidate for the major distortion on chromosome 5 in the CML322 family was identified. We attempted to map potential QTLs for trans-acting controllers of segregation distortion in individual populations but found none; it seems that selection operates directly on specific blocks of linked alleles.

We also tested whether specific regions of the genome show consistent segregation distortion favoring or disfavoring the B73 allele compared with other parental alleles (Fig. 3). This test includes all 4699 lines, and indeed, we found that 54% of the markers were distorted ( $P < 0.05$ ) by at least slight selection for or against B73 alleles, but few loci were under strong selection. Additionally, among-family segregation with a  $\chi^2$  test of B73 versus specific 25DL alleles showed that chromosome 4, containing both the *gal* and *su1* distortions, creates a large among-family distortion. However, we observed lower overall B73 versus 25DL distortion, most likely because of the different direction of bias of *gal* and *su1*, which cancel in the composite analysis. We saw little correspondence between flowering-time QTLs and regions of segregation distortion (Fig. 3).

Because these RIL families were derived from inbred lines, one might expect relatively few large single-locus effects for fitness and distortion. However, we expected epistasis to play a role in the creation of these new RILs, such that specific two-locus allelic combinations would be favored,

resulting in LD between certain unlinked loci. We surveyed two-locus LD for all pairs of markers on separate chromosomes, first, by comparing B73 alleles to all others from the complete set of 25 families. Marginally significant LD was observed between chromosome 6 and 7, but with a maximum  $r^2$  of only 0.005, a trivially small effect. Second, we tested LD within each population separately; among the 13.6 million tests, the highest level of LD was  $r^2 = 0.13$ , about what is expected by chance. Despite our tremendous diversity and statistical power, we saw virtually no evidence of epistatic effects on fitness in the NAM population.

The rate of evolution from natural selection or genetic gain from plant breeding for complex traits can be limited by repulsion phase linkages among favorable alleles [Hill-Robertson effect (25)]. Although maize has had a very large population size throughout most of its evolution, effective population sizes have been modest during the last century of inbred line development. One prediction of the Hill-Robertson effect is that favorable alleles have a higher chance of being in repulsion LD in regions with limited recombination. If such alleles exhibit dominance, then low-recombination pericentromeric regions should be under higher selective pressure to maintain heterozygosity. In NAM, residual heterozygosity averaged 4.1% for the 343 markers within 10 cM of the centromeres and averaged 3.2% for the 763 markers outside the centromeres, a 30% increase in pericentromeric regions. We found higher levels of heterozygosity near centromeres on all 10 chromosomes ( $P < 0.0004$ ) (Fig. 4), which suggested that selection favors heterozygosity in pericentromeric regions, perhaps because recombination in these regions has been insufficient to combine the optimal alleles. When B73 was mapped to identify heterotic regions (26), half of the largest QTL associated with centromeric regions, including a major QTL on chromosome 5 consisting of at least two distinct genes in repulsion phase linkage affecting yield (27). We observed high levels of residual heterozygosity in this region as well. We speculate that these data support the dominance theory of heterosis and that the increase in heterozygosity near centromeres is a consequence of heterosis. This heterosis is most likely the product of pseudo-overdominance that is most pronounced in peri-

centromeric regions, where the Hill-Robertson effect is strongest.

There were two striking and biologically significant discoveries in this project that impact maize breeding. The first was the extensive localized differences in recombination rates among families. We intend to genotype NAM at a much higher density to determine the structural bases of these differences. High-density genotyping in a large human pedigree has revealed extensive variation in use of recombination hotspots among individual families (28), similar to the variation we have revealed for maize. The second major finding was strong experimental support for the Hill-Robertson effect and the implications in understanding the basis of heterosis in maize. As similarly structured populations are developed for other model and crop species, it will be of interest to see if these findings are general or more specific to the demographic and breeding history of maize.

## References and Notes

1. T. F. Mackay, *Annu. Rev. Genet.* **35**, 303 (2001).
2. M. I. Tenaillon et al., *Proc. Natl. Acad. Sci. U.S.A.* **98**, 9161 (2001).
3. S. I. Wright et al., *Science* **308**, 1310 (2005).
4. S. A. Flint-Garcia et al., *Plant J.* **44**, 1054 (2005).
5. E. S. Lander, D. Botstein, *Genetics* **121**, 185 (1989).
6. D. L. Remington et al., *Proc. Natl. Acad. Sci. U.S.A.* **98**, 11479 (2001).
7. M. Zhang et al., *Genetics* **169**, 2305 (2005).
8. J. Yu, J. B. Holland, M. D. McMullen, E. S. Buckler, *Genetics* **178**, 539 (2008).
9. F. Wei et al., *PLoS Genet.* **3**, e123 (2007).
10. K. Liu et al., *Genetics* **165**, 2117 (2003).
11. E. S. Buckler et al., *Science* **325**, XXX (2009).
12. E. Esch, J. M. Szymanski, H. Yates, W. P. Pawlowski, E. S. Buckler, *Genetics* **177**, 1851 (2007).
13. W. D. Beavis, D. Grant, *Theor. Appl. Genet.* **82**, 636 (1991).
14. C. G. Williams, M. M. Goodman, C. W. Stuber, *Genetics* **141**, 1573 (1995).
15. H. Yao, P. S. Schnable, *Genetics* **170**, 1929 (2005).
16. H. K. Dooner, L. He, *Plant Cell* **20**, 249 (2008).
17. F. Chardon et al., *Genetics* **168**, 2169 (2004).
18. Materials and methods are available as supporting material on Science Online.
19. N. Sharopova et al., *Plant Mol. Biol.* **48**, 463 (2002).
20. A. M. Casa et al., *Proc. Natl. Acad. Sci. U.S.A.* **97**, 10083 (2000).
21. M. Falque et al., *Genetics* **170**, 1957 (2005).
22. R. A. Emerson, *Genetics* **19**, 137 (1934).
23. A. E. Longley, *Genetics* **46**, 641 (1961).
24. H. G. Nass, P. L. Crane, *Crop Sci.* **10**, 139 (1970).
25. W. G. Hill, A. Robertson, *Genet. Res.* **8**, 269 (1966).
26. C. W. Stuber, S. E. Lincoln, D. W. Wolff, T. Helentjaris, E. S. Lander, *Genetics* **132**, 823 (1992).
27. G. I. Graham, D. W. Wolff, C. W. Stuber, *Crop Sci.* **37**, 1601 (1997).
28. G. Coop, X. Wen, C. Ober, J. K. Pritchard, M. Przeworski, *Science* **319**, 1395 (2008).
29. Supported by National Science Foundation Award DBI0321467 and by research funds provided by USDA-ARS to M.D.M., E.S.B., and J.B.H.

## Supporting Online Material

www.sciencemag.org/cgi/content/full/325/5941/737/DC1  
Materials and Methods  
Figs. S1 to S5  
Tables S1 and S2  
References

31 March 2009; accepted 18 June 2009  
10.1126/science.1174320