

Supporting Online Material for

Genetic Properties of the Maize Nested Association Mapping Population

Michael D. McMullen,* Stephen Kresovich,* Hector Sanchez Villeda, Peter Bradbury,
Huihui Li, Qi Sun, Sherry Flint-Garcia, Jeffry Thornsberry, Charlotte Acharya,
Christopher Bottoms, Patrick Brown, Chris Browne, Magen Eller, Kate Guill, Carlos
Harjes,† Dallas Kroon, Nick Lepak, Sharon E. Mitchell, Brooke Peterson, Gael Pressoir,‡
Susan Romero, Marco Oropenza Rosas, Stella Salvo, Heather Yates, Mark Hanson,
Elizabeth Jones, Stephen Smith, Jeffrey C. Glaubitz, Major Goodman, Doreen Ware,
James B. Holland,* Edward S. Buckler*

*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.)

Published 7 August 2009, *Science* **325**, 737 (2009)
DOI: 10.1126/science.1174320

This PDF file includes

Materials and Methods
Figs. S1 to S5
References

Other Supporting Online Material for this manuscript includes the following:
(available at www.sciencemag.org/cgi/content/full/325/5941/737/DC1)

Two tables as .xls files.
Table S1. NAM genetic map
Table S2. Context sequences for SNPs for NAM genetic map

Materials and Methods

Development of RIL families.

The nested association mapping (NAM) genetic resource is comprised of 5000 RILs, composed of 25 families of 200 RILs per family. NAM was created on the basis of a reference design where one inbred line, B73, was crossed to 25 other diverse inbred lines and 200 RILs were derived from each cross. The other 25 parents; B97, CML52, CML69, CML103, CML228, CML247, CML277, CML322, CML333, Hp301, Il14H, Ki3, Ki11, Ky21, M37W, M162W, Mo18W, MS71, NC350, NC358, Oh43, Oh7B, P39, Tx303, and Tzi8 (Maize Molecular and Functional Diversity Project, <http://www.panzea.org>), were chosen from a set of 302 maize inbreds from around the world on the basis of simple sequence repeat allele data (*S1*, *S2*) to maximize the genetic diversity captured in the RIL families. Initial F₁ plants were self pollinated to generate S1 seed and lines were developed by single seed descent to the S5 stage. Although more than 200 S5 lines were produced for each family, 200 lines were sib-mated to increase seed. Each family was derived from three initial S1 ears, 1/3 of the lines for each family were developed with Columbia MO as the summer nursery and Puerto Rico as the winter nursery, 1/3 of the lines with Raleigh NC as the summer nursery and Homestead FL as the winter nursery, and 1/3 of the lines were made with Ithaca NY, Raleigh, NC and both winter nursery sites. It was the intent of this approach to develop lines of each family in differing environments in the hope of avoiding unintentional selection against specific genotypes that might occur in a single environment.

SNP panel for genotype analysis.

Tissue from up to four etiolated seedlings were harvested per line, lyophilized, ground in a Geno/Grinder 2000 (BT&C/OPS Diagnostics, Bridgewater, NJ) and then extracted with a standard (S3) cetyltrimethyl ammonium bromide extraction procedure. DNA was quantified using PicoGreen reagent Quant-iT™ PicoGreen ® dsDNA Reagent (Invitrogen, Carlsbad, CA). The genotypes of a total of 6825 RIL lines were determined using a panel of 1536 SNPs by the Illumina GoldenGate Assay system (S4) (Illumina, San Diego, CA). The majority of the SNP were chosen from alignments of sequences of the NAM parents from Maize Diversity Project database (www.panzea.org). To determine the demographic history of maize the Maize Diversity Project previously constructed sequence alignments of >3000 randomly chosen genes (S5, S6). In addition, our database includes sequence alignments of >1000 genes chosen as potential candidates to affect agronomic traits. For the origin of the SNPs, 974 were chosen from the random genes, 329 were chosen from candidate genes and 233 were chosen from alignments of sequence of inbred lines provided by Pioneer Hi-bred International. All sequence alignments are available at www.panzea.org, (GenBank accessions BV123534-BV144210, BV446558-BV447590, BV1063620-BV123527, BV725508-BV726468). The SNPs were chosen at sites where the B73 allele is rare and therefore polymorphic in a large number of families. The call areas for homozygous allele 1, heterozygous, and homozygous allele 2 were set on the basis of positions of the parental and F₁ control individuals in the SNP graph display of the Illumina BeadStudio software (Illumina, San Diego, CA). Each SNP was assigned a quality score from 1 to 4. Quality 4 SNPs were considered failed assays and were excluded from further analysis. These SNPs failed for a number of reasons including low overall signal intensity, too many failed individuals, no discrete clusters representing alleles, or no discrimination of alleles. There were 325 Quality 4 SNPs. There were 956 Quality 1 SNPs, representing high quality assays in which all parental and F₁ controls and the vast majority of the individuals fall unambiguously

into the three expected genotype classes and all individuals can be give a genotype call on the basis of a single cluster pattern. The remaining 255 SNP loci were of intermediate quality or more complex pattern and were assigned quality scores of 2 or 3. For some specific SNPs, some of the 25 DL parents and corresponding families segregated with the expected patterns, other parental individuals failed, or the 25 DL parents and corresponding F₁ individuals clustered in distinct positions on the SNP graph. These complex patterns are most likely due to secondary polymorphisms near the SNP being assayed. To correctly call the individuals for Quality 2 and 3 SNPs, 25 subprojects were defined, each corresponding to the parental controls, F₁ and RIL individuals corresponding to a specific family of lines. The allele cluster patterns for each of the 255 SNPs were set separately for each 25 RIL families on the basis of the position of B73, the specific 25 DL parent and corresponding F₁ individuals. This approach allows the best possible call for this SNP in a specific family and allows for removal of that SNP in specific RILs families where SNP calls would be problematic. The genotypes of a total of 1211 SNPs were accepted for use in constructing genetic map. Of these 26 markers were duplicated to allow SNP call error rates to be estimated. The error rate where a SNP was called a homozygous allele in one assay and a heterozygous allele in the other assay was on the order of 1%. However because the genetic mapping was done in an RIL self population structure, heterozygous call were converted to missing data, therefore miscalls of one homozygous class as a heterozygote will have no impact on map construction. The error rate in calling a SNP one homozygous allele in one assay and the other homozygous allele in the second assay was 1.7×10^{-5} .

The SNP allele calls were then used to ensure that the genotypes of the RIL lines were consistent with expectations on the basis of the parents of that family and five generations of self pollination. We examined in detail the 5000 RILs (200 lines/family) that had been grown in the large-scale field phenotype experiments in 2006 and 2007 (S7). Examination of genotypes revealed that one of the three original F₁ ears of the B73×Ki3 cross was not consistent with expected genotypes resulting in the elimination of 66 RILs for this family. We also identified 63 additional lines among the 25 RIL families with multiple alleles inconsistent with parental genotypes and these RILs were eliminated. In addition, there were 82 lines with the correct alleles but with greater than 8% heterozygous loci. These lines were also eliminated. There were 90 RILs where the individuals had low overall signal intensities in the SNP assays and SNP genotypes were not accepted for these lines. After exclusion of all the lines described above, 4699 RILs were accepted for construction of the NAM genetic map.

Construction of the NAM genetic map

The genetic map was constructed with the UNIX version of MAPMAKER 3.0 (S8). The code for this version has been locally modified to accept large data sets. All the map scores for the 4699 accepted RIL lines were concatenated into a single map score file. The B73 allele was designated as the “A” parent allele and the 25DL allele was designated as the “B” parent allele, heterozygous loci were converted to missing data “-”, as required by MAPMAKER when mapping in the RI self population structure. In addition, markers that were non-polymorphic in a particular family were also converted to missing data. Examination of resulting data file identified a few markers that within particular families were only polymorphic in the lines derived from a subset of the three initial S1 ears. Therefore each marker was checked for consistent segregation among subfamilies and non-segregating markers within particular

subfamilies were also converted to missing data. This is assumed to have been caused by a low level of residual heterozygous loci in the 25DL inbred parent plant used to make the initial F₁ crosses. In cases where two SNP were genotyped from the same initial amplicon we constructed a composite locus using as the base genotypes the SNP polymorphic in the larger number of families and then adding in the genotypes in families polymorphic at only the second SNP.

An initial framework map order was established for 115 loci. The order and chromosome position of these markers had been predetermined in a pilot experiment mapping on 94 individuals of the maize IBM population (S9). These markers were chosen to be evenly spaced across all chromosomes and to be polymorphic in at least 20 families. The order of these framework markers was defendable at LOD 3.0 by “ripple” analysis in MAPMAKER 3.0 in both the IBM population and in the initial NAM map. The remaining markers were placed to chromosome against this initial framework using the “assign” function in MAPMAKER 3.0 at LOD 10. Starting with the initial framework, markers were ordered within a chromosome with the “build” command, accepting placements at LOD 3, then LOD 2. Remaining markers were ordered into best position by the “try” command. The resulting chromosome maps were examined for markers with too many double crossovers or large numbers of crossovers in specific families indicating genotyping errors. These markers were eliminated from the data file and chromosomes were rebuilt from the initial framework until stable chromosomes were obtained. In addition to the composite maps, individual family genetic maps were constructed for each of the 25 families for the markers included in the final composite map. These maps were constructed with Mapmaker as above, except that no initial framework was used. Loci were assigned to chromosome by group and markers position within a chromosome obtained with the “order” function. Across all 25 populations there were only nine regions where a different order of closely placed markers was obtained for a specific family compared to the composite map, each with a unique order for only one family.

Analysis of segregation distortion

One difficulty in directly using the marker genotype to compare segregation distortion is the missing data caused by markers being monomorphic in specific populations. Therefore, to examine all markers of all families for evidence of segregation distortion, we used an imputed data file in which missing or monomorphic markers were extrapolated from flanking marker data (S7). Within each family and marker the proportion of B73 alleles was calculated and segregation distortion tested by χ^2 analysis against the 1:1 expectation of RIL population structure. For the composite analysis, the total proportion of B73 alleles across all families was tested against the expectation of 50% by χ^2 analysis (df = 1). For the across-family analysis, the number of B73 and 25DL alleles for each family are tested against a 1:1 expectation by χ^2 analysis (df = 49).

Analysis of family specific recombination rates

The problem of comparing recombination rates across families caused by monomorphic markers is similar to that for segregation distortion except that one cannot impute positions for recombination events. Therefore a sliding window analysis of intervals in which the ends of the intervals were not the same in each family was constructed as follows. The target interval length of 4 cM was assayed. Starting with one marker the marker closest to 4 cM away was selected and the number of recombination events in that interval was counted and compared to the expected number of recombination events for that interval from the composite NAM map. For families where both markers defining that interval were monomorphic, a

search was made to determine if a polymorphic marker pair could be obtained by replacing a monomorphic marker in that family with a polymorphic marker within 1 cM. If an interval with two polymorphic markers could be defined, then the number of recombination events in that interval were counted and compared to the expected number of recombination events for that interval from the composite NAM map. This procedure was done for all families, then the interval position was redefined by the average position of the left marker and average position of the right marker. Differences in recombination rates among the families for this new synthetic interval were tested by χ^2 analysis of observed total number of recombination events within a family versus that expected from the NAM composite map distance with df = (number of families for which intervals can be defined – 1).

Supplemental References

- S1. K. Liu *et al.*, *Genetics* **165**, 2117-2128 (2003).
- S2. J. Yu, J. B. Holland, M. D. McMullen, E. S. Buckler, *Genetics* **178**, 539 (2008).
- S3. J. J. Doyle, J. L. Doyle, *Phytochem. Bull.* **19**, 11-15 (1987).
- S4. J. B. Fan *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* **68**, 69-78 (2003).
- S5. S. I. Wright *et al.* *Science* **308**, 1310-1314 (2005).
- S6. M. Yamasaki *et al.*, *Plant Cell* **17**, 2859-2872 (2005).
- S7. E. S. Buckler *et al.*, *Science, accompanying paper*.
- S8. E. S. Lander *et al.*, *Genomics* **1**, 174-181 (1987).
- S9. M. Lee *et al.*, *Plant Mol. Biol.* **48**, 453-461 (2002).
- S10. S. A. Flint-Garcia *et al.*, *Plant J.* **44**, 1054-1064 (2005).
- S11. Names of products are necessary to report factually on available data: however, neither the USDA, nor any other participating institution guarantees or warrants the standard of the product and the use of the name does not imply approval of the product to the exclusion of others that may also be suitable.

Supplemental Figure Legends

Fig. S1. Development and genotyping of NAM population. (A) Strategy for development of the 5000 recombinant inbred lines for NAM. (B) Segregation of parental block among the NAM RIL is defined by 1106 SNP markers. (C) Polymorphisms within linkage blocks are obtained by projection of parental sequence onto the individual RILs allowing joint linkage-association to be performed with NAM.

Fig. S2. Genetic diversity of the NAM parental inbreds. The positions of the NAM inbred parental lines are indicated on a Fitch-Margoliash phylogenetic tree of 302 lines (S10). The figure also displays phenotypic variation in ear and plant morphology.

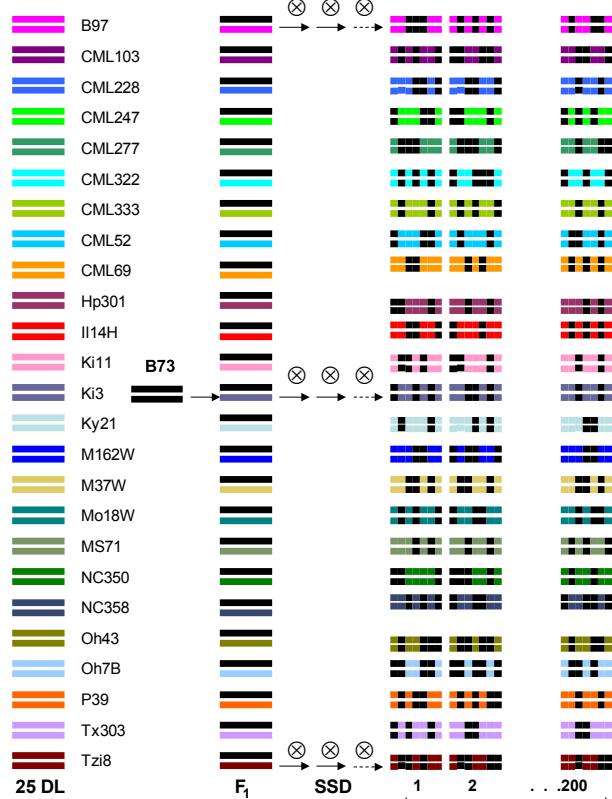
Fig. S3. Composite genetic linkage map for NAM. Genetic linkage map for the 10 chromosomes of maize for 1106 SNP loci for 4699 RIL for NAM. The genetic distance between markers in cM is indicated. The context sequences for all SNP assays along with the genotype file to reconstruct map are available from www.panzea.org.

Fig. S4. Differential recombination rates across NAM families. A sliding window analysis was performed to identify differential recombination rates among the NAM families. Synthetic intervals of ~ 4 cM were constructed as described in the above methods. The $-\log P$ of a χ^2 test comparing the actual number of crossover events versus expected on the basis of the genetic distance of the NAM composite map (Fig. S3) is indicated at the mid position of the interval tested. The bold dots indicate intervals significant at $P < 0.05$. The gold bar along the chromosome shows the approximate position of the centromere.

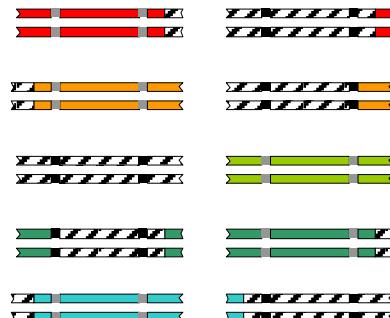
Fig. S5. Segregation distortion by line (A) and by marker (B) was overall very minor, although many of these minor differences do reflect significant departures from expectations. (A) Proportion of the markers that were contributed by each of the donor parents, coloring by type of donor. Sweet and popcorns, although temperate, were highlighted as they have known large distortion loci. (B) Distribution of donor contribution for the 1106 markers. Although the vast majority of markers are within 4% of expectation 50:50, more than half the markers show significant deviations ($P < 0.025$, two-tailed $P < 0.05$).

Figure S1.

A.



B.



C.

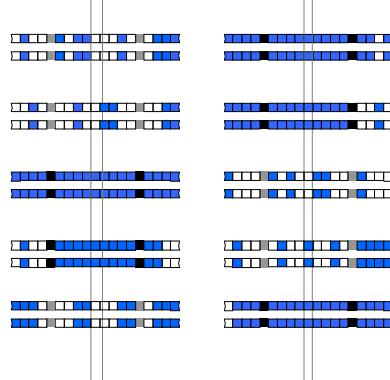
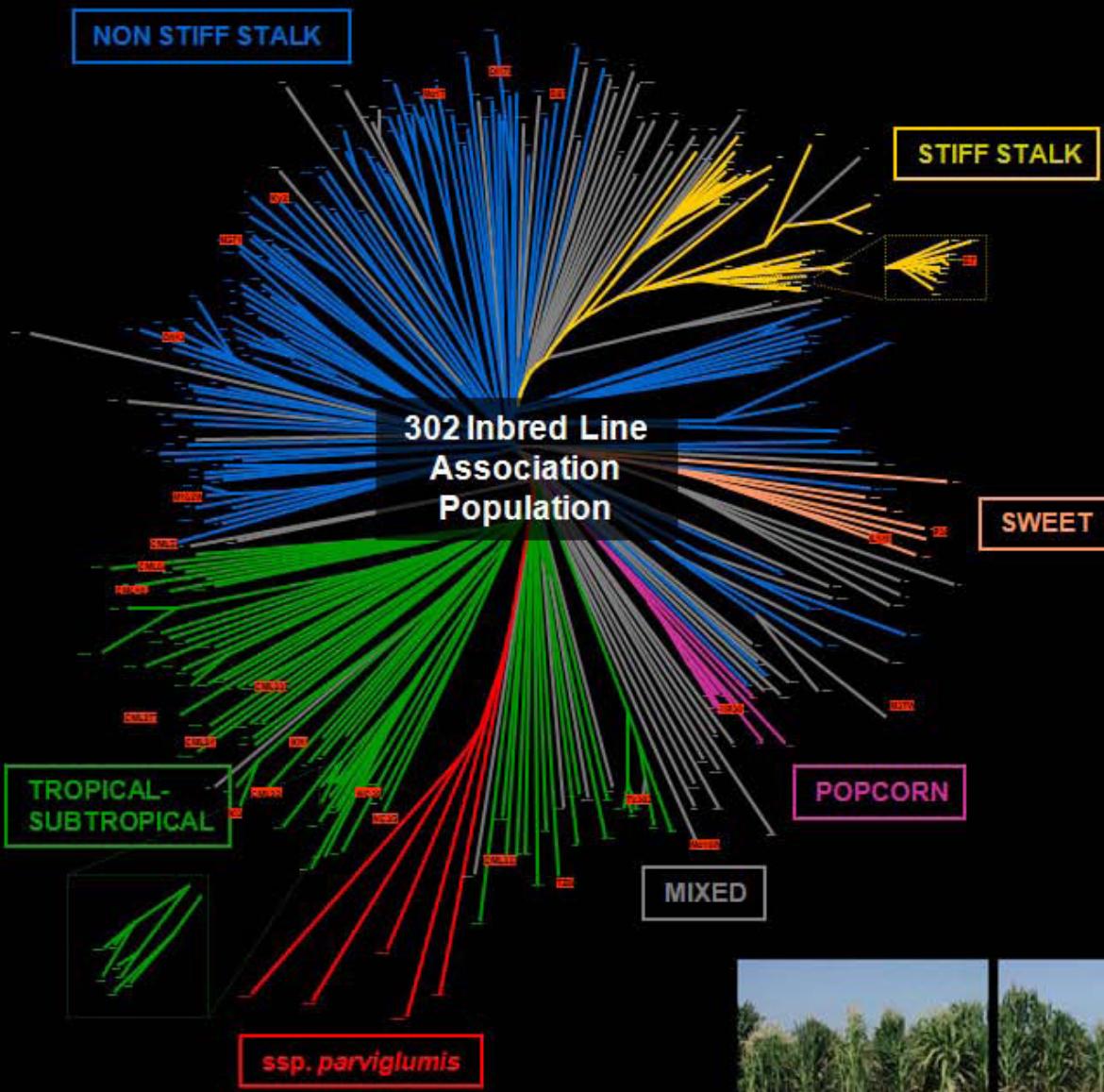


Figure S2.



Non-Stiff Stalk	Mixed	Tropical
B97	M37W	CML103 CML228
Ky21	Mo18W	CML247 CML277
M162W	Tx303	CML322 CML333
MS71		CML52 CML69
Oh43		Ki11 Ki3
Oh7B		NC350 NC358
		Tzi8
	Mo17 (IBM)	
	P39	

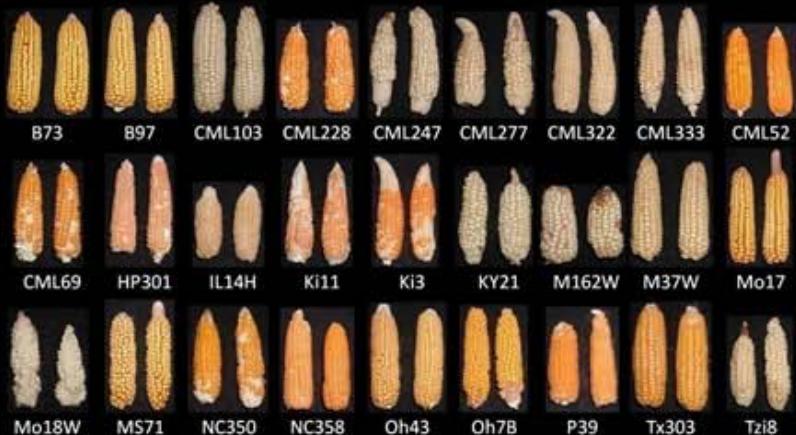


Figure S3.

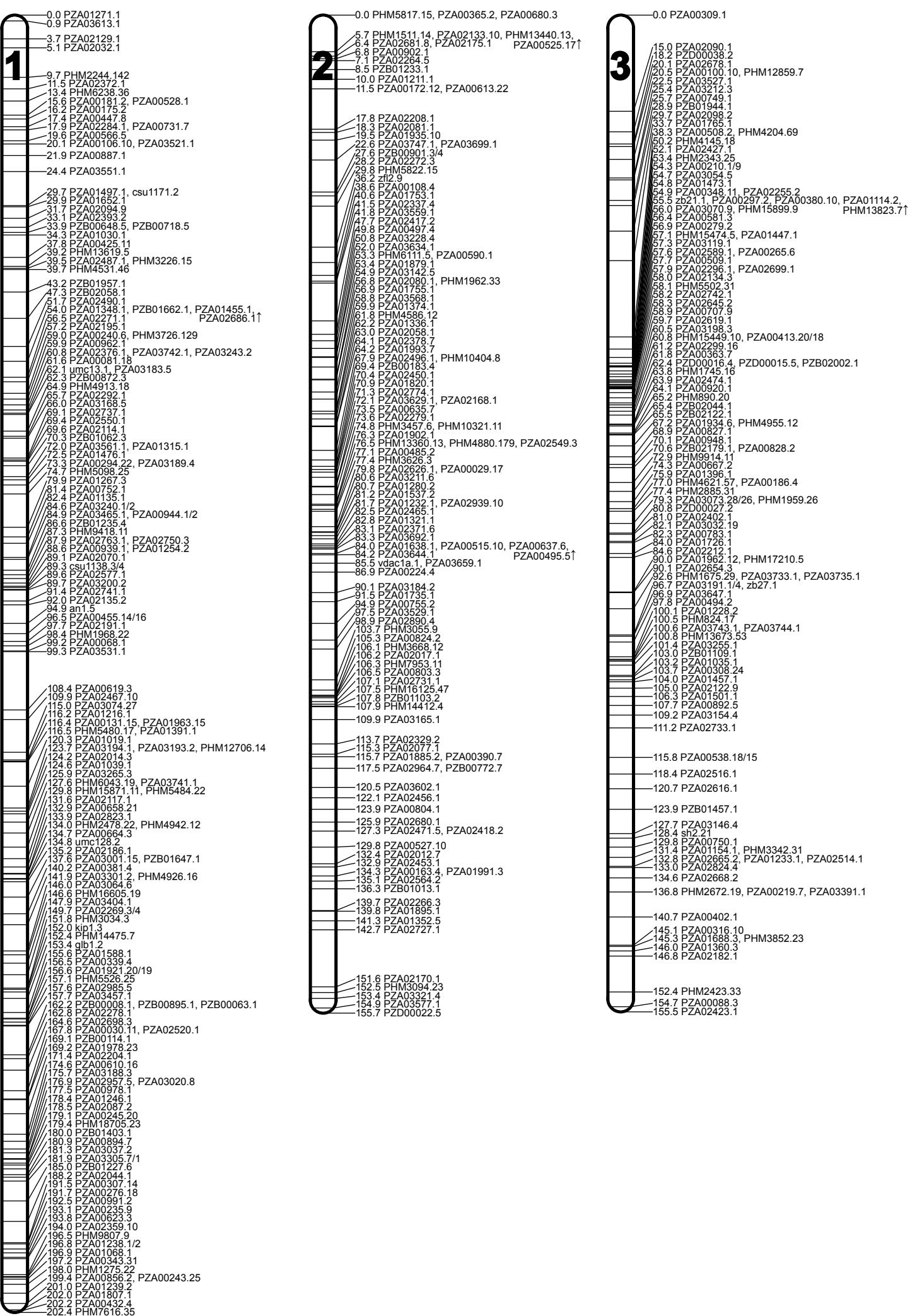


Figure S3 (continued).

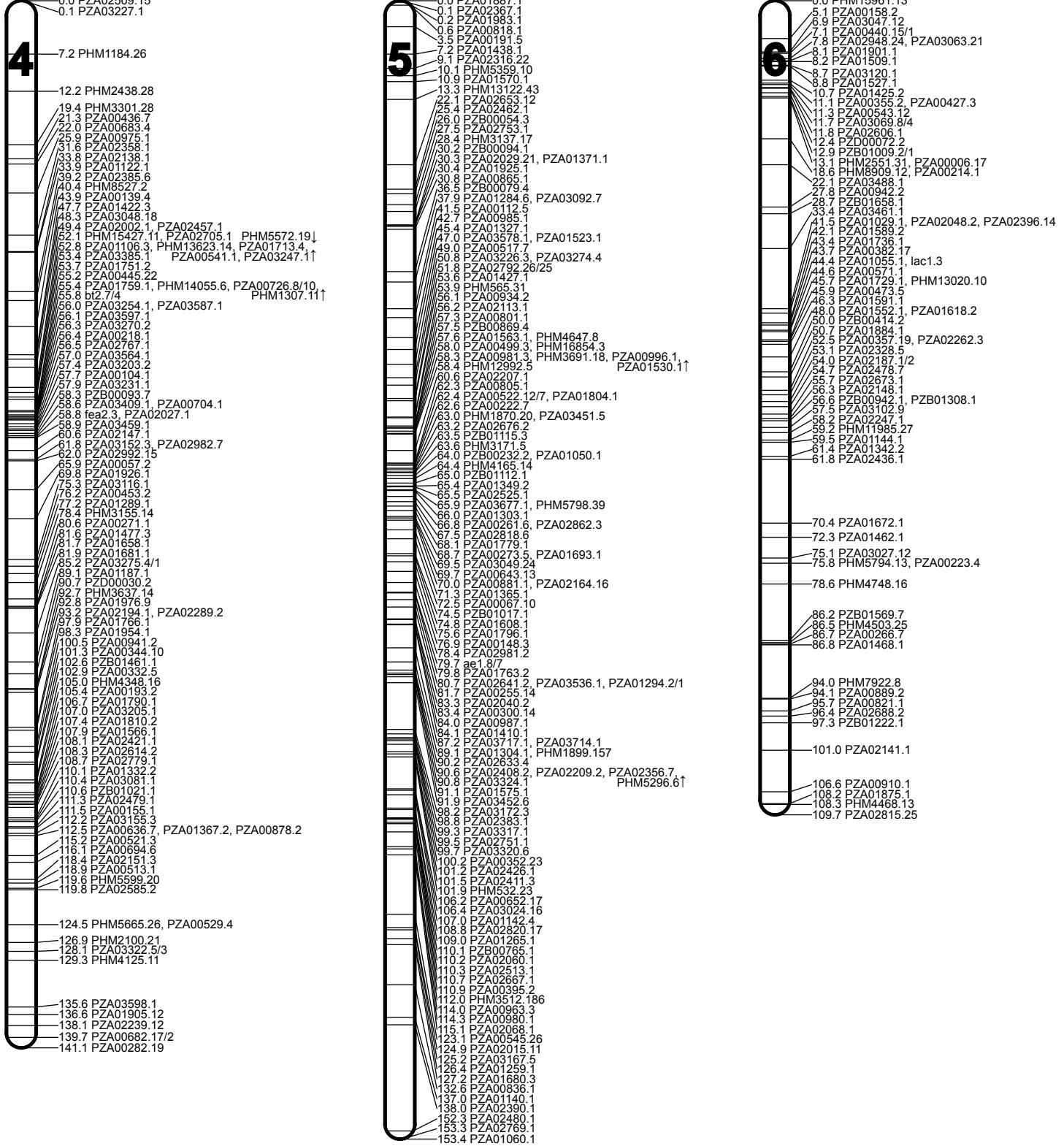


Figure S3 (continued).

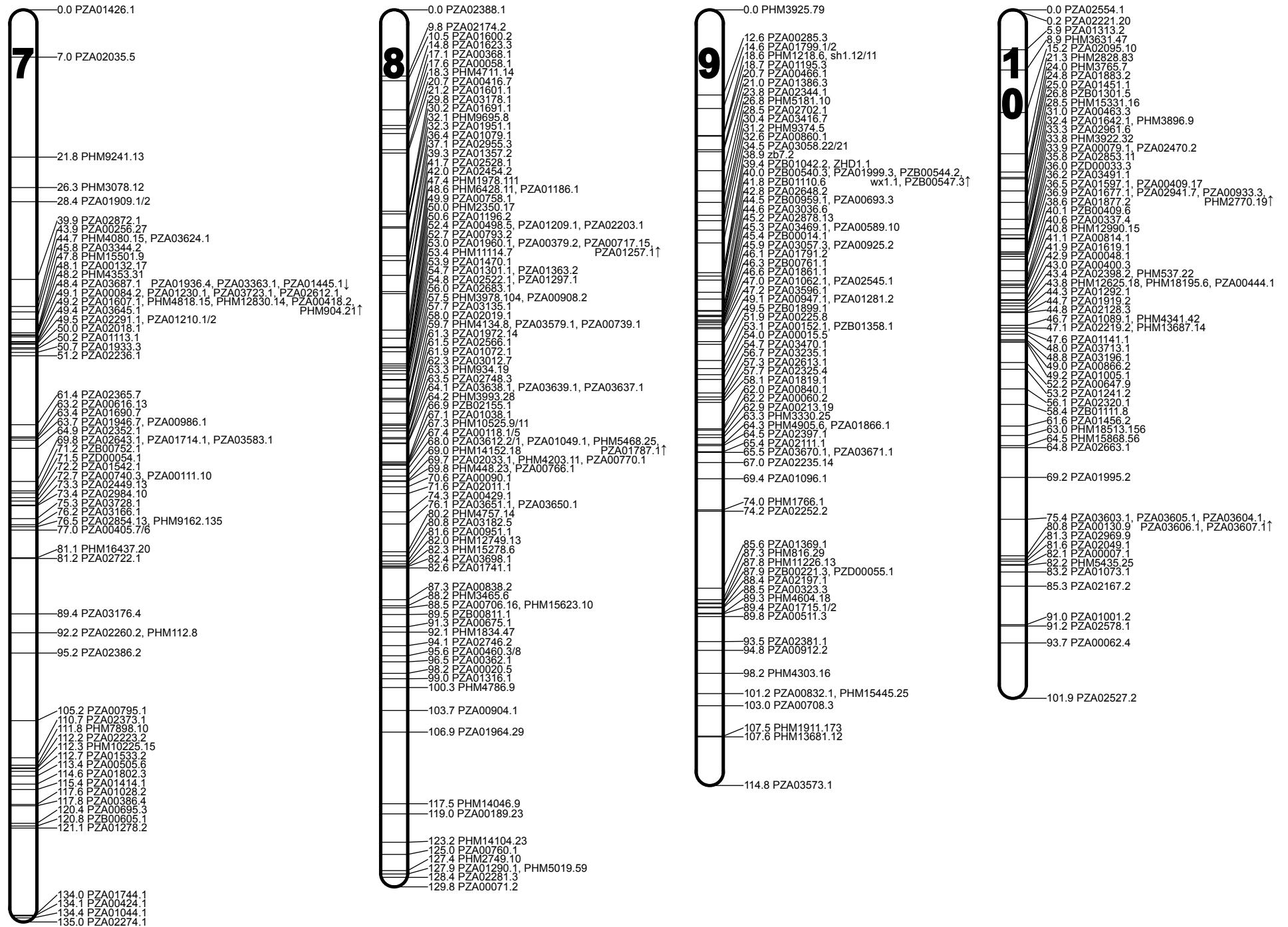


Figure S4.

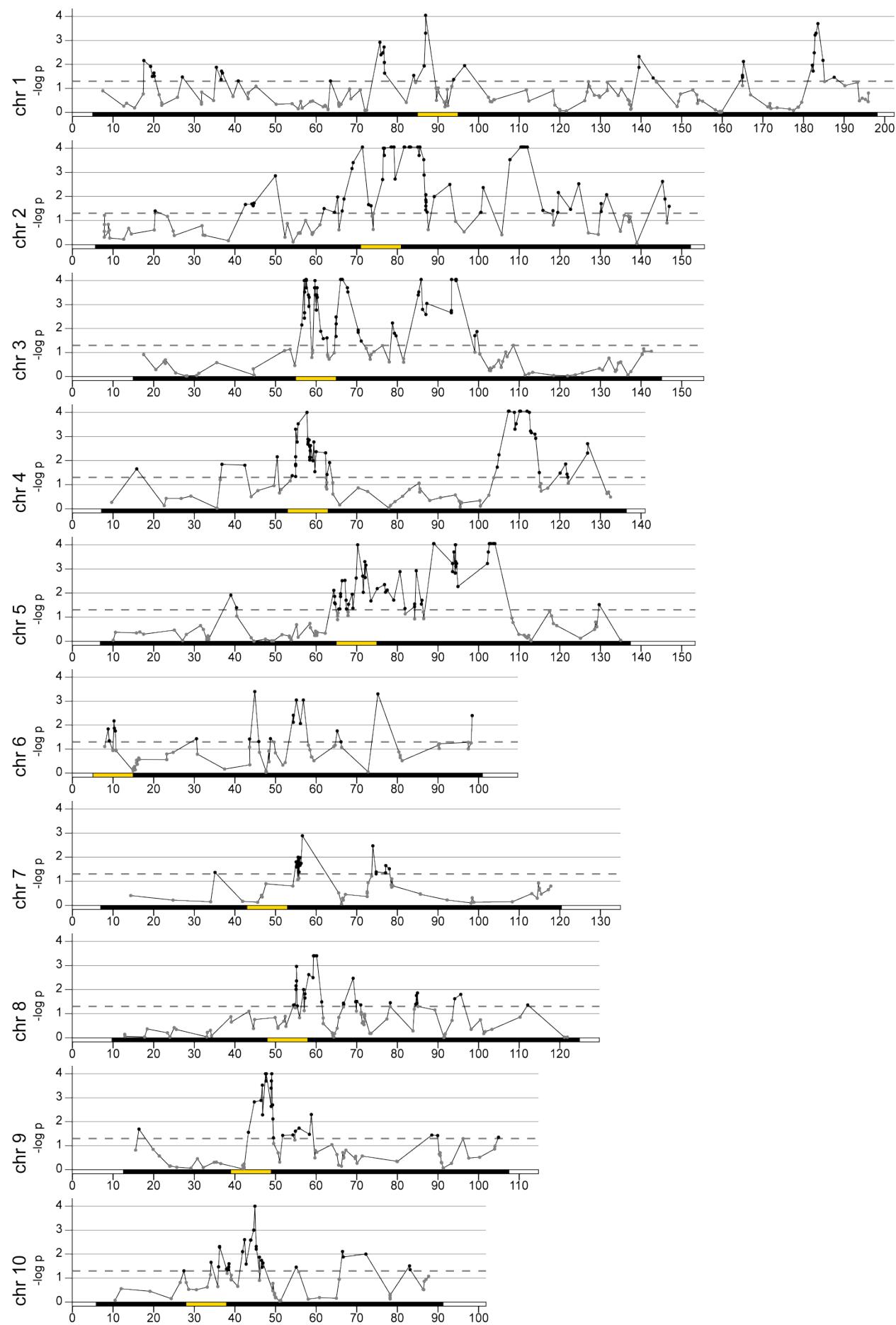
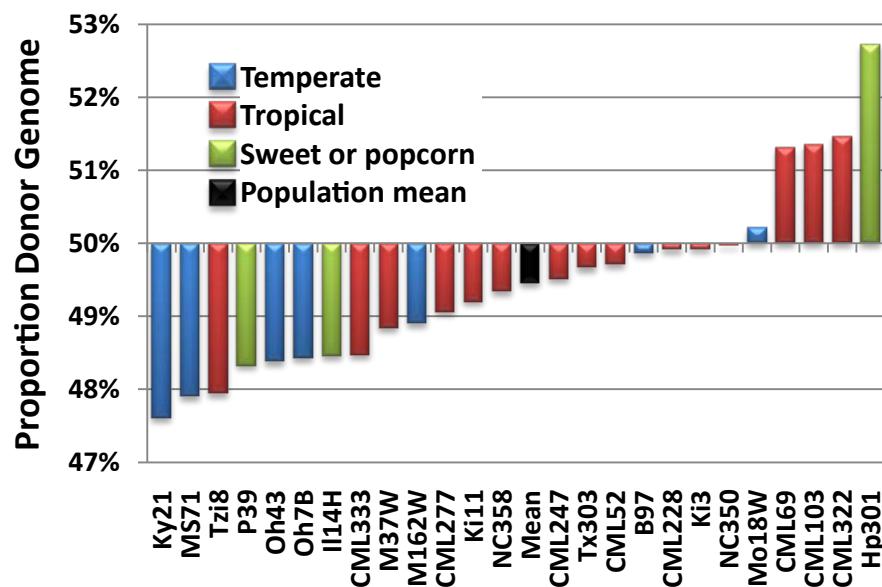


Figure S5.

A.



B.

