B73-Mo17 Near-Isogenic Lines Demonstrate Dispersed Structural Variation in Maize

Steven R. Eichten, Jillian M. Foerster, Natalia de Leon, Ying Kai, Cheng-Ting Yeh, Sanzhen Liu, Jeffrey A. Jeddeloh, Patrick S. Schnable, Shawn M. Kaeppler, and Nathan M. Springer*

Department of Plant Biology, University of Minnesota, Saint Paul, Minnesota 55108 (S.R.E., N.M.S.); Department of Agronomy, University of Wisconsin, Madison, Wisconsin 53706 (J.M.F., N.d.L., S.M.K.); Center for Plant Genomics, Iowa State University, Ames, Iowa 50011 (Y.K., C.-T.Y., S.L., P.S.S.); and Roche NimbleGen, Madison, Wisconsin 53719 (J.A.J.)

Recombinant inbred lines developed from the maize (Zea mays ssp. mays) inbreds B73 and Mo17 have been widely used to discover quantitative trait loci controlling a wide variety of phenotypic traits and as a resource to produce high-resolution genetic maps. These two parents were used to produce a set of near-isogenic lines (NILs) with small regions of introgression into both backgrounds. A novel array-based genotyping platform was used to score genotypes of over 7,000 loci in 100 NILs with B73 as the recurrent parent and 50 NILs with Mo17 as the recurrent parent. This population contains introgressions that cover the majority of the maize genome. The set of NILs displayed an excess of residual heterozygosity relative to the amount expected based on their pedigrees, and this excess residual heterozygosity is enriched in the low-recombination regions near the centromeres. The genotyping platform provided the ability to survey copy number variants that exist in more copies in Mo17 than in B73. The majority of these Mo17-specific duplications are located in unlinked positions throughout the genome. The utility of this population for the discovery and validation of quantitative trait loci was assessed through analysis of plant height variation.

Structured populations have been widely used to discover the genetic architecture underlying heritable variation for phenotypic traits. Structured populations are produced from the cross of two inbred plants, and independent families are derived from individual F2 plants by repeated self-pollination or doubled haploid production. Populations selected to near homozygosity are called recombinant inbred line (RIL) populations and consist of approximately equal proportions of the two parental lines that have been shuffled through meiotic recombination events. While populations such as RILs are statistically powerful for discovering relevant genomic intervals, they have limitations for the isolation and detailed characterization of individual quantitative trait loci (QTLs) that affect a phenotype of interest. Near-isogenic lines (NILs) and introgression lines are useful for QTL discovery and validation and as a resource to initiate positional cloning projects. In addition, these types of populations can be used to address questions regarding genome organization and genetic linkage.

The term “isogenic line” conceptually refers to a pair of inbred lines that differ by a single gene. NILs are not that pure and contain some proportion of a donor parent genome in a recurrent parent background. Originally, the term arose to account for the fact that linkage drag carried additional genes into backcross-derived lines selected on the basis of a single gene introgression in addition to recognition that random unlinked genomic fragments would be present. In this study, we are using the term NILs to describe a population of lines that are predominantly of the recurrent parent genotype and have one to a few segments from the donor parent. This type of population has also been referred to as introgression lines, but this term originated to describe backcross-derived lines containing segments from wild species (Eshed and Zamir, 1995). NILs have been developed for QTL analyses in maize (Zea mays ssp. mays) between B73 and Tx303 (Szalma et al., 2007) and B73 × Gaspe Flint (Salvi et al., 2011). In Arabidopsis (Arabidopsis thaliana), NILs were able to detect smaller effect QTLs than was possible using RILs, but they required more replications of phenotypic measurements and had limited genetic resolution (Keurentjes et al., 2007).

The two most widely used populations for QTL discovery in maize are the Intermated B73 × Mo17 (IBM) population (Lee et al., 2002) and Nested Association Mapping (NAM) populations (Yu et al., 2008; McMullen et al., 2009). B73 and Mo17 are the parents of...
for the IBM population. This population has been widely used for developing genetic maps (Sharopova et al., 2002; Yim et al., 2002; Fu et al., 2006; Liu et al., 2009) and for studying the genetic architecture of numerous traits (Mickelson et al., 2002; Hazen et al., 2003; Zhu et al., 2005, 2006; Balint-Kurti et al., 2007, 2008, 2010; Wassom et al., 2008; Ordas et al., 2009; Pressoir et al., 2009; Gustafson and de Leon 2010; Lorenz et al., 2010). The NAM population was developed by crossing 25 diverse inbreds to B73 and then developing a population of 200 RILs for each of these crosses, resulting in a total of 5,000 RILs that each contain approximately 50% B73. This population has been used to monitor the genetic architecture for flowering time (Buckler et al., 2009), leaf architecture (Tian et al., 2011), and resistance to southern (Kump et al., 2011) and northern (Poland et al., 2011) leaf blight. We developed NILs derived from B73 and Mo17 to provide a tool for validating QTLs as a resource for initiating fine-mapping experiments.

Obtaining distributed genotypic information on each line in a population(s) under study is essential for genetic research. There are numerous genotyping technologies available for maize, including RFLPs (Helenjärvi et al., 1988; Gardiner et al., 1993), simple sequence repeats (SSRs; Taramino and Tingey, 1996; Sharopova et al., 2002), insertion/deletion polymorphisms (Bhattamakki et al., 2002; Fu et al., 2006), and single-nucleotide polymorphisms (Ching et al., 2002; Barbazuk et al., 2007; Gore et al., 2009; Liu et al., 2009, 2010). Although these technologies have been successful, it is generally desirable to utilize a high-throughput technology that would provide data for a large number of markers in a single assay. There are a number of technologies for high-throughput genotyping (for review, see Gupta et al., 2008). These high-throughput approaches vary substantially in number of markers, amount of information required for development, accuracy, ease of application, and data analysis. Several groups have demonstrated the utility of microarray hybridizations of genomic DNA for high-throughput genotyping (Winzeler et al., 1998; Borevitz et al., 2003; Baird et al., 2008; Flibotte et al., 2009). Previously, we identified a number of long-oligonucleotide probes that exhibit variable hybridization to B73 and Mo17 (Springer et al., 2009). The recent completion of the maize genome reference sequence (Schnable et al., 2009) allowed for the physical mapping of each of these oligonucleotide probe sequences to the maize chromosomes. Genotyping of two intermated RILs through comparative genomic hybridization has been performed as a proof of concept in maize utilizing these probes (Fu et al., 2010). A subset of these probes was used to develop a custom platform for genotyping approximately 67,000 markers in our current NIL population.

The utility of this genotyping platform was demonstrated by determining high-resolution genotypes for 150 NILs developed from the inbred lines B73 and Mo17. This population provides an opportunity to test for elevated residual heterozygosity as noted by McMullen et al. (2009). In addition, our mapping platform also provides information on copy number variations (CNVs) that show increased copy number in Mo17 relative to the reference genome (B73). As a consequence of the technology and the genetic architecture of the NIL, the novel CNVs were able to be mapped. Surprisingly, many inbred-specific copy number gains do not represent linked duplications but are instead dispersed as unlinked CNVs throughout the genome.

RESULTS

Mapping and Analysis

NILs were developed from maize inbreds B73 and Mo17. NILs were developed via repeated backcrosses to the recurrent parental line for three generations and subsequent self-pollinations for four to six generations (Supplemental Fig. S1; Supplemental Table S1). Based on the number of backcrosses and self-pollinations, the expected proportion of the recurrent and donor parents within the genome can be calculated (Supplemental Fig. S1). This crossing scheme was expected to homozygose the majority (approximately 93%) of the genome in each line to the sequence of the recurrent parent. The remaining small (approximately 6%) proportion of the genome includes homozygous regions contributed by the donor parent and heterozygous regions. The 150 NILs within the population can be separated into two categories: 100 lines with B73 as the recurrent parent and Mo17 as the donor parent (B73-like NILs), and 50 lines with Mo17 as the recurrent parent and B73 as the donor parent (Mo17-like NILs).

NILs were genotyped using a custom long-oligonucleotide microarray developed by Roche NimbleGen. The array consisted of 12 identical subarrays each containing 135,000 probes, allowing 12 parallel two-color comparisons per slide resulting in 24 single-color analyses per slide. These 135,000 probes were selected from a previously described 2.1M feature array (Springer et al., 2009; Fu et al., 2010) and were rearrayed into the 135K layout. The current array includes 34,821 probes that had equivalent signals in previous B73 and Mo17 comparative hybridizations. These control probes were used for normalizing the hybridization data obtained from different lines. The array also includes 61,865 probes that show significantly higher hybridization to B73 than to Mo17. These likely represent sequences that contain multiple sequence polymorphisms in Mo17 or are missing in Mo17 (Springer et al., 2009). The hybridization signals from these probes can be used for genotyping (Fu et al., 2010). The remaining probes included 4,944 probes that had higher signal in Mo17 than in B73 and approximately 33,000 probes that exhibited polymorphic behavior in other genotypes (not analyzed for this study).
Each NIL was hybridized to this array in a single replicate, and the signal was normalized using the 34,821 probes that provide equivalent signal in B73 and Mo17. The normalized signal for the sample was then contrasted to a normalized B73 signal (that was determined by hybridization to the same slide). Each slide provided data for 20 NILs and two replicates of B73 and Mo17. The log₂(NIL/B73) of the normalized hybridization signals from the hybridization of genomic DNA to this array was used to genotype each of the NILs (Fig. 1). Two different approaches were utilized for processing of the resulting data. The first approach utilized the DNAcopy algorithm to identify segments of similarly behaving probes that represent B73, Mo17, or heterozygous genotypes (Fig. 1, A and C). The second approach was based on an average log₂ (NIL/B73) value calculated across features physically located within the boundaries defined by a bacterial artificial chromosome (BAC) clone. There are 7,313 BACs (43% of all BACs in a minimum tiling path) in the B73 reference genome (B73 RefGen_v1) that are represented by at least three polymorphic probes on the array. The average log₂ ratio for each BAC was then used to perform DNAcopy analysis to identify segments of BACs that are B73, Mo17, or heterozygous genotypes (Fig. 1, B and D). The analysis of the BAC averages has the advantage of using multiple measurements for each genotyping classification (Fig. 1). In addition, it is known that there is a potential for local (within-BAC) ambiguities in the assembly of the reference genome (Schnable et al., 2009), and the use of per-BAC calls eliminates the potential for complications of local ambiguities. Therefore, the per-BAC genotyping calls (Supplemental Table S2) were used for all subsequent analyses. However, the per-probe genotyping data can provide increased resolution near recombination events (Fig. 1, E–H), and we also provide these data (Supplemental Table S3), as they may be useful to individuals interested in specific genomic regions.

**Genetic Structure of NILs**

The genotype at each of the 7,313 BACs with at least three markers was determined in 150 NILs, including 100 B73-like and 50 Mo17-like NILs (Fig. 2). These NILs contain an average of 8.15 introgressed chromosomal segments (Table I). The average introgression is 20 centimorgan (cM) and 19.6 Mb in length. Based on the method by which these NILs were generated, each is expected to contain approximately 6% to 7% of the donor genome; the observed introgressed fraction is slightly higher (7.44%) than expected (Table I). Based on the population size of 100 B73-like NILs and

---

**Figure 1.** Hybridization log₂ ratio values provide lower overall signal for Mo17-like sequence compared with B73, allowing regions of introgression to be mapped. A to D, All 10 maize chromosomes. Blue and red markers indicate B73-like and Mo17-like calls, respectively. Black lines indicate segment lengths and average log₂ ratio values developed from DNAcopy. The resulting mapped segments are shown within (black boxes). A and C show all 62,995 array probes that develop a higher hybridization signal from B73 compared with Mo17. B and D contain probes averaged by their corresponding BAC and provide more robust log₂ ratio averages and clearer visualization. E to H, Detailed views of regions showing the resolution differences using a probe or an averaged BAC approach. Window sizes are indicated.

---

Each NIL was hybridized to this array in a single replicate, and the signal was normalized using the 34,821 probes that provide equivalent signal in B73 and Mo17. The normalized signal for the sample was then contrasted to a normalized B73 signal (that was determined by hybridization to the same slide). Each slide provided data for 20 NILs and two replicates of B73 and Mo17. The log₂(NIL/B73) of the normalized hybridization signals from the hybridization of genomic DNA to this array was used to genotype each of the NILs (Fig. 1). Two different approaches were utilized for processing of the resulting data. The first approach utilized the DNAcopy algorithm to identify segments of similarly behaving probes that represent B73, Mo17, or heterozygous genotypes (Fig. 1, A and C). The second approach was based on an average log₂ (NIL/B73) value calculated across features physically located within the boundaries defined by a bacterial artificial chromosome (BAC) clone. There are 7,313 BACs (43% of all BACs in a minimum tiling path) in the B73 reference genome (B73 RefGen_v1) that are represented by at least three polymorphic probes on the array. The average log₂ ratio for each BAC was then used to perform DNAcopy analysis to identify segments of BACs that are B73, Mo17, or heterozygous genotypes (Fig. 1, B and D). The analysis of the BAC averages has the advantage of using multiple measurements for each genotyping classification (Fig. 1). In addition, it is known that there is a potential for local (within-BAC) ambiguities in the assembly of the reference genome (Schnable et al., 2009), and the use of per-BAC calls eliminates the potential for complications of local ambiguities. Therefore, the per-BAC genotyping calls (Supplemental Table S2) were used for all subsequent analyses. However, the per-probe genotyping data can provide increased resolution near recombination events (Fig. 1, E–H), and we also provide these data (Supplemental Table S3), as they may be useful to individuals interested in specific genomic regions.

**Genetic Structure of NILs**

The genotype at each of the 7,313 BACs with at least three markers was determined in 150 NILs, including 100 B73-like and 50 Mo17-like NILs (Fig. 2). These NILs contain an average of 8.15 introgressed chromosomal segments (Table I). The average introgression is 20 centimorgan (cM) and 19.6 Mb in length. Based on the method by which these NILs were generated, each is expected to contain approximately 6% to 7% of the donor genome; the observed introgressed fraction is slightly higher (7.44%) than expected (Table I). Based on the population size of 100 B73-like NILs and
approximately 6% of the genome introgressed in each line, it is expected that each locus should be represented by an average of approximately six independently derived NILs. There are fewer Mo17-like NILs; therefore, it is expected that the average representation of B73 introgressions at each locus in this population will be approximately 3×. For each marker, the percentage of NILs that contain an introgression from the donor parent in each of the two NIL populations was plotted (Fig. 3). The average number of lines containing an introgression for any marker is 7.66× for the B73-like NILs and 3.79× for the Mo17-like NILs (Table I). All 7,313 markers are represented by at least one Mo17 introgression in the B73-like NILs, while over 96% of markers are represented by at least one B73 introgression in the Mo17-like NILs. Importantly, most genomic regions are represented in multiple lines. There are at least three different introgressions for 94.76% of the BACs in the B73-like NILs and for 62.85% of the BACs in the Mo17-like NILs. It was noted that many genomic regions that exhibited an elevated number of B73 introgressions had fewer than expected Mo17 introgressions, and vice versa (Fig. 3). For a control analysis, the B73-like NILs were randomly divided into two groups of 50 NILs to permit comparison with the 50 Mo17-like NILs. The correlation of the introgression number at each marker was determined. The subsamples of the B73-like NILs were not significantly different from each other but were significantly different ($\chi^2$, $P < 0.05$) from the introgression frequency in the Mo17-like population. The observation of biased representation in the NIL populations suggests that there are some loci with preferred inheritance of either the B73 or Mo17 allele.

Centromeric Regions Exhibit Elevated Levels of Residual Heterozygosity

A comparison of the physical and genetic coordinate-based visualizations of these lines reveals large physical regions of introgression (Supplemental Fig. S2). Regions spanning chromosome centromeres are a small proportion of the genetic map (Fig. 2). This result illustrates the lower frequency of recombination found within the centromeric region of chromosomes (Liu et al., 2009). It was observed that there were multiple examples of large heterozygous introgressions that encompassed the middle portions of chromosomes (Supplemental Fig. S2). This observation is reminiscent of the excess residual heterozygosity noted around centromeres in the NAM population (McMullen et al., 2009). Consequently, a more detailed analysis of residual heterozygosity within the NIL population was conducted.

Maize exhibits substantial levels of heterosis for a number of traits, including flowering time and yield (Zanoni and Dudley, 1989; Auger et al., 2005; Springer and Stupar, 2007; Flint-Garcia et al., 2009). Therefore, it is likely that any unintentional selection for healthy, early-flowering plants during the generation of the NILs will result in selection for higher than expected levels of residual heterozygosity. Levels of heterozygosity higher than expected were observed within both the B73-like and Mo17-like NILs ($\chi^2$, $P < 0.05$;
McMullen et al. (2009) noted an excess of residual heterozygosity in low-recombination peri-centromeric regions within the NAM population. A similar enrichment for residual heterozygosity in peri-centromeric regions was also noted within the NIL populations in this study (Fig. 4). Interestingly, while there is an elevated level of residual heterozygosity across the entire genome, the effect is most pronounced near the centromeres.

Mo17-Specific Amplifications Are Often Dispersed throughout the Genome

Studies of structural variation have revealed many examples of CNV and presence-absence variation between B73 and Mo17 (Springer et al., 2009; Belo´ et al., 2010). One class of CNV includes sequences that have more copies in Mo17 than in B73, hereafter termed Mo17-specific amplifications. Probes within these Mo17-specific amplification sequences exhibit higher levels of hybridization to Mo17 genomic DNA than to B73 genomic DNA. We were interested in determining the frequency of linked and unlinked amplifications in contributing to inbred-specific gain of a CNV. The microarray used for genotyping the NIL population included 4,944 (3.8%) probes that exhibit significantly higher signals for Mo17 than for B73 (Springer et al., 2009).

Several criteria were used to identify a subset of these probes that could be used for assessing whether Mo17-specific amplifications are genetically linked to the same location as the original B73 sequence or are located in unlinked genomic regions. We restricted the analysis to a series of 87 NILs for which we had the highest quality hybridizations. Furthermore, the analysis was restricted to the 505 probes that were covered by introgressions in both the B73-like and Mo17-like samples and those that had significant differences between the B73-like and Mo17-like samples. For each probe, four different classes of NILs were defined based on the NIL type (B73 or Mo17) and the genotype at the B73-based genomic position of the probe (B73 or Mo17). The B73-like NILs with the B73 genotype at the probe locus and Mo17-like NILs with the Mo17 genotype at the probe locus are the predominant classes and are used to define the parental signal distributions. The other classes (B73-like NILs with the Mo17 genotype at the probe locus and Mo17-like NILs with

Table 1. Genetic characterization of NILs

<table>
<thead>
<tr>
<th>Parameter</th>
<th>B73-Like NILs</th>
<th>Mo17-Like NILs</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of NILs</td>
<td>100</td>
<td>50</td>
<td>150</td>
</tr>
<tr>
<td>Average no. of introgressions per line</td>
<td>7.66</td>
<td>8.98</td>
<td>8.15</td>
</tr>
<tr>
<td>Average introgression size (kb)</td>
<td>19,597</td>
<td>19,868</td>
<td>19,698</td>
</tr>
<tr>
<td>Average introgression size (cm)</td>
<td>21.79</td>
<td>18.21</td>
<td>20.48</td>
</tr>
<tr>
<td>Average no. of selfs</td>
<td>4.26</td>
<td>5.21</td>
<td>4.57</td>
</tr>
<tr>
<td>Expected percentage of genome introgressed (% bp)</td>
<td>6.58</td>
<td>6.42</td>
<td>6.53</td>
</tr>
<tr>
<td>Average percentage of genome introgressed (% bp)</td>
<td>7.44</td>
<td>7.43</td>
<td>7.44</td>
</tr>
<tr>
<td>Average percentage of genome introgressed (% cM)</td>
<td>7.06</td>
<td>6.92</td>
<td>7.01</td>
</tr>
<tr>
<td>Expected heterozygosity (% cM)</td>
<td>0.70</td>
<td>0.35</td>
<td>0.55</td>
</tr>
<tr>
<td>Average heterozygosity (% bp)</td>
<td>1.56</td>
<td>1.28</td>
<td>1.42</td>
</tr>
<tr>
<td>Average heterozygosity (% cM)</td>
<td>1.34</td>
<td>1.13</td>
<td>1.27</td>
</tr>
</tbody>
</table>

Figure 3. Percentage of introgression per BAC within the B73-like (blue) or Mo17-like (red) NIL population by physical (A) or genetic (B) position. For every BAC, a count of how many lines contain an introgression at that location was developed. This count was divided by the total number of NIL lines within the class (B73 like or Mo17 like).
the B73 genotype at the probe locus) are relatively rare and provide the opportunity to test for linked versus unlinked amplification. Linked amplifications are expected to have a signal reflective of the genotype at the probe locus. Therefore, the observed signal of the test classes will depend upon the genotype at the probe locus, not the NIL type. Conversely, unlinked amplifications will exhibit signal based on the NIL type, not the locus genotype. Figure 5 depicts the signal distributions in each of the four classes for several examples of linked and unlinked amplification. Each probe was classified as a linked amplification, unlinked amplification, or unassigned based on the relative hybridization signal in the test classes relative to the parental classes (Table II). The consensus of these classes suggests a dominating prevalence of unlinked amplifications and relatively few linked amplifications. In several cases, there were multiple probes that were located within a larger Mo17-amplified CNV segment detected by Springer et al. (2009). The 505 probes represent a total of 390 segments, and 25 of these segments possessed at least three probes. In most of these examples, each of the probes within the segment exhibited consistent assignment as either linked or unlinked.

Utilization of NILs to Detect or Corroborate QTLs: Plant Height as an Example

NIL populations can be used for the discovery of QTLs or for the validation and dissection of QTLs discovered in a separate population (such as RILs or NAM populations). Here, we use plant height as a trial case to demonstrate both approaches. Significant phenotypic variation ($P < 0.05$) was observed for plant height among NIL lines for each population (B73-like and Mo17-like) when evaluated across the three environments (Supplemental Table S4). Plant height QTLs were identified in the NIL population and in two different RIL populations (for full description, see Supplemental Fig. S5).

It might often be the case that a researcher would seek to confirm previously detected QTLs using this NIL resource. We demonstrate the relative utility of this approach by testing for corroboration of QTLs previously detected in the IBM population that were evaluated in different environments than the NIL populations. We chose QTL intervals based on $1 - \log$ of the odds score QTL confidence intervals based on our analysis of the IBM population to test for the phenotypic differentiation of NILs that have a specific introgressed nonrecombinant segment compared with those that do not contain the introgressed segment. Based on such a comparison, a single QTL from the IBM population was confirmed on chromosome 1 (80,047,351–82,550,551 bp from AGP version 2) of the NIL population. Due to the lack of replication of specific nonrecombinant segments in the NIL population, confirmation of QTLs detected in the IBM population could not be conducted in some cases. At least two lines needed to contain an introgression to conduct this analysis. In this case, additional purification of these lines to include single introgressed segments would improve this type of analysis.

Confirmation of QTL models developed from RIL populations can be conducted by predicting the phenotypic performance of specific NILs using QTL models identified in the IBM population and comparing those with observed phenotypes. We predicted plant height deviation from the recurrent parent for all 150 NIL lines using allelic effects from the QTL model developed for the IBM population ($r^2 = 0.071$, $P < 0.05$; Supplemental Fig. S4). Because the heritability of plant height in the IBM population is 0.89, 89% is the maximum amount of variation that can be explained by this model. The predicted values are on average more extreme than the observed values, which is likely due to overestimation of allelic effects in the original analysis. The IBM model was formulated from phenotypic data taken in different years and environments than the NIL evaluation, which might often be the case with validation studies. Also, the 2009 and 2010 IBM data showed a significant year-by-genotype interaction ($P < 0.001$), which could contribute to the low predictability by the NIL populations (Supplemental Table S5). Validation of QTLs provides additional confidence in the repeatability of detection across environments and samples of a population. However, the lack of validation does not indicate that the original assessment was incorrect but rather that gamete, recombination, and environment sampling will have a substantial effect on the ability to consistently identify QTLs across populations derived from the same parent.

**DISCUSSION**

Numerous groups have used the IBM recombinant inbred population to discover QTLs in maize. In
addition, the recurrent use of B73 in the NAM population (McMullen et al., 2009) has led to the discovery of numerous QTLs in B73. We have developed a near-isogenic population that can be used for further validation and characterization of these QTLs. This genetic resource is expected to facilitate map-based cloning efforts to characterize the underlying basis for many QTLs in maize. The initial characterization of these NILs has led to intriguing observations about the genetic architecture of these lines, the genomic location of inbred-specific gain CNV, and the relative utility of NILs and RILs for QTL discovery.

Genetic Architecture of NILs

This population includes segmental introgressions of B73 and Mo17 into both backgrounds. When the frequency of introgression at each marker in both populations was compared, several regions with the opposing biases toward either the B73 or Mo17 allele were identified. These trends were confirmed by analyzing subsets of the B73-like NILs and by analysis of relative coverage for each of the bins. The most likely explanation is that there are several loci for which there is a preferred allele with the ability to confer increased fitness. A careful analysis of segregation distortion in the maize IBM RILs identified several such genomic regions that exhibit bias toward B73 or Mo17 haplotypes (Fu et al., 2006). Biased representation of some of these same regions that exhibited segregation distortion was observed, and Fu et al. (2006) suggested that some of this segregation distortion may be due to inadvertent selection for flowering time QTLs. The findings of their study suggest that the same mechanism may have contributed to the biased retention of these regions in our NIL population as well.

Maize exhibits substantial heterosis for a number of phenotypic characteristics in B73 × Mo17 hybrids

Table II. High frequency of unlinked CNVs

<table>
<thead>
<tr>
<th>Classification</th>
<th>B73-Like NIL with Mo17 at Probe(a)</th>
<th>Mo17-Like NIL with B73 at Probe(a)</th>
<th>Consensus(b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unlinked</td>
<td>430 (85%)</td>
<td>311 (62%)</td>
<td>279 (55%)</td>
</tr>
<tr>
<td>Linked</td>
<td>42 (8%)</td>
<td>120 (24%)</td>
<td>30 (6%)</td>
</tr>
<tr>
<td>Unassigned(b)</td>
<td>33 (7%)</td>
<td>74 (15%)</td>
<td>196 (39%)</td>
</tr>
</tbody>
</table>

\(a\)The signals for each probe were used to assign linked or unlinked based on the relationship to signals in the parental classes (for details, see “Materials and Methods”).

\(b\)The consensus calls from both B73-like and Mo17-like NILs are reported. If a call was only made in one population the consensus is unassigned.
(Zanoni and Dudley, 1989; Auger et al., 2005; Springer and Stupar, 2007). As NILs are produced, it is possible that individual plants within families that have higher levels of heterozygosity will exhibit higher levels of heterosis and therefore will be inadvertently selected for self-pollination because they are the healthiest and earliest flowering individuals. McMullen et al. (2009) noted an excess of residual heterozygosity in centromeric regions of the RILs produced for the NAM population. This was interpreted as support for the Hill-Robertson effects caused by repulsion-phase linkage of advantageous alleles (Hill and Robertson, 1966). Hill and Robertson (1966) discussed the implications of genetic linkage on the rate of genetic gain in breeding for complex traits, similar to predictions on the effect of recombination on evolutionary traits made previously (Fisher, 1930; Muller, 1932; Felsenstein, 1974). The so-called Hill-Robertson effect predicts repulsion-phase linkage of favorable alleles in regions with low recombination rates. Assuming that these favorable alleles exhibit dominant behavior, there would be selective pressure to maintain the heterozygosity of these low-recombination regions (McMullen et al., 2009). Many heterosis QTLs can be localized to low-recombination centromeric regions (Stuber, 1995), and in at least one case, fine-mapping revealed two favorable alleles in repulsion-phase linkage (Graham et al., 1997). McMullen et al. (2009) predicted that low-recombination pericentromeric regions may contribute disproportionately to heterosis through the Hill-Robertson effect of repulsion-phase linkages of favorable alleles.

A potential alternative explanation is that excess residual heterozygosity is observed near the centromeric regions of RIL or NIL populations simply due to the fact that relatively small genetic regions of heterozygosity include large physical regions with numerous genes. The regions surrounding maize centromeres exhibit very little recombination (Liu et al., 2009). Indeed, the middle third of the physical chromosome often contains less than 5% of the genetic distance of the chromosome. The gene number per Mb is substantially reduced in the region within 10 cM of the centromeres (11.63 genes per Mb within 10 cM of centromere compared with 25.8 genes per Mb on the remaining portions of chromosomes). However, the centromeres actually have a 4-fold higher density of genes per cM (56.1 genes per cM within 10 cM of centromere as compared with 12.98 genes per cM on the remainder of chromosomes). This means that small cM blocks of residual heterozygosity near centromeres will result in higher levels of per-gene heterozygosity. In our population, we noted quite strong enrichment for residual heterozygosity near centromeres. However, there was also an excess of residual heterozygosity on the chromosomal arms as well. It is possible that the inadvertent selection arising from crossing the earliest flowering, healthiest plants results in enrichment for residual heterozygosity genome wide but that this effect is strongest at the centromeres.

Genomic Location of CNVs

The array platform used for genotyping contained a series of probes that detect sequences with Mo17-specific amplifications. Analysis of the signals from these probe sequences in the NIL population allowed us to determine whether the Mo17-specific amplifications mapped to the same location as the original B73 sequence or whether they were located in unlinked genomic locations. We found that the majority of amplifications mapped to unlinked genomic locations. The fact that many of these Mo17-specific gain CNVs are located in unlinked genomic positions suggests that transposition or fractionation of amplified regions may be a major contributor to inbred-specific duplications in maize.

There is evidence for recombination- and replication-based mechanisms for CNV emergence (Innan and Kondrashov, 2010). These mechanisms are generally expected to result in linked (cis-localized) gain CNVs. Unlinked (trans-localized) amplifications can arise from mobilization via transposable elements (Bennetzen, 2005; Kim et al., 2008). Indeed, in one well-characterized example, there were several gene-like fragments that appear in some maize haplotypes but not others (Fu and Dooner, 2002), which were later characterized as Helitron-mediated duplications of sequences present on another chromosome (Lai et al., 2005). There are many examples of the movement of gene fragments by Mu elements (Jiang et al., 2004; Bennetzen, 2005) or by Helitron elements (Lai et al., 2005; Morgante et al., 2005; Du et al., 2009; Yang and Bennetzen, 2009). There is also evidence that these Helitron insertions are often polymorphic among different haplotypes (Lai et al., 2005; Du et al., 2009), which could lead to inbred-specific CNVs for genic sequences. Fractionation of amplified genomic regions could also lead to apparent inbred-specific gain CNVs. Maize is an ancient tetraploid with a number of syntenic regions that represent the ancestral chromosomes. However, in many cases, one member of the amplified gene pair has been lost such that the two syntenic regions have fractionated the original function (Langham et al., 2004; Freeling, 2009). In many cases, this fractionation process is biased such that one duplicate region experiences a higher rate of loss than the other region (Woodhouse et al., 2010). Ongoing fractionation of amplified genomic regions could lead to our observation of unlinked Mo17-specific duplications. If one member of the gene pair was lost in B73 but not in Mo17, it would result in an apparent unlinked Mo17 gain CNV. There would be a single copy of the gene in the B73 genome but Mo17 would contain both members of the gene pair, one within both of the syntenic regions. Parsimony suggests that the unlinked Mo17-specific sequences are unlikely to be the result of both polymorphic transposition events and the removal of one member of a gene pair in B73.
Contrasting QTL Discovery in RIL and NIL Populations

In general, NILs can be used for QTL detection and corroboration in three ways: (1) comparison of means across genotype classes using the entire population; (2) comparison of individual lines, or lines sharing a common introgression at a specific position, relative to the recurrent parent; and (3) evaluation of a genetic model developed from a separate population. All three approaches have strengths and weaknesses.

Marker-by-marker tests for significance of genotype mean differences across the entire population are most similar to the analysis used in other population structures. One shortcoming of this type of analysis is that the average number of individuals contributing to each mean is highly skewed between the donor parent type (few replications) and the recurrent parent genotype (many replications). For example, using the set of materials described in this paper, comparisons in the B73-like set would, on average, have six lines with the Mo17 homozygous genotype and 94 with the B73 homozygous genotype. The differential amount of replication of genotypes dramatically reduces power to detect QTLs relative to the same number of progeny in balanced population structures such as RILs (Kaespler, 1997). This disadvantage is offset for QTLs that are detected if the goal is high-resolution mapping, as the NIL is an excellent starting point for detailed analysis of specific QTLs. In addition, crossing pairs of NILs allows more specific and powerful tests of epistasis than genome-wide evaluation of all possible interactions.

Statistical comparison of the performance of individual NILs to the recurrent parent is a useful way to assess the effect of specific chromosome regions without growing the entire population, or it can be used to evaluate individuals within the population. If a NIL contains a single introgressed fragment, then a significant difference in performance unambiguously assigns the effect to that single region physically bounded by the recombination break points. In this set of lines, most lines have multiple introgressed fragments (on average approximately eight), so a performance difference in any line relative to the recurrent parent confounds the effect of multiple regions. This confounding can be ameliorated by further purifying the lines prior to phenotypic analysis or by implementing a decision tree similar to that proposed by Szalma et al. (2007).

An experimental goal might also be the validation of previously detected QTLs. Individual QTLs could be validated using the above approaches, although lack of validation might be due to statistical power, so interpretations are more relevant to confirm than to discount QTLs detected in other samples or population types. Another approach is to test whether a multiple-QTL model predicts performance within this independent sample. The model would likely be developed in a previous experiment and set of environments, although there might be scenarios where all materials (i.e., RIL and NIL populations) are grown in a single environment. Evaluation of different populations in different environments has the potential to reduce the predictive value of a model due to genotype by environment interactions. Decisions on the use of NIL and RIL populations will often be based on subsequent research goals, with NILs being of particular use as a starting point for fine-mapping of specific QTLs and to study epistatic interactions of targeted QTLs.

MATERIALS AND METHODS

Plant Materials

A set of 186 maize (Zea mays ssp. mays) B73-like and 70 Mo17-like NILs were developed from parental inbreds B73 and Mo17. Initial backcrosses and self-crosses were conducted in Madison, Wisconsin. Subsequent self-crosses were conducted at the University of Minnesota Agricultural Research Station in Falcon Heights. Supplemental Figure S1 provides an outline of the crossing scheme used to develop the NIL population. Genotype information was obtained for 150 of these lines, and seed is available from the Maize Genetics Stock Center (http://maizecoop.cropssci.uiuc.edu/) for each of these genotypes. Plant height measurements were also obtained for the IBM genotypes (Lee et al., 2002) and for the North Carolina Recombinant Inbred (NCR) B73-Mo17 population (Senior et al., 1996).

Array Design

A custom 12 × 135K long-oligonucleotide microarray was designed by Roche NimbleGen using sequence predicted from the B73 reference genome (Schnable et al., 2009). Probes were selected as a subset used for a previously described 2.1M feature array (Springer et al., 2009; Fu et al., 2010). The 2.1M feature array was used to identify several distinct categories of probe based on their hybridization behavior from B73 and Mo17 samples. These categories included 61,865 probes where the average signal should be higher in the B73 genotype when compared with the Mo17 genotype (B > M), 4,944 probes in which the average signal should be higher in Mo17 compared with B73 (B < M), and 34,222 probes in which the signal should be equal across both B73 and Mo17 (B = M). The 61,865 B>M probes located on all 10 maize chromosomes (Supplemental Table S6) were used for mapping and the 4,944 B < M probes were used to map Mo17-specific duplications.

DNA Labeling and Microarray Hybridization

DNA samples from NIL seedlings were isolated (Saghai-Maroof et al., 1984) from aboveground seedling tissue. DNA (1–2 µg) samples were labeled, amplified, and hybridized for 72 to 96 h at 42°C according to the array manufacturer’s protocol (NimbleGen Arrays User’s Guide: CGH Analysis version 5.1). Slides were washed and immediately scanned using the GenePix 4000B Scanner (Molecular Devices) according to the array manufacturer’s protocol. Array images and data were processed using NimbleScan software version 2.6. Samples showing a mean experimental:mean random probe hybridization ratio greater than 1.4 verified experimental integrity and provided arrays usable for analysis. A total of 154 samples (103 B73-like and 51 Mo17-like) from the initial 256 samples met this threshold and were used for subsequent analyses. There were four replicated genotypes, leading to a total of 100 unique B73-like and 50 unique Mo17-like genotypes.

Data Normalization

To minimize local signal variation found across any individual array, the DNAcopy algorithm was used to produce spatially normalized hybridization values for all probes for the 154 samples using NimbleScan version 2.6 (Roche NimbleGen). Probes that map to more than one location within the B73 genome, or that produced technical artifacts upon hybridization, were removed, leaving 129,597 total probes for analysis. To allow for interarray comparisons, the average hybridization values of the B = M probes for each array were calculated and normalized to a signal value of 5,500.
Mapping of NIL Lines

Each NIL sample from an array was compared with a control B73 genomic DNA hybridized to each array. Log base 2 ratios of raw (nonlog) NIL hybridization signal against B73 were developed for 62,995 B > M probes. To develop a map based on BAC calls, log₂(NIL/B73) values were averaged by their corresponding BACs. BACs containing less than three probes were removed from analysis, leaving 7,313 BAC values for mapping.

Mapped segments of B73-like or Mo17-like introgressions were first mapped for the set of 7,313 BACs due to the robust average ratio value compared with the single probe list. DNAcopy version 1.22.1 from the R Bioconductor package (Gentleman et al., 2004) was used under default parameters to locate and define segments within the mapping BACs. Segments were initially classified by the average log₂ ratio value of the segment. A segment was called B73 like if log₂(NIL/B73) was greater than −0.15 and was called Mo17 like if log₂(NIL/B73) was less than −0.8. Uncalled segments were called B73 like, Mo17 like, or heterogeneous based on visual inspection of the DNAcopy segmental output for each NIL sample. A final visual inspection was conducted to confirm initial calls. For any line, ambiguous segments were not called to prevent erroneous inclusion of introgressed regions.

To facilitate the analysis of heterozygous regions as well as Mo17-specific duplications, a more robust subset of mapped NILs was developed. This was needed, as overall variability across arrays provided samples that could be mapped using averaged probe values per BAC, but the confidence of any single probe measurement was low. NIL samples with a difference in segment mean of more than 0.7 between B73-like segments and Mo17-like segments were selected as more robust samples. Having a difference of at least 0.7 between segment means provides the highest quality data differentiating between B73-like and Mo17-like regions and allows heterogeneous regions (if present) to be more clearly distinguished. Using this cutoff, 87 of the initial 154 mapped NIL samples were selected for detailed analysis of heterozygosity and Mo17-specific duplications.

Phenotypic Measurements and Analysis

Three maize populations were used for analysis, the 150 B73 × Mo17 backcrossed NILs, 2/8 NCRI progeny of a different B73 × Mo17 cross, and the IBM population. Measurements for 100 B73-like and 50 Mo17-like NILs were collected for one replication at three environments, two at the University of Minnesota Agricultural Research Station (Minnesota environments 1 and 2) and one at the University of Wisconsin West Madison Agricultural Research Station, over the summer of 2010. The Minnesota environments were planted 3 weeks apart from each other in separate areas of the field, providing for two very distinct environmental conditions. The NCRI B73 × Mo17 population (Senior et al., 1996) was measured in 1997 at the West Madison Agricultural Research Station with two replicates. The IBM population was grown at the University of Wisconsin Arlington Agricultural Research Station in 2009 and 2010, with two replicates each year.

Plant height was recorded from the soil surface to the node subtending the flag leaf of mature stalks for all populations. The data were analyzed using PROC GLM of SAS version 9.2 (SAS Institute) with the linear model for each NIL population (B73 like and Mo17 like): Yijk = μ + Ei + Gk + ϵijk, where Yijk was an observation of the k genotype (G) in the i environment (E), μ was the overall mean, and ϵ was the residual error. Genotype and environment were considered random effects. The presence of significant Spearman rank correlations across environments allowed the analysis for all environments combined.

The IBM data were analyzed using PROC GLM of SAS version 9.2 (SAS Institute) with the following linear model: Yijk = μ + Yij + RYij/i + Gk + Yi + Ck + ϵijk, where Yijk was an observation of the k genotype (G) in the i replicate (R) within the j year (Y), μ was the overall mean, and ϵ was the residual error. All effects were considered random. Significant Spearman rank correlations across environments (years) allowed the analysis of averages across environments. Heritability on a family mean basis was calculated using IBM data with职工 heritability (Fehr, 1987). Repeatability for the NIL data was calculated using the following formula:

\[ R = \frac{\sigma^2(G)}{\sigma^2(E) + \sigma^2(G)} \]

where \( \sigma^2(G) \) is the genotypic variance, \( \sigma^2(Y) \) is the genotype-by-year variance, and \( \sigma^2(E) \) is the error variance (Fehr, 1987). Repeatability for the NIL data was calculated using the following formula:

\[ R = \frac{\sigma^2(G)}{\sigma^2(E) + \sigma^2(G)} \]

Mapping of QTLs

A set of 196 SSR and RFLP markers on the 208 NCRI progeny and 1,340 SSR and RFLP markers on the 302 RILs of the IBM population were used for composite interval mapping with Windows QTL Cartographer version 2.5 (Wang et al., 2010). QTL discovery analysis on the NIL population was conducted using single-marker analysis on the average plant height across environments by comparing the means of each homozygous class. A t test was used to determine significance. Bonferroni correction for multiple testing was conducted by dividing the α value of 0.05 by 852 nonredundant tests (i.e., one t test per nonrecombinant segment). The test was not performed on markers where either parental class was not represented by at least two individual NILs.

To evaluate the effect of specific introgressions, the phenotypic differentiation of lines that contain a nonrecombinant segment of donor parent from lines that did not contain the segment was calculated. The nonrecombinant segments were determined based on sets of contiguous markers spanning the IBM 1 = log of the odds score confidence interval. Independent t test analysis was conducted (p < 0.05) for each segment in each population (B73 like and Mo17 like). This analysis was conducted for the B73-like and Mo17-like NILs separately, as meaningful comparison involved lines with the specific introgression versus the corresponding specific recurrent parent.

To validate the IBM plant height QTL model, deviation from the recurrent parent was predicted on all NIL lines based on the sum of allelic effects from QTLs identified in the IBM population. For example, if a Mo17-like NIL had the B73 allele at an IBM QTL, the predicted deviation from Mo17 for that NIL was increased or decreased based on the effect of that QTL in the IBM population. If the Mo17-like NIL had the Mo17 allele at every IBM QTL, the predicted deviation was zero; therefore, that individual’s plant height would be expected to be equal to Mo17. Only data from the two replicates from Minnesota were used for this analysis.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure S1. Crossing scheme to generate NILs.
Supplemental Figure S2. NIL composition based on physical map.
Supplemental Figure S3. Discovery of plant height QTL.
Supplemental Figure S4. Validation of plant height QTL.
Supplemental Table S1. Summary for each NIL genotype.
Supplemental Table S2. BAC-based genotyping calls for all NILs.
Supplemental Table S3. Probe-based genotyping calls for all NILs.
Supplemental Table S4. ANOVA for plant height in NILs.
Supplemental Table S5. ANOVA for plant height in IBMs.
Supplemental Table S6. Chromosomal distribution of markers.

ACKNOWLEDGMENTS

We gratefully acknowledge the assistance of Peter Hermanson, Christian Thorvaldson, and Mac Garrett in performing crosses and collecting phenotypic data and Ruth Swanson-Wagner for assistance with hybridizations. The Minnesota Supercomputing Institute provided access to software packages used for data analysis.

Received February 18, 2011; accepted June 20, 2011; published June 24, 2011.

LITERATURE CITED


Mapping of QTL for resistance to the Mediterranean corn boron attack using the inbreded B73 × Mo17 (IBM) population of maize. Theor Appl Genet 119: 1451–1459


