Theor Appl Genet (2004) 109: 508–514
DOI 10.1007/s00122-004-1666-0

ORIGINAL PAPER

B. Parisseaux · R. Bernardo

In silico mapping of quantitative trait loci in maize

Received: 8 December 2003 / Accepted: 20 March 2004 / Published online: 19 May 2004

Abstract Quantitative trait loci (QTL) are most often detected through designed mapping experiments. An alternative approach is in silico mapping, whereby genes are detected using existing phenotypic and genomic databases. We explored the usefulness of in silico mapping via a mixed-model approach in maize (*Zea mays* L.). Specifically, our objective was to determine if the procedure gave results that were repeatable across populations. Multilocation data were obtained from the 1995–2002 hybrid testing program of Limagrain Genetics in Europe. Nine heterotic patterns comprised 22,774 single crosses. These single crosses were made from 1,266 inbreds that had data for 96 simple sequence repeat (SSR) markers. By a mixed-model approach, we estimated the general combining ability effects associated with marker alleles in each heterotic pattern. The numbers of marker loci with significant effects—37 for plant height, 24 for smut [*Ustilago maydis* (DC.) Cda.] resistance, and 44 for grain moisture—were consistent with previous results from designed mapping experiments. Each trait had many loci with small effects and few loci with large effects. For smut resistance, a marker in bin 8.05 on chromosome 8 had a significant effect in seven (out of a maximum of 18) instances. For this major QTL, the maximum effect of an allele substitution ranged from 5.4% to 41.9%, with an average of 22.0%. We conclude that in silico mapping via a mixed-model approach can detect associations that are repeatable across different populations. We speculate that in silico mapping will be more useful for gene discovery than for selection in plant breeding programs.

Communicated by H.C. Becker

B. Parisseaux
Limagrain Verneuil Holding, Département Recherche, Zone Industrielle de la Métairie,
49 160 Longué Jumelles, France

R. Bernardo (✉)
Department of Agronomy and Plant Genetics, University of Minnesota,
411 Borlaug Hall, 1991 Upper Buford Circle
St. Paul, MN, 55108, USA
e-mail: berna022@umn.edu
Tel.: +1-612-6256282
Fax: +1-612-6251268

Introduction

Quantitative trait loci (QTL) in plants have usually been detected through designed mapping experiments: an F_2 or backcross mapping population is first created by crossing two inbreds, and a set of progenies is evaluated in a set of locations and years (Dudley 1993; Kearsey and Farquhar 1998). An alternative approach is to map QTL from data that are routinely generated in a breeding program. "In silico mapping" is defined as the use of existing phenotypic and genomic databases for detecting genes (Grupe et al. 2001).

Public and private plant breeding programs in major crop species have accumulated massive amounts of phenotypic data for different traits, but these data are underutilized in QTL mapping. Compared with designed mapping experiments, in silico mapping for quantitative traits in plant breeding programs has four advantages. First, in silico mapping exploits large mapping populations. In maize, for example, thousands of experimental hybrids are evaluated each year (Smith et al. 1999). In contrast, the small populations (e.g., fewer than 500 progenies) often used in designed mapping experiments lead to a low power for detecting QTL (Melchinger et al. 1998), overestimation of QTL effects (Beavis 1994), and imprecise estimates of QTL location (van Ooijen 1992; Visscher et al. 1996). Second, the experimental hybrids or inbreds are evaluated in multiple, diverse environments. An experimental maize hybrid is typically evaluated in 20 environments; those that are eventually released as cultivars are evaluated in up to 1,500 location-year combinations (Smith et al. 1999). The use of many environments permits the sampling of a sufficient set of QTL × environment interactions so that the results would be applicable across a wide range of future environments. Third, the hybrids and inbreds tested typically comprise a

wide sample of the germplasm and genetic backgrounds that have been found useful. In contrast, only a narrow genetic background is exploited in designed mapping experiments that use F_2 or backcross populations. Fourth, the data used for in silico mapping are already available at no extra cost.

Offsetting these advantages are two main complications to in silico mapping. First, the performance data are highly unbalanced: some hybrids or inbreds are evaluated in one set of environments, whereas other hybrids or inbreds are evaluated in a different set of environments. This results from the screening process by which hybrids or inbreds that fail to perform well are discarded, those that perform well are subjected to more testing, and new hybrids or inbreds are tested for the first time each year. Second, the hybrids or inbreds do not comprise a single homogenous population. Any in silico mapping procedure would therefore have to account for pedigree relationships and differences in the genetic backgrounds among hybrids or inbreds.

The mixed-model approach, which leads to best linear unbiased predictions (BLUP) of random genetic effects and best linear unbiased estimates (BLUE) of fixed environmental effects, has been successfully adapted in plants (Bernardo 1994, 1996a; Panter and Allen 1995). The mixed-model approach integrates both phenotypic information and pedigree information in genetic evaluation. The integration of genomic data in mixed-model approach as originally proposed by Kennedy et al. (1992) is therefore a logical extension of BLUP and BLUE with phenotypic and pedigree data. Preliminary results in maize indicate that the mixed-model approach is potentially useful for in silico mapping of QTL (Bernardo 1998). One criterion for assessing the usefulness of in silico mapping is whether the procedure leads to repeatable results across different populations. With this criterion, our objective was to determine the usefulness of in silico mapping for detecting QTL across data sets produced by a maize breeding company.

Table 1 Estimates of variances and heritability (h^2) (ignoring marker data) in different heterotic patterns of maize

Group		Number of single crosses	V_R	$V_{GCA(1)}$	$V_{GCA(2)}$	V_{SCA}	V_{SCA}/V_G	h^2
1	2							
Plant height								
A	C	313	214.6	56.7	18.3	6.2	0.08[a]	0.71[b]
C	D	232	272.5	16.3	121.6	7.0	0.05	0.78
D	E	884	175.8	104.6	62.2	7.4	0.04	0.89
G	I	644	137.1	44.9	69.5	7.0	0.06	0.81
C	F	386	71.3	64.9	121.6	12.0	0.06	0.92
G	F	500	73.3	78.8	119.8	12.1	0.06	0.93
B	E	1,192	117.2	181.3	122.6	20.4	0.06	0.94
H	F	535	73.3	67.4	116.8	19.8	0.10	0.92
H	I	1,037	140.9	58.1	61.7	14.1	0.11	0.86
Smut resistance								
A	C	716	25.58	5.76	8.05	2.16	0.14	0.65
C	D	361	11.65	2.55	0.40	0.11	0.04	0.48
D	E	388	5.28	1.29	0.22	1.03	0.40	0.73
G	I	1,135	35.86	8.97	7.76	1.37	0.08	0.58
C	F	660	31.04	15.25	11.37	2.32	0.08	0.71
G	F	789	39.14	9.72	8.61	0.60	0.03	0.57
B	E	342	3.78	2.21	0.08	1.00	0.30	0.82
H	F	963	37.85	14.58	10.58	3.01	0.11	0.64
H	I	1,567	35.18	11.84	7.69	1.73	0.08	0.65
Grain moisture								
A	C	1,971	3.24	3.37	2.21	0.18	0.03	0.97
C	D	1,170	3.76	2.25	3.24	0.32	0.05	0.96
D	E	1,694	3.50	3.24	3.83	0.28	0.04	0.98
G	I	3,457	2.54	1.53	4.83	0.20	0.03	0.97
C	F	2,069	2.14	1.94	2.24	0.14	0.03	0.96
G	F	2,445	2.26	1.60	2.19	0.16	0.04	0.96
B	E	1,634	3.07	2.52	3.06	0.27	0.05	0.97
H	F	3,212	2.17	2.85	2.04	0.18	0.04	0.96
H	I	5,122	4.20	3.42	7.34	[c]		0.97

[a] $V_G=V_{GCA(1)}+V_{GCA(2)}+V_{SCA}$

[b] Heritability among single crosses on an entry-mean basis, calculated as $h^2=V_G/[V_G+V_R/$ (average number of locations for each single cross)]

[c] Not calculated due to an extremely large data set, i.e., 5,122 single crosses with grain-moisture data available

Materials and methods

Germplasm and pedigree, marker, and performance data sets

A total of 1,266 inbreds used in the Limagrain Genetics maize breeding program in Europe were classified into nine different heterotic groups, denoted by A to I. There were 89 inbreds in group A, 54 in B, 130 in C, 56 in D, 136 in E, 108 in F, 104 in G, 338 in H, and 251 in I. Groups A and I comprised flint-dent inbreds, F comprised flint inbreds, and the rest of the heterotic groups comprised dent inbreds. From pedigree records, the coefficient of coancestry among inbreds within each heterotic group was calculated by tabular analysis (Emik and Terrill 1949). Within each heterotic group, the minimum coefficient of coancestry between inbreds was zero, i.e., unrelated inbreds. The maximum coefficient of coancestry between inbreds was 0.53 in group B, 0.63 in groups C and G, and 0.75 in groups A, D, E, F, H, and I. The genetic model for mixed-model analysis of single crosses required the assumption that inbreds belonging to different heterotic groups were unrelated.

Each inbred was fingerprinted using 96 SSR markers spread across the maize genome. The size of the consensus map was 1,571 cM, and the number of markers per chromosome ranged from six on chromosome 10 to 15 on chromosome 1. The average number of bands per SSR marker was 12.7. Standard protocols (http://www.maizegdb.org/documentation/maizemap/ssr_protocols.php) were used for SSR analysis. For proprietary reasons we are unable to provide information on the specific SSR markers used as well as their map positions. This lack of information, however, did not hinder us in our objective of determining the repeatability of results from in silico mapping.

A heterotic pattern comprised a pair of complementary heterotic groups. Nine heterotic patterns were considered. The hybrid performance data set comprised results from multilocation yield trials conducted by Limagrain Genetics from 1995 to 2002. Each trial was conducted at 1–35 location(s) (with an average of six) in France, Germany, Hungary, and Italy in a randomized complete block design with one or two replications at each location. The performance data set was highly unbalanced across multilocation trials but (ignoring occasional missing plots) was balanced within each multilocation trial. Each data point in the subsequent analysis comprised the average performance of a single cross at several locations in a multilocation trial, rather than the performance at an individual location. Data for plant height (centimeters), resistance to common smut [*Ustilago maydis* (DC.) Cda., percentage of infected plants per plot], and grain moisture (percentage) were analyzed. For grain moisture, the number of tested single crosses ranged from 1,170 in the C × D pattern to 5,122 in the H × I pattern, for a total of 22,774 single crosses across the nine heterotic patterns (Table 1). However, not all of the 22,774 single crosses were evaluated for plant height or smut resistance.

Mixed-model analysis

Preliminary analysis without marker data, performed as outlined by Bernardo (1996a), indicated that specific combining ability was minor for grain moisture, plant height, and smut resistance (Table 1). Specific combining ability was therefore ignored in the analysis as recommended by Bernardo (1996b). Heritability among single-cross means was high for each trait, ranging from 0.71 to 0.94 for plant height, from 0.48 to 0.82 for smut resistance, and from 0.96 to 0.98 for grain moisture (Table 1).

Consider a heterotic pattern between two heterotic groups, group 1 and 2. Suppose n single crosses were made between n_1 inbreds from group 1 and n_2 inbreds from group 2. The single crosses were evaluated in t different multilocation yield trials, resulting in p total data points. A total of l_1 SSR marker alleles were present among the group 1 inbreds, whereas a total of l_2 marker alleles were present among the group 2 inbreds. With marker data, the linear model for the performance of single crosses was

$$y = X\beta + M_1\alpha_1 + M_2\alpha_2 + Z_1 g_1 + Z_2 g_2 + e$$

where $\mathbf{y}=p\times 1$ vector of observed performance for a given trait; $\beta = t \times 1$ vector of fixed effects associated with multilocation trials; $\alpha_1 = l_1 \times 1$ vector of general combining ability (GCA) effects associated with marker alleles in group 1; $\alpha_2 = l_2 \times 1$ vector of GCA effects associated with marker alleles in group 2; $\mathbf{g_1}=n_1\times 1$ vector of GCA effects not associated with marker alleles of group 1 inbreds; $\mathbf{g_2}=n_2\times 1$ vector of GCA effects not associated with marker alleles of group 2 inbreds; $\mathbf{e}=p\times 1$ vector of residual effects; and $\mathbf{X}$, $\mathbf{M_1}$, $\mathbf{M_2}$, $\mathbf{Z_1}$, and $\mathbf{Z_2}$ were incidence matrices of 1s and 0s relating y to β, α_1, α_2, g_1, and g_2, respectively.

The effects associated with multilocation yield trials (β) were assumed fixed, although the individual environments—whose effects were not specified in the model—were assumed random (Bernardo 1996a). As is assumed in designed mapping experiments for QTL and as proposed by Kennedy et al. (1992), the GCA effects of marker alleles(α_1 and α_2) were assumed fixed. The GCA not accounted for by the markers ($\mathbf{g_1}$ and $\mathbf{g_2}$) as well as $\mathbf{e}$ were assumed random. The variances of these random effects were $\mathrm{Var}(\mathbf{g_1}) = \mathbf{G_1}V_{\mathrm{GCA}(1)}$, $\mathrm{Var}(\mathbf{g_2}) = \mathbf{G_2}V_{\mathrm{GCA}(2)}$, and $\mathrm{Var}(\mathbf{e}) = \mathbf{R}V_{\mathrm{R}}$, where $\mathbf{G_1}=n_1\times n_1$ matrix of coefficients of coancestry among group 1 inbreds; $\mathbf{G_2}=n_2\times n_2$ matrix of coefficients of coancestry among group 2 inbreds; $\mathbf{R}=p\times p$ matrix where the off-diagonal elements were zero and the ith diagonal element was the reciprocal of the number of locations for the ith data point in $\mathbf{y}$; $V_{\mathrm{GCA}(1)}$= variance of GCA effects not associated with marker alleles in group 1; $V_{\mathrm{GCA}(2)}$ =variance of GCA effects not associated with marker alleles in group 2; and V_{R}=residual variance.

The BLUE of β, α_1, and α_2 and BLUP of $\mathbf{g_1}$ and $\mathbf{g_2}$ were obtained by solving the following mixed-model equations:

$$\begin{bmatrix} \hat{\beta} \\ \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \hat{\mathbf{g}}_1 \\ \hat{\mathbf{g}}_2 \end{bmatrix} = \begin{bmatrix} \mathrm{X}'\mathrm{R}^{-1}\mathrm{X} & \mathrm{X}'\mathrm{R}^{-1}\mathrm{M}_1 & \mathrm{X}'\mathrm{R}^{-1}\mathrm{M}_2 & \mathrm{X}'\mathrm{R}^{-1}\mathrm{Z}_1 & \mathrm{X}'\mathrm{R}^{-1}\mathrm{Z}_2 \\ & \mathrm{M}'_1\mathrm{R}^{-1}\mathrm{M}_1 & \mathrm{M}'_1\mathrm{R}^{-1}\mathrm{M}_2 & \mathrm{M}'_1\mathrm{R}^{-1}\mathrm{Z}_1 & \mathrm{M}'_1\mathrm{R}^{-1}\mathrm{Z}_2 \\ & & \mathrm{M}'_2\mathrm{R}^{-1}\mathrm{M}_2 & \mathrm{M}'_2\mathrm{R}^{-1}\mathrm{Z}_1 & \mathrm{M}'_2\mathrm{R}^{-1}\mathrm{Z}_2 \\ (\text{symmetric}) & & & \mathrm{Z}'_1\mathrm{R}^{-1}\mathrm{Z}_1+\theta_1 & \mathrm{Z}'_1\mathrm{R}^{-1}\mathrm{Z}_2 \\ & & & & \mathrm{Z}'_2\mathrm{R}^{-1}\mathrm{Z}_2+\theta_2 \end{bmatrix}^{-} \begin{bmatrix} \mathrm{X}'\mathrm{R}^{-1}\mathrm{y} \\ \mathrm{M}'_1\mathrm{R}^{-1}\mathrm{y} \\ \mathrm{M}'_2\mathrm{R}^{-1}\mathrm{y} \\ \mathrm{Z}'_1\mathrm{R}^{-1}\mathrm{y} \\ \mathrm{Z}'_2\mathrm{R}^{-1}\mathrm{y} \end{bmatrix}$$

$$= \begin{bmatrix} \mathrm{C}_{11} & \mathrm{C}_{12} & \mathrm{C}_{13} & \mathrm{C}_{14} & \mathrm{C}_{15} \\ & \mathrm{C}_{22} & \mathrm{C}_{23} & \mathrm{C}_{24} & \mathrm{C}_{25} \\ & & \mathrm{C}_{33} & \mathrm{C}_{34} & \mathrm{C}_{35} \\ & (\text{symmetric}) & & \mathrm{C}_{44} & \mathrm{C}_{45} \\ & & & & \mathrm{C}_{55} \end{bmatrix} \begin{bmatrix} \mathrm{X}'\mathrm{R}^{-1}\mathrm{y} \\ \mathrm{M}'_1\mathrm{R}^{-1}\mathrm{y} \\ \mathrm{M}'_2\mathrm{R}^{-1}\mathrm{y} \\ \mathrm{Z}'_1\mathrm{R}^{-1}\mathrm{y} \\ \mathrm{Z}'_2\mathrm{R}^{-1}\mathrm{y} \end{bmatrix}$$

where $\theta_1 = G_1^{-1}V_R/V_{GCA(1)}$ and $\theta_2 = G_2^{-1}V_R/V_{GCA(2)}$. Restricted maximum likelihood estimates of the variances were obtained by iterating on (Henderson 1985)

V_R=[$\mathbf{y'R^{-1}y}$−(solution vector) (right-hand side vector)]/[p −(number of estimable fixed effects)]
$V_{GCA(1)}$=[$\mathbf{g_1'G_1^{-1}g_1}$+V_R trace ($\mathbf{G_1^{-1}C_{44}}$)]/n_1
$V_{GCA(2)}$=[$\mathbf{g_2'G_2^{-1}g_2}$+V_R trace($\mathbf{G_2^{-1}C_{55}}$)]/n_2

GCA effects of marker alleles

For each heterotic pattern, GCA effects of marker alleles were analyzed in two steps. In the first step, marker loci with significant GCA effects were identified within each chromosome. We chose a backwards elimination procedure because it allowed the examination of the full model (i.e., all markers) for each chromosome. Because the overspecified effects approach (White and Hodge 1989, p. 306) was used in the model, the GCA effect of an individual marker allele was not estimable, but the difference in GCA effects between two marker alleles at the same locus was estimable. Significance (P=0.001) of pairwise differences in marker GCA effects within a locus was determined by z-tests, given that the variance of $\hat{\alpha}_1$ was $\mathbf{C_{22}}V_R$ and the variance of $\hat{\alpha}_2$ was $\mathbf{C_{33}}V_R$ (Henderson 1985). A marker locus was then declared significant if it had at least one significant within-locus pairwise difference.

In the second step, an across-genome analysis was performed using those markers that had significant GCA effects in the per-chromosome analyses. Marker loci with significant effects (P=0.05) were retained. At each significant locus, the maximum effect of an allele substitution was calculated as the maximum GCA effect minus the minimum GCA effect at the same locus.

We counted the number of times a marker locus was found significant across the nine heterotic patterns. A given marker locus can be significant twice because the tests for marker GCA effects were performed independently for each of the two heterotic groups in a heterotic pattern. With nine heterotic patterns, a marker locus can thus be significant up to 18 times in this study.

All of the analyses were performed using proprietary software, written by R. Bernardo in 1994–1996 and by B. Parisseaux in 2003 and owned by Limagrain Verneuil Holding. In silico mapping via mixed-model analysis was computer-intensive; a typical analysis for one chromosome and one trait required two to three days of computer time on a 1.2 GHz Pentium III machine with 512 MB of RAM.

Results

The number of SSR loci (out of 96) that had a significant GCA effect in at least one heterotic group was 37 for plant height, 24 for smut resistance, and 44 for grain moisture (Fig. 1). The number of significant marker loci on each chromosome ranged from zero on chromosome 2 to nine on chromosome 1 for plant height; from zero on chromosome 3 to six on chromosome 4 for smut

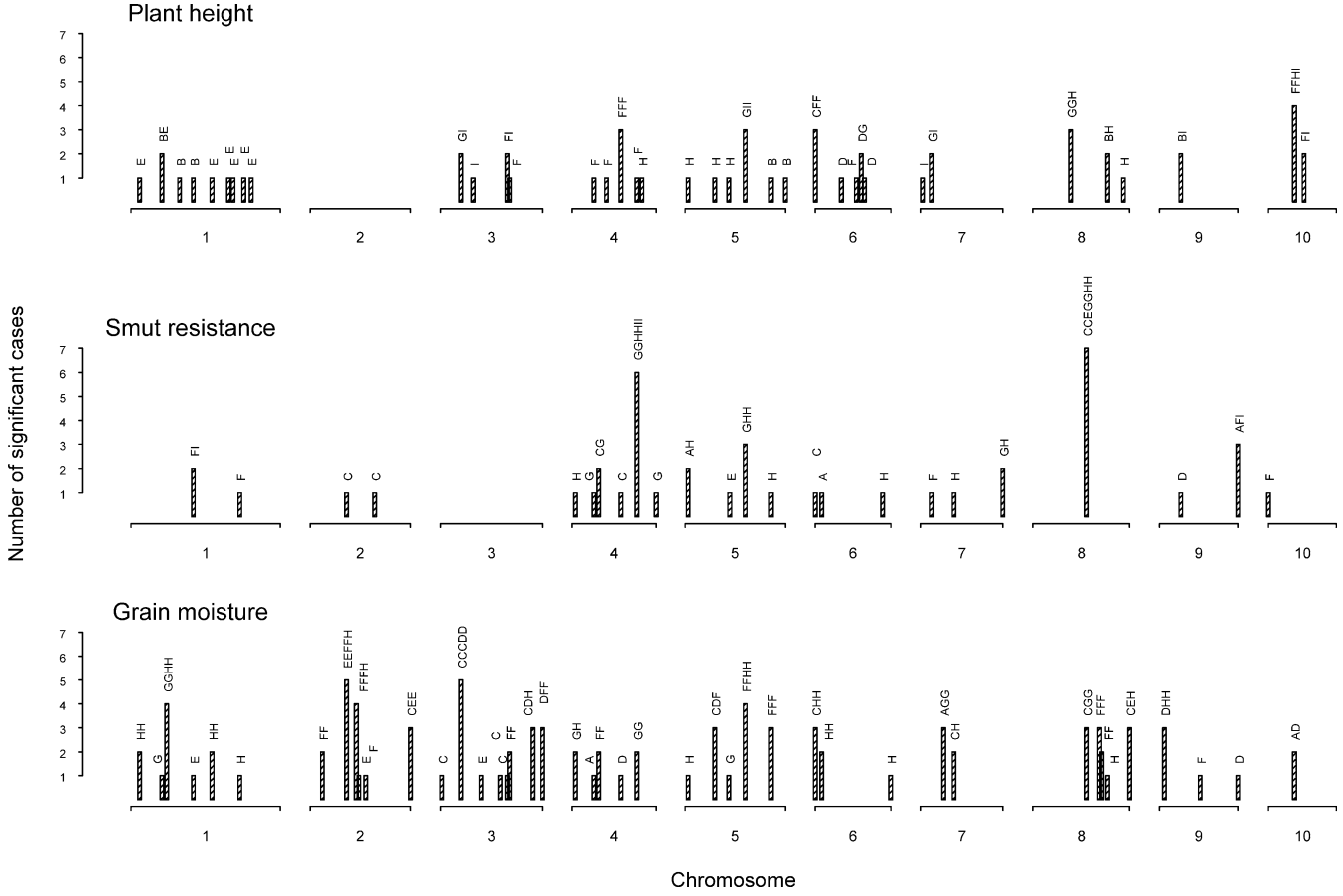

Fig. 1 Frequency of significant marker general combining ability (GCA) effects for plant height, resistance to common smut, and grain moisture in maize. Heterotic groups (denoted *A* to *I*) in which significant effects were detected are given *above each bar*. The chromosome sizes are drawn to scale on the *x-axis*

resistance; and from one on chromosome 10 to eight on chromosome 3 for grain moisture.

For plant height, most of the significant marker loci were detected only once or twice out of a possible maximum of 18 times (Fig. 1). For this trait, the maximum number of times a given marker had a significant effect was four, i.e., a marker on chromosome 10 that had a significant effect in two heterotic patterns that involved group F, one heterotic pattern that involved group H, and one heterotic pattern that involved group I (denoted by FFHI in Fig. 1). The results for grain moisture were more repeatable than those for plant height. For grain moisture, one marker on chromosome 2 (EEFFH) and one marker on chromosome 3 (CCCDD) were significant five times. The most striking example of repeatability was for smut resistance. One marker in bin 4.08 on chromosome 4 was significant six times (GGHHII), whereas one marker in bin 8.05 on chromosome 8 was significant seven times (CCEGGHH). This result for chromosome 8 was noteworthy because none of the nine other markers on the same chromosome had a significant effect in any heterotic group. We therefore consider this result as evidence of a major QTL for smut resistance on chromosome 8.

Some marker loci had significant effects in both groups of a heterotic pattern, whereas other marker loci had significant effects in only one group of a heterotic pattern. For example, one of the markers on chromosome 5 was significant for plant height in both groups of the G×I heterotic pattern. The estimated effects expressed as the maximum effect of an allele substitution at the locus were 17.3 cm in group G and 15.9 cm in group I. The five other markers on chromosome 5 that were significant for plant height had effects in one heterotic group only.

Marker loci on the same chromosome often had significant effects in certain heterotic groups but not in others. For plant height, the nine significant marker loci on chromosome 1 had effects detected only in heterotic groups E and B (Fig. 1). Out of these nine markers, seven had significant effects in group E and were all detected in the D×E heterotic pattern, whereas two had significant effects in group B and were both detected in the B×E heterotic pattern. For smut resistance, the two significant marker loci on chromosome 2 had effects detected only in group C. Four out of the six significant marker loci on chromosome 4 had effects detected in groups G and H. For grain moisture, the significant marker loci on chromosomes 1 and 6 were detected predominantly in group H.

The effects at marker loci varied. For the major QTL for smut resistance on chromosome 8, the maximum effect of an allele substitution ranged from 5.4% to 41.9%, with an average of 22.0%. All three traits studied had many loci with small effects and fewer loci with large effects (Fig. 2). In most cases, the maximum effect of an allele substitution was less than 35 cm for plant height, less than 15% for smut resistance, and less than 4% for grain moisture. There was clear evidence, however, of spurious effects for plant height. In seven instances, the maximum effect of an allele substitution was greater than 150 cm. This result was an artifact given that maize in France, Germany, Hungary, and Italy is generally 150–350 cm tall. An upper limit on the maximum effect of an allele substitution is therefore 350–150=200 cm, and it is unlikely that QTL would have effects that approach this limit.

Discussion

Number, location, and effects of QTL

The number of SSR markers with significant GCA effects was consistent with the level of complexity of the trait. We surmised that smut resistance was the simplest trait and grain moisture was the most complex trait and, correspondingly, we found that the number of significant markers was smallest for smut resistance and largest for grain moisture.

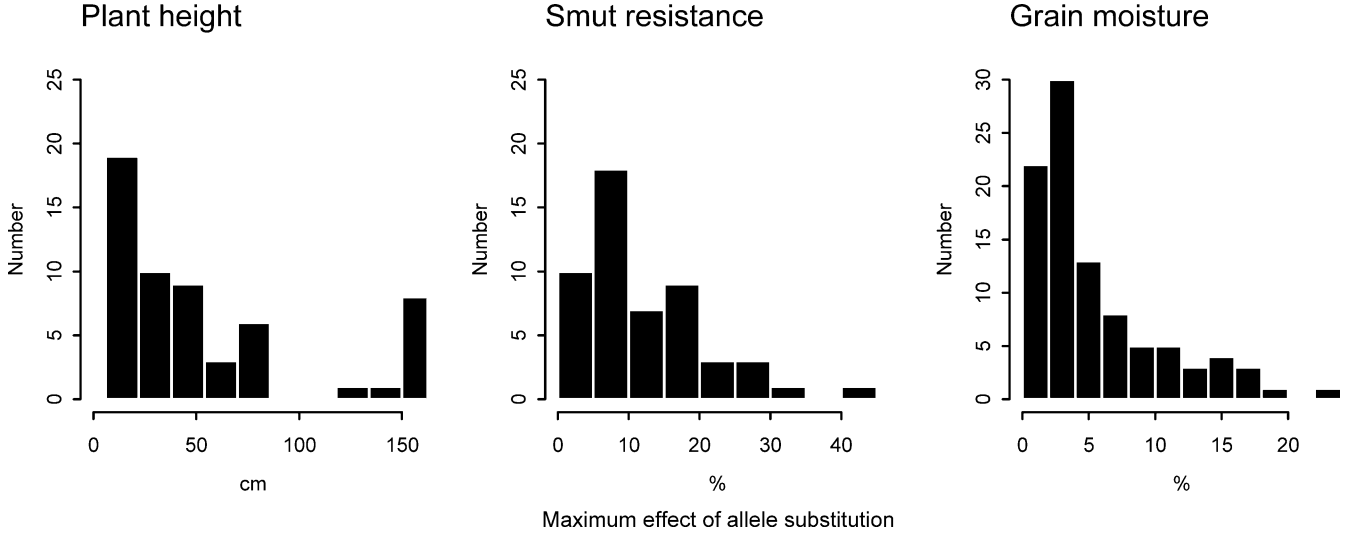

Fig. 2 Distribution of marker GCA effects for plant height, resistance to common smut, and grain moisture in maize. The maximum effect of an allele substitution was the maximum GCA effect minus the minimum GCA effect at the same marker locus

Because of differences in the germplasm used, the numbers of QTL we detected through in silico mapping were not directly comparable with those previously detected through designed mapping experiments. On the one hand, the wide range of germplasm sampled with in silico mapping enhances the detection of many QTL. On the other hand, mapping populations are often developed by crossing two parents that are widely divergent for a trait, e.g., susceptible parent and resistant parent for smut. A diverse mapping population also enhances the detection of many QTL. In the largest QTL mapping study published in maize (976 families from an F_2 population, genotyped with 172 markers and evaluated in 19 environments), Openshaw and Frascaroli (1997) detected 36 significant markers for plant height and 32 for grain moisture (data for smut resistance were absent). This result for plant height (36 QTL) was consistent with the number of significant markers we detected for plant height (37) via in silico mapping. For grain moisture, we detected a larger number of significant markers (44) than did Openshaw and Frascaroli (1997), perhaps because of a wider range of maturities sampled in our germplasm than in the single F_2 population used by Openshaw and Frascaroli (1997). For smut resistance, Lübberstedt et al. (1998) detected 19 significant markers across four populations, whereas Kerns et al. (1999) detected 22 significant markers in one population. These previous results were consistent with the number of significant markers (24) we detected for smut resistance.

Comparisons of QTL location among different studies are difficult because of differences in map sizes and in the sets of molecular markers used. Furthermore, meaningful comparisons are difficult to make because significant markers often span large chromosomal segments (e.g., chromosome 1 for plant height, Fig. 1), or many markers across the genome have significant effects so that a QTL detected in one study would tend to be also found in the same general chromosomal location in another study. Classic studies using chromosomal interchanges in maize (Burnham and Cartledge 1939; Saboe and Hayes 1941) indicated QTL for smut resistance on chromosome 8. The restriction fragment length polymorphism marker with the largest effect for smut resistance found by Kerns et al. (1999) was in bin 8.01 on chromosome 8. The major QTL we detected for smut resistance was in a different bin (8.05) on chromosome 8.

The distribution of gene effects we detected by in silico mapping (Fig. 2) was strikingly similar to previous summaries of gene effects from designed QTL mapping experiments (Kearsey and Farquhar 1998; Bernardo 2002, p. 310). As with previous studies, our results supported the model that quantitative traits are jointly controlled by many loci with small effects and few loci with large effects. But a few of our results for plant height (i.e., effects greater than 150 cm) indicate that caution is needed in interpreting large estimated effects. These spurious estimates were most likely due to multicolinearity or ill-conditioning in the data. By this we mean that because of linkage or statistical dependencies among the markers used, the effects at separate marker loci were difficult to estimate independently of each other. As Press et al. (1992, p. 56) noted, ill-conditioning "pulls the solution vector way off towards infinity along some direction" such that the estimates of effects become meaningless. We used standard methods to detect and reduce ill-conditioning (Press et al. 1992, p. 56) when solving the mixed-model equations, but more aggressive approaches seem needed.

Usefulness of in silico mapping via a mixed-model approach

The detection of a QTL requires that the QTL is segregating in the population being studied. This requirement has two important implications in our study. First, a marker found significant in one heterotic group but not in another does not necessarily imply an inconsistency due to the in silico mapping method. Rather, this result may have been due to QTL alleles being fixed in one heterotic group but not in another, i.e., a limitation of the germplasm studied rather than the method used. The detection of the same significant markers across heterotic groups does suggest, however, that in silico mapping via the mixed-model approach is useful. Second, the estimates of marker GCA effects were applicable only to the heterotic pattern in which the effects were estimated. Consider a marker allele (M_1) in a particular heterotic group (C). The GCA effects of M_1 are evaluated in the C×D and C×F heterotic patterns. As expected from a one-locus genetic model of testcross effects (Rawlings and Thompson 1962), the GCA effect of M_1 in group C would be different from the GCA effect of M_1 in group D. Furthermore, the GCA effect of M_1 in the C×D heterotic pattern would differ from its effect in the C×F heterotic pattern.

The two main phases in hybrid breeding are (1) development of inbreds and (2) identification of superior single-cross combinations among inbreds. Theoretical (Bernardo 1999) and empirical (Bernardo 1998) results indicate that QTL information is not useful in identifying superior single crosses. Empirical results in maize (Johnson 2001, 2004) indicate that, in the context of hybrid breeding, markers are useful mainly for improving the mean performance of a base population from which inbreds are later developed. The usefulness of in silico estimates of QTL effects for marker-assisted recurrent selection in a specific population needs to be studied. We speculate, however, that in silico mapping via a mixed-model approach will be more useful in gene discovery than in selection. In this study, the prime example of this potential for gene discovery was our detection of a major QTL for smut resistance on chromosome 8.

We used SSR markers because they were the marker system currently available. We envision that our approach could be used to detect associations at a finer genetic level (Mackay 2001), e.g., single nucleotide polymorphisms within candidate genes. Also, interval mapping could, in theory, be used in the in silico mapping procedure to estimate the location of QTL within a marker interval. But

the in silico mapping procedure without interval mapping is already computationally demanding. Expanding the methodology to include interval mapping would be computationally prohibitive.

In conclusion, in silico mapping via a mixed-model approach is useful from the standpoint that it can detect associations that are repeatable across different populations. The main practical advantage of this approach is that large data sets that already exist are exploited in QTL mapping. We are currently conducting simulation studies to determine the method's statistical power for detecting QTL both in self-pollinated crops and cross-pollinated crops, assuming different trait heritabilities, sample sizes, numbers of QTL, and marker densities.

References

Beavis WD (1994) The power and deceit of QTL mapping experiments: lessons from comparative QTL studies. Proc Corn Sorghum Ind Res Conf 49:250–266

Bernardo R (1994) Prediction of maize single-cross performance using RFLPs and information from related hybrids. Crop Sci 34:20–25

Bernardo R (1996a) Best linear unbiased prediction of maize single-cross performance. Crop Sci 36:50–56

Bernardo R (1996b) Testcross additive and dominance effects in best linear unbiased prediction of maize single-cross performance. Theor Appl Genet 93:1098–1102

Bernardo R (1998) Predicting the performance of untested single crosses: trait and marker data. In: Lamkey KR, Staub JE (eds) Concepts and breeding of heterosis in crop plants. Publ 25, Crop Science Society of America, Madison, Wis., pp 117–127

Bernardo R (1999) Marker-assisted best linear unbiased prediction of single-cross performance. Crop Sci 39:1277–1282

Bernardo R (2002) Breeding for quantitative traits in plants. Stemma Press, Woodbury, Minn.

Burnham CR, Cartledge JL (1939) Linkage relations between smut resistance and semi-sterility in maize. J Am Soc Agron 31:924–933

Dudley JW (1993) Molecular markers in plant improvement: manipulation of genes affecting quantitative traits. Crop Sci 33:660–668

Emik LO, Terrill CE (1949) Systematic procedures for calculating inbreeding coefficients. J Hered 40:51–55

Grupe A, Germer S, Usuka J, Aud D, Belknap JK, Klein RF, Ahluwalia MK, Higuchi R, Peltz G (2001) In silico mapping of complex disease-related traits in mice. Science 292:1915–1918

Henderson CR (1985) Best linear unbiased prediction of nonadditive genetic merits in noninbred populations. J Anim Sci 60:111–117

Johnson L (2001) Marker assisted sweet corn breeding: a model for specialty crops. Proc Corn Sorghum Ind Res Conf 56:25–30

Johnson GR (2004) Marker-assisted selection. Plant Breed Rev 24:293–309

Kearsey MJ, Farquhar AGL (1998) QTL analysis in plants; where are we now? Heredity 80:137–142

Kennedy BW, Quinton M, van Arendonk JAM (1992) Estimation of effects of single genes on quantitative traits. J Anim Sci 70:2000–2012

Kerns MR, Dudley JW, Rufener GK (1999) QTL for resistance to common rust and smut in maize. Maydica 44:37–45

Lübberstedt T, Klein D, Melchinger AE (1998) Comparative QTL mapping of resistance to *Ustilago maydis* across four populations of European flint-maize. Theor Appl Genet 97:1321–1330

Mackay TFC (2001) The genetic architecture of quantitative traits. Annu Rev Genet 35:303–339

Melchinger AE, Utz HF, Schön CC (1998) Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power for QTL detection and large bias in estimates of QTL effects. Genetics 149:383–403

Ooijen JW van (1992) Accuracy of mapping quantitative trait loci in autogamous species. Theor Appl Genet 84:803–811

Openshaw SJ, Frascaroli E (1997) QTL detection and marker-assisted selection for complex traits in maize. Proc Corn Sorghum Ind Res Conf 52:44–53

Panter DM, Allen FL (1995) Using best linear unbiased predictions to enhance breeding for yield in soybean: I. Choosing parents. Crop Sci 35:397–405

Press WH, Teukolshy SA, Vetterling WT, Flannery BP (1992) Numerical recipes in Fortran: the art of scientific computing, 2nd edn. Cambridge University Press, New York

Rawlings JO, Thompson DL (1962) Performance level as criterion for the choice of maize testers. Crop Sci 2:217–220

Saboe LC, Hayes HK (1941) Genetic studies of reactions to smut and firing in maize by means of chromosomal interchanges. J Am Soc Agron 33:463–470

Smith OS, Hoard K, Shaw F, Shaw R (1999) Prediction of single-cross performance. In: Coors JG, Pandey S (eds) The genetics and exploitation of heterosis in crops. American Society of Agronomy, Crop Science Society of America, Madison pp 277–285

Visscher PM, Thompson R, Haley CS (1996) Confidence intervals in QTL mapping by bootstrapping. Genetics 143:1013–1020

White TL, Hodge GR (1989) Predicting breeding values with applications in forest tree improvement. Kluwer, Dordrecht