

# Comparison of Mating Designs for Establishing Nested Association Mapping Populations in Maize and *Arabidopsis thaliana*

Benjamin Stich<sup>1</sup>

Max Planck Institute for Plant Breeding Research, 50829 Köln, Germany

Manuscript received August 11, 2009

Accepted for publication October 1, 2009

## ABSTRACT

The nested association mapping (NAM) strategy promises to combine the advantages of linkage mapping and association mapping. The objectives of my research were to (i) investigate by computer simulations the power and type I error rate for detecting quantitative trait loci (QTL) with additive effects using recombinant inbred line (RIL) populations of maize derived from various mating designs, (ii) compare these estimates to those obtained for RIL populations of *Arabidopsis thaliana*, (iii) examine for both species the optimum number of inbreds used as parents of the NAM populations, and (iv) provide on the basis of the results of these two model species a general guideline for the design of NAM populations in other plant species. The computer simulations were based on empirical data of a set of 26 diverse maize inbred lines and a set of 20 *A. thaliana* inbreds both representing a large part of the genetic diversity of the corresponding species. I observed considerable differences in the power for QTL detection between NAM populations of the same size but created on the basis of different crossing schemes. This finding illustrated the potential to improve the power for QTL detection without increasing the total resources necessary for a QTL mapping experiment. Furthermore, my results clearly indicated that it is advantageous to create NAM populations from a large number of parental inbreds.

**M**ANY traits that are important for fitness and agricultural value of plants are quantitative traits. Such traits are affected by many genes, the environment, and interactions between genes and the environment (HOLLAND 2007). In plants, quantitative trait locus (QTL) mapping is a key tool for studying the genetic architecture of quantitative traits (YANO 2001). This method enables the estimation of (i) the number of genome regions affecting a trait, (ii) the distribution of gene effects, and (iii) the relative importance of additive and nonadditive gene action.

Until now, most of the plant QTL mapping studies have been based on linkage mapping methods using individual biparental populations. The major limitations of such approaches are a poor resolution in detecting QTL and that with biparental crosses of inbred lines only two alleles at any given locus can be studied simultaneously (FLINT-GARCIA *et al.* 2005). Association mapping methods, which are successfully applied in human genetics to detect genes coding for human diseases (*e.g.*, WILLER *et al.* 2008), promise to overcome these limitations (KRAAKMAN *et al.* 2004). However, in

comparison with linkage mapping approaches, association mapping approaches have only a low power to detect QTL in genomewide scans (YU and BUCKLER 2006).

The nested association mapping (NAM) strategy proposed by YU *et al.* (2008) uses recombinant inbred line (RIL) populations derived from several crosses of parental inbreds. Due to diminishing chances of recombination over short genetic distance and a given number of generations, the genomes of these RILs are mosaics of chromosomal segments of their parental genomes. Consequently, within the chromosomal segments, the linkage disequilibrium (LD) information across the parental inbreds is maintained. Thus, if diverse parental inbreds are used, LD decays within the chromosomal segments of the RILs over a short physical distance (WILSON *et al.* 2004). Therefore, the NAM strategy allows to exploit both recent and ancient recombination and, thus, will show a high mapping resolution (YU *et al.* 2008). Furthermore, due to the balanced design underlying the proposed mapping strategy as well as the systematic reshuffling of the genomes of the parental inbreds during RIL development, NAM populations are expected to show a high power to detect QTL in genomewide approaches (BUCKLER *et al.* 2009).

Exploitation of the advantages of the NAM strategy requires developing, genotyping, and phenotyping of RIL populations from several crosses of diverse parental

Supporting information is available online at <http://www.genetics.org/cgi/content/full/genetics.109.108449/DC1>.

<sup>1</sup>Address for correspondence: Max Planck Institute for Plant Breeding Research, Carl-von-Linné-Weg 10, 50829 Köln, Germany.  
E-mail: stich@mpiz-koeln.mpg.de

inbreds. This, however, requires large financial resources (*cf.* YU *et al.* 2008). Therefore, it is mandatory that the available resources are spent in an optimum way.

STICH *et al.* (2009) examined the optimum allocation of resources for NAM in maize with respect to the number of RILs derived from the reference design as well as the number of environments and replications per environment used for phenotypic evaluation. The power for QTL detection, however, is expected to be influenced not only by these factors but also by the crossing scheme from which RIL populations are derived. To my knowledge, no study has so far compared RIL populations derived from various mating designs regarding the power for detecting QTL with additive effects. Furthermore, no information is available on the optimum number of inbreds used as parents of the NAM populations.

For *Arabidopsis thaliana*, more advanced genomic tools are available than for most other plant species (*e.g.*, ALONSO *et al.* 2003; CLARK *et al.* 2007). This fact increases the prospects of success of NAM approaches. However, *A. thaliana* differs from maize with respect to the genome size and the allele frequency, which both have the potential to influence the power for QTL detection. Nevertheless, to my knowledge, no study has so far examined the power of NAM in *A. thaliana*.

The objectives of my research were to (i) investigate by computer simulations the power and type I error rate for detecting QTL with additive effects using RIL populations of maize derived from various mating designs, (ii) compare these estimates to those obtained for RIL populations of *A. thaliana*, (iii) examine for both species the optimum number of inbreds used as parents of the NAM populations, and (iv) provide on the basis of the results of these two model species a general guideline for the design of NAM populations in other plant species.

## MATERIALS AND METHODS

**Simulations:** The computer simulations were based on empirical data of 653 single-nucleotide polymorphism (SNP) markers of 26 diverse maize inbred lines, namely B73, B97, CML52, CML69, CML103, CML228, CML247, CML277, CML322, CML333, Hp301, IL14H, Ki3, Ki11, Ky21, M37W, M162W, Mo18W, MS71, NC350, NC358, Oh7b, Oh43, P39, Tx303, and Tzi8 (YU *et al.* 2008). These inbreds were selected on the basis of 100 simple sequence repeat markers from a worldwide sample of 260 maize inbreds to capture the maximum genetic diversity (LIU *et al.* 2003). Details about SNP discovery, detection, and mapping were described by YU *et al.* (2008).

Furthermore, I used for my study empirical data of 653 SNP markers of 20 *A. thaliana* inbreds, namely Bay-0, Bor-4, Br-0, Bur-0, C24, Col-0, Cvi-0, Est-1, Fei-0, Got-7, Ler-1, Lov-5, Nfa-8, Rrs-7, Rrs-10, Sha, Tamm-2, Ts-1, Tsu-1, and Van-0 (CLARK *et al.* 2007). These inbreds were selected on the basis of polymorphisms in 876 genomewide distributed fragments from a sample of 96 *A. thaliana* genotypes to capture the maximum

TABLE 1

Number of crosses  $N_C$  underlying the segregating populations derived from reference (REF), diallel (DIA), factorial (FCT), single round-robin (SRR), double round-robin (DRR), reduced round-robin (RRR), independent round-robin (IRR), and distance-based (DB) designs

Mating design	Maize $N_C$	<i>A. thaliana</i> $N_C$
REF	25	19
DIA	325	190
FCT	169	100
SRR	26	20
DRR	52	40
RRR	25	20
IRR	13	10
DB <sub>15</sub>	50	30
DB <sub>30</sub>	100	60
DB <sub>60</sub>	200	120

For a detailed description of the designs, see MATERIALS AND METHODS.

genetic diversity (NORDBORG *et al.* 2005). The 653 SNP markers were selected from a set of 648,570 nonredundant SNP markers (MBML2 data set; CLARK *et al.* 2007; KIM *et al.* 2007; [ftp://ftp.arabidopsis.org/Polymorphisms/Perlegen\\_Array\\_Resequencing\\_Data\\_2007/SNP\\_predictions/](ftp://ftp.arabidopsis.org/Polymorphisms/Perlegen_Array_Resequencing_Data_2007/SNP_predictions/)) to uniformly cover the chromosomes (supporting information, File S1). Genetic map positions for these SNPs were lacking. Therefore, a linear model was applied to project the physical map position of the SNPs on the genetic map of SINGER *et al.* (2006).

*Mating designs evaluated:* The  $I = 26$  maize inbreds and the  $I = 20$  *A. thaliana* inbreds were used to examine 10 different mating designs using computer simulations (Figure S1). RILs were derived from the crosses of each design, where each RIL was assumed to be derived from a distinct  $F_2$  plant through single-seed descent with selfing to the  $F_6$  generation.

For the reference (REF) design in maize, RIL populations were derived from the crosses between B73 and the 25 diverse inbreds (MCMULLEN *et al.* 2009) (Table 1). In *A. thaliana*, RIL populations were derived from the crosses between Col-0 and the 19 inbred lines. For the diallel (DIA) design, a RIL population was derived from each of the crosses in the diallel (method 4; GRIFFING 1956) among all *I* maize or *A. thaliana* parental inbreds. For the factorial (FCT) design, the 26 maize inbreds or the 20 *A. thaliana* inbreds were randomly partitioned into two subsets of equal size and a RIL population was derived from each cross between the two sets of inbreds (COMSTOCK and ROBINSON 1948).

For the single round-robin (SRR) design (VERHOEVEN *et al.* 2006), RILs were derived from each of the chain crosses, *i.e.*, inbred 1  $\times$  inbred 2, inbred 2  $\times$  inbred 3, ..., inbred  $I$   $\times$  inbred 1. For the double round-robin (DRR) design, a RIL population was derived from each of the double-chain crosses, *i.e.*, inbred 1  $\times$  inbred 2, inbred 1  $\times$  inbred 3, inbred 2  $\times$  inbred 3, inbred 2  $\times$  inbred 4, ..., inbred  $I$   $\times$  inbred 1. For reduced round robin (RRR), RILs were derived from the reduced double-chain crosses, *i.e.*, inbred 1  $\times$  inbred 2, inbred 1  $\times$  inbred 3, inbred 2  $\times$  inbred 3, inbred 3  $\times$  inbred 4, inbred 5  $\times$  inbred 6, ..., inbred  $(I - 1)$   $\times$  inbred  $I$ . For the independent round-robin (IRR) design, a RIL population was derived from each of the independent chain crosses among the  $I$  inbreds, *i.e.*, inbred 1  $\times$  inbred 2, inbred 3  $\times$  inbred 4, ..., inbred  $(I - 1)$   $\times$  inbred  $I$ .

The data sets for the distance-based designs  $DB_p$  were established by selecting from all crosses in a diallel among the  $I$  inbreds the  $p\%$  combinations of parental inbreds, which show, on the basis of all marker loci, the maximum genetic dissimilarity (NEI and LI 1979). In the current study, the designs  $DB_{15}$ ,  $DB_{30}$ , and  $DB_{60}$  were examined.

In addition to the above-described simulations with  $I = 26$  and 20 parental inbreds for maize and *A. thaliana*, respectively, I examined exemplarily for the REF and DIA designs scenarios with the same total number of RILs  $N$  derived from crosses between a reduced number of parents  $I$ : For maize, I set  $I = 23, 20, \dots, 5$  and for *A. thaliana*  $I = 17, 14, \dots, 5$ . In the simulations of the REF design, the same reference parent (B73; Col-0) was chosen as initially described and the remaining  $I - 1$  parents were randomly selected from the entire set of parental inbreds. In contrast, for the DIA design,  $I$  parents were randomly selected from the entire set of parents.

For each of the above-described mating designs, which differ with respect to the number of crosses  $N_C$  (Table 1), I assumed a total number of RILs  $N = 1250, 2500$ , or 5000. The number of RILs per cross  $N_p$  was calculated as follows: In scenarios with the number of remaining RILs  $r = N \bmod N_C = 0$ ,  $N_p = N/N_C$ . In contrast, in scenarios with  $r \neq 0$ , I chose for  $r$  populations  $N_p = N/N_C + 1$ , whereas for the remaining  $N_C - r$  populations  $N_p = N/N_C$ .

**Definition of genotypic and phenotypic values:** A total of 100 simulation runs were performed for each of the examined mating designs. For each run, three subsets of SNPs ( $l = 25, 50, 100$ ) were sampled at random without replacement from the linkage map and were defined as QTL. The SNP markers of my study are biallelic and, thus, the 25 diverse maize inbreds or the 19 *A. thaliana* inbreds used as parents carry either the same allele as the reference parent (B73; Col-0) or the non-reference parent allele. At each QTL, one allele was assigned the genotypic effect zero whereas the genotypic effect of the other allele was drawn randomly without replacement from the geometric series  $l(1 - a)[1, a, a^2, a^3, \dots, a^{l-1}]$ , with  $a = 0.90$  (25 QTL),  $a = 0.96$  (50 QTL), or  $a = 0.99$  (100 QTL) (LANDE and THOMPSON 1990). Genotypic values of the inbreds were determined by summing up the effects of the individual alleles.

From the genotypic values of the RILs of each cross, the genotypic variance within the cross  $\sigma_g^2$  was calculated. For the progenies of each cross, the phenotypic values were generated by adding a realization from a normally distributed variable  $N(0, ((1 - h^2)/h^2)\sigma_g^2)$  to the genotypic values, where  $h^2$  denotes the heritability on an entry-mean basis. On the basis of previous empirical studies, I examined  $h^2$  values of 0.5 and 0.8 (FLINT-GARCIA *et al.* 2005). All simulations were performed with software PLABSOFT (MAURER *et al.* 2008), which is implemented as an extension of the statistical software R (R DEVELOPMENT CORE TEAM 2004).

**Statistical analyses:** The comparison of statistical analyses concerning the power  $1 - \beta^*$  requires an equal empirical type I error rate  $\alpha^*$ . To meet this requirement, I applied the following two-step procedure for QTL detection, which corresponds to that described by STICH *et al.* (2009). In a first step, stepwise multiple linear regression implemented in PLABQTL (UTZ and MELCHINGER 1996) was used to select a set of cofactors based on the SCHWARZ (1978) Bayesian criterion, using the model

$$y = \mu + \sum_i b_i x_i + e,$$

where  $y$  is the vector of the phenotypic values of all RILs,  $\mu$  is the intercept,  $b_i$  is the regression coefficient of the  $i$ th marker locus,  $x_i$  is an incidence vector of the genotypes of the RILs at the  $i$ th marker, and  $e$  is the vector of residual errors. I assumed

that all RILs are genotyped with such a high number of markers that each QTL has a marker that is in complete LD with the QTL. Therefore, all SNPs, inclusive of those treated as QTL, were included in the QTL detection procedure.

In the second step, I calculated a  $P$ -value for the association of each marker  $q$  with the phenotypic value for an  $F$ -test with a full model against a reduced model,

$$y = \mu + b_q x_q + \sum_{c \neq q} b_c x_c + e,$$

where  $b_q$  ( $b_c$ ) is the regression coefficient of the  $q$ th marker locus (or  $c$ th cofactor) and  $x_q$  ( $x_c$ ) is an incidence vector of the genotypes of the RILs at the  $q$ th marker ( $c$ th cofactor). The  $c \neq q$  in the above formula indicates that from the total set of cofactors only those cofactors are used in the  $F$ -test of a specific marker that are not identical to the marker under consideration. This constraint is inevitable to detect also those QTL for which a cofactor was selected in the first step.

In addition to the above-described procedure for QTL detection, I used a procedure that accounts for the structure of the simulated RIL populations by including the mean value of the RILs derived from each cross (*cf.* YU *et al.* 2008) in the model of each of the two above-described steps.

For each combination of  $N$ ,  $l$ , and  $h^2$  examined for each mating design, the nominal  $\alpha$ -level was chosen in such a way that the empirical type I error rate  $\alpha^*$  was 0.01 (Table S1). Due to the fact that none of the simulated QTL was monomorphic in any of the examined scenarios, the power for QTL detection ( $1 - \beta^*$ ) was calculated on the basis of this  $\alpha$ -level as the average proportion of QTL correctly identified from the total number of QTL  $l$ .

## RESULTS

For maize, the average map distance between the 653 SNP markers was 2.4 cM, whereas for *A. thaliana* the average map distance was 0.6 cM. The pairwise genetic dissimilarity among the 26 maize inbreds ranged from 0.25 to 0.42, where for *A. thaliana* values between 0.16 and 0.31 were observed. For maize, the average frequency of the allele of the reference parent B73 was 0.81 in the RILs of the REF design and ranged from 0.63 to 0.66 in the RILs of all other designs. In contrast, the frequency of the allele of the *A. thaliana* reference parent Col-0 was 0.89 in the RILs of the REF design and ranged from 0.76 to 0.79 in the RILs of all other designs.

For the 1250 RILs derived from the REF design of maize, a power to detect QTL  $1 - \beta^*$  of 0.603 was observed for the scenario with  $l = 25$  QTL and  $h^2 = 0.5$  (Table 2). The duplication or quadruplication of the number of QTL from 25 to 50 or 100 resulted in a decrease of  $1 - \beta^*$  to about three-fourths or one-third of the initial value, respectively. For  $l = 25$  QTL, an increase of  $h^2$  from 0.5 to 0.8 resulted in an increase of  $1 - \beta^*$  of about one half, where this increase was even more pronounced for  $l = 50$  and 100 than for  $l = 25$ . The duplication or quadruplication of the number of RILs  $N$  from 1250 to 2500 or 5000 resulted in a small increase of  $1 - \beta^*$  for  $l = 25$ , a medium increase for  $l = 50$ , and a large increase for  $l = 100$ . Furthermore, the increase of

TABLE 2

Power to detect QTL ( $\alpha^* = 0.01$ ) and the corresponding standard error for different numbers  $N$  of maize recombinant inbred lines derived from different mating designs: reference (REF), diallel (DIA), factorial (FCT), single round-robin (SRR), double round-robin (DRR), reduced round-robin (RRR), independent round-robin (IRR), and distance-based (DB) designs

Mating design	25 QTL		50 QTL		100 QTL	
	$h^2 = 0.5$	$h^2 = 0.8$	$h^2 = 0.5$	$h^2 = 0.8$	$h^2 = 0.5$	$h^2 = 0.8$
$N = 1250$						
REF	0.603 (0.0090)	0.961 (0.0037)	0.459 (0.0078)	0.934 (0.0042)	0.189 (0.0060)	0.888 (0.0047)
DIA	0.694 (0.0079)	0.981 (0.0028)	0.590 (0.0058)	0.979 (0.0024)	0.381 (0.0085)	0.978 (0.0014)
FCT	0.706 (0.0083)	0.986 (0.0023)	0.591 (0.0072)	0.978 (0.0021)	0.376 (0.0078)	0.978 (0.0016)
SRR	0.678 (0.0082)	0.986 (0.0023)	0.574 (0.0071)	0.977 (0.0021)	0.348 (0.0079)	0.976 (0.0019)
DRR	0.692 (0.0073)	0.982 (0.0028)	0.585 (0.0068)	0.976 (0.0022)	0.373 (0.0082)	0.979 (0.0016)
RRR	0.677 (0.0080)	0.982 (0.0030)	0.553 (0.0072)	0.972 (0.0028)	0.328 (0.0076)	0.967 (0.0042)
IRR	0.661 (0.0086)	0.978 (0.0033)	0.535 (0.0071)	0.965 (0.0030)	0.302 (0.0078)	0.970 (0.0018)
DB <sub>15</sub>	0.658 (0.0076)	0.973 (0.003)	0.529 (0.0075)	0.968 (0.0026)	0.284 (0.0073)	0.954 (0.0025)
DB <sub>30</sub>	0.692 (0.0077)	0.983 (0.0024)	0.559 (0.0067)	0.973 (0.0023)	0.330 (0.0067)	0.972 (0.0017)
DB <sub>60</sub>	0.686 (0.0084)	0.985 (0.0025)	0.589 (0.0061)	0.977 (0.0023)	0.352 (0.0084)	0.979 (0.0015)
$N = 2500$						
REF	0.734 (0.0072)	0.986 (0.0024)	0.600 (0.0073)	0.977 (0.0021)	0.317 (0.0072)	0.968 (0.0021)
DIA	0.816 (0.0073)	0.992 (0.0018)	0.753 (0.0059)	0.993 (0.0013)	0.639 (0.0079)	0.992 (0.0010)
FCT	0.799 (0.0067)	0.993 (0.0015)	0.737 (0.0062)	0.992 (0.0013)	0.631 (0.0084)	0.992 (0.0008)
SRR	0.790 (0.0080)	0.994 (0.0015)	0.734 (0.0071)	0.993 (0.0014)	0.596 (0.0084)	0.993 (0.0007)
DRR	0.808 (0.0068)	0.994 (0.0014)	0.756 (0.0063)	0.991 (0.0015)	0.636 (0.0079)	0.993 (0.0008)
RRR	0.796 (0.0076)	0.992 (0.0017)	0.718 (0.0059)	0.987 (0.0018)	0.600 (0.0082)	0.992 (0.0010)
IRR	0.768 (0.0078)	0.991 (0.0019)	0.698 (0.0064)	0.991 (0.0014)	0.546 (0.0100)	0.991 (0.0010)
DB <sub>15</sub>	0.773 (0.0078)	0.992 (0.0017)	0.686 (0.0066)	0.986 (0.0017)	0.482 (0.0091)	0.986 (0.0011)
DB <sub>30</sub>	0.805 (0.0064)	0.992 (0.0017)	0.720 (0.0064)	0.989 (0.0012)	0.563 (0.0093)	0.987 (0.0011)
DB <sub>60</sub>	0.810 (0.0067)	0.996 (0.0014)	0.749 (0.0060)	0.994 (0.0011)	0.626 (0.0079)	0.993 (0.0008)
$N = 5000$						
REF	0.829 (0.0079)	0.992 (0.0017)	0.754 (0.0074)	0.993 (0.0013)	0.515 (0.0090)	0.991 (0.0010)
DIA	0.910 (0.0064)	0.997 (0.0010)	0.883 (0.0052)	0.998 (0.0005)	0.866 (0.0051)	0.997 (0.0006)
FCT	0.900 (0.0060)	0.997 (0.0010)	0.871 (0.0049)	0.997 (0.0008)	0.858 (0.0053)	0.998 (0.0005)
SRR	0.901 (0.0061)	0.997 (0.0010)	0.866 (0.0048)	0.997 (0.0007)	0.842 (0.0065)	0.998 (0.0006)
DRR	0.918 (0.0044)	0.998 (0.0009)	0.872 (0.0045)	0.998 (0.0008)	0.844 (0.0057)	0.997 (0.0005)
RRR	0.890 (0.0066)	0.996 (0.0013)	0.857 (0.0057)	0.996 (0.0008)	0.816 (0.0073)	0.996 (0.0006)
IRR	0.889 (0.0066)	0.996 (0.0013)	0.836 (0.0059)	0.996 (0.0010)	0.772 (0.0076)	0.996 (0.0006)
DB <sub>15</sub>	0.886 (0.0065)	0.995 (0.0013)	0.827 (0.0059)	0.994 (0.0011)	0.753 (0.0073)	0.994 (0.0007)
DB <sub>30</sub>	0.897 (0.0050)	0.997 (0.0012)	0.851 (0.0056)	0.994 (0.0012)	0.792 (0.0071)	0.997 (0.0005)
DB <sub>60</sub>	0.909 (0.0052)	0.997 (0.0011)	0.878 (0.0054)	0.997 (0.0008)	0.848 (0.0058)	0.997 (0.0005)

$1 - \beta^*$  was more pronounced for  $h^2 = 0.5$  than for  $h^2 = 0.8$ . Across all scenarios of the REF design of maize, the nominal  $\alpha$ -level ranged from 0.0002 ( $N = 1250$ ;  $l = 100$ ;  $h^2 = 0.5$ ) to 0.0133 ( $N = 5000$ ;  $l = 25$ ;  $h^2 = 0.8$ ) (Table S1).

The  $1 - \beta^*$  trends observed for RIL populations derived from the non-REF designs upon changes of  $l$ ,  $h^2$ , and  $N$  were similar to that found for the REF design (Table 2). For  $N = 1250$ , the ranking of the various designs with respect to the  $1 - \beta^*$  values across all examined levels of  $l$  and  $h^2$  was FCT > DIA > DRR > DB<sub>60</sub> > SRR > DB<sub>30</sub> > RRR > IRR > DB<sub>15</sub> > REF. The duplication or quadruplication of  $N$  from 1250 to 2500 or 5000 resulted in a shift of FCT to rank four and of DB<sub>30</sub> to rank seven. The trends of the nominal  $\alpha$ -level observed for RIL populations derived from the non-REF designs upon changes of  $l$ ,  $h^2$ , and  $N$  were similar to that found for the REF design.

Across all examined scenarios, the  $1 - \beta^*$  values observed for *A. thaliana* were between one-tenth and one-fifth lower than those observed for maize (Table 3). The  $1 - \beta^*$  trends observed for *A. thaliana* RIL populations derived from all designs upon changes of  $l$ ,  $h^2$ , and  $N$  were similar to those found for maize. For *A. thaliana*, the ranking of the examined designs with respect to the  $1 - \beta^*$  values across all examined levels of  $l$ ,  $h^2$ , and  $N$  was FCT > DIA > DB<sub>60</sub> > DRR > SRR > DB<sub>30</sub> > RRR > DB<sub>15</sub> > IRR > REF. The trends of the nominal  $\alpha$ -level observed for *A. thaliana* RIL populations were similar to those found for maize.

Decreasing the number of parental inbreds involved in the development of RIL populations of maize resulted in a decrease of  $1 - \beta^*$  (Figure 1). This decrease of  $1 - \beta^*$  was more pronounced for *A. thaliana* than for maize. For both species, the decrease of  $1 - \beta^*$  upon the



TABLE 3

Power to detect QTL ( $\alpha^* = 0.01$ ) and the corresponding standard error for different numbers  $N$  of *Arabidopsis thaliana* recombinant inbred lines derived from different mating designs: reference (REF), diallel (DIA), factorial (FCT), single round-robin (SRR), double round-robin (DRR), reduced round-robin (RRR), independent round-robin (IRR), and distance-based (DB) designs

Mating design	25 QTL		50 QTL		100 QTL	
	$h^2 = 0.5$	$h^2 = 0.8$	$h^2 = 0.5$	$h^2 = 0.8$	$h^2 = 0.5$	$h^2 = 0.8$
$N = 1250$						
REF	0.514 (0.0108)	0.905 (0.0067)	0.337 (0.0074)	0.854 (0.0065)	0.137 (0.0044)	0.626 (0.0088)
DIA	0.615 (0.0095)	0.966 (0.0039)	0.492 (0.0072)	0.958 (0.0034)	0.264 (0.0063)	0.941 (0.0040)
FCT	0.643 (0.0103)	0.969 (0.0038)	0.488 (0.0078)	0.954 (0.0035)	0.250 (0.0057)	0.938 (0.0036)
SRR	0.609 (0.0095)	0.966 (0.0036)	0.460 (0.0075)	0.956 (0.0030)	0.231 (0.0054)	0.929 (0.0039)
DRR	0.624 (0.0082)	0.968 (0.0039)	0.467 (0.0081)	0.955 (0.0031)	0.241 (0.0063)	0.940 (0.0038)
RRR	0.597 (0.0085)	0.961 (0.0038)	0.446 (0.0063)	0.937 (0.0039)	0.211 (0.0057)	0.915 (0.0042)
IRR	0.576 (0.0107)	0.948 (0.0052)	0.406 (0.0079)	0.921 (0.0052)	0.170 (0.0055)	0.856 (0.0071)
DB <sub>15</sub>	0.576 (0.0101)	0.930 (0.0053)	0.421 (0.0067)	0.911 (0.0041)	0.201 (0.0062)	0.891 (0.0042)
DB <sub>30</sub>	0.603 (0.0089)	0.957 (0.0047)	0.450 (0.0066)	0.938 (0.0035)	0.221 (0.0063)	0.923 (0.0033)
DB <sub>60</sub>	0.627 (0.0083)	0.972 (0.0033)	0.481 (0.0077)	0.953 (0.0033)	0.257 (0.0066)	0.948 (0.0030)
$N = 2500$						
REF	0.640 (0.0101)	0.958 (0.0043)	0.473 (0.0082)	0.937 (0.0040)	0.218 (0.0050)	0.831 (0.0065)
DIA	0.766 (0.0084)	0.991 (0.0020)	0.636 (0.0076)	0.984 (0.0019)	0.430 (0.0077)	0.987 (0.0012)
FCT	0.754 (0.0077)	0.990 (0.0020)	0.655 (0.0079)	0.989 (0.0014)	0.436 (0.0078)	0.986 (0.0013)
SRR	0.741 (0.0090)	0.988 (0.0023)	0.622 (0.0078)	0.984 (0.0018)	0.397 (0.0067)	0.985 (0.0013)
DRR	0.744 (0.0085)	0.990 (0.0023)	0.632 (0.0065)	0.983 (0.0019)	0.401 (0.0068)	0.984 (0.0014)
RRR	0.736 (0.0088)	0.986 (0.0022)	0.608 (0.0079)	0.980 (0.0024)	0.366 (0.0073)	0.978 (0.0016)
IRR	0.709 (0.0096)	0.983 (0.0028)	0.547 (0.0088)	0.975 (0.0023)	0.302 (0.0060)	0.967 (0.0023)
DB <sub>15</sub>	0.692 (0.0091)	0.958 (0.0039)	0.571 (0.0074)	0.949 (0.0032)	0.358 (0.0088)	0.948 (0.0023)
DB <sub>30</sub>	0.728 (0.0093)	0.981 (0.0028)	0.608 (0.0064)	0.980 (0.0018)	0.383 (0.0079)	0.976 (0.0015)
DB <sub>60</sub>	0.750 (0.0081)	0.991 (0.0020)	0.650 (0.0071)	0.987 (0.0016)	0.414 (0.0083)	0.987 (0.0012)
$N = 5000$						
REF	0.744 (0.0094)	0.983 (0.0026)	0.632 (0.0076)	0.976 (0.0021)	0.338 (0.0067)	0.947 (0.0029)
DIA	0.863 (0.0070)	0.997 (0.0012)	0.796 (0.0070)	0.995 (0.0011)	0.667 (0.0085)	0.996 (0.0006)
FCT	0.864 (0.0068)	0.999 (0.0007)	0.804 (0.0060)	0.996 (0.0010)	0.660 (0.0083)	0.996 (0.0005)
SRR	0.851 (0.0069)	0.997 (0.0013)	0.783 (0.0074)	0.993 (0.0012)	0.626 (0.0084)	0.996 (0.0006)
DRR	0.854 (0.0067)	0.998 (0.0009)	0.785 (0.0072)	0.993 (0.0011)	0.629 (0.0079)	0.996 (0.0006)
RRR	0.842 (0.0072)	0.996 (0.0014)	0.751 (0.0075)	0.993 (0.0013)	0.592 (0.0092)	0.995 (0.0008)
IRR	0.830 (0.0084)	0.992 (0.0020)	0.707 (0.0083)	0.993 (0.0015)	0.486 (0.0098)	0.990 (0.0012)
DB <sub>15</sub>	0.807 (0.0083)	0.965 (0.0037)	0.728 (0.0069)	0.965 (0.0026)	0.577 (0.0095)	0.963 (0.0017)
DB <sub>30</sub>	0.844 (0.0070)	0.994 (0.0015)	0.765 (0.0070)	0.994 (0.0011)	0.610 (0.0098)	0.993 (0.0007)
DB <sub>60</sub>	0.866 (0.0062)	0.996 (0.0012)	0.794 (0.0059)	0.996 (0.0010)	0.652 (0.0094)	0.996 (0.0006)

reduction of  $I$  was stronger for scenarios with a low number of QTL and high values for  $h^2$  than vice versa (Figure 1, A and B). Across all examined levels of  $l$  and  $h^2$ , the decrease of  $1 - \beta^*$  upon the reduction of  $I$  was slightly more pronounced for the DIA design than for the REF design, whereas  $N$  did not influence this trend.

Across all examined scenarios, the  $1 - \beta^*$  values observed for maize as well as *A. thaliana* on the basis of the QTL detection procedure that takes into account the structure of the RIL populations were between one-third and one-fourth lower than those observed for the QTL detection that ignores this structure (data not shown). For the former QTL detection procedure, the  $1 - \beta^*$  trends observed with respect to the examined mating designs as well  $l$ ,  $h^2$ , and  $N$  were similar to those found for the latter QTL detection procedure.

## DISCUSSION

In contrast to previous joint linkage and LD studies, which focused on mining existing mapping population in pedigrees or heterogeneous stocks (*e.g.*, MEUWISSEN *et al.* 2002), the NAM strategy proposed by Yu *et al.* (2008) aims to create an integrated mapping population specifically designed for a full genome scan for QTL. One idea of this strategy is that with common-parent-specific (CPS) markers genotyped for the parental inbreds and the RILs, the inheritance of chromosome segments nested within two adjacent CPS markers can be inferred through linkage. Genotyping the founders with additional high-density markers enables the projection of genetic information, capturing LD information, from the parental inbreds to the RILs.

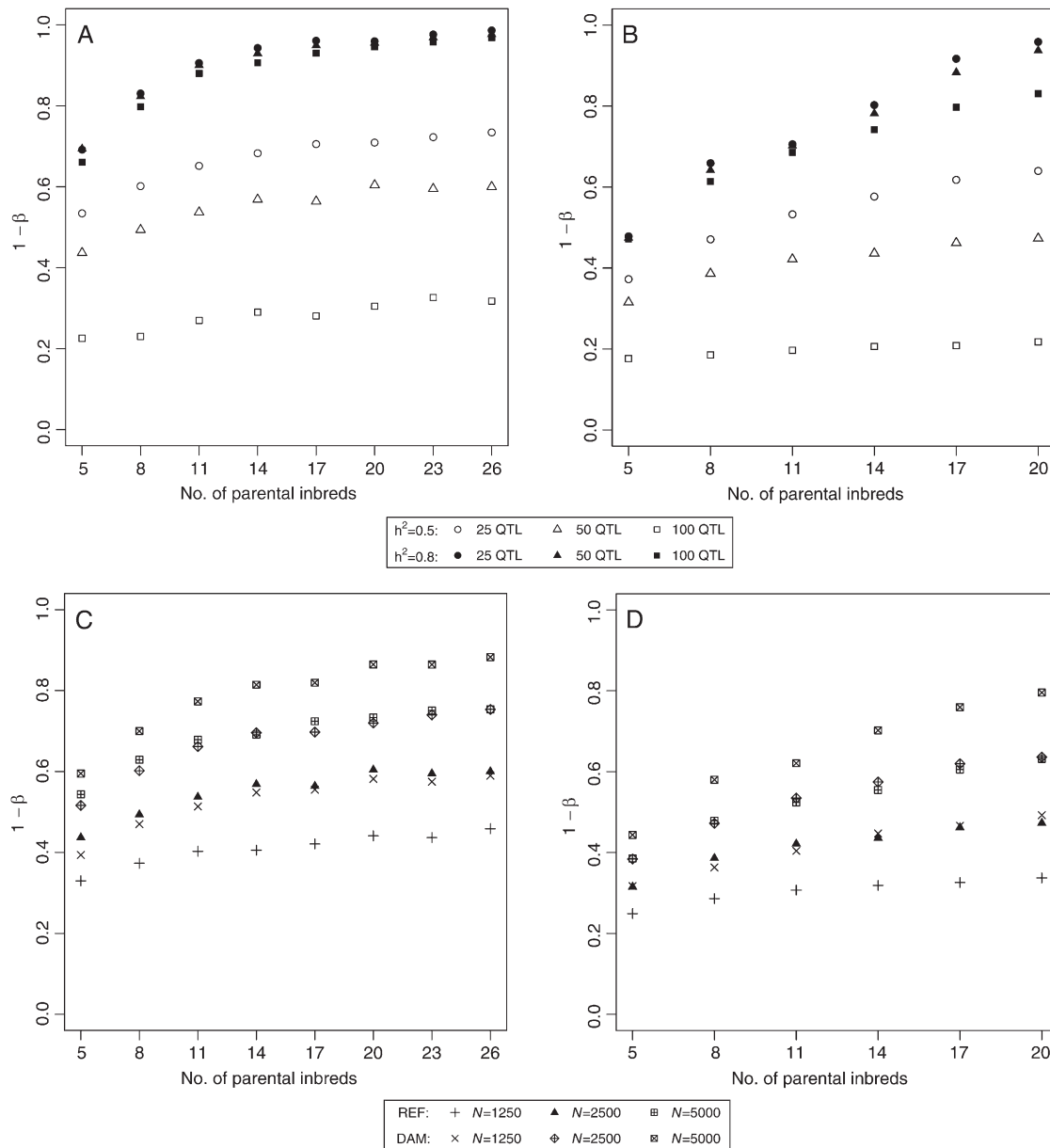


FIGURE 1.—Power to detect QTL ( $\alpha^* = 0.01$ ) of maize (A and C) and *A. thaliana* (B and D) recombinant inbred line (RIL) populations. (A and B)  $N = 2500$  RILs were derived from the reference (REF) design using various numbers of parental inbreds. (C and D) Different numbers of RILs  $N$  were derived from the REF and diallel (DIA) designs using various numbers of parental inbreds and assuming 50 QTL and  $h^2 = 0.5$ .

This approach is expected to allow high-resolution QTL mapping with a relatively low number of markers in the RILs.

However, using other designs than the REF design, the CPS marker strategy is not straightforward to implement. Therefore, in the current study, I assumed that all RILs are genotyped with such a high number of markers that each QTL has a marker that is in complete LD with the QTL. Due to the fast progress of genome sequencing techniques (SHENDURE *et al.* 2004), this is a realistic assumption in the foreseeable future. Therefore, it will be possible to exploit both recent and

ancient recombination in RIL populations derived from other mating designs than the REF design.

**QTL detection procedures for NAM populations:** A NAM population consists of a large number of segregating populations (Table 1). The existence of alleles that are specific for some of the segregating populations can lead to experimentwide LD between the causal gene and some unlinked markers. This, however, has the potential to increase the rate of false positive associations when applying QTL detection procedures that neglect the structure of the NAM population (YU *et al.* 2008). Therefore, I used in addition to such a QTL

detection procedure a procedure that accounts for the structure of the simulated RIL populations by including the mean value of the RILs derived from each cross as a covariate. However, I observed for none of the examined mating designs a considerable difference with respect to the nominal  $\alpha$ -level, which is required to obtain an empirical  $\alpha$ -level of 0.01, between the two examined QTL detection procedures. This finding suggested that the above-mentioned issue of experiment-wide LD between the causal gene and some unlinked markers might be neglected in the current study. Furthermore, because I observed for the QTL detection procedure neglecting the structure of the NAM population considerably higher  $1 - \beta^*$  estimates than for the QTL detection procedure accounting for it, only the results of the former method are discussed below. Nevertheless, further research on the most appropriate QTL detection procedure for NAM populations is required.

**Power to detect QTL under different mating designs in maize:** Across all the examined scenarios of maize, the lowest power  $1 - \beta^*$  was observed for the RILs derived from the REF design (Table 2). This observation is in accordance with results of STICH *et al.* (2007), who compared RIL populations derived from different designs with respect to their power to detect three-way epistatic interactions. These findings might be attributable to the fact that the average frequency of the common parent allele was closer to 1 for RILs derived from the REF design than for all other designs. Crossing schemes that result in RILs with an average allele frequency strongly deviating from 0.5 have a low power to detect QTL because the probability that some QTL alleles are present in only a very low number of RILs is maximized (VERHOEVEN *et al.* 2006).

Despite this disadvantage of the REF design, the project “molecular and functional diversity of the maize genome” applied this crossing scheme to establish a NAM population in maize (YU *et al.* 2008). The main advantage of this crossing scheme is that crossing the 25 diverse inbreds to the inbred B73, which is well adapted to U.S. environmental conditions, facilitates the development as well as the phenotyping of RILs within the United States (YU *et al.* 2008). This issue, however, might be of lower importance for the choice of the most appropriate crossing scheme to establish NAM populations (i) based on another set of parental maize inbreds as well as (ii) for other plant species with a lower genetic diversity than that of the parental inbreds used in the project “molecular and functional diversity of the maize genome.” Therefore, the  $1 - \beta^*$  estimates observed for the other crossing schemes are discussed below.

My results revealed a lower power  $1 - \beta^*$  for the DB designs than for the DIA, FCT, and DRR crossing schemes (Table 2). This observation is in contrast to results of STICH *et al.* (2007), who observed for optimally allocated DB designs a higher power to detect three-way

epistatic interactions than for the DIA crossing scheme. The development of RIL populations from pairs of parental inbreds, which were selected to maximize the pairwise genetic dissimilarity, increases indeed the average probability that QTL are segregating. Nevertheless, such a selection can also lead to the fixation of some QTL, because in contrast to the other designs not all parental inbreds are used for the establishment of the segregating populations. In scenarios with a low power for QTL detection such as that of STICH *et al.* (2007), the fixation of some QTL has only marginal effects and, thus, the increased average probability that QTL are segregating of DB designs can be used. However, in scenarios with a power  $1 - \beta^*$  close to 1, like in the present study, it is indispensable that all QTL are polymorphic. Thus, balanced crossing schemes such as DIA, FCT, or DRR might be superior to DB designs in scenarios with a high power to detect QTL and vice versa.

Across all scenarios of maize, my results revealed a higher power  $1 - \beta^*$  for RILs derived from the crossing schemes DIA, FCT, and DRR than for the SRR, RRR, and IRR mating designs (Table 2). This observation cannot be explained by differences in allele frequencies, as the RILs derived from all designs except REF showed similar allele frequencies. Partly, my finding might be explained by the large number of small populations derived from the former designs (Table 1). This explanation is in contrast to results of VERHOEVEN *et al.* (2006). The different findings can be explained by the different assumptions underlying the simulations. First, VERHOEVEN *et al.* (2006) detected QTL within individual RIL populations whereas in my study QTL were detected across all RIL populations. Second, VERHOEVEN *et al.* (2006) assumed a distinct allele for each parental inbred. In these cases, large numbers of small populations show, due to the increased probability that some QTL alleles have only a very small class size, a lower power to detect QTL than do a small number of large populations. However, the assumptions made by VERHOEVEN *et al.* (2006) ignore the fact that for real data not all QTL segregate in every population (XU 1996). In my study, this fact was considered by using SNP data of parental inbreds as a basis of the simulations. Consequently, the mating designs resulting in a large number of small populations have indeed the above-mentioned disadvantage but this is compensated by the large number of individuals within populations segregating for the QTL.

For maize, the results of my study indicated that the DIA and FCT crossing schemes result in the highest power for QTL detection. However, to establish the number of crosses required for these mating designs might be realistic only for an outcrossing species such as maize. In contrast, the number of crosses necessary for the DRR mating design are considerably lower than that required for the DIA and FCT crossing schemes (Table

1). Nevertheless, for all three mating designs similar  $1 - \beta^*$  estimates were observed. These findings suggested that the DRR crossing scheme might be the most appropriate design to establish NAM populations in autogamous or partial autogamous species such as barley, wheat, or rapeseed.

**Factors influencing the power for QTL detection and the relative performance of crossing schemes to establish NAM populations:** Theoretical considerations suggest that the power for QTL detection  $1 - \beta^*$  but also the relative performance of different crossing schemes to establish NAM populations might be influenced by (i) the plant species examined, (ii) the genetic architecture of the trait under consideration, (iii) the total population size, (iv) the inbreeding procedure, and (v) the number of parental genotypes used.

*Plant species:* Across all crossing schemes, lower  $1 - \beta^*$  estimates were observed for *A. thaliana* NAM populations than for maize NAM populations of similar size (Tables 2 and 3). This finding might be due to the higher frequency of the reference allele in *A. thaliana* compared with the same design of maize. Another explanation might be the four times lower average map distance between the SNP markers in *A. thaliana* compared with that in maize. Thereby, in *A. thaliana*, the recombination between the markers is reduced, which is expected to increase the type I error rate. This decreases the power for QTL detection, however, when fixing the empirical type I error rate as described in MATERIALS AND METHODS.

My results revealed only slight differences between the rankings of the various crossing schemes with respect to  $1 - \beta^*$  for maize and *A. thaliana*. This observation suggested that my conclusions regarding the most appropriate crossing scheme might also be valid for other plant species.

*Genetic architecture of the trait:* A higher-power  $1 - \beta^*$  was observed for traits influenced by a low number of QTL than for traits influenced by a high number of QTL (Tables 2 and 3). Similarly, increasing  $h^2$  from 0.5 to 0.8 resulted for all examined designs and all numbers of QTL in a considerably higher power to detect QTL. These observations are in accordance with quantitative genetic theory and previous studies (*e.g.*, VAN OOIJEN 1992; BEAVIS 1994; FALCONER and MACKAY 1996) and can be explained by the fact that in the former case each QTL explains a higher proportion of the phenotypic variance than in the latter.

The ranking of the various crossing schemes differed slightly among the three QTL scenarios as well as between the two heritability scenarios. However, the observed differences followed no clear trend.

*Total population size:* Across all examined designs, a higher power for QTL detection was observed for populations with a higher number of entries (Tables 2 and 3). This observation is in accordance with results of SCHÖN *et al.* (2004) and can be explained by the fact

that in this case the allele effects are estimated more precisely.

The ranking of the various crossing schemes differed only slightly among the three levels examined for the total number of RILs. Therefore, I expect that my findings are valid for a broad range of total population sizes.

*Inbreeding procedure:* The use of inbred genotypes in QTL mapping experiments has several advantages (BURR *et al.* 1988; LANDER and BOTSTEIN 1989). Due to the short generation time, such individuals are created for *A. thaliana* by repeated self-pollination. Most crop species, however, have considerably longer generation times. Therefore, the creation of fully homozygous genotypes in one step via doubled-haploid (DH) induction (JENSEN 1975; BAJAJ 1977; BORDES *et al.* 1997; WENZEL *et al.* 1977) is an interesting alternative to repeated self-pollination and, thus, was examined in my study (data not shown).

My results revealed a power for QTL detection of DH populations derived from  $F_1$  genotypes that is similar to that observed for RIL populations of identical size. However, I observed across all the examined scenarios a considerably higher power for DH populations that were derived from  $F_2$  genotypes. This observation might be explained by the additional recombinations that occurred before the induction of DHs (*cf.* BERNARDO 2009). Nevertheless, the use of RILs in a NAM context is justified if the ultimate objective of the experiment is to clone the QTL. In this case, the use of heterogenous inbred families (TUINSTRAN *et al.* 1997) derived from RILs proved to be a powerful tool (*cf.* FRIDMAN *et al.* 2000).

*Number of parental genotypes:* Across all examined scenarios of maize and Arabidopsis, I observed a higher power  $1 - \beta^*$  for NAM populations that were established using a high number of parental inbreds than a low number (Figure 1). This finding can be explained by the fact that a higher number of parental inbreds increases the number of polymorphic QTL. Therefore, my findings indicate to use a high number of parental inbreds for the creation of NAM populations.

For *A. thaliana*, a linear increase of  $1 - \beta^*$  was observed with an increase of the parental inbreds from 5 to 20. In contrast, for maize, the increase of the power for QTL detection was high for an increase of the parental inbreds from 5 to 14 but was considerably lower for an increase from 14 to 26. These observations might be due to the higher number of rare alleles in *A. thaliana* compared with maize.

For the REF as well as the DIA crossing scheme, I observed a similar increase of  $1 - \beta^*$  with an increasing number of parental inbreds. This finding suggested that the ranking of the crossing schemes with respect to  $1 - \beta^*$  is not or only marginally influenced by the number of parental inbreds used.

**Conclusions:** My finding of considerable differences in  $1 - \beta^*$  estimates between NAM populations of the



same size but created on the basis of different crossing schemes illustrated the potential to improve the power for QTL detection without increasing the total resources necessary for a QTL mapping experiment. For maize as well as *A. thaliana*, I observed the highest power for QTL detection for the DIA and FCT crossing schemes. However, for species with a high genetic diversity, such as maize, it will be difficult to generate high-quality phenotypic values in field trials with RIL populations derived from crosses between diverse material. Furthermore, these designs require creation of a high number of crosses, which might be difficult in autogamous or partial autogamous species such as barley, wheat, or rapeseed. For these species, the DRR crossing scheme might be a promising alternative, because it requires the creation of only a relatively low number of crosses, while almost the same  $1 - \beta^*$  estimates were observed as the DIA and FCT designs. Finally, my results clearly indicated that it is advantageous to create NAM populations from a large number of parental inbreds.

I thank Edward S. Buckler and Detlef Weigel for providing the genotypic information for the maize and *A. thaliana* inbreds, respectively. Furthermore, I thank Maarten Koornneef for critical reading of the manuscript. I thank the Plant Computational Biology group of the Max Planck Institute for Plant Breeding Research for use of their computer cluster as well as the associate editor and two anonymous reviewers for their valuable suggestions. Funding for this work was provided by the Max Planck Society.

#### LITERATURE CITED

- ALONSO, J. M., A. N. STEPANOVA, T. J. LEISSE, C. J. KIM, H. CHEN *et al.*, 2003 Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* **301**: 653–657.
- BAJAJ, Y. P. S., 1977 In vitro induction of haploids in wheat (*Triticum aestivum* L.). *Crop Improv.* **4**: 54–64.
- BEAVIS, W. D., 1994 The power and deceit of QTL experiments: lessons from comparative QTL studies, pp. 250–266 in *49th Annual Corn and Sorghum Industry Research Conference*. American Seed Trade Association, Washington, DC.
- BERNARDO, R., 2009 Should maize doubled haploids be induced among F<sub>1</sub> or F<sub>2</sub> plants? *Theor. Appl. Genet.* **119**: 255–262.
- BORDES, J., R. D. DE VAULX, A. LAPIERRE and M. POLLACSEK, 1997 Haplodiploidization of maize (*Zea mays* L.) through induced gynogenesis assisted by glossy markers and its use in breeding. *Agronomie* **17**: 291–297.
- BUCKLER, E. S., J. M. HOLLAND, P. J. BRADBURY, C. B. ACHARYA, P. J. BROWN *et al.*, 2009 The genetic architecture of maize flowering time. *Science* **325**: 714–718.
- BURR, B., F. A. BURR, K. H. THOMPSON, M. C. ALBERTSEN and C. W. STUBER, 1988 Gene mapping with recombinant inbreds in maize. *Genetics* **118**: 519–526.
- CLARK, R. M., G. SCHWEIKERT, C. TOOMAJIAN, S. OSSOWSKI, G. ZELLER *et al.*, 2007 Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. *Science* **317**: 338–342.
- COMSTOCK, R. E., and H. F. ROBINSON, 1948 The components of genetic variance in populations of biparental progenies and their use in estimating the average degree of dominance. *Biometrics* **4**: 254–266.
- FALCONER, D. S., and T. F. C. MACKAY, 1996 *Introduction to Quantitative Genetics*, Ed. 4. Longman Group, London.
- FLINT-GARCIA, S. A., A. THUILLET, J. YU, G. PRESSOIR, S. M. ROMERO *et al.*, 2005 Maize association population: a high resolution platform for QTL dissection. *Plant J.* **44**: 1054–1064.
- FRIDMAN, E., T. PLEBAN and D. ZAMIR, 2000 A recombination hotspot delimits a wild-species quantitative trait locus for tomato sugar content to 484 bp within an invertase gene. *Proc. Natl. Acad. Sci. USA* **97**: 4718–4723.
- GRIFFING, B., 1956 Concept of general and specific combining ability in relation to diallel crossing systems. *Aust. J. Biol. Sci.* **9**: 463–493.
- HOLLAND, J. B., 2007 Genetic architecture of complex traits in plants. *Curr. Opin. Plant Biol.* **10**: 156–161.
- JENSEN, C. J., 1975 Barley monolpoids and double monolpoids: techniques and experience, pp. 316–345 in *Barley Genetics IV*, edited by H. GAUL. Thieme, München, Germany.
- KIM, S., V. PLAGNOL, T. T. HU, C. TOOMAJIAN, R. M. CLARK *et al.*, 2007 Recombination and linkage disequilibrium in *Arabidopsis thaliana*. *Nat. Genet.* **39**: 1151–1155.
- KRAAKMAN, A. T. W., R. E. NIKS, P. M. M. VAN DEN BERG, P. STAM and F. A. EEUWIJK, 2004 Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivar. *Genetics* **168**: 435–446.
- LANDE, R., and R. THOMPSON, 1990 Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* **124**: 743–756.
- LANDER, E. S., and D. BOTSTEIN, 1989 Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**: 185–199.
- LIU, K., M. GOODMAN, S. MUSE, J. S. SMITH, E. BUCKLER *et al.*, 2003 Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites. *Genetics* **165**: 2117–2128.
- MAURER, H. P., A. E. MELCHINGER and M. FRISCH, 2008 Population genetical simulation and data analysis with Plabsoft. *Euphytica* **161**: 133–139.
- McMULLEN, M. M., S. KRESOVICH, H. SANCHEZ VILLEDA, P. BRADBURY, H. LI *et al.*, 2009 Genetic properties of the maize nested association mapping population. *Science* **325**: 737–740.
- MEUWISSEN, T. H., A. KARLSEN, S. LIEN, I. OLSAKER and M. E. GODDARD, 2002 Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping. *Genetics* **161**: 373–379.
- NEI, M., and W. H. LI, 1979 Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. USA* **76**: 5269–5273.
- NORDBORG, M., T. T. HU, Y. ISHINO, J. JHAVERI, C. TOOMAJIAN *et al.*, 2005 The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biol.* **3**: e196.
- R DEVELOPMENT CORE TEAM, 2004 *R: A Language and Environment for Statistical Computing*. Vienna.
- SCHÖN, C. C., H. F. UTZ, S. GROH, B. TRUBERG, S. OPENSHAW *et al.*, 2004 Quantitative trait locus mapping based on resampling in a vast maize testcross experiment and its relevance to quantitative genetics for complex traits. *Genetics* **167**: 485–498.
- SCHWARZ, G., 1978 Estimating the dimension of a model. *Ann. Stat.* **6**: 461–464.
- SHENDURE, J., R. D. MITRA, C. VARMA and G. M. CHURCH, 2004 Advanced sequencing technologies: methods and goals. *Nat. Rev. Genet.* **5**: 335–344.
- SINGER, T., Y. FAN, H. S. CHANG, T. ZHU, S. P. HAZEN *et al.*, 2006 A high-resolution map of *Arabidopsis* recombinant inbred lines by whole-genome exon array hybridization. *PLoS Genet.* **2**: e144.
- STICH, B., J. YU, A. E. MELCHINGER, H.-P. PIEPHO, H. F. UTZ *et al.*, 2007 Power to detect higher-order epistatic interactions in a metabolic pathway using a new mapping strategy. *Genetics* **176**: 563–570.
- STICH, B., H. F. UTZ, H.-P. PIEPHO, H. P. MAURER and A. E. MELCHINGER, 2009 Optimum allocation of resources for QTL detection using a nested association mapping strategy in maize. *Theor. Appl. Genet.* (in press).
- TUINSTRAN, M. R., G. EJETA and P. B. GOLDSBROUGH, 1997 Heterogeneous inbred family (HIF) analysis: a method for developing near-isogenic lines that differ at quantitative trait loci. *Theor. Appl. Genet.* **95**: 1005–1011.
- UTZ, H. F., and A. E. MELCHINGER, 1996 PLABQTL: a program for composite interval mapping of QTL. *J. Quant. Trait Loci* **2**: 1–5.
- VAN OOIJEN, J. W., 1992 Accuracy of mapping quantitative trait loci in autogamous species. *Theor. Appl. Genet.* **84**: 803–811.
- VERHOEVEN, K. J. F., J.-L. JANNINK and L. M. MCINTYRE, 2006 Using mating designs to uncover QTL and the genetic architecture of complex traits. *Heredity* **96**: 139–149.

- WENZEL, G., F. HOFFMAN and E. THOMAS, 1977 Anther culture as a breeding tool in rape. I. Ploidy level and phenotype of androgenetic plants. *Z. Pflanzenzücht.* **78**: 149–155.
- WILLER, C. J., S. SANNA, A. U. JACKSON, A. SCUTERI, L. L. BONNYCASTLE *et al.*, 2008 Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat. Genet.* **40**: 161–169.
- WILSON, L. M., S. R. WHITT, A. M. IBÁÑEZ, T. R. ROCHEFORD, M. M. GOODMAN *et al.*, 2004 Dissection of maize kernel composition and starch production by candidate gene association. *Plant Cell* **16**: 2719–2733.
- XU, S., 1996 Mapping quantitative trait loci using four-way crosses. *Genet. Res.* **68**: 175–181.
- YANO, M., 2001 Genetic and molecular dissection of naturally occurring variation. *Curr. Opin. Plant Biol.* **4**: 130–135.
- YU, J., and E. BUCKLER, 2006 Genetic association mapping and genome organization of maize. *Curr. Opin. Biotech.* **17**: 155–160.
- YU, J., J. B. HOLLAND, M. D. McMULLEN and E. S. BUCKLER, 2008 Power analysis of an integrated mapping strategy: nested association mapping. *Genetics* **138**: 539–551.

Communicating editor: H. ZHAO

# GENETICS

Supporting Information

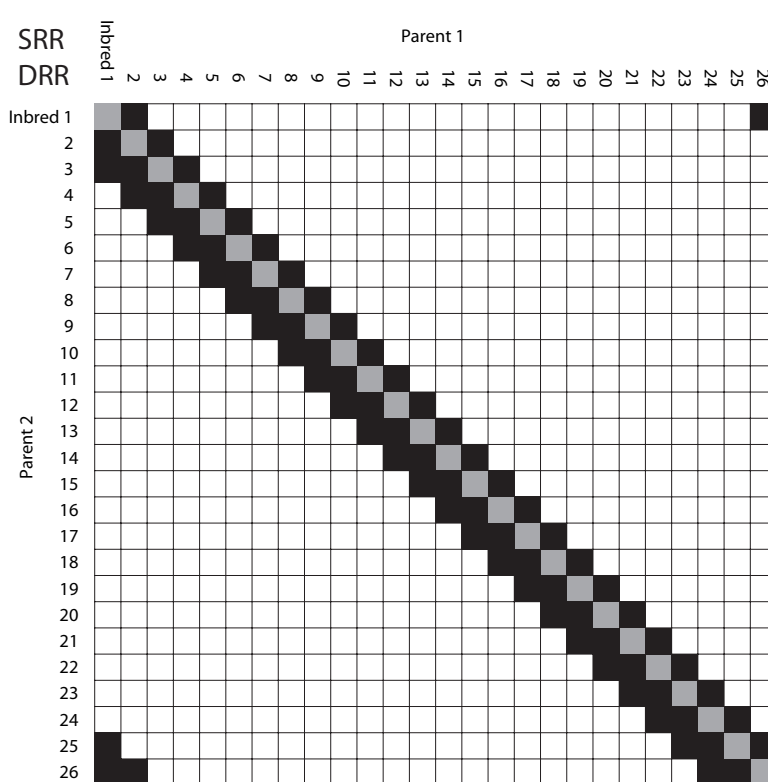
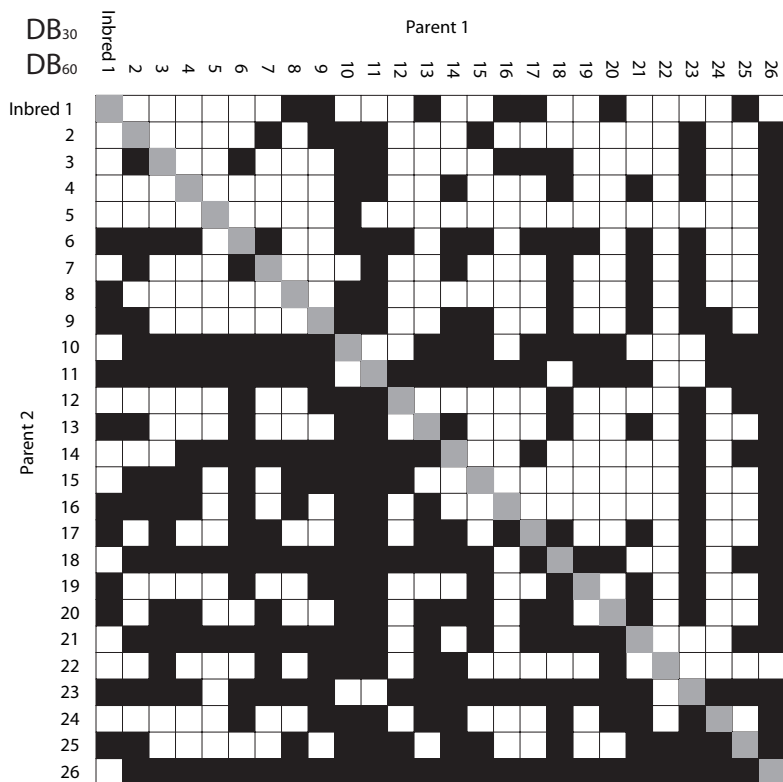
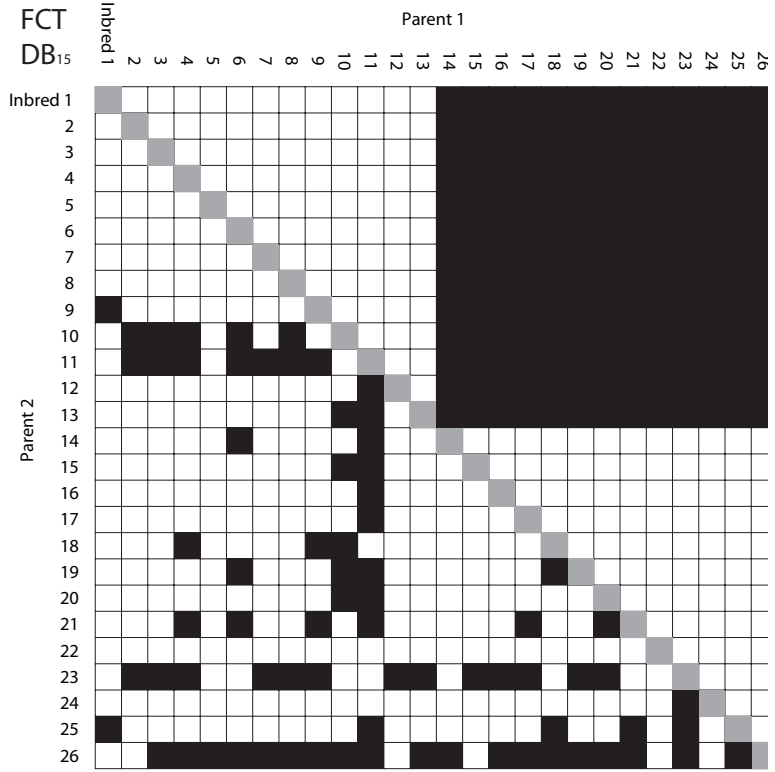
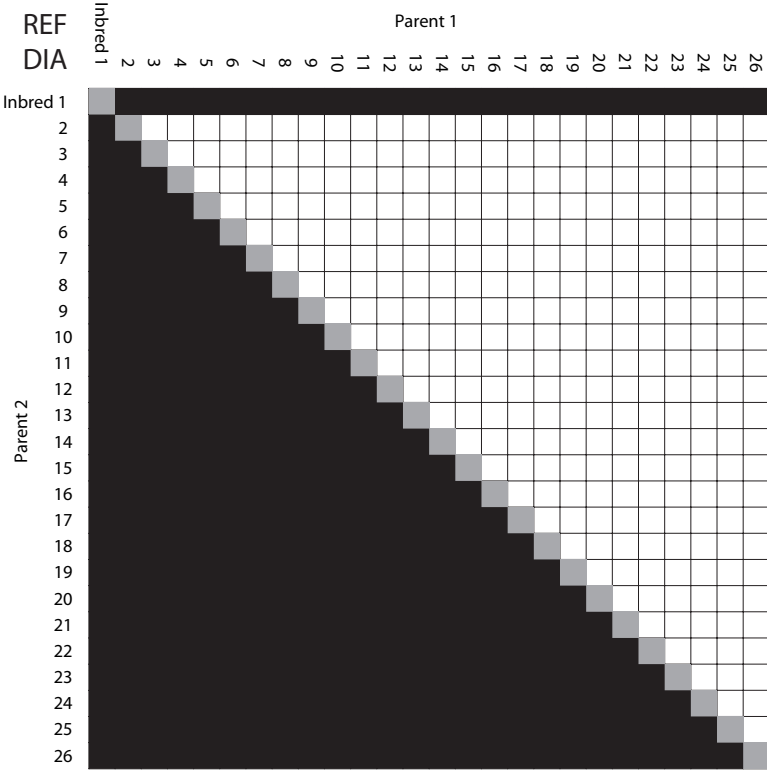
<http://www.genetics.org/cgi/content/full/genetics.109.108449/DC1>

**Comparison of Mating Designs for Establishing Nested Association  
Mapping Populations in Maize and *Arabidopsis thaliana***

**Benjamin Stich**

Copyright © 2009 by the Genetics Society of America

DOI: 10.1534/genetics.109.108449





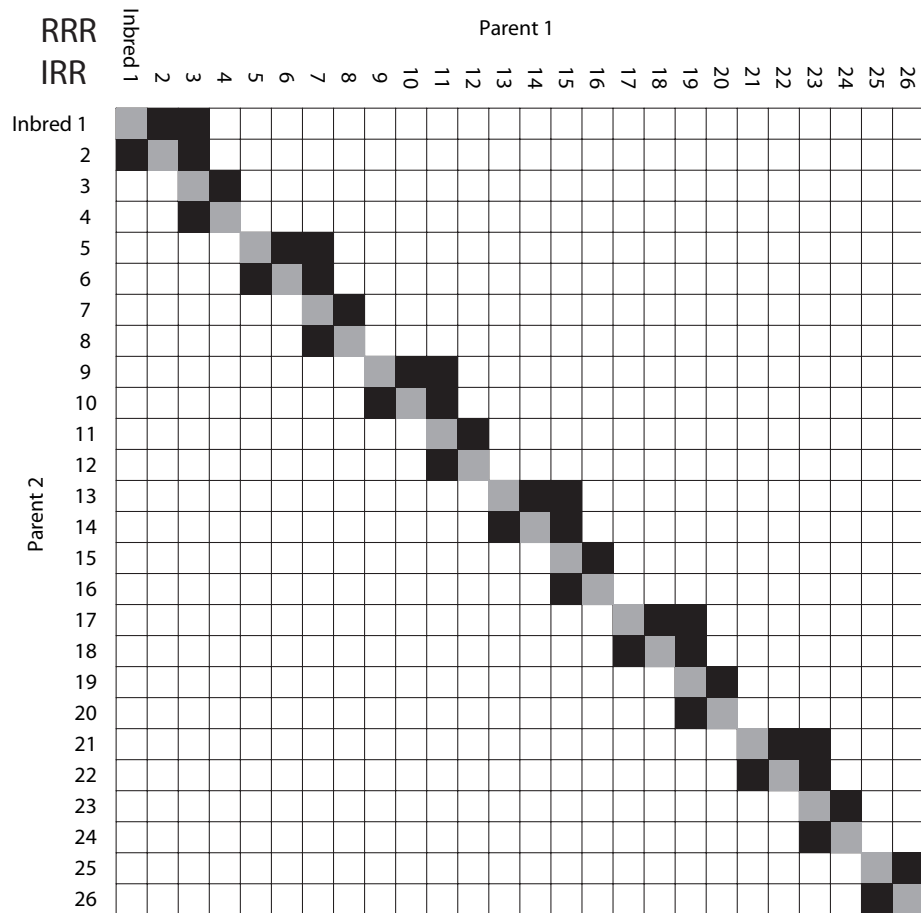


FIGURE S1.—Mating designs examined in my study: Reference (REF), diallel (DIA), factorial (FCT), distance-based (DB), single round-robin (SRR), double round-robin (DRR), reduced round-robin (RRR), and independent round-robin (IRR) (DB) design. Black squares indicate the crosses from which a segregating population was derived.

**File S1****List of single nucleotide polymorphisms of *Arabidopsis thaliana* used in my study**

PERL0000013 PERL0000340 PERL0001224 PERL0002506 PERL0003960 PERL0005218 PERL0006265 PERL0007033  
PERL0008141 PERL0009147 PERL0010317 PERL0011472 PERL0012452 PERL0013496 PERL0014567 PERL0015748  
PERL0016790 PERL0018078 PERL0019120 PERL0020202 PERL0022118 PERL0023775 PERL0024721 PERL0025663  
PERL0027067 PERL0028498 PERL0029809 PERL0031513 PERL0032709 PERL0033700 PERL0034890 PERL0036066  
PERL0037438 PERL0039548 PERL0041187 PERL0042507 PERL0044551 PERL0045914 PERL0046935 PERL0048342  
PERL0049634 PERL0050357 PERL0052629 PERL0053843 PERL0055010 PERL0056230 PERL0057451 PERL0059209  
PERL0060871 PERL0063029 PERL0065573 PERL0067547 PERL0069358 PERL0070555 PERL0071709 PERL0073587  
PERL0074875 PERL0076092 PERL0077395 PERL0078788 PERL0079737 PERL0081310 PERL0082698 PERL0083757  
PERL0085910 PERL0088184 PERL0090833 PERL0094252 PERL0096331 PERL0097970 PERL0100124 PERL0102247  
PERL0105803 PERL0107833 PERL0109316 PERL0111572 PERL0114103 PERL0116421 PERL0119522 PERL0122067  
PERL0123763 PERL0125466 PERL0130015 PERL0133125 PERL0135933 PERL0137585 PERL0139671 PERL0141831  
PERL0143503 PERL0146421 PERL0148888 PERL0151975 PERL0154049 PERL0155355 PERL0156961 PERL0158464  
PERL0159792 PERL0161172 PERL0162350 PERL0163633 PERL0164740 PERL0166042 PERL0167475 PERL0168733  
PERL0170123 PERL0171888 PERL0172584 PERL0173734 PERL0174632 PERL0176100 PERL0177766 PERL0179573  
PERL0181760 PERL0184513 PERL0187441 PERL0189698 PERL0191305 PERL0193482 PERL0195826 PERL0197857  
PERL0200167 PERL0202526 PERL0204631 PERL0206238 PERL0208341 PERL0211113 PERL0213906 PERL0216220  
PERL0218976 PERL0221150 PERL0223348 PERL0225282 PERL0227866 PERL0229748 PERL0231261 PERL0232543  
PERL0233966 PERL0235333 PERL0236918 PERL0239115 PERL0240945 PERL0242213 PERL0243785 PERL0245397  
PERL0246884 PERL0248055 PERL0250470 PERL0251769 PERL0252788 PERL0253988 PERL0254863 PERL0255756  
PERL0256911 PERL0257779 PERL0258670 PERL0259468 PERL0260691 PERL0261767 PERL0262930 PERL0264117  
PERL0265609 PERL0266517 PERL0269216 PERL0269538 PERL0270429 PERL0271944 PERL0273072 PERL0275837  
PERL0277047 PERL0278893 PERL0280757 PERL0282401 PERL0284382 PERL0286458 PERL0289601 PERL0291899  
PERL0294182 PERL0296370 PERL0298369 PERL0300941 PERL0304513 PERL0306043 PERL0310008 PERL0312295  
PERL0315729 PERL0318803 PERL0321692 PERL0324326 PERL0326798 PERL0328390 PERL0329781 PERL0331429  
PERL0333297 PERL0335113 PERL0336631 PERL0338149 PERL0340031 PERL0341599 PERL0343092 PERL0344515  
PERL0345841 PERL0347159 PERL0348750 PERL0349363 PERL0350544 PERL0351768 PERL0353003 PERL0353925  
PERL0354826 PERL0355660 PERL0356740 PERL0357867 PERL0359109 PERL0360581 PERL0361772 PERL0363568  
PERL0364638 PERL0366048 PERL0367540 PERL0368999 PERL0370297 PERL0371122 PERL0371785 PERL0372642  
PERL0373961 PERL0375179 PERL0376043 PERL0376924 PERL0377783 PERL0378501 PERL0379324 PERL0380339  
PERL0381278 PERL0382622 PERL0383491 PERL0384341 PERL0385079 PERL0386236 PERL0386965 PERL0387834  
PERL0388733 PERL0389793 PERL0390568 PERL0391557 PERL0392632 PERL0393545 PERL0394521 PERL0395362  
PERL0396655 PERL0397784 PERL0398659 PERL0399140 PERL0400986 PERL0402119 PERL0403749 PERL0404908  
PERL0406622 PERL0408132 PERL0409235 PERL0410016 PERL0410863 PERL0411690 PERL0412881 PERL0413540  
PERL0414615 PERL0416059 PERL0417132 PERL0418329 PERL0419694 PERL0421044 PERL0422442 PERL0424186  
PERL0425503 PERL0426599 PERL0427848 PERL0429204 PERL0430650 PERL0431516 PERL0432753 PERL0433948  
PERL0434714 PERL0436220 PERL0437379 PERL0438725 PERL0439963 PERL0441818 PERL0443585 PERL0444826  
PERL0446053 PERL0447436 PERL0448880 PERL0450235 PERL0451521 PERL0453106 PERL0454188 PERL0455768  
PERL0457232 PERL0458941 PERL0460645 PERL0462072 PERL0463684 PERL0465842 PERL0467200 PERL0468765  
PERL0469971 PERL0471678 PERL0473070 PERL0474605 PERL0475865 PERL0477544 PERL0479058 PERL0481440  
PERL0483217 PERL0484713 PERL0486265 PERL0488616 PERL0492371 PERL0495569 PERL0497374 PERL0500354

PERL0501948 PERL0503458 PERL0506509 PERL0508983 PERL0512311 PERL0514597 PERL0517612 PERL0520681  
PERL0523932 PERL0527086 PERL0531451 PERL0535113 PERL0537338 PERL0540675 PERL0543912 PERL0554217  
PERL0557823 PERL0563369 PERL0568569 PERL0572371 PERL0575903 PERL0579784 PERL0584131 PERL0587701  
PERL0591196 PERL0593667 PERL0597323 PERL0599401 PERL0601882 PERL0604277 PERL0607497 PERL0610536  
PERL0613142 PERL0614639 PERL0616648 PERL0618491 PERL0620090 PERL0621445 PERL0623512 PERL0624959  
PERL0626387 PERL0627507 PERL0628499 PERL0629796 PERL0631358 PERL0633047 PERL0634378 PERL0635631  
PERL0637123 PERL0638496 PERL0640154 PERL0641673 PERL0642918 PERL0644739 PERL0646056 PERL0647073  
PERL0648238 PERL0649641 PERL0650565 PERL0652328 PERL0653447 PERL0654431 PERL0655780 PERL0657618  
PERL0657883 PERL0658501 PERL0659964 PERL0661209 PERL0662505 PERL0664456 PERL0666234 PERL0667720  
PERL0669450 PERL0671106 PERL0672283 PERL0674032 PERL0677372 PERL0679611 PERL0681173 PERL0683560  
PERL0685627 PERL0688352 PERL0689909 PERL0691560 PERL0693654 PERL0696138 PERL0698263 PERL0699352  
PERL0702542 PERL0707647 PERL0710102 PERL0712766 PERL0714609 PERL0716056 PERL0717953 PERL0719295  
PERL0721467 PERL0723918 PERL0725628 PERL0727455 PERL0730155 PERL0732558 PERL0734616 PERL0737066  
PERL0739772 PERL0742322 PERL0744124 PERL0746219 PERL0748499 PERL0750962 PERL0753682 PERL0755286  
PERL0757562 PERL0759638 PERL0761748 PERL0763309 PERL0765918 PERL0768391 PERL0770007 PERL0771584  
PERL0775604 PERL0777437 PERL0779829 PERL0781559 PERL0783149 PERL0784606 PERL0786358 PERL0789281  
PERL0792435 PERL0794740 PERL0796953 PERL0798668 PERL0800387 PERL0801382 PERL0802869 PERL0804611  
PERL0806002 PERL0807763 PERL0809207 PERL0810676 PERL0811712 PERL0813020 PERL0814266 PERL0815456  
PERL0816363 PERL0817514 PERL0818605 PERL0819694 PERL0821167 PERL0822200 PERL0823374 PERL0824061  
PERL0825425 PERL0826219 PERL0827169 PERL0827842 PERL0828503 PERL0829280 PERL0830132 PERL0830908  
PERL0831942 PERL0832688 PERL0833517 PERL0834373 PERL0835446 PERL0836576 PERL0837863 PERL0838930  
PERL0839985 PERL0840818 PERL0841626 PERL0842436 PERL0843763 PERL0844845 PERL0845784 PERL0847394  
PERL0848410 PERL0849149 PERL0850106 PERL0851122 PERL0852079 PERL0853156 PERL0854788 PERL0856095  
PERL0858243 PERL0860031 PERL0862324 PERL0863846 PERL0865191 PERL0866389 PERL0867520 PERL0868541  
PERL0869722 PERL0870562 PERL0872017 PERL0873353 PERL0874630 PERL0875579 PERL0876648 PERL0877881  
PERL0879265 PERL0880626 PERL0882028 PERL0883699 PERL0885135 PERL0886198 PERL0887053 PERL0888022  
PERL0889035 PERL0890064 PERL0891081 PERL0892379 PERL0893302 PERL0894873 PERL0896048 PERL0897523  
PERL0899280 PERL0901330 PERL0902684 PERL0904032 PERL0906157 PERL0907385 PERL0909016 PERL0911784  
PERL0913224 PERL0914729 PERL0916480 PERL0917611 PERL0919240 PERL0920843 PERL0922929 PERL0924788  
PERL0927219 PERL0929091 PERL0931154 PERL0932984 PERL0935524 PERL0937408 PERL0939627 PERL0942226  
PERL0944163 PERL0946952 PERL0949754 PERL0951938 PERL0953776 PERL0956001 PERL0958223 PERL0959841  
PERL0962428 PERL0964811 PERL0966759 PERL0969302 PERL0972285 PERL0975247 PERL0976259 PERL0977961  
PERL0980027 PERL0982469 PERL0984750 PERL0987030 PERL0990723 PERL0993040 PERL0995768 PERL0998942  
PERL1002262 PERL1005229 PERL1007526 PERL1009660 PERL1011755 PERL1013450 PERL1015217 PERL1017011  
PERL1018671 PERL1021975 PERL1024258 PERL1026723 PERL1030098 PERL1033642 PERL1036502 PERL1038822  
PERL1041455 PERL1044264 PERL1047902 PERL1052002 PERL1055017 PERL1057554 PERL1059469 PERL1060579  
PERL1062296 PERL1064297 PERL1066329 PERL1068612 PERL1069943 PERL1071295 PERL1072878 PERL1074739  
PERL1076091 PERL1077408 PERL1079427 PERL1081038 PERL1082734 PERL1084419 PERL1085792 PERL1086844  
PERL1088414 PERL1090200 PERL1091726 PERL1092728 PERL1094438 PERL1095830 PERL1096737 PERL1097715  
PERL1098819 PERL1100053 PERL1101054 PERL1102096 PERL1103906 PERL1105953 PERL1107420 PERL1108372  
PERL1109218 PERL1110418 PERL1111171 PERL1112247 PERL1113488 PERL1114791 PERL1115964 PERL1117197  
PERL1118352 PERL1119628 PERL1120917 PERL1122556 PERL1124181

TABLE S1

**Nominal  $\alpha$  level required to obtain an empirical type I error rate ( $\alpha^*$ ) of 0.01 with  $N$  recombinant inbred lines derived from different mating designs: Reference (REF), diallel (DIA), factorial (FCT), single round-robin (SRR), double round-robin (DRR), reduced round-robin (RRR), independent round-robin (IRR), and distance-based (DB) designs.**

Mating design	Maize						<i>Arabidopsis thaliana</i>					
	25 QTL		50 QTL		100 QTL		25 QTL		50 QTL		100 QTL	
	$h^2 = 0.5$	$h^2 = 0.8$	$h^2 = 0.5$	$h^2 = 0.8$	$h^2 = 0.5$	$h^2 = 0.8$	$h^2 = 0.5$	$h^2 = 0.8$	$h^2 = 0.5$	$h^2 = 0.8$	$h^2 = 0.5$	$h^2 = 0.8$
<i>N = 1250</i>												
REF	0.0091	0.0133	0.0045	0.0111	0.0002	0.0037	0.0061	0.0076	0.0022	0.0043	0.0006	0.0004
DIA	0.0101	0.0128	0.0037	0.0115	0.0004	0.0056	0.0080	0.0110	0.0025	0.0096	0.0003	0.0045
FCT	0.0098	0.0134	0.0041	0.0097	0.0004	0.0056	0.0081	0.0117	0.0026	0.0083	0.0002	0.0038
SRR	0.0086	0.0114	0.0043	0.0101	0.0003	0.0059	0.0069	0.0108	0.0018	0.0089	0.0001	0.0028
DRR	0.0097	0.0118	0.0043	0.0109	0.0003	0.0061	0.0068	0.0130	0.0022	0.0084	0.0001	0.0037
RRR	0.0085	0.0138	0.0026	0.0095	0.0002	0.0057	0.0072	0.0108	0.0015	0.0069	0.0001	0.0025
IRR	0.0098	0.0133	0.0034	0.0088	0.0003	0.0063	0.0064	0.0100	0.0008	0.0065	0.0001	0.0013
DB <sub>15</sub>	0.0102	0.0146	0.0031	0.0112	0.0004	0.0060	0.0087	0.0148	0.0028	0.0097	0.0003	0.0042
DB <sub>30</sub>	0.0119	0.0153	0.0039	0.0116	0.0003	0.0066	0.0077	0.0131	0.0027	0.0091	0.0001	0.0047
DB <sub>60</sub>	0.0095	0.0136	0.0041	0.0097	0.0003	0.0074	0.0084	0.0129	0.0027	0.0095	0.0004	0.0052
<i>N = 2500</i>												
REF	0.0098	0.0141	0.0046	0.0117	0.0002	0.0100	0.0067	0.0093	0.0023	0.0072	0.0002	0.0015
DIA	0.0113	0.0145	0.0069	0.0118	0.0010	0.0090	0.0098	0.0137	0.0035	0.0114	0.0002	0.0084
FCT	0.0108	0.0137	0.0060	0.0120	0.0010	0.0089	0.0089	0.0132	0.0038	0.0120	0.0003	0.0090
SRR	0.0111	0.0140	0.0065	0.0108	0.0008	0.0084	0.0092	0.0123	0.0026	0.0110	0.0001	0.0077
DRR	0.0111	0.0141	0.0067	0.0112	0.0011	0.0080	0.0087	0.0137	0.0033	0.0121	0.0002	0.0079
RRR	0.0122	0.0127	0.0046	0.0114	0.0007	0.0075	0.0090	0.0124	0.0034	0.0098	0.0001	0.0072
IRR	0.0096	0.0129	0.0057	0.0114	0.0006	0.0086	0.0067	0.0118	0.0017	0.0102	0.0001	0.0056
DB <sub>15</sub>	0.0109	0.0148	0.0051	0.0132	0.0003	0.0105	0.0101	0.0155	0.0050	0.0135	0.0004	0.0083
DB <sub>30</sub>	0.0114	0.0162	0.0053	0.0128	0.0012	0.0106	0.0100	0.0132	0.0045	0.0122	0.0004	0.0090
DB <sub>60</sub>	0.0121	0.0145	0.0066	0.0117	0.0009	0.0084	0.0098	0.0142	0.0043	0.0133	0.0003	0.0092
<i>N = 5000</i>												
REF	0.0107	0.0145	0.0070	0.0126	0.0008	0.0117	0.0070	0.0106	0.0036	0.0085	0.0003	0.0048
DIA	0.0125	0.0147	0.0086	0.0129	0.0037	0.0096	0.0092	0.0131	0.0053	0.0116	0.0008	0.0097
FCT	0.0111	0.0132	0.0072	0.0116	0.0038	0.0093	0.0099	0.0134	0.0055	0.0121	0.0009	0.0102
SRR	0.0129	0.0132	0.0076	0.0109	0.0032	0.0089	0.0092	0.0122	0.0031	0.0108	0.0007	0.0090
DRR	0.0133	0.0143	0.0080	0.0119	0.0033	0.0096	0.0091	0.0129	0.0048	0.0104	0.0006	0.0103
RRR	0.0102	0.0131	0.0070	0.0117	0.0020	0.0092	0.0094	0.0124	0.0039	0.0110	0.0004	0.0098
IRR	0.0126	0.0130	0.0069	0.0104	0.0014	0.0101	0.0092	0.0114	0.0034	0.0096	0.0001	0.0070
DB <sub>15</sub>	0.0118	0.0138	0.0069	0.0128	0.0023	0.0109	0.0109	0.0140	0.0070	0.0126	0.0010	0.0107
DB <sub>30</sub>	0.0108	0.0153	0.0076	0.0129	0.0024	0.0110	0.0110	0.0141	0.0068	0.0133	0.0011	0.0113
DB <sub>60</sub>	0.0122	0.0128	0.0091	0.0119	0.0039	0.0099	0.0111	0.0139	0.0067	0.0121	0.0012	0.0112



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.