

Using mating designs to uncover QTL and the genetic architecture of complex traits

KJF Verhoeven^{1,3}, J-L Jannink^{2,3} and LM McIntyre^{1,3}

¹Computational Genomics and Department of Agronomy, Purdue University, Lilly Hall of Life Sciences, 915 W State Street, West Lafayette, IN 47907-2054, USA; ²Department of Agronomy, Iowa State University, Ames, IA 50011-1010, USA

Analysis of quantitative trait loci (QTL) affecting complex traits is often pursued in single-cross experiments. For most purposes, including breeding, some assessment is desired of the generalizability of the QTL findings and of the overall genetic architecture of the trait. Single-cross experiments provide a poor basis for these purposes, as comparison across experiments is hampered by segregation of different allelic combinations among different parents and by context-dependent effects of QTL. To overcome this problem, we combined the benefits of QTL analysis (to identify genomic regions affecting trait variation) and classic diallel analysis (to obtain insight into the general inheritance of the trait) by analyzing multiple mapping families that are connected via shared parents. We first provide a theoretical derivation of main (general combining ability (GCA)) and interaction (specific combining ability (SCA)) effects on F_2 family means relative to variance components in a randomly mating

reference population. Then, using computer simulations to generate F_2 families derived from 10 inbred parents in different partial-diallel designs, we show that QTL can be detected and that the residual among-family variance can be analyzed. Standard diallel analysis methods are applied in order to reveal the presence and mode of action (in terms of GCA and SCA) of undetected polygenes. Given a fixed experiment size (total number of individuals), we demonstrate that QTL detection and estimation of the genetic architecture of polygenic effects are competing goals, which should be explicitly accounted for in the experimental design. Our approach provides a general strategy for exploring the genetic architecture, as well as the QTL underlying variation in quantitative traits.

Heredity (2006) **96**, 139–149. doi:10.1038/sj.hdy.6800763; published online 23 November 2005

Keywords: diallel; QTL; simulations; power; GCA; SCA

Introduction

Analysis of quantitative trait loci (QTL) has progressed substantially from the early days of single marker analysis. Methodology now ranges from regression-based (multiple) interval mapping to likelihood-based full Bayesian approaches (eg Haley and Knott, 1992; Jansen, 1993; Jansen and Stam, 1994; Zeng, 1994; Satagopan *et al*, 1996; Kao *et al*, 1999; Carlborg *et al*, 2000; Corander and Sillanpää, 2002; Jansen *et al*, 2003). In addition to normally distributed traits, approaches have been developed specifically for binary traits (Yi and Xu, 2000; McIntyre *et al*, 2001; Yi and Xu, 2002b; Coffman *et al*, 2005) and for ordinal and categorical traits (Rao and Li, 2000; Yi *et al*, 2004). Upon detecting QTL in a single experiment, a key issue of interest is the generalizability of the findings. Many authors have compared results for different mapping populations based upon relative QTL position (Welz and Geiger, 2000; Kolb *et al*, 2001; Kamoshita *et al*, 2002; Simko, 2002; Clancy *et al*, 2003; Flint-Garcia *et al*, 2003; Toojinda *et al*, 2003; Tuberosa *et al*, 2003; Chardon *et al*, 2004). These comparisons have led to

approaches in meta-analysis for QTL results (Goffinet and Gerber, 2000; Khatkar *et al*, 2004), and bioinformatic tools are being developed to facilitate this effort (Arcade *et al*, 2004; Sawkins *et al*, 2004). Findings often indicate concordance of intervals for some QTL, but typically a considerable proportion of family-specific QTL appear to be identified.

Comparing QTL findings from different families fails to account explicitly for segregation of different allelic combinations among different parents and the context dependency of QTL effects. These deficiencies immediately lead to questions about the generalizability of QTL findings from experimental populations to breeding populations. Previous quantitative genetic studies, focusing on the general (GCA) and specific combining abilities (SCA) of parents for a particular complex trait (Griffing, 1956), were popular precisely because they allowed an estimation of the genetic architecture of the phenotype, with inference extending back to the reference population. This provided an understanding of the utility of selection programs for the trait of interest (Hallauer and Miranda, 1988).

Modern breeding therefore faces this problem: efficient marker-assisted selection requires knowledge of the value of a QTL in the context of the (population-wide) genetic architecture of the trait, but traditional QTL study designs do not generate this knowledge. Classic quantitative genetic studies provide insight into the general behavior of a trait, but do not identify

Correspondence: LM McIntyre, Computational Genomics and Department of Agronomy, Purdue University, Lilly Hall of Life Sciences, 915 W State Street, West Lafayette, IN 47907-2054, USA.

E-mail: lmcintyre@purdue.edu

³All authors contributed equally to this work

Received 16 March 2005; accepted 22 July 2005; published online 23 November 2005

specific genomic regions which can be selected and introgressed.

The benefits of both approaches can be combined by analyzing multiple recombinant families in a mating design with shared parents. Gilbert (1985a,b) pioneered this approach, partitioning single-gene effects from overall genomic (polygenic) effects in diallel crosses. We propose that using populations (eg F_2 or RIL) developed from a common group of parents and constructed using classic mating designs to jointly infer the combining abilities and QTL location is a synergistic combination of tools available to the modern breeder.

Important progress along this road has been made in theoretical studies on QTL detection methods in interconnected families, exploring segregation probabilities among more complex crosses (Xu, 1996), the efficiency of different statistical methods for QTL detection (Rebai and Goffinet, 1993; Liu and Zeng, 2000; Rebai and Goffinet, 2000; Yi and Xu, 2002a; Jannink and Wu, 2003); the number of parents in the mating design (Muranty, 1996), and the trade-off between the number of families and family size (Xie *et al*, 1998; Xu, 1998; Wu and Jannink, 2004). These studies have shown that using populations with shared parents leads to more generalizable inferences about QTL, and can lead to increased QTL detection power. In addition to these theoretical efforts, large populations of RILs with shared parental lines are being constructed (eg the NSF-funded project 'Molecular and Functional Diversity in the Maize Genome', NSF DBI 0321467).

Prior studies indicate that to produce generalizable results on the genetic architecture of complex traits several parents are needed (see, for instance Wu and Jannink, 2004). In this paper, we address the question of how best to develop RIL populations in an interconnected design with a given number of parents. The classic diallel design, where every parent is mated with every other parent, is extremely labor intensive. In combining ability analysis, it has long been established that partial-diallel designs (eg NC design II, factorial (FCT) designs, or circulant designs) may be preferable, as they are less labor intensive and can still result in considerable power (Kempthorne and Curnow, 1961; Dhillon and Singh, 1978).

In this study, we investigate the efficiency of several partial-diallel mating designs for the joint analysis of QTL and combining abilities. To relate QTL results obtained in a diallel of F_2 families back to variance components of a randomly mating reference population, we first provide a theoretical derivation of the among- and between-family variances for F_2 families, and F_2 extensions of the GCA and SCA variance components. We then use simulation techniques to determine the trade-offs between the number of crosses and the number of individuals examined within each cross under several different patterns of inheritance. We explore how different designs affect QTL detection power, and power for inferences about the underlying genetic architecture of the trait.

Methods

Simulations

We simulated diploid F_2 progeny derived from mating F_1 hybrids between 10 inbred parents in a half-diallel

mating design without selfs. Note that, as is common practice, we refer here to inbred 'parents' when the inbreds are in fact the grandparents of the F_2 progeny evaluated. Inbred parents were assumed to derive from a randomly mated reference population. This design yielded 45 F_2 families. At each locus the 10 parents all carried a distinct allele, resulting in 55 recognizable diploid genotypes per locus across the set of 45 F_2 families (10 homozygous and 45 heterozygous genotypes). Each F_2 family consisted of 500 individuals, and we performed 1000 replicate simulations.

We used an infinite alleles model for the reference population such that all QTL alleles and all marker alleles were different for each parent. This model has the advantage of simplicity and agrees well with standard quantitative genetic variance component models. QTL effects were simulated by specifying the additive, dominance, and additive-by-additive epistatic variance contributions of each locus (or pair of loci in the case of epistasis) to the total phenotypic trait variance. Genetic variances were specified with reference to a randomly mating population from which the inbred founding parents were assumed to have been extracted. In all simulations, genetic variances in the reference population contributed a total of 30% of phenotypic variance, leaving a residual error variance of 70%. Genetic effects were simulated as follows. For each QTL, 10 additive effects were sampled from a standard normal distribution. Each effect, denoted a_i , with $i = 1, \dots, 10$, corresponded to the allele carried by one of the founding inbred parents. The effects were standardized to have zero mean and a variance equal to half that specified for the QTL. Thus, in a randomly-mating population, a QTL with those alleles at equal frequencies would have generated the specified additive variance. Note that this simulation approach deviates from what would happen in a repeated sampling of an infinite alleles population. QTL variance in our simulation was constant from one simulation run to the next and always equaled the specified variance. In a repeated sampling of reference population, the sampled QTL variance would change from one run to the next. Our approach simplifies the interpretation of the results because of the constant QTL variance it provided. For dominance, 55 effects, denoted d_{ij} with $i, j = 1, \dots, 10$ and $d_{ij} = d_{ji}$, were sampled from a standard normal distribution. Each effect corresponded to the difference between the value conferred by a given diploid genotype and the value predicted by the additive effects of the genotype's two alleles. These effects were standardized so that $\sum_i d_{ij} = 0$ for any j , $\sum_j d_{ij} = 0$ for any i , and the variance of the d_{ij} was equal to that specified for the QTL. A consequence of this sampling scheme is that the variance of the d_{ii} , called the variance of homozygous dominance deviations (Edwards and Lamkey, 2002), is expected to be equal to the variance of the d_{ij} as a whole. In addition, the covariance between additive effects and homozygous dominance deviations is expected to be zero. While neither of these assumptions may hold *in vivo* (Edwards and Lamkey, 2002), they provide a useful starting point for exploring the question at hand. For additive-by-additive epistasis, 100 effects, denoted aa_{ij} with $i, j = 1, \dots, 10$, were sampled from a standard normal distribution. These effects were standardized so that $\sum_i aa_{ij} = 0$

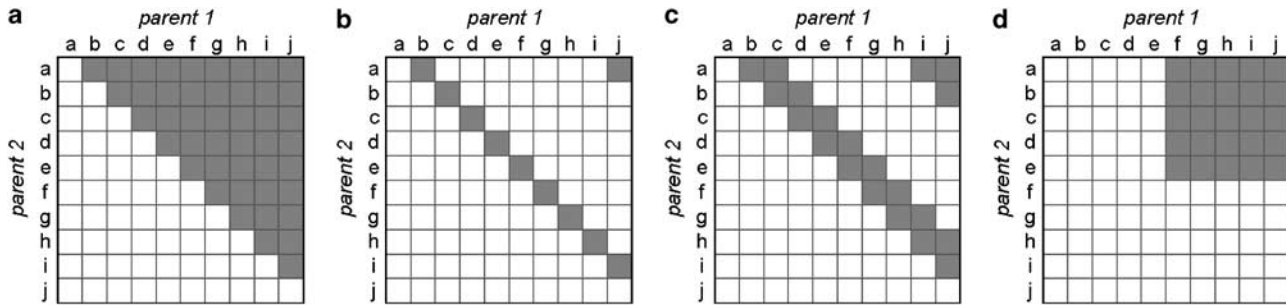


Figure 1 Different partial-diallel designs considered in our simulations. Shaded cells indicate crosses that are included in the analysis. (a) Half diallel; (b) SRR; (c) DRR; (d) FCT.

for any j , $\sum_i aa_{ij} = 0$ for any i , and the variance of the aa_{ij} was equal to one-fourth of that specified for the QTL pair. Each effect was associated with the two-locus haplotype receiving allele i at the first QTL of the pair and allele j at the second QTL. The effect corresponded to the difference between the value conferred by the haplotype and the value conferred by the additive effects of the haplotype's two alleles. The variance was again set so that the specified additive-by-additive epistasis would be generated by the QTL pair in a randomly mating reference population. Having sampled the genetic effects, the genotypic value of each progeny was generated by first simulating the progeny's QTL genotype, and then summing the genetic effects associated with that genotype.

Simulated genomes consisted of eight unlinked QTL. Two additive-effect QTL were adjacent to marker loci and each contributed 7.5% of the phenotypic variation. Marker loci were at the QTL position itself (M_2, M_5), flanking the QTL at 5 cM (M_1, M_4) and at 10 cM (M_3, M_6) according to the map $M_1M_2M_3$ on one linkage group and $M_4M_5M_6$ on another. Markers were multiallelic and informative in all crosses. The six remaining QTL together contributed 15% of the phenotypic variance and introduced unmarked or *polygenic* variation in trait values. We explored three different partitions of the polygenic variance into its additive, dominance and epistatic components: only additive variance (15%: 0%: 0% for additive, dominance, and epistatic, respectively); only dominance and epistatic (0%: 6%: 9%); and additive plus dominance plus epistatic variance (6%: 6%: 3%). In all cases, additive and dominance variances were partitioned evenly across the six polygenes, and the epistatic variance was partitioned evenly across three pairs of polygenes. One additional linkage group with three linked markers ($M_7M_8M_9$) without QTL was also simulated.

We examined all 45 F_2 families in our analysis of the half-diallel design. We compared this design to three other partial-diallel designs: the single-round robin (SRR), double-round robin (DRR), and FCT (Figure 1). The SRR, DRR, and FCT designs were constructed by subsetting the half diallel. This ensured that the comparisons across designs were made on the same data. Each design was analyzed over a range of family sizes (10, 25, 50, 100, 200, or 500 individuals per F_2 family included in the analysis). The total size of a given experiment is the product of the number of families by the number of individuals analyzed per family.

QTL detection and analysis of combining abilities

We discuss the following ANOVA models for detection of marked QTL:

$$Y_{fmk} = \mu + M_m + \varepsilon_{fmk} \quad (1)$$

$$Y_{fmk} = \mu + \phi_f + M_{m(f)} + \varepsilon_{fmk} \quad (2)$$

where Y is the phenotype for the quantitative trait of interest, μ is the overall mean, M is the effect of marker m nested within family f in Model 2, and ϕ is the effect of family f ($f = 1, \dots, n_1$), where n_1 is the number of families (10, 20, 25, 45).

The family effect can be partitioned into an additive component that is attributed to parental main effects and a nonadditive component that is attributed to parental interactions. These are F_2 extensions of the GCA and SCA variance components, which are traditionally defined with respect to the F_1 progeny (Sprague and Tatum, 1942):

$$Y_{ijk} = \mu + G_i + G_j + S_{ij} + \varepsilon_{ijk} \quad (3)$$

where G_i and G_j are the GCA effects of the parents i and j and S is the effect of the parental combination (Griffing, 1956; Lynch and Walsh, 1998). Sums of squares associated with the GCA effect were obtained by including binary indicator variables in the design matrix that coded the involvement of each parent in the cross (see Johnson and King, 1998). Following a hierarchical decomposition of variance inherent to diallel designs, we used type I, sequential, analyses to first obtain the sums of squares associated with GCA and then those associated with SCA (after accounting for GCA). Significance tests were carried out according to standard diallel theory: GCA mean squares were tested over SCA mean squares, and SCA mean squares were tested over the model error mean squares (Lynch and Walsh, 1998). All tests were conducted at a nominal significance level of $\alpha = 0.05$. Note that Model 3 cannot be fit for the SRR as GCA and SCA effects are only separately estimable when the number of families generated exceeds the number of parents used in the mating design. Therefore, only Model 2 was fit for the SRR data.

After Model 2 has led to the detection of QTL (M_2 and M_5), it can be reordered to test for residual genetic effects not associated with the marked QTL. We refer to loci generating such residual effects as polygenes (Mather, 1941). Their presence is examined using Model 4:

$$Y_{fk} = \mu + M_2 + M_5 + \phi_f + \varepsilon_{fk} \quad (4)$$

where M_2 and M_5 are the markers at the identified QTL positions. As with Model 1, marker effects are fitted before family structure is taken into account. The effects of the two marked QTL are first removed (type I sums of squares) and the family effects (f) estimated are residual effects, leading to a test of the null hypothesis that no undetected polygenic effect is present. A rejection of this null hypothesis indicates the presence of unmarked polygenic effects. As in Model 3, the family effect can be partitioned into its GCA and SCA components:

$$Y_{ijmk} = \mu + M_2 + M_5 + G_i + G_j + S_{ij} + \varepsilon_{ijmk} \quad (5)$$

which splits the residual family effect into a GCA effect due to parental main effects and an SCA effect due to parental interactions. If the effects considered in the models were orthogonal, then the order of the model would be irrelevant, and Model 4 would give identical answers to Model 2. However, the lack of orthogonality in these models makes the consideration of both Model 2 and Model 4 relevant. Note that all the models considered here do not include epistatic terms. A more general approach would be to account for the interactions among loci, as well as their main effects. In the case where the mode of action of the QTL is unknown, models that include epistatic terms are to be preferred. As our simulations only included additive QTL, for demonstration purposes, we did not include these terms in our models.

When parental effects are considered random, the GCA and SCA components of variance can also be estimated. As a demonstration, for a single case (half-diallel design, polygenes contributing only dominance and epistatic variance, 500 individuals per F_2 family) we estimated the variance components by fitting a family effects model, $Y_{fjk} = \mu + \phi_f + \varepsilon_{fjk}$, and Model 3 using REML (Searle *et al*, 1992; Zhu and Weir, 1996). The estimated variances were compared to the expected variances of GCA and SCA effects according to the theoretical decomposition of among- F_2 family variance. While different approaches have been developed for the analysis of diallel designs (Hayman, 1954a,b; Griffing, 1956; Gardner and Eberhart, 1966), partial diallels (Kempthorne and Curnow, 1961; eg Fyfe and Gilbert, 1963; Viana *et al*, 1999) and extensions of analysis to F_2 generations (Jinks, 1956; Hill *et al*, 2001), for the purposes of demonstration we follow the approach of Cockerham (1983) to obtain covariances among relatives derived by self-fertilization. All models were fit and tested in SAS (SAS Institute Cary, NC, USA).

Results

Decomposition of among- F_2 family variance

The following decomposition applies to among- F_2 family variances in the absence of QTL analysis, as well as to among-family variances that are residual once QTL main effects have been removed. Cockerham (1983) gave formulas for obtaining covariances between relatives derived from self-fertilization caused by single-locus effects. The relevance of these results to our context stems from the following reasoning. Assume a randomly mating reference population from which inbreds are extracted, then crossed in some form of diallel. The resulting F_1 progeny genotypes occur in the diallel in

equal frequencies as they would in the reference population. The F_1 are subsequently self-fertilized to obtain F_2 progeny. Cockerham's (1983) results provide the relevant single-locus total among-progeny variance components and among-family variance components. We present these results first, then extend them to two-loci additive-by-additive epistasis.

Denoting the progeny inbreeding coefficient F_g , the total among-progeny variance is

$$\begin{aligned} V_T &= (1 + F_g)V_A + (1 - F_g)V_D \\ &\quad + 4F_gD_1 + F_gD_2 + F_g(1 - F_g)H^* \\ &= \frac{3}{2}V_A + \frac{1}{2}V_D + 2D_1 + \frac{1}{2}D_2 + \frac{1}{4}H^* \end{aligned}$$

The among-family variance is

$$\begin{aligned} V_B &= V_A + (1 - F_g)^2V_D + 2F_gD_1 + \frac{1}{2}F_g^2D_2 + 0H^* \\ &= V_A + \frac{1}{4}V_D + D_1 + \frac{1}{8}D_2 \end{aligned}$$

By subtraction, we can obtain the within-family variance as

$$\begin{aligned} V_W &= \frac{1}{2}V_A + F_g(1 - F_g)V_D + 2F_gD_1 \\ &\quad + F_g(1 - \frac{1}{2}F_g)D_2 + F_g(1 - F_g)H^* \\ &= \frac{1}{2}V_A + \frac{1}{4}V_D + D_1 + \frac{3}{8}D_2 + \frac{1}{4}H^* \end{aligned}$$

To extend these results for covariances caused by two-loci additive-by-additive effects, define $\theta_{tgg'}$ as the probability that, for both of two unlinked loci, a random allele from g is IBD to a random allele from g' , where g and g' are the products of self-fertilizing their common ancestor that was itself the product of t generations of self-fertilization. The resemblance between g and g' due to additive-by-additive epistasis is $4\theta_{tgg'}^2 V_{AA}$. Three cases can be distinguished. First, with probability $(1 - F_t)^2$, the common ancestor was not inbred at either locus. In that case, the probability of IBD between random alleles from g and g' is $\frac{1}{2}$ at each locus, and since the loci are unlinked, they segregate independently and $\theta_{tgg'}^2 = \frac{1}{4}$. Second, with probability $F_t(1 - F_t)$, the common ancestor is inbred at one locus, but not the other. In that case $\theta_{tgg'}^2 = \frac{1}{2}$. Finally, with probability F_t^2 , the common ancestor is inbred at both loci, in which case $\theta_{tgg'}^2 = 1$. Summing these conditional probabilities gives the unconditional probability $\theta_{tgg'}^2 = \frac{1}{4}(1 + F_t)^2$. Thus, for independent loci, $\theta_{tgg'}^2$ is the square of the $\theta_{tgg'}$ coefficient defined by Cockerham (1983). In our context, setting t to zero and one provides the coefficients for among-family and total among-progeny variance, respectively. Adding these terms to the results above gives

$$\begin{aligned} V_T &= \frac{3}{2}V_A + \frac{1}{2}V_D + 2D_1 + \frac{1}{2}D_2 + \frac{1}{4}H^* + \frac{9}{4}V_{AA} \\ V_B &= V_A + \frac{1}{4}V_D + D_1 + \frac{1}{8}D_2 + V_{AA} \\ V_W &= \frac{1}{2}V_A + \frac{1}{4}V_D + D_1 + \frac{3}{8}D_2 + \frac{1}{4}H^* + \frac{5}{4}V_{AA} \end{aligned}$$

The among-family variance can be partitioned further into components caused by GCA and SCA. The necessary coefficients of identity to deduce the GCA component follow. Denote by g and g' progeny derived from the inbred line crosses $A \times B$ and $A \times C$, respectively, after g generations of selfing. The lines g and g' share one common inbred ancestor, and their covariance will therefore provide the GCA variance. Denote by F_g ,

$F_{g'}$, $\theta_{gg'}$, $\gamma_{gg'}$, $\Delta_{g-g'}$, $\delta_{gg'}$, and $\delta_{g+g'}$ the inbreeding coefficients of g and g' , their coefficients of coancestry, and their three- and four-gene coefficients of identity for the relationship defined here, and using a notation similar to Cockerham (1983). A two-locus coefficient of coancestry $\theta_{gg'}$, similar to the coefficient $\theta_{igg'}$ described above, will account for additive-by-additive epistatic contributions to the GCA variance. The inbreeding coefficient is given by the usual $F_g = 1 - (\frac{1}{2})^g$. The coefficient of coancestry $\theta_{gg'} = \frac{1}{4}$ irrespective of the levels of inbreeding g and g' . For unlinked loci, $\theta_{gg'}^2 = \frac{1}{16}$. For $\gamma_{gg'}$ there is a probability of F_g that g is inbred, and if it is inbred, a probability of $\frac{1}{4}$ that a random gene from g' is IBD to it. Therefore, $\gamma_{gg'} = \frac{1}{4}F_g$. Similarly, $\gamma_{gg'} = \frac{1}{4}F_{g'}$. There is a probability of $F_g F_{g'}$ that both g and g' are inbred at a locus, so $\Delta_{g-g'} = F_g F_{g'}$. If both g and g' are inbred at a locus, there is a probability of $\frac{1}{4}$ that they are also IBD to each other, so $\delta_{gg'} = \frac{1}{4}F_g F_{g'}$. Finally, $\delta_{g+g'}$ is the half the probability that neither g nor g' are inbred and that two pairs of alleles between g and g' are IBD. Since g and g' have only one common ancestor, this probability is zero. Specifying these coefficients to the situation of F_2 families, we find that the covariance between progeny with one common inbred ancestor is

$$\begin{aligned} V_{GCA} &= \text{cov}(g, g') \\ &= \frac{1}{2}V_A + 0V_D + \frac{1}{2}(F_g + F_{g'})D_1 \\ &\quad + \frac{1}{4}F_g F_{g'} D_2 + 0H^* + \frac{1}{4}V_{AA} \\ &= \frac{1}{2}V_A + \frac{1}{2}D_1 + \frac{1}{16}D_2 + \frac{1}{4}V_{AA} \end{aligned}$$

Since each family has two parents, the among-family variance consists of twice the GCA variance plus the SCA variance. The SCA variance is obtained by subtraction:

$$\begin{aligned} V_{SCA} &= V_B - 2V_{GCA} \\ &= 0V_A + (1 - F_g)^2 V_D + 0D_1 + 0D_2 + 0H^* + \frac{1}{2}V_{AA} \\ &= \frac{1}{4}V_D + \frac{1}{2}V_{AA} \end{aligned}$$

Dr Wyman Nyquist independently obtained these same results (pers. comm.). As with traditional diallel analyses, GCA variance is largely associated with additive QTL effects, while only nonadditive genetic effects contribute to SCA variance, including dominance, additive-by-additive, and additive-by-dominance effects. Note that the additive, dominance, and epistatic variances simulated for each QTL were specified with reference to a randomly-mating population and contribute to GCA and SCA differently when measured in the F_1 versus the F_2 generation, due to the generation of inbreeding resulting from intermating F_1 individuals (Jinks, 1956; Hill et al, 2001).

In our simulations, V_A , V_D , and V_{AA} are specified. Given a finite number of inbred parents used in the diallel, the variance of the inbred dominance deviations, D_2 , will depend on the deviations sampled in each simulation run. In the expectation, however, it will be equal to V_D . As we sample additive and dominance effects independently, the expectation of D_1 is zero. However, for a specific simulation, D_1 may be either positive or negative due to the specific effects sampled. Given that the expectation of the average inbred dominance deviation is zero in our simulations, the expectation of the squared average, H^* , is equal to the

variance of the average. That variance depends on the number of inbred parents included in the diallel, n_{Par} , and is V_D/n_{Par} . In practice, the variance contributed by that component will generally be small.

QTL analysis

As expected, a naive search for marker-trait associations that failed to take into account the family structure of the F_2 progeny (Model 1) resulted in highly significant effects of all markers, irrespective of linkage to a QTL (Figure 2). In the half-diallel design, for example, tests for differences among the 55 genotypic classes of any marker on the neutral linkage group (without QTL) yielded a significant result in at least 94% of the replicate simulations, even at the lowest family size. When the family structure was appropriately accounted for (Model 2), the type I error estimated by examining unlinked markers was found to be consistent with the threshold set for significance testing (0.05, see Figure 2b). All further analyses, therefore, follow Model 2.

For a fixed number of individuals per F_2 family, designs with more families have more power to detect the marked QTL, as the total experiment size increases with the number of families (Table 1, Figure 3a). In contrast, when the total experimental size (number of F_2 families times the number of individuals per family) was

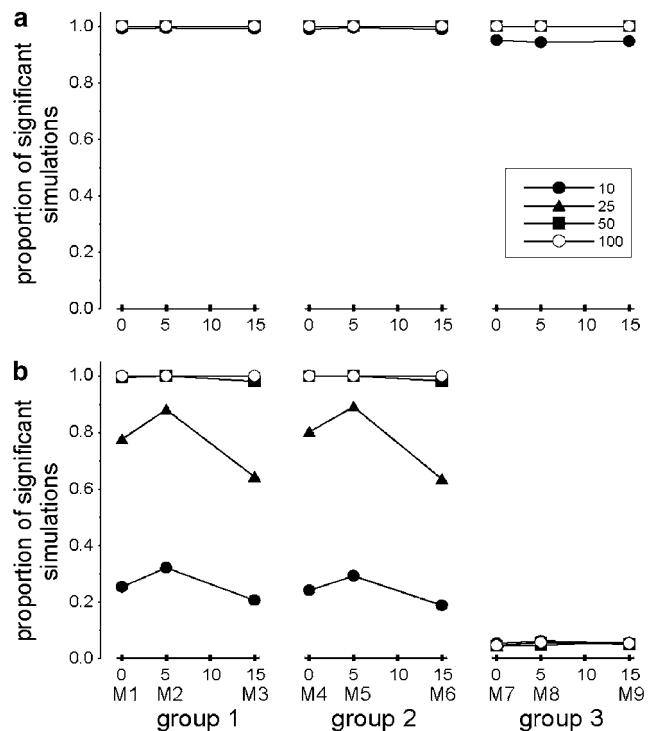


Figure 2 Marker-trait analysis with and without accounting for the family structure of the F_2 progeny, for analyses including different numbers of individuals per F_2 family. For each of nine markers the proportion of significant simulations is given (1000 replicated simulations; $\alpha = 0.05$). QTL were present on groups 1 and 2 (at M_2 and M_5), but not on group 3. (a) Naive search that does not take into account the family grouping of F_2 individuals; (b) nested approach, where marker effects are tested within families. Results are shown for the half-diallel design, under purely nonadditive polygene action.

Table 1 QTL analysis: power to detect the two marked QTL at a fixed number of 10, 25, and 50 individuals per F₂ family, in different partial-diallel designs and with different modes of polygene action

Diallel design (no. of families)	Polygene action		
	Additive	Nonadditive	Both
Family size	10 · 25 · 50	10 · 25 · 50	10 · 25 · 50
<i>Half diallel (45)</i>			
QTL 1	0.38 · 0.93 · 1.00	0.32 · 0.88 · 1.00	0.34 · 0.91 · 1.00
QTL 2	0.36 · 0.94 · 1.00	0.29 · 0.89 · 1.00	0.31 · 0.91 · 1.00
<i>Factorial (25)</i>			
QTL 1	0.28 · 0.77 · 0.99	0.23 · 0.69 · 0.99	0.24 · 0.71 · 0.99
QTL 2	0.25 · 0.77 · 0.99	0.22 · 0.69 · 0.98	0.22 · 0.72 · 0.99
<i>Double-round robin (20)</i>			
QTL 1	0.24 · 0.69 · 0.97	0.20 · 0.63 · 0.95	0.22 · 0.64 · 0.97
QTL 2	0.23 · 0.67 · 0.97	0.19 · 0.65 · 0.96	0.20 · 0.66 · 0.96
<i>Single-round robin (10)</i>			
QTL 1	0.16 · 0.47 · 0.83	0.15 · 0.41 · 0.78	0.16 · 0.45 · 0.82
QTL 2	0.16 · 0.46 · 0.83	0.14 · 0.41 · 0.79	0.16 · 0.43 · 0.80

The proportion of replicated simulations (out of 1000) is given, in which the marker that was completely linked with the QTL showed a significant association with the phenotype at the $\alpha = 0.05$ level.

fixed, the partial-diallel designs with fewer and larger families (eg single and double round-robin designs) were more powerful in detecting QTL than designs with more and smaller families (Figure 3b). For instance, at a fixed total experimental size of 500 individuals, an additive QTL that accounted for 7.5% of the phenotypic variance was picked up in about 80% of the replicated simulations in the SRR design, as opposed to only about 35% in the half-diallel design. Power was affected slightly by the mode of action of the polygenes; marked QTL were more readily detected if polygenes acted additively (Table 1).

Exploring undetected polygene effects

Power to detect residual polygenic effects was dependent on the number of families as well as the total experiment size (Figure 4). To obtain high power to detect these residual effects, large experiment sizes were needed, containing between 1000 and 2000 progeny. Neither the mode of action of the polygenes (additive versus nonadditive) nor the specific partial-diallel design had a clear effect on the power to detect this effect (Figure 4). Further, Model 5 correctly identified the mode of polygene action: (1) a significant GCA effect but no SCA effect when only additive polygenes were specified; (2) a significant SCA effect but no GCA effect when only nonadditive polygenes were specified; and (3) significant GCA and SCA effects when polygenes contributed additive as well as nonadditive variance (Figure 4). Fitting the marker effects did, however, absorb part of the total genetic variance attributable to polygenic effects, as illustrated by the fact that SCA was detected less easily after fitting the marker effects even when the linked QTL themselves did not contribute any nonadditive variance (Figure 5).

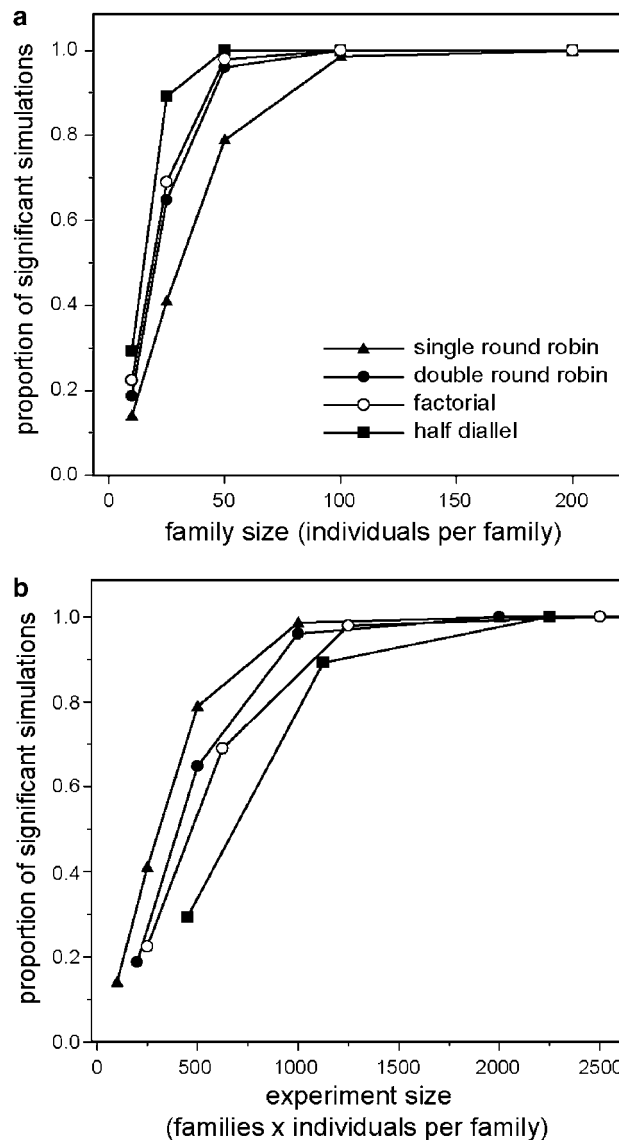


Figure 3 Power of different partial-diallel designs to detect an additive QTL that contributes 7.5% of the phenotypic variance, expressed as the proportion of simulations (out of 1000) in which the completely linked marker showed a significant association with the phenotype at the $\alpha = 0.05$ level. Each diallel design was analyzed with 10, 25, 50, 100, and 200 numbers of F₂ individuals per family, yielding different total experimental sizes for the designs that include different numbers of families. Results are shown for one of the two specified QTL, under purely nonadditive polygene action. (a) Comparison of designs at equal numbers of individuals per family; (b) Comparison of designs at equal total experimental sizes.

Among-F₂ family variance components

Variance components were estimated in PROC MIXED using the REML option (SAS Institute, Cary, NC, USA) by sequentially fitting the model terms. That is, the GCA is fit first and the residuals from the GCA model are then used as the dependent variable in the SCA analysis. The model is fit sequentially because of the structure of the GCA and SCA terms. Since PROC MIXED estimates type III variance components, a single model including GCA and SCA will be problematic because these terms are not

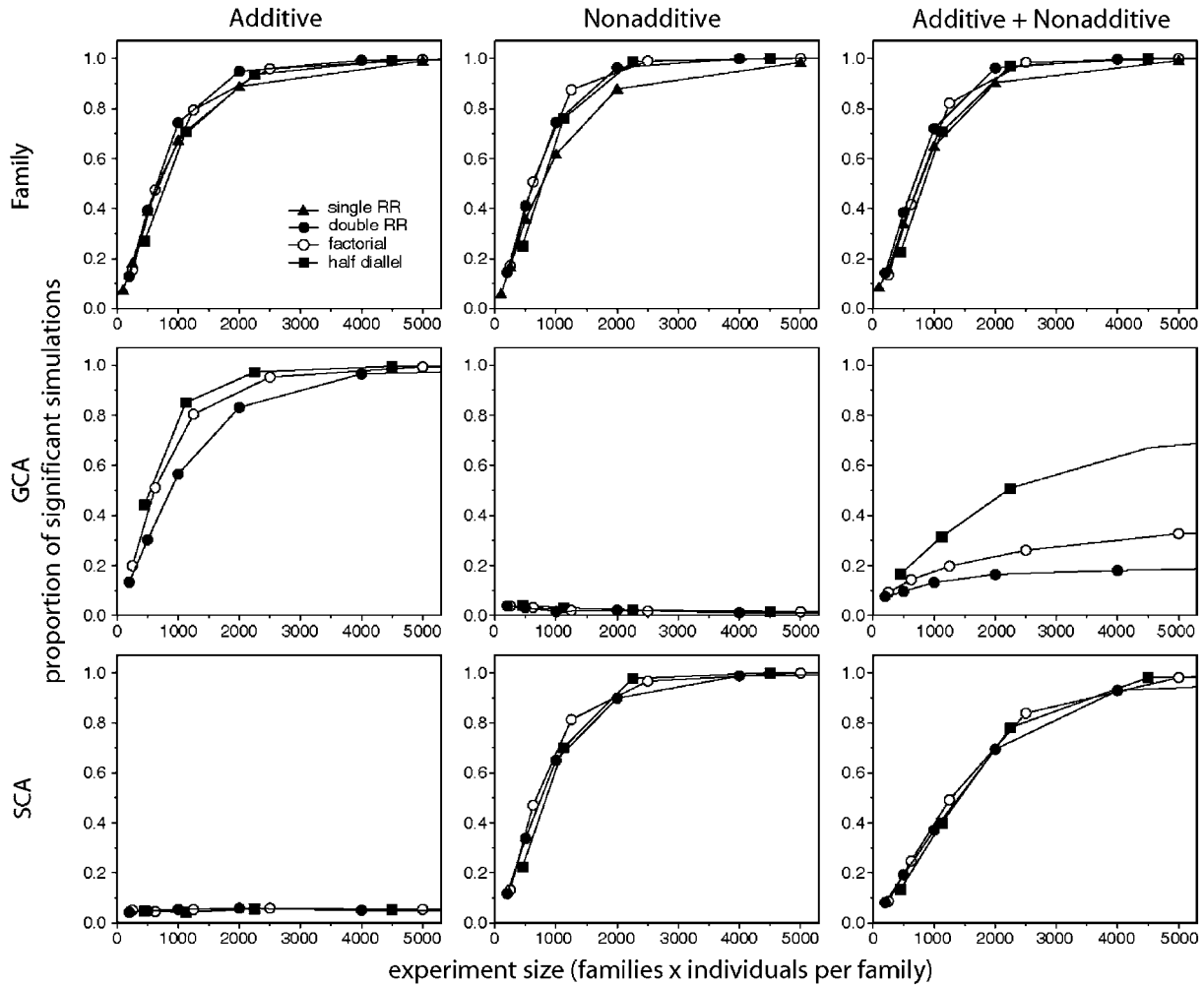


Figure 4 Analysis of residual among-family variance after fitting the two detected QTL. Upper panels show the proportion of replicated simulations (out of 1000) in which the among-family variance was significant at the $\alpha = 0.05$ level, in different diallel designs and for different total experimental sizes. In the middle and lower panels, the among-family variance is partitioned into its components GCA and SCA.

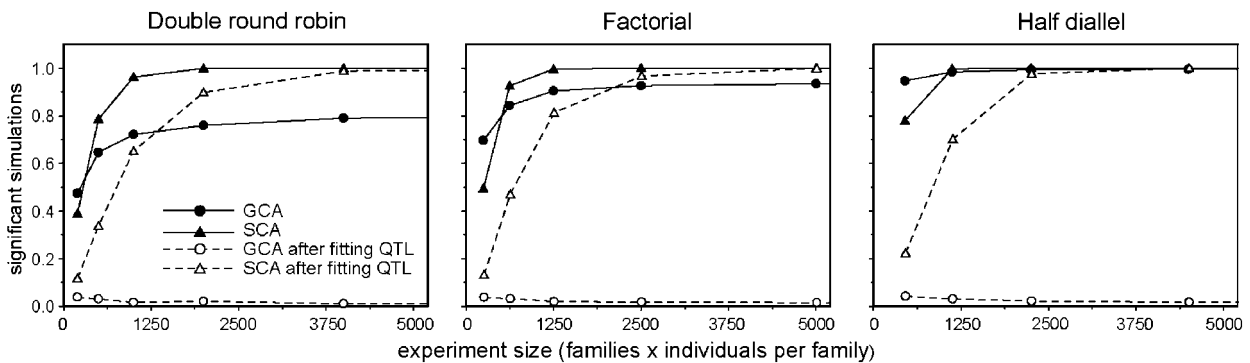


Figure 5 Proportion of replicated simulations (out of 1000) in which GCA (circles) and SCA (triangles) effects were significant at the $\alpha = 0.05$ level, in different diallel designs and for different total experimental sizes. Dashed lines show results for models that account for effects of the detected QTL before testing GCA and SCA effects. Solid lines show results for models that do not include detected QTL. Results are shown for simulations where the polygenes contributed only nonadditive variance; marked QTL contributed only additive variance.

orthogonal. Estimated variances were concordant with expectations, but appear to slightly underestimate their expected values, although the empirical confidence interval does contain the true parameter value (Table 2).

Discussion

Joint analysis of multiple interconnected families can be a powerful strategy to extend QTL analysis beyond the limitations of single crosses (Rebai and Goffinet, 1993;

Table 2 Expected and observed variance components in the F₂ generation

Component	Expected		Average (SD)
	General	Simulated case	
$V_{\text{between families}}$	$V_A + \frac{3}{8}V_D + V_{AA}$	= 26.25	22.04 (6.47)
V_{GCA}	$\frac{1}{2}V_A + \frac{1}{16}V_D + \frac{1}{4}V_{AA}$	= 10.13	9.52 (6.92)
V_{SCA}	$\frac{1}{4}V_D + \frac{1}{2}V_{AA}$	= 6.00	5.87 (1.38)

Expectations are based on QTL variances specified with respect to a randomly mating base population, and assume random pairing of F₁ individuals to generate F₂ families. The table shows the case where marked QTL contribute additive variance and polygenes contribute only dominance and additive-by-additive variance ($V_A=15$; $V_D=6$; $V_{AA}=9$). Estimated values are from the half-diallel design with 500 individuals per F₂ family, and are averaged over 1000 replicated simulations. The standard deviation over replicated simulations is also given.

Muranty, 1996; Rebaï and Goffinet, 2000; Jannink and Jansen, 2001). We find that, given a fixed number of parents and total experimental size, some partial-diallel designs are more efficient than others in detecting QTL. Strikingly, we find that using the diallel as a basis for developing interconnected families allows for combining QTL analysis with traditional analysis of GCA and SCA, providing insight into the genetic architecture of the trait and revealing the mode of action of undetected QTL.

QTL detection power in different designs

For markers unlinked to QTL, associations with the phenotype arise (see Figure 2a) because of linkage disequilibrium generated by the mating design. Alleles at unlinked loci will be in linkage equilibrium *within* families, but they will be in linkage disequilibrium *across* families. If we denote by M₁ and Q₁ alleles from a marker and an unlinked QTL carried by an arbitrary inbred parent P₁, linkage equilibrium means that $P(Q_1 | M_1) = P(Q_1)$. In gametes carrying M₁, $P(Q_1 | M_1) = \frac{1}{2}$. Within families of P₁, $P(Q_1) = \frac{1}{2}$, and linkage equilibrium holds. Across all families, however, it is generally not true that $P(Q_1) = \frac{1}{2}$. For example, with 10 parents in the design and assuming that each parent contributes to an equal number of families of equal size, $P(Q_1) = 0.1$, such that Q₁ and M₁ are associated more strongly than expected under equilibrium, which will generate a spurious association between marker and phenotype. To avoid this problem, it is necessary to test for QTL association by nesting marker effects within families. This is clearly demonstrated by our findings using Model 1 (Figure 2). Note that, if *all* genetic variance is accounted for by marked QTL in a multiple QTL model, no residual genetic variation will be available to associate with unlinked markers, and there would be no need to account for family structure. While this situation is theoretically possible, it seems unlikely in practice, given that many small QTL are likely to remain undetected. In addition, non-nuclear effects and interactions between nuclear and extra-nuclear effects may also be detected as a polygene effect in these designs.

Comparing different partial-diallel mating designs using outbred parents, Muranty (1996) showed that, when the number of families and the number of

offspring per family are held constant, QTL detection power is affected by the number of parents used in the design, but the arrangement of the parents in specific diallel designs has little impact. Based on this result, we explored QTL detection power with a fixed number of parents in order to compare different mating designs under the assumption that all designs sample the same genetic variation (10 parents). For a fixed experiment size, we found that QTL detection power was greatest for the mating design with the fewest but largest families (the SRR), while power was lowest for the design with the most, smallest families (the half diallel). This is in agreement with previous findings (eg Soller and Genizi, 1978; Wu and Jannink, 2004). In small families, the stochastic nature of segregation may cause some QTL or marker genotypes to be represented by very few (or zero) individuals. For example, for an F₂ family size of 10, the probability that one of the homozygote genotypes is missing is 10.9%. When family size increases to 20, this probability drops to 0.6%, but the probability that one of the homozygote genotypes is represented by only one individual is still 4.8%. In such situations the family in question contributes little information about the marker effect, reducing power of the QTL analysis. As all our designs involved 10 parents, the minimum number of families was 10 (in the SRR design). Further reduction in the number of families, by reducing the number of parents involved in the crossing design, may further increase the power to detect QTL segregating in the pedigree, but may decrease QTL detection efficiency due to insufficient sampling of QTL alleles of the base population from which the parents were derived (Muranty, 1996; Wu and Jannink, 2004).

We focused on detecting additive QTL that each explained 7.5% of total variance, with sample sizes ranging from 10 to 500 per F₂ family. Power of QTL detection was considerable even with small family sizes, due to the ability to borrow power across families and due to the relatively simple QTL effects that we simulated. For exact *localization* of QTL (as opposed to *detection*) the effects of small family size may be different, as the power that can be borrowed across families will depend on the degree that recombination probabilities between markers are similar between the families. Also, with more complex (eg epistatic) QTL models more segregants per family will need to be examined. We also use a very straightforward single-marker analysis for QTL detection; however, the approach described is general and can be used in combination with any QTL detection procedure. The important point is to account for the family structure in the testing procedure.

Another simplification of our analysis is that we assumed an infinite alleles model such that for both markers and QTL all parents were assumed to carry unique alleles. Clearly, having completely informative markers will increase QTL detection power relative to the more realistic case in which markers will not be informative in all families. The use of multiple linked markers in a region, denser marker maps and interval mapping approaches can mitigate this problem. Also, realistically, individual QTL may not segregate in all families as in the simulations. In the simulations, though the two parents of a family will always carry distinct QTL alleles, the difference in their allelic effects may be small, mimicking a nonsegregating situation. The overall

variance of allelic effects will have a much greater influence on QTL detection power than whether that variance is composed of an 'infinite' allelic series or of a discrete set of few alleles.

We noted a modest but consistent effect of the polygenes' mode of action on QTL detection power: power was lower when polygenes contributed nonadditive variance (Table 1). This is caused by the differential contribution of V_A , V_D , and V_{AA} to within- F_2 family variance: 15% nonadditive variance, specified with reference to a randomly-mating base population, contributes more variance within F_2 families than 15% additive variance does (see Decomposition of F_2 family variance). In the QTL detection models, polygenic variance contributes to the error variance, making the QTL more readily detectable when polygenes act additively.

Detecting polygenic effects

Having detected QTL and statistically controlled for their effects, the remaining genetic variation was explored using traditional quantitative genetic approaches to assess the modes of gene action of undetected QTL. A first step in this process involves determining whether there is residual variation unaccounted for by the marked QTL, as indicated by a significant family effect in Model 4. An important result of our simulations is that the detection of this effect does not depend on the mode of action of the unmarked polygenes (Figure 4). Having detected residual (polygenic) variance, it becomes meaningful to determine the mode of gene action generating the variance by partitioning it into components attributable to GCA and SCA effects. Knowing the predominant mode of polygene action will give insight into the feasibility of identifying their map locations through further experimentation. In particular, additive-effect loci will be more easily identifiable than polygenic loci that generate purely dominant or epistatic variances.

Of note, when marker effects are fit before family effects, as is necessary to make inferences on residual polygenic effects, a portion of the polygenic variance is absorbed in the marker classes due to mating-design-wide linkage disequilibrium between polygenes and markers (see Discussion above). This affects the analysis of polygenic effects (see Figure 5). Even so, in the simple regression-based approach proposed here, polygene action is still detectable and can even be partitioned into additive and nonadditive components. For the half-diallel, DRR, and FCT designs, we tested whether the polygenic effect was predominantly additive (partitioned to GCA) or predominantly dominant or epistatic (partitioned to SCA). Not surprisingly, we find that designs in which each parent contributes to more families (more connected designs) are more efficient in the GCA/SCA analysis of this polygenic variance than less connected designs (Kempthorne and Curnow, 1961).

For a fixed experiment size, more connected designs have fewer individuals per family than less connected designs. As a result, the power for detection of the QTL is lower. The potential gain of insight that more connected designs provide into the mode of action of polygenes must be balanced against the need for an adequate number of segregants per family for QTL detection power. If the primary goal of the study is to detect the relevant QTL, our results suggest that a SRR design is a

good initial choice – although subsequent experiments using more connected designs would be required to obtain a comprehensive understanding of the QTL and polygene effects. If, in contrast, the main goal is to understand the inheritance of the trait with an initial estimate of QTL location, one of the more connected designs will be a better choice.

Variance component estimations

Given specification of V_A , V_D , and V_{AA} of the QTL and polygenes in our simulations, we derived theoretical expectations for the GCA and SCA variances in the F_2 . The empirically observed variances appeared to deviate from these expectations; in particular, among-family variance was slightly lower than predicted (Table 2). Since marker effects were not included in the models estimating variance components in Table 2, absorption of family effects by markers does not explain this underestimation. The theoretical expectations were derived assuming random mating, but the partial-diallel mating designs are not random insofar as selfing is not allowed. The exclusion of selfing means that parents were paired using sampling without replacement, which generates a negative covariance between the effects of parents of a family. This negative covariance would be expected to depress the among-family variance. Support for this hypothesis comes from the fact that GCA variance showed a smaller deviation from expectation than overall family variance. That is, predicted family variance from the GCA/SCA analysis is $2 \times 9.52 + 5.87 = 24.91$, which is greater than the directly estimated family variance of 22.04 (Table 2). Each parent's GCA effect is estimated while controlling for the GCA of its mates, which, to a large extent, eliminates any negative covariance.

Conclusion

We demonstrate that by taking a classical quantitative genetic approach to mapping we can both locate specific genomic regions associated with a phenotype, and dissect the genetic architecture of the trait. Having multiple populations in the study ensures that QTL findings are more generalizable than findings from single-cross experiments. By combining diallel mating designs with QTL mapping, the identification of specific loci responsible for trait variation and the assessment of the generalizability of the findings will allow breeders to decide how to focus their efforts. In addition, the quantitative assessment of the amount and type of unexplained genetic variation in the reference population can serve as a guide to further experimentation.

As multiple mapping populations are expensive to generate, co-ordinated efforts to create suitable populations will be crucial. We show here that efforts to create populations in a design that is suited to GCA and SCA estimation (or to augment the existing designs to such an interconnected design) can greatly increase the value of the mapping resource and therefore should be strongly encouraged.

Acknowledgements

We thank Jiao Song for her help with running and analyzing the simulations, and Wyman Nyquist for helpful discussions. This study is based upon work

supported by the National Science Foundation under Grant No. 9904704.

References

- Arcade A, Labourdette A, Falque M, Mangin B, Chardon F, Charcosset A *et al* (2004). BioMercator: integrating genetic maps and QTL towards discovery of candidate genes. *Bioinformatics* **20**: 2324–2326.
- Carlborg O, Andersson L, Kinghorn B (2000). The use of a genetic algorithm for simultaneous mapping of multiple interacting quantitative trait loci. *Genetics* **155**: 2003–2010.
- Chardon F, Virlon B, Moreau L, Falque M, Joets J, Decousset L *et al* (2004). Genetic architecture of flowering time in maize as inferred from quantitative trait loci meta-analysis and synteny conservation with the rice genome. *Genetics* **168**: 2169–2185.
- Clancy JA, Han F, Ullrich SE (2003). Comparative mapping of beta-amylase activity QTLs among three barley crosses. *Crop Sci* **43**: 1043–1052.
- Cockerham CC (1983). Covariances of relatives from self-fertilization. *Crop Sci* **23**: 1177–1180.
- Coffman CJ, Doerge RW, Simonsen KL, Nichols KD C, Wolfinger R, McIntyre LM (2005). Model selection in binary trait locus mapping. *Genetics* **170**: 1281–1297.
- Corander J, Sillanpää MJ (2002). A unified approach to joint modeling of multiple quantitative and qualitative traits in gene mapping. *J Theor Biol* **218**: 435–446.
- Dhillon BS, Singh J (1978). Evaluation of circulant partial diallel crosses in maize. *Theor Appl Genet* **52**: 29–37.
- Edwards JW, Lamkey KR (2002). Quantitative genetics of inbreeding in a synthetic maize population. *Crop Sci* **42**: 1094–1104.
- Flint-Garcia SA, Jampatong C, Darrah LL, McMullen MD (2003). Quantitative trait locus analysis of stalk strength in four maize populations. *Crop Sci* **43**: 13–22.
- Fyfe JL, Gilbert N (1963). Partial diallel crosses. *Biometrics* **19**: 278–286.
- Gardner CO, Eberhart SA (1966). Analysis and interpretation of the variety cross diallel and related populations. *Biometrics* **22**: 439–452.
- Gilbert DG (1985a). Estimating single gene effects on quantitative traits. 1. A diallel method applied to Est 6 in *D. melanogaster*. *Theor Appl Genet* **69**: 625–629.
- Gilbert DG (1985b). Estimating single gene effects on quantitative traits. 2. Statistical properties of five experimental methods. *Theor Appl Genet* **69**: 631–636.
- Goffinet B, Gerber S (2000). Quantitative trait loci: a meta-analysis. *Genetics* **155**: 463–473.
- Griffing B (1956). Concept of general and specific combining ability in relation to diallel crossing systems. *Aust J Biol Sci* **9**: 463–493.
- Haley CS, Knott SA (1992). A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* **69**: 315–324.
- Hallauer AR, Miranda JB (1988). *Quantitative Genetics in Maize Breeding*, 2nd edn. Iowa State University Press: Ames, IA.
- Hayman BI (1954a). The analysis of variance of diallel tables. *Biometrics* **10**: 235–244.
- Hayman BI (1954b). The theory and analysis of diallel crosses. I. *Genetics* **39**: 789–809.
- Hill J, Wagoire WW, Ortiz R, Stølen O (2001). Analysis of a combined F₁/F₂ diallel cross in wheat. *Theor Appl Genet* **102**: 1076–1081.
- Jannink JL, Jansen R (2001). Mapping epistatic quantitative trait loci with one-dimensional genome searches. *Genetics* **157**: 445–454.
- Jannink JL, Wu XL (2003). Estimating allelic number and identity in state of QTLs in interconnected families. *Genet Res* **81**: 133–144.
- Jansen RC (1993). Interval mapping of multiple quantitative trait loci. *Genetics* **135**: 205–211.
- Jansen RC, Jannink JL, Beavis WD (2003). Mapping quantitative trait loci in plant breeding populations: use of parental haplotype sharing. *Crop Sci* **43**: 829–834.
- Jansen RC, Stam P (1994). High-resolution of quantitative traits into multiple loci via interval mapping. *Genetics* **136**: 1447–1455.
- Jinks JL (1956). The F₂ and backcross generations from a set of diallel crosses. *Heredity* **10**: 1–30.
- Johnson GR, King JN (1998). Analysis of half diallel mating designs. *Silvae Genet* **47**: 74–79.
- Kamoshita A, Wade LJ, Ali ML, Pathan MS, Zhang J, Sarkarung S *et al* (2002). Mapping QTLs for root morphology of a rice population adapted to rainfed lowland conditions. *Theor Appl Genet* **104**: 880–893.
- Kao CH, Zeng ZB, Teasdale RD (1999). Multiple interval mapping for quantitative trait loci. *Genetics* **152**: 1203–1216.
- Kempthorne O, Curnow RN (1961). The partial diallel cross. *Biometrics* **17**: 229–250.
- Khatkar MS, Thomson PC, Tammen I, Raadsma HW (2004). Quantitative trait loci mapping in dairy cattle: review and meta-analysis. *Genet Sel Evol* **36**: 163–190.
- Kolb FL, Bai GH, Muehlbauer GJ, Anderson JA, Smith KP, Fedak G (2001). Host plant resistance genes for fusarium head blight: mapping and manipulation with molecular markers. *Crop Sci* **41**: 611–619.
- Liu YF, Zeng ZB (2000). A general mixture model approach for mapping quantitative trait loci from diverse cross designs involving multiple inbred lines. *Genet Res* **75**: 345–355.
- Lynch M, Walsh B (1998). *Genetics and Analysis of Quantitative Traits*. Sinauer: Sunderland.
- Mather K (1941). Variation and selection of polygenic characters. *J Genet* **41**: 159–193.
- McIntyre LM, Coffman CJ, Doerge RW (2001). Detection and localization of a single binary trait locus in experimental populations. *Genet Res* **78**: 79–92.
- Muranty H (1996). Power of tests for quantitative trait loci detection using full-sib families in different schemes. *Heredity* **76**: 156–165.
- Rao SQ, Li X (2000). Strategies for genetic mapping of categorical traits. *Genetica* **109**: 183–197.
- Rebaï A, Goffinet B (1993). Power of tests for QTL detection using replicated progenies derived from a diallel cross. *Theor Appl Genet* **86**: 1014–1022.
- Rebaï A, Goffinet B (2000). More about quantitative trait locus mapping with diallel designs. *Genet Res* **75**: 243–247.
- Satagopan JM, Yandell YS, Newton MA, Osborn TC (1996). A Bayesian approach to detect quantitative trait loci using Markov chain Monte Carlo. *Genetics* **144**: 805–816.
- Sawkins MC, Farmer AD, Hoisington D, Sullivan J, Tolopko A, Jiang Z *et al* (2004). Comparative Map and Trait Viewer (CMTV): an integrated bioinformatic tool to construct consensus maps and compare QTL and functional genomics data across genomes and experiments. *Plant Mol Biol* **56**: 465–480.
- Searle SR, Casella G, McCulloch CE (1992). *Variance Components*. John Wiley & Sons: New York.
- Simko I (2002). Comparative analysis of quantitative trait loci for foliage resistance to *Phytophthora infestans* in tuber-bearing Solanum species. *Am J Potato Res* **79**: 125–132.
- Soller M, Genizi A (1978). The efficiency of experimental designs for the detection of linkage between a marker locus and a locus affecting a quantitative trait in segregating populations. *Biometrics* **34**: 47–55.
- Sprague GF, Tatum LA (1942). General versus specific combining ability in crosses of corn. *J Am Soc Agron* **34**: 923–932.
- Toojinda T, Siangliw M, Tragoonrun S, Vanavichit A (2003). Molecular genetics of submergence tolerance in rice: QTL analysis of key traits. *Ann Bot* **91**: 243–253.

- Tuberosa R, Salvi S, Sanguineti MC, Maccaferri M, Giuliani S, Landi P (2003). Searching for quantitative trait loci controlling root traits in maize: a critical appraisal. *Plant Soil* **255**: 35–54.
- Viana JMS, Cruz CD, Cardoso AA (1999). Theory and analysis of partial diallel crosses. *Genet Mol Biol* **22**: 591–599.
- Welz HG, Geiger HH (2000). Genes for resistance to northern corn leaf blight in diverse maize populations. *Plant Breed* **119**: 1–14.
- Wu XL, Jannink JL (2004). Optimal sampling of a population to determine QTL location, variance, and allelic number. *Theor Appl Genet* **108**: 1434–1442.
- Xie CQ, Gessler DDG, Xu SZ (1998). Combining different line crosses for mapping quantitative trait loci using the identical by descent-based variance component method. *Genetics* **149**: 1139–1146.
- Xu S (1996). Mapping quantitative trait loci using four-way crosses. *Genet Res* **68**: 175–181.
- Xu S (1998). Mapping quantitative trait loci using multiple families of line crosses. *Genetics* **148**: 517–524.
- Yi NJ, Xu SZ (2000). Bayesian mapping of quantitative trait loci for complex binary traits. *Genetics* **155**: 1391–1403.
- Yi NJ, Xu SZ (2002a). Linkage analysis of quantitative trait loci in multiple line crosses. *Genetica* **114**: 217–230.
- Yi NJ, Xu SZ (2002b). Mapping quantitative trait loci with epistatic effects. *Genet Res* **79**: 185–198.
- Yi NJ, Xu SZ, George V, Allison DB (2004). Mapping multiple quantitative trait loci for ordinal traits. *Behav Genet* **34**: 3–15.
- Zeng ZB (1994). Precision mapping of quantitative trait loci. *Genetics* **136**: 1457–1468.
- Zhu J, Weir BS (1996). Mixed model approaches for diallel analysis based on a bio-model. *Genet Res* **68**: 233–240.