Theor Appl Genet (1998) 96: 551–557

M. Pérez-Enciso

Sequential bulked typing: a rapid approach for detecting QTLs

Received: 15 September 1997 / Accepted: 6 October 1997

Abstract A strategy of DNA pooling aimed at identifying markers linked to quantitative trait loci (QTLs), 'Sequential Bulked Typing' (SBT), is presented. The method proposed consists in pooling DNA from consecutive pairs of individuals ranked phenotypically, i.e., pools are formed with individuals ranked (1st, 2nd), (3rd, 4th), … ,(N-1st, Nth). The N/2 pools are subsequently amplified using the polymerase chain reaction (PCR). If the whole population is typed the number of PCRs per marker is halved with respect to individual typing (IT). But if this strategy is combined with selective genotyping of extreme individuals savings can be further increased. Two extreme cases are considered: in the first one (SBT^0), it is assumed that only presence or absence of a given allele can be ascertained in a pool; in the second one (SBT^1), it is further assumed that differences between allele band intensities can be distinguished. The theory to estimate by maximum likelihood the QTL effect and its position with respect to flanking markers is presented. The behaviour of IT and SBT was studied using stochastic computer simulation in backcross and F_2 populations. Three percentages of subpattern distinction (0, 50 and 100%) two population sizes ($n = 1200$ and 600) and two QTL effects (a = 0.1 and 0.25 standard deviations) were considered. SBT^1 had the same power as individual genotyping at half the genotyping costs in all situations studied. Accuracy of QTL location is not increased with a dense number of markers, as opposed to individual typing. As a result DNA pooling is not useful for accurate location of the QTL but rather to pick up genome regions containing QTLs of at least moderate effect. The theory developed provides the general theoretical framework to deal with any DNA pooling strategy aimed at detecting QTLs.

Key words DNA pooling · QTL mapping · Selective genotyping

Communicated by H. C. Becker

M. Pérez-Enciso
Centre UdL-IRTA, Rovira Roure 177, 25198 Lleida, Spain
Fax: (34) 73 23 83 01
E-mail: enciso@etsea.udl.es

Introduction

A large number of DNA polymorphisms have been uncovered by recent advances in molecular techniques. Most of these polymorphisms are thought to be neutral, but they do provide the main tools by which to locate genes affecting traits of economic interest, which are usually quantitative traits. Since the pioneering work of Patterson et al. (1988) a large number of experiments have located chromosome regions in plants that contain genes affecting quantitative traits, the so-called 'quantitative trait loci' (QTLs). Given the amount of work involved in these types of experiments, accurate methods for selecting chromosome regions of 'potential interest' in a reasonable time are very much needed.

Two main strategies aimed at achieving this purpose have been envisaged: selective genotyping (Lander and Botstein 1989) and pooling DNA from several individuals (Giovannoni et al. 1991; Michelmore et al. 1991). By pooling DNA from resistant versus susceptible individuals several authors have been able to locate genes responsible for disease resistance in plants (Michelmore et al. 1991; Poulsen et al. 1995; Villar et al. 1996; Tartarini 1996). In a trait showing continuous variation, a generalization is pooling the individuals with the highest and lowest performance (Darvasi and Soller 1994) and looking for allelic bands specific to each tail. Yet, the difference between alternative genotypes has to be very large, at least more than two standard deviations. Otherwise, the probability of finding both genotypes is very high along all of the

phenotypic range. The critical difference in quantitative traits with respect to dichotomous traits is that in the former the phenotypic distributions of alternative genotypes overlap, and thus pooling strategies are much less powerful.

Darvasi and Soller (1994) proposed that if actual frequencies of the marker alleles can be inferred in the amplified pool, bulked analysis can be directly applied to quantitative traits. They suggested that allele frequencies can be estimated from the quantitative densitometry of allelic bands (Pacek et al. 1993). The general performance of this method depends critically on the error associated to quantifying allelic frequencies by densitometry, which poses serious technical difficulties. First, allele frequencies in the amplified product need not necessarily correspond to those in the original mixture due to differential allelic amplification. For instance, Taylor et al. (1994) found that only 39 out of 226 microsatellites studied were suitable for pool amplification in a cross between CAST/Ei and MEV mice strains. Second, optimizing the bulk polymerase chain reaction (PCR) protocol for a large number of markers can be time consuming and subject to irreproducible errors. The experimental error has to be assessed in advance for each marker and allelic frequency, which is time-consuming. Because of the technical skills needed, results may not be comparable between laboratories.

In this work I present an alternative pooling strategy, called 'sequential bulked typing' (SBT), aimed at reducing the number of PCRs needed to detect QTLs but which does not assume that pool allele frequencies are accurately estimated. The theory has been developed for backcross (BC) and F_2 crosses between inbred lines, but the same principles can be applied to other schemes like dihaploid lines or within-family designs. The performance of the method will be compared with individual genotyping (IT) using classical interval mapping (Lander and Botstein 1989).

Theory

The method

The method proposed consists of pooling DNA from consecutive pairs of individuals ranked phenotypically, i.e., pools are formed with individuals ranked (1st, 2nd), (3rd, 4th), … , (N-1st, Nth). The N/2 pools are subsequently amplified using PCR. If the whole population is typed, the number of PCRs per marker is halved with respect to individual typing, but if this strategy is combined with the selective genotyping of extreme individuals, savings can be further increased. I have only considered pools formed by consecutive pairs of individuals, but any number of individuals can be pooled together and pools can in principle be made up of different number of individuals. In general, however, the informativeness of pool amplification will decrease rapidly as the number of individuals per pool increase.

Consider two completely inbred lines with alternative marker alleles *M* and *m* fixed in each line. All possible electrophoresis bulk patterns from a $MM \times Mm$ backcross and a $Mm \times Mm$ F_2 cross are shown in Fig. 1. Pattern A (C) results from pools with both individuals having genotypes *MM* (*mm*), whereas in pattern B the two bands, corresponding to both alleles, are observed. In the least favourable case, only presence or absence of a given band can be detected, and thus only patterns A, B–, and C are identified. I will denote this case by SBT^o. A more optimistic view is that, in addition to presence versus absence, differences in band intensities between pools can be identified. In this case a number of 'B' subpatterns are distinguished, two in a BC and three in a F_2 cross (Fig. 1). Note that the possibility of distinguishing subpatterns requires fewer assumptions than estimating allele frequencies, as subpattern identification will not be affected by differential amplification provided that the same allele tends to be amplified predominantly. Besides, any, previous labour to assess pool amplification error is minimum because subpatterns are assigned simply by comparing the band intensities corresponding to the different bulks. The effect of subpattern misassignment is studied below. I will denote by SBT^1 the case where subpatterns B_1, B_2 or B_3 can be assigned in all 'B' bulks. In general, SBT^X indicates that subpatterns can be assigned in x% of the bulks. If flanking markers are analysed, up to nine (BC) and 25 (F2) patterns can be distinguished. Note that in analysing DNA pools we deal with 'patterns' rather than with genotypes as in individual typing.

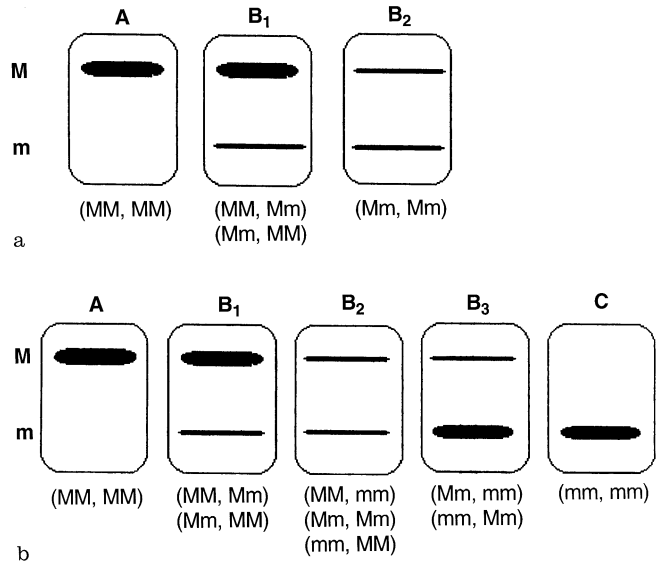

Fig. 1a, b Electrophoresis patterns that can be distinguished in a $Mm \times MM$ backcross (**a**) and in a $Mm \times Mm$ F_2 cross (**b**) when a DNA pool of two individuals is amplified. The possible individual genotype combinations are listed *below* each diagram

The likelihoods

I will follow the usual convention of modelling the phenotypic variation of a quantitative trait by a normal distribution. Consider a trait y which follows a normal distribution conditionally on each QTL genotype, g. The marginal phenotypic distribution is thus a mixture of normals, the number of components being equal to the number of possible QTL genotypes. Further assume that the trait is standardized such that the variance within QTL genotype is one. Thus $p(y|g_j) = N(\mu_j,1)$, where g_1, g_2, and g_3 correspond to QTL genotypes *GG*, *Gg*, and *gg*, respectively; and $\mu_1 = a$, $\mu_2 = d$, $\mu_3 = -a$, following classical notation (Falconer and Mackay 1996). In the backcross, d cannot be estimated and I will assume instead $\mu_1 = a/2$ and $\mu_2 = -a/2$.

In classical interval mapping and individual typing genetic markers are scored throughout the genome and the likelihood is computed along the intervals delimited between adjacent markers. The likelihood can be written as

$$L_{IT} \propto p(y, M) = p(y|M)\, p(M)$$

$$= \prod_{i=1}^{N} \left[\sum_{j=1}^{n_g} p\,(y_i \mid g_j)\, p\,(g_j \mid M_i) \right] p(M_i) \quad (1)$$

(Lander and Botstein 1989) where M is the marker genotype, N is the number of individuals, n_g is the number of possible genotypes, two in a backcross or three in a F_2 cross; $p(g|M)$ is the probability of the QTL genotype given marker information, either single or flanking markers, which can be easily obtained from transmission rules given the distances between loci. (Note: I will use $p(\cdot)$ to generally refer to probability or density functions). I will assume that genetic distances between markers are known with reasonable accuracy such that they need not be estimated jointly with the QTL effect and position.

The likelihood with SBT is given by

$$L_{SBT} \propto p(y, patt) = \prod_{i=1}^{N/2} p\,(patt_i | y_{i1}, y_{i2})\, p(y_{i1})\, p(y_{i2}) \quad (2)$$

where $patt_i$ is the electrophoresis pattern of ith bulk at the interval marker considered, and subscripts i1 and i2 refers to first and second individual of ith bulk, respectively. Note that in Eq. 2 the product is over N/2 rather than over N as in Eq. 1. Above, $p(y) = \sum_{j=1}^{n_g} p\,(y|g_j)\, p\,(g_j)$ is a mixture of normals with p(g) equal to 0.5 for $g = g_1, g_2$ in a BC and 0.25, 0.50 and 0.25 for $g = g_1, g_2$, and g_3, respectively, in a F_2. The term $p(patt|y_1, y_2)$ is computed as follows. In a backcross,

$$p(patt = \text{'A'} | y_1, y_2) = p(M = \text{'MM'} | y_1) \cdot p(M = \text{'MM'} | y_2) \quad (3a)$$

and, in the case of SBT^0,

$$\begin{aligned} p(patt = \text{'B'} | y_1, y_2) &= p(M = \text{'MM'} | y_1) \cdot p(M = \text{'M m'} | y_2) \\ &\quad + p(M = \text{'Mm'} | y_1) \cdot p(M = \text{'Mm'} | y_2) \\ &\quad + p(M = \text{'Mm'} | y_1) \cdot p(M = \text{'MM'} | y_2) \\ &= 1 - p(patt = \text{'A'} | y_1, y_2) \end{aligned} \quad (3b)$$

Similar expressions can be derived for SBT^1 and F_2 crosses. Above, p(M|y) is the probability of an individual having a given marker genotype conditional on its phenotype. If the marker is unlinked to any gene affecting the trait, this is simply the frequency of the marker genotype in the population. Otherwise, the marker genotype and the phenotypic values will be correlated. In general,

$$p(M|y) = \sum_{j=1}^{n_g} p(M | g_j)\, p(g_j | y) \quad (4)$$

where $p(M|g)$ is the probability of the marker's genotype given the QTL genotype, which can be obtained from rules identical to those of p(g|M) in Eq. 1, and p(g|y) is the probability of an individual having a given QTL genotype conditional on its phenotype. In a backcross,

$$p(g = \text{'GG'} | y) = \phi(y - a/2) / [\phi(y - a/2) + \phi(y + a/2)] \quad (5a)$$

$$p(g = \text{'Gg'} | y) = \phi(y + a/2) / [\phi(y - a/2) + \phi(y + a/2)] \quad (5b)$$

where $\phi(\cdot)$ is the standard normal density function. Extension to flanking markers is straightforward. The equivalent equations to 5a and 5b for a F_2 cross are

$$p(g = \text{'GG'} | y) = \phi(y - a) / [\phi(y - a) + 2\,\phi(y - d) + \phi(y + a)] \quad (6a)$$

$$p(g = \text{'Gg'} | y) = 2\,\phi(y - d) / [\phi(y - a) + 2\,\phi(y - d) + \phi(y + a)] \quad (6b)$$

$$p(g = \text{'gg'} | y) = \phi(y + a) / [\phi(y - a) + 2\,\phi(y - d) + \phi(y + a)] \quad (6c)$$

since there are three QTL genotypes with expected frequencies 1:2:1.

If only individuals with phenotype below $y = z_1$ and above $y = z_2$ are genotyped (selective genotyping), p(y|g) in Eq. 1 is a truncated normal distribution with a density function given by

$$p(y | g_i) = \phi(y - \mu_i) / [\Phi(z_1 - \mu_i) + \Phi(-z_2 + \mu_i)] \quad (7)$$

(Darvasi and Soller 1992). Selective genotyping can be combined with SBT by genotyping only extreme pools. In this case, truncation is taken into account by specifying truncation points in p(y) (Eq. 2).

Simulation study

The behaviour of IT and SBT was explored using stochastic computer simulation in backcross and F_2 populations. Different percentages of subpattern distinction, SBT^1, $SBT^{0.5}$, and SBT^0, were considered. Two population sizes (n = 1200 and 600) and two QTL effects (a = 0.1 and 0.25 SD) were considered. In the F_2, both complete additivity (d = 0) and complete dominance (d = a) were simulated but as the results were similar only the results with complete dominance action are shown. Besides full genotyping, selective typing of 10%, 20%, and 40% extreme individuals, half in each tail, was considered. Two situations were studied, either that the marker studied is analysed individually (e.g., it has not been mapped), or that a number of markers are analysed jointly. In the former case, it was assumed that the QTL was located exactly at the marker. In the latter case, the QTL was simulated at position 25 cM of a 100-cM chromosome. Markers were evenly spaced along the chromosome every 50, 10, or 1 cM; i.e., 3, 11, or 101 markers, respectively, were genotyped. Between 300 and 500 simulation replicates were run for each case. Maximum likelihood estimates of a and d were obtained by means of the Newton-Raphson algorithm, which provides directly the observed information matrix and, consequently, an approximate error of the estimates. Convergence was very fast, usually in less than five rounds. The presence of linkage between a marker and a QTL was tested by means of a *t*-test using the asymptotic distribution of maximum likelihood estimates. The type-I error probability was set to 0.05.

Results and discussion

All of the methods used here provided unbiased estimates of additive and dominance parameters in all of the situations studied (results not shown), as was expected from the asymptotic properties of maximum likelihood estimates, although the standard errors of the estimates were different for each method. The performance in terms of power of IT, SBT^1, $SBT^{0.5}$, and SBT^0 for the combination of parameters described is presented in Fig. 2 for a BC design when the marker was analysed individually. The power of IT and SBT^1 was almost identical in all situations studied. For instance, for n = 1200, a = 0.1, 40% genotyped, power was 0.484 and 0.482 for IT and SBT^1, respectively. Thus, a single line for both methods was drawn in the figures. For the remaining SBT strategies, the relative advantage of SBT over IT increased with population size and QTL effect (Fig. 2b). This is precisely the situation where pooling will be more useful, as it is

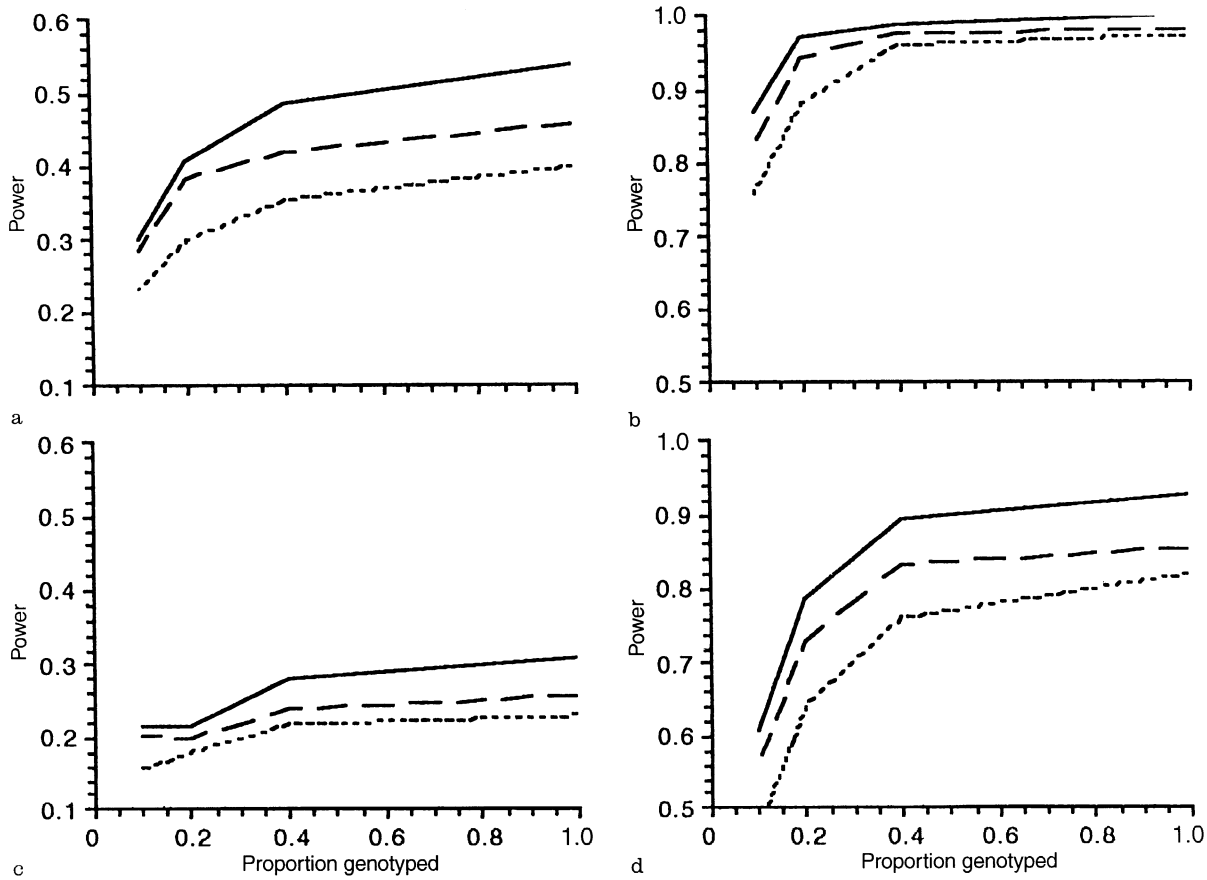

Fig. 2a–d Power attained in a backcross with individual typing or SBT^1 (*continuous line*), $SBT^{0.5}$ (*dashed line*), and SBT^0 (*dotted line*) for different population sizes (n) and QTL additive effects (a): **a** $n = 1200$, a = 0.1 SD; **b** $n = 1200$, a = 0.25 SD; **c** $n = 600$, a = 0.1 SD; (**d**) $n = 600$, a = 0.25 SD

more useful to be able to limit genotyping work in large populations than in moderate- or small-sized ones. In addition, DNA pooling will be primarily used as an initial screen to detect large-effect QTLs and discard 'uninteresting' genome regions, whereas the identification of small-effect genes would require individual genotyping. At moderate QTL effects and smaller population sizes, the possibility of identifying B subpatterns becomes more critical in order to limit the number PCRs. For instance, for n = 600 and a = 0.25 SD $SBT^{0.5}$ is more effective than IT given an equal number of PCRs, but SBT^0 is more effective than IT only for very extreme selective genotyping (Fig. 2d). For a = 0.1, DNA pooling is cost-effective only if the percentage of bulks with subpatterns identified is at least 50% (Fig. 1a, b).

The performance of bulked typing is inversely related to the number of possible genotypic combinations within B-pattern bulks. Thus, a priori one should expect SBT to be more useful in BC than in F_2 crosses because the B pattern is less informative in the latter cross (Fig. 1). However, the difference between extreme phenotypes is approximately doubled in the F_2, with the result that F_2 designs are more powerful than BC given an equal number of individuals (Fig. 3 vs. Fig. 2a, c). In fact, for a = 0.25 SD power was 1 or nearly 1 for all of the methods and both population sizes considered, except for SBT^0 (results not presented). As in BC, the power in estimating the additive effect with SBT^1 was identical to that with IT.

The above considerations make the possibility of distinguishing subpatterns, at least in part of the bulks, an important issue. As argued before, this is technically much more feasible than estimating allele frequencies from DNA pools of several individuals and it should not be affected by differential amplification. A matter of concern is the robustness of SBT^1, as there always exists a risk of misassignment within B subpatterns. The impact of misassignment was studied by simulating that subpatterns were assigned at random in a given percentage of bulks. The results are presented in Table 1 for a BC and a F_2 cross. As expected, misassignment caused a downward bias in the estimates and lowered the power; however, these effects were negligible unless the proportion of errors exceeded 10% or 20%. Therefore, a sensible strategy would be to compare the estimates obtained with SBT^0 and SBT^1: very different values would be a symptom of misassignment. Otherwise, SBT^1 estimates should be preferred, as they have smaller errors.

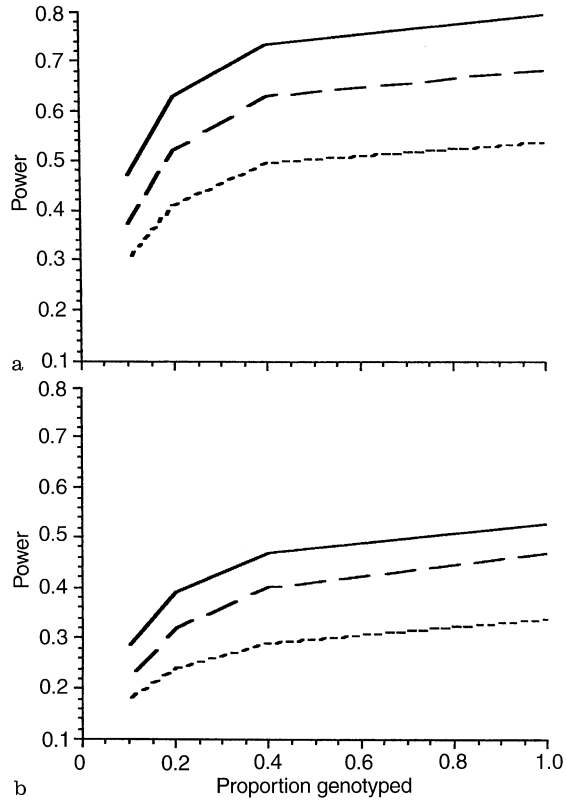

Fig. 3 Power attained for the additive effect in a F_2 cross with individual typing or SBT^1 (*continuous line*), $SBT^{0.5}$ (*dashed line*), and SBT^0 (*dotted line*) for different population sizes: **a** $n = 1200$, **b** $n = 600$. In all cases complete dominance was assumed (a = d) as well as a QTL additive effect a = 0.1 SD

Table 1 Effect of errors in assigning B subpatterns on the QTL effect estimates (â) and on power (β) for two designs: a backcross (600 individuals, QTL effect = 0.25 SD) and a F_2 cross (1200 individuals, QTL effect = 0.10 SD) populations. The proportion genotyped was the 40% extreme

Percentage	Backcross		F_2	
errors	â	β	â	β
0	0.250	0.91	0.100	0.78
1	0.249	0.90	0.102	0.77
5	0.247	0.86	0.099	0.76
10	0.247	0.86	0.096	0.70
20	0.225	0.78	0.090	0.68
30	0.216	0.78	0.084	0.64
50	0.183	0.64	0.071	0.54

The theory developed here allows us to estimate dominance as well as additive effects in an appropriate design, i.e., in a F_2 cross. Yet most of our information on the dominance effect comes from the heterozygous genotypes, and these are pooled in the less-informative pattern, B. It is not possible to distinguish between *Mm*, *Mm* pools, which contribute mostly to the estimate of the dominance effect, and *MM*, *mm* pools, which contribute to the estimate of the additive effect. This is illustrated in Table 2, where it can be seen that power with SBT was not affected by the possibility of distinguishing subpatterns. Power for the dominance effect and the additive effect was similar for SBT^0, as can be seen with comparisons between Table 2 and Fig. 3. All in all, DNA pooling is a strategy that is more efficient in estimating additive effect than dominance effect.

Table 2 Power for the dominance effect in a F_2 cross using individual (IT) and sequential bulked typing (SBT) for population size *n* and selective genotyping of the q% extreme progeny; complete dominance and a QTL effect of 0.1 SD were assumed

n	q (%)	IT	SBT^1	SBT^0
1200	100	0.79	0.64	0.63
	40	0.78	0.63	0.61
600	100	0.54	0.43	0.42
	40	0.48	0.29	0.25

A second set of simulations were run in order to study the effect of marker spacing, under the assumption that there is already an existing genetic map for the species under consideration and that distances between markers are known or can be estimated with reasonable accuracy. Results for different marker spacings and either full genotyping or 40% selective genotyping are presented in Table 3 for a BC design and a QTL effect a = 0.25 SD. The most important aspect to notice is that marker spacing and selective genotyping affects the estimates of the QTL effect and its position in a different way. Estimates of QTL effect are unbiased irrespective of marker spacing and selective genotyping, although the associated error is inversely related to the number of individuals and, to a lesser extent, by the number of markers. In contrast, estimates of QTL position are much more sensitive to selective genotyping. Whereas most of the information on the QTL effect is conveyed by individuals departing from the mean, QTL position accuracy depends primarily on the number of recombinants between markers, which is proportional to the number of individuals, irrespective of their phenotype. With fewer individuals genotyped, power decreases and location estimates tend to be scattered more or less at random along the chromosome. On average, the mean is 'regressed' towards the middle of the chromosome (here 50 cM). This effect is more noticeable with SBT than with individual typing (Table 3). An a priori surprising result is that a dense marker spacing even deteriorated location estimates. The explanation is that in pool analyses, the larger the number of markers, the larger the number of marker phase genotypes that have to be inferred, e.g., whether *MMNN*, *mmnn* or *MMnn*, *mmNN* where *M* and *N* are flanking markers—genotype combinations are more likely. Consequently, the chances of errors increase without the reward of a more precise location estimate, as this precision is only moderately increased with very

Table 3 Effect of selective genotyping of the q% extreme progeny and of marker spacing on the QTL additive effect estimate (â) and its position estimate (δ) in a backcross design with 600 individuals, according to different typing strategies: IT, individual typing; SBT^x, sequential bulked typing when x percent of the B subpatterns can be assigned. The simulated QTL effect and position were 0.25 SD and 25 cM, respectively. Empirical standard errors in parentheses

q (%)	Number of markers	â			δ		
		IT	SBT^1	SBT^0	IT	SBT^1	SBT^0
100	3	0.258 (0.107)	0.254 (0.156)	0.228 (0.112)	0.307 (0.233)	0.318 (0.304)	0.444 (0.273)
	11	0.273 (0.081)	0.254 (0.156)	0.168 (0.111)	0.288 (0.184)	0.360 (0.311)	0.502 (0.304)
	101	0.270 (0.099)	0.226 (0.149)	0.145 (0.128)	0.269 (0.142)	0.357 (0.325)	0.478 (0.327)
40	3	0.263 (0.122)	0.265 (0.190)	0.235 (0.154)	0.386 (0.273)	0.406 (0.339)	0.450 (0.270)
	11	0.255 (0.118)	0.235 (0.152)	0.164 (0.141)	0.310 (0.202)	0.431 (0.280)	0.509 (0.299)
	101	0.275 (0.107)	0.258 (0.186)	0.171 (0.137)	0.291 (0.175)	0.372 (0.310)	0.479 (0.323)

dense mapping. For instance, with individual typing a 33-fold increase in genotyping work only reduces the standard error of position estimate by 40%. Darvasi et al. (1993) showed that neither power nor accuracy increased in IT with a marker spacing below 20 cM. This phenomenon is even more remarkable with DNA pooling. This has important practical consequences, as it suggests that typing markers spaced closer than 50 cM is not worth the work if SBT, or any other DNA pooling strategy, is employed. Power for QTL detection was not increased either (results not presented). Note, however, that the QTL effect should be less biased than its position and thus the (economic) benefit of an eventual identification of the gene can be balanced in advance against the total cost of fine-mapping the relevant region.

Finally, it should be mentioned that selective genotyping and DNA pooling strategies are optimum for a single-trait analysis. In practice, experiments are carried out bearing in mind that even if a trait is of particular relevance, other traits will be recorded and analysed. A loss of power is expected in the analyses of these secondary traits with SBT. Some simulations were run in order to explore this issue. As expected, the QTL effect errors were larger but, interestingly, estimates were still unbiased (results not shown). The worst case is when a second trait is completely uncorrelated with the trait on which ranking is based. For instance in a BC, $n = 600$, $q = 100\%$, the power for an uncorrelated trait QTL with SBT^1 (SBT^0) is 0.67 (0.54), and 0.20 (0.16), for $a = 0.25$ and 0.10 SD, respectively. These figures are about 25–30% lower than for the main trait (Fig. 2c, d). It seems that SBT will enable the largest effect QTLs for a number of traits to be pinpointed simultaneously if these are not completely uncorrelated.

Conclusion

A large part of the time and costs involved in QTL detection experiments is expended in genotyping. DNA pooling is a powerful strategy by which to reduce the genotyping effort for mapping purposes and for the identification of markers linked to dichotomous traits. Yet, its relevance on the analysis of quantitative traits is less well studied. The theory developed in this work provides the general theoretical framework to deal with any DNA pooling strategy aimed at detecting markers linked to quantitative traits. It allows us to utilize information from flanking markers and obtain estimates for both the additive and dominant effects. Besides the cases studied, several other situations, like pools consisting of more than two individuals or more than one QTL models, can be accommodated. In practice, the optimum strategy may combine individual and pool typing. Again, this poses no methodological problem. Two main conclusions can be drawn from this study.

First, if the main purpose of the experiment is to estimate the QTL effect associated with a given set of markers, then SBT has the same power as individual genotyping at half the genotyping costs, provided that the subpatterns can be distinguished. Comparing the estimates obtained with SBT^0 and SBT^1 will provide insight into the extent of pattern misassignment. The smaller the percentage of bulks for which the subpatterns are correctly identified, the larger the minimum QTL effect that is to be detected with SBT.

Second, DNA pooling is not useful for accurate location of the QTL. A reasonable strategy seems rather to use SBT with widely spaced markers, which would allow us to pick up the 'interesting' genome regions, those containing QTLs of at least moderate effect. In a second stage, individual typing with a large number of markers within each promising region might be unavoidable if the gene is to be located with precision. But, alternatively, as our understanding of the physiology of 'complex' quantitative traits is deepened and as genetic maps become more densely populated with codifying sequences, the candidate gene approach will become more attractive. SBT would allow us to pinpoint a series of regions hopefully containing a (small)

number of candidate genes. Their direct effects will be checked in a second stage. SBT provides a compromise between individual typing for a large number of markers, which is highly onerous, and DNA pooling of both extreme tails (Darvasi and Soller 1994), which is likely to be very imprecise.

Acknowledgements I thank for useful discussions C. Haley, C. Benito, J. Gavora, I. Jackson, and J. Medrano. This work was supported by CICYT grant no. AGF96-2510, by Spanish–British "Acción Integrada" no. HB95-70, and by an OECD fellowship.

References

Darvasi A, Soller M (1992) Selective genotyping for determination of linkage between a marker locus and a quantitative trait locus. Theor Appl Genet 85:353–359

Darvasi A, Soller M (1994) Selective DNA pooling for determination of linkage between a molecular marker and a quantitative trait locus. Genetics 138:1365–1373

Darvasi A, Weinreb A, Minke V, Weller JI, Soller M (1993) Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. Genetics 134: 943–951

Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics, 4th edn. Longman, Essex

Giovannoni JJ, Wing RA, Ganal MW, Tanksley SD, (1991) Isolation of molecular markers from specific chromosomal intervals using DNA pools from existing mapping populations. Nucleic Acids Res 19:6553–6558

Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121: 185–199.

Michelmore RW, Paran I, Kesseli RV (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. Proc Natl Acad Sci USA 88:9828–9832

Pacek P, Sajantila A, Syvanen AC (1993) Determination of allele frequencies at loci with length polymorphism by quantitative analysis of DNA amplified from pooled samples. PCR Methods Applic 2:313–317

Patterson AH, Lander ES, Hewitt JD, Petersson S, Lincoln SE, Tanksley SD (1988) Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335:721–726

Poulsen DME, Henry RJ, Johnston RP, Irwin JAG, Rees RG (1995) The use of bulk segregant analysis to identify a RAPD marker linked to leaf rust resistance in barley. Theor Appl Genet 91:270–273

Tartarini S (1996) RAPD markers linked to the *vf* gene for scab resistance in apple. Theor Appl Genet 92:803–810

Taylor BA, Navin A, Phillips SJ (1994) PCR–amplification of simple sequence repeat variants from pooled DAN samples for rapidly mapping new mutations in mouse. Genomics 21:626–632

Villar M, Lefebvre F, Bradshaw HD, Teissier du Cross E (1996) Molecular genetics of rust resistance in poplars (*Melampsora larici-populina* Kleb/*Populus* sp.) by bulked segregant analysis in a 2×2 factorial mating design. Genetics 143:531–536