Theor Appl Genet (1994) 88:355–361


G.-L. Wang · A. H. Paterson

# Assessment of DNA pooling strategies for mapping of QTLs

Received: 7 January 1993 / Accepted: 9 September 1993

**Abstract** The synthesis of "DNA pools" from segregating populations is an efficient strategy for identifying DNA markers closely linked to genes or genomic regions of interest. To-date, DNA pooling based solely upon phenotypic information, or "bulked segregant analysis", has been employed only in the analysis of simply-inherited traits. We have assessed the utility of phenotype-based DNA pools for "tagging" (e.g., identifying DNA markers closely-linked to) quantitative trait loci (QTLs), segregating in the presence of other such loci, and expressing phenotypes which are influenced by the environment. Theoretical estimates suggest that QTL alleles with phenotypic effects of 0.75–1.0 standard deviations (SD), or larger, should be detectable in backcross (BC), $F_2$ and recombinant inbred (RI) or doubled haploid (DH) populations of manageable size (100–200 plants/lines). However, *post hoc* analysis of three data sets, used in QTL mapping of tomato and rice, indicate that the majority of QTLs identified had allele effects of less than 0.75 SD, and thus could not be easily tagged in DNA pools. Segregation distortion can have a large effect on the allelic composition of DNA pools, necessitating the use of more individuals in the pools to minimize false positive and false negative results. In general, we suggest that use of phenotype-based DNA pools might be successful in tagging QTLs of very large effect, but is unlikely to permit comprehensive identification of the majority of QTLs affecting a complex trait. DNA pools constructed from a priori information should, however, be useful in identifying new DNA markers for regions of the genome known to contain QTLs.

**Key words** DNA pools · Bulk segregant analysis · RFLPs · RAPDs · QTL mapping · Genetic mapping

Communicated by A. L. Kahler

G.-L. Wang · A. H. Paterson (✉)
Department of Soil and Crop Sciences, Texas A & M University, College Station, TX 77843, USA

## Introduction

DNA markers have become an indispensable tool in plant genetics, and are of particular value in gene mapping and marker-assisted selection. However, the large numbers of individuals required for gene mapping experiments hinder the utilization of these tools. Mapping of simply-inherited traits can be streamlined by the synthesis of "DNA pools" from individuals sharing a common phenotype (Michelmore et al. 1991) or genotype at an interval delineated by two RFLPs (Giovannoni et al. 1991). The principle of DNA pooling is the grouping together of informative individuals so that a particular genomic region carrying a "selectable marker" (morphological mutation, or DNA marker) of interest can be studied in a randomized genetic background of unlinked loci. Trait mapping using DNA pools is amenable to RFLPs or other DNA markers, but can be streamlined through the use of RAPDs (Williams et al. 1990) which minimize needs for genomic DNA, Southern blotting and radioisotopes. Using DNA pooling methods, Michelmore et al. (1991) identified three RAPD markers in lettuce linked to a gene for resistance to downy mildew. Giovannoni et al. (1991) identified two RAPD markers tightly linked to loci affecting ripening and jointless stem in tomato.

While simply-inherited loci altering morphology or disease reaction can be of great importance, measures of agricultural productivity, e.g., yield, quality and horizontal resistance to various stresses, are usually influenced by many genes. The restriction fragment length polymorphism (RFLP) technique has been successfully used to map individual genetic factors or quantitative trait loci (QTLs) associated with these traits in several crops (Nienhuis et al. 1987; Paterson et al. 1988, 1990, 1991; Tanksley and Hewitt 1988; Weller et al. 1988; Grant et al. 1989; Keim et al. 1990; Stuber et al. 1992; Wang et al. 1993). While such experiments are accurate and repeatable, they require large populations, and are labor-intensive and tedious.

Our objective in the present work was to determine whether a phenotype influenced by multiple genetic loci, together with the environment, could serve as a "selectable marker" for the construction of DNA pools which would permit enrichment for DNA markers near genes influencing the target phenotype. It is well-documented that the most extreme individuals in a phenotypic distribution carry a disproportionately large amount of genetic information useful for mapping QTLs (Lander and Botstein 1989; Darvasi and Soller 1992). However, when a phenotype is influenced by multiple genetic loci, and environment, individuals can have extreme phenotypes due to different sets of QTLs, or due to non-genetic factors. Thus, the task is to distinguish a genetic signal against a background of environmental noise. Factors which determine the power to accomplish this task include the magnitude of the difference between "subpopulations" (carrying different QTL alleles), the extent of mixing of sub-populations, and the number of individuals studied (Mendell et al. 1991).

In this paper, we evaluate the possibility of using DNA pooling strategies for QTL mapping in $F_2$, backcross (BC), recombinant inbred (RI), and doubled haploid (DH) populations. Predictions about the likelihood of finding DNA polymorphisms linked to the target QTL in phenotype-based DNA pools are tested by *post hoc* analysis of large plant populations previously used for QTL mapping. The effects of population size, fraction of population selected, magnitude of phenotypic effect of individual QTL alleles (QTL allele effect), and effects of both dominance and deviations from Mendelian segregation ratios, are considered. General concordance of *post hoc* analyses with theoretical predictions supports the conclusion that phenotype-based DNA pools can be reliably used to tag occasional QTLs of very large effect, but may frequently fail to tag many QTLs of smaller effect which collectively explain a large portion of the genetic variation in a trait.

## Materials and methods

The work described in this paper involved two discrete components. First, using defined parameters of the normal distribution (Selby 1973), we determined expectations for cross-contamination of DNA pools under a variety of circumstances, e.g. different population sizes, different QTL effects, different degrees of segregation distortion.

To exemplify how this was done, consider a population consisting of two genotypes, genotype A and genotype B, at a particular QTL in a backcross population. We assume that the population is derived from a cross between homozygous parents, and segregates for only two alleles per locus. The two genotype groups have equal sample size and are each normally distributed with a common variance, $\sigma^2$, but different means, $\mu_A$ and $\mu_B$ (Fig. 1). The phenotypic effect of this QTL is expressed in units of standard deviations, and calculated from $[(\mu_A - \mu_B)/\sigma]/n$, where n is the number of alleles per locus which differ between genotypes A and B (one for a backcross population, two for an $F_2$, DH, or RI). For example, if a QTL allele at a particular locus has an effect of 0.75 standard deviations in a backcross population then, based on the genotype at that locus, the backcross population can be divided into two normally-distributed populations (A and B) whose means differ by 0.75 SD. If individuals are selected whose phenotypes are so extreme that they fall more than 1.0 SD outside the means of population (e.g., genotype) A, then by definition these fall more than 1.75 SD outside the mean of population B (Fig. 1). The fraction of population A which falls outside 1.0 SD in one tail can be taken directly from published tables (Selby 1973), as can the fraction of B which falls outside 1.75 SD. If populations A and B contain equal numbers of individuals, then the allele composition of the resulting DNA pool can be directly calculated from the ratio %B/(%A + %B). Adjustments must be made for heterozygous genotypes "contaminating" otherwise homozygous pools, and vice versa (e.g., backcross or $F_2$ populations), since in these cases, "contaminants" only contain one "contaminating allele". Similarly, adjustments must be made when the numbers of individuals in the A and B populations differ (e.g., $F_2$ dominant locus case, or deviation from Mendelian segregation), weighting the contribution of the populations to the overall allele composition in proportion to the expected population size.

DNA pools to be screened with DNA markers are constructed from the most extreme individuals in the overall population (e.g., A plus B). A successful result (e.g., new DNA markers diagnostic of the trait) is expected when the individuals at one extreme of the distribution are mostly of genotype A, and those at the other extreme are mostly of genotype B, at a particular QTL. We considered $\leq 5\%$ contamination of pools with the alternate allele small enough to indicate a successful result, approximately the level found by Michelmore et al. (1991) to render the rare allele undetectable in similar RAPD-PCR amplifications. Using RFLPs, a slightly higher level of contamination could probably be tolerated (Michelmore et al. 1991) – thus the data presented here represents a conservative appraisal of the feasibility of mapping QTLs in DNA pools with RFLPs. Finally, throughout this work we have defined the A and B pools as containing ten individuals since ten plants in a pool is sufficient to avoid detecting false positive markers, even with moderate deviations from Mendelian segregation (See Fig. 4 and related text). Variations in the number of plants per pool may be warranted in some circumstances, but affect the size of the polymorphic region and consequently the likelihood of a success (Michelmore et al. 1991), so we have held this number constant.

The second component of this work was to determine if the theoretical expectations are corroborated by empirical data. This was done by *post hoc* analysis of large BC (Paterson et al. 1988), $F_2$ (Paterson et al. 1991) and RI (Wang et al. 1993) populations which had been completely genotyped with RFLPs, and employed to map a large number of QTLs. The *post hoc* analysis involved sorting a population based on phenotype, and determining the allele composi-

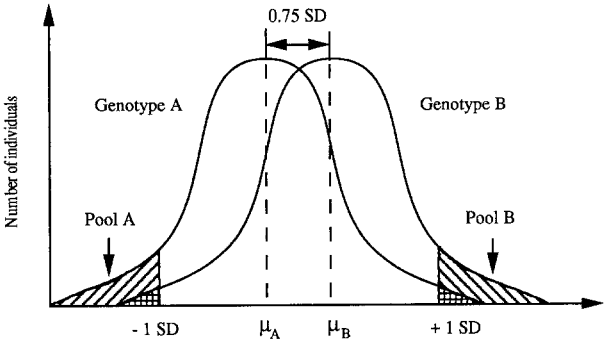



**Fig. 1** Illustration of using DNA pooling strategies for mapping a quantitative trait in a backcross population. A QTL controlling the trait has an allele effect of 0.75 standard deviations (SD). Consequently, the population can be thought of as two subpopulations differing in genotype at this QTL, and differing in mean phenotype by 0.75 SD. If one selects individuals more than 1.0 SD outside the mean of genotype A to form pool A, those falling 1.75 SD outside the mean of genotype B will also be chosen and will "contaminate" pool A. If the QTL has larger effect (> 0.75 SD) or more extreme individuals are selected (> 1.0 SD), a clean pool A and pool B can be distinguished

tion of the hypothetical DNA pools which would result from this (prepared as described above), by examining RFLP genotypes of these individuals at markers near mapped QTLs.

## Results

### Backcross population

Backcross populations are better than $F_2$s, but poorer than RI or DH populations, for tagging QTLs using phenotype-based DNA pools. At any locus in a backcross population, two subpopulations of putatively equal size occur; the genotype of one subpopulation is the same as the recurrent parent ($xx$); the genotype of the other is heterozygous ($Xx$). Using dominant markers such as RAPDs (Williams et al. 1990), only those bands unique to the $XX$ parent will be informative; thus on average only about half of the available polymorphisms will be useful. Consequently, one may have to screen twice as many markers to obtain the required polymorphisms. A contaminant individual in the $xx$ pool which is from the $Xx$ subpopulation has 50% of the $X$ allele so that this individual has only half the contamination of a homozygous contaminant, e.g., the $XX$ individual. Therefore, the contamination rate in a backcross population is much lower than that in other populations. Assuming there is no segregation distortion, a QTL with an allele effect of $\geq 0.75$ SD might be identified in a backcross population of about 100 plants (Fig. 2).

### $F_2$ population

$F_2$ populations appear to be a poor choice for tagging QTLs using DNA pools. Under other circumstances, the presence of all three possible genotypes at a locus permits one to distinguish additivity from dominance, and the presence of two informative gametes in each indvidual affords more recombinational information than might otherwise be available. However, these factors complicate the analysis of DNA pools.

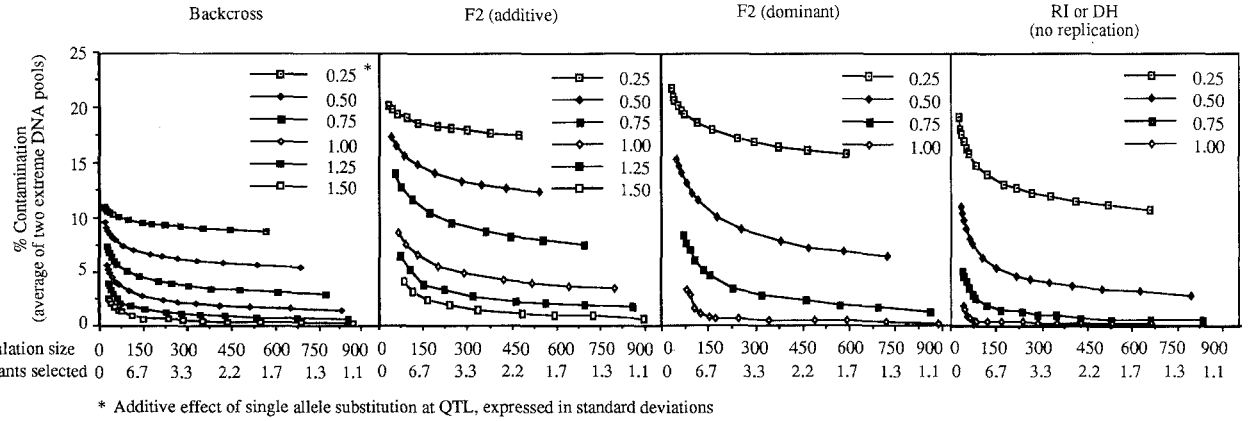
In an $F_2$ population, the utility of DNA pooling strategies is affected by gene action; that is, whether the dosage effect at a QTL is "additive" or "dominant". As described above, for strictly additive QTLs (phenotype of heterozygote = midparent mean), the relationship between contamination rate and population size at different QTL allele effects is as shown in Fig. 2. For QTL alleles with a phenotypic effect less than 1.0 SD, contamination from the heterozygote and contrasting homozygote are both appreciable. Contamination is less for QTL alleles with a phenotypic effect greater than 1.0 SD because most contaminants are from heterozygous class. Strictly additive QTLs with an allele effect of 0.75 SD can be detected in an $F_2$ population of about 3 300 plants (Fig. 2).

If a QTL has complete dominant gene action (phenotype of heterozygote = homozygote) in an $F_2$ population, the phenotypic values of $XX$ and $Xx$ individuals are the same, and only two subpopulations must be considered at an expected ratio of 3:1. As for the backcross, only half of the dominant polymorphic markers are informative, specifically those bands unique to the $XX$ parent. Contamination in the $aa$ pool is higher than that in the additive case because the dominant class ($XX$ and $Xx$) is three times larger than the recessive ($xx$) class. However, the average phenotype of the $XX + Xx$ pool differs from the average phenotype of the $xx$ pool by twice the additive effect of the allele (because of dominance); thus the separation between pools is equivalent to the recombinant inbred or doubled haploid case. A QTL with an allele effect of 0.75 SD requires, on average, a population of 140 or more plants to be detected (Fig. 2).

### RI or DH lines

RI or DH lines are a good choice for tagging QTLs using DNA pools. RI or DH populations contain two subpopulations at a locus, each homozygous, with the same genotype as one of the two parents. All DNA polymor-

**Fig. 2** Relationship between contamination rate in DNA pools and population size (or percent of population selected) in BC, $F_2$ and RI or DH lines. Note that RI or DH values are for individual plants, and assume no replication. For a discussion on the effects of replication, see text




* Additive effect of single allele substitution at QTL, expressed in standard deviations

phisms are informative (with both dominant and codominant markers). QTLs with allele effects of 0.75 SD can be detected in populations of about 40 lines (Fig. 2).

We note that RI or DH lines can be evaluated in several environments with replications. Such replication reduces the environmental "noise", thereby increasing the proportion of phenotypic variation attributed to QTLs. Consequently, by using replicated traits, QTLs would be detected in smaller populations than our theoretical expectations would indicate. We have not factored this into our results, the gains from replication will vary substantially from experiment to experiment, in relation to the proportion of phenotypic variance due to non-genetic factors (environment, measurement error).

## Effect of deviation from Mendelian ratios

It was noted previously that a dominant QTL in an $F_2$ population resulted in two subpopulations with a 3:1 ratio. Deviations in Mendelian segregation ratios create a similar situation. Segreation distortion is particularly prominent in wide crosses, such as those often used for RFLP mapping. The relationship between contamination rate and population size in the "rare" pool and "common" pool is calculated at different skewing levels for QTLs having an 0.5 SD effect in an RI population (Fig. 3). The "rare" pool has a much higher contamination rate, and a larger population size is needed to identify QTLs than in a normally-segregating population. However, there is less contamination in the "common" pool (B), than in a normally-segregating population. A clean pool B can be obtained from a population of 200 plants, if a QTL has an allele effect of as little as 0.25 SD. As the skewing increases, the contamination rate in pool B decreases.

Segregation distortion increases the likelihood of false positive polymorphism, requiring use of a larger number of individuals in the pools to "homogenize" genomic regions outside the target. A highly-skewed locus might be polymorphic by chance if the DNA pools are not big enough. The relationship between allele frequency and the number of individuals needed in DNA pools is shown in Fig. 4. For example, at least 13 plants were required to form the DNA pools when allele frequencies of some loci are $\leq 0.20$ in the population. The number of individuals needed in pools to avoid false positives is determined by segregation ratios at the most deviant loci in the genome, not just the locus under selection.

As additional individuals are added to the pools, the length of the chromosome segment which remains polymorphic between the pools is likely to shrink, making it somewhat more difficult to find linked markers – however the markers which are found will be closer to the target gene (Michelmore et al. 1991).

## Post hoc analysis of DNA pooling in a backcross population of tomato

A backcross population of 237 plants, derived from an interspecific cross between *Lycopersicon esculentum* cv UC82B and *L. chmielewskii* accession LA 1028, was used in mapping QTLs for fruit mass, soluble solids concen-

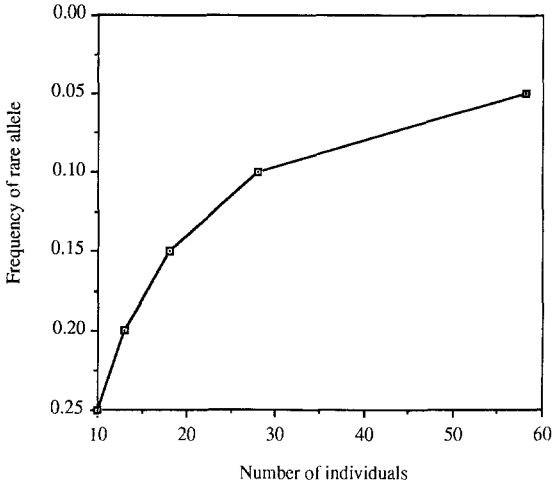



**Fig. 4** Number of individuals required in DNA pools to avoid identification of false positive markers (e.g., a 95% likelihood that a rare allele is presented in both DNA pools), under different degrees of segregation distortion

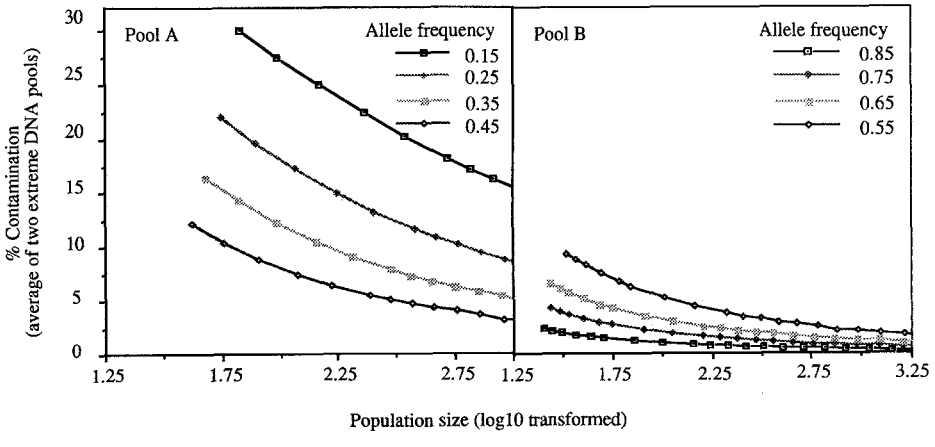



**Fig. 3** Relationship between contamination rate and population size at different degrees of segregation distortion for a QTL with an allele effect of 0.5 SD in a recombinant inbred population. Genotype A has a superior phenotype and is over-represented in the population. Pool A is formed in which ten individuals have the highest desired phenotype. Pool B is formed in which ten individuals have the lowest desired phenotype

tration and pH (Paterson et al. 1988). A total of 15 QTLs were identified, with allele effects ranging from 0.48 to 1.25 SD (Fig. 5). The correlation between QTL allele effects and the percent of contamination in DNA pools is highly significant ($r = 0.796$**). Our theoretical prediction was that QTLs with an allele effect that $> 0.75$ SD could be tagged in a backcross population of this size. The *SP* locus, which has a large effect on soluble solids concentration (1.25 SD) and pH (1.0 SD), had 0% contamination in the DNA pools. Marker CD34, linked to a QTL with 0.9 SD effect on fruit pH, had 5% contamination in the DNA pools. These were the only QTLs which could have been detected by DNA pools, and also the only QTLs with effects $> 0.75$ SD in this population, corroborating our prediction.

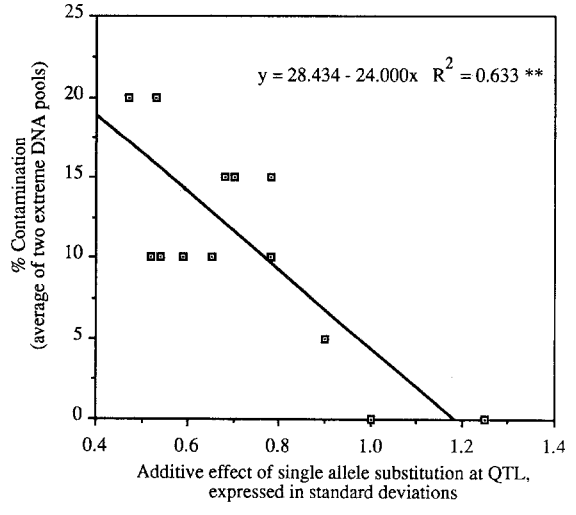
## Post hoc analysis of DNA pooling in an $F_2$ population of tomato

An $F_2$ population of 350 plants, derived from a wide cross between *L. esculentum* Mill and *L. cheesmanii*, was used to determine the chromosomal locations of QTLs affecting fruit size, soluble solids concentration and pH (Paterson et al. 1991). A total of 14 QTLs was detected, with effects of 0.25–0.80 SD. Our theoretical prediction was that in an $F_2$ population of this size, only QTLs with an effect $\geq 1.0$ SD might be tagged. None of the QTLs found in this population yielded an uncontaminated DNA pool, consistent with our prediction.

## Post hoc analysis of DNA pooling in an RI population of rice

One-hundred-and-thirty-nine RI lines derived from an *Oryza sativa* ssp. *japonica* and *O. sativa* ssp. *indica* cross, Moroberekan X CO39, were used to map QTLs encoding partial resistance to rice blast. Nine QTLs were found to be associated with lesion number, with allele effects of 0.32–0.88 SD. Pools from the highly-resistant lines (Moroberekan phenotype) had a higher contamination rate than expected because of extreme segregation distortion in the RI population. The lowest contamination in DNA pools was 10%. However, most of the pools from the highly-susceptible lines (CO39 phenotype) were uncontaminated, which is consistent with the result of theoretical estimations (Fig. 3). The correlation between QTL allele effects and pool contaminations was not significant, probably due to differences in segregation distortion of each QTL.

**Fig. 5** Relationship between QTL allele effects and contamination rate in a backcross population of tomato (*Lycopersicon esculentum* cv UC82B and *L. chmielewskii* accession LA 1028, Paterson et al. 1988)



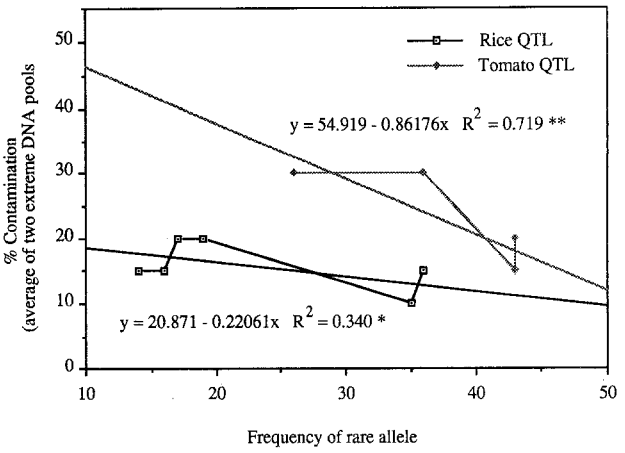




**Fig. 6** Relationship between segregation distortion and contamination rate in DNA pools in a tomato backcross (*Lycopersicon esculentum* cv UC82B and *L. chmielewskii* accession LA 1028, 237 individuals, Paterson et al. 1988) and a rice recombinant inbred population [Moroberekan (*Oryza sativa* ssp. *japonica*) × CO39 (*O. sativa* ssp. *indica*), 131 lines, Wang et al. 1993]

## Relationship between segregation distortion and contamination rate in tomato and rice

Segregation distortion was observed in both tomato and rice populations. As the skewness (deviation from Mendelian ratio) increases, the contamination rate also increases (Fig. 6). Marker CD46, linked to pH in the backcross population of tomato and having an allele effect of 0.8 SD, was highly skewed (only 26% of individuals were heterozygotes versus an expected 50%). Based on its allele effect and the size of the backcross population, our prediction of DNA contamination in pool A is 19%. The observed contamination of pool A at this locus was 20%. Marker RG333 was associated with partial resistance to rice blast, and its segregation was highly skewed (36% of individuals were homozygous for the *japonica* allele) in the RI population. The observed contamination of pool A was 20%, slightly higher than our prediction of 17%.

## Discussion

The success of DNA pooling strategies for tagging QTLs based solely upon phenotypic information (e.g., bulked segregant analysis) is dependent on the magnitude of the phenotypic effect of individual QTLs, the population size sampled, and the influence of non-genetic factors on the phenotype. A qualitative trait with 100% penetrance is the ideal case for using this method (Giovannoni et al. 1991; Michelmore et al. 1991). To tag genes underlying continuous variation (QTLs) using phenotype-based DNA pools, several precautions are recommended, including (1) the use of crosses in which extreme variation is observed; (2) the use of large populations, (Fig. 2); (3) the use of homozygous populations, i.e., RI or DH lines; (4) the replication of phenotypic evaluations, facilitated by the use of homozygous populations, but also possible by using $F_2$ or BC-derived lines for phenotype evaluations (e.g., Paterson et al. 1991). In general, RI or DH populations are preferable; however, the ease with which BC populations are made in many species may favor their use in some cases. $F_2$ populations appear to be a poor choice for tagging QTLs based solely upon phenotypic information.

By selecting individuals with the highest and lowest phenotypic values to form DNA pools, QTLs with very large effects on phenotype ($\geq 1.0$ SD) might be routinely identified if sufficiently large populations are studied. However, QTLs which account for much of the genetic variation in a trait will probably escape detection. For example, our results found that one QTL explaining about 20% of the phenotypic variation in tomato soluble solids, and two QTLs collectively explaining about 24% of the variation in tomato fruit pH, could have been tagged in phenotype-based DNA pools. However, additional QTLs explaining about 25% of the phenotypic variation in soluble solids, about 24% of the variation in fruit pH, and 58% of the variation in fruit mass, would have escaped detection. Only 13%, 9%, and 11% of variation in these traits, respectively, was attributable to non-genetic factors. Subsequent RFLP mapping experiments (Paterson et al. 1990) permitted detection of yet additional QTLs with very small effects which fell below the significance threshold of the initial RFLP mapping experiment upon which these results were based (Paterson et al. 1988). General concordance of *post hoc* analyses with theoretical predictions supports the conclusion that phenotype-based DNA pools can be reliably used to tag occasional QTLs of very large effect, but may frequently fail to tag many QTLs of smaller effect which collectively explain a large portion of the genetic variation in a trait. This is consistent with the expectations of Mendell et al. (1991) which indicate that very large populations ($>482$ individuals) are necessary to have even a 50% ability for distinguishing mixtures of populations differing by even as much as two standard deviations, with 95% confidence. (Our population size estimates are significantly lower, since we worked with averages rather than 95% confidence limits).

Segregation distortion increases the contamination rate of the DNA pool corresponding to the "rare" genotype, but reduces contamination of the "common" genotype pool. While the proportions of the two populations which fall in the wrong pool are determined by the population variance and tail proportion selected (see Fig. 1), the absolute number of contaminants in a pool is influenced by the number of individuals in each of the two populations. Severe segregation distortion in the rice population we examined appears to explain the failure of phenotype-based pooling to identify QTLs with effects which otherwise would be large enough to detect. One way to find QTLs using DNA pooling strategies in a skewed population might be to use $F_1$ DNA in place of the "rare" genotype pool. This is similar to using a backcross population, where one pool is heterozygous. However, as in backcross populations, only half the informative markers which are present in the donor parent could be used. Further, this requires prior knowledge of the nature and extent of segregation distortion in the population. As previously discussed, the "common" pool should include enough individuals to avoid finding many false positive markers (Fig. 4).

One issue which has received relatively little attention is the size of the genomic region which is being targetted. If DNA markers are used to create pools (Giovannoni et al. 1991), this is known from prior mapping. However, if phenotypes are used, the size of the polymorphic region varies with the number of individuals used in the pool (Michelmore et al. 1991), and with the population structure. For example, ten $F_2$ individuals carry 20 potentially-informative gametes, leaving a smaller polymorphic region than ten BC or DH individuals which have only ten informative gametes. The length of the polymorphic region in pools from RI ("single-seed descent") individuals is similar to that of $F_2$ individuals, since RI individuals each carry ten informative gametes but after twice as many opportunities for recombination (Taylor 1978). While a larger polymorphic region is a bigger target, a smaller region yields markers more closely linked to the target gene. In practice, other considerations probably outweigh the differences in size of the polymorphic region.

Experimental aspects of DNA pooling warrant discussion. Two considerations are paramount – the DNA of each individual needs to be equally represented in DNA pools, and the experiment needs to include measures for independent verification of putative new markers. Equal DNA concentrations may be assured simply by extracting DNA from each member of the pool individually, and quantitating. This is of particular concern regarding PCR-based assays, and DNA pooling procedures using RFLPs appear to be less sensitive to small variations in DNA concentration (Churchill et al. 1993). Verification of new markers is more difficult – our own protocol is to repeat putative positive results from PCR-based assays twice with a new DNA

preparation, then demonstrate co-segregation of the marker with the phenotype. For RFLPs, we proceed directly to verifying co-segregation, without the repetitions. Nonetheless, DNA from a population segregating for both the phenotype and the relevant markers should be available.

Finally, a more straightforward but quite valuable application of DNA pooling is for identifying new DNA markers near previously-mapped QTLs, utilizing DNA marker genotypes to create pools (Giovannoni et al. 1991). High-density mapping of genomic regions harboring QTLs serves two valuable purposes. First, given a large number of markers to choose from, there is a higher probability of finding at least one informative marker near the QTL, especially in an elite cross where polymorphism is limiting. Second, additional markers facilitate high-resolution mapping (Paterson et al. 1990), and ultimately map-based cloning of the QTL.

**Acknowledgements** We thank R. Wing, S. Kowalski, and two anonymous reviewers for valuable comments, and S. Tanksley (Cornell University) and R. Nelson (International Rice Research Institute) for the use of data generated under support of USDA (88-37262-3921), BARD (US-1388-87) (to S.T.), IRRI and Rockefeller Foundation (to R.N. and G.L.W.). This work was also supported by the Texas Agric. Expt. Stn., Texas Higher Education Coordinating Board (999902-148 to A.H.P., and R. Wing), and USDA-NRICGP (92-37300-7654 to A.H.P., R. Wing and K. Schertz).

## References

Allard RW (1960) Principles of plant breeding. Wiley, New York

Churchill GA, Giovannoni JJ, Tanksley SD (1993) Pooled sampling makes high-resolution mapping practical with DNA markers. Proc Natl Acad Sci USA 90:16–20

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

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

Grant D, Nelson W, Blair D, Katt ML, Martich J, Smith JSC, Bowen SL, Tenborg RA, Fincher RR, Meier R, Beavis WD, Smith OS (1989) Identification and genetic localization of loci affecting yield and other agronomic traits in maize. In: Current communications in molecular biology. Cold Spring Harbor Laboratory, New York, pp 131–134

Keim P, Diers BW, Olson TC, Shoemaker RC (1990) RFLP mapping in soybean: association between marker loci and variation in quantitative traits. Genetic 126:735–742

Mendell NR, Thode HC, Finch SJ (1991) The likelihood ratio test for the two-component normal mixture problem: powder and sample size analysis. Biometrics 47:1143–1148

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

Nienhuis J, Helentjaris T, Slocum M, Ruggero B, Schaefer a (1987) Restriction fragment length polymorphism analysis of loci associated with insect resistance in tomato. Crop Sci 27: 797–803

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

Paterson AH, Deverna JW, Lanini B, Tanksley SD (1990) Fine mapping of quantitative trait loci using selected overlapping recombinant chromosomes, in an interspecific cross of tomato. Genetics 124:735–742

Paterson AH, Damon S, Hewitt JD, Zamir D, Rabinowich HD, Lincoln SE, Lander ES, Tanksley SD (1991) Mendelian factors underlying quantitative traits in tomato: comparison across species, generations, and environments. Genetics 127:181–197

Selby SM (1973) Standard mathematical tables, 23rd ed. CRC Press, Boca Raton, Florida, pp 548–591

Stuber CW, Lincoln SE, Wolff DW, Helentjaris T, Lander ES (1992) Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. Genetics 132:823–839

Tanksley SD, Hewitt JD (1988) Use of molecular markers in breeding for soluble solids in tomato–a re-examination. Theor Appl Genet 75:811–823

Taylor, B (1978) Recombinant inbred strains: use in gene mapping. In: Morse, H (ed) Origins of inbred mice. Academic Press, New York, pp 423–438

Wang GL, Mackill DJ, Bonman JM, McCouch SR, Nelson RJ (1993) RFLP mapping of genes conferring complete and partial resistance to blast in a durably-resistant rice cultivar. Genetics (in press)

Weller JI, Soller M, Brody T (1988) Linkage analysis of quantitative traits in an interspecific cross of tomato (*L. esculentum* × *L. pimpinellifolium*) by means of genetic markers. Genetics 118:329–339

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res 18:6531–6535