Advanced Intercross Lines, an Experimental Population for Fine Genetic Mapping

A. Darvasi and M. Soller

Department of Genetics, The Alexander Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, 91904 Jerusalem, Israel

Manuscript received February 27, 1995
Accepted for publication August 2, 1995

ABSTRACT

An advanced intercrossed line (AIL) is an experimental population that can provide more accurate estimates of quantitative trait loci (QTL) map location than conventional mapping populations. An AIL is produced by randomly and sequentially intercrossing a population that initially originated from a cross between two inbred lines or some variant thereof. This provides increasing probability of recombination between any two loci. Consequently, the genetic length of the entire genome is stretched, providing increased mapping resolution. In this way, for example, with the same population size and QTL effect, a 95% confidence interval of QTL map location of 20 cM in the F_2 is reduced fivefold after eight additional random mating generations (F_{10}). Simulation results showed that to obtain the anticipated reduction in the confidence interval, breeding population size of the AIL in all generations should comprise an effective number of ≥ 100 individuals. It is proposed that AILs derived from crosses between known inbred lines may be a useful resource for fine genetic mapping.

THE normal range of phenotypic variation in a wide variety of physiological and morphological traits has a polygenic basis and is quantitative in nature. That is, trait variation is determined by a number of loci, with allele substitution at each locus having a relatively small effect on trait value; trait expression is also affected by macro- and microenvironmental factors. Traits of this nature are termed "complex traits" or "quantitative traits" and the individual loci affecting trait expression are usually termed "quantitative trait loci" (QTL).

THODAY (1961) was the first to estimate map location of a QTL using a pair of flanking markers. Since then, several statistical methods have been developed that can exploit the information provided by larger numbers of markers and more complete genome maps. In the main, these methods are based on regression analysis and maximum likelihood, and they provide statistical tests for the presence of a QTL and estimates of the QTL parameters (*i.e.*, gene effect, dominance and map location), (WELLER 1986; JENSEN 1989; LANDER and BOTSTEIN 1989; KNAPP *et al.* 1990; HALEY and KNOTT 1992; DARVASI *et al.* 1993; JANSEN 1993; ZENG 1993, 1994; JANSEN and STAM 1994). However, these methods are not able to efficiently utilize the increasing ability to saturate a given chromosomal region with very closely spaced markers. In particular, with the usual F_2, BC, half-sib or full-sib experimental designs and populations of reasonable size, even when using an infinite number of markers, a QTL of moderate effect can only be assigned to a map location in a rather broad chromosomal region (DARVASI *et al.* 1993). This is due primarily to the lack of sufficient recombinational events in small chromosomal regions, even in large F_2 or BC populations.

In the present study, we propose a novel type of experimental population, specifically intended to exploit the power of inbred lines and saturated genetic maps, to provide fine mapping of QTL of moderate effect in experimental populations of reasonable size. The proposed population, termed an "advanced intercross line" (AIL), is based on the well-known principle that continued intercrossing of a population, will reduce linkage disequilibrium and cause the proportion of recombinants between any linked loci to asymptotically approach 0.5 (FALCONER 1989). An AIL is initiated by a cross between two inbred lines, and derived by sequentially and randomly intercrossing each generation, until advanced intercross generations are attained. In an AIL, the many recombinational events required for fine mapping of QTL are accumulated in a single relatively small population over the course of many generations rather than by producing and examining many progeny in a single large F_2 or BC generation. In this study, we show that an appropriately formed AIL can provide a three- to fivefold reduction in the confidence interval of QTL map location as compared with a F_2 or BC population, without any increase in the number of individuals phenotyped or genotyped. The only requirement is that breeding population size of the AIL should not fall below an effective number of 100 individuals per generation.

Corresponding author: Ariel Darvasi, Department of Genetics, The Hebrew University of Jerusalem, 91904 Jerusalem, Israel.
E-mail: arield@cc.huji.ac.il

THEORY

An AIL is produced from a F_2 population generated by crossing two inbred lines assumed homozygous for alternative alleles at a series of QTL and marker loci. The following generations, F_3, F_4, F_5 . . . , are sequentially produced by randomly intercrossing the previous generation. For QTL mapping purposes, only individuals from one of the later generations are phenotyped and genotyped; the previous generations, termed the "breeding generations," are reared and reproduced only.

Proportion of recombinants in an AIL: The expected proportion of recombinant haplotypes, r_t, between two loci, *A* and *B*, in the F_t generation of an AIL, will equal the proportion of recombinant haplotypes in the previous generation, r_{t-1}, **plus** the net increase in recombinant haplotypes as a result of recombination in the double heterozygotes consisting of the two original haplotypes (say, *AB/ab*, which produces new recombinants), **less** recombination in double heterozygotes consisting of two recombinant haplotypes (say, *Ab/aB*, which regenerates the original parental haplotypes). If the proportion of recombinant haplotypes in the $t-1$ generation is r_{t-1}, the proportion of parental-type double heterozygotes will be $(1-r_{t-1})^2/2$ and the proportion of recombinant-type double heterozygotes will be $r_{t-1}^2/2$. Consequently,

$$r_t = r_{t-1} + \tfrac{1}{2}r(1-r_{t-1})^2 - \tfrac{1}{2}rr_{t-1}^2 = \tfrac{1}{2}r + r_{t-1}(1-r) \quad (1)$$

where r is the proportion of recombination in the F_2 generation. This is similar to the expression for proportion of recombinants in a random mating population, starting from an initial level of linkage disequilibrium (FALCONER 1989). This was also independently derived by W. BEAVIS (personal communication) to investigate the influence of random mating on recombination among marker loci in maize. From Equation (1), r_t can readily be derived, as a function of r alone, giving:

$$r_t = \frac{1-(1-r)^{t-2}(1-2r)}{2} \quad (2)$$

Obtaining r as a function of r_t using Equation (2) can only be done numerically. However, when dealing with relative small values of r, Equation (2) can be accurately approximated using a first order Taylor's expansion, giving:

$$r_t = \frac{rt}{2} \quad (3)$$

so that,

$$r = \frac{2r_t}{t} \quad (4)$$

In an experiment carried out in Maize, proportion of recombination in a fifth random mating generation, was on average as expected from Equation (2) (W. BEAVIS, personal communication).

QTL mapping accuracy: QTL mapping accuracy is expressed as the confidence interval, with a certain confidence level, of QTL map location. Consider a F_2 population, derived from a pair of inbred lines, that has been phenotyped and genotyped with a given marker spacing and in which a linkage analysis is performed. In this case, the confidence interval for map location of an independently segregating QTL depends on the following parameters: length of the chromosome, QTL location relative to chromosome ends, marker spacing, experimental population size, and standardized gene effect at the QTL (DARVASI *et al.* 1993). The width of the confidence interval is defined in units of proportion of recombination as this is the only unit for distance between loci that can be estimated in a linkage experiment. Distances in units of proportion of recombination can be transformed to cM units using a mapping function (BAILEY 1961).

In an AIL, *if* (i) all of the parameters listed above that influence confidence interval of QTL map location are identical to those of a F_2 population and (ii) the proportion of recombinants is binomially distributed (as in a F_2 population) with an expectation of r_t, as given in Equation (2), *then* the width of the confidence interval of QTL map location in the AIL, represented as a proportion of recombination, will necessarily be the same as in a F_2. Therefore, given a confidence interval, C, in the F_t generation (where C is the distance from a given QTL to one end of the confidence interval, in proportion of recombination units), the corresponding confidence interval on the scale of the F_2 generation, C', can be approximated using Equation (4), as:

$$C' = \frac{C}{t/2} \quad (5)$$

That is, with advancing generations, the confidence interval is reduced by a factor of $t/2$. To obtain the corresponding confidence interval M', in cM, C' is transformed to cM using Haldane's mapping function (HALDANE 1919) and doubled to represent the total confidence interval length. The confidence interval can also be obtained without the use of Taylor's approximation as given in Equation (4), by a numerical calculation of r as a function of r_t from Equation (2).

Assumptions: As stated here, Equation (5) represents the reduction in confidence interval of QTL map location provided by an AIL *only* if the assumptions listed above hold. The validity of these assumptions is now considered. Marker spacing, in proportion of recombination units, and sample size are parameters controlled in the experiment. Thus they can be kept equal for the F_2 population and the AIL. It should be noted, however, that equal marker spacing in proportion of recombination, implies higher marker density in a

given physical region for an AIL as compared with the corresponding F_2 population. Additional aspects of marker spacing in AIL are discussed in detail in the *Statistical power and marker spacing* section.

Equal chromosome length and QTL map location between F_2 and the AIL can be obtained only by assuming a chromosome of infinite length. This assumption is required because in the AIL the relevant chromosome length increases. Binomial distribution of recombinants in the AIL can be obtained by assuming an infinite number of individuals in the breeding generations. In this case, sampling variation between generations will not contribute to the variance of recombinants in the AIL. Consequently, the variance of the number of recombinants will be as in a F_2 population. Infinite chromosome length and infinite number of individuals in the breeding generations are clearly not a realistic case. These are theoretical assumptions that make the theory correct. A simulation (see next section) study was carried out therefore to investigate the influence of finite chromosome length and finite sample size in the breeding generations on the confidence interval of QTL map location. The simulations will provide information on whether reasonable chromosome length (*i.e.*, 100 cM) is "close" enough to infinite, and which breeding population size is "close" enough to infinite in order that the theory, which assumes infinite sizes for both parameters, will still hold.

The assumption of constant QTL effect will hold only when all QTL affecting the trait of interest are unlinked. When some of the QTL affecting trait value, other than the QTL being mapped, are linked, the genetic variance will change due to disassociation of the linked QTL. As a result, the standardized QTL gene effect can be altered in the course of forming an AIL. This, however, is expected to have only a small to moderate effect on QTL gene effect, as discussed in the section on *linked QTL*. The case where the mapped QTL itself is in linkage with another QTL is also considered in that section.

Simulation study: We have previously shown through a simulation study, in a backcross design (Darvasi *et al.* 1993), that a 95% confidence interval of QTL map location, using an *infinite number of markers,* is a close approximation of the 95% confidence interval of a maximum likelihood estimate (MLE) of QTL map location obtained with *moderately spaced markers* (10–20 cM). This is a very useful result, because in the "infinite number of markers" model, the QTL is always at a marker so that the simulation need not estimate QTL location with respect to flanking markers. Consequently a 95% confidence of QTL map location, using an infinite number of markers is much easier to simulate than the confidence interval using moderately spaced markers. Note that the above result also shows that once a moderate marker spacing is achieved (*i.e.*, a proportion of recombination of ~0.10–0.15 between adjacent markers) further increase in marker density does not further decrease the confidence interval of QTL map location. The implicit assumption is that it will be possible to achieve a marker spacing with $r_t = 0.10$ or so between adjacent markers in the AIL, so that confidence interval of a QTL mapped in the AIL will not be limited by marker spacing. Consequently, confidence interval of QTL map location in an AIL can also be closely approximated by the readily calculated confidence interval with an infinite number of markers. On this basis, and because of the simplicity of its execution, we now use the concept of an infinite number of markers, solely as a calculating device, to approximate the effect of the AIL design on confidence interval of QTL map location, for realistic situations, where number of markers is limited, but marker density in the region of interest is on the order of $r_t = 0.10$. All simulations assumed a 100-cM chromosome with a QTL located at its midpoint. A F_2 population was generated, as a cross between two inbred lines with alternative alleles for all markers and for the QTL. To simulate a situation of an infinite number of markers a marker was placed every 0.1 cM. The quantitative trait was assumed to have a normal distribution with equal variances for all QTL genotypes and standardized gene effects of $-d$, h and d for the QTL genotypes qq, Qq and QQ, respectively. Subsequent breeding generations, consisting of N_b individuals each, were generated by randomly choosing two individuals from the previous generation to serve as the parents of each new individual in the following generation. From each parent, a single gamete was generated on the assumption of no recombination interference. Thus, sampling was from a binomial distribution with expectation of 0.001 to define a recombination event between any two adjacent markers. The two gametes obtained represented the new offspring. At generation t, N test individuals were produced and a maximum likelihood estimator for QTL map location was obtained. Empirical confidence intervals were constructed from the simulations as detailed in Darvasi *et al.* (1993).

The parameter combinations for the simulation study consisted of: (i) $N_b = N = 500$, $d = 0.75$; (ii) $N_b = N = 1500$, $d = 0.25$; (iii) $N_b = N = 100$, $d = 1.0$; (iv) $N_b = 50$, $N = 1500$, $d = 0.25$; (v) $N_b = 100$, $N = 1500$, $d = 0.25$. The first parameter combination was chosen to investigate a case where the initial confidence interval is small; thus chromosome length is not expected to influence the confidence interval. Parameter combinations (ii) and (iii) were chosen to provide similar confidence intervals in a F_2 population achieved by different combinations of gene effect and population size. The specific effect of small breeding and test population size as compared with gene effect, on confidence interval in an AIL can then be determined. The two final parameter combinations, (iv) and (v), were chosen to further investigate the specific influence of breeding population size.

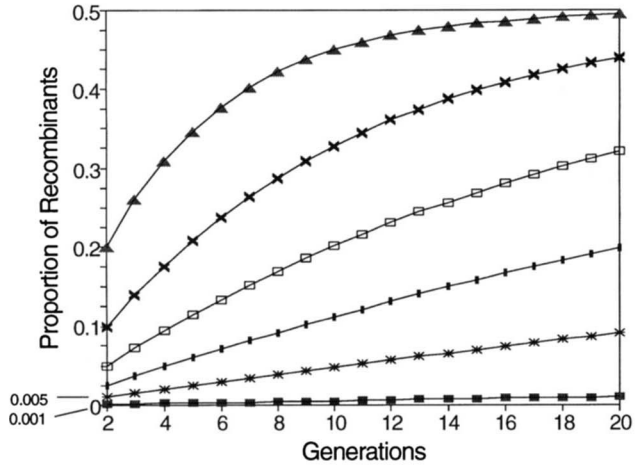

FIGURE 1.—The proportion of recombinant haplotypes as a function of generation number according to initial proportion of recombinant haplotypes in the F_2 generation (initial proportion of recombinants: ■, 0.001; ∗, 0.005; ˣ, 0.025; □, 0.05; ×, 0.1; ▲, 0.2).

At each of the above parameter combinations, 1000 replicate simulations were carried out using a value of $h = 0$. The breeding population sizes (N_b) of all generations except the last were equal, but only the last generation (test population size, N) was analyzed. For each parameter combination, a 95% confidence interval was empirically obtained from the 1000 replicate simulations (for details see DARVASI *et al.* 1993).

Linked QTL: Although the basic nature of quantitative traits is that they are affected by multiple genes, fine structure analysis (THOMPSON 1975), patterns of segregation following hybridization (LANDE 1981) and mapping results from plant and animal populations suggest that the number of chromosomal regions differentiating two inbred lines with respect to a particular trait is not necessarily large. As shown by MCMILLAN and ROBERTSON (1974) and SOLLER *et al.* (1979), this can result from a few genes of moderate to large effect or from many genes of small effect. Thus, current data do not enable us to distinguish between these possibilities. Consequently, it seems reasonable to proceed on the simpler assumption that experimental results represent QTL of moderate to large effect that are randomly dispersed throughout the genome. On this basis, simple Poisson considerations show that the large majority of QTL will be found standing alone. Thus the analysis in this study should probably hold for a large part of QTL in a particular AIL. More to the point, it is precisely the fine mapping potential of an AIL that may allow this question to be resolved.

Even when assuming that most of the QTL are independent, the presence of some linked QTL will have an influence on the standardized gene effect of all QTL, because the approach to linkage equilibrium of linked QTL, in the course of advanced generations, will affect genetic variance. Nevertheless, the effect of the transition to linkage equilibrium on genetic variance is mixed. If the crossed populations were derived from populations originally at linkage equilibrium, then one can assume that an approximately equal number of linked QTL are found in coupling (*i.e.*, alleles of the same direction of effect are linked) as in repulsion (alleles of opposite direction of effect are linked). When linked QTL are in coupling, recombination will decrease genetic variance in the F_t generation as compared with F_2 generation. When the linked QTL are in repulsion, recombination will increase genetic variance in the F_t generation as compared with the F_2 generation. Overall, then, one would expect only a minor change in genetic variance. However, in many instances crosses will be carried out between parental lines chosen to differ in the quantitative trait of interest. Consequently, linked QTL are likely to be found predominantly in coupling. In this case, total genetic variance in the F_t generation will be less than in the F_2 generation. Thus, relative gene effect at the QTL will increase, and confidence interval of QTL map location in the AIL will decrease even more relative to that in a F_2.

Statistical power and marker spacing: As previously stated the infinite number of markers model provides approximate estimates of confidence intervals for the situation where a moderate marker spacing is used. Obviously, in an actual experiment, a finite number of markers will be used with a given marker spacing. The same markers in an AIL will correspond to a wider marker spacing when measured in proportion of recombination units. Thus, AIL will require considerably higher marker density to provide equivalent statistical power for QTL *detection* as compared with a F_2 population of equal size. This exemplifies the difference in population design required for QTL detection, as opposed to QTL mapping (DARVASI *et al.* 1993).

Once a QTL is detected, however, an AIL does not require genotyping more densely spaced markers than a F_2 to achieve an equivalent confidence interval of QTL map location. This follows from the fact that genotyping with a smaller marker spacing than the 95% confidence interval itself, does not significantly increase accuracy (DARVASI *et al.* 1993). Thus, for example, if the resolving power of an experiment (*i.e.*, the 95% confidence interval of QTL map location using an infinite number of markers) in an F_2 population of, say, 1500 individuals is 1 cM; then genotyping with a marker spacing narrower than 1 cM will not significantly increase mapping accuracy. In an AIL, the same confidence interval, measured by *physical* length, will correspond to a genetic distance of, say, 5 cM, and can be obtained with, say, 500 individuals. In this case, genotyping with a marker spacing narrower than 5 cM, in the AIL, will not increase mapping accuracy. Note, however, that the 5-cM marker spacing in the AIL corresponds to the same *physical* marker spacing as the 1 cM marker spacing in the F_2.

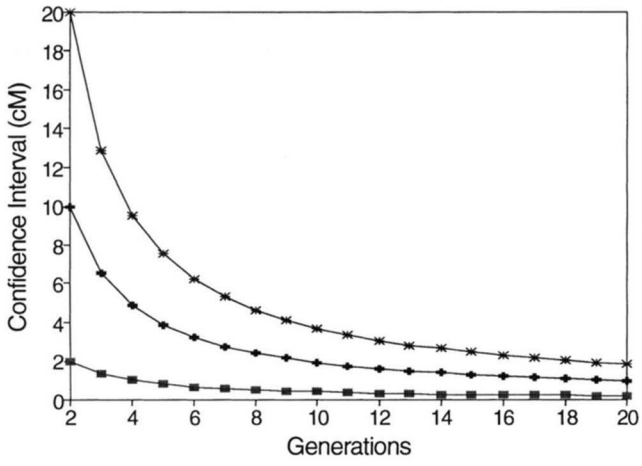

FIGURE 2.—The theoretical decrease in confidence interval of QTL map location as a function of generation number for several initial confidence intervals in the F_2 generation (■, 2 cM; ˣ, 10 cM; ∗, 20 cM).

NUMERICAL RESULTS

Figure 1 presents the expected proportion of recombinant haplotypes as a function of the number of intercross generations according to the initial proportion of recombination in the F_2 generation. It can be seen that when the initial proportion of recombination is at all appreciable, say, $r > 0.10$, the expected proportion of recombinant haplotypes increases rapidly, and asymptotically approaches 0.50 by 10–20 generations. When the initial proportion of recombination is small, say, $r < 0.05$ the proportion of recombinant haplotypes increases virtually linearly through 20 generations. For example, after 20 generations, initial proportions of recombination of $r = 0.01$ and 0.005 give proportions of recombinant haplotypes of $r_{20} = 0.091$ and 0.048, respectively.

Figure 2 presents the expected value of the confidence interval in an AIL, as a function of the number of generations, for several initial values of the confidence interval in the founder F_2 population. The values presented here are obtained from expression (5), transformed to cM and doubled. The expected confidence interval decreases dramatically in the early generations. However after ~10 generations, further reduction in the confidence interval is moderate and linear. Thus, a F_2 confidence interval of 20 cM is reduced to 3.7 cM (more than fivefold!) after eight additional generations, (F_{10}), but only to 1.8 cM (only a further twofold) after an additional 10 generations (F_{20}).

It was found that in all cases the confidence interval calculated through the use of Taylor's approximation was virtually identical to the confidence interval obtained through exact numerical calculations from Equation (2) (data not shown).

Figure 3, A–C, is based on the simulation studies. Each panel shows two curves describing the decrease in the confidence interval as a function of the number of generations; one curve representing the exact numerical theoretical calculation based on Equation (2), and the second curve representing the results obtained

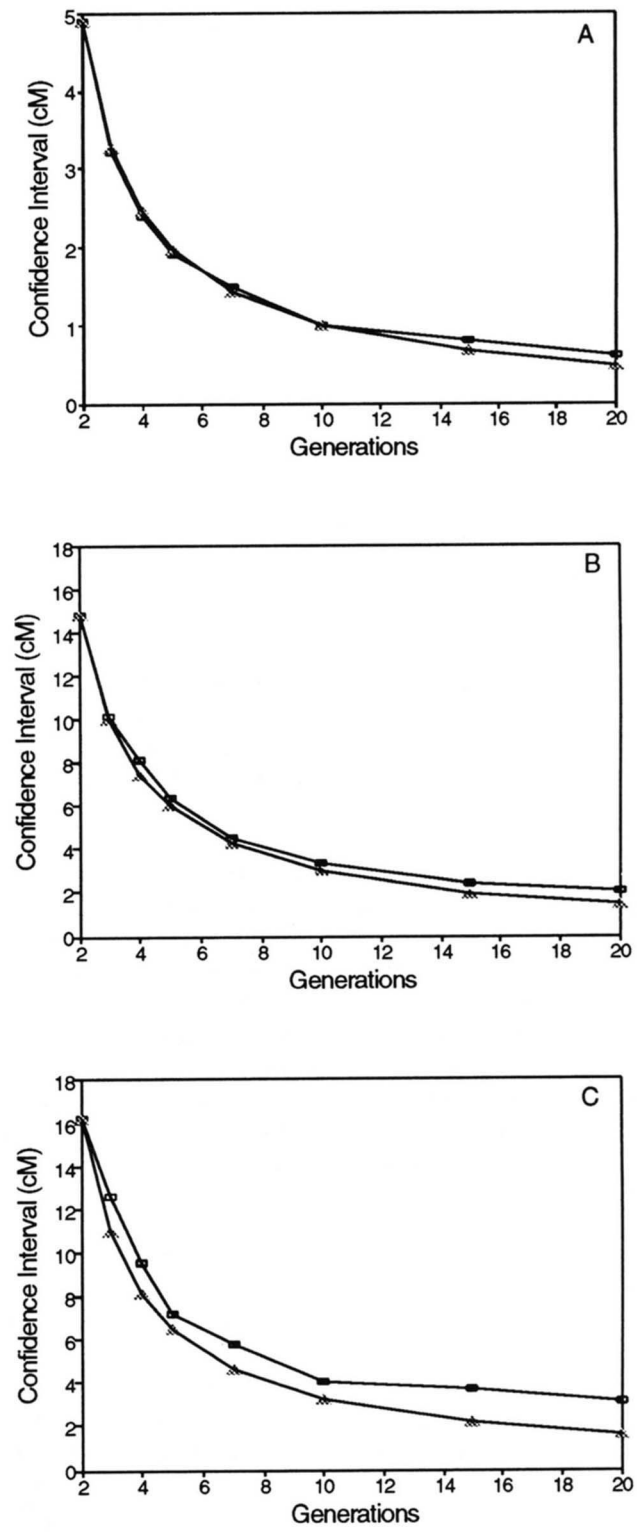

FIGURE 3.—Theoretical values (□) and simulation results (▲) for confidence interval of QTL map location as a function of generation number, for various combinations of gene effect (d), breeding population size (N_b), and test population size (N): (A) $d = 0.75$, $N_b = N = 500$; (B) $d = 0.25$, $N_b = N = 1500$; (C) $d = 1.0$, $N_b = N = 100$.

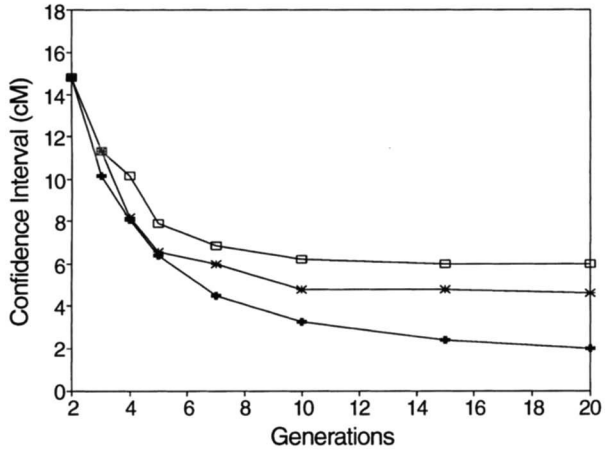

FIGURE 4.—Simulation results for confidence interval of QTL map location as a function of generation number for gene effect, $d = 0.25$; test population size, $N = 1500$; and three values of breeding population size, $N_b = 50$ (□), $N_b = 100$ (*), and $N_b = 1500$ (+).

from 1000 replicated simulations. Each pair of curves represent a different combination of gene effect and population size: $d = 0.75$, $N_b = N = 500$ (Figure 3A); $d = 0.25$, $N_b = N = 1500$ (Figure 3B); and $d = 1.0$, $N_b = N = 100$ (Figure 3C). As expected, for large breeding population sizes (Figure 3, A and B), the simulation values are extremely close to the theoretical values. For a smaller breeding population size, however, the difference between the theoretical values and the simulations was more noticeable, particularly in the later generations (Figure 3C). This was the case, even though the initial confidence interval for the simulations was about the same as for the simulation of Figure 3B. From this, we conclude that in the later generations of an AIL, it is primarily breeding population size, rather than gene effect, that determines the correspondence between the theoretical expectations and those that will be obtained in practice.

In the above simulations, breeding and test population sizes were the same ($N_b = N$), so that their effects on confidence interval are confounded. To disentangle these effects, Figure 4 presents the confidence interval obtained by simulation for the case where $d = 0.25$, $N = 1500$ and breeding population sizes were $N_b = 50$, 100 and 1500 ($N_b = 1500$ is repeated from Figure 3B). Until the F_5, reduction in breeding population size to $N_b = 100$ did not influence the confidence interval of map location, whereas a reduction of the breeding population to $N_b = 50$ slightly increased the confidence interval. However, when breeding population size was small, further advanced generations did not provide a further reduction in the confidence interval. For example, after 10 generations, an initial confidence interval of 14.8 cM was reduced to 3.3 cM using a large breeding population ($N_b = 1500$), but only to 4.8 and 6.1 cM when using breeding population sizes of $N_b = 100$ and $N_b = 50$, respectively.

DISCUSSION

The result of this study show that for given QTL gene-effect and experimental population size, an existing AIL can significantly increase the accuracy of estimated QTL map location. For example, for a QTL with a gene effect of $d = 0.25$ and a test population size of $N = 1500$, a confidence interval of 14.8 cM is obtained in the F_2 generation. This is reduced to ~5 cM in an AIL that has reached the F_{10} generation or later. The only requirement is that the AILs have been reproduced with an effective number of ≥100 breeding individuals. This requirement on breeding population size was obtained from the simulations and is necessary to practically eliminate the influence of accumulated variation in proportion of recombination throughout the breeding generations.

An AIL, with its greatly increased proportion of recombination between any two loci, provides an excellent tool for disentangling and individually mapping closely linked QTL. Quantifying the advantages of an AIL over a F_2 for mapping linked QTL is beyond the scope of this study. Nevertheless, it is worth noting that recently developed sophisticated methods aimed at mapping linked QTL in a F_2 population (*e.g.,* ZENG 1993, 1994) could be equally applied in an AIL. The only difference would be that the problematic linkage between QTL will be significantly reduced. For example, two linked QTL at a distance of 5 cM, which in practice cannot be separated in a F_2 population, will be at corresponding genetic distances of 26 cM in an AIL at F_{10} and 51 cM in an AIL at F_{20}! At these distances, the above-mentioned method will be able to separate them with ease.

As pointed out in the introduction, the greater efficiency of an AIL, relative to a F_2, for determining map location of QTL derives from the increased number of recombinant events found between any two loci in an AIL. It is this that allows more precise localization of the QTL to a specific chromosomal region. However, this very same increase in recombinational events reduces the effect due to the QTL that will be associated with any particular marker contrast. In particular, the contrast between alternative homozygotes at a given linked marker, MM-mm, will equal $2(1 - 2r)d$ in a F_2 generation, but only $2(1 - 2r_t)d$ in the F_t generation. Thus, in the F_t generation, the quantitative effect associated with a particular marker will be less than that in the F_2 generation. Consequently, power for QTL detection (with respect to a **given** marker) is less in an AIL than in a F_2, while precision of estimation of QTL map location (with respect to potentially saturated map) is greater in an AIL than in a F_2. This is unusual in that most experimental parameters (*e.g.,* population size, gene effect) have similar effect on power as on confidence interval of QTL map location. This is, however, but a further example to the separation between the effect

of experimental factors on confidence interval of QTL map location and on power for detection of a QTL, that has been remarked upon previously (DARVASI *et al.* 1993).

Because of the decreased effectiveness of an AIL for detecting the presence of a QTL in a given chromosomal region, when the initial founder lines of the AIL are available, it may be possible to cross these to produce a F_2 or BC population for initial identification of the QTL. This would be followed by fine mapping in the AIL. In most instances, however, it will be possible to carry out initial QTL mapping on the AIL, with only a small increase in absolute genotyping costs as compared with the equivalent F_2 population. As shown in Equation (3), the total proportion of recombination between two closely linked markers in generation t, will be increased by a factor of $t/2$ relative to the proportion of recombination in the F_2 generation. Thus for an AIL at the F_{10} stage, about five times as many markers are required to obtain similar statistical power as in the F_2 population. Nevertheless, the initial assignment of QTL to chromosomes can be achieved with relatively small population size and wide marker spacing (DARVASI and SOLLER 1994a). Only this number of individuals would need to be genotyped with respect to markers covering the entire genome. The entire population would be subsequently genotyped only for those markers covering relevant QTL containing chromosomal regions.

Furthermore, because fine mapping in an AIL would generally be applied to a single trait at any one time, selective genotyping (LEBOWITZ *et al.* 1987; LANDER and BOTSTEIN 1989; DARVASI and SOLLER 1992), possibly with DNA pooling (HILLEL *et al.* 1990; MICHELMORE *et al.* 1991; CHURCHILL *et al.* 1993; PACEK *et al.* 1993; DARVASI and SOLLER 1994b; KHATIB *et al.* 1994), could be used in the initial mapping step. In this case, although about five times more genotypings will still be required for initial mapping in the AIL, the overall number of genotypings required is relatively small so that the absolute difference in costs as compared with a F_2 population will also be small (equivalent, say, to the difference between 200 as compared with 1000 genotypings).

AIL can be compared instructively with recombinant inbred lines (RIL). RIL are produced from a F_2 population by a series of self-fertilizations or through continued brother-sister mating. This process, as in AIL, increases the proportion of recombination, r^*, between any two loci, relative to the proportion of recombination, r, in a F_2 population. For self-fertilization, $r^* = 2r/(1 + 2r)$ and for brother-sister mating $r^* = 4r/(1 + 6r)$, (HALDANE and WADDINGTON 1931). This amount of recombination in a set of RIL is approximately equivalent to the expected amount of recombination present in an AIL at the F_4 and F_8 generation, respectively. Thus, a set of RIL presents many of the map-expanding properties of an AIL. For reasonable statistical power in QTL mapping, however, a set consisting of a large number of RIL is required (SOLLER and BECKMANN 1990). Consequently, a single AIL is as effective for fine mapping as a very large number of RIL, yet is not more expensive to produce than a single RIL. In some instances, however, where the trait of interest is influenced by a small number of QTL with low heritability, RIL can be an efficient mapping tool, because RIL significantly reduce the environmental variance. In such cases, it may be worth considering intercalating two to six generations of random mating before initiating self-fertilization or brother-sister mating, to further increase the amount of recombination.

AIL are particularly applicable to species with a short generation cycle that can be easily reproduced by intercrossing and for which inbred lines exist. Inbred lines of mice are the outstanding example, because these can be readily crossed and subsequently intercrossed without great effort. At four generations a year, each generation consisting of 50–100 single pair matings, an AIL(10) (obtained from the F_{10} generation) could be produced in 3 years. Inbred lines of chicken also exist, but most have been formed within the White Leghorn breed and hence may not differ sufficiently in genetic architecture for most traits of interest to be useful for QTL mapping. Inbred lines of corn, and cultivars and accession lines of most annual selfing plant species, are also candidates for AIL production. In this case, production of an AIL(10) would take longer time due to the annual reproduction period of most plant species; although, depending on photoperiod requirements, two to three generations a year could be obtained by rearing the plants in warm climates, by alternating reproduction between Southern and Northern hemispheres, or by use of a phytotron.

The time scale required for production of an AIL applies, of course, only to the first time that such a population is produced. In plant species, an AIL, once produced, can be stored in the form of seeds, and test populations produced as needed. In mice, once an AIL(10), say, has been produced, it can then be maintained by reproducing at a much slower rate. Indeed, for standard lines and crosses, one could maintain a population consisting of the two parental lines and two advanced generations that are about 10 generations apart. When one of the AIL reaches the 10th generation, the reproduction cycle would be slowed, and a new AIL would be initiated. Consequently, there would always be an AIL of the 10th generation or later, available. Because most of the increment in mapping accuracy is obtained in the first 10 generations of an AIL, this scheme provides an efficient long-term genetic resource population for fine mapping any trait that has a genetically different architecture in the original parent lines. Indeed, for a relative small investment, such leapfrog AIL could be produced between pairs of the most important inbred lines of mice, as resource populations for fine mapping in this species. It should be noted

that once an AIL exists, its advantages apply equally to simple Mendelian traits, as to quantitative traits. Thus, an AIL could also serve as a useful resource for fine mapping of marker loci or known genes. In particular, methods developed for saturating specific chromosomal regions with marker loci (CHURCHILL *et al.* 1993; LISITSYN 1994), could be used in an AIL to map the markers with increased accuracy.

When an existing AIL is not available, the question arises whether to pursue fine mapping through increased population size at the F_2 level, or to develop an AIL(10), say, specifically for this purpose. The most important consideration in making such a decision would be the total number of individuals reared and phenotyped under each strategy. When marker density is not a limitation, the confidence interval of QTL map location appears to be inversely proportional to experimental population size (DARVASI *et al.* 1993). On this principle, simple calculations show that an AIL will generally reduce the total number of individuals that must be reared for fine mapping by a factor of two or three; and the number of animals that must be phenotyped by a factor of three or four, relative to a F_2 population providing equivalent mapping accuracy. For example, a QTL with gene effect of $d = 0.25$ can be mapped with a 95% confidence interval of 5 cM using a F_2 population size of ~4500, where all individuals are reared and genotyped (unpublished simulations). Alternatively, the same confidence interval can be achieved using an AIL(10) with 1500 individuals. Assuming a breeding population size of 100, this sums to a total of 2300 individuals reared and 1500 phenotyped. When rearing and phenotyping costs are low, it may be preferable to simply produce and phenotype a larger number of F_2 individuals. Otherwise, it may be useful to invest the time required to produce an AIL. Thus, in an experiment with corn or mice aimed at simple morphological traits, simply increasing F_2 numbers might be the method of choice. But for experiments involving traits that are more difficult to evaluate, such as water-use efficiency in plants or complex behaviors in mice, it may be more efficient to produce an AIL. This may also be the case when total rearing facilities are limited, so that a number of reproductive seasons are required to achieve large numbers. In this case, one might well accumulate numbers through progressive intercrossing, rather than by repeating a F_2 or backcross. Calculating confidence intervals of QTL map location for data obtained over a progressive series of generations will be an intriguing, but certainly not insurmountable, statistical exercise.

AILs can also be formed in species where inbred lines are not available by initiating the AIL with a cross between a single male and a single female, forming a large F_1 population (through multiple ovulation and embryo transplantation, if necessary), and then continuing by repeated intercrossing at the earliest possible age. A population formed in this way would contain four independent chromosomes when initiated, as compared with the two chromosomes of a F_1 formed by crossing two inbred lines. The usual QTL mapping paradigms, developed for a F_2 population, can readily be adapted to this situation (R. FERNANDO, personal communication). Within this context, the ability of AIL to increase the number of recombinational events between any two loci, ought to lead to a decrease in the confidence interval of QTL map location, comparable with that found in the present study with respect to a F_2 population. Once formed, such a population could serve as a general fine-mapping resource for the species in question.

LITERATURE CITED

BAILEY, N. T. J., 1961 *Introduction to the Mathematical Theory of Genetic Linkage.* Oxford University Press, London.

CHURCHILL, G. A., J. J. GIOVANNONI and S. D. TANKSLEY, 1993 Pooled-sampling makes high-resolution mapping practical with DNA markers. Proc. Natl. Acad. Sci. USA **90:** 16–20.

DARVASI, A., and M. SOLLER, 1992 Selective genotyping for determination of linkage between a marker locus and a quantitative trait locus. Theor. Appl. Genet. **85:** 353–359.

DARVASI, A., and M. SOLLER, 1994a Optimum spacing of genetic markers for determining linkage between marker loci and quantitative trait loci. Theor. Appl. Genet. **89:** 351–357.

DARVASI, A., and M. SOLLER, 1994b Selective DNA pooling for determination of linkage between a molecular marker and a quantitative trait locus. Genetics **138:** 1365–1373.

DARVASI, A., A. WEINREB, V. MINKE, J. I. WELLER and M. SOLLER, 1993 Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. Genetics **134:** 943–951.

FALCONER, D. S., 1989 *Introduction to Quantitative Genetics,* Ed. 3. Longman, New York.

HALDANE, J. B. S., 1919 The combination of linkage values and the calculation of distance between the loci of linked factors. J. Genet. **8:** 299–309.

HALDANE, J. B. S., and C. H. WADDINGTON, 1931 Inbreeding and linkage. Genetics **16:** 357–374.

HALEY, C. S., and S. A. KNOTT, 1992 A simple regression method for mapping quantitative loci in line crosses using flanking markers. Heredity **69:** 315–324.

HILLEL, J., R. AVNER, C. BAXTER-JONES, E. A. DUNNINGTON, A. CAHANER *et al.,* 1990 DNA fingerprints from blood mixes in chickens and turkeys. Anim. Biotechnol. **1:** 201–204.

JANSEN, R. C., 1993 Interval mapping of multiple quantitative trait loci. Genetics **135:** 205–211.

JANSEN, R. C., and P. STAM, 1994 High resolution of quantitative traits into multiple loci via interval mapping. Genetics **136:** 1447–1455.

JENSEN, J., 1989 Estimation of recombination parameters between a quantitative trait locus (QTL) and two marker loci. Theor. Appl. Genet. **78:** 613–618.

KHATIB, H., A. DARVASI, Y. PLOTSKI and M. SOLLER, 1994 Determining relative microsatellite allele frequencies in pooled DNA samples. PCR Methods Applic. **4:** 13–18.

KNAPP, S. J., W. C. BRIDGES and D. BIRKES, 1990 Mapping quantitative trait loci using molecular marker linkage map. Theor. Appl. Genet. **79:** 583–592.

LANDE, R., 1981 The minimum number of genes contributing to quantitative variation between and within populations. Genetics **99:** 541–553.

LANDER, E. S., and D. BOTSTEIN, 1989 Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics **121:** 185–199.

LEBOWITZ, R. J., M. SOLLER and J. S. BECKMANN, 1987 Trait-based analyses for the detection of linkage between marker loci and quantitative trait loci in crosses between inbred lines. Theor. Appl. Genet. **73:** 556–562.

LISITSYN, N., 1994 Direct isolation of polymorphic markers linked to a trait by genetically directed representational difference analysis. Nat. Genet. **6:** 946–951.

MCMILLAN, I., and A. ROBERTSON, 1974 The power of methods for detection of major genes affecting quantitative characters. Heredity **32:** 349–356.

MICHELMORE, R. W., I. PARAN and R. V. KESSELI, 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., A. SAJANTILA and A. C. SYVANEN, 1993 Determination of allele frequencies at loci with length polymorphism by quantitative analysis of DNA amplified from pooled samples. PCR Methods Applic. **2:** 313–362.

SAX, K., 1923 Association of size difference with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. Genetics **8:** 552–560.

SOLLER, M., and J. S. BECKMANN, 1990 Marker-based mapping of quantitative trait loci using replicated progenies. Theor. Appl. Genet. **80:** 205–208.

SOLLER, M., T. BRODY and A. GENIZI, 1979 The expected distribution of marker-linked quantitative effects in crosses between inbred lines. Heredity **43:** 179–190.

THODAY, J. M., 1961 Location of polygenes. Nature **191:** 368–370.

THOMPSON, J. N., 1975 Quantitative variation and gene number. Nature **258:** 665–668.

WELLER, J. I., 1986 Maximum likelihood techniques for the mapping and analysis of quantitative trait loci with the aid of genetic markers. Biometrics **42:** 627–640.

ZENG, Z. B., 1993 Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. Proc. Natl. Acad. Sci. USA **90:** 10972–10976.

ZENG, Z. B., 1994 Precision mapping of quantitative trait loci. Genetics **136:** 1457–1468.

Communicating editor: Z.-B. ZENG