MAPPING RESOLUTION OF QUANTITATIVE TRAIT LOCI IN ADVANCED INTERCROSSED GENERATIONS

A. Darvasi and M. Soller
Dept. of Genetics, The Alexander Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, 91904 Jerusalem, Israel.

SUMMARY

Advanced intercrossed generations (AIG) is an experimental design that can provide more accurate estimates of QTL map location by means of linkage analysis. AIG are produced by randomly and sequentially intercrossing a previous generation, which initially originated from an F_2 population. This provides increasing probability of recombination between any two loci. Consequently, the genetic length of the entire genome is stretched, providing a better mapping resolution. A 95% confidence interval of QTL map location of 20 cM in the F_2 is reduced to 3.7 cM after 8 additional generations (F_{10}), analyzing the same population size for the same QTL. AIG derived from crosses between known inbred lined may be an attractive resource for fine mapping of QTL.

INTRODUCTION

The continuing development of new classes of polymorphic DNA level genetic markers provides a firm basis for extending methods for detection and mapping of quantitative trait loci (QTL) to a wide variety of agricultural and experimental species. These methods are mainly based on regression analysis and maximum likelihood (Darvasi et al., 1993; Lander and Botstein, 1989; Knapp et al., 1990; Jensen, 1989; Haley and Knott, 1992; Weller, 1986)). However, these methods are not able to efficiently utilize the increasing ability to saturate a given chromosomal region with very closely spaced markers. With the usual F_2, BC, half-sib or full-sib experimental designs, and populations of reasonable size, even using an infinite number of markers, a QTL of moderate effect can be assigned to a map location in a rather broad chromosomal region, only (Darvasi et al., 1993). In the present study the use of advanced intercross generations (AIG), produced within a given population, by sequentially and randomly intercrossing the previous generations is proposed as a means of more effectively exploiting the large number of markers that can be obtained in a given chromosomal region, to provide more accurate estimates of the QTL map location.

THEORY

Advanced intercross generations (AIG) are produced from an F_2 population generated from two inbred lines assumed homozygous for alternative alleles at a series of QTL and marker loci. The AIG, F_3, F_4, F_5..., are sequentially produced by randomly intercrossing the previous generation. For QTL mapping purposes, only the final generation is phenotyped and genotyped, the previous generations are raised and reproduced, only.

Proportion of recombinants in advanced intercross generations

The expected proportion of recombinant haplotypes, between two loci, A and B, in the F_t generation, r_t, will equal:

$$r_t = r_{t-1} + \frac{1}{2} r(1-r_{t-1})^2 - \frac{1}{2} r r_{t-1}^2 = \frac{1}{2} r + r_{t-1}(1-r) \qquad (1)$$

where r, the proportion of recombination in the F_2 generation, is equal to 1/2 the proportion of crossover events per meiosis between the two loci. From this equation r_t can readily be derived, as a function of r alone, giving:

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

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:

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

QTL mapping accuracy

We have previously shown through simulation studies that the 95% confidence interval of QTL map location, using an infinite number of markers is a close approximation of the 95% confidence interval of the maximum likelihood estimate (MLE) of QTL map location obtained with moderately spaced markers (10-20 cM) (Darvasi et al., 1993). On this basis, we now use the concept of an infinite number of markers, in order to explore the effects of the AIG design on confidence interval of QTL map location.

Consider an F_2 population for which C, in proportion of recombination units, is the distance from a given QTL, to one end of the confidence interval of the QTL map location. On the assumption of an infinite number of markers, C will depend only on the confidence level, the population size and the QTL gene effect. Since these are unaffected by the number of intercross generations, after t generations of large sample sizes the same proportion of recombinant haplotypes, C, is expected to define the same confidence interval. However, as shown in equation (3) a given proportion of recombinant haplotypes, r_t, measured in the F_t generation corresponds to a much smaller proportion of recombinant haplotypes in the F_2 generation. Consequently, given a confidence interval C in the F_t generation, the corresponding confidence interval on the scale of the F_2 generation, C', will equal:

$$C' = \frac{2C}{t} \qquad (4)$$

To obtain the corresponding confidence interval M', in cM, which represents the mapping resolution, C' is transformed to cM using Haldane's mapping function (Haldane, 1919) and doubled to represent the total confidence interval length.

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, $r>0.1$, the expected proportion of recombinant haplotypes increases rapidly, and asymptotically approaches 0.5 by 10 to 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. Thus, when defining map distances in units of the proportion of recombinant haplotypes between two loci, AIG is equivalent to stretching the chromosome with each additional generation.

Figure 2 presents the expected value of the confidence interval (M') as a function of the number of generations, for several initial values of the confidence interval (M) in the founder F_2 population. 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, for example, a confidence interval of 20 cM in the F_2 is reduced to 3.7 cM after 8 additional generations, (F_{10}), but only to 1.8 cM after another 10 generations (F_{20}).

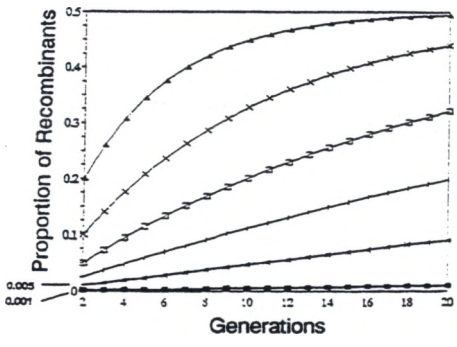

Figure 1: The proportion of recombinant haplotypes as a function of the number of the intercrossed generations.

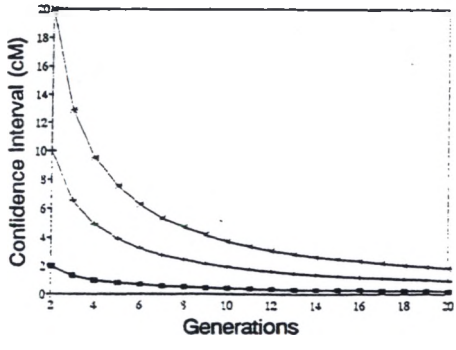

Figure 2: Confidence interval, in cM, as a function of the number of intercrossed generations.

DISCUSSION

The AIG design can significantly increase the accuracy of estimating QTL map location for a given number of individuals phenotyped and genotyped. This design is of particular interest for species that can be easily reproduced by intercrossing. Once the population reaches the desired advanced generation, a larger population can be produced for the genetic analysis, i.e., genotyping and phenotyping.

AIG will be particularly advantageous when phenotyping costs are relatively high, as it is the case for behavioral traits and some disease resistant traits. This will also hold for cases were a large number of traits are analyzed in the same population.

The AIG design is appropriate as a natural second stage in QTL mapping. In a first stage, say an F_2 analysis, a QTL can be located to a rather broad region of the chromosome. Locating the QTL to a more specific region of the chromosome by the use of AIG requires the availability of polymorphic markers in that region. At the present stage a relatively saturated genetic map exists in experimental species only for mice. The AIG design, therefore, presents a reasonable strategy where the time that the advanced generations are produced can be utilized for the search of genetic markers in the chromosomal regions of interest, as found in the F_2 analysis.

An AIG population can serve as a resource for QTL fine mapping. One can maintain a population consisting of the two parental lines and two advanced generations that are 10 generations apart one from the other. Once a population reaches the F_{10} generation, the older AIG is discarded and a new AIG is produced from the parental lines. Consequently, one will always have an AIG population of at least F_{10}. Since most of the accuracy increment is obtained in the first 10 generations, this schema provides an efficient long term genetic resource population for fine mapping any trait for which phenotypic differences are found in the original parental lines. Producing an AIG for a given pair of inbred lines, would be much less onerous than producing a set of recombinant lines, yet would be a much more powerful resource for purposes of gene mapping. Thus, it may be worthwhile to produce AIG between the most commonly used inbred lines of mice, as a general resource for the mapping community.

REFERENCES

DARVASI, A., WEINREB, A., MINKE, V., WELLER, J.I. AND SOLLER, M. (1993) Genetics, 134:943-951.

DARVASI, A. and SOLLER, M. (1992) Theor. Appl. Genet., 84:803-811.

HALDANE, J.B.S. (1919) J. Genet, 8:299-309.

HALEY, C.S. and KNOTT, S.A. (1992) Heredity, 69:315-324.

JENSEN, J. (1989) Theor. Appl. Genet., 78:613-618.

KNAPP, S.J., BRIDGES, W.C. and BIRKES, D. (1990) Theor. Appl. Genet., 79:583-592.

LANDER, E.S. and BOTSTEIN, D. (1989) Genetics, 121:185-199.

LEBOWITZ, R.J., SOLLER, M. and BECKMANN, J.S. (1987)Theor. Appl. Genet., 73:556-562.

WELLER, J.I. (1986) Biometrics, 42:627-640.