Genet. Res., Camb. (2002), **80**, *pp.* 231–236. *With* 1 *figure.* © 2002 Cambridge University Press
DOI: 10.1017/S001667230200589X Printed in the United Kingdom

A haplotypic approach to founder-origin probabilities and outbred QTL analysis

M. HUMBERTO REYES-VALDÉS[1]* AND CLAIRE G. WILLIAMS[2]
[1] *Departamento de Fitomejoramiento, Universidad Autónoma Agraria Antonio Narro, Buenavista, Saltillo, Coah, Mexico, C.P. 25315*
[2] *Faculty of Genetics, Texas A&M University, College Station, TX 77843-2135, USA*

(*Received 15 February 2002 and in revised form 10 June and 22 July 2002*)

Summary

Founder-origin probability methods are used to trace specific chromosomal segments in individual offspring. A haplotypic method was developed for calculating founder-origin probabilities in three-generation outbred pedigrees suited to quantitative trait locus (QTL) analysis. Estimators for expected founder-origin proportions were derived for a linkage group segment, an entire linkage group and a complete haplotype. If the founders are truly outbred, the haplotypic method gives a close approximation when compared with the Haley *et al.* (1994) method that simultaneously uses all marker information for QTL analysis, and it is less computationally demanding. The chief limitation of the haplotypic method is that some information in two-allele intercross marker-type configurations is ignored. Informativeness of marker arrays is discussed in the framework of founder-origin probabilities and proportions. The haplotypic method can be extended to more complex pedigrees with additional generations.

1. Introduction

An offspring genome within a three-generation pedigree can be viewed as a mosaic of chromosomal segments contributed by individual founders. Tracing the founder origin for each segment in a genomic mosaic can be used to relate allelic or haplotypic origin to phenotypic values. Founder-origin probabilities apply to chromosome points, inferring descent from a given founder. In pedigrees derived from two inbred lines it is straightforward to use markers to estimate founder-origin probabilities along each linkage group. Tracing founder origins becomes complex with outbred pedigrees for three reasons: (i) there are multiple, highly heterozygous founders, (ii) there can be twice as many alleles as the number of founders and (iii) marker-type configurations are heterogeneous for each linkage group.

For generalization to outbred pedigrees, founder-origin probabilities or proportions is more appropriate than using line-origin terminology. Estimating founder-origin probabilities is considered concomitant with mapping quantitative trait loci (QTLs) (Martínez & Curnow, 1992; Haley & Knott, 1992; Zeng, 1994). Such probabilities have been used to analyse three-generation outbred pedigrees for QTL mapping (Haley *et al.*, 1994; Knott *et al.*, 1997). Founder-origin probabilities can be combined with multiple regression, maximum likelihood and IBD approaches for QTL analysis. They can also be incorporated into a mixed model for interval mapping (Nagamine & Haley, 2001).

We developed a haplotype-based method for estimating founder-origin probabilities, expected founder-origin proportions and their conditional variances. Rather than using all marker information simultaneously (Haley *et al.*, 1994), the haplotypic method uses adjacent fully informative haplotypic markers for each individual offspring in reference to a given founder.

2. Genetic model

(i) *Founder-origin probabilities and proportions using the haplotypic method*

Founder-origin probabilities apply to chromosome points, inferring descent from a given founder. A

* Corresponding author.

founder-origin proportion is expressed as the ratio of a founder genome in a chromosome segment, an entire chromosome or an entire haplotype for a given individual offspring, relative to its total map length. Founder-origin estimations have been extended to the outbred case from a model for marker-based introgression through backcrossing (Reyes-Valdés, 2000). A short description of this model using a two-allele case follows.

In the classic backcross design, a recurrent parent is crossed with a donor parent. Their F_1 offspring is crossed to the recurrent parent, generating the first backcross (BC_1). Assume that selected individuals possess the entire set of marker alleles present in the recurrent parent together with the marker alleles associated with the segments(s) to be introgressed from the donor parent. Depending on the target genome chosen for selection, there are two marker-locus classes: recurrent markers (**R**) and donor markers (**D**). The two marker-locus classes together with the chromosome ends (**E**) delineate three chromosome landmarks: chromosome ends, recurrent markers and donor markers. These three landmark types in turn define six intervals: end–end, end–recurrent, end–donor, recurrent–recurrent, recurrent–donor and donor–donor.

Functions for calculating the probability of donor genome at any point within any of those six intervals for a given backcross generation are provided by Reyes-Valdés (2000). Those formulae, adjusted for BC_1 and absence of chiasma interference, are presented in Table 1. They are represented by $f_T(x)$, where T denotes a pair of contiguous chromosome landmarks (**R**, **D** or **E**) neighbouring a chromosome position x. If x coincides with the position of a marker of type **R** or **D**, respectively, then we have $f_R(r)=0$ and $f_D(d)=1$. For simplicity, the probability of donor genome at the map point x is defined as $g(x)=f_T(x)$.

The expectation $E(G)$ of the proportion G of donor genome in a linkage group is defined as follows:

$$E(G)=\frac{1}{L}\int_0^L g(x)\,\mathrm{d}x, \tag{1}$$

where L is the map length of the chromosome or chromosome segment in morgans.

The proportion G of donor genome has the following conditional variance (V_G) given an array of donor and recurrent markers:

$$V_G=\frac{1}{L^2}\int_0^L\left[g(x)\int_0^L g(y|x)\,\mathrm{d}y\right]\mathrm{d}x-E(G)^2, \tag{2}$$

where $g(y|x)$ is the probability of donor genome at the map position y given donor genome at the map position x. Note that $g(y|x)$ is identical to $g(y)$ if the positions x and y belong to different intervals defined by fully informative markers. Because analytical solutions are not available for all situations, a numerical method must be used to solve the integrals. To obtain the expected proportion of donor genome at the entire haplotype level, an average proportion is calculated across linkage groups then weighted by the respective map length of each linkage group. More detailed discussion has been published on donor genome introgression (Stam & Zeven, 1981; Reyes-Valdés, 2000). Similarly, Visscher (1996) developed an approach based on selection index theory to calculate the variance of the proportion of donor genome explained by markers using regression theory in the framework of the classical selection index (Hazel, 1943).

Table 1. *Formulae to calculate the probability of donor genome in a given position x of a chromosome in non-recurrent parent gametes forming the first backcross*

Interval	Formula
end–end	$f_{EE}(x)=\frac{1}{2}$
end–recurrent	$f_{ER}(x)=\delta(\lvert x-r\rvert)$
end–donor	$f_{ED}(x)=[1-\delta(\lvert x-d\rvert)]$
recurrent–recurrent	$f_{RR}(x)=\dfrac{\delta(\lvert x-r1\rvert)\delta(\lvert x-r2\rvert)}{[1-\delta(\lvert r1-r2\rvert)]}$
recurrent–donor	$f_{RD}(x)=\dfrac{\delta(\lvert r-x\rvert)[1-\delta(\lvert x-d\rvert)]}{\delta(\lvert d-r\rvert)}$
donor–donor	$f_{DD}(x)=1-\dfrac{\delta(\lvert x-d1\rvert)\delta(\lvert x-d2\rvert)}{1-\delta(\lvert d1-d2\rvert)}$

Absence of interference is assumed. The symbols r and d represent the positions of a donor (**D**) and a recurrent (**R**) marker locus, respectively, whereas δ is the Haldane's (1919) inverse mapping function.

(ii) *Extending the haplotypic method for founder-origin probabilities to a three-generation outbred pedigree*

A case study of an outbred three-generation pedigree with four founders is viewed here as a set of backcross experiments. Let A_1, A_2, A_3 and A_4 be the four grandparents or founders of the pedigree. The parents are defined as P_{12} and P_{34} with the subscripts denoting their respective founders. The individual offspring are $O_1, O_2, \ldots, O_n$, where n is their sibship size (Fig. 1). Consider founder A_1 as the donor parent in a backcross experiment so that A_2, A_3 and A_4 are collectively viewed as the recurrent parent. Thus P_{12} is the F_1 in our virtual backcross and $O_1, O_2, \ldots, O_n$ are the first backcross generation. Founders A_2, A_3 and A_4 are not assumed to be genetically identical.

Assume that a fully informative marker allele for offspring O_i is inherited from P_{12} and that it is known

https://doi.org/10.1017/S001667230200589X Published online by Cambridge University Press

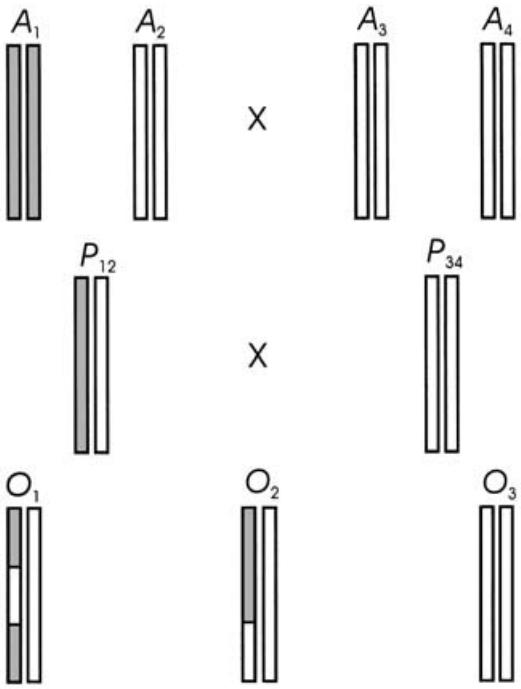

Fig. 1. Schematic representation of chromosome states in a three-generation pedigree, in which A_1, the reference founder, is viewed as a donor parent in a backcross experiment. The parent P_{12} is regarded as an F_1, whereas the progeny members O_i are considered as members of a backcross 1. In the same fashion, each founder can be viewed, one at a time, as a donor parent and a model for marker-based introgression applied to calculate probabilities of founder origin in the offspring individuals.

whether the marker allele descended from the founder A_1. Furthermore, if A_1 contributed the allele then it is classified as a donor (**D**) marker. If A_1 did not contribute the allele then the marker locus is classified as a recurrent (**R**) marker. If chromosome ends (**E**) are included as landmarks, a chromosome for the reference individual may have, for example, an array of marker types designated as [**E D R E**]. Using the backcross model, the probability of any given x contributed by founder A_1 can be calculated for this individual offspring O_i. The same process is conducted with reference to each of the other three founders. Note that there are only two possible independent probabilities for a given chromosome site: let $g(x, A_i)$ be the probability that a given genome site x in a gamete from P_{ij} came from founder A_i; thus $g(x, A_2) = 1 - g(x, A_1)$, and $g(x, A_4) = 1 - g(x, A_3)$.

(iii) *Identification of fully informative haplotypic markers and their definition as donor or recurrent markers*

Fully informative markers for each individual offspring must be identified and then defined either as donor (**D**) or recurrent markers (**R**). Let $\mathbf{P}_{ij}$ be a set of marker alleles for a given locus of P_{ij} and $N(\mathbf{P}_{ij})$ its number of elements. Note that $N(\mathbf{P}_{ij}) = 2$ for a heterozygous parent and $N(\mathbf{P}_{ij}) = 1$ for a homozygous parent. The same notation will be used for other members of the pedigree.

Let $\mathbf{S}_i$ be the set of alleles that could have been inherited by an offspring from parent P_{ij}. When considering A_1 as a donor parent, the process to classify a marker locus starts with finding $\mathbf{S}_1$ as follows:

$$\mathbf{S}_1 = \begin{cases} \mathbf{P_{12}} \cap \mathbf{O}_k \text{ when } N(\mathbf{P_{12}} \cap \mathbf{O}_k) = 1 \\ \mathbf{O}_k - \mathbf{P_{34}} \text{ when } [N(\mathbf{P_{34}} \cap \mathbf{O}_k) = 1] \\ \quad \text{and } [N(\mathbf{P_{12}} \cap \mathbf{O}_k) \neq 1] \\ \mathbf{P_{12}} \text{ otherwise} \end{cases} \tag{3}$$

Next, it is determined whether an allele was contributed by founder A_1. Let $\mathbf{S}_1^*$ be the set of alleles that parent P_{12} could have inherited from founder A_1:

$$\mathbf{S}_1^* = \begin{cases} \mathbf{A_1} \cap \mathbf{P_{12}} \text{ when } N(\mathbf{A_1} \cap \mathbf{P_{12}}) = 1 \\ \mathbf{P_{12}} - \mathbf{A_2} \text{ when } [N(\mathbf{A_2} \cap \mathbf{P_{12}}) = 1] \\ \quad \text{and } [N(\mathbf{A_1} \cap \mathbf{P_{12}}) \neq 1] \\ \mathbf{A_1} \text{ otherwise} \end{cases} \tag{4}$$

The necessary and sufficient conditions to consider a marker locus to be fully informative on a haplotype basis are (i) $N(\mathbf{S_1}) = N(\mathbf{S_1^*}) = 1$, and (ii) $N(\mathbf{P_{12}}) = \mathbf{2}$. Once the conditions are fulfilled, $\mathbf{S}_1^* = \mathbf{S}_1$ will indicate that the marker locus is a donor marker. On the other hand, if $\mathbf{S}_1^* \neq \mathbf{S}_1$, the marker locus is defined as a recurrent marker. This decision process is conducted for every locus in every offspring with reference to founders A_1 and A_3. For the latter case (A_3), the sets are:

$$\mathbf{S}_3 = \begin{cases} \mathbf{P_{34}} \cap \mathbf{O}_k \text{ when } N(\mathbf{P_{34}} \cap \mathbf{O}_k) = 1 \\ \mathbf{O}_k - \mathbf{P_{12}} \text{ when } [N(\mathbf{P_{12}} \cap \mathbf{O}_k) = 1] \\ \quad \text{and } [N(\mathbf{P_{34}} \cap \mathbf{O}_k) \neq 1] \\ \mathbf{P_{34}} \text{ otherwise} \end{cases} \tag{5}$$

$$\mathbf{S}_3^* = \begin{cases} \mathbf{A_3} \cap \mathbf{P_{34}} \text{ when } N(\mathbf{A_3} \cap \mathbf{P_{34}}) = 1 \\ \mathbf{P_{34}} - \mathbf{A_4} \text{ when } [N(\mathbf{A_4} \cap \mathbf{P_{34}}) = 1] \\ \quad \text{and } [N(\mathbf{A_3} \cap \mathbf{P_{34}}) \neq 1] \\ \mathbf{A_3} \text{ otherwise} \end{cases} \tag{6}$$

When marker allele data are missing, due either to dominance or to scoring failure, there may be several possible set definitions for one or more members of the pedigree. For instance, O_k can have possible set definitions $\mathbf{O}_{k1}$ and $\mathbf{O}_{k2}$. In this case, a marker locus will be considered fully informative if and only if for both possibilities the set-based algorithm can consistently identify the marker as either donor or recurrent. Total or partial genotypic reconstruction in founders and parents is needed to provide set definitions. This genotypic reconstruction can be aided by a program written during this research in Mathematica

https://doi.org/10.1017/S001667230200589X Published online by Cambridge University Press

language (Wolfram, 1999) that generates sets of possible parental marker-genotype doublets with respect to each progeny member. The genotypic reconstruction is based on Mendelian inheritance testing, selecting only those parental genotype combinations that are consistent with the whole offspring. The data analysis is then followed by another Mathematica routine that checks combinations of four founders, two parents and one offspring. The routine works as follows: (i) all possible marker-genotype possibilities are considered for each given combination of founders, parents and one offspring separately, and the resulting genotypes are sent to the next step; (ii) all possible Mendelian crosses in the mating combinations are performed, considering all potential genotypes; (iii) genotype possibilities with Mendelian inconsistencies are sequentially eliminated, thus reconstructing offspring marker genotypes; and (iv) consistent combinations are tested and informative loci are selected by the set-based algorithm, and defined as either donor or recurrent markers. If the consistent combinations give different solutions, the locus under analysis is treated as uninformative.

Genotype reconstruction and Mendelian inheritance checks are necessary when missing data occur or some dominant markers are used. However, the sole use of dominant markers is not recommended for the application of the haplotypic method because only recessive marker genotypes can be considered informative in the offspring in order to estimate founder-origin probabilities. An exception to this situation is the parental backcross configuration where one parent is a heterozygote and the other is a recessive homozygote. In this case both possible marker phenotypes in the progeny can be used to trace founder origin.

The resulting data for each offspring haplotype and linkage group will be an array of marker-type configurations and positions with reference to each founder. The following three-marker array is given as an example:

$$\begin{bmatrix} \mathbf{E} & \mathbf{R} & \mathbf{D} & \mathbf{R} & \mathbf{E} \\ 0 & 0 & 0{\cdot}1 & 0{\cdot}8 & 0{\cdot}8 \end{bmatrix},$$

where the first row is the marker array along a chromosome. The second row is the respective array of positions of those marker loci in the linkage group expressed in morgans. For each offspring there are two such arrays and it is possible for the two arrays to have a different set of informative markers.

For a given marker locus, let A_iA_j indicate that an offspring inherited one allele from founder A_i and the other allele from A_j. If male and female haplotypes are considered simultaneously, one or more founder-origin combinations can be regarded as possible, e.g. $\{A_1A_3, A_1A_4\}$. However, the haplotypic method ignores marker information indicating the sets of possible founder-origin combinations $\{A_1A_3, A_2A_4\}$ and $\{A_1A_4, A_2A_3\}$. This will happen when the marker locus has a parental two-allele intercross configuration and the offspring is heterozygous, making the haplotypic method approximate rather than exact. Thus, the haplotypic method is not suitable for a classical intercross between two inbred lines, because F_2 heterozygotes are considered to be uninformative. It is best suited to outbred pedigrees with a high proportion of informative multiple-allele codominant markers.

(iv) *Estimation of founder-origin probabilities from fully informative haplotypic markers*

Once the arrays of donor and recurrent markers have been defined for each individual offspring, the probabilities of donor genome can be estimated for any given point of a linkage map with reference to each founder (Table 1). A unit probability would indicate that the gamete contributing to the given individual carried donor genome at the point x, which means that such a point is located within a genomic segment descending from the reference founder. Conversely, a probability value of 0 indicates that the chromosomal segment including point x was not contributed by the reference founder.

Using the haplotypic method, probabilities are calculated from fully informative, flanking haplotypic markers or a single haplotypic marker when searching beyond markers at the linkage group end. The haplotypic method can be extended to estimating the proportion of donor genome in a chromosome segment or an entire chromosome (1). Conditional variances can also be estimated for these proportions (2).

When combining probability information from two haplotypes or individuals, e.g. calculating multiple founder-origin or IBD probabilities, the process becomes multipoint if different sets of adjacent informative markers are used for each haplotype.

(v) *Use of founder-origin probabilities in QTL mapping*

General agreement is expected between the haplotype-based method and the simultaneous use of all marker information for QTL analysis (Haley *et al.*, 1994; Knott *et al.*, 1997) if there are few markers in two-allele intercross configuration. Founder-origin probabilities are calculated at regular intervals along a linkage group in the case of regression-based QTL analysis. These estimates are then used as independent variables with phenotypic values as the dependent variable in regression analyses. If we assume that one founder is fixed for a QTL allele that is different from those QTL alleles in the three remaining founders, a simple linear regression can be performed along the

https://doi.org/10.1017/S001667230200589X Published online by Cambridge University Press

linkage group on the founder-origin probability, following Martinez & Curnow (1992). A significant peak for the test statistic indicates a putative QTL and the regression coefficient provides an estimate of additive or haplotypic effect. A threshold for the test statistic can be obtained by the permutation test (Churchill & Doerge, 1994) and confidence intervals for map positions can be obtained by bootstrap (Visscher *et al*., 1996).

Another possible configuration for founders is given, in terms of QTL genotypes as: A_1: Q_1Q_1, A_2: Q_2Q_2, A_3: Q_1Q_1, A_4: Q_2Q_2. In this case, multiple regression is applied to probabilities of haplotypic combinations. Founder-origin probabilities can be multiplied to obtain the probabilities of haplotype combinations if only fully informative haplotypic markers are used. The variable combinations needed to obtain approximate coefficients to estimate additive (α) and dominance (θ) effects are as follows:

$$\alpha\text{: } g(x, A_1)g(x, A_3) - g(x, A_2)g(x, A_4), \tag{7}$$

$$\theta\text{: } g(x, A_1)g(x, A_4) + g(x, A_2)g(x, A_3). \tag{8}$$

When more than two QTL alleles may be segregating in the mapping population, a model for maternal, paternal and maternal × paternal effects can be fitted (Knott *et al*., 1997). In this case, approximate probabilities of founder-origin combinations are generated by simple multiplication of haplotypic founder-origin probabilities.

(vi) *Informativeness of a marker array*

The marker information content in linkage maps is important for relating haplotypic origin to phenotypic values. Knott *et al*. (1997) define an informativeness statistic (Ψ) for average map point information based on the offspring of a three-generation outbred pedigree. This statistic can be extended to other types of mapping populations. With reference to founders A_1 and A_2, average informativeness or Ψ, at map position x, is expressed as follows (Knott *et al*., 1997):

$$\Psi = \frac{1}{n}\sum_{i=1}^{n} 4[g_i(x, A_1) - 0{\cdot}5]^2. \tag{9}$$

Each term in the summation is:

$$\begin{aligned} 4[g_i(x, A_1) - 0{\cdot}5]^2 &= 4[g_i(x, A_1)^2 - g_i(x, A_1) + 0{\cdot}25] \\ &= [0{\cdot}25 - g_i(x, A_1)(1 - g_i(x, A_1))]/0{\cdot}25. \end{aligned}$$

The value of 0·25 shown within brackets is the variance of a Bernoulli random variable that equals 1 when the given haplotype carries genome from A_1 with a probability of 0·5 at the chromosome point x, and equals 0 otherwise. Therefore, the 0·25 value is the maximum conditional variance of A_1 founder genome at a given genome location. This value describes the situation for unmarked chromosomes. The product $g(1-g)$ is the variance of a Bernoulli variable with parameter g, which in this case is the probability of A_1 calculated with marker information. Thus, the informativeness statistic of Knott *et al*. (1997) measures the reduction in conditional variance due to the use of markers to calculate a probability of founder origin in a given map location.

Within the framework of the polygenic model for QTL analysis (Knott *et al*., 1997), the informativeness of an individual haplotypic marker array, at the linkage group level, can be calculated in an analogous way:

$$H = \frac{V_T - V_G}{V_T}, \tag{10}$$

where V_G is the conditional variance of a founder-origin proportion (2) and V_T is the total variance of the proportion of founder origin in a linkage group calculated with the same function (2) in the case of an unmarked chromosome (end–end). Since $V_T - V_G$ is the variance explained by the genetic markers, the expectation E[H] is equivalent to the parameter R^2 used by Visscher (1996) for the proportion of the variance in genetic composition of backcross generations explained by markers. The procedure to calculate E[H] differs from the calculation of R^2, where a linear relationship between marker indices and the proportion of non-recurrent genome is established in the framework of selection indices (Visscher, 1996). Here variance calculations are based on numerical integration to obtain E[H].

As an example, the optimum marker distribution of two marker loci for QTL analysis under the polygenic model (Knott *et al*., 1997) was analysed by maximizing the expected index E[H]. It was calculated as the sum of the products of the probability of each phase (coupling and repulsion) by the respective index H. The expectation E[H] was calculated for different combinations of marker positions. The best marker array was two marker loci symmetrically located around the centre of the linkage map, with 45 cM of distance for a linkage group of 100 cM so that the markers are at 27·5 and 72·5 cM. These positions are identical to the values reported by Visscher (1996) for the optimum marker location in backcross breeding.

The index H is bounded by the limits 0 and 1, where a value of 0 shows absence of marker information as in the case of an unmarked chromosome. A value of 1 shows the maximum possible to be attained. As the marker saturation increases, H approaches 1. Extension of this measure to haplotypic pairs rather than haplotypes themselves requires the use of the variances

https://doi.org/10.1017/S001667230200589X Published online by Cambridge University Press

of founder-origin combinations, not founder-origin variances for single haplotypes.

3. Concluding remarks

The haplotype-based method is effective for approximating founder-origin probabilities in the case of a three-generation outbred pedigree with multiple allele marker loci. The haplotypic method can be used to calculate expectations of founder-origin proportions for chromosome segments, entire linkage groups and an entire haplotype in addition to their conditional variances. Multiplication of haplotypic founder-origin probabilities provides approximate probabilities of haplotype combinations. The haplotypic method is less computationally demanding than the all-marker based approach, with the loss of information directly related to the proportion of two-allele intercross marker-type configurations.

The haplotypic method for founder-origin probabilities can be extended to more complex inbred or outbred pedigrees. The general type of pedigree that can be extended using the haplotypic method is the so-called zero-loop pedigree (Cannings *et al*., 1978), which is a tree of individuals and marriages. If one of the parents of an individual offspring is also a founder itself then the founder-origin probability for the haplotype coming from that founder is equal to 1. If the founder is a grandparent, then the haplotypic method as described in this study applies because each offspring is treated as a first-generation backcross. If the founder is a great-grandparent, the offspring member must be treated as a second-generation backcross using multi-generation backcross equations, or recurrently using first-generation backcross formulae. However, efficient allele or haplotype tracing from offspring to founders requires codominant, multiple-allele markers.

We thank Dr David Gwaze for his excellent suggestions and applications of this model, and two anonymous reviewers who greatly enhanced the quality of this paper. This work was sponsored by the US Department of Energy Grant #DE-FC07-00ID13877 (to C.G.W.) and Universidad Autónoma Agraria Antonio Narro, Mexico (M.H.R.V.). All the calculations were programmed in the Mathematica system of Wolfram Research, Inc.

References

Cannings, C., Thompson, E. A. & Skolnick, M. H. (1978). Probability functions on complex pedigrees. *Advances in Applied Probability* **10**, 26–61.

Churchill, G. A. & Doerge, R. W. (1994). Empirical threshold values for quantitative trait mapping. *Genetics* **138**, 963–971.

Haldane, J. B. S. (1919). The combination of linkage values, and the calculation of distances between the loci of linked factors. *Journal of Genetics* **8**, 299–309.

Haley, C. S. & Knott, S. A. (1992). A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* **69**, 315–324.

Haley, C. S., Knott, S. A. & Elsen, J. M. (1994). Mapping quantitative trait loci in crosses between outbred lines using least squares. *Genetics* **136**, 1195–1207.

Hazel, L. N. (1943). The genetic basis for constructing selection indices. *Genetics* **28**, 476–490.

Knott, S. A., Neale, D. B., Sewell, M. M. & Haley, C. S. (1997). Multiple marker mapping of quantitative trait loci in an outbred pedigree of loblolly pine. *Theoretical and Applied Genetics* **94**, 810–820.

Martínez, O. & Curnow, R. N. (1992). Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. *Theoretical and Applied Genetics* **85**, 480–488.

Nagamine, Y. & Haley, C. S. (2001). Using the mixed model for interval mapping of quantitative trait loci in outbred line crosses. *Genetical Research* **77**, 199–207.

Reyes-Valdés, M. H. (2000). A model for marker-based selection in gene introgression breeding programs. *Crop Science* **40**, 91–98.

Stam, P. & Zeven, A. C. (1981). The theoretical proportions of the donor genome in near-isogenic lines of self-fertilizers bred by backcrossing. *Euphytica* **30**, 227–238.

Visscher, P. M. (1996). Proportion of the variation in genetic composition in backcrossing programs explained by genetic markers. *Journal of Heredity* **87**, 136–138.

Visscher, P. M., Thompson, R. & Haley C. S. (1996). Confidence intervals in QTL mapping by bootstrapping. *Genetics* **143**, 1013–1020.

Wolfram, S. (1999). *The Mathematica book*. Champaign, IL: Wolfram Media/Oakleigh, Melbourne: Cambridge University Press.

Zeng, Z. B. (1994). Precision mapping of quantitative trait loci. *Genetics* **136**, 1457–1468.