The calculation of recombination frequencies in crosses of allogamous plant species with applications to linkage mapping

E. RITTER[1] AND F. SALAMINI[2]*

[1] *CIMA-Arkaute, Apartado 46, E-01080 Vitoria-Gasteiz, Spain*
[2] *Max-Planck-Institute for Plant Breeding Research, Carl-von-Linne-Weg 10, D-50829 Köln, Germany*

Summary

The recombination frequency (r) between two loci defined by conventional or molecular markers can be estimated by solving proper Maximum Likelihood equations. These are based on expected and observed marker class frequencies in the progeny of a cross, and are specific for each allelic configuration of the parents[1]. In a cross between two diploid parents up to four different alleles, besides a null allele, can be detected at one locus. This defines in each parent, considering a locus A, nine basic allelic configurations based on two allelic marker fragments (A_i/A_j), one single marker allele and a null allele $(A_i/A0)$, or just null alleles $(A0/A0)$. With respect to two loci the consideration of all possible diploid allelic configurations in the parents of a cross allows the detection of 21 different expected marker class distributions producing estimates of r in the progeny. General formulas for calculating the ML equations and the corresponding information functions have been developed for the 21 marker class distributions. Simplified formulas have been also derived and the relative efficiency of the information functions compared. As expected, in the majority of cases, allelic marker configurations give more precise estimates of linkage values than single marker configurations. A method for the construction of linkage maps based on two point estimates, linkage subgroups and allelic bridges is presented. The method is an improvement on an original proposal by Ritter et al. (1990).

1. Introduction

Molecular markers have led to the development of genetic linkage maps in numerous plant and animal species. In mapping, pairwise linkage tests are performed and recombination frequencies between segregating markers are calculated. Linked polymorphisms are clustered and aligned in linkage groups depending on their recombination frequencies. Formulas for calculating recombination frequencies between loci

---

* Corresponding author.
(1) In this paper the term allelic configuration describes in the parents of a cross both the existence of different alleles at two linked loci and the phase of coupling or repulsion of specific alleles. In this sense, when symbols for alleles at two loci follow each other on the same line (e.g. A1B1) they are considered to be linked in coupling. Allelic configurations, as above defined, are different from the concept of a 'marker class' (Ritter et al, 1990 and see also in this paper). In this case, the class defines the phenotype, with respect to marker alleles, of a group of genotypes with the same marker alleles in common, found in the progeny of a cross (Example: the marker class A1A2B1B2 can include individual genotypes having the configurations A1B1/A2B2, A2B1/A1B2, A1B2/A2B1 and A2B2/A1B1).

segregating with different allelic configurations (1) in the parents have been provided by Fisher (1921), Fisher & Balmakund (1928), Mather (1938), Allard (1956), Bailey (1961), and Ritter et al. (1990).

The computer program MAPMAKER was developed to map in F2s, in backcrosses and in other progenies derived from inbred lines (Lander et al. 1987). Additional and different types of progenies can be handled with programs like LINKAGE-1 (Suiter et al. 1983), LINKEM (Vowden & Ridout, 1994) or JOINMAP (Stam, 1993). The latter allows the combination of maps resulting from different datasets. Ritter et al. (1990) discussed linkage mapping in crosses between heterozygous parents. This approach led to the construction of linkage maps in potato (Gebhardt et al. 1989; 1991) and sugar beet (Barzen et al. 1992).

The present paper extends the approach of Ritter et al. (1990) by considering genotype formation more generally during mating, and analysing the frequency of the genotypes, represented by classes of markers, in the progenies resulting from all possible allelic

Table 1. *Mating table, expected marker classes in the progeny, frequency of each marker class and their derivatives for the cross with the parental allelic configuration A1B1/A2B2 × A3B3/A4B4*

GT(*)	GF(**)	Parental gametes and their frequency		Recombinant gametes and their frequency	
		A3B3 $\frac{1}{2}(1-r)$	A4B4 $\frac{1}{2}(1-r)$	A3B4 $\frac{1}{2}r$	A4B3 $\frac{1}{2}r$
Parental gametes		FT I(f) Marker classes		FT II(f) Marker classes	
A1B1	$\frac{1}{2}(1-r)$	A1A3B1B3	A1A4B1B4	A1A3B1B4	A1A4B1B3
A2B2	$\frac{1}{2}(1-r)$	A2A3B2B3	A2A4B2B4	A2A3B2B4	A2A4B2B3
gametes		FT II(f) Marker classes		FT III(f) Marker classes	
A1B2	$\frac{1}{2}r$	A1A3B2B3	A1A4B2B4	A1A3B2B4	A1A4B2B3
A2B1	$\frac{1}{2}r$	A2A3B1B3	A2A4B1B4	A2A3B1B4	A2A4B1B3

FT(f)	Expected frequency for each marker class	Derivative
I	$\frac{1}{4}(1-r)^2 = t1^{(+)}$	$\frac{1}{2}(r-1) = t1'^{(v)}$
II	$\frac{1}{4}(r-r^2) = t2^{(+)}$	$\frac{1}{4}(1-2r) = t2'^{(v)}$
III	$\frac{1}{4}r^2 = t3^{(+)}$	$\frac{1}{2}r = t3'^{(v)}$

(*) GT = Gametes.
(**) GF = Gamete frequency.
(f) FT = Type of expected frequency.
  I = no recombination in both parents.
  II = recombination in one parent.
  III = recombination in both parents.
(+)(v) see also formula (3).

Table 2. *Marker class phenotypes (I–IV) and corresponding genotypes in the progeny, depending on the allelic configurations of the parents at one locus A*

Allelic configuration of parents		Marker class phenotypes and, in brackets, the genotypes of the progeny			
Name of configuration	Description	I	II	III	IV
One-allelic configuations					
A1/A0 × A0/A0	One individual(*) allele and three null alleles	A1 (A1/A0)	A0 (A0/A0)		
A1/A0 × A1/A0	One common marker(*) allele and two null alleles	A1 (A1/A1) (A1/A0)	A0 (A0/A0)		
Two-allelic configurations					
A1/A2 × A0/A0	Two allelic markers in one parent, two null alleles in the other	A1A0 (A1/A0)	A2A0 (A1/A0)		
A1/A2 × A1/A0	One marker allele common to both parents, one allelic individual marker in one parent and a null allele in the other	A1A2 (A1/A2)	A1A0 (A1/A1) (A1/A0)	A2A0 (A2A0)	
A1/A2 × A1/A2	Two allelic markers common to both parents	A1A2 (A1/A2)	A1A0 (A1/A1)	A2A0 (A2A2)	
A1/A0 × A2/A0	One individual marker and a null allele in each parent	A1A2 (A1/A2)	A1A0 (A1/A0)	A2A0 (A2/A0)	A0A0 (A0/A0)
Three-allelic configurations					
A1/A2 × A3/A0	Two allelic markers in one parent, a different individual marker and a null allele in the other parent	A1A3 (A1/A3)	A1A0 (A1/A0)	A2A3 (A2/A3)	A2A0 (A2/A0)
A1/A2 × A1/A3	Three allelic markers, one is common to both parents	A1A1 (A1/A1)	A1A3 (A1/A3)	A2A1 (A2/A1)	A2A3 (A2/A3)
Four-allel configuration					
A1/A2 × A3/A4	Four different allelic markers	A1A3 (A1/A3)	A1A4 (A1/A4)	A2A3 (A2/A3)	A2A4 (A2/A4)

(*) see the definitions given in the text.

Table 3. *Marker class distributions (MCD) in the progeny of crosses between diploid parents; all possible allelic configurations at two loci are considered*

A Progenies consisting of four marker class phenotypes (single marker configurations)

Case No(*): 1, 2, 3, 4, 5

Ex(**):

Case 1 — A1B1/A0B0 × A0B0/A0B0

C(†)	F(†)	M(†)
A1B1	2-2-0	m_{11}
A0B0	2-2-0	m_{12}
A1B0	0-2-2	m_{21}
A0B1	0-2-2	m_{22}

Case 2 — A1B1/A0B0 × A0B1/A0B0

C	F	M
A1B1	2-3-1	m_{11}
A0B1	1-3-2	m_{21}
A0B0	1-1-0	m_{31}
A1B0	1-1-1	m_{41}

Case 3 — A1B1/A0B0 × A0B1/A0B1

C	F	M
A1B1	2-4-3	m_{11}
A1B0	1-2-0	m_{21}
A0B1	1-2-0	m_{22}
A0B0	0-0-1	m_{31}

Case 4 — A1B0/A0B1 × A0B1/A0B1

C	F	M
A1B1	3-4-2	m_{11}
A0B1	0-2-1	m_{21}
A1B0	0-2-1	m_{22}
A0B0	1-0-0	m_{31}

Case 5 — A1B1/A0B0 × A1B0/A0B1

C	F	M
A1B1	2-5-2	m_{11}
A1B0	1-1-1	m_{21}
A0B1	1-1-1	m_{22}
A0B0	0-1-0	m_{31}

B Progenies consisting of six marker class phenotypes

Case 6 — A1B1/A0B2 × A0B1/A0B0

C	F	M
A1B1	2-2-0	m_{11}
A0B1B2	1-1-0	m_{21}
A0B2	1-1-0	m_{22}
A0B1	0-2-2	m_{31}
A1B1B2	0-1-1	m_{41}
A1B2	0-1-1	m_{42}

Case 7 — A1B1/A0B2 × A1B2/A0B0

C	F	M
A1B1B2	1-2-0	m_{11}
A1B1	1-1-1	m_{21}
A1B2	1-1-1	m_{22}
A0B0	1-0-0	m_{31}
A1B0	0-2-1	m_{41}
A0B2	0-1-1	m_{42}

Case 8 — A1B2/A0B1 × A1B1/A0B0

C	F	M
A1B1	1-3-2	m_{11}
A1A2B1B2	1-2-0	m_{21}
A2B2	1-1-0	m_{22}
A1B1B2	1-1-0	m_{31}
A1B2	0-1-1	m_{41}
A1A2B1	0-1-1	m_{51}
A2B1	0-1-1	m_{52}
A1A2B1	0-1-0	m_{61}
A2B1B2	0-1-0	

Case 9 — A1B1/A0B2 × A0B1/A0B2

C	F	M
A1B1B2	1-2-1	m_{11}
A0B1B2	1-2-1	m_{21}
A1B1	1-1-0	m_{31}
A0B2	1-1-0	m_{41}
A1B2	0-1-1	m_{51}
A0B1	0-1-1	m_{61}

Case 10 — A1B1/A0B2 × A0B1/A0B2

C	F	M
A1B1B2	2-2-2	m_{11}
A1B1	1-2-0	m_{12}
A0B2	1-0-0	m_{21}
A1B2	0-2-1	m_{22}
A0B1B2	0-2-0	m_{31}
A0B1	0-0-1	m_{32}

Case 11 — A1B2/A0B1 × A1B1/A0B2

C	F	M
A1B1B2	1-4-1	m_{11}
A1B2	1-1-1	m_{21}
A1B1	1-1-1	m_{31}
A0B1B2	1-0-1	m_{41}
A0B1	0-1-0	m_{51}
A0B2	0-1-0	m_{61}

C Progenies consisting of eight marker class phenotypes

Case 12 — A1B1/A0B0 × A0B2/A0B0

C	F	M
A1B1B2	1-1-0	m_{11}
A1B1	1-1-0	m_{12}
A0B2	1-1-0	m_{13}
A0B0	1-1-0	m_{14}
A1B2	0-1-1	m_{21}
A1B0	0-1-1	m_{22}
A0B1B2	0-1-1	m_{23}
A0B1	0-1-1	m_{24}

Case 13 — A1B2/A0B0 × A1B1/A0B0

C	F	M
A1B1B2	1-2-0	m_{11}
A1B1	1-1-1	m_{21}
A1B2	1-1-1	m_{31}
A0B0	1-0-0	m_{41}
A1B0	0-2-1	m_{51}
A0B2	0-1-1	m_{52}
A0B1	0-1-1	m_{61}
A0B1B2	0-1-1	m_{61}

Progenies consisting of nine marker class phenotypes

Case 14 — A1B1/A2B2 × A1B1/A0B0

C	F	M
A1B1	2-2-0	m_{11}
A1A2B1B2	1-0-0	m_{21}
A2B2	1-0-0	m_{22}
A1B1B2	0-1-1	m_{31}
A1B2	0-1-1	m_{32}
A1A2B1	0-1-1	m_{33}
A2B1	0-1-1	m_{34}
A1A2B2	0-1-0	m_{41}
A2B1B2	0-1-0	m_{42}

Case 15 — A1B2/A2B1 × A1B1/A0B0

C	F	M
A1B1	2-2-0	m_{11}
A1B2	1-0-0	m_{21}
A1A2B1	1-0-0	m_{22}
A2B1	0-1-1	m_{31}
A1B1B2	0-1-1	m_{32}
A1A2B1B2	0-1-1	m_{33}
A2B2	0-1-1	m_{34}
A1A2B2	0-1-0	m_{41}
A2B1B2	0-1-0	m_{42}

Case 16 — A1B1/A2B2 × A1B1/A0B2

C	F	M
A1B1B2	2-2-2	m_{11}
A1B1	1-2-0	m_{12}
A1A2B1B2	1-0-0	m_{13}
A2B2	1-0-0	m_{14}
A2B1B2	0-2-1	m_{21}
A1B2	0-2-0	m_{22}
A1A2B2	0-2-0	m_{23}
A1A2B1	0-0-1	m_{31}
A2B1	0-0-1	m_{32}

Case 17 — A1B1/A2B2 × A1B1/A2B2

C	F	M
A1A2B1B2	2-0-2	m_{11} (#)
A1B1	1-0-0	m_{21}
A2B2	1-0-0	m_{31}
A1B1B2	0-2-0	m_{41}
A1A2B1	0-2-0	m_{51}
A1A2B2	0-2-0	m_{61}
A2B1B2	0-2-0	m_{71}
A1B2	0-0-1	m_{72}
A2B1	0-0-1	m_{81}

Case 18 — A1B2/A2B1 × A1B1/A2B2

C	F	M
A1B1B2	1-0-1	m_{31}
A1A2B2	1-0-1	m_{32}
A1A2B1	1-0-1	m_{33}
A2B1B2	1-0-1	m_{34}
A1A2B1B2	0-4-0	
A1B2	0-1-0	
A2B1	0-1-0	
A2B2	0-1-0	
A0B1B2	0-1-0	

D Progenies of crosses consisting of twelve marker class phenotypes

Case 19 — A1B1/A2B0 × A1B2/A0B0

C	F	M
A1B1B2	1-1-0	m_{11}
A1B1	1-1-0	m_{12}
A1A2B2	1-0-0	m_{13}
A2B0	1-0-0	m_{14}
A1B2	0-1-1	m_{21}
A1B0	0-1-1	m_{22}
A1A2B1	0-1-1	m_{31}
A2B0	0-1-1	m_{32}

Case 20 — A1B1/A2B0 × A1B2/A2B0

C	F	M
A1A2B0	1-0-1	m_{41}
A1A2B2	1-0-1	m_{42}
A1B1B2	1-0-0	m_{43}
A2B0	1-0-0	m_{44}
A1A2B1B2	0-2-0	m_{51}
A1A2B0	0-2-0	m_{52}

Progeny of a cross consisting of sixteen marker class phenotypes

Case 21 — A1B1/A0B0 × A2B2/A0B0

C	F	M
A1A2B1B2	1-0-0	m_{41}
A1B1	1-0-0	m_{42}
A2B2	1-0-0	m_{43}
A0B0	1-0-0	m_{44}
A1A2B1	0-1-1	m_{51}
A1B1B2	0-1-0	m_{52}

(*) case number of marker class distribution (see table 4); (**) Ex = example for the corresponding configuration type; (†) C = marker classes (for the example); F = frequency code (see text for details); (#) uninformative marker class.

M = marker class code; (#) uninformative marker class.

configurations in the parents of a cross, to derive estimates of r.

2. Methods and results

(i) *General concepts*

An unbiased estimator of the recombination frequency r between two segregating markers is obtained by solving the maximum likelihood equation for r (Fisher, 1921):

$$\frac{\delta ln L(r)}{\delta r} = \sum_j Z_j \frac{1}{p_j} \frac{\delta p_j}{\delta r} = 0 \qquad (1)$$

where p_j are the expected frequencies of each marker class and Z_j are the observed frequencies of these marker classes. The standard error *SE* of the estimate *R* for r is obtained from the variance $V(R)$ with:

$$SE = \sqrt{V(R)} = \frac{1}{\sqrt{I_r}} \text{ where } I_r = n \sum_j \frac{1}{p_j} \left(\frac{\delta p_j}{\delta r}\right)^2 \qquad (2)$$

and n represents the number of offspring . The information function I_r measures the quality of the estimate (Mather, 1938).

The calculation of the estimates can be facilitated using mating and calculation tables as described by Ritter *et al.* (1990). In the progeny of a cross, different marker classes are expected depending on the allelic configuration in the parents. Their corresponding frequencies can be derived from the mating tables. Three types of expected frequencies (FT) have been considered (in table 1 an example is given for the cross A1B1/A2B2 × A3B3/A4B4): Type I, where, in the progeny of the cross, the genotypes are formed by gametes which are non-recombinant in both parents; Type II, where recombination occurs in either one of the parents; and Type III, where recombination occurs in both parents. The expected frequency of each marker class is obtained by summing the frequencies with which a marker class occurs in the mating table. The number of marker classes, their expected frequencies and, therefore, the proper formula to be used to obtain an estimate R of r, depend on the specific allelic configuration of the parents.

(ii) *Allelic configurations*

When two diploid parents are considered, four different marker alleles can exist at a particular locus. These can be denoted by *A1* to *A4* for locus *A* and *B1* to *B4* for locus *B*. Null alleles can also exist, denoted by the symbols *A0* and *B0*. At a locus, either a single marker may segregate (e.g. one marker allele and a null allele are present in either one or both parents) or two allelic markers may segregate (both alleles of a locus are represented by different markers). When a single marker segregates there is a need to differentiate between the case when the marker is only present in one parent (we define this situation as a case of 'individual marker') or when the marker occurs in both parents ('common marker'). In the putative diploid parents of a cross, nine basic informative allelic configurations involving one to four allelic markers and null alleles can exist at a locus. They comprise one-allelic configurations [*A1/A0*(= *P1*) × *A0/A0*(= *P2*) and *A1/A0* × *A1/A0*], two-allelic configurations [*A1/A2* × *A0/A0*, *A1/A2* × *A1/A0*, *A1/A2* × *A1/A2* and *A1/A0* × *A2/A0*], three-allelic configurations [*A1/A2* × *A3/A0* and *A1/A2* × *A1/A3*] and one four-allele configuration [*A1/A2* × *A3/A4*]. Other configurations can be converted into these basic ones by renumbering the alleles, swapping their position on homologous chromosomes or by exchanging the order of the two parents when describing a specific cross. The basic configurations will generate two, three or four phenotypic marker classes in the progeny, which may be composed of different genotypes (see table 2). The same situation holds true for a second locus *B*, so that for the two loci *A* and *B*, progenies can be derived with four, six, eight, nine, twelve and sixteen different marker classes, depending on the genotype of the parents (see also table 3). Although only nine basic marker configurations exist at each locus, if the alleles present in the heterozygous state, in one or both parents, are linked in coupling or repulsion, different marker class distributions (MCD) are generated. Under conditions of random mating, these different marker class distributions can be individualized (table 3). The marker class distributions based only on a single marker allele at each locus *A* and *B* consist of four classes of marker phenotypes (table 3, part A) and are of particular interest for linkage mapping (see comments on fig. 1). The other marker class distributions consist of six, eight, twelve and sixteen marker classes, and are reported in tables 3B,C and D, respectively.

Mating tables have been computed for all possible allelic configurations in two diploid parents of a cross. For each allelic configuration the marker classes present in the progenies of the cross and their expected frequencies were derived. During the calculation of MCDs for all possible configurations, it became evident that several allelic configurations lead to identical MCDs. For instance, all crosses with allelic configurations generating four different genotypes at each locus (last four lines and last four columns in table 4), originate progenies with 16 marker class phenotypes (case number 21 in table 3). A total of 21 different marker class distributions (21 case numbers) exist, which are reported in table 3 from case 1 to case 21. (Each case corresponds to the allelic configuration(s) mentioned in table 4; for example, case No 1 can be obtained from the crosses A1B1/A0B0 × A0B0/A0B0, or A1B1/A0B2 × A0B0/A0B0,

Table 4. *Marker class distributions resulting from the allelic configurations at two loci in the parents of a cross. Two or more different allelic configurations may generate the same marker class distribution.The numbers of times that this occurs is given in brackets*

	LOCUS B								
PG(*)	2 P1 P2	2 P1 P2	2 P1 P2	3 P1 P2	3 P1 P2	4 P1 P2	4 P1 P2	4 P1 P2	4 P1 P2
LOCUS A PC(**) P1 P2	B1 B0 / B0 B0	B1 B1 / B0 B0	B1 B0 / B2 B0	B1 B1 / B2 B0	B1 B1 / B2 B2	B1 B2 / B0 B0	B1 B3 / B2 B0	B1 B1 / B2 B3	B1 B3 / B2 B4
	CASE NUMBER(f)								
2 A1 A0 / A0 A0	1 (2)	2 (2)	1 (2)	6 (2)	9 (2)	12 (2)	12 (2)	12 (2)	12 (2)
2 A1 A1 / A0 A0	—	3 4 5 c r cr(+)	2 (2)	7 8 (2) (2)	10 11 (2) (1)	13 (4)	13 (4)	13 (4)	13 (4)
2 A1 A0 / A2 A0	—	—	1 (2)	6 (2)	9 (2)	12 (2)	12 (2)	12 (2)	12 (2)
3 A1 A1 / A2 A0	—	—	—	14 15 (2) (2)	16 (4)	19 (4)	19 (4)	19 (4)	19 (4)
3 A1 A1 / A2 A2	—	—	—	—	17 18 (2) (1)	20 (4)	20 (4)	20 (4)	20 (4)
4 A1 A2 / A0 A0	—	—	—	—	—	21 (4)	21 (4)	21 (4)	21 (4)
4 A1 A3 / A2 A0	—	—	—	—	—	—	21 (4)	21 (4)	21 (4)
4 A1 A1 / A2 A3	—	—	—	—	—	—	—	21 (4)	21 (4)
4 A1 A3 / A2 A4	—	—	—	—	—	—	—	—	21 (4)

(*) PG = number of different genotypes with respect to one locus in the derived progeny.

(**) PC = allelic configuration for the loci *A* and *B* in the parents P1 and P2.

(f) The case numbers indicated here correspond to the 21 listed in table 3.

(+) c = both alleles common to the two parents are linked in coupling; r = alleles common to both parents are in repulsion; cr = alleles in coupling in one parent and in repulsion in the other.

or A1B1/A2B2 × A0B0/A0B0.) In table 3, one example of the possible allelic configurations is given for each case.

Each of the three digit code given, in table 3 (columns F), to each marker class (columns C) represents the times that the expected frequency of the specific class – defined in table 1 as FT I or II or III – has to be summed to calculate the total expected frequency of that marker class. For example, the code 2–3–1 indicates that a specific marker class appears 2 times with frequency FT I, 3 with FT II and 1 with FT III. The frequency code 1–2–1 is not informative because for this marker class the frequency types sum to a constant value of 1 for which the derivative would be zero. This marker class is omitted in the ML equation for calculating r by using formula (1).

A specific comment is needed to further clarify the relationship between MCDs and allelic configurations.

In fact, for some allelic configurations (five in total) up to three different MCDs exist, depending on the linkage phase of the alleles at the two loci. For such cases specific calculation formulas for each MCD have been generated (table 4).

The allelic configurations in table 4, include all possible crosses among two parents and are sufficient to describe all linkage relationships among the alleles of two loci. However, they cannot be directly predicted from the molecular phenotypes of the two parents which generate a specific cross. The parental configurations, however, can be inferred from the molecular phenotypes of the progeny of the cross. The approach to the solution of this problem relies on the capacity to 1) determine which fragments present on a gel represent alternative alleles at one locus, and 2) to determine the phase of linkage of the alleles at two loci for both homologous chromosomes in both

Table 5. *Formulas and maximum likelihood equations for the 21 MCDs derived from the general formula* (4)

No of MCD[*]	Formula or ML equation

1
$$R^{[**]} = \frac{m_{21} + m_{22}}{n}$$

2
$$\frac{m_{11}}{(r-2)} + \frac{m_{21}}{(r+1)} + \frac{m_{31}}{(r-1)} + \frac{m_{41}}{r} = 0$$

3
$$R = 1 - \sqrt{S}$$

4
$$R = \sqrt{S}$$

5
$$R = \tfrac{1}{2} - \sqrt{\tfrac{1}{4} - S}$$

$$R = \chi - \sqrt{X^2} + \frac{2m_{31}}{n}$$

$$X = \frac{m_{11} - 2m_{21} - 2m_{22} - m_{31}}{2n}$$

6
$$R = \frac{m_{31} + m_{41} + m_{42}}{n}$$

7
$$\frac{m_{11}}{r-2} + \frac{(2r-1)m_{21}}{r^2 - r + 1} + \frac{2m_{31}}{r-1} + \frac{2(1-r)m_{41}}{r(2-r)} + \frac{m_{51}}{r} + \frac{(1-2r)m_{61}}{r(1-r)} = 0$$

8
$$\frac{m_{11}}{r+1} + \frac{2rm_{21}}{r^2 - 1} + \frac{(2r-1)m_{31}}{r^2 - r + 1} + \frac{m_{41}}{r-1} + \frac{(1-2r)m_{51}}{r(1-r)} + \frac{2m_{61}}{r} = 0$$

9
$$R = \frac{m_{31} + m_{32}}{m_{21} + m_{22} + m_{31} + m_{32}}$$

10
$$\frac{(2r-1)m_{11}}{r^2 - r + 1} + \frac{2rm_{21}}{r^2 - 1} + \frac{2m_{31}}{r-1} + \frac{2(1-r)m_{41}}{r(2-r)} + \frac{(1-2r)m_{51}}{r(1-r)} + \frac{2m_{61}}{r} = 0$$

11
$$\frac{2m_{11}}{2r^2 - 2r - 1} + \frac{m_{21} + m_{22}}{r^2 - r + 1} + \frac{2m_{31}}{2r^2 - 2r + 1} + \frac{m_{41} + m_{42}}{r(r-1)} = 0$$

12
$$R = \frac{m_{21} + m_{22} + m_{23} + m_{24}}{n}$$

13
$$\frac{2rm_{11}}{r^2 - 1} + \frac{(2r-1)(m_{21} + m_{22})}{r^2 - r + 1} + \frac{2m_{31}}{r-1} + \frac{2(1-r)m_{41}}{r(2-r)} + \frac{(1-2r)(m_{51} + m_{52})}{r(1-r)} + \frac{2m_{61}}{r} = 0$$

14
$$R = \frac{(m_{31} + m_{32} + m_{33} + m_{34}) + (m_{41} + m_{42})}{m_{11} + (m_{31} + m_{32} + m_{33} + m_{34}) + 2(m_{21} + m_{22} + m_{41} + m_{42})}$$

15
$$R = \frac{m_{21} + m_{31} + m_{32} + 2(m_{41} + (m_{42})}{m_{21} + (m_{11} + m_{12} + m_{13} + m_{14}) + 2(m_{31} + m_{32} + m_{41} + m_{42})}$$

16
$$\frac{m_{21} + 2m_{41}}{r-1} + \frac{2(2r-1)m_{31}}{2r^2 - 2r + 1} + \frac{(2r-1)(m_{51} + m_{71} + m_{72})}{r(r-1)} + \frac{m_{61} + 2m_{81}}{r} = 0$$

17
$$\frac{2(2r-1)m_{11}}{2r^2 - 2r + 1} + \frac{2(m_{21} + m_{22})}{r-1} + \frac{(1-2r)(m_{31} + m_{32} + m_{33} + m_{34})}{r(1-r)} + \frac{2(m_{41} + m_{42})}{r} = 0$$

18
$$R = \tfrac{1}{2} - \sqrt{\tfrac{1}{4} - \frac{m_{21} + m_{31} + m_{32} + m_{33} + m_{34}}{2n}}$$

19
$$R = \frac{(m_{31} + m_{32}) + (m_{41} + m_{42} + m_{43} + m_{44}) + 2(m_{51} + m_{52})}{(m_{11} + m_{12} + m_{31} + m_{32}) + 2(m_{21} + m_{22} + m_{41} + m_{42} + m_{43} + m_{44} + m_{51} + m_{52})}$$

20
$$\frac{2(2r-1)(m_{11} + m_{12})}{2r^2 - 2r + 1} + \frac{2(m_{21} + m_{22})}{r-1} + \frac{(1-2r)(m_{31} + m_{32} + m_{41} + m_{42} + m_{43} + m_{44})}{r(1-r)} + \frac{2(m_{51} + m_{52})}{r} = 0$$

21
$$R = \frac{2(m_{41} + m_{42} + m_{43} + m_{44}) + (m_{21} + m_{22} + m_{23} + m_{24} + m_{25} + m_{26} + m_{27} + m_{28})}{2n}$$

[*] MCD = marker class distribution.
[**] R = estimate of r; r = *recombination frequency*; m_{ij} = marker class codes from table 3.

parents. The solution is based on the *observed frequencies* of the marker classes in the progeny.

The association of fragments with alleles may be difficult since molecular probes can reveal multiple banding patterns (Gebhardt *et al.* 1991, Görg *et al.* 1992). In such cases, two (or more) fragments describe the same allele when they are always present together or both absent in single plants of a segregating

Table 6. *Information functions of the estimates of recombination frequencies for all possible MCD*

No. of MCD[*]	Information function	Relative efficiency[**] for r equal to			
		0.01	0.05	0.15	0.30
1, 6, 12	$\dfrac{n}{r(1-r)}$	1.00	1.00	1.00	1.00
2	$\dfrac{n(1+2r-2r^2)}{2r(1-r^2)(2-r)}$	0.25	0.27	0.29	0.32
3	$\dfrac{2n(3-4r+2r^2)}{r(3-2r+r_2)(2-r)}$	0.99	0.94	0.83	0.65
4	$\dfrac{2n(1+2r^2)}{(2+r^2)(1-r^2)}$	0.01	0.05	0.13	0.26
5	$\dfrac{n(1+2r-2r^2)(1-2r)^2}{2r(1-r)(1-r+r^2)(2+r-r^2)}$	0.25	0.23	0.17	0.07
7	$\dfrac{n(8-10r^3+24r^2-21r)}{4r(2-r)(1-r)(1-r+r^2)}$	0.99	0.94	0.83	0.67
8	$\dfrac{n(10r^3-6r^2+3r+1)}{4r(1-r+r^2)(1-r^2)}$	0.26	0.28	0.34	0.40
9	$\dfrac{n}{2r(1-r)}$	0.50	0.50	0.50	0.50
10, 13	$\dfrac{n(4-5r-3r^2+16r^3-8r^4)}{2r(1-r^2)(1-r+r^2)(2-r)}$	0.99	0.96	0.87	0.74
11	$\dfrac{n(8r^2(r-1)^2(4r^2-4r+3)-1)}{2r(r-1)(1-r+r^2)(1-4r^2+8r^3-4r^4)}$	0.50	0.50	0.41	0.18
14	$\dfrac{n(3-r)}{2r(1-r)}$	1.50	1.48	1.43	1.35
15	$\dfrac{n(2+r)}{2r(1-r)}$	1.01	1.03	1.08	1.15
16	$\dfrac{n(14r^2-14r+5)}{4r(1-2r+2r^2)(1-r)}$	1.24	1.20	1.08	0.89
17, 20	$\dfrac{2n(1-3r+3r^2)}{r(1-2r+2r^2)(1-r)}$	1.98	1.90	1.66	1.28
18	$\dfrac{2n(2r-1)^2}{r(1-2r+2r^2)(1-r)}$	1.96	1.79	1.32	0.55
19	$\dfrac{3n}{2r(1-r)}$	1.50	1.50	1.50	1.50
21	$\dfrac{2n}{r(1-r)}$	2.00	2.00	2.00	2.00

[*] MCD = marker class distribution, n = sample size.
[**] the relative efficiency of each MCD is expressed as a fraction or multiple of the information functions of MCD No. 1.

progeny. In contrast, they represent two different alleles of the same locus if either one or the other is present in each progeny member. This kind of inference can be extended to other configurations. The allelic configurations where one 'individual' and one 'common marker' (see before) are present in the two parents, should result in progeny where all genotypes have marker fragments. The same holds true for a configuration involving two common fragments.

Different individual markers from different parents can only be identified as alleles if a common allelic marker segregates in the cross.

For single marker loci, the phase of linkage between alleles at two loci in both homologous chromosomes of both parents can be determined by applying the formulas given by Ritter *et al.* (1990). From the observed marker class frequencies, the proper MCD – which consist, in such cases, of only four marker

A) The putative loci mapping to homologous chromosomes of hypothetical parents P1 and P2 are indicated with L1 to L10.

B) Alleles of P1 and P2 are the following:

	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10
P1	X1,Y1	O1,X2	O2,Y2	O3,V1	O4,V2	O5,n	Z1,n	V3,Z2	V4,Z3	O6,V5
P2	X1,Y1	θ1,X2	Y2,n	V1,n	θ2,V2	θ3,n	Z1,n	V3,Z2	V4,Z3	V5,n

C) Detected Linkage Subgroups
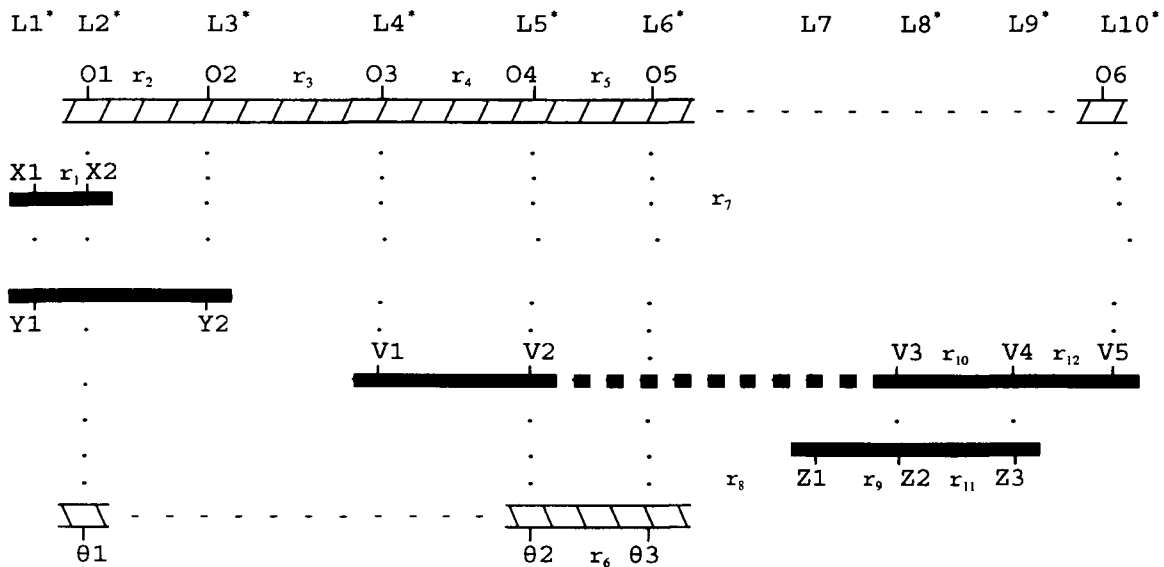
Linkage subgroups:

▱ based on alleles of P1   ▨ based on alleles of P2   ▬ based on alleles of P1 and P2

* Loci with markers segregating without recombination and used as bridge between different linkage subgroups

D) Estimation of recombiantion values

Interval		Allelic configuration at the two loci considered			
		Single markers		Alternative allelic markers	
	value of r	no of MCD	Markers used	no of MCD	Markers used
L1 - L2	r_1	3	X1 X2	20	X1 Y1 - θ1 O1X2
L2 - L3	r_2	1	O1 O2	19	θ1 O1 X2 - O2 Y2
L3 - L4	r_3	1	O2 O3	15	O2 Y2 - O3 V1
L4 - L5	r_4	1	O3 O4	19	O3 V1 - O4 θ2V2
L5 - L6	$(r_5+r_6)/2$	1	O4 O5/θ2 θ3	21	O4 θ2 V2 - O5 θ3
L6 - L7	$(r_7+r_8)/2$	2	O5 Z1/θ3 Z1	13	O5 θ3 - Z1
L7 - L8	r_9	3	Z1 Z2	10	Z1 - V3 Z2
L8 - L9	$(r_{10}+r_{11})/2$	3	V3 V4/Z2 Z3	17	V3 Z2 - V4 Z3
L9 - L10	r_{12}	3	V4 V5	16	V4 Z3 - V5 O6

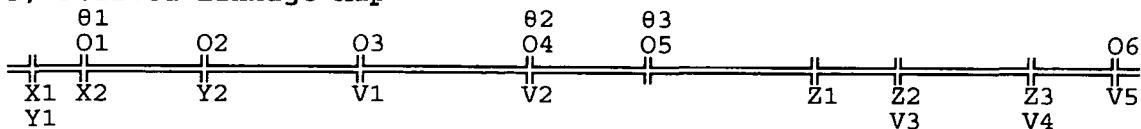**D) Derived Linkage Map**

Figure 1: Use of allelic two-point estimates of r in linkage mapping. A hypothetical chromosome has loci L1 to L10 (**A**) for which several alleles exist. Alleles (**B**) include individual markers (like θ_i, o_i), common markers (like x_i, y_i, v_i, z_i) and

classes – can be deduced and applying the proper formula should give recombination frequencies within the valid range of 0 to 0·5. All the allelic configurations based on single markers at two loci can be identified unambiguously using this method, with the exception of the configurations corresponding to the cases No 4 and 5 of Table 3 [see footnote ($\neq$)]. These can not be distinguished from each other based on the distribution of genotypes in the progeny of the cross. These two cases give in-range but different recombination values for a fixed set of observed marker class frequencies. In such cases, other loci flanking loci A and B and segregating in both parents allow the differentiation between configurations 4 and 5.

(iii) *Formulas for calculating recombination frequencies and information functions*

The recombination frequency between two loci can be calculated by solving the Maximum Likelihood equation given in formula (1). Based on the additive character of the three types of expected frequencies (table 1, lower section), formula (1) can be rewritten as:

$$\frac{\delta lnL(r)}{\delta r} = \sum_j Z_j \frac{n_{1j}t'_1 + n_{2j}t'_2 + n_{3j}t'_3}{n_{1j}t_1 + n_{2j}t_2 + n_{3j}t_3} = 0 \qquad (3)$$

where t_i and t'_i represent FTs and their corresponding derivatives, respectively, and n_{ij} represent the frequency with which each of the three FTs i occurs for each marker class j, as specified by the values of the three digits given in table 3 for the F columns. After including the formulas of table 1 (lower section) and simplifying, the following equation is obtained:

$$\frac{\delta lnL(r)}{\delta r} = \sum_j Z_j \frac{2r(n_{1j} - n_{2j} + n_{3j}) + (n_{2j} - 2n_{1j})}{r^2(n_{1j} - n_{2j} + n_{3j}) + r(n_{2j} - 2n_{1j}) + n_{1j}} = 0 \qquad (4)$$

The information function I_r [formula (2)] can be rewritten as:

$$I_r = \sum_j \frac{\frac{1}{4}(2r(n_{1j} - n_{2j} + n_{3j}) + (n_{2j} - 2n_{1j}))^2}{r^2(n_{1j} - n_{2j} + n_{3j}) + r(n_{2j} - 2n_{1j}) + n_{1j}} = 0 \qquad (5)$$

For $r = 0$ and when the first digit of the frequency code (table 3, column F) is zero, formulas (4) and (5) are not defined. Such situations correspond to the absence of recombination, which is easily detectable as discussed above.

By assigning to the two parents the correct allelic configurations, and after having identified the type and frequency code (column F in table 3) of the marker classes, the general formulas (4) and (5) can be converted, as in tables 5 and 6 which present solutions specific for the 21 MCDs. MCDs 2, 7, 8, 10, 11, 13, 16, 17, and 20 allow the calculation of r based on Maximum Likelihood equations, while MCDs 1, 3, 4, 5, 6, 9, 12, 14, 15, 18, 19 and 21 allow the calculation of R, an estimate of r.

3. Discussion

Allard (1956) has presented numerous ML equations for F2 and backcross data, distinguishing in the progeny between homo and heterozygous individuals and/or differentiating between the coupling and repulsion phases of alleles at different loci. He also considered complete and incomplete dominance, complementary gene effects and the case of incomplete data. The present paper is a more radical approach because it analyzes all possible parental allelic configurations of two loci A and B, represented by segregating dominant single markers and/or codominant allelic markers which may exist for any type of cross. In the absence of gametic or zygotic disturbances, molecular markers segregate in the progeny of a cross with only 1:1 or 3:1 ratios, coupling and repulsion phases are indistinguishable and, in most cases, homozygous and heterozygous individuals cannot be identified in the progeny.

Nevertheless, the number of different allelic markers present at two loci in the parents and the relative state of allele linkage leads to 21 different frequency distributions of marker classes in the progeny. Based on these distributions the recombination frequency between two loci can be calculated for each MCD by applying the general formula (3) or the simplified formulas given in table 5. Ritter *et al.* (1990) have already presented a subset of these calculation formulas and ML equations; those numbered as 1 to 5, 9, 10, 11, 17, 18, 21. For the MCDs missing in the paper cited, we provide new formulas and equations.

Using the algorithms given, recombination frequencies and information functions can easily be calculated based on the approach presented. These have already been converted into a computer program (Ritter, unpublished data). To calculate the recombination frequency when considering a segregating progeny of a cross from parents with unknown allelic configurations, the strategy is as follows:

1. Based on the progeny, the determination of the allelic configuration in the parents of a cross for a given set of allelic markers from two loci.

null alleles (n). Linkage subgroups (C) have to be established based on MCDs with case numbers 1 and 3 (these originate from parental allelic configurations having individual or common alleles in both parents). The linkage subgroups are connected via allelic bridges (dotted vertical lines in C) as proposed by Ritter *et al.* (1990). Allelic marker estimates of higher precision instead of single marker estimates (D) can be used for obtaining the final linkage map given in E. (See table 6; details concerning this figure are also given in text.)

2. The derivation of the type and frequency code of expected marker classes based on parental configurations from mating tables.

3. The determination of the observed marker class frequencies [Z_j in formulas (1),(3),(4)].

4. The calculation of the recombination frequency between the two loci and of the related information function by inserting the corresponding frequency codes into the formulas (4) and (5). These can be solved by using different approaches such as Newton's method, the secant method or the methods described by Press *et al.* (1989). The formulas given in table 5 and 6 are of help to the end user because they simplify the procedures needed to calculate r.

The use of formulas based on several parental configurations of allelic markers may have different applications in linkage mapping based on molecular markers. One evident advantage is the rapid detection of possible scoring errors indicated by the appearance of marker classes which should not occur given a specific allelic configuration.

As shown in table 6, specific allelic configurations have higher reliability (= higher values of their information function) than other configurations, particularly than those involving single markers. To estimate precisely the distance of a marker locus from a particular gene of interest, or when integrating a new locus into an existing linkage map, it is evident that the use of allelic markers increases the precision of the estimate of the recombination frequency.

As pointed out by Ritter *et al.* (1990), the use of mixtures of estimates obtained from different allelic configurations for aligning linked markers may lead to contradictions and unsafe alignments due to different degrees of accuracy of the individual estimates. A method proposed to overcome this problem (Ritter *et al.* 1990) makes use of linkage subgroups based on single marker configurations corresponding to MCD cases No. 1 and 3 of table 3 (this paper). The MCD No. 3 is particularly useful because in this case both parents of the cross have common fragment markers, allowing the creation of allelic bridges (see figure 2 of Ritter *et al.*, 1990). Figure 1 of this paper shows a hypothetical example of integrated mapping. The group of loci given in fig. 1A is represented by different individual markers like O_i descending from parent 1, θ_i descending from parent 2 and markers common to both parents like X_i, Y_i, V_i and Z_i. Individual markers of each parent and common markers are considered separately when estimating r values. First, individual markers linked in coupling or repulsion and common markers linked in coupling are assigned to particular linkage subgroups. The markers defining each linkage subgroup are linearly ordered based on their recombination frequencies (described before, see Ritter *et al.* 1990). Linkage subgroups based on individual and common fragments are then connected based on markers identified as alleles of

particular loci. These loci are characterized by the absence of recombination between them and can be *bona fide* considered as defining a single genetic locus whose alleles represent bridges among linkage subgroups. As shown in fig. 1C, for example, O4, V2 and θ2 represent alleles of locus L5 which connect three different linkage subgroups. The relative orientation of these linkage subgroups can be determined based on the recombination frequencies between loci defined by allelic or non allelic markers.

As proposed by Ritter at al. (1990), the final linkage map (fig. 1E) is based on estimates of r between single markers, or on average estimates obtained for identical intervals (columns 2 and 3 of fig. 1D). Alternatively, r can be obtained based on allelic marker configurations as shown in columns 5 and 6 of fig. 1D. For example, the distance between loci L1 and L2 can be determined either by using simply the estimate r_1, or considering the allelic marker configurations at the two loci, by using MCD No. 20 which involves allelic markers X1 and Y1 of locus L1 and O1, X2 and θ1 of locus L2. The advantage of the calculations using allelic markers is that they use specific estimates of r which have higher precision.

References

Allard, R. W. (1956). Formulas and tables to facilitate the calculation of recombination values in heredity. *Hilgardia* **24**, 235–278.

Bailey, N. T. J. (1961). *Introduction to the mathematical theory of linkage.* Oxford: Clarendon Press.

Barzen, E., Mechelke, W., Ritter, E., Seitzer, J. F. & Salamini, F. (1992). RFLP markers for sugar beet breeding: chromosomal linkage maps and location of major genes for rhizomania resistance, monogermy and hypocotyl colour. *Plant Journal* **2(4)**, 601–611.

Fisher, R. A. (1921). On the mathematical foundations of theoretical statistics. *Philos. trans. Royal Society of London* A **122**, 309–368.

Fisher, R. A. & Balmakund, B. (1928). The estimation of linkage from the offspring of selfed heterozygotes. *Journal of Genetics* **20**, 79–92.

Gebhardt, C., Ritter, E., Debener, T., Schachtschabel, U., Walkemeier, B., Uhrig, H. & Salamini, F. (1989). RFLP analysis and linkage mapping in Solanum tuberosum. *Theoretical and Applied Genetics* **78**, 65–75.

Gebhardt, C., Ritter, E., Barone, A., Debener, T., Walkemeier, B., Schachtschabel, U., Kaufmann, H., Thompson, R. D., Bonierbale, M. W., Tanksley, S. D. & Salamini, F. (1991). RFLP maps of potato and their alignment with the homoeologous tomato genome. *Theoretical and Applied Genetics* **83**, 49–57.

Görg, R, Schachtschabel, U., Ritter, E., Salamini, F. & Gebhardt, C. (1992). Discrimination among 136 tetraploid potato varieties by fingerprints using highly polymorphic DNA markers. Crop Science 32: 815–819.

Lander, E. S., Green, P., Abrahamson, J., Barlow, A., Daly, M. J., Lincoln, S. E. & Newburg, L. (1987). Mapmaker: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* **1**, 174–181.

Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. (1989). *Numerical Recipes (FORTRAN Version).* Cambridge: Cambridge University Press.

Ritter, E., Gebhardt, C. & Salamini, F. (1990). Estimation of recombination frequencies and construction of RFLP linkage maps in plants from crosses between heterozygous parents. *Genetics* **125**, 645–654.

Stam, P. (1993). Construction of integrated genetic linkage maps by means of a new computer package: JOINMAP. *Plant Journal* **3(5)**, 739–744.

Suiter, K. A., Wendel, J. F. & Case, J. S. (1983). LINKAGE-1: a PASCAL computer program for the detection and analysis of genetic linkage. *Journal of Heredity* **74**, 203–204.

Vowden, C. & Ridout, M. (1994). *LINKEM a program for genetic linkage analysis*. Kent, UK: Horticulture Research International East Malling, West Malling.