A general algorithm to compute multilocus genotype frequencies under various mating systems

Hospital, F.* : Station de Génétique Végétale, INRA/UPS/INA-PG, Ferme du Moulon, F-91190 Gif sur Yvette, France
Dillmann, C. : GEVES, La Minière, F-78285 Guyancourt Cedex, France
and
Melchinger, A. E. : Universität Hohenheim, Institut für Pflanzenzüchtung, Saatgutforschung und Populationsgenetik, D-70593 Stuttgart, Germany

Draft version: May 15, 2001

Running head:
Multilocus genotype frequencies
Keywords:
Population genetics, quantitative genetics, genetic linkage analysis, Mendelian segregation, mating systems.
Corresponding author:
Frédéric Hospital
Station de Génétique Végétale, INRA/UPS/INA-PG
Ferme du Moulon, 91190 GIF SUR YVETTE France
Tel: (33)(1) 69 33 23 36
Fax: (33)(1) 69 33 23 40
E-mail: `fred@moulon.inra.fr`

*To whom reprint requests should be sent

Abstract

This paper provides a general method to derive algebraic expressions of genotype frequencies for multiple loci under various mating systems including random mating, backcrossing, selfing, and full-sib mating. For each mating system, general equations are presented. In the case of three loci, comprehensive tables provide recurrence equations for genotype frequencies under random or self mating, and expected genotype frequencies after two generations of full-sib mating. Our results should prove useful in genetic linkage analysis.

Introduction

Theoretical problems in population or quantitative genetics often require that the expected genotype frequencies at two or even more loci are known. This is the case for example in genetic linkage analysis and marker-assisted selection. Obtaining such algebraic expressions is in most cases theoretically possible, but in practice it is very laborious, when more than two loci or more than two successive generations are considered. Hence, only few results corresponding to some specific cases are available in the literature. Haldane and Waddington (1931) presented complete recurrence equations for genotype frequencies at two loci under self fertilization or full-sib mating and derived asymptotic expressions for the recombination fraction. Allard (1956) tabulated comprehensive values for calculation of recombination fractions in progeny of an F_1 hybrid resulting from the cross of two homozygous inbred strains. Feldman *et al.* (1974) gave recurrence equations for genotype frequencies at three loci under random mating with selection. Snape (1988) made use of Haldane and Waddington's recurrence equations and studied recombination frequency estimates in single-seed descent populations. The computation of three-locus genotype frequencies for the interval mapping of quantitative trait loci was performed for F_2 populations (Haley and Knott, 1992; Luo and Kearsey, 1992) and for backcross populations (Martinez and Curnow, 1992). Knott and Haley (1992) handled the case of full-sib families without giving explicit formulae for genotype frequencies. Visscher and Thompson (1995) gave expressions for haplotype frequencies under backcrossing.

The practical difficulty of writing complex algebraic expressions without errors can be overcome today by using computer programs performing symbolic calculations. We derive here a general method to obtain closed expressions for genotype frequencies at any number of linked loci with such a program and apply it to provide recurrence equations and complete expressions for genotype frequencies at initial generations under random mating, backcrossing, self fertilization, or full-sib mating with no selection. The results for selfing or full-sib mating can be used to obtain genotype frequencies in recombinant inbred strains derived by either mating scheme.

System and methods

This method was originally designed for use with the software package Mathematica version 2.2 (Wolfram, 1988), and we applied it to obtain symbolic expressions in the situations described in the Algorithm section. The Mathematica notebooks may be obtained upon request by sending your electronic address to the corresponding author. The computing time depends on the number of loci, and on the number of successive generations taken into account. The

algorithm could also be implemented in any available language to provide numeric results. In the latter case, computation may be faster.

Definitions

Let n be the total number of loci. Since in most cases the studied populations will be the progeny of a cross between two inbred strains, we consider the case of biallelic loci. There are $N = 2^n$ possible gamete types. Each gamete type may then be represented by a decimal integer i ranging from 1 to N. If we designate the two alleles at each locus as numbers 0 and 1, each gamete can also be represented by a binary number written with n digits ranging from $\overbrace{0\cdots0}^{n}$ to $\overbrace{1\cdots1}^{n}$. This binary representation of gametes is equivalent to the set representation used by Geiringer (1944), Schnell (1961) and Christiansen (1987, 1989), but binary representation of gametes is more convenient here for automatic computations. The correspondence between decimal and binary indexing of gametes is provided in Appendix A by equations (A.1) and (A.2). An example of both systems of indices in the case of three loci is:

Binary	000	001	010	011	100	101	110	111
Decimal	1	2	3	4	5	6	7	8

(1)

Let (x, y) denote the genotype formed by the union of (maternal) gamete x and (paternal) gamete y. We denote the probability that genotype (x, y) produces the gamete i after meiosis as $P_{x,y}[i]$. A method for the automatic computation of P for any number of loci is described in Appendix A.

Let $f_{x,y}(t)$ be the frequency of the genotype (x, y) at generation t $(1 \leq x \leq y \leq N)$, the recurrence relationship with genotype frequencies at the previous generation $(t-1)$ is obtained by combining the gametic probabilities in different ways depending on the mating system as described below.

For some mating systems, we provide tables for the case of three loci. The relevant parameter in all tables is r_l $(1 \leq l \leq n-1)$, the recombination rate between adjacent loci l and $(l+1)$. For full-sib mating, recombination rates in males and females were allowed to be possibly different, so that r_l is replaced by r_l^m (recombination in males) or r_l^f (recombination in females) for each l.

Algorithm

The formulae given in this section do not require any assumption about interference in recombination. Absence of interference is assumed for the calculation of probabilities R_k (see Appendix A, equation (A.6)) and, hence, is also assumed in the tables giving explicit results (see Discussion).

Hybrid populations

Consider the hybrid population $\mathcal{P}^{\mathcal{A}\times\mathcal{B}}$ obtained by randomly crossing individuals from a population $\mathcal{P}^{\mathcal{A}}$ to those from a population $\mathcal{P}^{\mathcal{B}}$. This situation is relevant to hybrid breeding, and is a general case of which random mating and backcrossing are special cases (see below).

Let $f^A_{x,y}$, $f^B_{x,y}$ and $f^{A\times B}_{x,y}$ be the frequency of genotype (x,y) in $\mathcal{P}^{\mathcal{A}}$, $\mathcal{P}^{\mathcal{B}}$ and $\mathcal{P}^{\mathcal{A}\times\mathcal{B}}$, respectively. We have:

$$\begin{aligned} f^{A\times B}_{i,j}(t) &= \left(\sum_{x=1}^{N}\sum_{y=x}^{N} P_{x,y}[i]\, f^A_{x,y}(t-1)\right)\left(\sum_{u=1}^{N}\sum_{v=u}^{N} P_{u,v}[j]\, f^B_{u,v}(t-1)\right) + \\ &\quad \delta(i,j)\left(\sum_{x=1}^{N}\sum_{y=x}^{N} P_{x,y}[j]\, f^A_{x,y}(t-1)\right)\left(\sum_{u=1}^{N}\sum_{v=u}^{N} P_{u,v}[i]\, f^B_{u,v}(t-1)\right) \qquad (2) \\ &= \sum_{x=1}^{N}\sum_{y=x}^{N}\sum_{u=1}^{N}\sum_{v=u}^{N}\Bigg(\Big(P_{x,y}[i]\, P_{u,v}[j] + \delta(i,j)\, P_{x,y}[j]\, P_{u,v}[i] \Big) \\ &\quad f^A_{x,y}(t-1)\, f^B_{u,v}(t-1) \Bigg) \qquad (3) \end{aligned}$$

where $\delta(i,j)$ is such that:

$$\delta(i,j) = \begin{cases} 0 & \text{if } i = j \\ 1 & \text{if } i \neq j \end{cases} \qquad (4)$$

Random mating

In the case of random mating, genotype frequencies may be obtained from equation (3) by setting $\mathcal{P}^{\mathcal{A}} = \mathcal{P}^{\mathcal{B}}$. The gametes produced by each genotype are pooled prior to mating, so that the recurrence relationship may be obtained at the level of gamete frequencies. Let $q_x(t)$ be the frequency of gamete type x which form generation t. We have:

$$q_i(t) = \sum_{x=1}^{N}\sum_{y=x}^{N}\left(2^{\delta(x,y)}\, P_{x,y}[i]\, q_x(t-1)\, q_y(t-1)\right) \qquad (5)$$

Recurrence relationships on gamete frequencies for three loci are given in Table 1. It can be compared to Table 1 in Feldman *et al.* (1974) dealing with a symmetric viability selection model.

— Table 1 around here —

The genotype frequencies are then simply obtained by

$$\begin{aligned} f_{i,j}(t) &= 2^{\delta(i,j)}\, q_i(t)\, q_j(t) \qquad (6) \\ &= 2^{\delta(i,j)} \sum_{x=1}^{N}\sum_{y=x}^{N}\sum_{u=1}^{N}\sum_{v=u}^{N}\Big(P_{x,y}[i]\, P_{u,v}[j]\, f_{x,y}(t-1)\, f_{u,v}(t-1) \Big) \qquad (7) \end{aligned}$$

Backcrossing

In the case of backcrossing, let (b,b') be the genotype of the recurrent parent, and let $\mathcal{B}$ be the set of the indices of all possible gametes produced by the recurrent parent. Again, the recurrence relationship on genotype frequencies can be obtained from equation (3). Consider that $\mathcal{P}^{\mathcal{B}}$ is reduced to the single genotype (b,b'), and that $\mathcal{P}^{\mathcal{A}}$ is the donor population ($\mathcal{P}^{\mathcal{A}}(t) =$

$\mathcal{P}^{\mathcal{A}}(t-1) \times \mathcal{P}^{\mathcal{B}}$). Genotype frequencies in the two populations are such that:

$$\text{in } \mathcal{P}^{\mathcal{A}} \quad : \quad f^A_{x,y} = 0 \quad \text{if } x \notin \mathcal{B} \text{ and } y \notin \mathcal{B} \text{ (except for the initial parent)}$$

$$\text{in } \mathcal{P}^{\mathcal{B}} \quad : \quad f^B_{u,v} = \begin{cases} 1 & \text{if } (u,v) = (b,b') \\ 0 & \text{otherwise} \end{cases}$$

We then have the recurrence relationship for genotype frequencies in $\mathcal{P}^{\mathcal{A}}$:

$$\begin{aligned} f_{i,j}(t) &= \left(\sum_{x \notin \mathcal{B}} \sum_{\substack{y \in \mathcal{B} \\ y \geq x}} \Big(P_{x,y}[i]\, P_{b,b'}[j] + \delta(i,j)\, P_{x,y}[j]\, P_{b,b'}[i] \Big)\, f_{x,y}(t-1) \right) + \\ &\qquad \left(\sum_{x \in \mathcal{B}} \sum_{y \geq x} \Big(P_{x,y}[i]\, P_{b,b'}[j] + \delta(i,j)\, P_{x,y}[j]\, P_{b,b'}[i] \Big)\, f_{x,y}(t-1) \right) \qquad (8) \\ &= \sum_{x=1}^{N} \sum_{y \in \mathcal{B}} \left(\Big(P_{x,y}[i]\, P_{b,b'}[j] + \delta(i,j)\, P_{x,y}[j]\, P_{b,b'}[i] \Big)\, f_{\min(x,y),\max(x,y)}(t-1) \right) \qquad (9) \end{aligned}$$

Genotype frequencies at two or three loci when both parents are homozygous are given in Visscher and Thompson (1995). Note that our results extend to the case when parents are not homozygous: recurrence relationships for genotype frequencies in the case of backcrossing to any population $\mathcal{P}^{\mathcal{B}}$ can also be derived from equation (3).

Self-fertilization

Under selfing, offspring genotype frequencies must be first computed for each parent genotype and then summed up. The recurrence relationship for genotype frequencies is:

$$f_{i,j}(t) = 2^{\delta(i,j)} \left(\sum_{x=1}^{N} \sum_{y=x}^{N} P_{x,y}[i]\, P_{x,y}[j]\, f_{x,y}(t-1) \right) \qquad (10)$$

Genotype frequencies in recombinant inbred strains derived by repeated self mating can be obtained numerically by iterating equation (10) until the equilibrium is reached.

In the progeny of a cross between two inbred strains, some genotype frequencies may be equal at each generation, due to symmetry. Consider the set of all possible genotype frequencies $\{f_{i,j}\}_{1 \leq i \leq j \leq N}$ as the elements of a triangular matrix where maternal gamete indices (i) are on rows, and paternal gamete indices (j) are on columns. The first diagonal D_1 defined by $i = j$ ($1 \leq i \leq N$) contains the frequencies of the N genotypes that are homozygous at all loci. The second diagonal D_2 defined by $j = N+1-i$ ($1 \leq i \leq N/2$) contains the frequencies of the $N/2$ genotypes that are heterozygous at all loci. During meiosis, each recombination event always produces two gamete types with the same frequency. Hence, depending on the genotype frequencies in the original F_1 population, some genotypes may have the same frequency at any following generation. Obviously, if the triangular matrix of genotype frequencies in the original F_1 population is symmetrical with respect to D_2, this symmetry will remain valid at any generation of selfing. Let i' be the gamete type symmetrical to gamete type i. We define i' by:

$$i' = N + 1 - i \qquad (11)$$

for any i. We then have the symmetry on genotype frequencies:

$$f_{i,j} = f_{j',i'} \tag{12}$$

for any genotype (i, j) $(1 \leq i \leq j \leq N)$.

For example, the symmetries defined by equation (12) hold in the case of three loci when the original cross is $\frac{000}{000} \times \frac{111}{111}$, so that the F_1 population contains only the genotype $\frac{000}{111}$. The frequencies of the 36 possible genotypes can then be described by only 20 parameters (denoted f_i^*). These parameters and the corresponding genotypes are presented in Table 2 along with recurrence relationships on the f_i^* parameters for three loci.

— Table 2 around here —

Under self-fertilization with the same initial cross, some additional symmetries exist within each side of D_2 on the triangular matrix. These symmetries may be determined *a priori* by computing the *recombination score* of each genotype, as is done in our Mathematica notebook. We define the recombination score of a gamete type as the number of recombination events needed between each pair of adjacent loci to derive this gamete from the original F_1. The recombination score is written as a $(n-1)$ digit. For example, at three loci, if the original F_1 is $\frac{000}{111}$, the recombination score of gamete type 011 would be 10 (one recombination between locus 1 and locus 2, no recombination between locus 2 and locus 3). We then define the recombination score of a genotype (i, j) as the sum of the recombination scores of i and j, times an arbitrary sign. We chose to give a positive recombination score to genotypes with the first locus being homozygous (for either of the two alleles), and a negative score to genotypes with the first locus being heterozygous. For example, the recombination score of genotype $\frac{010}{011}$ would be +21. If two genotypes have the same recombination score their frequencies are equal at each generation. At three loci, this would reduce the number of f_i^* parameters needed to describe all possible genotype frequencies from 20 to 18 ($f_6^* = f_{12}^*$; $f_9^* = f_{13}^*$). These additional symmetries were not taken into account in Table 2, so that the same indexing may also be used for full-sib mating (see below). Note that the symmetries given by the computation of the recombination scores include the symmetries given by equation (12).

Full-sib mating

It is not possible to obtain a recurrence relationship for genotype frequencies for the case of full-sib mating. Yet, it is possible to derive such a relationship for the frequencies of couples of genotypes (*i.e.* couples of individuals). Let $(x, y; u, v)$ be the couple of the (female) genotype (x, y) and the (male) genotype (u, v). Let $G_{x,y;u,v}$ be the matrix of frequencies of all possible *genotypes* produced by the couple $(x, y; u, v)$, so that $G_{x,y;u,v}[i, j]$ is the probability that genotype (x, y) produces the gamete i and that genotype (u, v) produces the gamete j. If we allow recombination frequencies in males and females to be possibly different, we have:

$$G_{x,y;u,v} = \left(P^f_{x,y}\right)' . P^m_{u,v} \tag{13}$$

where the prime denotes transposition and P^f and P^m are row vectors obtained as in equations (A.6) and (A.8) by replacing r_l by r_l^f (recombination in females) and r_l^m (recombination in males), respectively. Let $h_{x,y;u,v}(t)$ be the frequency of the couple $(x, y; u, v)$ at generation t, we then have the recurrence relationship for couple frequencies:

$$h_{i,j;k,l}(t) = \sum_{x=1}^{N}\sum_{y=x}^{N}\sum_{u=1}^{N}\sum_{v=u}^{N} \Bigg(\Big(G_{x,y;u,v}[i,j] + \delta(i,j)\, G_{x,y;u,v}[j,i] \Big) \Big(G_{x,y;u,v}[k,l] + \delta(k,l)\, G_{x,y;u,v}[l,k] \Big) h_{x,y;u,v}(t-1) \Bigg) \quad (14)$$

The frequency of a given genotype is then obtained by:

$$f_{i,j}(t) = \sum_{x=1}^{N}\sum_{y=x}^{N}\sum_{u=1}^{N}\sum_{v=u}^{N} \Bigg(\Big(G_{x,y;u,v}[i,j] + \delta(i,j)\, G_{x,y;u,v}[j,i] \Big) h_{x,y;u,v}(t-1) \Bigg) \quad (15)$$

Genotype frequencies in recombinant inbred strains derived by repeated full-sib mating can be obtained numerically by iterating equation (15) until the equilibrium is reached.

The symmetries defined by equations (11) and (12) apply under full-sib mating for genotype frequencies. They also induce symmetries on couple frequencies. In addition, the frequency of the couple female$_a$ × male$_b$ is equal to the frequency of the couple female$_b$ × male$_a$, provided this symmetry existed also in the initial generation. Hence, the following symmetries hold for couple frequencies at any generation:

$$h_{i,j;k,l} = h_{l',k';j',i'} = h_{j',i';l',k'} = h_{k,l;i,j} \quad (16)$$

At three loci, when only the couple $\frac{000}{111} \times \frac{000}{111}$ is present in the population at the first generation, these symmetries reduce the number of couples to be considered from 1296 to 360, and the number of genotypes from 36 to 20. Genotype frequencies can then be represented by the same starred parameters f_i^* given Table 2. A table containing recurrence equations for three loci was too big to be presented here, but it is possible to derive such a table with our Mathematica notebook. We only provide genotype frequencies for the second generation (Table 3). Note that for homogeneous recombination rates ($r_l^f = r_l^m$), these frequencies simplify to the frequencies for the F_2 generation under selfing.

— Table 3 around here —

At two loci, recurrence relationships on couples frequencies provided by equation (14) were checked against the corresponding table in Haldane and Waddington (1931, eqn 3.1). This revealed some typographical errors in Haldane and Waddington's table. The corrected formulae are given in Appendix B.

Discussion

We have considered the case of biallelic loci. The extension to the case of more than two alleles per locus is straightforward and requires a few minor modifications: 2 should be replaced by m in the definitions of N, γ and g (see Appendix A).

The examples given in the tables were obtained under the hypothesis of no interference in recombination. But, whether there is interference or not is only relevant to the definition of

R_k (Appendix A, equation (A.6)). Hence, interference can be taken into account by modifying this equation only. For example, R in equation (A.6) can be replaced by the strictly equivalent function γ defined by Schnell (1961, equation 4), where linkage values can be derived from any available map function assuming interference.

We hope that this paper provides a clear and explicit basis, which will avoid a large number of geneticists having to go through laborious calculations in the course of their work. Also, it is easy to include our formulae in computer programs concerning genetic linkage analysis.

The possible applications of our work are manifold. For example, genetic linkage analysis is often performed after several generations of selfing (plant breeding) or full-sib mating (animal breeding), so that the studied population can be a F_3, a F_4 or a population of recombinant inbreds (F_6 to F_∞). Implementing our algorithm would then provide exact values for the expected frequencies of marker-QTL haplotypes at the specified generation, and hence improve the precision of interval mapping of Quantitative Trait Loci, or the estimation of recombination rates. Also, in some situations (*e.g.* , backcrossing over several successive generations), not only the genotype at the two nearest markers on each side of the putative QTL is informative, but also the genotype at more distant markers. Interval mapping of QTL could then be extended to multiple loci by using our algorithm, or true multipoint tests at more than three loci could be performed in linkage analysis. This is also relevant for any situation where expected genotype frequencies at many loci given a genetic map are needed (*e.g.* , marker-assisted selection, graphical genotypes). More generally, our algorithm can be useful in various types of numerical simulation programs dealing with population or quantitative genetics.

Acknowledgements

We thank I. Goldringer, P. Brabant and one anonymous reviewer for helpful comments on earlier versions of the manuscript. The Mathematica software package was supported by AIP INRA No 93/4924 *via* the MMM group.

References

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

Christiansen, F.B. (1987) The deviation from linkage equilibrium with multiple loci varying in a stepping-stone cline. *J. Genet.*, **66**, 45–67.

Christiansen, F.B. (1989) Linkage equilibrium in multi-locus genotypic frequencies with mixed selfing and random mating. *Theor. Appl. Genet.*, **35**, 307–336.

Feldman, M.W., Franklin, I. and Thomson, G.J. (1974) Selection in complex genetic systems I. The symmetric equilibria of the three-locus symmetric viability model. *Genetics*, **76**, 135–162.

Geiringer, H. (1944) On the probability theory of linkage in Mendelian heredity. *Ann. Math. Stat.*, **15**, 25–57.

Haldane, J.B.S. and Waddington, C.H. (1931) Inbreeding and linkage. *Genetics*, **16**, 357–374.

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

Knott, S.A. and Haley, C.S. (1992) Maximum likelihood mapping of quantitative trait loci using full-sib families. *Genetics*, **132**, 1211–1222.

Luo, Z.W. and Kearsey, M.J. (1992) Interval mapping of quantitative trait loci in an F_2 population. *Heredity*, **69**, 236–242.

Martinez, O. and Curnow, R.N. (1992) Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. *Theor. Appl. Genet.*, **85**, 480–488.

Schnell, F.W. (1961) Some general formulations of linkage effects in inbreeding. *Genetics*, **46**, 947–957.

Snape, J.W. (1988) The detection and estimation of linkage using doubled haploids or single seed descent populations. *Theor. Appl. Genet.*, **76**, 125–128.

Visscher, P.M. and Thompson, R. (1995) Haplotype frequencies of linked loci in backcross populations derived from inbred lines. *Heredity*, **75**, 644–649.

Wolfram, S. (1988) *Mathematica, a system for doing mathematics by computer.* Addison-Wesley Publishing Company, Inc, Redwood City, California.

Appendix A
Automatic computation of multilocus segregations

We will use the following notations: $V[k]$ is the k-th element of a given vector V, $M[k,l]$ is the element of a given matrix M at row k and column l and $M[k,\bullet]$ is the vector formed by the k-th row of matrix M.

We consider n biallelic loci. There are $N = 2^n$ possible gamete types and $N(N+1)/2$ possible genotypes. We need an indexed representation of all possible gamete types and genotypes, and a representation of all possible recombination events relating the genotypes to the gamete types. We use either a decimal, or a binary indexing of gamete types (see Definitions section).

We define the general function Base such that $\text{Base}_b^n(i,l)$ gives the l-th digit of the representation of i in base b with a total number of n digits. The binary representation of gamete with decimal index i ($1 \leq i \leq N$) is stored in a row vector γ_i of length n. The relationship between the decimal and the binary representation of the gamete is obtained by considering that the vector γ_i contains the n digits of the binary representation of $(i-1)$, so that the allele of gamete i at locus l (the l-th element of γ_i) is obtained by:

$$\gamma_i[l] = \text{Base}_2^n(i-1,l) \quad \text{for } 1 \leq l \leq n \tag{A.1}$$

The set of vectors $\{\gamma_i\}_{1 \leq i \leq N}$ is then the set of all possible gamete types.

Conversely the decimal index is obtained by $i = g(\gamma_i)$, with

$$g(\gamma_i) = 1 + \sum_{k=0}^{n-1} 2^k \, \gamma_i[n-k] \tag{A.2}$$

An example of both decimal and binary indexing of gametes is given in the text for the three locus case (equation (1)).

Let $\omega_{i,j}$ be the $2 \times n$ matrix containing the binary representation of genotype (i,j) formed by the union of (maternal) gamete i and (paternal) gamete j.

$$\omega_{i,j} = \begin{bmatrix} [\ \gamma_i\ ] \\ [\ \gamma_j\ ] \end{bmatrix} \tag{A.3}$$

During each meiosis, there is either recombination between two successive loci, or not, regardless of the genotype at these loci. Then, the number of gamete types produced during each meiosis depends on the genotype, since different recombination events may produce the same gamete type (if at least one locus is homozygous), but at this point we treat separately the gametes produced by different recombination events. Taking 1 as maternal origin (i) and 2 as paternal origin (j), the set of all possible recombination events, regardless of the genotype, is represented by the $2^n \times n$ matrix Φ such that

$$\Phi[i,l] = 1 + \mathrm{Base}_2^n(i-1,l) \tag{A.4}$$

(note that $\Phi[i,l] = 1 + \gamma_i[l]$ only in the biallelic case). Then, Φ can be used as a filter to read $\omega_{i,j}$ and provide the set $\Gamma_{i,j}$ of all gametes produced by a given genotype (i,j) during meiosis, in a form suitable for the following calculation of recombination frequencies. $\Gamma_{i,j}$ can be written as a $2^n \times n$ matrix such that:

$$\Gamma_{i,j}[k,l] = \omega_{i,j}[\Phi[k,l],l] \quad \text{for} \begin{cases} 1 \le k \le 2^n \\ 1 \le l \le n \end{cases} \tag{A.5}$$

As previously noticed, several rows of the matrix Γ may be identical, but they will sum up during recombination calculation.

Assuming no interference, the probability R_k associated to each row k of matrix Γ is computed as

$$R_k = \frac{1}{2}\left(\prod_{l=1}^{n-1}\left[\,\rho_{k,l}\,r_l + (1-\rho_{k,l})\,(1-r_l)\right]\right) \tag{A.6}$$

where r_l is the recombination frequency between locus l and locus $(l+1)$ $(1 \le l \le n-1)$ and ρ is such that

$$\rho_{k,l} = |\,\Phi[k,l] - \Phi[k,l+1]\,| \tag{A.7}$$

Now, we need to sum up the probabilities R_k corresponding to identical gametes, and to order it correspondingly with our indexing of all possible gamete types (i from 1 to N). With the definitions above, $g\left(\Gamma_{i,j}[k,\bullet]\right)$ is the decimal index of gamete type produced by genotype (i,j) with probability R_k. Let I be the identity matrix of size N, so that $I[i,\bullet]$ is the row vector containing a 1 at position i and 0's at other positions. The ordered frequencies of all the possible gamete types produced by genotype (i,j) is then given by the row vector $P_{i,j}$ of length N such that

$$P_{i,j} = \sum_{k=1}^{2^n} R_k\, I[\,g\left(\Gamma_{i,j}[k,\bullet]\right),\bullet\,] \tag{A.8}$$

Appendix B
Recurrence relationships for couples frequencies at two loci under full-sib mating.

We present here a correction of Haldane & Waddington (1931) eqn 3.1 in the notations of these authors. Only the equations that using our method (equation (14)) were found to differ from the results in the publication of Haldane & Waddington are given. Please refer to the

original article for the other equations and the definition of the parameters.

$$
\begin{aligned}
G_{n+1} &= \frac{1}{16}\,(Q + \alpha\,\beta\,U + \gamma\,\delta\,U + \alpha\,\beta\,V + \gamma\,\delta\,V + \alpha\,\beta\,\gamma\,\delta\,W + 2\,\alpha\,\beta\,\gamma\,\delta\,X + \\
&\quad \alpha\,\beta\,\gamma\,\delta\,Y) \\
H_{n+1} &= \frac{1}{2}\,H + \frac{1}{4}\,(\alpha\,\beta + \gamma\,\delta)\,L + \frac{1}{4}\,(\alpha\,\beta + \gamma\,\delta)\,N + \frac{1}{8}\,Q + \frac{1}{8}\,R + \frac{1}{16}\,(\alpha^2 + 2\,\alpha\,\beta + \\
&\quad \gamma^2 + 2\,\gamma\,\delta)\,U + \frac{1}{16}\,(2\,\alpha\,\beta + \beta^2 + 2\,\gamma\,\delta + \delta^2)\,V + \frac{1}{16}\,\alpha\,\gamma\,(\beta\,\gamma + \alpha\,\delta)\,W + \\
&\quad \frac{1}{16}\,(\beta\,\gamma + \alpha\,\delta)\,(\alpha\,\gamma + \beta\,\delta)\,X + \frac{1}{16}\,\beta\,\delta\,(\beta\,\gamma + \alpha\,\delta)\,Y \\
I_{n+1} &= \frac{1}{2}\,I + \frac{1}{4}\,(\alpha\,\beta + \gamma\,\delta)\,M + \frac{1}{4}\,(\alpha\,\beta + \gamma\,\delta)\,P + \frac{1}{8}\,Q + \frac{1}{8}\,S + \frac{1}{16}\,(2\,\alpha\,\beta + \beta^2 + \\
&\quad 2\,\gamma\,\delta + \delta^2)\,U + \frac{1}{16}\,(\alpha^2 + 2\,\alpha\,\beta + \gamma^2 + 2\,\gamma\,\delta)\,V + \frac{1}{16}\,\beta\,\delta\,(\beta\,\gamma + \alpha\,\delta)\,W + \\
&\quad \frac{1}{16}\,(\beta\,\gamma + \alpha\,\delta)\,(\alpha\,\gamma + \beta\,\delta)\,X + \frac{1}{16}\,\alpha\,\gamma\,(\beta\,\gamma + \alpha\,\delta)\,Y \\
M_{n+1} &= \frac{1}{4}\,(\alpha^2 + \gamma^2)\,M + \frac{1}{4}\,(\beta^2 + \delta^2)\,P + \frac{1}{8}\,(\beta^2 + \delta^2)\,U + \frac{1}{8}\,(\alpha^2 + \gamma^2)\,V + \\
&\quad \frac{1}{8}\,\beta^2\,\delta^2\,W + \frac{1}{8}\,(\beta^2\,\gamma^2 + \alpha^2\,\delta^2)\,X + \frac{1}{8}\,\alpha^2\,\gamma^2\,Y \\
Q_{n+1} &= 2\,G + \frac{1}{2}\,H + \frac{1}{2}\,I + \frac{1}{2}\,J + \frac{1}{2}\,K + \frac{1}{4}\,(\beta^2 + \delta^2)\,L + \frac{1}{4}\,(\beta^2 + \delta^2)\,M + \\
&\quad \frac{1}{4}\,(\alpha^2 + \gamma^2)\,N + \frac{1}{4}\,(\alpha^2 + \gamma^2)\,P + \frac{1}{4}\,Q + \frac{1}{8}\,R + \frac{1}{8}\,S + \frac{1}{8}\,T + \\
&\quad \frac{1}{8}\,(\alpha + \beta^2 + \gamma + \delta^2)\,U + \frac{1}{8}\,(\alpha + \beta^2 + \gamma + \delta^2)\,V + \frac{1}{16}\,(\beta\,\gamma + \alpha\,\delta)^2\,W + \\
&\quad \frac{1}{8}\,(\alpha\,\gamma + \beta\,\delta)^2\,X + \frac{1}{16}\,(\beta\,\gamma + \alpha\,\delta)^2\,Y \\
T_{n+1} &= \frac{1}{8}\,T + \frac{1}{8}\,(\alpha\,\beta + \gamma\,\delta)\,U + \frac{1}{8}\,(\alpha\,\beta + \gamma\,\delta)\,V + \frac{1}{16}\,(\beta\,\gamma + \alpha\,\delta)^2\,W + \\
&\quad \frac{1}{8}\,(\alpha\,\gamma + \beta\,\delta)^2\,X + \frac{1}{16}\,(\beta\,\gamma + \alpha\,\delta)^2\,Y \\
X_{n+1} &= \frac{1}{4}\,T + \frac{1}{4}\,(\alpha\,\beta + \gamma\,\delta)\,U + \frac{1}{4}\,(\alpha\,\beta + \gamma\,\delta)\,V + \frac{1}{4}\,\alpha\,\beta\,\gamma\,\delta\,W + \frac{1}{2}\,\alpha\,\beta\,\gamma\,\delta\,X + \\
&\quad \frac{1}{4}\,\alpha\,\beta\,\gamma\,\delta\,Y
\end{aligned}
$$

Table 1 Recurrence relationships for gamete frequencies at three loci under random mating. For each gamete type i shown in first column, second column gives the corresponding frequency q_i, and last column gives recurrence equation for $q_i(t+1)$ in terms of the $q_j(t)$'s where t is omitted.

Gamete	Freq.	Recurrence equation
000	q_1	$q_1 + r_1\,(-q_1q_6 - q_1q_7 - q_1q_8 + q_2q_5 + q_3q_5 + q_4q_5) + r_2\,(-q_1q_4 - q_1q_6 - q_1q_8 + q_2q_3 + q_2q_5 + q_2q_7) + r_1\,r_2\,(2\,q_1q_6 + q_1q_8 - 2\,q_2q_5 - q_2q_7 + q_3q_6 - q_4q_5)$
001	q_2	$q_2 + r_1\,(q_1q_6 - q_2q_5 - q_2q_7 - q_2q_8 + q_3q_6 + q_4q_6) + r_2\,(q_1q_4 + q_1q_6 + q_1q_8 - q_2q_3 - q_2q_5 - q_2q_7) + r_1\,r_2\,(-2\,q_1q_6 - q_1q_8 + 2\,q_2q_5 + q_2q_7 - q_3q_6 + q_4q_5)$
010	q_3	$q_3 + r_1\,(q_1q_7 + q_2q_7 - q_3q_5 - q_3q_6 - q_3q_8 + q_4q_7) + r_2\,(q_1q_4 - q_2q_3 - q_3q_6 - q_3q_8 + q_4q_5 + q_4q_7) + r_1\,r_2\,(q_1q_8 - q_2q_7 + q_3q_6 + 2\,q_3q_8 - q_4q_5 - 2\,q_4q_7)$
011	q_4	$q_4 + r_1\,(q_1q_8 + q_2q_8 + q_3q_8 - q_4q_5 - q_4q_6 - q_4q_7) + r_2\,(-q_1q_4 + q_2q_3 + q_3q_6 + q_3q_8 - q_4q_5 - q_4q_7) + r_1\,r_2\,(-q_1q_8 + q_2q_7 - q_3q_6 - 2\,q_3q_8 + q_4q_5 + 2\,q_4q_7)$
100	q_5	$q_5 + r_1\,(q_1q_6 + q_1q_7 + q_1q_8 - q_2q_5 - q_3q_5 - q_4q_5) + r_2\,(q_1q_6 - q_2q_5 + q_3q_6 - q_4q_5 - q_5q_8 + q_6q_7) + r_1\,r_2\,(-2\,q_1q_6 - q_1q_8 + 2\,q_2q_5 + q_2q_7 - q_3q_6 + q_4q_5)$
101	q_6	$q_6 + r_1\,(-q_1q_6 + q_2q_5 + q_2q_7 + q_2q_8 - q_3q_6 - q_4q_6) + r_2\,(-q_1q_6 + q_2q_5 - q_3q_6 + q_4q_5 + q_5q_8 - q_6q_7) + r_1\,r_2\,(2\,q_1q_6 + q_1q_8 - 2\,q_2q_5 - q_2q_7 + q_3q_6 - q_4q_5)$
110	q_7	$q_7 + r_1\,(-q_1q_7 - q_2q_7 + q_3q_5 + q_3q_6 + q_3q_8 - q_4q_7) + r_2\,(q_1q_8 - q_2q_7 + q_3q_8 - q_4q_7 + q_5q_8 - q_6q_7) + r_1\,r_2\,(-q_1q_8 + q_2q_7 - q_3q_6 - 2\,q_3q_8 + q_4q_5 + 2\,q_4q_7)$
111	q_8	$q_8 + r_1\,(-q_1q_8 - q_2q_8 - q_3q_8 + q_4q_5 + q_4q_6 + q_4q_7) + r_2\,(-q_1q_8 + q_2q_7 - q_3q_8 + q_4q_7 - q_5q_8 + q_6q_7) + r_1\,r_2\,(q_1q_8 - q_2q_7 + q_3q_6 + 2\,q_3q_8 - q_4q_5 - 2\,q_4q_7)$

r_l: recombination rate between adjacent loci l and $(l+1)$.

Table 2 Recurrence relations for genotype frequencies at three loci under selfing. For each genotype shown in first column, second column gives the corresponding f_i^* parameter, and last column gives recurrence equation for $f_i^*(t+1)$ in terms of the $f_j^*(t)$'s where t is omitted.

Genot.	Freq.	Recurrence equation
$\frac{000}{000}, \frac{111}{111}$	f_1^*	$\frac{1}{4}\left(4\,f_1^* + f_5^* + f_6^* + {s_2}^2\,f_7^* + f_8^* + {s_{13}}^2\,f_9^* + {s_1}^2\,f_{10}^* + {r_2}^2\,f_{11}^* + {r_{13}}^2\,f_{13}^* + {r_1}^2\,f_{16}^* + {r_1}^2\,{s_2}^2\,f_{17}^* + {r_1}^2\,{r_2}^2\,f_{18}^* + {r_2}^2\,{s_1}^2\,f_{19}^* + {s_1}^2\,{s_2}^2\,f_{20}^*\right)$
$\frac{001}{001}, \frac{110}{110}$	f_2^*	$\frac{1}{4}\left(4\,f_2^* + f_5^* + {r_2}^2\,f_7^* + {r_{13}}^2\,f_9^* + {s_1}^2\,f_{10}^* + {s_2}^2\,f_{11}^* + f_{12}^* + {s_{13}}^2\,f_{13}^* + f_{14}^* + {r_1}^2\,f_{16}^* + {r_1}^2\,{r_2}^2\,f_{17}^* + {r_1}^2\,{s_2}^2\,f_{18}^* + {s_1}^2\,{s_2}^2\,f_{19}^* + {r_2}^2\,{s_1}^2\,f_{20}^*\right)$
$\frac{010}{010}, \frac{101}{101}$	f_3^*	$\frac{1}{4}\left(4\,f_3^* + f_6^* + {r_2}^2\,f_7^* + {s_{13}}^2\,f_9^* + {r_1}^2\,f_{10}^* + {s_2}^2\,f_{11}^* + {r_{13}}^2\,f_{13}^* + f_{14}^* + f_{15}^* + {s_1}^2\,f_{16}^* + {r_2}^2\,{s_1}^2\,f_{17}^* + {s_1}^2\,{s_2}^2\,f_{18}^* + {r_1}^2\,{s_2}^2\,f_{19}^* + {r_1}^2\,{r_2}^2\,f_{20}^*\right)$
$\frac{011}{011}, \frac{100}{100}$	f_4^*	$\frac{1}{4}\left(4\,f_4^* + {s_2}^2\,f_7^* + f_8^* + {r_{13}}^2\,f_9^* + {r_1}^2\,f_{10}^* + {r_2}^2\,f_{11}^* + f_{12}^* + {s_{13}}^2\,f_{13}^* + f_{15}^* + {s_1}^2\,f_{16}^* + {s_1}^2\,{s_2}^2\,f_{17}^* + {r_2}^2\,{s_1}^2\,f_{18}^* + {r_1}^2\,{r_2}^2\,f_{19}^* + {r_1}^2\,{s_2}^2\,f_{20}^*\right)$
$\frac{000}{001}, \frac{110}{111}$	f_5^*	$\frac{1}{2}\left(f_5^* + r_2\,s_2\,f_7^* + r_{13}\,s_{13}\,f_9^* + r_2\,s_2\,f_{11}^* + r_{13}\,s_{13}\,f_{13}^* + {r_1}^2\,r_2\,s_2\,f_{17}^* + {r_1}^2\,r_2\,s_2\,f_{18}^* + r_2\,{s_1}^2\,s_2\,f_{19}^* + r_2\,{s_1}^2\,s_2\,f_{20}^*\right)$
$\frac{000}{010}, \frac{101}{111}$	f_6^*	$\frac{1}{2}\left(f_6^* + r_2\,s_2\,f_7^* + r_1\,s_1\,f_{10}^* + r_2\,s_2\,f_{11}^* + r_1\,s_1\,f_{16}^* + r_1\,r_2\,s_1\,s_2\,f_{17}^* + r_1\,r_2\,s_1\,s_2\,f_{18}^* + r_1\,r_2\,s_1\,s_2\,f_{19}^* + r_1\,r_2\,s_1\,s_2\,f_{20}^*\right)$
$\frac{000}{011}, \frac{100}{111}$	f_7^*	$\frac{1}{2}\left({s_2}^2\,f_7^* + {r_2}^2\,f_{11}^* + r_1\,s_1\,{s_2}^2\,f_{17}^* + r_1\,{r_2}^2\,s_1\,f_{18}^* + r_1\,{r_2}^2\,s_1\,f_{19}^* + r_1\,s_1\,{s_2}^2\,f_{20}^*\right)$
$\frac{000}{100}, \frac{011}{111}$	f_8^*	$\frac{1}{2}\left(f_8^* + r_{13}\,s_{13}\,f_9^* + r_1\,s_1\,f_{10}^* + r_{13}\,s_{13}\,f_{13}^* + r_1\,s_1\,f_{16}^* + r_1\,s_1\,{s_2}^2\,f_{17}^* + r_1\,{r_2}^2\,s_1\,f_{18}^* + r_1\,{r_2}^2\,s_1\,f_{19}^* + r_1\,s_1\,{s_2}^2\,f_{20}^*\right)$
$\frac{000}{101}, \frac{010}{111}$	f_9^*	$\frac{1}{2}\left({s_{13}}^2\,f_9^* + {r_{13}}^2\,f_{13}^* + r_1\,r_2\,s_1\,s_2\,f_{17}^* + r_1\,r_2\,s_1\,s_2\,f_{18}^* + r_1\,r_2\,s_1\,s_2\,f_{19}^* + r_1\,r_2\,s_1\,s_2\,f_{20}^*\right)$
$\frac{000}{110}, \frac{001}{111}$	f_{10}^*	$\frac{1}{2}\left({s_1}^2\,f_{10}^* + {r_1}^2\,f_{16}^* + {r_1}^2\,r_2\,s_2\,f_{17}^* + {r_1}^2\,r_2\,s_2\,f_{18}^* + r_2\,{s_1}^2\,s_2\,f_{19}^* + r_2\,{s_1}^2\,s_2\,f_{20}^*\right)$
$\frac{001}{010}, \frac{101}{110}$	f_{11}^*	$\frac{1}{2}\left({r_2}^2\,f_7^* + {s_2}^2\,f_{11}^* + r_1\,{r_2}^2\,s_1\,f_{17}^* + r_1\,s_1\,{s_2}^2\,f_{18}^* + r_1\,s_1\,{s_2}^2\,f_{19}^* + r_1\,{r_2}^2\,s_1\,f_{20}^*\right)$
$\frac{001}{011}, \frac{100}{110}$	f_{12}^*	$\frac{1}{2}\left(r_2\,s_2\,f_7^* + r_1\,s_1\,f_{10}^* + r_2\,s_2\,f_{11}^* + f_{12}^* + r_1\,s_1\,f_{16}^* + r_1\,r_2\,s_1\,s_2\,f_{17}^* + r_1\,r_2\,s_1\,s_2\,f_{18}^* + r_1\,r_2\,s_1\,s_2\,f_{19}^* + r_1\,r_2\,s_1\,s_2\,f_{20}^*\right)$
$\frac{001}{100}, \frac{011}{110}$	f_{13}^*	$\frac{1}{2}\left({r_{13}}^2\,f_9^* + {s_{13}}^2\,f_{13}^* + r_1\,r_2\,s_1\,s_2\,f_{17}^* + r_1\,r_2\,s_1\,s_2\,f_{18}^* + r_1\,r_2\,s_1\,s_2\,f_{19}^* + r_1\,r_2\,s_1\,s_2\,f_{20}^*\right)$
$\frac{001}{101}, \frac{010}{110}$	f_{14}^*	$\frac{1}{2}\left(r_{13}\,s_{13}\,f_9^* + r_1\,s_1\,f_{10}^* + r_{13}\,s_{13}\,f_{13}^* + f_{14}^* + r_1\,s_1\,f_{16}^* + r_1\,{r_2}^2\,s_1\,f_{17}^* + r_1\,s_1\,{s_2}^2\,f_{18}^* + r_1\,s_1\,{s_2}^2\,f_{19}^* + r_1\,{r_2}^2\,s_1\,f_{20}^*\right)$
$\frac{010}{011}, \frac{100}{101}$	f_{15}^*	$\frac{1}{2}\left(r_2\,s_2\,f_7^* + r_{13}\,s_{13}\,f_9^* + r_2\,s_2\,f_{11}^* + r_{13}\,s_{13}\,f_{13}^* + f_{15}^* + r_2\,{s_1}^2\,s_2\,f_{17}^* + r_2\,{s_1}^2\,s_2\,f_{18}^* + {r_1}^2\,r_2\,s_2\,f_{19}^* + {r_1}^2\,r_2\,s_2\,f_{20}^*\right)$
$\frac{010}{100}, \frac{011}{101}$	f_{16}^*	$\frac{1}{2}\left({r_1}^2\,f_{10}^* + {s_1}^2\,f_{16}^* + r_2\,{s_1}^2\,s_2\,f_{17}^* + r_2\,{s_1}^2\,s_2\,f_{18}^* + {r_1}^2\,r_2\,s_2\,f_{19}^* + {r_1}^2\,r_2\,s_2\,f_{20}^*\right)$
$\frac{011}{100}$	f_{17}^*	$\frac{1}{2}\left({s_1}^2\,{s_2}^2\,f_{17}^* + {r_2}^2\,{s_1}^2\,f_{18}^* + {r_1}^2\,{r_2}^2\,f_{19}^* + {r_1}^2\,{s_2}^2\,f_{20}^*\right)$
$\frac{010}{101}$	f_{18}^*	$\frac{1}{2}\left({r_2}^2\,{s_1}^2\,f_{17}^* + {s_1}^2\,{s_2}^2\,f_{18}^* + {r_1}^2\,{s_2}^2\,f_{19}^* + {r_1}^2\,{r_2}^2\,f_{20}^*\right)$
$\frac{001}{110}$	f_{19}^*	$\frac{1}{2}\left({r_1}^2\,{r_2}^2\,f_{17}^* + {r_1}^2\,{s_2}^2\,f_{18}^* + {s_1}^2\,{s_2}^2\,f_{19}^* + {r_2}^2\,{s_1}^2\,f_{20}^*\right)$
$\frac{000}{111}$	f_{20}^*	$\frac{1}{2}\left({r_1}^2\,{s_2}^2\,f_{17}^* + {r_1}^2\,{r_2}^2\,f_{18}^* + {r_2}^2\,{s_1}^2\,f_{19}^* + {s_1}^2\,{s_2}^2\,f_{20}^*\right)$

Notations: $s_l = 1 - r_l$; $r_{13} = r_1 + r_2 - 2\,r_1\,r_2$; $s_{13} = 1 - r_{13}$.

Table 3 Genotype frequencies at three loci in the second generation under full-sib mating. The genotypes corresponding to the f_i^* parameters are the same as in Table 2.

$$
\begin{aligned}
f_1^* &= \tfrac{1}{4}\, s_1^f\, s_2^f\, s_1^m\, s_2^m \\
f_2^* &= \tfrac{1}{4}\, r_2^f\, r_2^m\, s_1^f\, s_1^m \\
f_3^* &= \tfrac{1}{4}\, r_1^f\, r_2^f\, r_1^m\, r_2^m \\
f_4^* &= \tfrac{1}{4}\, r_1^f\, r_1^m\, s_2^f\, s_2^m \\
f_5^* &= \tfrac{1}{4}\left(r_2^f + r_2^m - 2\, r_2^f\, r_2^m\right) s_1^f\, s_1^m \\
f_6^* &= \tfrac{1}{4}\left(r_1^m\, r_2^m\, s_1^f\, s_2^f + r_1^f\, r_2^f\, s_1^m\, s_2^m\right) \\
f_7^* &= \tfrac{1}{4}\left(r_1^f + r_1^m - 2\, r_1^f\, r_1^m\right) s_2^f\, s_2^m \\
f_8^* &= \tfrac{1}{4}\left(r_1^f + r_1^m - 2\, r_1^f\, r_1^m\right) s_2^f\, s_2^m \\
f_9^* &= \tfrac{1}{4}\left(r_1^m\, r_2^m\, s_1^f\, s_2^f + r_1^f\, r_2^f\, s_1^m\, s_2^m\right) \\
f_{10}^* &= \tfrac{1}{4}\left(r_2^f + r_2^m - 2\, r_2^f\, r_2^m\right) s_1^f\, s_1^m
\end{aligned}
\qquad
\begin{aligned}
f_{11}^* &= \tfrac{1}{4}\left(r_1^f + r_1^m - 2\, r_1^f\, r_1^m\right) r_2^f\, r_2^m \\
f_{12}^* &= \tfrac{1}{4}\left(r_1^f\, r_2^m\, s_2^f\, s_1^m + r_2^f\, r_1^m\, s_1^f\, s_2^m\right) \\
f_{13}^* &= \tfrac{1}{4}\left(r_1^f\, r_2^m\, s_2^f\, s_1^m + r_2^f\, r_1^m\, s_1^f\, s_2^m\right) \\
f_{14}^* &= \tfrac{1}{4}\left(r_1^f + r_1^m - 2\, r_1^f\, r_1^m\right) r_2^f\, r_2^m \\
f_{15}^* &= \tfrac{1}{4}\left(r_2^f + r_2^m - 2\, r_2^f\, r_2^m\right) r_1^f\, r_1^m \\
f_{16}^* &= \tfrac{1}{4}\left(r_2^f + r_2^m - 2\, r_2^f\, r_2^m\right) r_1^f\, r_1^m \\
f_{17}^* &= \tfrac{1}{2}\, r_1^f\, r_1^m\, s_2^f\, s_2^m \\
f_{18}^* &= \tfrac{1}{2}\, r_1^f\, r_2^f\, r_1^m\, r_2^m \\
f_{19}^* &= \tfrac{1}{2}\, r_2^f\, r_2^m\, s_1^f\, s_1^m \\
f_{20}^* &= \tfrac{1}{2}\, s_1^f\, s_2^f\, s_1^m\, s_2^m
\end{aligned}
$$

r_l^f (r_l^m): recombination rate in females (males) between adjacent loci l and $(l+1)$; $s_l = 1 - r_l$.