A Computational Approach to Developing Mathematical Models of Polyploid Meiosis

Marc Rehmsmeier[1]

Gregor Mendel Institute of Molecular Plant Biology GmbH, 1030 Vienna, Austria, and Computational Biology Unit, Uni Computing, Uni Research, 5008 Bergen, Norway

ABSTRACT Mathematical models of meiosis that relate offspring to parental genotypes through parameters such as meiotic recombination frequency have been difficult to develop for polyploids. Existing models have limitations with respect to their analytic potential, their compatibility with insights into mechanistic aspects of meiosis, and their treatment of model parameters in terms of parameter dependencies. In this article I put forward a computational approach to the probabilistic modeling of meiosis. A computer program enumerates all possible paths through the phases of replication, pairing, recombination, and segregation, while keeping track of the probabilities of the paths according to the various parameters involved. Probabilities for classes of genotypes or phenotypes are added, and the resulting formulas are simplified by the symbolic-computation system Mathematica. An example application to autotetraploids results in a model that remedies the limitations of previous models mentioned above. In addition to the immediate implications, the computational approach presented here can be expected to be useful through opening avenues for modeling a host of processes, including meiosis in higher-order ploidies.

MEIOTIC recombination is the exchange of chromosomal regions between homologous chromosomes during meiosis, a specialized type of cell division in sexually reproducing eukaryotes (Lao and Hunter 2010; Lichten and De Massy 2011; Pradillo and Santos 2011). Meiotic recombination ensures the proper segregation of chromosomes through the formation of crossovers between interhomologous DNA strands (Lichten and De Massy 2011) and increases genetic diversity among offspring (Yanowitz 2010).

Polyploidy is the accumulation of additional complete sets of chromosomes and is widespread among flowering plants, including important crop species (Brownfield and Köhler 2011; Jackson *et al.* 2011). While evidence that polyploidy has promoted speciation or morphological diversification is lacking (Otto 2007), the possibility has been acknowledged (Comai 2005; Otto 2007), and it has recently been suggested that increased rates of meiotic recombination might have contributed to the evolutionary success of polyploid plants (Pecinka *et al.* 2011).

Tetraploids are generally classified as allotetraploids or autotetraploids. Allotetraploids such as *Arabidopsis suecica* (Pecinka *et al.* 2011) are derived from two different (diploid) species (Brownfield and Köhler 2011) and are traditionally considered to show pairing between homologous chromosomes only, although pairing between homeologous chromosomes may occur (Yang *et al.* 2013). Autotetraploid genomes are derived from a single species (Brownfield and Köhler 2011) and generally show no pairing preference among the four homologous chromosomes, although again preferential pairing might occur (R. Wu *et al.* 2001).

Mathematical models of meiosis that relate the genotypes of offspring to those of their parents at single- or multiple-marker locations through parameters such as the meiotic recombination frequency between a marker pair are easily formulated for diploids and allopolyploids, but have been difficult to develop for autopolyploids. The difficulties lie in the more complex meiosis in autopolyploids, in which not only bivalents are formed—pairs of homologous chromosomes, parts of which are then exchanged later through crossovers—but possibly also higher-order valents such as quadrivalents. In quadrivalents, each chromosomal region is still paired with only one other chromosome, but pairing
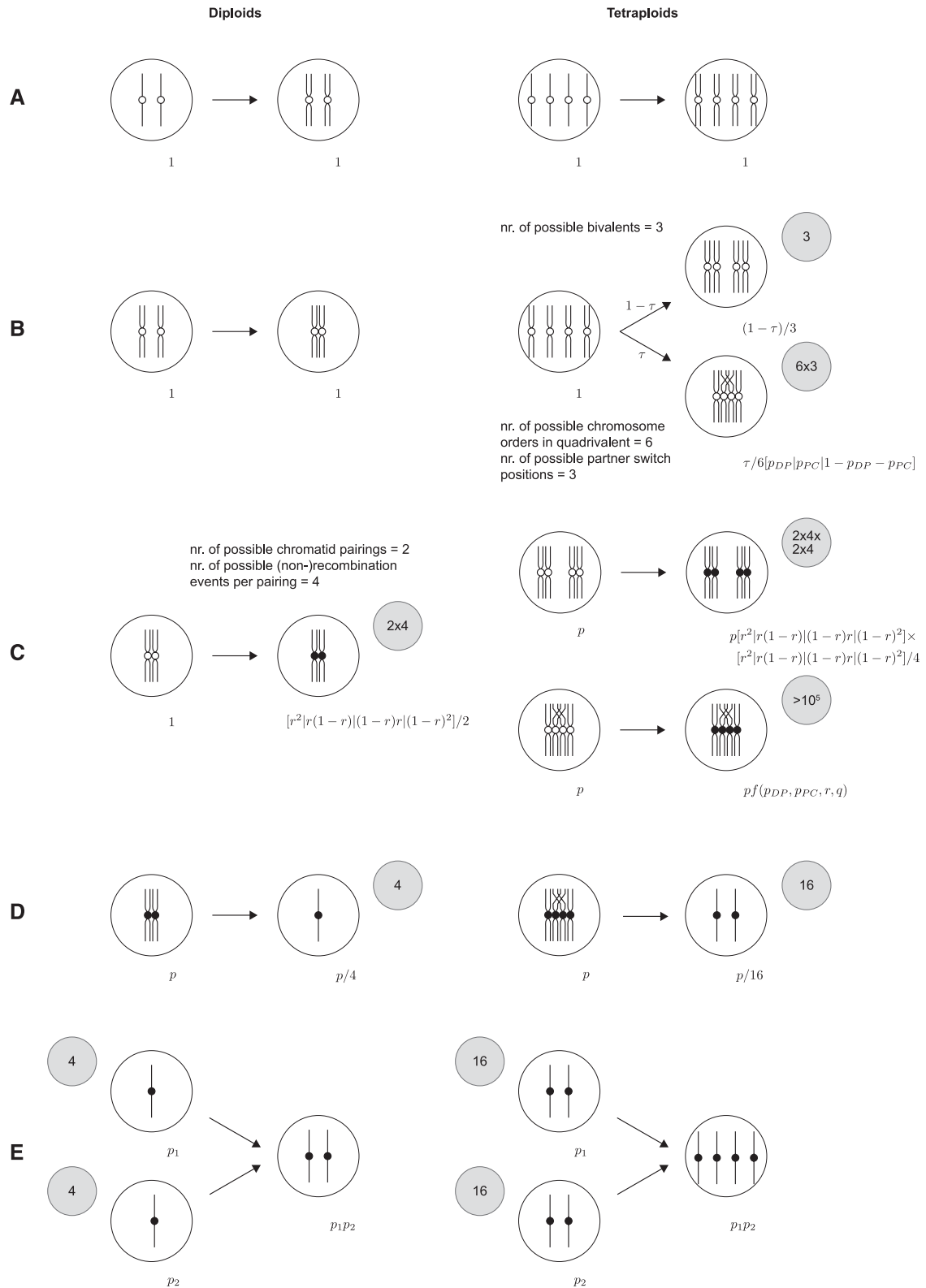
Figure 1 (A–E) Models of meiotic recombination in diploids and autotetraploids. The left column shows diploids and the right column tetraploids. (A) Replication of chromosomes. (B) Pairing of chromosomes into bivalents (diploids, tetraploids) and quadrivalents (tetraploids). (C) Recombination of paired chromosomes. See also Figure 2. (D) Distribution of (possibly) recombined chromosomes into gametes. (E) Crossing of gametes into seeds. Shaded circles show overall numbers of possibilities in that step. 1, p, p_1, and p_2 on the left sides of steps denote probabilities from previous steps. Formulas on the right sides of steps denote probabilities up to and including the current step. Alternatives are separated by vertical bars in square

partners can switch in a position-dependent manner. Such pairing-partner switches can lead to the phenomenon of double reduction, in which homologous segments of sister chromatids—"sister factors" (Levings and Alexander 1966)—end up in the same gamete (Mather 1935; Haynes and Douches 1993). Double reduction of a locus requires homologous chromosomes to form a quadrivalent and a recombination between the locus and the centromere (Mather 1936; Luo *et al.* 2004). Estimates of coefficients of double reduction can provide insights into the extent of quadrivalent formation (see, *e.g.*, Wang and Luo 2012), and estimates of the recombination frequency r are necessary for the construction of genetic maps (see, *e.g.*, Ripol *et al.* 1999).

In 1947, Fisher described 11 modes of tetraploid gamete formation for two linked loci (Fisher 1947), but it was not until over half a century later that frequencies of these gamete types were expressed in terms of recombination and double-reduction parameters under a tetrasomic model of inheritance that takes the formation of quadrivalents into account (S. S. Wu *et al.* 2001; Luo *et al.* 2004; Wu and Ma 2005; Lu *et al.* 2012). The resulting sets of equations, however, have at least three limitations: First, the models are descriptive rather than analytical; that is, they do not explain the probability distribution of gamete modes in terms of biological (here meiotic) mechanisms; instead, gamete modes are described in terms of observables, that is, whether a gamete is recombined or not and double reduced or not. Second, as a consequence, the resulting models might not be compatible with any given mechanistic model of meiosis, either agreed upon for a particular species or hypothesized with the aim of further testing. Third, the parameterizations of probability distributions treat recombination rates and coefficients of double reduction as independent, thus putting aside the fact that recombination is required to allow double reduction to take place.

In this article I put forward a computational approach to the probabilistic modeling of meiosis, with an example application to autotetraploids. The approach takes account of the major mechanistic aspects of autotetraploid meiosis, including mixed bivalent and quadrivalent pairings, pairing-partner switches, and recombination on a chromatid level. A computer program enumerates all possible paths through the phases of replication, pairing, recombination, and segregation, while keeping track of the probabilities of the paths according to the various parameters involved. Probabilities for classes of genotypes or phenotypes are added, and the resulting formulas are simplified by the symbolic-computation system Mathematica (Wolfram Research, Inc. 2010).

The approach is applied to the case of extreme autotetraploids with tetrasomic inheritance, the mixed formation of bivalents and quadrivalents, two linked marker loci on the same arm of a chromosome, and one partner switch per quadrivalent. Model parameters model the ratio of quadrivalent to bivalent formation (τ), the location of the partner switch (between the centromere and the centromere-proximal locus, between the two loci, or outside these two regions), the recombination frequency between the centromere and the centromere-proximal locus (q), and the recombination frequency between the two loci (r).

The resulting model is profoundly different from previous ones in several aspects, including the distribution of probability mass across gamete modes, most strikingly by one of the modes having zero probability. The coefficient of double reduction at the centromere-proximal locus, α, and the coefficient of double reduction at the centromere-distal locus, β, are not independent variables, but are fully explained through the other parameters. Further symbolic elimination of the partner switch location parameters turns the five-parameter model into a four-parameter model of r, q, α, and β; the mixing parameter τ disappears, showing that the ratio of quadrivalent to bivalent formation does not influence the frequencies of gamete modes beyond α and β.

The resulting model is also applied to a recent study of recombination frequencies in tetraploids based on a seed-fluorescence assay (Pecinka *et al.* 2011) and in a simulation study, in which it is compared to previously published models.

In addition to the immediate implications described, the computational approach presented here can be expected to be useful through opening avenues for modeling a host of processes, among them manifestations of meiosis that contain multiple-partner switches, crossover and chromatid interference, and meiosis in higher-order ploidies, such as hexaploidy and beyond.

Materials and Methods

The definition of the 11 gamete modes is taken from Table 1 in Luo *et al.* (2004), but see also Table 10 in Fisher (1947), there in a different order. With A denoting the centromere-proximal locus and B the centromere-distal locus, and with $1 \leq i, j, k, l \leq 4$ (i, j, k, l all different) denoting the four copies of the marker-carrying chromosome, the 11 gamete modes are defined as

1: A_iB_i / A_iB_i
2: A_iB_j / A_iB_j

---

brackets, r is the frequency of recombination between the two marker loci, q is the recombination frequency between the centromere and the centromere-proximal marker, τ is the quadrivalent-to-bivalent ratio, and p_{DP} and p_{PC} are the probabilities of having a partner switch between the markers and between the centromere-proximal marker and the centromere, respectively. $f(p_{DP}, p_{PC}, r, q)$ denotes a function of the parameters indicated.
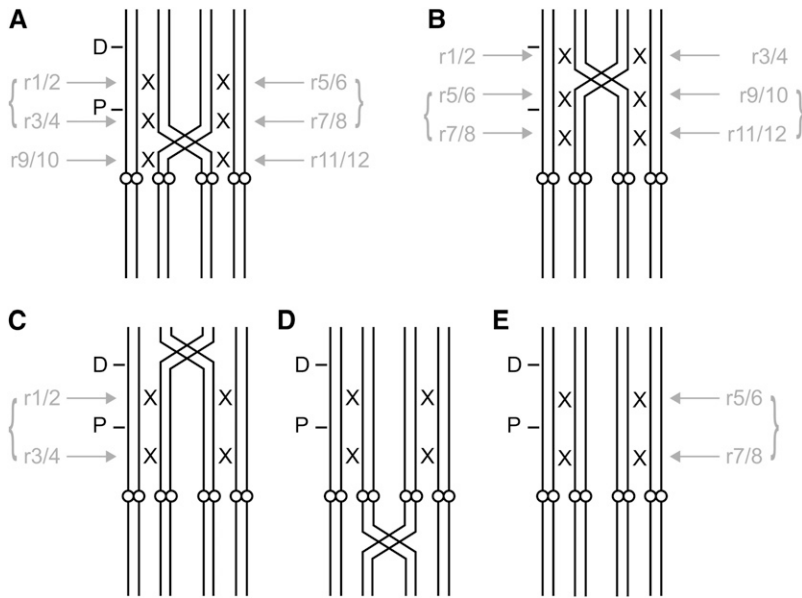
Figure 2 (A–E) Chromosome pairing and recombination events. (A) Quadrivalent with partner switch between centromere and centromere-proximal locus. (B) Quadrivalent with partner switch between loci. (C and D) Quadrivalents with partner switch outside the two regions in A and B, effectively equivalent to E. (E) Two bivalents (no partner switch). D, centromere-distal locus; P, centromere-proximal locus; open circles, centromeres. X's denote possible recombination events (two events per X). Shaded text and arrows describe the order of recombination events taken account of in the computer program. Curly brackets indicate that chromatid partner choice is not independent. Annotation of recombination events is implicitly the same in C–E.

3: A_iB_i / A_iB_j
4: A_iB_j / A_iB_k
5: A_iB_i / A_jB_i
6: A_iB_j / A_kB_j
7: A_iB_i / A_jB_j
8: A_iB_i / A_jB_k
9: A_iB_j / A_jB_i
10: A_iB_j / A_jB_k
11: A_iB_j / A_kB_l

with the two columns of A's and B's showing the two chromosomes of the gamete.

The computational procedure, as described in detail in the *Results* section, has been implemented in the functional programming language Haskell. The computer program outputs sums of probabilities for gamete modes and other classes of interest that are then simplified by the symbolic-computation system Mathematica (Wolfram Research, Inc. 2010). The program together with ASCII text versions of the formulas presented in this article are available from the author.

Results

A five-parameter model

The approach put forward here is to computationally enumerate all possible paths of meiosis through the phases of replication, pairing, recombination, and segregation, while keeping track of the probabilities of the paths according to the various parameters involved. Probabilities for classes of genotypes or phenotypes are added, and the resulting formulas are simplified by the symbolic-computation system Mathematica. As an example application, I present the development of a model of meiosis in an autotetraploid with two linked marker loci on the same arm of a chromosome, taking into account mixed bivalent and quadrivalent pairings, pairing-partner switches, and

recombination on a chromatid level (see Figure 1, together with a comparison to diploid meiosis).

Autotetraploids are modeled here as extreme autotetraploids, that is, without pairing preferences between chromosomes. Chromosomes can form either quadrivalents or bivalents, following a quadrivalence-to-bivalence ratio τ.

Consistent with a model of quadrivalent formation in which chromosome pairing starts at the ends of chromosomes, a (single) pairing-partner switch can occur such that one end of a chromosome pairs with one chromosome and the other end pairs with a different chromosome. Upon completion of pairing, this manifests itself in a partner switch, the location of which is either between the two markers (occurring with probability p_{DP}, D for "distal" and P for "proximal"), between the centromere-proximal marker and the centromere (occurring with probability p_{PC}), or outside these regions (occurring with probability $1 - p_{DP} - p_{PC}$). For simplicity, in the first two cases, partner switches are modeled to be in the middle of the respective regions; the third case is treated as equivalent to the formation of two bivalents, and the location of the switch is irrelevant.

Chromatids of paired (segments of) chromosomes are modeled as pairing randomly, with sister chromatids not being allowed to pair. Depending on the valency and, in the case of a quadrivalent, on the position of the partner switch, up to three (effective) recombination events per chromatid can occur (see Figure 2). If the region between proximal and distal markers is partner switch free, recombination between these loci occurs with frequency r. If the partner switch is in this region, two recombinations can occur, one in each of the halves defined by the partner switch, with frequency $r' = (1 - \sqrt{1 - 2r})/2$. Similarly, for the region between centromere-proximal marker and centromere, either one recombination event with frequency q or two events with frequency $q' = (1 - \sqrt{1 - 2q})/2$ each can occur. The half-recombination frequencies r' and q' are solutions of

Table 1 Mean estimates of recombination frequency r and standard deviations of estimates (in parentheses), relative to true values

		Relative mean estimated r			
r	q	Five-parameter model	Luo *et al.* (2004)	S. S. Wu *et al.* (2001)	Wu and Ma (2005)
0.01	0.01	0.97 (0.025)	0.91 (0.023)	0.91 (0.023)	0.91 (0.023)
0.01	0.10	0.99 (0.024)	0.91 (0.022)	0.91 (0.022)	0.91 (0.022)
0.01	0.30	1.00 (0.025)	0.92 (0.022)	0.92 (0.022)	0.92 (0.022)
0.10	0.01	0.99 (0.022)	0.95 (0.020)	0.95 (0.020)	0.95 (0.020)
0.10	0.10	0.99 (0.022)	0.95 (0.020)	0.95 (0.020)	0.95 (0.020)
0.10	0.30	1.00 (0.022)	0.95 (0.021)	0.95 (0.021)	0.95 (0.021)
0.30	0.01	1.00 (0.023)	1.03 (0.023)	1.04 (0.024)	1.03 (0.023)
0.30	0.10	1.00 (0.022)	1.03 (0.022)	1.04 (0.024)	1.03 (0.022)
0.30	0.30	1.00 (0.020)	1.04 (0.022)	1.05 (0.023)	1.04 (0.022)

r and q are varied as indicated; τ, p_{DP}, and p_{PC} are fixed to 0.67, 0.21, and 0.34, respectively.

$r = 2r'(1 - r')$ and $q = 2q'(1 - q')$, respectively, expressing the fact that a recombination in the whole interval effectively occurs if and only if a recombination occurs in any of the halves but not in the other. In a partner switch free region, modeling two half-recombinations with frequency r' would be equivalent to one recombination with frequency r and likewise for q' and q. Every region or half-region of a chromatid can undergo recombination with the respective region of the pairing partner. Recombination events are modeled as being independent. Every region or half-region can be subject to at most one (effective) recombination event. Since recombination events in the distal-marker telomere region and the arm that does not carry the markers have no effect on marker distributions, they can be ignored here.

Paired centromeres together with their attached chromatids are randomly distributed toward opposite poles (modeling anaphase I), and the two chromatids attached to a common centromere are again randomly distributed toward different poles (modeling anaphase II). This results in four gametes, each carrying two chromatids that are not sister chromatids. Note though that through recombination, parts of chromatids that were originally sister chromatids (sister factors) can now end up in the same gamete.

Sorting the resulting gametic products into gamete modes (see *Materials and Methods* for a definition of modes), and summing up probabilities per mode, results, after simplification with Mathematica, in the following five-parameter model, with r being the recombination frequency between the two markers, q the recombination frequency between the centromere and the centromere-proximal marker, τ the quadrivalent-to-bivalent ratio, p_{DP} the probability of a pairing-partner switch between the two markers, p_{PC} the probability of a pairing-partner switch between the centromere and the centromere-proximal marker, p_1–p_{11} the probabilities of the 11 modes, and r' the half-recombination frequency as defined above:

$$P_1 = P_{PC}\left(\frac{q}{4} - \frac{qr}{2} + \frac{qr^2}{4}\right)\tau$$

$$P_2 = \frac{P_{PC}\,qr^2\tau}{4}$$

$$P_3 = P_{PC}\left(\frac{qr}{2} - \frac{qr^2}{2}\right)\tau$$

$$P_4 = 0$$

$$P_5 = P_{PC}\left(\frac{r}{2} - \frac{qr}{2} - \frac{r^2}{2} + \frac{qr2}{2}\right)\tau + P_{DP}\left(\frac{r}{4} - \frac{rr'}{4}\right)\tau$$

$$P_6 = \frac{P_{DP}\,rr'\tau}{4}$$

$$P_7 = 1 - 2r + r^2 + P_{PC}\left(-\frac{q}{4} + \frac{r}{4} + \frac{qr}{2} - \frac{r^2}{4} - \frac{qr^2}{4}\right)\tau + P_{DP}\left(\frac{5r}{8} - \frac{13r^2}{16} - r' + \frac{7rr'}{8}\right)\tau$$

$$P_8 = 2r - 2r^2 + P_{PC}\left(-r + r^2\right)\tau + P_{DP}\left(-\frac{r}{2} + \frac{13r^2}{8} - \frac{3rr'}{2}\right)\tau$$

$$P_9 = P_{PC}\left(\frac{r}{4} + \frac{r^2}{4} - \frac{qr2}{4}\right)\tau + P_{DP}\left(-\frac{r}{8} + \frac{3r^2}{16} + \frac{r'}{2} - \frac{5rr'}{8}\right)\tau$$

$$P_{10} = P_{DP}\left(\frac{-3r^2}{8} + \frac{3rr'}{2}\right)\tau$$

$$P_{11} = r^2 - \frac{P_{PC}\,r^2\tau}{2} + P_{DP}\left(-\frac{r}{4} - \frac{5r^2}{8} + \frac{r'}{2} - \frac{rr'}{4}\right)\tau.$$

Coefficients of double reduction

Instead of sorting gametic products into gamete modes we can also classify them according to whether they are double reduced, in the centromere-proximal marker locus, in the centromere-distal marker locus, or in both marker loci. Summing up probabilities defines the coefficients of double reduction at the proximal locus, α, and at the distal locus, β, as

$$\alpha = \frac{P_{PC}q\tau}{4} \tag{1}$$

$$\beta = \frac{P_{DP}r\tau}{4} + P_{PC}\left(\frac{q}{4} + \frac{r}{4} - qr - \frac{r^2}{2} + qr^2\right)\tau. \tag{2}$$

Table 2 Mean estimates of the coefficient of double reduction α and standard deviations of estimates (in parentheses), relative to true values

			Relative mean estimated α		
r	q	α	Five-parameter model	Luo *et al.* (2004)	S. S. Wu *et al.* (2001)
0.01	0.01	0.00057	1.07 (0.029)	1.07 (0.029)	1.11 (0.033)
0.01	0.10	0.00570	0.97 (0.027)	0.97 (0.028)	0.98 (0.031)
0.01	0.30	0.01708	0.97 (0.028)	0.98 (0.028)	0.99 (0.032)
0.10	0.01	0.00057	1.07 (0.029)	1.07 (0.029)	1.11 (0.033)
0.10	0.10	0.00570	0.97 (0.028)	0.97 (0.028)	0.98 (0.031)
0.10	0.30	0.01708	0.97 (0.028)	0.98 (0.028)	0.99 (0.032)
0.30	0.01	0.00057	1.07 (0.029)	1.07 (0.029)	1.11 (0.033)
0.30	0.10	0.00570	0.97 (0.027)	0.97 (0.028)	0.98 (0.031)
0.30	0.30	0.01708	0.97 (0.028)	0.98 (0.028)	0.99 (0.032)

r and q are varied as indicated; τ, p_{DP}, and p_{PC} are fixed to 0.67, 0.21, and 0.34, respectively; the true α is determined by q, τ, and p_{PC} through Equation 1.

Elimination of parameters

The above probabilities of the 11 gamete modes are defined in terms of the five parameters r, q, τ, p_{DP}, and p_{PC}. Since the coefficients of double reduction, α and β, are also defined in terms of p_{DP} and p_{PC}, we can solve these two equations for p_{DP} and p_{PC} and substitute the solutions in the gamete mode equations, effectively eliminating p_{DP} and p_{PC}, and making α and β parameters. This procedure results in the following model, with p_1^e–p_{11}^e being the probabilities of the 11 modes, with parameters eliminated (indicated by the superscript e) and again r' as defined above:

$$p_1^e = \alpha(1 - 2r + r^2)$$

$$p_2^e = \alpha r^2$$

$$p_3^e = \alpha(2r - 2r^2)$$

$$p_4^e = 0$$

$$p_5^e = \beta(1 - r') + \alpha\left(-1 + 2r - 2r^2 + r' - 4rr' + \frac{2rr'}{q} + 4r^2r' - \frac{2r^2r'}{q}\right)$$

$$p_6^e = \beta r' + \alpha\left(-r' + 4rr' - \frac{2rr'}{q} - 4r^2r' + \frac{2r^2r'}{q}\right)$$

$$p_7^e = 1 - 2r + r^2 + \beta\left(\frac{5}{2} - \frac{13r}{4} + \frac{7r'}{2} - \frac{4r'}{r}\right) + \alpha\left(-\frac{7}{2} + \frac{61r}{4} - \frac{4r}{q} - 24r^2 + \frac{21r^2}{2q} + 13r^3 - \frac{13r^3}{2q} - \frac{39r'}{2} + \frac{8r'}{q} + \frac{4r'}{r} + 30rr' - \frac{15rr'}{q} - 14r^2r' + \frac{7r^2r'}{q}\right)$$

$$p_8^e = 2r - 2r^2 + \beta\left(-2 + \frac{13r}{2} - 6r'\right) + \alpha\left(2 - \frac{29r}{2} + 34r^2 - \frac{13r^2}{q} - 26r^3 + \frac{13r^3}{q} + 6r' - 24rr' + \frac{12rr'}{q} + 24r^2r' - \frac{12r^2r'}{q}\right)$$

$$p_9^e = \beta\left(-\frac{1}{2} + \frac{3r}{4} - \frac{5r'}{2} + \frac{2r'}{r}\right) + \alpha\left(\frac{1}{2} - \frac{11r}{4} + \frac{2r}{q} + 4r^2 - \frac{3r^2}{2q} - 3r^3 + \frac{3r^3}{2q} + \frac{21r'}{2} - \frac{4r'}{q} - \frac{2r'}{r} - 18rr' + \frac{9rr'}{q} + 10r^2r' - \frac{5r^2r'}{q}\right)$$

$$p_{10}^e = \beta\left(\frac{-3r}{2} + 6r'\right) + \alpha\left(\frac{3r}{2} - 6r^2 + \frac{3r^2}{q} + 6r^3 - \frac{3r^3}{q} - 6r' + 24rr' - \frac{12rr'}{q} - 24r^2r' + \frac{12r^2r'}{q}\right)$$

$$p_{11}^e = r^2 + \beta\left(-1 - \frac{5r}{2} - r' + \frac{2r'}{r}\right) + \alpha\left(1 - \frac{3r}{2} + \frac{2r}{q} - 6r^2 + \frac{r^2}{q} + 10r^3 - \frac{5r^3}{q} + 9r' - \frac{4r'}{q} - \frac{2r'}{r} - 12rr' + \frac{6rr'}{q} + 4r^2r' - \frac{2r^2r'}{q}\right).$$

Surprisingly, the quadrivalent-to-bivalent ratio τ also disappears, indicating that it does not influence gamete mode probabilities beyond the coefficients of double reduction α and β. The resulting model is now one of four parameters, r, q, α, and β. Note, however, that α and β are not independent parameters, but depend on r and q, and that probabilities are undefined for $r = 0$ or $q = 0$. For $r = 0$, limits can be found by setting $r = 0$ in the original p_1–p_{11} and eliminating parameters again, resulting in the probabilities of all gamete modes becoming zero, except modes 1 and 7:

$$\lim_{r \to 0} p_1^e = \alpha$$

$$\lim_{r \to 0} p_7^e = 1 - \alpha.$$

Note that $\alpha = \beta$ in this case. Limits for $q = 0$ cannot be found in the same way, since α becomes also zero, and, as a consequence, p_{DP} and p_{PC} cannot be both eliminated.

Table 3 Mean estimates of the coefficient of double reduction β and standard deviations of estimates (in parentheses), relative to true values

			Relative mean estimated β		
r	q	β	Five-parameter model	Luo et al. (2004)	S. S. Wu et al. (2001)
0.01	0.01	0.0020	0.98 (0.029)	3.3 (0.036)	0.92 (0.032)
0.01	0.10	0.0069	0.98 (0.025)	1.6 (0.031)	0.94 (0.030)
0.01	0.30	0.0179	0.99 (0.030)	1.2 (0.030)	0.99 (0.032)
0.10	0.01	0.0141	0.98 (0.028)	4.2 (0.032)	0.97 (0.031)
0.10	0.10	0.0174	1.00 (0.029)	3.7 (0.030)	0.99 (0.031)
0.10	0.30	0.0247	1.00 (0.029)	2.9 (0.029)	1.01 (0.030)
0.30	0.01	0.0346	0.98 (0.030)	4.7 (0.026)	1.02 (0.032)
0.30	0.10	0.0354	0.98 (0.030)	4.7 (0.025)	1.02 (0.032)
0.30	0.30	0.0372	0.99 (0.029)	4.6 (0.023)	1.01 (0.033)

r and q are varied as indicated; τ, p_{DP}, and p_{PC} are fixed to 0.67, 0.21, and 0.34, respectively; the true β is determined by r, q, τ, p_{DP}, and p_{PC} through Equation 2.

Application to a seed fluorescence assay

Melamed-Bessudo et al. (2005) introduced transgenic A. thaliana lines carrying two fluorescent seed markers that the authors used to estimate meiotic recombination frequencies (MRFs) between the two marker loci in diploid Arabidopsis plants. In Pecinka et al. (2011), the procedure was applied to tetraploidized Arabidopsis plants, with one copy of chromosome 3 carrying the markers on its top arm (green centromere-distal and red centromere-proximal), and MRFs were found to be significantly higher than in diploids. However, MRFs were estimated on the basis of a diploid model of meiosis, which is equivalent to a tetraploid model in which only bivalents form (see Wang and Luo 2012 and Rehmsmeier 2012 for an in-depth discussion). For a mixed bivalent–quadrivalent model we can specialize the five-parameter gamete model to the following equations (either by computationally sorting gametes into the various seed fluorescence classes or by manually adding probabilities from corresponding gamete modes), which describe the probabilities of seeds that contain only the red-fluorescent marker (p_r, "red only"), of seeds that contain only the green-fluorescent marker (p_g, "green only"), of seeds that contain both markers (p_y, "yellow"), and of seeds that contain none of the markers (p_b, "brown"):

$$P_r = \frac{r}{2} + P_{PC}\left(-\frac{r}{8} - \frac{qr}{8} - \frac{r^2}{8} + \frac{3qr^2}{16}\right)\tau$$
$$+ P_{DP}\left(\frac{-3r}{16} + \frac{r'}{4} - \frac{rr'}{16}\right)\tau$$

$$P_g = \frac{r}{2} + P_{PC}\left(-\frac{r}{4} + \frac{qr}{8} - \frac{qr^2}{16}\right)\tau + P_{DP}\left(-\frac{r}{4} + \frac{r'}{4} - \frac{rr'}{16}\right)\tau$$

$$P_y = \frac{1}{2} - \frac{r}{2} + P_{PC}\left(-\frac{q}{16} + \frac{r}{8} + \frac{qr}{8} + \frac{r^2}{8} - \frac{3qr^2}{16}\right)\tau$$
$$+ P_{DP}\left(\frac{3r}{16} - \frac{r'}{4} + \frac{rr'}{16}\right)\tau$$

$$p_b = \frac{1}{2} - \frac{r}{2} + P_{PC}\left(\frac{q}{16} + \frac{r}{4} - \frac{qr}{8} + \frac{qr^2}{16}\right)\tau$$

$$+ P_{DP}\left(\frac{r}{4} - \frac{r'}{4} + \frac{rr'}{16}\right)\tau.$$

Eliminating p_{DP} and p_{PC} (and implicitly τ), as was done above for the 11 gamete modes, results in the following four-parameter model of seed fluorescences:

$$p_r^e = \frac{r}{2} + \beta\left(-\frac{3}{4} - \frac{r'}{4} + \frac{r'}{r}\right)$$
$$+ \alpha\left(\frac{3}{4} - \frac{7r}{2} + \frac{r}{q} + \frac{15r^2}{4} - \frac{2r^2}{q} + \frac{17r'}{4} - \frac{2r'}{q} - \frac{r'}{r}\right.$$
$$\left. - 5rr' + \frac{5rr'}{2q} + r^2r' - \frac{r^2r'}{2q}\right)$$

$$p_g^e = \frac{r}{2} + \beta\left(-1 - \frac{r'}{4} + \frac{r'}{r}\right)$$
$$+ \alpha\left(1 - \frac{7r}{2} + \frac{r}{q} + \frac{15r^2}{4} - \frac{2r^2}{q} + \frac{17r'}{4} - \frac{2r'}{q} - \frac{r'}{r}\right.$$
$$\left. - 5rr' + \frac{5rr'}{2q} + r^2r' - \frac{r^2r'}{2q}\right)$$

$$p_y^e = \frac{1}{2} - \frac{r}{2} + \beta\left(\frac{3}{4} + \frac{r'}{4} - \frac{r'}{r}\right)$$
$$+ \alpha\left(-1 + \frac{7r}{2} - \frac{r}{q} - \frac{15r^2}{4} + \frac{2r^2}{q} - \frac{17r'}{4} + \frac{2r'}{q}\right.$$
$$\left. + \frac{r'}{r} + 5rr' - \frac{5rr'}{2q} - r^2r' + \frac{r^2r'}{2q}\right)$$

$$p_b^e = \frac{1}{2} - \frac{r}{2} + \beta\left(1 + \frac{r'}{4} - \frac{r'}{r}\right)$$
$$+ \alpha\left(-\frac{3}{4} + \frac{7r}{2} - \frac{r}{q} - \frac{15r^2}{4} + \frac{2r^2}{q} - \frac{17r'}{4} + \frac{2r'}{q}\right.$$
$$\left. + \frac{r'}{r} + 5rr' - \frac{5rr'}{2q} - r^2r' + \frac{r^2r'}{2q}\right).$$

Table 4 Mean estimates of recombination frequency q and standard deviations of estimates (in parentheses), relative to true values

r	q	Relative mean estimated q, five-parameter model
0.01	0.01	25.6 (2.27)
0.01	0.10	4.0 (0.57)
0.01	0.30	1.4 (0.26)
0.10	0.01	4.7 (1.24)
0.10	0.10	1.8 (0.55)
0.10	0.30	1.0 (0.23)
0.30	0.01	1.4 (0.18)
0.30	0.10	1.1 (0.19)
0.30	0.30	1.1 (0.19)

r and q are varied as indicated; τ, p_{DP}, and p_{PC} are fixed to 0.67, 0.21, and 0.34, respectively.

Note again that that probabilities are undefined for $r = 0$ or $q = 0$, but that limits for $r = 0$ can be found by setting $r = 0$ in the original p_r, p_g, p_y, and p_b and eliminating parameters again, resulting in the following equations:

$$\lim_{r \to 0} p_r^e = 0$$

$$\lim_{r \to 0} p_g^e = 0$$

$$\lim_{r \to 0} p_y^e = \frac{2 - \alpha}{4}$$

$$\lim_{r \to 0} p_b^e = \frac{2 + \alpha}{4}.$$

Note again that $\alpha = \beta$ in this case. Also again, limits for $q = 0$ cannot be found in the same way, since α becomes also zero, and, as a consequence, p_{DP} and p_{PC} cannot be both eliminated.

While the four-parameter model of seed fluorescences cannot be solved for the four parameters—since there are only three independent equations, the four equations summing up to unity—one can find the relationships

$$p_g^e - p_r^e = \frac{\alpha - \beta}{4}$$

$$p_b^e - p_y^e = \frac{\alpha + \beta}{4},$$

allowing estimation of the coefficients of double reduction α and β from seed count differences. Note that these same equations can also be derived from other models, such as the one in Luo et al. (2004) as used in Rehmsmeier (2012) (see Supporting Information, File S1), suggesting that this particular seed fluorescence assay is not sensitive to the underlying model of meiosis when coefficients of double reduction are to be determined. This, however, does not hold in cases where two or three copies of a chromosome carry the markers (see File S1).

Seed fluorescence probabilities for selfed plants can be defined by combining gamete probabilities, following the procedure in Wang and Luo (2012) and Rehmsmeier (2012).

Simulations

To understand the implications of the five-parameter model with respect to parameter estimation, I simulated gamete formation with fully informative markers (see, e.g., S. S. Wu et al. 2001) according to the five-parameter model and estimated parameters with the same model, with Luo et al.'s model (Luo et al. 2004) and with Wu et al.'s model (S. S. Wu et al. 2001; Wu and Ma 2005). Parameters r and q covered ranges from small to large values (0.01, 0.1, and 0.3), and τ, p_{DP}, and p_{PC} were fixed to 0.67, 0.21, and 0.34, respectively. The sample size was 1000, each sampling was repeated 100 times, and means and standard deviations were calculated from these 100 repetitions. Since gamete modes 7 and 9 cannot be distinguished by their genotypes, nor can modes 8 and 10 (see, e.g., S. S. Wu et al. 2001), the respective counts were added, as were the corresponding probability functions. S. S. Wu et al. (2001) and Wu and Ma (2005) give slightly different equations for the recombination frequency r (denoted θ in S. S. Wu et al. 2001), and both versions were used here. Parameters for the five-parameter model and for Luo et al.'s model were estimated with a maximum-likelihood approach, using R's optim function in conjunction with the BFGS method and gradient functions (R Core Team 2012). For Wu et al.'s model, the recombination frequency r was directly estimated with a closed-form solution to their equations describing r (Equation 6 in S. S. Wu et al. 2001 and Equations 4, 5, and 8 in Wu and Ma 2005), and coefficients of double reduction, α and β, were calculated directly as sums of observed gamete frequencies [$\hat{\alpha} = f_1 + f_2 + f_3 + f_4$, $\hat{\beta} = f_1 + f_2 + f_5 + f_6$ (S. S. Wu et al. 2001)]. For Luo et al.'s model, β was calculated from r and α according to Equation 1 in Luo et al. (2004). True values of α and β and their estimates from the five-parameter model were calculated from the (true or estimated, respectively) r, q, τ, p_{DP}, and p_{PC}.

Table 1 shows mean estimates of the recombination frequency r and standard deviations of the estimates, expressed relative to the true values. Mean estimates for the five-parameter model are mainly perfect (1.0) or near perfect (0.99); for very small values of r and q, the mean estimate is slightly lower, at 0.97. The mean estimates for Luo et al.'s and Wu et al.'s models are between 0.91 and 1.03 of the true values. Mean estimates for Luo et al.'s and Wu et al.'s models are identical, and mean estimates for S. S. Wu et al. (2001) and Wu and Ma (2005) are nearly identical. Relative standard deviations are similar for all methods and always very small (≤ 0.025).

Table 2 shows mean estimates of the coefficient of double reduction at the centromere-proximal locus, α, and standard deviations of the estimates, expressed relative to the true values. Mean estimates for the five-parameter model and for Luo et al.'s model are between 0.97 and 1.07 and are mainly identical. Estimates with the direct method (sums of

Table 5 Mean estimates of the quadrivalent-to-bivalent ratio τ, of the probability of a pairing-partner switch between the two markers p_{DP}, and of the probability of a pairing-partner switch between the centromere and the centromere-proximal marker p_{PC}, and standard deviations of estimates (in parentheses), relative to true values

		Five-parameter model		
r	q	Relative mean estimated τ	Relative mean estimated p_{DP}	Relative mean estimated p_{PC}
0.01	0.01	0.93 (0.43)	2.68 (0.82)	0.41 (0.40)
0.01	0.10	1.05 (0.38)	2.15 (0.79)	0.62 (0.41)
0.01	0.30	1.11 (0.30)	1.49 (0.69)	0.91 (0.39)
0.10	0.01	0.82 (0.25)	0.74 (0.58)	1.52 (0.47)
0.10	0.10	0.90 (0.29)	0.95 (0.65)	1.35 (0.51)
0.10	0.30	0.83 (0.26)	0.69 (0.51)	1.58 (0.46)
0.30	0.01	0.92 (0.25)	1.11 (0.32)	1.21 (0.34)
0.30	0.10	0.96 (0.26)	1.01 (0.36)	1.20 (0.40)
0.30	0.30	1.05 (0.26)	0.92 (0.29)	1.08 (0.36)

r and q are varied as indicated; τ, p_{DP}, and p_{PC} are fixed to 0.67, 0.21, and 0.34, respectively.

gamete mode frequencies) are between 0.98 and 1.11. Relative standard deviations are small for all methods, although slightly larger for the direct method.

Table 3 shows mean estimates of the coefficient of double reduction at the centromere-distal locus, β, and standard deviations of the estimates, expressed relative to the true values. Mean estimates from the five-parameter model are between 0.98 and 1.00. Mean estimates from sums of gamete mode frequencies are between 0.92 and 1.02. Estimates from Luo *et al.*'s model are between 1.2 and 4.7. Standard deviations are small for all methods.

Table 4 shows mean estimates of the recombination frequency q and standard deviations of the estimates, expressed relative to the true values. Mean estimates are accurate for the majority of cases, but deviate strongly from the true values for small values of r and q.

Table 5 shows mean estimates of the quadrivalent-to-bivalent ratio τ, the probability of a pairing-partner switch between the two markers p_{DP}, and the probability of a pairing-partner switch between the centromere and the centromere-proximal marker p_{PC} and standard deviations of the estimates, expressed relative to the true values. Mean estimates of τ are accurate, with values between 0.83 and 1.11, but have standard deviations of mainly between 0.25 and 0.3. Mean estimates of p_{DP} and p_{PC} are inaccurate for small values of r and q (r and $q = 0.01$), with values of 2.7 and 0.41. Mean estimates of p_{DP} and p_{PC} are accurate for larger values of r and q, with values between 0.69 and 1.6 for r and $q \geq 0.1$ and values of 0.92 and 1.1 for r and $q = 0.3$. Standard deviations for estimates of p_{DP} and p_{PC} are large, with values between 0.29 and 0.82.

Discussion

Owing to its complexities, polyploid meiosis has resisted the development of mathematical models. Indeed, Luo *et al.* point out that a tedious analysis was necessary to be able to express probabilities of gamete modes as functions of meiotic recombination frequency and the coefficient of double reduction (Luo *et al.* 2004). Any attempt to manually

formulate the complex model of tetraploid meiosis presented here—one that is consistent with postulated mechanistic behaviors such as the formation of pairing-partner switches—would be doomed. The computational approach to developing mathematical models of polyploid meiosis that is presented here makes complex models feasible to begin with.

In addition to the methodological advances, the resulting model of autotetraploid meiosis gives unexpected insights into the process of double reduction and its consequences for model parameterization and gamete mode probabilities. Luo *et al.*'s model (Luo *et al.* 2004) is a model of two parameters, the frequency of recombination between the two loci, r, and the coefficient of double reduction at the centromere-proximal locus, α. For double reduction at the centromere-proximal locus to occur, a partner switch between the centromere and the locus is necessary, and the probability of such a switch is, in addition to other requirements, expressed by α. Luo *et al.* couple the coefficient of double reduction at the centromere-distal locus, β, to α by $\beta = (\alpha(3 - 4r)^2 + 2r(3 - 2r))/9$ (Equation 1 in Luo *et al.* 2004). Double reduction at that locus, however, requires a partner switch between it and the centromere—the probability of which cannot be controlled by α and r alone (see Wu and Ma 2005 for a related discussion). In fact, letting one locus coincide with the centromere, thus forcing α to be zero, would make β entirely defined by r.

In response to this problem, Wu and Ma (2005) developed a model that is parameterized with β as well as with α and r. The model presented here contains a further parameter, the frequency of recombination between the centromere and the centromere-proximal locus, q. This additional parameter is naturally introduced when the possibility of a pairing-partner switch between the centromere and the centromere-proximal locus is taken into account. Indeed, upon setting the probability of such a pairing-partner switch, p_{PC} above, to zero, q disappears again from the equations, which can easily be seen in the definitions of p_1–p_{11} by noting that q occurs only with p_{PC} as a factor.
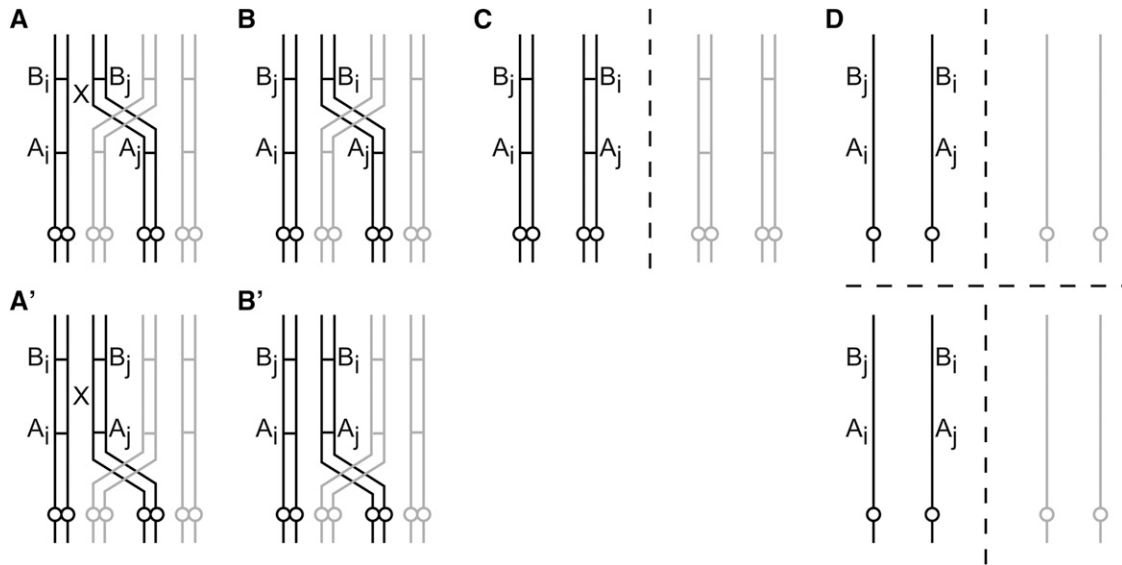
Figure 3 (A–D) Gamete mode 9 through a partner switch (but without double reduction). Partner switch is between the two loci, *A* and *B*. Circles denote centromeres. X denotes two recombinations. (A) A quadrivalent forms, and the two chromosomes (solid lines) recombine twice. (B) The recombined chromosomes. The distal markers, *B*, have been swapped. (C) Chromosomes move apart in anaphase I. (D) Chromatids move apart in anaphase II. Two gametes now have configuration A_iB_j/A_jB_i (gamete mode 9). A single recombination leads to one gamete of mode 9. A similar process is shown (with one or two recombinations) for the partner switch being between the centromere and the centromere-proximal locus (A' and B', continuing with C and D).

In the model presented here, the coefficients of double reduction, α and β, are not independent parameters, unlike α in Luo *et al.* (2004) and α and β in Wu and Ma (2005), but are defined in terms of the quadrivalent-to-bivalent ratio τ, the probabilities of partner switches, p_{DP} and p_{PC}, and the recombination frequencies r and q. The coefficients of double reduction are thus decomposed into what constitutes them mechanistically (in terms of τ, p_{DP}, p_{PC}, r, and q), allowing for a deeper understanding of these individual aspects.

The analytic nature of the coefficients of double reduction has surprising consequences for the dependence of gamete mode probabilities on these coefficients. When setting α and β to zero in the model presented here, all modes except modes 7, 8, and 11 reduce to probability zero, in contrast to the model of Luo *et al.*, in which, among others, modes 9 and 10 remain nonzero as well (mode 9 being described by the genotype A_iB_j/A_jB_i and mode 10 by A_iB_j/A_jB_k, with *A* denoting the centromere-proximal marker; *B* denoting the centromere-distal marker; the "/" distinguishing the two chromosomes; and *i*, *j*, and *k* denoting different chromosomes). The probability of mode 9 becoming zero appears counterintuitive at first glance, since this mode is not double reduced in either of the two loci. However, the manifestation of this mode requires a pairing-partner switch between the centromere and the centromere-distal locus (see Figure 3). Since here α and β are defined in terms of recombination frequencies r and q, pairing-partner switch probabilities p_{DP} and p_{PC}, and the quadrivalent-to-bivalent ratio τ, setting α and β to zero while leaving r and q untouched is equivalent to setting either p_{DP} and p_{PC} or τ to

zero, thus prohibiting the necessary formation of a pairing-partner switch. This becomes also apparent from the definition of p_9 in the five-parameter model. The requirement of a partner switch holds similarly for mode 10 (see Figure 4).

A further striking result of the model presented here is not only the generally different distribution of probability mass among the gamete modes, but also the fact that gamete mode 4 (A_iB_j/A_iB_k) has probability zero, even in the general case of mixed quadrivalent and bivalent formation. Since A_i is double reduced, there must be a partner switch between the centromere and the centromere-proximal locus. On the other hand, the two sister segments containing B_i have to cross over with two different chromosomes, *j* and *k*, which would require a second partner switch, between the two loci. Indeed, in a model that allows two partner switches, gamete mode 4 has nonzero probability (see File S2).

When the probabilities of pairing-partner switches, p_{DP} and p_{PC}, are eliminated and α and β are introduced, the quadrivalent-to-bivalent ratio parameter τ also disappears, indicating that τ does not contribute to the probabilities of gamete modes beyond its contribution to the coefficients of double reduction. In other words, gamete mode probabilities are fully defined in terms of recombination frequencies and coefficients of double reduction. Interestingly, this does not hold for a model in which a second partner switch can be formed (see File S3).

Double-reduction parameters in Luo *et al.* (2004) and Wu and Ma (2005) are conditioned on complete quadrivalency, thus being inconsistent with a definition of double reduction
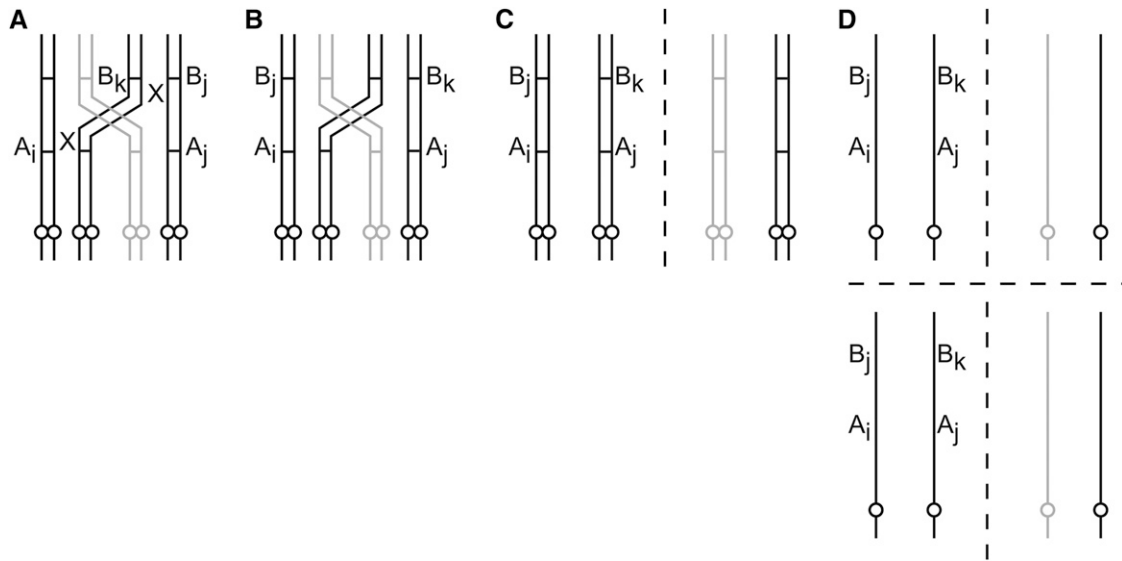
Figure 4 (A–D) Gamete mode 10 through a partner switch (but without double reduction). Partner switch is between the two loci, *A* and *B*. Circles denote centromeres. X's denote two recombinations each. (A) A quadrivalent forms, and the chromosomes (solid lines) recombine twice in each of the positions indicated by X. (B) The recombined chromosomes. The distal markers, *B*, have been swapped. (C) Chromosomes move apart in anaphase I. (D) Chromatids move apart in anaphase II. Two gametes have now configuration A_iB_j/A_jB_k (gamete mode 10). A single recombination leads to one gamete of mode 10.

that does not make assumptions in this respect. Estimates of double reduction can thus be expected to be biased when the formation of bivalents is not negligible. In contrast, in the model presented here, coefficients of double reduction explicitly depend on the quadrivalent-to-bivalent ratio τ and are thus not conditioned on any particular valency.

The results of the simulations show that estimates under the five-parameter model of the recombination frequency r are very accurate and more so than estimates under existing models. Estimates of the coefficient of double reduction α are similarly accurate for all methods. Interestingly, this is also the case for the five-parameter model, even though α is not estimated directly, but calculated from the estimates of the five parameters. Estimates of the coefficient of double reduction β are very accurate for the five-parameter model and similar to the direct method of S. S. Wu *et al.* (2001). Estimates of β under Luo *et al.*'s model, by using their Equation 1, deviate considerably from the true values, demonstrating that β cannot be accurately expressed through Luo *et al.*'s Equation 1 by α and r alone when gamete mode frequencies follow the five-parameter model presented here. Estimates of q, the recombination frequency between the centromere and the centromere-proximal locus, deviate strongly from the true value for very small values of r and q, such as r and $q = 0.01$. Closer inspection shows that for these values, the estimated q is frequently close to 1/2 (see Figure S1), explaining the relative mean estimate of ∼0.25. For such small values of r and q, several gamete modes have exceedingly small probabilities. For example, for $r = 0.01$ and $q = 0.01$ (and $\tau = 0.67$, $p_{DP} = 0.21$, $p_{PC} = 0.34$), the probability of gamete mode 2 is $<5.7e$-8. As a consequence, extremely large sample sizes would be necessary to generate

gamete counts for all modes and to allow for an accurate estimation of q. Nevertheless, for larger values of r and q, mean estimates of q are accurate. Assuming that recombination frequency grows monotonously with distance from the centromere, estimates of q could assist in genetic-map construction by directly suggesting a linear order of markers on a chromosome arm.

In this article I have put forward a computational approach to the probabilistic modeling of meiosis, with an example application to autotetraploids. The relative simplicity of automatically deriving such a model from a computational description of its underlying mechanisms and the multitude of aspects in which the resulting model distinguishes itself from previously published ones demonstrate the usefulness of the approach. Possibilities now arise for the development of more complex models, such as a tetraploid three-locus model with 107 gamete modes (Fisher 1947), hexaploid two-locus models with two mixing ratios (one hexavalent *vs.* one quadrivalent and one bivalent *vs.* three bivalents) and 40 gamete modes (Fisher 1947), or models of ploidies of even higher order.

It is worth noting that the approach presented here relieves the developer of mathematical models from the burden of having to argue about properties of gamete modes in terms of model parameters. Instead, the modeler can concentrate on a computational description of the process and leave the construction and simplification of expressions to the computer. The resulting mathematical models simply emerge automatically. Not only does this approach allow one to build mathematical models that would be impossible to build manually, but also it frees creative energy that can be invested in the definition of the structure of the model

and in the interpretation of the resulting mathematical formulas.

Acknowledgments

Literature Cited

Brownfield, L., and C. Köhler, 2011 Unreduced gamete formation in plants: mechanisms and prospects. J. Exp. Bot. 62(5): 1659–1668.

Comai, L., 2005 The advantages and disadvantages of being polyploid. Nat. Rev. Genet. 6(11): 836–846.

Fisher, R. A., 1947 The theory of linkage in polysomic inheritance. Philos. Trans. R. Soc. Lond. B Biol. Sci. 233(594): 55–87.

Haynes, K. G., and D. S. Douches, 1993 Estimation of the coefficient of double reduction in the cultivated tetraploid potato. Theor. Appl. Genet. 85(6): 857–862.

Jackson, S. A., A. Iwata, S. H. Lee, J. Schmutz, and R. Shoemaker, 2011 Sequencing crop genomes: approaches and applications. New Phytol. 191(4): 915–925.

Lao, J. P., and N. Hunter, 2010 Trying to avoid your sister. PLoS Biol. 8(10): e1000519.

Levings, C. S., and D. E. Alexander, 1966 Double reduction in autotetraploid maize. Genetics 54(6): 1297–1305.

Lichten, M., and B. de Massy, 2011 The impressionistic landscape of meiotic recombination. Cell 147(2): 267–270.

Lu, Y., X. Yang, C. Tong, X. Li, S. Feng *et al.*, 2012 A multivalent three-point linkage analysis model of autotetraploids. Brief. Bioinform. (in press).

Luo, Z. W., R. M. Zhang, and M. J. Kearsey, 2004 Theoretical basis for genetic linkage analysis in autotetraploid species. Proc. Natl. Acad. Sci. USA 101(18): 7040–7045.

Mather, K., 1935 Reductional and equational separation of the chromosomes in bivalents and multivalents. J. Genet. 30(1): 53–78.

Mather, K., 1936 Segregation and linkage in autotetraploids. J. Genet. 32(2): 287–314.

Melamed-Bessudo, C., E. Yehuda, A. R. Stuitje, and A. A. Levy, 2005 A new seed-based assay for meiotic recombination in Arabidopsis thaliana. Plant J. 43(3): 458–466.

Otto, S. P., 2007 The evolutionary consequences of polyploidy. Cell 131(3): 452–462.

Pecinka, A., W. Fang, M. Rehmsmeier, A. A. Levy, and O. Mittelsten Scheid, 2011 Polyploidization increases meiotic recombination frequency in Arabidopsis. BMC Biol. 9: 24.

Pradillo, M., and J. L. Santos, 2011 The template choice decision in meiosis: Is the sister important? Chromosoma 120(5): 447–454.

R Core Team, 2012 R: A *Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna. ISBN 3-900051-07-0, http://www.R-project.org.

Rehmsmeier, M., 2012 Response to Wang and Luo. BMC Biol. 10: 32.

Ripol, M. I., G. A. Churchill, J. A. da Silva, and M. Sorrells, 1999 Statistical aspects of genetic mapping in autopolyploids. Gene 235(1–2): 31–41.

Wang, L., and Z. Luo, 2012 Polyploidization increases meiotic recombination frequency in Arabidopsis: a close look at statistical modeling and data analysis. BMC Biol. 10: 30.

Wolfram Research, Inc., 2010 Mathematica, Version 8.0. Champaign, IL.

Wu, R., and C. X. Ma, 2005 A general framework for statistical linkage analysis in multivalent tetraploids. Genetics 170: 899–907.

Wu, R., M. Gallo-Meagher, R. C. Littell, and Z. B. Zeng, 2001 A general polyploid model for analyzing gene segregation in outcrossing tetraploid species. Genetics 159: 869–882.

Wu, S. S., R. Wu, C. X. Ma, Z. B. Zeng, M. C. K. Yang *et al.*, 2001 A multivalent pairing model of linkage analysis in autotetraploids. Genetics 159: 1339–1350.

Yang, X., Y. Lv, X. Pang, C. Tong, Z. Wang *et al.*, 2013 A unifying framework for bivalent multilocus linkage analysis of allotetraploids. Brief. Bioinform. 14(1): 96–108.

Yanowitz, J., 2010 Meiosis: making a break for it. Curr. Opin. Cell Biol. 22(6): 744–751.

Communicating editor: J. B. Holland

GENETICS

A Computational Approach to Developing Mathematical Models of Polyploid Meiosis
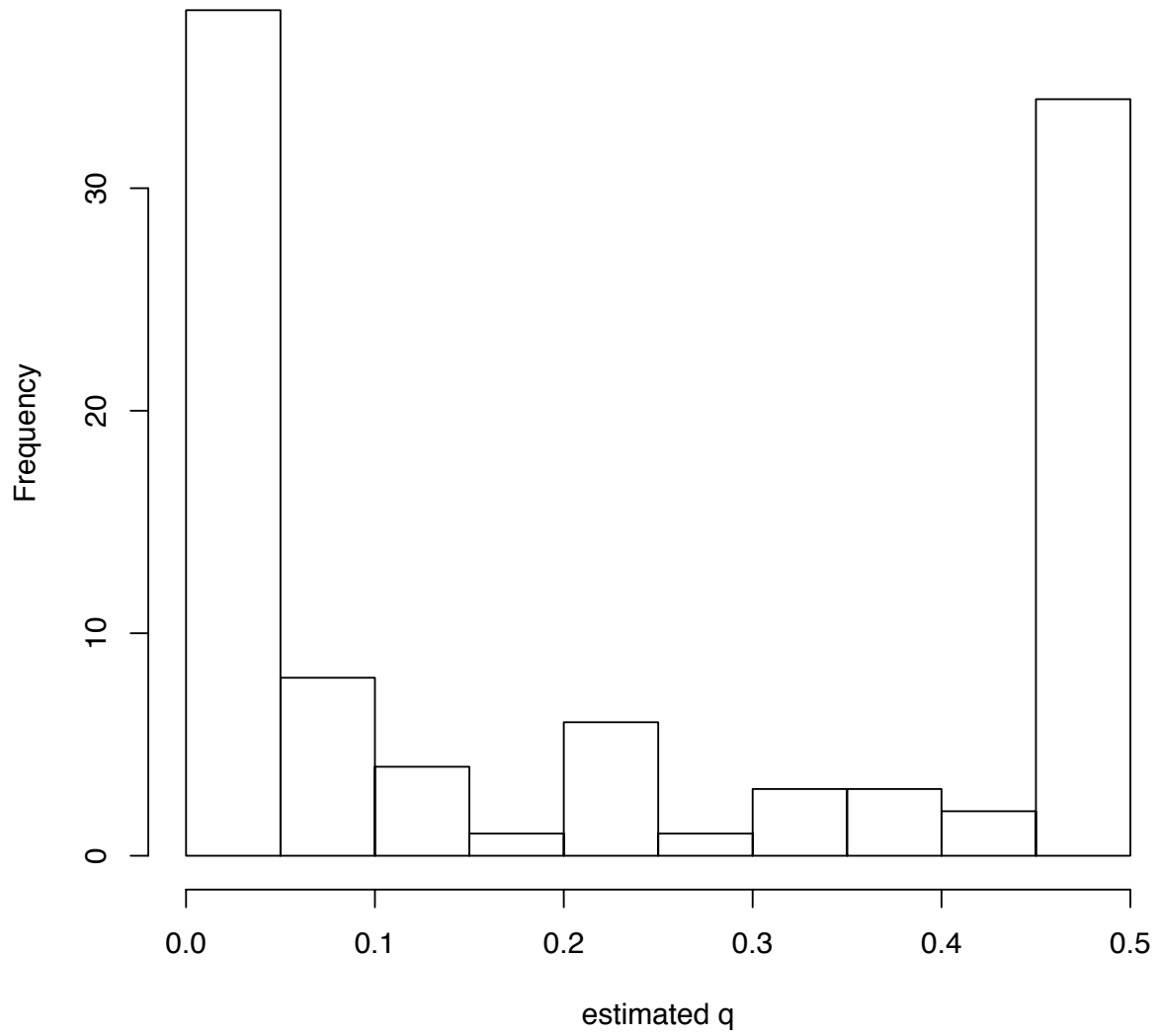
Marc Rehmsmeier

Figure S1 Histogram of q estimates

M. Rehmsmeier

Derivation of equations for seed-count differences in the seed fluorescence assay

Luo *et al.*'s equations (indicated here by superscript L) from Table 1 in (Luo *et al.* 2004) are

$$p_1^L = \alpha(1-r)^2$$

$$p_2^L = \alpha \frac{r^2}{3}$$

$$p_3^L = 2\alpha r(1-r)$$

$$p_4^L = 2\alpha \frac{r^2}{3}$$

$$p_5^L = \frac{2(1-\alpha)r(1-r)}{3}$$

$$p_6^L = \frac{2(1-\alpha)r^2}{9}$$

$$p_7^L = (1-\alpha)(1-r)^2$$

$$p_8^L = \frac{4(1-\alpha)r(1-r)}{3}$$

$$p_9^L = \frac{(1-\alpha)r^2}{9}$$

$$p_{10}^L = \frac{4(1-\alpha)r^2}{9}$$

$$p_{11}^L = \frac{2(1-\alpha)r^2}{9}$$

Probabilities for gamete marker classes (red-only, green-only, yellow/both, brown/none) can be defined by summing up the propabilities of gamete modes, weighted by the frequencies of the class in question among all genotypes of the respective mode (see also ref. 19). For each mode, the weight is the ratio of the number of instances of that gamete mode that are compatible with the class in question (red-only etc.) and the overall number of instances for that mode. The latter numbers are independent of actual genotypes and can for example be taken from Table 1 in ref. 14. The former numbers depend on the genotype. In the single-copy case (genotype RG/XY/XY/XY, with XY representing wild-type chromosomes), these numbers ("multipliers") are:

$$M_g^S = (0,3,3,6,0,3,0,6,0,6,6)$$

$$M_r^S = (0,3,0,3,3,6,0,6,0,6,6)$$

$$M_y^S = (1,0,3,0,3,0,3,6,3,6,0)$$

M. Rehmsmeier

$$M_b^S = (3,6,6,3,6,3,3,6,3,6,0)$$

which defines probabilities of color classes (the single-copy case indicated by the superscript S) as:

$$p_g^{LS} = \frac{0\,p_1^L}{4} + \frac{3\,p_2^L}{12} + \frac{3\,p_3^L}{12} + \frac{6\,p_4^L}{12} + \frac{0\,p_5^L}{12} + \frac{3\,p_6^L}{12} + \frac{0\,p_7^L}{6} + \frac{6\,p_8^L}{24} + \frac{0\,p_9^L}{6} + \frac{6\,p_{10}^L}{24} + \frac{6\,p_{11}^L}{12} + 0$$

$$p_r^{LS} = \frac{0\,p_1^L}{4} + \frac{3\,p_2^L}{12} + \frac{0\,p_3^L}{12} + \frac{3\,p_4^L}{12} + \frac{3\,p_5^L}{12} + \frac{6\,p_6^L}{12} + \frac{0\,p_7^L}{6} + \frac{6\,p_8^L}{24} + \frac{0\,p_9^L}{6} + \frac{6\,p_{10}^L}{24} + \frac{6\,p_{11}^L}{12} + 0$$

$$p_y^{LS} = \frac{1\,p_1^L}{4} + \frac{0\,p_2^L}{12} + \frac{3\,p_3^L}{12} + \frac{0\,p_4^L}{12} + \frac{3\,p_5^L}{12} + \frac{0\,p_6^L}{12} + \frac{3\,p_7^L}{6} + \frac{6\,p_8^L}{24} + \frac{3\,p_9^L}{6} + \frac{6\,p_{10}^L}{24} + \frac{0\,p_{11}^L}{12} + 0$$

$$p_b^{LS} = \frac{3\,p_1^L}{4} + \frac{6\,p_2^L}{12} + \frac{6\,p_3^L}{12} + \frac{3\,p_4^L}{12} + \frac{6\,p_5^L}{12} + \frac{3\,p_6^L}{12} + \frac{3\,p_7^L}{6} + \frac{6\,p_8^L}{24} + \frac{3\,p_9^L}{6} + \frac{6\,p_{10}^L}{24} + \frac{0\,p_{11}^L}{12} + 0$$

which simplify to

$$p_g^{LS} = -\frac{(2+\alpha)(-6+r)r}{36}$$

$$p_r^{LS} = \frac{r\,(6-2\,r+\alpha\,(-6+5\,r))}{12}$$

$$p_y^{LS} = \frac{\alpha\,(-3+6\,r-5\,r^2)+2\,(3-3\,r+r^2)}{12}$$

$$p_b^{LS} = \frac{(2+\alpha)(-3+r)^2}{36}$$

With $r = (3\,(-1+4\,\alpha+sqrt((-1+4\,\alpha)(-1+4\,\beta))))/(-4+16\,\alpha)$ from Eq. 1 in ref. 14 we obtain the following equations for seed-count differences (assuming a backcross with wildtype plants, such that seed fluorescence probabilities are the same as gamete probabilities):

$$p_g^{LS} - p_r^{LS} = \frac{\alpha-\beta}{4}$$

$$p_b^{LS} - p_y^{LS} = \frac{\alpha+\beta}{4}$$

which are the same as those derived for the model presented in this article (see main text).

In the double-copy case (genotype RG/RG/XY/XY) we have the following multipliers:

M. Rehmsmeier

$$M_g^D = (0,2,2,4,0,1,0,2,0,2,1)$$

$$M_r^D = (0,2,0,1,2,4,0,2,0,2,1)$$

$$M_y^D = (1,1,3,2,3,2,3,8,3,8,4)$$

$$M_b^D = (2,2,2,0,2,0,1,0,1,0,0)$$

leading to these gamete color frequencies (the superscript D indicates the double-copy case), specialized from the Luo *et al.* model (indicated again by the superscript L):

$$p_g^{LD} = -\frac{r\left(-14 + \alpha(-22 + r) + 5\,r\right)}{63}$$

$$p_r^{LD} = \frac{r\left(26 - 11\,r + \alpha(-26 + 23\,r)\right)}{63}$$

$$p_y^{LD} = \frac{189 - 82\,r + 42\,r^2 + \alpha(-105 + 130\,r - 114\,r^2)}{252}$$

$$p_b^{LD} = \frac{63 - 78\,r + 22\,r^2 + \alpha(105 - 114\,r + 26\,r^2)}{252}$$

With the above definition of r again, we obtain the following seed-count differences in the double-copy case, again assuming a backcross with wildtype plants, such that seed fluorescence probabilities are the same as gamete probabilities:

$$p_g^{LD} - p_r^{LD} = \frac{-1 + 10\,\alpha - 6\,\beta + sqrt\big((-1 + 4\,\alpha)(-1 + 4\,\beta)\big)}{28}$$

$$p_b^{LD} - p_y^{LD} = \frac{97 + 282\,\alpha^2 - 30\,\beta - 13\,sqrt\big((-1 + 4\,\alpha)(-1 + 4\,\beta)\big) + \alpha\left(-481 + 210\,\beta - 17\,sqrt\big((-1 + 4\,\alpha)(-1 + 4\,\beta)\big)\right)}{168\,(-1 + 4\,\alpha)}$$

These equations are different from equations derived from the model presented in this article which are:

$$p_g^{De} - p_r^{De} = \frac{\alpha - \beta}{3}$$

$$p_b^{De} - p_y^{De} = \frac{\alpha + \beta - 2}{3}$$

M. Rehmsmeier

where p_g^{De}, p_r^{De}, p_y^{De}, and p_b^{De} can be derived from p_1^e to p_{11}^e, the double-copy multiplicators M_g^D, M_r^D, M_y^D, and M_b^D, and the gamete mode frequencies (see derivation of p_g^{LS} etc. above), or directly by sorting gametes according to color in the computational procedure as:

$$p_r^{eD} = \frac{r}{3} - \frac{r^2}{6} + \beta\left(-\frac{1}{6} + \frac{5\,r}{12} + \frac{r'}{3\,r}\right)$$

$$+ \alpha\left(\frac{1}{6} - \frac{7\,r}{4} + \frac{r}{3\,q} + \frac{10\,r^2}{3} - \frac{3\,r^2}{2\,q} - \frac{5\,r^3}{3} + \frac{5\,r^3}{6\,q} + \frac{4\,r'}{3} - \frac{2\,r'}{3\,q} - \frac{r'}{3\,r} - \frac{4\,r\,r'}{3} + \frac{2\,r\,r'}{3\,q}\right)$$

$$p_g^{eD} = \frac{r}{3} - \frac{r^2}{6} + \beta\left(-\frac{1}{2} + \frac{5\,r}{12} + \frac{r'}{3\,r}\right)$$

$$+ \alpha\left(\frac{1}{2} - \frac{7\,r}{4} + \frac{r}{3\,q} + \frac{10\,r^2}{3} - \frac{3\,r^2}{2\,q} - \frac{5\,r^3}{3} + \frac{5\,r^3}{6\,q} + \frac{4\,r'}{3} - \frac{2\,r'}{3\,q} - \frac{r'}{3\,r} - \frac{4\,r\,r'}{3} + \frac{2\,r\,r'}{3\,q}\right)$$

$$p_y^{eD} = \frac{5}{6} - \frac{r}{3} + \frac{r^2}{6} + \beta\left(\frac{1}{6} - \frac{5\,r}{12} - \frac{r'}{3\,r}\right)$$

$$+ \alpha\left(-\frac{1}{2} + \frac{7\,r}{4} - \frac{r}{3\,q} - \frac{10\,r^2}{3} + \frac{3\,r^2}{2\,q} + \frac{5\,r^3}{3} - \frac{5\,r^3}{6\,q} - \frac{4\,r'}{3} + \frac{2\,r'}{3\,q} + \frac{r'}{3\,r} + \frac{4\,r\,r'}{3} - \frac{2\,r\,r'}{3\,q}\right)$$

$$p_b^{eD} = \frac{1}{6} - \frac{r}{3} + \frac{r^2}{6} + \beta\left(\frac{1}{2} - \frac{5\,r}{12} - \frac{r'}{3\,r}\right)$$

$$+ \alpha\left(-\frac{1}{6} + \frac{7\,r}{4} - \frac{r}{3\,q} - \frac{10\,r^2}{3} + \frac{3\,r^2}{2\,q} + \frac{5\,r^3}{3} - \frac{5\,r^3}{6\,q} - \frac{4\,r'}{3} + \frac{2\,r'}{3\,q} + \frac{r'}{3\,r} + \frac{4\,r\,r'}{3} - \frac{2\,r\,r'}{3\,q}\right)$$

For the triple-copy case (genotype GR/GR/GR/XY) we have the following multipliers:

$$M_g^T = (0,1,1,1,0,0,0,0,0,0,0)$$

$$M_r^T = (0,1,0,0,1,1,0,0,0,0,0)$$

$$M_y^T = (1,1,2,2,2,2,2,4,2,4,2)$$

$$M_b^T = (1,0,0,0,0,0,0,0,0,0,0)$$

With these, we obtain from Luo *et al.*'s model:

$$p_g^{LT} - p_r^{LT} = \frac{\alpha - \beta}{3}$$

$$p_b^{LT} - p_y^{LT} = \frac{34\,\alpha^2 - 8\,(-3 + \beta) + \alpha(-109 + 50\,\beta - 3\,sqrt((-1 + 4\,\alpha)\,(-1 + 4\,\beta)))}{-24 + 96\,\alpha}$$

M. Rehmsmeier

and from the model presented in this article:

$$p_g^{Te} - p_r^{Te} = \frac{\alpha - \beta}{4}$$

$$p_b^{Te} - p_y^{Te} = \frac{\alpha + \beta - 4}{4}$$

Probabilities of the eleven gamete modes and coefficients of double reduction when a second partner switch is

allowed

The superscript S indicates that a second partner switch is allowed. The occurrence of two simultaneous partner

switches is modeled such that they are not in the same region, that is they have to have at least one marker

between them. The eleven gamete modes:

$$p_1^S = \frac{p_{PC}q\,(4 - 8\,r + 4\,r^2)\tau}{16} + \frac{p_{DP}p_{PC}q\,(2\,r - 3\,r^2 - 4\,r' + 4\,r\,r')\tau}{16}$$

$$p_2^S = \frac{p_{PC}q\,r^2\tau}{4} + \frac{p_{DP}p_{PC}q\,(-2\,r - r^2 + 4\,r' - 4\,r\,r')\tau}{16}$$

$$p_3^S = -\frac{p_{PC}q\,r\,(-4 + 4\,r)\tau}{8} - \frac{p_{DP}p_{PC}q\,r\,(-3\,r + 4\,r')\tau}{8}$$

$$p_4^S = \frac{p_{DP}p_{PC}(-(q\,r^2) + 4\,q\,r\,r')\tau}{8}$$

$$p_5^S = \frac{p_{DP}p_{PC}r\,(-4 + 4\,q + 4\,r - 5\,q\,r)\tau}{8} + \frac{p_{PC}r\,(4 - 4\,q - 4\,r + 4\,q\,r)\tau}{8} + \frac{p_{DP}r\,(2 - 2\,r')\tau}{8}$$

$$p_6^S = -\frac{p_{DP}p_{PC}q\,r^2\tau}{8} + \frac{p_{DP}r\,r'\tau}{4}$$

$$p_7^S = \frac{16 - 32\,r + 16\,r^2}{16} + \frac{p_{PC}(-4\,q + 4\,r + 8\,q\,r - 4\,r^2 - 4\,q\,r^2)\tau}{16}$$
$$+ \frac{p_{DP}(10\,r - 13\,r^2 - 16\,r' + 14\,r\,r')\tau}{16}$$
$$+ \frac{p_{DP}p_{PC}(-4\,r + 4\,r^2 + 2\,q\,r^2 + 4\,q\,r' - 6\,q\,r\,r')\tau}{16}$$

$$p_8^S = \frac{(16 - 16\,r)r}{8} + \frac{p_{PC}r\,(-8 + 8\,r)\tau}{8} + \frac{p_{DP}r\,(-4 + 13\,r - 12\,r')\tau}{8}$$
$$+ \frac{p_{DP}p_{PC}r\,(8 - 6\,q - 8\,r + 3\,q\,r + 6\,q\,r')\tau}{8}$$

$$p_9^S = \frac{p_{PC}(4\,r + 4\,r^2 - 4\,q\,r^2)\tau}{16} + \frac{p_{DP}(-2\,r + 3\,r^2 + 8\,r' - 10\,r\,r')\tau}{16}$$
$$+ \frac{p_{DP}p_{PC}(-4\,r + 6\,q\,r - 4\,r^2 + 2\,q\,r^2 - 8\,q\,r' + 6\,q\,r\,r')\tau}{16}$$

$$p_{10}^S = \frac{-3\,p_{DP}(r^2 - 4\,r\,r')\tau}{8} - \frac{3\,p_{DP}p_{PC}(-(q\,r^2) + 2\,q\,r\,r')\tau}{8}$$

$$p_{11}^S = r^2 - \frac{p_{PC}r^2\tau}{2} + \frac{p_{DP}p_{PC}(-(q\,r) + 4\,r^2 - 2\,q\,r^2 + 2\,q\,r')\tau}{8} + \frac{p_{DP}(-2\,r - 5\,r^2 + 4\,r' - 2\,r\,r')\tau}{8}$$

Coefficients of double reduction:

$$\alpha^S = \frac{p_{PC}q\,\tau}{4}$$

$$\beta^S = \frac{\left(p_{DP}r \,+\, p_{PC}\big(2\,(-1 \,+\, p_{DP})(-1 \,+\, r)r \,-\, q\,(-1 \,+\, 2\,r)(1 \,+\, 2\,(-1 \,+\, p_{DP})r)\big)\right)\tau}{4}$$

Probabilities of the eleven gamete modes when a second partner switch is allowed, with p_{DP} and p_{PC} eliminated

Superscript S indicates that a second partner switch is allowed; superscript e indicates that p_{DP} and p_{PC} are eliminated.

$$p_1^{Se} = \frac{\alpha\left(\alpha\left(\begin{array}{c}(-2 + 4\,q)r^4 - 4\,q\,r' + r^2(-20 + 21\,q + 16\,r' - 32\,q\,r') + 2\,r^3(7 - 10\,q - 4\,r' + 8\,q\,r')\\ + r\,(8 - 6\,q - 8\,r' + 20\,q\,r')\end{array}\right) + q\,(\beta(-2\,r + 3\,r^2 + 4\,r' - 4\,r\,r') - (-1 + r)^2 r\,\tau)\right)}{r\,(8\,\alpha(1 - r + q\,(-1 + 2\,r)) - q\,\tau)}$$

p_2^{Se}

$$= \frac{\alpha\left(\alpha\left(-2\,(-1 + r)r\,(3\,r^2 + 4\,r' - 2\,r\,(1 + 2\,r')) + q\,(-1 + 2\,r)(6\,r^3 - 4\,r' + 2\,r\,(1 + 6\,r') - r^2(3 + 8\,r'))\right) + q\,(\beta(2\,r + r^2 - 4\,r' + 4\,r\,r') - r^3\tau)\right)}{r\,(8\,\alpha(1 - r + q\,(-1 + 2\,r)) - q\,\tau)}$$

$$p_3^{Se} = \frac{-2\,\alpha\left(\alpha\left((-2 + 4\,q)r^3 + 4\,q\,r' + r\,(-8 + 5\,q + 8\,r' - 16\,q\,r') + 2\,r^2(5 - 6\,q - 4\,r' + 8\,q\,r')\right) + q\,(\beta(3\,r - 4\,r') - (-1 + r)r\,\tau)\right)}{8\,\alpha(1 - r + q\,(-1 + 2\,r)) - q\,\tau}$$

$$p_4^{Se} = \frac{-2\,\alpha(-(\beta q) + \alpha(q\,(1 - 2\,r)^2 - 2\,(-1 + r)r))(r - 4\,r')}{8\,\alpha(1 - r + q\,(-1 + 2\,r)) - q\,\tau}$$

p_5^{Se}

$$= \frac{\alpha^2\left(q\,(8 - 26\,r + 24\,r^2 - 8\,r^3) + 4\,(-2 + 6\,r - 5\,r^2 + r^3)\right) + \beta q\,(-1 + r')\tau + \alpha\left(2\,\beta(4 - 4\,r + q\,(-4 + 5\,r)) + \left(2\,(-1 + r)r\,r' - q\,(-1 + r\,(2 - 4\,r') + r' + r^2(-2 + 4\,r'))\right)\tau\right)}{8\,\alpha(1 - r + q\,(-1 + 2\,r)) - q\,\tau}$$

$$p_6^{Se} = \frac{(\beta q - \alpha(q\,(1 - 2\,r)^2 - 2\,(-1 + r)r))(2\,\alpha r - r'\tau)}{8\,\alpha(1 - r + q\,(-1 + 2\,r)) - q\,\tau}$$

M. Rehmsmeier

p_7^{Se}

$$= \frac{\begin{array}{c} -8\,\alpha^2\left(\begin{array}{c}(-2+4\,q)r^4-2\,q\,r'+r^2(-18+15\,q+10\,r'-20\,q\,r')+2\,r^3\left(7-3\,r'+q\,(-8+6\,r')\right)\\+r\left(6-4\,r'+q\,(-4+11\,r')\right)\end{array}\right)\\+q\left((8+13\,\beta)r^2-4\,r^3+16\,\beta r'-2\,r\left(2+\beta(5+7\,r')\right)\right)\tau\\+\alpha\left(\begin{array}{c}-16\,q\,r'(\beta+\tau)-2\,(-1+2\,q)r^4(-16+13\,\tau)+2\,r\left(16+4\,\beta(2+3\,q\,r')-16\,r'\tau+q\,(-16+7\,\tau+39\,r'\tau)\right)\\-r^2(96+8\,\beta(2+q)-16\,\tau-60\,r'\tau+q\,(-128+61\,\tau+120\,r'\tau))+2\,r^3\left(\begin{array}{c}48-7\,(3+2\,r')\tau\\+4\,q\,(-20+(12+7\,r')\tau)\end{array}\right)\end{array}\right)\end{array}}{4\,r\,(8\,\alpha(1-r+q\,(-1+2\,r))-q\,\tau)}$$

$$p_8^{Se}=\frac{\begin{array}{c}4\,\alpha^2\left(3\,q\,(1-2\,r)^2(-2+r+2\,r')-2\,(-1+r)\left(4+3\,r^2+2\,r\,(-7+3\,r')\right)\right)\\+q\left(4\,(-1+r)r+\beta(4-13\,r+12\,r')\right)\tau\\-\alpha\left(\begin{array}{c}4\,\beta\left(8-8\,r+3\,q\,(-2+r+2\,r')\right)+4\,q\,(1+3\,r')\tau-2\,(-1+2\,q)r^3(-16+13\,\tau)\\+r\left(q\,(32-29\,\tau-48\,r'\tau)+8\,(-4+3\,r'\tau)\right)+r^2\left(64-26\,\tau-24\,r'\tau+q\,(-96+68\,\tau+48\,r'\tau)\right)\end{array}\right)\end{array}}{2\,(8\,\alpha(1-r+q\,(-1+2\,r))-q\,\tau)}$$

$$p_9^{Se}=\frac{\begin{array}{c}-8\,\alpha^2\left(\begin{array}{c}(-2+4\,q)r^4+4\,q\,r'+r^2(-8+11\,q-14\,r'+28\,q\,r')-2\,r^3\left(-2-3\,r'+6\,q\,(1+r')\right)\\+r\left(2+8\,r'-q\,(3+19\,r')\right)\end{array}\right)\\+\beta q\left(-3\,r^2-8\,r'+2\,r\,(1+5\,r')\right)\tau\\+\alpha\left(\begin{array}{c}-8\,\beta\left((-2+q)r^2-4\,q\,r'+r\,(-2+3\,q\,(1+r'))\right)\\+\left(\begin{array}{c}2\,r\left(-3\,r^3+8\,r'-2\,r\,(2+9\,r')+r^2(3+10\,r')\right)\\+q\left(12\,r^4+8\,r'-8\,r^3(2+5\,r')-2\,r\,(1+21\,r')+r^2(11+72\,r')\right)\end{array}\right)\tau\end{array}\right)\end{array}}{4\,r\,(8\,\alpha(1-r+q\,(-1+2\,r))-q\,\tau)}$$

$$p_{10}^{Se}=\frac{3\left(-(\beta q)+\alpha(q\,(1-2\,r)^2-2\,(-1+r)r)\right)(4\,\alpha r-8\,\alpha r'-r\,\tau+4\,r'\tau)}{2\,(8\,\alpha(1-r+q\,(-1+2\,r))-q\,\tau)}$$

p_{11}^{Se}

$$=\frac{\begin{array}{c}-4\,\alpha^2\left(q\,(1-2\,r)^2(r+2\,r^2-2\,r')-2\,r\,(r-5\,r^2+2\,r^3+2\,r'-2\,r\,r')\right)\\+q\left(-2\,r^3+\beta(5\,r^2-4\,r'+2\,r\,(1+r'))\right)\tau\\+\alpha\left(\begin{array}{c}4\,\beta\left(-4\,r^2+q\,(r+2\,r^2-2\,r')\right)+4\,q\,r'\tau-2\,r\,(q-4\,r'+9\,q\,r')\tau\\+r^2(-4+3\,q-12\,r'+24\,q\,r')\tau-2\,(-1+2\,q)r^4(-8+5\,\tau)-2\,r^3\left(-8+\tau-2\,r'\tau+q\,(8-6\,\tau+4\,r'\tau)\right)\end{array}\right)\end{array}}{2\,r\,(8\,\alpha(1-r+q\,(-1+2\,r))-q\,\tau)}$$

M. Rehmsmeier