

Genotypic Frequencies at Equilibrium for Polysomic Inheritance Under Double-Reduction

Kang Huang,* Tongcheng Wang,* Derek W. Dunn,* Pei Zhang,* Xiaoxiao Cao,* Rucong Liu,* and Baoguo Li*^{†,1}

*Shaanxi Key Laboratory for Animal Conservation, College of Life Sciences, Northwest University, Xi'an, China, 710069

and [†]Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, China, 650223

ORCID IDs: 0000-0002-8357-117X (K.H.); 0000-0001-5909-1224 (D.W.D.); 0000-0003-4433-382X (P.Z.); 0000-0001-7430-3889 (B.L.)

ABSTRACT Polyploids are organisms whose genomes consist of more than two complete sets of chromosomes. Both autopolyploids and allopolyploids may display polysomic inheritance. A peculiarity of polysomic inheritance is multivalent formation during meiosis resulting in double-reduction, which occurs when sister chromatid fragments segregate into the same gamete. Double-reduction can result in gametes carrying identical-by-descent alleles and slightly increasing homozygosity. This will cause the genotypic frequencies to deviate from expected values and will thus bias the results of standard population genetic analytical methods used in molecular ecology and selective breeding. In this study, we extend existing double-reduction models to account for any even level of ploidy, and derive the symbolic expressions for genotypic frequencies via two methods. Inbreeding coefficients and heterozygosity under double-reduction and inbreeding are also calculated. Numerical solutions obtained by computer simulations are compared with analytical solutions predicted by the model to validate the model.

KEYWORDS

Polysomic inheritance
double-reduction
genotypic frequency
inbreeding coefficient
heterozygosity

Polyploids are organisms whose genomes consist of more than two complete sets of chromosomes (Madlung 2013). They represent a significant proportion of plant species, with 30–80% of angiosperm species showing at least some polyploidy (Burow *et al.* 2001) and most lineages showing evidence of paleopolyploidy (Otto 2007). Polyploid plants can arise spontaneously in nature by several mechanisms, including meiotic or mitotic failures, and fusion of unreduced gametes (Comai 2005).

There are two distinct mechanisms of genome duplication that result in polyploidy: allopolyploidy and autopolyploidy. Autopolyploids are thought to usually arise within a species by the doubling of structurally similar homologous genomes, whereas allopolyploids arise via interspecific hybridization and subsequent doubling of non-homologous genomes (Parisod *et al.* 2010). Both autopolyploids and allopolyploids

can be found among both wild and domesticated plant species. Although rare, polyploidy is also found in a few species of vertebrates such as some salmonid fish (Limborg *et al.* 2017), the weather loach (*Misgurnus anguillicaudatus*) (Zhou *et al.* 2016), the common carp (*Cyprinus carpio*) (David *et al.* 2003), and the African clawed frog (*Xenopus laevis*) (Session *et al.* 2016).

In autopolyploids, more than two homologous chromosomes can pair at meiosis, resulting in the formation of multivalents and polysomic inheritance (Rieger *et al.* 1968). A peculiarity of polysomic inheritance is the possibility that a gamete inherits a single gene copy twice, termed double-reduction (Butruille and Boiteux 2000). For example, an autotetraploid individual *ABCD* produces a gamete *AA*. In prophase I, crossovers can happen between the locus and the centromere, resulting in an exchange of chromatid fragments between pairing chromosomes. In a multivalent configuration, the separation of chromosomes can be either disjunctive or nondisjunctive (De Silva *et al.* 2005). For nondisjunctive separation, chromosomes that were previously paired in Prophase I segregate into a single secondary oocyte. Double-reduction occurs when fragments of sister chromatids further segregate into a single gamete.

Double-reduction arises from a combination of three major events during meiosis: crossing-over between non-sister chromatids, an appropriate pattern of disjunction, and the subsequent migration of the chromosomal segments carrying a pair of sister chromatids to the same gamete (Darlington 1929; Haldane 1930). Multivalent chromosome

Copyright © 2019 Huang *et al.*

doi: <https://doi.org/10.1534/g3.119.400132>

Manuscript received January 26, 2019; accepted for publication March 20, 2019; published Early Online March 25, 2019.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Supplemental material available at Figshare: <https://doi.org/10.25387/g3.7883150>.

¹Corresponding author: E-mail: baoguo@nwu.edu.cn.

pairing during meiosis can also occur in allopolyploids, resulting in a mixed inheritance pattern across loci in the genome, termed segmental allopolyploidy (Stebbins 1950).

Geneticists have developed several mathematical models to simulate double-reduction. For instance, for tetrasomic inheritance, the rate of double-reduction α is assumed to have a minimum value of 0 under *random chromosome segregation* (RCS, Muller 1914). This increases to 1/7 with *pure random chromatid segregation* (PRCS, Haldane 1930), and to 1/6 with *complete equational segregation* (CES, Mather 1935). A diagram of these models can be found in Figure 1.

Double-reduction results in gametes carrying *identical-by-descent* (IBD) alleles and slightly increased homozygosity (Hardy 2016). Experiments aimed at estimating the frequency of double-reduction in autotetraploids have yielded values ranging from 0 to almost 0.30 (Wu *et al.* 2001). It has a similar effect as inbreeding, and can cause a deviation in genotypic frequencies from the Hardy-Weinberg equilibrium (Luo *et al.* 2006). Fisher (1943) first studied the Mid gene of *Lythrum salicaria*, and produced an analytical method to estimate genotypic frequencies for tetrasomic inheritance for a biallelic locus. Geiringer (1949) extended this work to a multiallelic locus for tetrasomic inheritance. However, the genotypic frequencies for polysomic inheritance for higher levels of ploidy and for a multiallelic locus have been neglected.

Genotypic frequencies are important in population genetics, molecular ecology and molecular breeding, with the *Hardy-Weinberg equilibrium*

(HWE) being the basal standard for analytical methods in disomic inheritance. Several methods assume that loci conform to the HWE for the calculations of certain model parameters or statistics, for example, the likelihood of a genotype (population assignment, Bayesian clustering, parentage analysis), the statistics measuring the deviation from expectations (linkage disequilibrium test, differentiation test, heterozygote deficiency/excess test, bottle-neck effect test), and other parameters (*e.g.*, individual inbreeding coefficient, relatedness coefficient, kinship coefficient, QTL mapping). Although some analytical methods have been developed for organisms with polysomic inheritance (*e.g.*, Clark and Jasieniuk 2011; Hardy and Vekemans 2002; Huang *et al.* 2015; Meirmans and van Tienderen 2004; Pritchard *et al.* 2000), none of these methods accounts for double-reduction, and thus in the presence of double-reduction will produce misleading results. For example, in tetrasomic inheritance, $ABCD \times EFGH$ may produce an offspring $AAEE$, and both parents are excluded as the true parents in parentage analysis.

Due to the genotypic ambiguity of polyploid species, the correct genotype cannot be obtained because some genotypes share a common electrophoretic band type in PCR-based co-dominant markers (*e.g.*, microsatellites) and the allelic dosage is unknown (Hardy 2016). For example, in autotetraploids, the three genotypes $AAAB$, $AABB$ and $ABBB$ have a same electrophoretic band type of AB . We define the set of alleles within an individual as its phenotype.

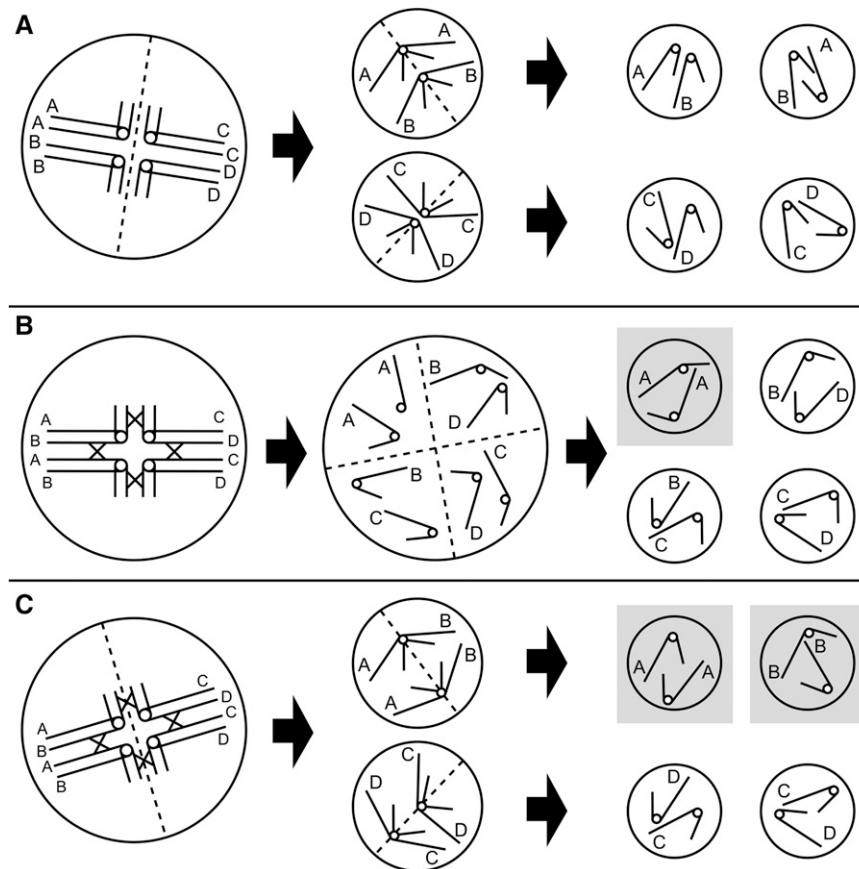


Figure 1 Diagrams of double-reduction models under tetrasomic inheritance. The leftmost column shows the primary oocytes, the middle column shows the secondary oocytes (A and C) or tetrad (B) and the rightmost column shows the gametes. The gametes with a gray background carry identical-by-double-reduction alleles. Dashed lines denote cellular fission, solid lines denote the arms of the chromosome, and the circles connecting solid lines denote the centromere. The target locus is located in the long arm of the chromosome and the identical-by-descent alleles are denoted by the same letter. (A) Random chromosome segregation (RCS) ignores the crossover between the target locus and the centromere (Muller 1914). In the absence of crossing over, gametes may originate from any combination of homologous chromosomes, and two sister chromatids never sort into the same gamete (Parisod *et al.* 2010); (B) Pure random chromatid segregation (PRCS) accounts for the crossing over between the target locus and the centromere, and assumes the chromatids behave independently and randomly segregate into gametes (Haldane 1930). When sister chromatids segregate into the same gamete, double-reduction occurs. The probability that the two chromatids in a gamete are sister chromatids is 4 (the number of sister chromatid pairs) divided by $\binom{8}{2}$ (the number of ways to sample two chromatids from eight chromatids), which is equal to 1/7. (C) Complete equational segregation (CES), homologous chromosomes

pair and chromatids exchange via recombination (Mather 1935). The whole arms of sister chromatids are exchanged into different chromosomes. The probability that two homologous chromosomes within a single secondary oocyte were previously paired at a target locus in prophase I is 1/3. In this case, these sister chromatid fragments will segregate in a single gamete at a ratio of 1/2, so the rate of double-reduction is thus 1/6 for tetrasomic inheritance.

Here, we extend an existing double-reduction model to incorporate the recombination rate into CES, and show two methods to derive the symbolic expressions for genotypic frequencies at equilibrium. To enable other researchers make similar calculations, we make available a C++ source code that enables the calculation of genotypic/phenotypic frequencies of gametes/zygotes at equilibrium to a maximum ploidy level of ten.

DOUBLE-REDUCTION MODELS

In this paper, we assume that the population size is sufficiently large and there are no differences in selection between both zygotes and gametes. All zygotes will thus have an equal opportunity to mature and reproduce, and all gametes will have an equal opportunity to become fertilized. Therefore, the zygote frequencies are equal to the genotypic frequencies of reproducing individuals, and as a result, we use the zygote frequencies to indicate the genotypic frequencies of reproducing individuals. The double-reduction rates (*i.e.*, alpha) are assumed to be equal among reproducing individuals (*e.g.*, in females and males). Mutation and migration are not included in our model, whereas inbreeding is included for analyzing the influence of inbreeding and double-reduction on the inbreeding coefficient and for estimating heterozygosity. Organisms with odd levels of ploidy are not considered because of their inherent inability to produce euploid gametes — these organisms are sterile.

Transitional probability from zygotes to gametes

The transitional probability from a zygote to a gamete is the probability that a zygote produces a gamete of a certain genotype, which can be used to find the double-reduction rates of PRCS and to derive the analytical expression of gamete frequencies in the first step of the non-linear method (see Non-linear method section). Such a probability can also be applied to some population genetics studies such as parentage analysis, and the calculation of segregation ratios in mapping populations (see Discussion).

If the rate of double-reduction is greater than zero, the gametes will carry *identical-by-double-reduction* (IBDR) alleles. For tetrasomic and hexasomic inheritance, there are only two and three allele copies within a gamete, respectively. Hence, it can be found at most one pair of IBDR alleles within a gamete, and a single parameter can be used to measure the rate of double-reduction.

In polysomic inheritance with a higher level of ploidy, there may be more than one pair of IBDR alleles in a gamete. Therefore, it is necessary to add some additional parameters in the segregation ratios. The transitional probability from zygotes to gametes for tetrasomic and hexasomic inheritances at a biallelic locus (the *Mid* gene in *Lythrum salicaria*) has been studied previously by Fisher and Mather (1943). Here, we extend the work of Fisher and Mather and use a general mathematical expression of this transitional probability. There are at most $\lfloor v/4 \rfloor$ pairs of IBDR alleles within a gamete, where v is the ploidy level of the resulting zygote. Let G and g be the genotypes of a zygote and a gamete, respectively, and let α_i be the probability that the gamete carries i pairs of IBDR alleles, then $\sum_{i=0}^{\lfloor v/4 \rfloor} \alpha_i = 1$. Suppose that h is the number of alleles at target locus, then the probability that G produces g is the following weighted sum:

$$T(g|G) = \sum_{i=0}^{\lfloor v/4 \rfloor} \sum_{j_1+j_2+\dots+j_h=i} \frac{\prod_{k=1}^h \delta_{jk} \binom{n_k}{j_k} \binom{n_k - j_k}{m_k - 2j_k} \alpha_i}{\binom{v}{i} \binom{v-i}{v/2-2i}} \quad (1)$$

Where n_k and m_k are the numbers of copies of the k^{th} allele (say A_k) in G and g , respectively, and j_k is the number of IBDR allele pairs

$A_k - A_k$ in g , $k = 1, 2, \dots, h$, then $\sum_{k=1}^h n_k = v$, $\sum_{k=1}^h m_k = v/2$ and $\sum_{k=1}^h j_k = i$.

The symbol δ_{jk} in Equation (1) is a binary variable, which is used to limit the range of j_k , so as to avoid some invalid situations. Note that the combination $\binom{n_k}{j_k}$ implies $0 \leq j_k \leq n_k$, and $\binom{n_k - j_k}{m_k - 2j_k}$ implies $0 \leq m_k - 2j_k \leq n_k - j_k$, or equivalently it implies $j_k \leq m_k/2$ and $m_k - n_k \leq j_k$. Hence $\delta_{jk} = 1$ if $\max(0, m_k - n_k) \leq j_k \leq \min(n_k, m_k/2)$, otherwise $\delta_{jk} = 0$.

In Equation (1), for every i ($i = 0, 1, 2, \dots, \lfloor v/4 \rfloor$), two events need consideration. First, is the event that g carries i pairs of IBDR alleles. Second, is the event that those i pairs of IBDR alleles are distributed into h kinds of alleles. Because both types of event are mutually exclusive, the products of the probabilities of the two events stated above are summed to obtain $T(g|G)$.

Each chromosome is duplicated during meiosis, so each allele in G becomes a pair of duplicated alleles. For the k^{th} kind of allele, first we sample j_k pairs from the n_k pairs of duplicated alleles $A_k - A_k$, then there are $\binom{n_k}{j_k}$ ways (where $0 \leq j_k \leq n_k$), and so the first combination in the numerator of Equation (1) is obtained. We then sample the non-IBDR alleles to form g . The remaining is $n_k - j_k$ pairs of duplicated alleles $A_k - A_k$, which still requires $m_k - 2j_k$ copies of A_k . Then we sample $m_k - 2j_k$ pairs of duplicated alleles $A_k - A_k$, and so there are $\binom{n_k - j_k}{m_k - 2j_k}$ ways (where $0 \leq m_k - 2j_k \leq n_k - j_k$), which obtains the second combination in the numerator of Equation (1). Moreover, if we sample one copy of A_k in each pair of the $m_k - 2j_k$ pairs of duplicated alleles, then there are altogether $2^{m_k - 2j_k}$ possible combinations. Because each sampling combination is independent for different kinds of alleles, the total number of combinations is the product

$$\prod_{k=1}^h 2^{m_k - 2j_k} \delta_{jk} \binom{n_k}{j_k} \binom{n_k - j_k}{m_k - 2j_k} = 2^{v/2 - 2i} \prod_{k=1}^h \delta_{jk} \binom{n_k}{j_k} \binom{n_k - j_k}{m_k - 2j_k}.$$

In the following example, we consider the total number of combinations to produce a gamete consisting of i pairs of IBDR alleles without accounting for the specific genotypes G and g . Similar to the above process, we first sample i pairs from the v pairs of duplicated alleles, and these alleles will become the IBDR alleles in g . We then sample $v/2 - 2i$ pairs from the remaining $v - i$ pairs, and then one allele in each pair of the $v/2 - 2i$ pairs is further sampled to form g . The total number of allele combinations is $2^{v/2 - 2i} \binom{v}{i} \binom{v-i}{v/2-2i}$. Here the coefficient $2^{v/2 - 2i}$ is eliminated in the fraction in Equation (1).

Equation (1) is a general expression of the transitional probability, and can be applied at multiallelic loci and at any even level of ploidy. As an example, the transitional probability from zygotes to gametes in octosomic inheritance at a biallelic locus is shown in Table 1. Another example is shown in Appendix A, which is the derivation of $T(BBBB | ABBBBBB) = (4\alpha_0 + 5\alpha_1 + 6\alpha_2)/8$.

Random chromosome segregation

The random chromosome segregation model ignores the crossover between the target locus and the centromere, and the rate of double-reduction (α) is equal to zero (Muller 1914). In the absence of crossing over, gametes may originate from any combination of homologous chromosomes, and two sister chromatids never sort into the same gamete (Parisod *et al.* 2010, Figure 1A). Therefore, genotypic frequencies concur with the HWE. Here, the HWE is expanded to account for polysomic inheritance, in which the alleles in a genotype are independent and randomly appear according to their frequencies. Thus, the genotypic frequency under the RCS can be expressed as

■ Table 1 Transitional probability from zygotes to gametes in octosomic inheritance at a biallelic locus

Zygote	Gamete					Divisor
	AAAA	AAAB	AABB	ABBB	BBBB	
AAAAAAA	1	—	—	—	—	1
AAAAAAB	$4\alpha_0 + 5\alpha_1 + 6\alpha_2$	$4\alpha_0 + 2\alpha_1$	$\alpha_1 + 2\alpha_2$	—	—	8
AAAAABB	$6\alpha_0 + 10\alpha_1 + 15\alpha_2$	$16\alpha_0 + 10\alpha_1$	$6\alpha_0 + 6\alpha_1 + 12\alpha_2$	$2\alpha_1$	α_2	28
AAAABBB	$4\alpha_0 + 10\alpha_1 + 20\alpha_2$	$24\alpha_0 + 20\alpha_1$	$24\alpha_0 + 15\alpha_1 + 30\alpha_2$	$4\alpha_0 + 10\alpha_1$	$\alpha_1 + 6\alpha_2$	56
AAAABBB	$\alpha_0 + 5\alpha_1 + 15\alpha_2$	$16\alpha_0 + 20\alpha_1$	$36\alpha_0 + 20\alpha_1 + 40\alpha_2$	$16\alpha_0 + 20\alpha_1$	$\alpha_0 + 5\alpha_1 + 15\alpha_2$	70
AAABBBB	$\alpha_1 + 6\alpha_2$	$4\alpha_0 + 10\alpha_1$	$24\alpha_0 + 15\alpha_1 + 30\alpha_2$	$24\alpha_0 + 20\alpha_1$	$4\alpha_0 + 10\alpha_1 + 20\alpha_2$	56
AABBBBB	α_2	$2\alpha_1$	$6\alpha_0 + 6\alpha_1 + 12\alpha_2$	$16\alpha_0 + 10\alpha_1$	$6\alpha_0 + 10\alpha_1 + 15\alpha_2$	28
ABBBBBB	—	—	$\alpha_1 + 2\alpha_2$	$4\alpha_0 + 2\alpha_1$	$4\alpha_0 + 5\alpha_1 + 6\alpha_2$	8
BBBBBBB	—	—	—	—	1	1

$$\Pr(G|RCS) = \frac{v!}{n_1!n_2!\dots n_h!} \prod_{k=1}^h p_k^{n_k}$$

Where p_k denotes the frequency of A_k . For example, the frequency of a tetraploid genotype $AABC$ is $12p_A^2p_Bp_C$.

When the conditions of the RCS are met, $\alpha_i = 0$ whenever $i > 0$, then Equation (1) can be simplified to

$$T(g|G, RCS) = \left[\prod_{k=1}^h \binom{n_k}{m_k} \right] / \binom{v}{v/2} \quad (2)$$

Pure random chromatid segregation

Pure random chromatid segregation accounts for crossing over and assumes the chromatids behave independently in meiotic anaphases, and randomly segregate into gametes (Haldane, 1930, Figure 1B). Then the probability that a zygote G produces a gamete g at a multiallelic locus can be derived by sampling $v/2$ chromatids

(or allele copies) in a total number of $2v$ chromatids (or allele copies). Because the number of chromatids is twice the number of chromosomes, the transitional probability $T(g|G, PRCS)$ can be modified from Equation (2) by duplicating the terms n_k and v in those combinations as follows:

$$T(g|G, PRCS) = \left[\prod_{k=1}^h \binom{2n_k}{m_k} \right] / \binom{2v}{v/2} \quad (3)$$

The value of alpha of PRCS can be derived by simulating the meiosis, first by sampling $v/2 - i$ IBDR allele pairs out of v IBDR allele pairs, and then by segregating i IBDR allele pairs and $v/2 - 2i$ non-IBDR alleles into the gamete. Therefore, the value of α_i is given by

$$\alpha_i = \binom{v}{v/2 - i} \binom{v/2 - i}{i} 2^{v/2 - 2i} / \binom{2v}{v/2}$$

The values of alpha under PRCS, ranging from tetrasomic to dodecasomic inheritance, are shown in Table 2. The expected number of IBDR allele pairs in a gamete $\lambda = \sum_i \alpha_i = v(v - 2)/(16v - 8)$.

■ Table 2 The probabilities of a gamete carrying i pairs of identical-by-double-reduction alleles (α_i) and the inbreeding coefficient in an outcrossed population with pure random chromatid segregation (PRCS), complete equational segregation (CES) or partial equational segregation (PES)

Model	Inheritance	α_1	α_2	α_3	F	f		
PRCS	Tetrasomic	$\frac{1}{7}$			$\frac{1}{15}$	0.0667	$\frac{1}{5}$	0.2000
	Hexasomic	$\frac{3}{11}$			$\frac{2}{35}$	0.0571	$\frac{1}{7}$	0.1429
	Octosomic	$\frac{24}{65}$	$\frac{1}{65}$		$\frac{1}{21}$	0.0476	$\frac{1}{9}$	0.1111
	Decasomic	$\frac{140}{323}$	$\frac{15}{323}$		$\frac{4}{99}$	0.0404	$\frac{1}{11}$	0.0909
	Dodecasomic	$\frac{1440}{3059}$	$\frac{270}{3059}$	$\frac{5}{3059}$	$\frac{5}{143}$	0.0350	$\frac{1}{13}$	0.0769
CES	Tetrasomic	$\frac{1}{6}$			$\frac{1}{13}$	0.0769	$\frac{3}{13}$	0.2308
	Hexasomic	$\frac{3}{10}$			$\frac{1}{16}$	0.0625	$\frac{5}{32}$	0.1563
	Octosomic	$\frac{27}{70}$	$\frac{3}{140}$		$\frac{3}{59}$	0.0508	$\frac{7}{59}$	0.1186
	Decasomic	$\frac{55}{126}$	$\frac{5}{84}$		$\frac{2}{47}$	0.0426	$\frac{9}{94}$	0.0957
	Dodecasomic	$\frac{285}{616}$	$\frac{65}{616}$	$\frac{5}{1848}$	$\frac{5}{137}$	0.0365	$\frac{11}{137}$	0.0803
PES	Tetrasomic	$\frac{r_s}{6}$			$\frac{r_s}{12+r_s}$	0.0400	$\frac{3r_s}{12+r_s}$	0.1200
	Hexasomic	$\frac{3r_s}{10}$			$\frac{r_s}{15+r_s}$	0.0323	$\frac{5r_s}{30+2r_s}$	0.0806
	Octosomic	$\frac{3r_s(10-r_s)}{70}$	$\frac{3r_s^2}{140}$		$\frac{3r_s}{56+3r_s}$	0.0261	$\frac{7r_s}{56+3r_s}$	0.0609
	Decasomic	$\frac{5r_s(14-3r_s)}{126}$	$\frac{5r_s^2}{84}$		$\frac{2r_s}{45+2r_s}$	0.0217	$\frac{9r_s}{90+4r_s}$	0.0489
	Dodecasomic	$\frac{5r_s(84-28r_s+r_s^2)}{616}$	$\frac{5r_s^2(14-r_s)}{616}$	$\frac{5r_s^3}{1848}$	$\frac{5r_s}{132+5r_s}$	0.0186	$\frac{11r_s}{132+5r_s}$	0.0409

Where r_s denotes the single chromatid recombination rate between target locus and the centromere, F and f denote the inbreeding coefficients in zygotes and gametes, respectively. The numerical solutions of F and f are given in the rightmost column. For PES, these numerical solutions are calculated with $r_s = 0.5$.

Complete equational segregation

The *complete equational segregation* (CES) assumes that the whole arms of the two pairing chromatids are exchanged between pairing chromosomes (Mather 1935). In Figure 1C, the leftmost two chromosomes in the primary oocyte are paired in the long arms, as well as the rightmost two chromosomes. In Metaphase I, the chromosomes randomly segregate into the secondary oocytes. If the pairing chromosomes segregate in the same secondary oocyte, then the duplicated alleles may further segregate into a single gamete.

The production of double-reduction gametes requires the fulfillment of two conditions: (i) the chromosomes paired at target locus segregate into the same secondary oocyte, and (ii) the duplicated alleles segregate into the same gamete. In order to construct a mathematical model of CES, we simulate this meiosis, and the values of alpha will be derived by using a two-step process.

First, we model the probability, μ_j , that j pairs of chromosomes paired at the target locus (PCPs) segregate into a secondary oocyte.

There are $v/2$ PCPs in the primary oocyte, so we have $\binom{v/2}{j}$ combinations to sample j pairs. Next, for the remaining $v/2 - j$ PCPs, it still requires $v/2 - 2j$ chromosomes that are not paired with each other at target locus during prophase I to form the secondary oocyte. We therefore (i) sample $v/2 - 2j$ PCPs from the remaining $v/2 - j$ PCPs (there are $\binom{v/2 - j}{v/2 - 2j}$ combinations), and (ii) sample one chromosome from each PCP in the previously sampled $v/2 - 2j$ PCPs (there are $2^{v/2 - 2j}$ chromosome combinations). The product of the above three numbers of combinations is divided by the total number of segregation modes in Metaphase I, i.e., $\binom{v}{v/2}$, to derive μ_j . The process described above can be summarized by the following expression:

$$\mu_j = 2^{v/2 - 2j} \binom{v/2}{j} \binom{v/2 - j}{v/2 - 2j} / \binom{v}{v/2} \quad (0 \leq j \leq \lfloor v/4 \rfloor). \quad (4)$$

Second, we model the probability, α_i , that i pairs of duplicated alleles segregate into a single gamete. There are j PCPs in the second oocyte, and each has four alleles (e.g., the alleles in both chromosomes are A and B in the top second oocyte of Figure 1C). There are four segregation modes for each PCP in Anaphase II, where two of them (AA and BB) produce gametes with IBDR alleles and the other two (AB and BA) pass non-IBDR alleles into the gamete. In order to produce a gamete consisting of i pairs of IBDR alleles, i PCPs are sampled and segregate IBDR alleles. Following the description above, we have $\binom{j}{i} 2^i$ combinations of sampling i PCPs that pass IBDR alleles into a gamete from j PCPs.

Next, the remaining $j - i$ PCPs pass non-IBDR alleles, and so there are 2^{j-i} combinations. In each of the remaining $v/2 - 2j$ chromosomes that are not paired with each other during prophase I, one of the two alleles segregates into the gamete, and hence there are $2^{v/2 - 2j}$ allele combinations. The product of these numbers of combinations is divided by the total number of segregation modes in Metaphase II, i.e., $2^{v/2}$, to obtain the expression $\binom{j}{i} \frac{2^i 2^{j-i} 2^{v/2 - 2j}}{2^{v/2}}$, i.e., $2^{-j} \binom{j}{i}$, which is the value of α_i conditional on j , where $i \leq j \leq \lfloor v/4 \rfloor$. Because the events that the secondary oocyte contains different numbers of PCPs are mutually exclusive, their probabilities are weighted by μ_j to obtain the weighted sum, i.e., α_i , whose expression is as follows:

$$\alpha_i = \sum_{j=i}^{\lfloor v/4 \rfloor} 2^{-j} \mu_j \binom{j}{i}. \quad (5)$$

Using Equation (5), the values of alpha of the CES can be derived for different inheritance modes, and these are presented in Table 2. The example deriving the values of alpha under octosomic inheritance is given in Appendix B. The expected number of IBDR allele pairs in a gamete $\lambda = \sum_i \alpha_i = v(v-2)/(16v-16)$.

Partial equational segregation

The *recombination fraction* (denoted by r) between two loci is defined as the ratio of the number of recombined gametes to the total number of gametes produced. The maximum recombination fraction for diploids is $r = 0.5$ (Xu 2013). However, in autopolyploids, because there are $v/2$ PCPs, the maximum recombination fraction for an autopolyploid organism is $r = 1 - 0.5^{v/2}$, e.g., $r = 0.75$ for autotetraploids.

We used the *single chromatid recombination rate* (denoted by r_s) between the target locus and the centromere (r_c) to evaluate the degree of recombination in polysomic inheritance, which is the probability that an allele in a chromatid is exchanged with the pairing chromatid. Hence, r and r_s can be converted to each other by the formula $r = 1 - (1 - r_s)^{v/2}$. Moreover, r_s can be calculated by Haldane's mapping function $r_s = \frac{1}{2} [1 - \exp(-2d/100)]$ where d is the distance between the target locus and the centromere (the unit of length is the centimorgan).

In the CES model, the two alleles in a pair of pairing chromatids are assumed to be exchanged at a probability of 100%. This 'ideal' condition is unlikely to occur in nature. We thus usually consider that the maximum single chromatid recombination rate is $r_s = 50\%$, and that the distance between the target locus and the centromere is extremely long. Although Mather (1935) incorporated the recombination rate into his model, assuming 100% recombination has been widely used in similar studies (e.g., Butruille and Boiteux 2000; Wu *et al.* 2001).

We incorporated the single chromatid recombination rate r_s into the CES to allow a greater generalization, which allows for the situation that only some fragments of the two pairing chromatids are exchanged between pairing chromosomes. Such a model is called the *partial equational segregation* (PES) model.

In the first step for CES, we model the probability μ_j that j PCPs segregate into a secondary oocyte, which remains unchanged in the PES.

Assuming that the alleles in each PCP are exchanged at a probability of r_s between pairing chromatids and those j PCPs are exchanged independently, then the number of *exchanged PCPs* (EPCPs) is drawn from the binomial distribution, and so the probability ν_k that the second oocyte contains k EPCPs is

$$\nu_k = \sum_{j=k}^{\lfloor v/4 \rfloor} \mu_j \binom{j}{k} r_s^k (1 - r_s)^{j-k}. \quad (6)$$

It is noteworthy that only the EPCPs are able to produce IBDR alleles under a PES model. Following the process of the second step for CES, we derive the alpha for PES by: (i) sampling i EPCPs from k EPCPs (there are $\binom{k}{i}$ ways to do this); (ii) sampling IBDR alleles from those i EPCPs (there are 2^i ways to do this); (iii) sampling non-IBDR alleles from the remaining $k - i$ EPCPs (there are 2^{k-i} ways to do this); (iv) sampling the remaining non-IBDR alleles from the remaining $v/2 - 2k$ chromosomes (there are $2^{v/2 - 2k}$ ways to do this). The product of the numbers of ways to do the above four steps is now

divided by the total number of segregation modes in Metaphase II, *i.e.*, $2^{v/2}$, to obtain the expression $\binom{k}{i} \frac{2^{i2^{k-i}v/2-2k}}{2^{v/2}}$, *i.e.*, $2^{-k} \binom{k}{i}$, which is the value of α_i conditional on k , where $i \leq k \leq \lfloor v/4 \rfloor$. Moreover, the events that the secondary oocyte contains different numbers of EPCPs are mutually exclusive. Therefore, the weighted sum of their probabilities with ν_k as their weights are calculated to obtain α_i under PES, whose expression is as follows:

$$\alpha_i = \sum_{k=i}^{\lfloor v/4 \rfloor} 2^{-k} \nu_k \binom{k}{i}, \quad (7)$$

where ν_k is given in Equation (6).

Using Equation (7), the values of alpha in the PES model for tetrasomic to dodecasomic inheritance can be derived, and the results are presented in Table 2. An example deriving the values of alpha under octosomic inheritance is also given in Appendix B. The expected number of IBDR allele pairs in a gamete is $\lambda = \sum_{i=0}^{\lfloor v/4 \rfloor} \alpha_i = r_s v(v-2)/(16v-16)$.

THE GENOTYPIC FREQUENCIES AT EQUILIBRIUM

Here we present two methods to derive the genotypic frequencies at a multiallelic locus for higher levels of ploidy. The first method, the non-linear method, is modified from Fisher's (1943) method and uses the transitional probability from zygotes to gametes to account for the multiallelic loci. Because this method is computationally difficult for a ploidy level higher than six under current conditions, we develop an alternative method to derive the genotypic frequency up to and including decasomic inheritance. The second method, the linear method, uses the solution of the non-linear method at a biallelic locus as the initial solution, and at each step a novel allele is segregated from an existing allele and the genotypic frequencies are updated.

Non-linear method

For this method, the total probability formula together with Equation (1) are used to derive the gamete frequencies and the multiplication theorem of probability is used to derive the zygote frequencies. If the genotypic frequencies reach equilibrium, they will not change subsequently, and the system of equations that we will construct can be solved by adding some constraints. In the following text, we denote GFG for *genotypic frequencies of gametes*, and GFZ for *genotypic frequencies of zygotes*. Our system of non-linear equations will be presented by the following two steps.

- i. Simulating meiosis. Given a gamete g , if the genotypic frequencies reach equilibrium, then the GFG is determined by the total probability formula, *i.e.*, the sum of the transition probabilities $T(g|G_i)$ weighted by GFZ, symbolically

$$\Pr(g) = \sum_i \Pr(G_i) T(g|G_i). \quad (8)$$

Here G_i is taken from all zygote genotypes and $T(g|G_i)$ can be obtained by Equation (1). Because there are $\binom{v/2+h-1}{v/2}$ gamete genotypes, Equation (8) represents $\binom{v/2+h-1}{v/2}$ equations.

- ii. Simulating fertilization. The GFZ in the next generation is determined by the multiplication theorem of probability with GFG as the weight. Under equilibrium, the GFG and GFZ are

constant across generations, then for a zygote G , we can establish an equation as follows:

$$\Pr(G) = \sum_i \Pr(g_i) \Pr(G \setminus g_i), \quad (9)$$

where G is regarded as a multiset, g_i is taken from all possible different subsets of G with $v/2$ elements and the symbol \setminus denotes the operation of set difference. Here, a *multiset* is a generalized set, whose elements can be repeated (*e.g.*, if $G = AAAB$), it can be regarded as a multiset consisting of four elements, *i.e.*, $G = \{A, A, A, B\}$, and its different subsets containing two elements are $\{A, A\}$ and $\{A, B\}$. Moreover, if $g = AA$ is regarded as $g = \{A, A\}$, then $G \setminus g = \{A, B\}$. Because there are $\binom{v+h-1}{v}$ zygote genotypes, Equation (9) represents $\binom{v+h-1}{v}$ equations. If the double-reduction ratios in two parents are different, then the genotypic frequencies of egg and sperm should be treated separately in Equation (9), *e.g.*, $\Pr(g_i)$ and $\Pr(G \setminus g_i)$ are replaced by the frequencies of an egg g_i and a sperm $G \setminus g_i$, respectively, and their frequencies are calculated by substituting different double-reduction parameters into Equation (8).

The process that transforms the allele frequencies into the GFG can be described by a linear substitution in the sense that every allele frequency can be expressed as a linear combination of the GFG. By substituting the equations determined by Equation (9) into the equations determined by Equation (8), we obtain a system of non-linear equations with the GFG as unknowns. This system of equations together with those expressions of the linear combinations mentioned above still forms a system of non-linear equations with the GFG as unknowns and the allele frequencies as parameters. For a lower level of ploidy, the latter system of non-linear equations can be solved, whose solution is with parameters. An example is given in Appendix C, which shows how to find the solution of the latter system of non-linear equations (*i.e.*, how to derive the expressions of GFG with parameters) for tetrasomic inheritance at a triallelic locus.

In Appendix C, for the case of GFG, the whole gamete genotypes are AA, BB, CC and AB, AC, BC . We choose AA and AB as representatives of the genotype pattern, then each genotype can be classified into one of these patterns. The partial solution determined by the representatives is called the *generalized form* of the solution mentioned above, whose expression is given in Equation (A3). For the case of GFZ, the notion of generalized form can be similarly defined, whose expression is given in Equation (A4).

For each gamete with $v/2$ allele copies, the number of alleles h at the target locus should at least be $v/2$ to allow for all genotypic patterns to be displayed. For example, the full gamete heterozygote $ABCD$ under octosomic inheritance can be observed at a tetra-allelic locus. Furthermore, if $h \geq v/2 + 1$, the generalized form for all genotypic patterns of GFG can be obtained. Because the allele frequencies at each locus sum to unity, the number of degrees of freedom is equal to $h - 1$. The full gamete heterozygotes can therefore be expressed by $h - 1$ allele frequencies.

Using the tetrasomic inheritance under RCS as an example, the frequency of a gamete $g = AB$ can be solved at a biallelic locus, and the solution is $2p_A(1 - p_A)$. Moreover, its solution with the generalized form at a triallelic locus can also be obtained, with the answer being $2p_A p_B$.

The number of equations determined by Equation (8) or Equation (9) increases exponentially with increasing levels of ploidy. For example, for the case of Equation (8), the number equals 6, 20, 70, 252 and 924 for tetrasomic inheritance to dodecasomic inheritance, while the

corresponding sequence is 15, 84, 495, 3003 and 18564 for the case of Equation (9). The two numerical sequences increase by about 3.5 and 6.0 times for each level, respectively. Moreover, each expression also becomes more complex as the ploidy level increases. For the case of dodecasomic inheritance, it takes multiple gigabytes of space to store these expressions.

Therefore, the system of non-linear equations becomes computationally difficult for higher levels of ploidy. Using this method, we solved the GFG and GFZ for a maximum ploidy level of six on a workstation with two Intel Xeon E5 2696 V3 CPUs, which have 36 cores in total.

Linear method

An alternative method to solve the genotypic frequencies at equilibrium is via the linear method. For a working example we focus on a protein-coding gene in a panmictic population at equilibrium with an infinite number of individuals. Suppose that there are three alleles with a unique DNA sequence, denoted by A , B and C , and let p_A , p_B and p_C be their frequencies. Due to the degeneration of codons, two alleles (*e.g.*, B and C) may code for the same protein. If isozyme electrophoresis is used for genotyping, then these two alleles cannot be distinguished, so both of them will, for example, be typed as B . However, if DNA sequencing is used, both alleles can be distinguished.

Therefore, it can be inferred that the number of alleles of this gene depends on how it is typed, with both biallelic and triallelic states conforming to equilibrium assumptions (*i.e.*, large population size, random mating, no mutation, no selection and no migration) and genotypic frequencies should concur with the equilibrium.

This scenario can also be expressed as follows: (i) originally, the locus is biallelic and the distribution of genotypes is at equilibrium; (ii) at each step, one novel allele mutates from an existing allele, the mutation rate is constant among different copies of this allele in all genotypes, and the IBD alleles in a genotype mutate simultaneously; (iii) therefore, after the mutation at each step, the genotypic frequency will still concur with the equilibrium; (iv) the mutations are repeated until there are $v/2 + 1$ alleles to obtain the generalized form of GFG. The following deductions are therefore derived.

Deduction (i): For both GFZ and GFG, the frequency of a genotype that contains only the unchanged alleles remains unchanged. In fact, the analytical expression of genotypic frequencies can be regarded as a function of the allele frequencies (refer to Equations (A3) and (A4) in Appendix C). Therefore, when a genotype contains only the unchanged alleles, the corresponding expression of the function is invariable in the process of mutation.

Deduction (ii): For both GFZ and GFG, the summations of the frequencies before and after the mutation for each genotype that has the same composition of unchanged alleles are equal (*e.g.*, $P_{AABB,2} = P_{AABB,3} + P_{AABC,3} + P_{AACC,3}$ where the last digit in every subscript denotes the number of alleles observed). This fact can be interpreted as several genotypes being merged into one when the locus switches from a triallelic state to a biallelic state.

Deduction (iii): For both GFZ and GFG, the ratio of frequencies of changed alleles in the genotypes that have the same composition of unchanged alleles is equal to the ratio of frequencies of changed alleles in the population. For example, if the changed alleles are B and C , we have $P_{AAAB,3} : P_{AAAC,3} = p_{B,3} : p_{C,3}$ and $2P_{AABB,3} + P_{AABC,3} : 2P_{AACC,3} + P_{AABC,3} = p_{B,3} : p_{C,3}$. Because the mutation rate is constant in different copies of B in all genotypes, a fixed proportion $\frac{p_{C,3}}{p_{B,3} + p_{C,3}}$ of B are mutated to C .

Deduction (iv): For two genotypes with the same pattern, we can use the expression of one genotype frequency to derive that of the other. For example, if the analytical expression $P_{AAAB,3} = f(p_{A,3}, p_{B,3})$ of the function is given, by replacing $p_{A,3}$ with $p_{B,3}$, and $p_{B,3}$ with $p_{C,3}$, it follows the expression of $P_{BBBC,3}$, that is $P_{BBBC,3} = f(p_{B,3}, p_{C,3})$.

According to Deduction (iv), if a generalized form for the GFG or GFZ is given, we can write down the whole expressions of the GFG or GFZ.

Using these deductions as stated above, the GFG can be solved step-by-step. At each step, one novel allele is separated from an existing allele. The procedure begins with a biallelic locus, which can be solved by the non-linear method. After the number of alleles reaches $v/2 + 1$, the generalized form of GFG is obtained, and then the GFZ can be derived by using Equation (9).

Using hexasomic inheritance as an example, we assume that the GFG at a biallelic locus is solved by the non-linear method, *i.e.*, the expressions of $P_{AAA,2}$, $P_{AAB,2}$ and $P_{ABB,2}$ are derived. If a new allele (*e.g.*, C) is segregated from B , we obtain $P_{AAA,3} = P_{AAA,2}$ via Deduction (i), and $P_{AAB,3} = \frac{p_{B,3}}{p_{B,3} + p_{C,3}} P_{AAB,2}$ via Deduction (iii). Additionally, $P_{BBB,3}$ and $P_{CCC,3}$ can be derived from $P_{AAA,3}$ via Deduction (iv). Similarly, $P_{AAC,3}$, $P_{ABB,3}$, $P_{BBC,3}$, $P_{ACC,3}$ and $P_{BCC,3}$ can be derived from $P_{AAB,3}$. Furthermore, $P_{ABC,3}$ will be obtained from $P_{ABB,2} - P_{ACC,3} - P_{ABB,3}$ via Deduction (ii). Because the degrees of freedom are currently two, one more step is required. In this step, P_{ABC} can be updated by Deduction (iii), and expressed by $P_{ABC,4} = \frac{p_{C,4}}{p_{C,4} + p_{D,4}} P_{ABC,3}$. Finally, P_{AAA} and P_{AAB} remain unchanged because of Deduction (i), and the remaining GFG will be obtained via Deduction (iv). The results of GFG and GFZ for hexasomic inheritance are presented in Appendix D.

The linear method can also be characterized by a system $\mathbf{Ax} = \mathbf{b}$ of linear equations. Using the first mutation in hexasomic inheritance as an example, the coefficient matrix \mathbf{A} is established (see Table 3 for details, in which p , q and r denote $p_{A,3}$, $p_{B,3}$ and $p_{C,3}$, respectively). The elements in \mathbf{b} can be derived from the above four deductions. For example, the first seven equations of $\mathbf{Ax} = \mathbf{b}$ are

$$P_{AAA,3} = P_{AAA,2}, \quad \text{via Deduction (i)}$$

$$P_{AAB,3} + P_{AAC,3} = P_{AAB,2}, \quad \text{via Deduction (ii)}$$

$$rP_{AAB,3} - qP_{AAC,3} = 0, \quad \text{via Deduction (iii)}$$

$$P_{ABB,3} + P_{ABC,3} + P_{ACC,3} = P_{ABB,2}, \quad \text{via Deduction (ii)}$$

$$2rP_{ABB,3} + (r - q)P_{ABC,3} - 2qP_{ACC,3} = 0. \quad \text{via Deduction (iii)}$$

$$P_{BBB,3} + P_{BBC,3} + P_{BCC,3} + P_{CCC,3} = P_{BBB,2}, \quad \text{via Deduction (ii)}$$

$$3rP_{BBB,3} + (2r - q)P_{BBC,3} + (r - 2q)P_{BCC,3} - 3qP_{CCC,3} = 0. \quad \text{via Deduction (iii)}$$

Then the first seven elements in \mathbf{b} are $P_{AAA,2}$, $P_{AAB,2}$, 0 , $P_{ABB,2}$, 0 , $P_{BBB,2}$, 0 . For these seven equations, one is derived by Deduction (i), where one is number of unchanged genotypes; three are derived by Deductions (ii) and (iii), respectively, where 3 is the number of changed genotypes. Therefore, with the first three deductions, we are able to establish the core equations. Furthermore,

■ **Table 3 Matrix A of the linear method of hexasomic inheritance at triallelic locus**

Genotypic frequencies of gamete									
$P_{AAA,3}$	$P_{AAB,3}$	$P_{AAC,3}$	$P_{ABB,3}$	$P_{ABC,3}$	$P_{ACC,3}$	$P_{BBB,3}$	$P_{BBC,3}$	$P_{BCC,3}$	$P_{CCC,3}$
1									
	1	1							
	r	$-q$							
			1	1	1				
			$2r$	$r - q$	$-2q$				
						1	1	1	1
						$3r$	$2r - q$	$r - 2q$	$-3q$
						1			
			1				1		
			r				$-p$		
	1			1				1	
	$2r$			$r - p$				$-2p$	
1		1			1				1
$3r$		$2r - p$			$r - 2p$				$-3p$
						1		1	1
						q		$-p$	
		1		1			1		
		$2q$		$q - p$			$-2p$		
1	1		1			1			
$3q$	$2q - p$		$q - 2p$			$-3p$			

Where p , q and r denote the frequencies of alleles A , B and C at a triallelic locus, respectively. The blank elements are zero.

by Deduction (iv) we can write down the remaining equations and the total number of equations is h times of the number of core equations.

The numbers of rows and columns of \mathbf{A} are 21 and 10, respectively, which means that there are 21 equations and 10 unknowns (*i.e.*, the whole GFG after mutation, *e.g.*, $P_{AAA,3}$ and $P_{AAB,3}$ and so on.) in this system $\mathbf{Ax} = \mathbf{b}$. There are thus many non-independent equations. The rank of \mathbf{A} can be derived by symbolic calculation, whose value is 10, equal to the number of unknowns. Therefore, our system of linear equations has a unique solution: $\mathbf{x} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}$.

The numbers of rows and columns of \mathbf{A} are respectively 39 and 28 for dodecasomic inheritance at a triallelic locus, and the rank of \mathbf{A} is 27 by symbolic calculation. Unfortunately, because there are 28 unknowns, our system $\mathbf{Ax} = \mathbf{b}$ is underdetermined and has an infinite number of solutions.

Using this method, we are able to derive the GFZ and GFG from tetrasomic to decasomic inheritance. For the cases of tetrasomic and hexasomic inheritance, these results are identical to the solutions obtained from the non-linear method. Subsequent expressions become more complex with increasing levels of ploidy, so we only present the results for hexasomic inheritance in Appendix D.

As an alternative, these expressions are placed in a 660 KB C++ source code for a library, and the source code is available in our supplementary material. The phenotypic frequencies can then be calculated by taking the sum of the frequencies of all possible genotypes that yield this phenotype, which are also given in the source code.

Inbreeding coefficient and heterozygosity

Double-reduction has a similar effect as inbreeding, both not only resulting in IBD alleles in the zygote but also slightly increasing homozygosity (Hardy 2016). Here, we assess the impact of inbreeding and double-reduction on the inbreeding coefficient. The concept of inbreeding is relative to a reference population, which by definition is without inbreeding or relatedness, with all alleles in the reference population defined as not being IBD (Lynch and Walsh 1998).

The *inbreeding coefficient* in polysomic inheritance is defined as the probability of sampling two IBD alleles from a genotype without replacement (Huang *et al.* 2015). In polyploids, gametes also have multiple alleles at a target locus, and the inbreeding coefficient and heterozygosity can also be applied to gametes to measure the degree of IBD or *identical-by-state* (IBS) allele pairs (Barone *et al.* 1995).

After fertilization, there are $\binom{v}{2}$ pairs of alleles in a zygote, where $\binom{v/2}{2}$ pairs of alleles are from one gamete, and $v^2/4$ pairs of alleles are from different gametes. The numbers of IBD allele pairs in both types of allele pairs are derived as follows.

The number of IBDR allele pairs in the $\binom{v/2}{2}$ allele pairs from the same gamete is the expected value λ of allele pairs, where $\lambda = \sum_{i=0}^{\lfloor v/4 \rfloor} i \alpha_i$. In the remaining allele pairs, the number of IBD allele pairs inherited from the parent is a proportion F of those remaining, where F is the inbreeding coefficient in zygotes of the previous generation. Therefore, there are $\lambda + F \left[\binom{v/2}{2} - \lambda \right]$ pairs of IBD alleles in each gamete. Hence the inbreeding coefficient in the gamete is:

$$f = \frac{\lambda + F \left[\binom{v/2}{2} - \lambda \right]}{\binom{v/2}{2}}. \quad (10)$$

For the $v^2/4$ allele pairs from two different fertilizing gametes, the mating individuals in the presence of inbreeding may also share IBD alleles. The *co-ancestry coefficient* is used to measure the degree of kinship in mating individuals, where the *co-ancestry coefficient*, denoted by θ , is defined as the probability that two alleles, one randomly sampled from each individual, are IBD (Jacquard 1972). Then there is a value $\theta v^2/4$ expected IBD allele pairs between fertilizing gametes.

From the above derivation, the number of pairs of IBD alleles in the zygote is in total $2\lambda + 2F \left[\binom{v/2}{2} - \lambda \right] + \theta v^2/4$. Therefore, the inbreeding coefficient for zygotes of the current generation F' is as follows:

$$F' = \frac{2\lambda + 2F \left[\binom{v/2}{2} - \lambda \right] + \theta v^2/4}{\binom{v}{2}}.$$

The formula is the transition function of inbreeding coefficients. If the genotypic frequencies reach equilibrium, these inbreeding coefficients will be constant across generations. By replacing F' with F in the transition function, the inbreeding coefficient F at equilibrium can be solved, whose expression is as follows:

$$F = \frac{8\lambda + \theta v^2}{8\lambda + v^2}. \quad (11)$$

For an outcrossed population, $\theta = 0$. Then, by substituting Equation (11) for Equation (10), the inbreeding coefficients F and f can be obtained, whose values at equilibrium for different double-reduction models are shown in the rightmost two columns of Table 2.

If the relationships between mating individuals can be classified into several types (e.g., self: selfing, parent-offspring: backcross), then the co-ancestry coefficient θ can be derived by a weighted average of the co-ancestry coefficients of those types of relationships. An example to derive the expressions of these co-ancestry coefficients as well as the inbreeding coefficient F is given in Appendix E.

Following the modification of the definition of inbreeding coefficients, the *heterozygosity* in polysomic inheritance is defined as the probability of sampling two non-IBS alleles from a genotype without replacement (Hardy 2016). For the two alleles sampled, if they are IBD, then these alleles cannot be IBS; otherwise they are independent and are non-IBS alleles at a probability of $1 - \sum_i^h p_i$. Let H be the heterozygosity of zygotes and h be that of gametes, then:

$$H = (1 - F) \left(1 - \sum_i^h p_i \right),$$

$$h = (1 - f) \left(1 - \sum_i^h p_i \right). \quad (12)$$

Where F is the inbreeding coefficient of zygotes and f is that of gametes. These expressions are identical to Nei's (1977) estimator of inbreeding coefficients, i.e., $F_{IS} = \frac{H_S - H_I}{H_S}$, where H_I and H_S are the observed and expected heterozygosities, respectively.

Simulation

The numerical solutions of genotypic and phenotypic frequencies of gametes and zygotes calculated by computer simulations will be used to compare with the analytical solutions so as to validate our models.

A population is generated with its initial genotypic frequency drawn from the HWE. Because we do not store the specific genotype for each individual, but only the frequency of each genotype, the population size can be considered as infinite. To display all kinds of genotype patterns, the number of alleles h should be greater than or equal to the ploidy level. For example, the complete heterozygote in tetrasomic inheritance

$ABCD$ cannot be observed at a triallelic locus. For simplicity, suppose that the allele frequencies are uniformly distributed, i.e., the frequency of any allele is equal to $1/h$, so that the frequencies of genotypes or phenotypes with a same pattern are equal (e.g., $P_{AAAB} = P_{BBBC}$).

The double-reduction models are simulated to calculate the gamete frequencies. First, for PRCS, the alleles in the zygotes are duplicated and then randomly segregate into four gametes. Second, for CES, the chromosomes are randomly paired and the alleles are exchanged between pairing chromosomes, and then the chromosomes randomly segregate into two secondary oocytes, finally the alleles within the same chromosomes randomly segregate into two gametes. Third, for PES, the chromosomes are randomly paired and the alleles are exchanged between paired chromosomes at a probability of r_s , and the remaining steps are the same as CES.

For a zygote, there are many possible ways to generate gametes. We enumerate each possibility and calculate the weighted average of gamete frequencies according to the probability of each possibility. The frequencies of identical gametes produced from different zygotes are weighed again by the zygote frequencies so as to obtain the gamete frequencies in the population.

The gametes are randomly merged to simulate fertilization, then the zygote frequency in the next generation can be calculated. The population is reproduced for 100 generations to converge the genotypic frequencies. The simulation program can be found in the electronic supplementary materials.

There are many potential genotypes and phenotypes in a population, and it is thus impractical to show all of these frequencies in the manuscript. As an alternative, the phenotypic frequencies of zygote patterns are shown in Table 4, and the genotypic frequencies of gamete patterns are shown in Table 5. As a comparison, the corresponding analytical solutions of model predictions are also shown. It is clear from the data shown in Tables 4 and 5 that both of numerical and analytical solutions of phenotypic frequencies of zygotes are approximately equal and their differences are due to the rounding-off error, as are the genotypic frequencies of gametes obtained via different approaches.

DATA AVAILABILITY

Material S1

The appendices.

Material S2

The simulation computer program to obtain the numerical solutions of genotypic and phenotypic frequencies of zygotes and the genotypic frequencies of gametes. The program can be run on Windows platforms and requires the .Net Framework V3.5 runtime library.

Material S3

The C++ source code of the genotypic and phenotypic frequencies of polysomic inheritance under double-reduction. The prefix of functions is GFZ (genotypic frequencies of zygotes), GFG (genotypic frequencies of gametes), PFZ (phenotypic frequencies of zygotes), or PFG (phenotypic frequencies of gametes). An example that is the function in the second case (i.e., $G = AAAB$) in Equation (A4) is given by

```
double GFZ4_iiij(double a1, double pi, double pj)
{
    double pi2 = pi*pi;
    return (8*(-1+a1)*pi2*(-3*a1+2*(-1+a1)*pi)*pj)/
        pow(2+a1,2);
}
```

Table 4. The numerical and analytical solutions of phenotypic frequencies obtained by numerical simulation and model predication. The number of alleles at this locus is equal to the ploidy level and their frequencies are uniformly distributed.

Ploidy Level	Phenotype Pattern	PRCS		CES		PES ($r_s = 0.5$)	
		Numerical	Analytical	Numerical	Analytical	Numerical	Analytical
4	A	1.0000×10^{-2}	1	1.1187×10^{-2}	121	7.2250×10^{-3}	289
	AB	7.0000×10^{-2}	7	7.2300×10^{-2}	10816	6.3950×10^{-2}	40000
	ABC	1.2000×10^{-1}	100	1.1649×10^{-1}	391	1.2870×10^{-1}	1279
	ABCD	6.0000×10^{-2}	25	5.5473×10^{-2}	5408	7.2600×10^{-2}	20000
6	A	2.5942×10^{-4}	3	3.0092×10^{-4}	2704	1.1491×10^{-4}	1287
	AB	4.3851×10^{-3}	50	4.7641×10^{-3}	75	2.8338×10^{-3}	363
	ABC	1.7954×10^{-2}	2039184	1.8512×10^{-2}	1352	1.5254×10^{-2}	5000
	ABCD	2.9388×10^{-2}	59976	2.8841×10^{-2}	566820864	3.1578×10^{-2}	946729
8	ABCDE	2.1185×10^{-2}	263	1.9806×10^{-2}	1350199	2.8157×10^{-2}	8238629824
	ABCDEF	5.6493×10^{-3}	59976	5.0416×10^{-3}	283410432	9.1174×10^{-3}	114428192
	ABCDEF	5.5456×10^{-6}	1017	6.8974×10^{-6}	291465	1.3534×10^{-6}	521645
	ABCDEF	1.8796×10^{-4}	56644	2.1435×10^{-4}	170569	7.7138×10^{-5}	5714096
10	ABCDEF	1.4242×10^{-3}	2497	1.5336×10^{-3}	1350199	8.3918×10^{-4}	379641343623698421760000
	ABCDEF	4.4981×10^{-3}	84966	4.6241×10^{-3}	283410432	3.5890×10^{-3}	14642346312283931623
	ABCDEF	7.2113×10^{-3}	300	7.1175×10^{-3}	15745024	4.7785×10^{-3}	189820671511849210880000
	ABCDEF	6.2074×10^{-3}	14161	5.9010×10^{-3}	227055	1.0850×10^{-3}	159292370706287141369
	ABCDEF	2.7424×10^{-3}	80	2.5160×10^{-3}	7872512	1.2618×10^{-8}	170318790423175116659
	ABCDEF	4.9012×10^{-4}	14161	4.3461×10^{-4}	155925	1.5416×10^{-6}	35929647497090903787
	ABCDEF	9.9572×10^{-8}	22080601	1.2994×10^{-7}	19845	3.0191×10^{-5}	47455167877962302720000
	ABCDEF	6.3347×10^{-6}	39816530864	7.5433×10^{-6}	3936256	4.8492×10^{-3}	47455167877962302720000
	ABCDEF	7.9681×10^{-5}	374186951	8.9303×10^{-5}	415212550274159479	1.7961×10^{-3}	35929647497090903787
	ABCDEF	4.1178×10^{-4}	472567345	4.3996×10^{-4}	14855586629293839325	2.2915×10^{-3}	20024123292350185869
	ABCDEF	1.1179×10^{-3}	746239705	1.1472×10^{-3}	22393137876378735	2.2915×10^{-3}	2372758393898115136000
	ABCDEF	1.7699×10^{-3}	165902211936	1.7538×10^{-3}	48426989441478428672	1.7315×10^{-3}	1186379196949057568000

■ **Table 5** The numerical and analytical solutions of gamete frequencies obtained by numerical simulation and model predication. The number of alleles at this locus is equal to the ploidy level and their frequencies are uniformly distributed.

Ploidy Level	Gamete Pattern	PRCS		CES		PES ($r_s = 0.5$)	
		Numerical	Analytical	Numerical	Analytical	Numerical	Analytical
4	AA	1.0000×10^{-1}	$\frac{1}{10}$	1.0577×10^{-1}	$\frac{11}{104}$	8.5000×10^{-2}	$\frac{17}{200}$
	AB	1.0000×10^{-1}	$\frac{1}{10}$	9.6154×10^{-2}	$\frac{5}{52}$	1.1000×10^{-1}	$\frac{11}{100}$
6	AAA	1.6106×10^{-2}	$\frac{23}{1428}$	1.7347×10^{-2}	$\frac{413}{23808}$	1.0720×10^{-2}	$\frac{973}{90768}$
	AAB	1.8908×10^{-2}	$\frac{9}{476}$	1.9279×10^{-2}	$\frac{153}{7936}$	1.6955×10^{-2}	$\frac{513}{30256}$
	ABC	1.6807×10^{-2}	$\frac{2}{119}$	1.5877×10^{-2}	$\frac{63}{3968}$	2.1351×10^{-2}	$\frac{323}{15128}$
8	AAAA	2.3549×10^{-3}	$\frac{4699}{1995408}$	2.6263×10^{-3}	$\frac{163467203}{62242730624}$	1.1634×10^{-3}	$\frac{716811013}{616150422400}$
	AAAB	2.8786×10^{-3}	$\frac{359}{124713}$	2.9714×10^{-3}	$\frac{46236677}{15560682656}$	2.0281×10^{-3}	$\frac{312402211}{154037605600}$
	AABB	3.6414×10^{-3}	$\frac{1211}{332568}$	3.9091×10^{-3}	$\frac{121657401}{31121365312}$	2.5527×10^{-3}	$\frac{786411807}{308075211200}$
	AABC	3.1713×10^{-3}	$\frac{791}{249426}$	3.1471×10^{-3}	$\frac{48971403}{15560682656}$	3.1551×10^{-3}	$\frac{97200513}{30807521120}$
	ABCD	2.6461×10^{-3}	$\frac{110}{41571}$	2.4917×10^{-3}	$\frac{19386393}{7780341328}$	3.9370×10^{-3}	$\frac{303223311}{77018802800}$
10	AAAAA	3.1555×10^{-4}	$\frac{2706677}{8577650312}$	3.6047×10^{-4}	$\frac{28428076219}{78864251224000}$	1.1233×10^{-4}	$\frac{317519939737}{28266871609216000}$
	AAAAB	4.0811×10^{-4}	$\frac{71441}{175054088}$	4.3549×10^{-4}	$\frac{6868978459}{15772850244800}$	2.1295×10^{-4}	$\frac{1203913445713}{5653374321843200}$
	AAABB	5.5441×10^{-4}	$\frac{2377759}{4288825156}$	6.0075×10^{-4}	$\frac{23688711503}{39432125612000}$	3.0058×10^{-4}	$\frac{4248298069913}{14133435804608000}$
	AAABC	4.7366×10^{-4}	$\frac{72551}{153172327}$	4.7477×10^{-4}	$\frac{9360681119}{19716062806000}$	3.6809×10^{-4}	$\frac{2601170012999}{7066717902304000}$
	AABBC	6.0212×10^{-4}	$\frac{2582389}{4288825156}$	6.2503×10^{-4}	$\frac{24646078029}{39432125612000}$	4.6039×10^{-4}	$\frac{6506930361039}{14133435804608000}$
	AABCD	5.0189×10^{-4}	$\frac{538128}{1072206289}$	4.8982×10^{-4}	$\frac{1931448081}{3943212561200}$	5.6757×10^{-4}	$\frac{160435060233}{282668716092160}$
	ABCDE	4.0219×10^{-4}	$\frac{431232}{1072206289}$	3.8000×10^{-4}	$\frac{749214021}{1971606280600}$	7.0494×10^{-4}	$\frac{99631848687}{141334358046080}$

The authors state that all data necessary for confirming the conclusions presented in the article are represented fully within the article. Supplemental material available at Figshare: <https://doi.org/10.25387/g3.7883150>.

DISCUSSION

Double-reduction models

Here, we describe existing double-reductions models and extend them to the systems that account for any even level of ploidy. However, some of these models assume ideal situations that are unrealistic in nature. The CES model assumes that the single chromatid recombination rate r_s is one at any locus, but the maximum value of r_s is 50% if the locus is located extremely far from the centromere. We incorporate the single chromatid recombination rate into CES, and define this as the partial equational segregation (PES).

Although PES probably reflects more closely to natural patterns of occurrence and may thus more accurately explain real genotypic frequencies, other models (RCS, PRCS and CES) are still useful. Researchers are able to apply them to calculate the genotypic frequencies in their applications without estimating alpha. In contrast, the PES needs an additional parameter r_s to calculate the genotypic frequencies, and thus has one more degree of freedom if r_s is estimated from the same data. Although PES may perform better in calculating the likelihood from genotypic/phenotypic data, it is also penalized more via the AIC (Akaike information criterion, Akaike 1974) or the BIC (Bayesian information criterion, Schwarz 1978).

In addition, all of these models (RCS, PRCS, CES and PES) assume that the chromosomes must form multivalents during prophase I. However, meiosis of polyploids can be much more complex. Some allopolyploid species display a mode of inheritance intermediate between disomic and polysomic at some loci (segmental allopolyploids, Stift *et al.* 2008), and some autopolyploid species also form bivalent, univalent and other types of valent

during meiosis (Qu *et al.* 1998; Lloyd and Bomblies 2016). The formation of different types of valent may influence the sterility of the gametes or seeds (Swaminathan and Sulbha 1959; Crowley and Rees 1968; Solís Neffa and Fernández 2000). Such a model can hardly be established if we consider so many factors, *e.g.*, fertility, proportion and double-reduction ratio for each kind of valent.

Regardless of how complex the nature of meiosis, different models will yield the same result: the IBDR alleles are in the fertile gamete. Therefore, all of these models can be incorporated into a generalized framework, *i.e.*, using the values of alpha to express a model. Meanwhile, comparing with RCS, PRCS and CES models, the number of degrees of freedom of such a model increases only by $\lfloor v/4 \rfloor$.

Genotypic frequencies, inbreeding coefficient and heterozygosity

Two methods are proposed to solve the genotypic equilibrium under double-reduction at a locus with an arbitrary number of alleles. The first method is the non-linear method, which can theoretically be applied to any even level of ploidy. However, it becomes computationally difficult for organisms with high levels of ploidy because these organisms have too many potential genotypes (*e.g.*, there are 19494 non-linear equations in dodecasomic inheritance). Although the system of non-linear equations for a higher level of ploidy is computationally difficult and we can currently only solve the genotypic frequencies for hexasomic inheritance, it will be possible with more powerful computers and more advanced algorithms.

The second method is the linear method, which requires more steps but less computational power. This uses the solution obtained by the non-linear method at a biallelic locus as the initial solution, and at each step it is assumed that one novel allele mutates from an existing allele. Some novel genotypes will be separated from existing genotypes, while the genotypic frequencies will still concur with the equilibrium. Using the four deductions, the equations of the genotypic frequencies of

gametes are initially established by the total probability formula, and then the equations of the genotypic frequencies of zygotes are established by the multiplication theorem of probability. Unfortunately, the constraints are insufficient to obtain a unique solution for ploidy levels greater than or equal to 12. Although this is not perfect, our method is more than adequate to answer most current research questions on this subject.

Double-reduction results in gametes carrying IBD alleles and slightly increased homozygosity (Hardy 2016). It has similar effects as inbreeding, and can cause a deviation in genotypic frequencies from the Hardy-Weinberg equilibrium (Luo *et al.* 2006). By modifying the definition of the inbreeding coefficient and heterozygosity in polysomic inheritance, the symbolic expressions measuring how double-reduction and inbreeding influence the inbreeding coefficient and heterozygosity are derived. We found that double-reduction can cause a maximum inbreeding coefficient being 0.0769 (in CES) under tetrasomic inheritance in an outcrossed population. For the same double-reduction model, the double-reduction ratio is increased while the inbreeding coefficient is reduced with increasing levels of ploidy (Table 2).

In the presence of double-reduction and/or inbreeding, the theoretical observed heterozygosity can be calculated from the inbreeding coefficient, and is identical to Nei's (1977) estimator of the inbreeding coefficient. This means that the inbreeding coefficient in polysomic inheritance can still be estimated by previous estimators, the same as for other *F*-statistics, although some modifications should be made to account for the finite sample size (*e.g.*, Robertson and Hill 1984; Weir and Cockerham 1984).

Applications of zygote frequencies

Genotypic frequencies can be used for many applications, such as for population genetics, molecular ecology, molecular breeding, and so on. Some of the applications of zygote frequencies are listed as follows.

- i. Estimating the allele frequency from phenotypes: the probability that observing a genotype conditional on a phenotype can be derived from the GFZ and Bayes formula. By counting the number of copies of each allele in each possible genotype and using the conditional probability as a weight, the allele frequencies can be updated from an initial solution with an Expectation Maximization algorithm (*e.g.*, De Silva *et al.* 2005; Kalinowski and Taper 2006).
- ii. Based on the allele frequency, many subsequent analyses can be performed. For example, genetic diversity analysis (*e.g.*, Rousset 2008), genetic distance analysis (*e.g.*, Nei 1972), population differentiation analysis (*e.g.*, Cockerham 1973), analysis of molecular variances (Excoffier and Lischer 2010), principal coordination analysis (*e.g.*, Peakall and Smouse 2012), and hierarchy clustering (*e.g.*, Odong *et al.* 2011).
- iii. Population assignment and Bayesian clustering: by calculating the product of multilocus genotypic frequencies, the probability of randomly sampling an individual with certain multilocus genotypes from a population can be obtained (*i.e.*, likelihood). The individual is assigned to the population with maximum likelihood (*e.g.*, Peakall and Smouse 2012). In the non-admixture model of Bayesian clustering, the population origin of each individual is randomly drawn from the populations according to the posterior probability that the individual originated from each population, and is updated iteratively in a Markov Chain Monte Carlo (MCMC) algorithm (Pritchard *et al.* 2000). The posterior probability is calculated from the multilocus genotypic frequencies of

the individual from each population by the Bayes formula (Pritchard *et al.* 2000).

- iv. Equilibrium test: as a Hardy-Weinberg equilibrium test, this test is able to evaluate whether the distribution of genotypes/phenotypes concurs with the equilibrium state for a particular double-reduction model. The observed number of each genotype/phenotype can be obtained by counts from the data, and the expected number can be obtained by the GFZ or PFZ. A Chi-square goodness-of-fit test or a Fisher's *G* test can be used for this application (*e.g.*, Rousset 2008).
- v. Linkage disequilibrium test: the contingency table of a pair of loci is established, with each row denoting a genotype of the first locus with each column denoting that of the second locus. The expected number in each cell under the hypothesis that the two loci are under linkage equilibrium is the product of the GFZ at these two loci (*e.g.*, Rousset 2008).
- vi. Estimating individual inbreeding coefficient, relatedness coefficient or kinship coefficient from phenotypes: method-of-moment estimators (*e.g.*, Hardy and Vekemans 2002; Huang *et al.* 2014) or the maximum-likelihood estimator (*e.g.*, Huang *et al.* 2015) can be used to estimate each of these coefficients between possible genotypes, with the estimates being weighted according to the conditional probability in application (i) above.

Applications of gamete frequencies

There are also some potential applications of the transitional probability from zygotes to gametes and the gamete frequencies, in addition to solve the values of alpha for PRCS (Equation (3)) and to derive the genotypic frequencies at equilibrium (Equation (8)).

- i. Parentage analysis: the transitional probability that a parent or a pair of parents produce an offspring is used in the calculation of likelihoods of the following two hypotheses: (a) the alleged father is the true father of the offspring and (b) the alleged father is unrelated to the offspring and is randomly sampled from the population (Marshall *et al.* 1998; Kalinowski *et al.* 2007). This transitional probability is a summation of the products of two gamete frequencies that form the offspring's genotype, where the frequency is equal to $\Pr(g | G_a)\Pr(G_o \setminus g | G_m)$ if the mother's genotype is known, or equal to $\Pr(g | G_a)\Pr(G_o \setminus g)$ if the mother's genotype is unknown, in which G_o , G_a and G_m denote the genotypes of the offspring, alleged father and true mother, respectively.
- ii. Deriving the segregation ratio in mapping populations, or deriving the genotypic frequencies for populations which are not at equilibrium: we use an F_2 population in tetrasomic inheritance as an example. Assuming the genotypes of F_1 individuals are *AABB*, using Equation (1), the gamete ratio is $AA : AB : BB = 1 + 2\alpha_1 : 4 - 4\alpha_1 : 1 + 2\alpha_1$. From the multiplication theorem, the segregation ratio of the F_2 population is $AAAA : AAAB : AABB : ABBB : BBBB = (1 + 2\alpha_1)^2 : 8(1 + \alpha_1 - 2\alpha_1^2) : 6(3 - 4\alpha_1 + 4\alpha_1^2) : 8(1 + \alpha_1 - 2\alpha_1^2) : (1 + 2\alpha_1)^2$.
- iii. Estimating the values of alpha from mapping populations: by equating the observed segregation ratio with the expected value, a system of non-linear equations can be established. The values of alpha can be solved by the least-squares method.
- iv. Perform an equilibrium test or a linkage disequilibrium test: the deviation of the observed gamete frequencies from expected can be measured by either a χ^2 or a *G* statistic by using either a Chi-square test or a Fisher's *G*-test, respectively. With the cumulative distribution function of a Chi-squared distribution, the significance can be calculated.

ACKNOWLEDGMENTS

We would like to thank the two anonymous reviewers for their suggestions and comments. This study was funded by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB31020302), the Natural Science Foundation of China (31770411, 31730104, 31572278 and 31770425), and the Young Elite Scientists Sponsorship Program by CAST (2017QNRC001), the National Key Program of Research and Development, Ministry of Science and Technology (2016YFC0503200), and the Natural Science Basic Research Plan in Shaanxi Province of China (2018JM3024). DWD is supported by a Shaanxi Province Talents 100 Fellowship.

LITERATURE CITED

- Akaike, H., 1974 A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* 19: 716–723. <https://doi.org/10.1109/TAC.1974.1100705>
- Barone, A., C. Gebhardt, and L. Frusciant, 1995 Heterozygosity in 2n gametes of potato evaluated by RFLP markers. *Theor. Appl. Genet.* 91: 98–104. <https://doi.org/10.1007/BF00220864>
- Burrow, M. D., C. E. Simpson, J. L. Starr, and A. H. Paterson, 2001 Transmission genetics of chromatin from a synthetic amphidiploid to cultivated peanut (*Arachis hypogaea* L.): broadening the gene pool of a monophyletic polyploid species. *Genetics* 159: 823–837.
- Butruille, D. V., and L. S. Boiteux, 2000 Selection-mutation balance in polysomic tetraploids: impact of double reduction and gametophytic selection on the frequency and subchromosomal localization of deleterious mutations. *Proc. Natl. Acad. Sci. USA* 97: 6608–6613. <https://doi.org/10.1073/pnas.100101097>
- Clark, L. V., and M. Jasieniuk, 2011 polysat: an R package for polyploid microsatellite analysis. *Mol. Ecol. Resour.* 11: 562–566. <https://doi.org/10.1111/j.1755-0998.2011.02985.x>
- Cockerham, C. C., 1973 Analyses of gene frequencies. *Genetics* 74: 679–700.
- Comai, L., 2005 The advantages and disadvantages of being polyploid. *Nat. Rev. Genet.* 6: 836–846. <https://doi.org/10.1038/nrg1711>
- Crowley, J. G., and H. Rees, 1968 Fertility and selection in tetraploid *Lolium*. *Chromosoma* 24: 300–308. <https://doi.org/10.1007/BF00336197>
- Darlington, C. D., 1929 Chromosome behaviour and structural hybridity in the Tridacnidae. *J. Genet.* 21: 207–286. <https://doi.org/10.1007/BF02984208>
- David, L., S. Blum, M. W. Feldman, U. Lavi, and J. Hillel, 2003 Recent duplication of the common carp (*Cyprinus carpio* L.) genome as revealed by analyses of microsatellite loci. *Mol. Biol. Evol.* 20: 1425–1434. <https://doi.org/10.1093/molbev/msg173>
- De Silva, H. N., A. J. Hall, E. Rikkerink, M. A. McNeillage, and L. G. Fraser, 2005 Estimation of allele frequencies in polyploids under certain patterns of inheritance. *Heredity* 95: 327–334. <https://doi.org/10.1038/sj.hdy.6800728>
- Excoffier, L., and H. E. L. Lischer, 2010 arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* 10: 564–567. <https://doi.org/10.1111/j.1755-0998.2010.02847.x>
- Fisher, R. A., 1943 Allowance for double reduction in the calculation of genotype frequencies with polysomic inheritance. *Ann. Hum. Genet.* 12: 169–171.
- Fisher, R. A., and K. Mather, 1943 The inheritance of style length in *Lythrum salicaria*. *Ann. Eugen.* 12: 1–23. <https://doi.org/10.1111/j.1469-1809.1943.tb02307.x>
- Geiringer, H., 1949 Chromatid segregation of tetraploids and hexaploids. *Genetics* 34: 665–684.
- Haldane, J. B. S., 1930 Theoretical genetics of autopolyploids. *J. Genet.* 22: 359–372. <https://doi.org/10.1007/BF02984197>
- Hardy, O. J., 2016 Population genetics of autopolyploids under a mixed mating model and the estimation of selfing rate. *Mol. Ecol. Resour.* 16: 103–117. <https://doi.org/10.1111/1755-0998.12431>
- Hardy, O. J., and X. Vekemans, 2002 spagedi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol. Ecol. Notes* 2: 618–620. <https://doi.org/10.1046/j.1471-8286.2002.00305.x>
- Huang, K., S. T. Guo, X. G. Qi, P. Zhang, and B. G. Li, 2015 A maximum-likelihood estimation of pairwise relatedness for autopolyploids. *Heredity* 114: 133–142. <https://doi.org/10.1038/hdy.2014.88>
- Huang, K., K. Ritland, S. T. Guo, M. Shattuck, and B. G. Li, 2014 A pairwise relatedness estimator for polyploids. *Mol. Ecol. Resour.* 14: 734–744. <https://doi.org/10.1111/1755-0998.12217>
- Jacquard, A., 1972 Genetic information given by a relative. *Biometrics* 28: 1101–1114. <https://doi.org/10.2307/2528643>
- Kalinowski, S. T., and M. L. Taper, 2006 Maximum likelihood estimation of the frequency of null alleles at microsatellite loci. *Conserv. Genet.* 7: 991–995. <https://doi.org/10.1007/s10592-006-9134-9>
- Kalinowski, S. T., M. L. Taper, and T. C. Marshall, 2007 Revising how the computer program cervus accommodates genotyping error increases success in paternity assignment. *Mol. Ecol.* 16: 1099–1106. <https://doi.org/10.1111/j.1365-294X.2007.03089.x>
- Limborg, M. T., W. A. Larson, L. W. Seeb, and J. E. Seeb, 2017 Screening of duplicated loci reveals hidden divergence patterns in a complex salmonid genome. *Mol. Ecol.* 26: 4509–4522. <https://doi.org/10.1111/mec.14201>
- Lloyd, A., and K. Bomblies, 2016 Meiosis in autopolyploid and allopolyploid *Arabidopsis*. *Curr. Opin. Plant Biol.* 30: 116–122. <https://doi.org/10.1016/j.pbi.2016.02.004>
- Luo, Z. W., Z. E. Zhang, R. M. Zhang, M. Pandey, O. Gailing *et al.*, 2006 Modeling population genetic data in autotetraploid species. *Genetics* 172: 639–646. <https://doi.org/10.1534/genetics.105.044974>
- Lynch, M., and B. Walsh, 1998 *Genetics and analysis of quantitative traits*, Sinauer Associates, Sunderland.
- Madlung, A., 2013 Polyploidy and its effect on evolutionary success: old questions revisited with new tools. *Heredity* 110: 99–104. <https://doi.org/10.1038/hdy.2012.79>
- Marshall, T. C., J. Slate, L. E. B. Kruuk, and J. M. Pemberton, 1998 Statistical confidence for likelihood-based paternity inference in natural populations. *Mol. Ecol.* 7: 639–655. <https://doi.org/10.1046/j.1365-294x.1998.00374.x>
- Mather, K., 1935 Reductional and equational separation of the chromosomes in bivalents and multivalents. *J. Genet.* 30: 53–78. <https://doi.org/10.1007/BF02982205>
- Meirmans, P. G., and P. H. van Tienderen, 2004 genotype and genodive: two programs for the analysis of genetic diversity of asexual organisms. *Mol. Ecol. Notes* 4: 792–794. <https://doi.org/10.1111/j.1471-8286.2004.00770.x>
- Muller, H. J., 1914 A new mode of segregation in Gregory's tetraploid primulas. *Am. Nat.* 48: 508–512. <https://doi.org/10.1086/279426>
- Nei, M., 1972 Genetic distance between populations. *Am. Nat.* 106: 283–292. <https://doi.org/10.1086/282771>
- Nei, M., 1977 F-statistics and analysis of gene diversity in subdivided populations. *Ann. Hum. Genet.* 41: 225–233. <https://doi.org/10.1111/j.1469-1809.1977.tb01918.x>
- Odong, T. L., J. van Heerwaarden, J. Jansen, T. J. van Hintum, and F. A. van Eeuwijk, 2011 Determination of genetic structure of germplasm collections: are traditional hierarchical clustering methods appropriate for molecular marker data? *Theor. Appl. Genet.* 123: 195–205. <https://doi.org/10.1007/s00122-011-1576-x>
- Otto, S. P., 2007 The evolutionary consequences of polyploidy. *Cell* 131: 452–462. <https://doi.org/10.1016/j.cell.2007.10.022>
- Parisod, C., R. Holderegger, and C. Brochmann, 2010 Evolutionary consequences of autopolyploidy. *New Phytol.* 186: 5–17. <https://doi.org/10.1111/j.1469-8137.2009.03142.x>
- Peakall, R., and P. E. Smouse, 2012 genalex 6.5: genetic analysis in Excel. Population genetic software for teaching and research update. *Bioinformatics* 28: 2537–2539. <https://doi.org/10.1093/bioinformatics/bts460>
- Pritchard, J. K., M. Stephens, and P. Donnelly, 2000 Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.
- Qu, L., J. Hancock, and J. Whallon, 1998 Evolution in an autopolyploid group displaying predominantly bivalent pairing at meiosis: genomic similarity of diploid *Vaccinium darrowi* and autotetraploid V.

- corymbosum* (Ericaceae). *Am. J. Bot.* 85: 698–703. <https://doi.org/10.2307/2446540>
- Rieger, R., A. Michaelis, and M. M. Green, 1968 *A glossary of genetics and cytogenetics: Classical and molecular*, Springer-Verlag, New York. <https://doi.org/10.1007/978-3-662-01012-9>
- Robertson, A., and W. G. Hill, 1984 Deviations from Hardy-Weinberg proportions: sampling variances and use in estimation of inbreeding coefficients. *Genetics* 107: 703–718.
- Rousset, F., 2008 genepop007: a complete re-implementation of the genepop software for Windows and Linux. *Mol. Ecol. Resour.* 8: 103–106. <https://doi.org/10.1111/j.1471-8286.2007.01931.x>
- Schwarz, G., 1978 Estimating the dimension of a model. *Ann. Stat.* 6: 461–464. <https://doi.org/10.1214/aos/1176344136>
- Session, A. M., Y. Uno, T. Kwon, J. A. Chapman, A. Toyoda *et al.*, 2016 Genome evolution in the allotetraploid frog *Xenopus laevis*. *Nature* 538: 336–343. <https://doi.org/10.1038/nature19840>
- Solís Neffa, V. G., and A. Fernández, 2000 Chromosome studies in *Turnera* (Turneraceae). *Genet. Mol. Biol.* 23: 925–930. <https://doi.org/10.1590/S1415-47572000000400037>
- Stebbins, G. L., 1950 *Variation and Evolution in Plants*, Columbia University Press, New York.
- Stift, M., C. Berenos, P. Kuperus, and P. H. van Tienderen, 2008 Segregation models for disomic, tetrasomic and intermediate inheritance in tetraploids: a general procedure applied to rorippa (yellow cress) microsatellite data. *Genetics* 179: 2113–2123. <https://doi.org/10.1534/genetics.107.085027>
- Swaminathan, M. S., and K. Sulbha, 1959 Multivalent frequency and seed fertility in raw and evolved tetraploids of *Brassica campestris* var. *toria*. *Mol. Genet. Genomics* 90: 385–392. <https://doi.org/10.1007/BF00888813>
- Weir, B., and C. C. Cockerham, 1984 Estimating *F*-statistics for the analysis of population structure. *Evolution* 38: 1358–1370. <https://doi.org/10.1111/j.1558-5646.1984.tb05657.x>
- Wu, R. L., 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.
- Xu, S. Z., 2013 *Principles of Statistical Genomics*, Springer New York, New York. <https://doi.org/10.1007/978-0-387-70807-2>
- Zhou, H., T. Y. Ma, R. Zhang, Q. Z. Xu, F. Shen *et al.*, 2016 Analysis of different ploidy and parent-offspring genomic DNA methylation in the loach *Misgurnus anguillicaudatus*. *Int. J. Mol. Sci.* 17: 1299. <https://doi.org/10.3390/ijms17081299>

Communicating editor: Y. Kim