The infinitesimal model for autotetraploid and mixed-ploidy populations

Arthur Zwaenepoel[1,*]

1. University of Lille, CNRS, UMR 8198 – Evo-Eco-Paleo, F-59000 Lille, France
[*]arthur.zwaenepoel@univ-lille.fr

Abstract

We define the infinitesimal model of quantitative genetics (*sensu* Barton *et al.* (2017)) for the inheritance of an additive quantitative trait in mixed-ploidy populations consisting of diploid, triploid and autotetraploid individuals producing haploid and diploid gametes. We implement efficient simulation methods and use these to study the quantitative genetics of mixed-ploidy populations and the establishment of tetraploids after an environmental challenge and in a new habitat.

1 Introduction

2 Model and Methods

We first develop the infinitesimal model (in the sense of Barton *et al.* (2017), i.e. the 'Gaussian descendants' approximation (Turelli 2017)) for an autotetraploid population. We then consider mixed-ploidy populations in which diploids, triploids and tetraploids coexist and interbreed through the production of haploid and diploid gametes that combine freely.

2.1 The infinitesimal model

Consider a population which expresses a quantitative trait determined by a large number of additive loci of small effect. The infinitesimal model approximates the inheritance of such a trait by assuming that the trait value Z_{ij} of a random offspring from parents with trait values z_i and z_j follows a Gaussian distribution with mean equal to the midparent value and variance which is independent of the mean:

$$Z_{ij} \sim \text{Normal}\left(\frac{z_i + z_j}{2}, V_{ij}\right) \tag{1}$$

Here, V_{ij} is referred to as the *segregation variance* in family (i, j). This is the variation generated among offspring from the same parental pair due to random Mendelian segregation in meiosis. This approximation can be justified as arising from the limit where the number of loci determining the trait tends to infinity (Barton *et al.* 2017). An alternative, and for our purposes useful, way to characterize the model in diploids and polyploids is to write

$Z_{ij} = Y_i + Y_j$, where Y_i and Y_j are independent Gaussian random variables $Y_i \sim \text{Normal}\left(\frac{z_i}{2}, V_i\right)$ (and similarly for Y_j). We refer to Y_i as the (random) gametic value of individual i, and to V_i as the *gametic segregation variance* of individual i. This formulation is helpful in that it highlights that Mendelian segregation occurs independently in both parents when gametes are produced, which then combine additively to determine the offspring trait value.

If we assume a population consisting of infinitely many unrelated individuals, expressing a trait with genetic variance V and segregation variance V_0, we find that under random mating the variance in the offspring generation is

$$\begin{aligned} V' &= \mathbb{E}[\text{Var}[Z_{ij}|Z_i, Z_j]] + \text{Var}[\mathbb{E}[Z_{ij}|Z_i, Z_j]] \\ &= \mathbb{E}[V_{ij}] + \text{Var}\left(\frac{Z_i + Z_j}{2}\right) \\ &= V_0 + \frac{V}{2} \ . \end{aligned} \tag{2}$$

So that at equilibrium ($V' = V$), we have $V = 2V_0$ (Barton *et al.* 2017). If we denote by X the contribution to the trait value associated with a randomly sampled haploid genome from the population, we also have that, at equilibrium, $V = 2V_0 = mV_x$, where $V_x = \text{Var}[X]$. Note that no assumptions regarding ploidy level have been made in the definition of the model. Indeed, as long as all individuals are of the same *even* ploidy level, eq. 1 and eq. 2 apply equally well to the inheritance of an additive trait in polyploids as in diploids. Two factors however do lead to differences with the standard diploid model due to their effect on V_i: (1) the potential occurrence of double reduction in polyploids affects the relationship between V_x and V_0 and (2) the evolution of the segregation variance over time in finite populations is affected by the ploidy level.

2.2 The effect of double reduction on the segregation variance

When an autotetraploid forms quadrivalents during prophase I, a form of 'internal inbreeding' may occur as a result of the phenomenon called *double reduction* (Lynch and Walsh 1998 p. 57). Double reduction happens when, as a result of recombination, replicated gene copies on sister chromatids move to the same pole during anaphase I, as illustrated in fig. 1. In the example shown in fig. 1, one of the four generated gametes is AA, which would not occur in the ordinary bivalent meiosis, because in that case, paired chromosomes (involved in cross-overs) are separated during anaphase I. The frequency of double reduction at a locus in the presence of multivalent formation is hence determined by the frequency at which that locus is involved in a cross-over (which depends on the distance to the centromere), and has an upper bound at 1/6 (Stift *et al.* 2008).

When double reduction occurs, an $ABCD$ genotype would generate 10 distinct gametes, as opposed to 6 in when chromosomes form bivalents. As a result, the segregation variance is increased by double reduction. For a random genotype $X_1X_2X_3X_4$, we can find the gametic

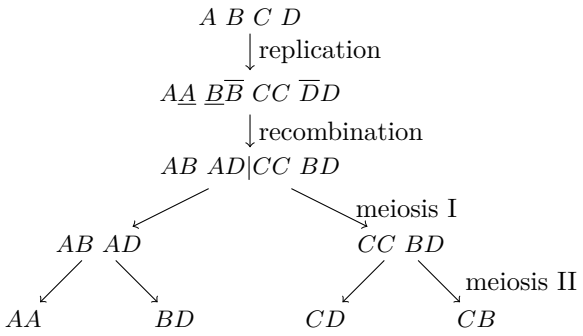

Figure 1: Schematic illustration of a meiotic division in an autotetraploid leading to double reduction at a locus with genotype $ABCD$. Two recombination events are assumed to occur at the locus (denoted by the bars).

segregation variance contributed by a locus when double reduction happens as

$$\begin{aligned}\mathbb{E}[\mathrm{Var}[Y|X_1,X_2,X_3,X_4]] &= \mathrm{Var}[Y] - \mathrm{Var}[\mathbb{E}[Y|X_1,X_2,X_3,X_4]] \\ &= \mathrm{Var}[2X] - \mathrm{Var}\left[\frac{1}{4}(2X_1+2X_2+2X_3+2X_4)\right] \\ &= 4v_x - \frac{1}{4}4v_x = 3v_x\end{aligned}$$

where X denotes the additive effect of a random allele at the locus drawn from the reference population and $v_x = \mathrm{Var}[X]$. In the absence of double reduction we have

$$\mathbb{E}[\mathrm{Var}[Y|X_1,X_2,X_3,X_4]] = 2\mathrm{Var}[X] - \mathrm{Var}\left[\frac{1}{6}\sum_{i=1}^{3}\sum_{j=i+1}^{4}(X_i+X_j)\right] = v_x$$

Assuming that the probability of double reduction at any locus is α (also α_4 below), and summing over independent loci, we find that the gametic segregation variance in the presence of double reduction should be

$$\frac{V_0}{2} = (1-\alpha)V_x + 3\alpha V_x = V_x(1+2\alpha). \tag{3}$$

2.3 Inbreeding in autotetraploids

In a finite population the segregation variance will decrease over time due to inbreeding. Indeed, in a diploid individual for instance, when the two gene copies at a locus are identical, segregation will not generate any variation. The rate of inbreeding depends on the ploidy level (Moody *et al.* 1993; Arnold *et al.* 2012), and hence the rate at which the segregation variance declines *does* so too. Let F_i the inbreeding coefficient of individual i, defined as the probability that two *distinct* gene copies at a locus in individual i are identical by descent (IBD) with respect to some reference population, and let Φ_{ij} be the coancestry coefficient for individuals i and j, defined as the probability that two genes sampled independently from i and j at a given locus are IBD. In diploids, one can show that the gametic segregation variance V_i is decreased by a factor $1-F_i$ relative to the segregation variance in the reference

population $V_0/2$ (Barton *et al.* 2017). In polyploids, however, F_i does not suffice to describe the state of homozygosity in individual i. In tetraploids, for instance, we have five distinct homozygosity states, which we can symbolically represent as $abcd, aabc, aabb, aaab$ and $aaaa$ (in general, the number of homozygosity states grows according to the partition function $(1, 2, 3, 5, 7, 11, 15, 22, \dots)$). Representing the probability of being in these five increasingly homozygous states as $\delta_1, \dots, \delta_5$, we find that the gametic segregation variance is reduced by a factor

$$\phi = \delta_1 + \left(1 - \frac{1}{6}\right)\delta_2 + \left(1 - \frac{1}{3}\right)\delta_3 + \left(1 - \frac{1}{2}\right)\delta_4$$

which is precisely $1 - F_i$, as in diploids (see also Moody *et al.* (1993)). This shows that we do not need to track the array of homozygosity coefficients in order to compute the segregation variance in a tetraploid family, but only require the inbreeding coefficients of the parents. This is a consequence of the fact that, in tetraploids, gametes are diploid. For higher ploidy levels, one would need to track higher order identity coefficients. We shall hence not consider higher ploidy levels in this paper.

In order to efficiently conduct individual-based simulations under the infinitesimal model, we hence need to be able to track the F_i and Φ_{ij}. Denoting the parents of individual i by k and l, the recursion for the inbreeding coefficients is

$$\begin{aligned} F_i &= \Phi_{kl} && \text{in diploids} \\ F_i &= \frac{1}{6}(F_k^* + F_l^* + 4\Phi_{kl}) && \text{in tetraploids} \end{aligned} \tag{4}$$

where $F_k^* = \alpha + (1 - \alpha)F_k$. The recursion for the tetraploid case follows from considering three cases: either (1) the two genes sampled in individual i both came from the gamete contributed by parent k, which happens with probability 1/6, in which case they are IBD with probability F_k^*; or (2) as in (1) but from parent l; or (3) with probability 2/3 the two genes came from different gametes, in which case they are IBD with probability Φ_{kl}. The recursion for the coancestry coefficients in m-ploids is given by

$$\begin{aligned} \Phi_{ii} &= \frac{1}{m}\left(1 + (m-1)F_i\right) \\ \Phi_{ij} &= \sum_k \sum_l P_{ik} P_{jl} \Phi_{kl} && i \neq j \end{aligned} \tag{5}$$

where the sums are over individuals in the parental population, and where $P_{ik} \in \{0, \frac{1}{2}, 1\}$ is the probability that a gene copy in i is derived from parent k. When dealing with discrete generations, P_{ik} values can be conveniently represented in a $N(t) \times N(t-1)$ matrix, where $N(t)$ is the population size in generation t, so that $\Phi(t) = P\Phi(t-1)P^T$, where $\Phi(t)$ is the matrix of coancestry coefficients in generation t (Barton *et al.* 2017).

2.4 The infinitesimal model in mixed-ploidy populations

When considering mixed-ploidy populations, we need to consider how equilibrium variances in the model scale with ploidy level, and how offspring of different ploidy levels are derived from a parental pair. Throughout the following paragraphs, we specialize to a diploid-triploid-tetraploid complex. A glossary of the relevant notation can be found in tbl. 2.

2.4.1 Scaling of the genetic variance across ploidy levels

The equilibrium phenotypic variance V under the infinitesimal model was derived above as $2V_0$, where V_0 is the segregation variance (eq. 2). As noted below eq. 2, this entails that for m-ploids, $2V_0 = mV_{x,m}$, where $V_{x,m}$ is the variance of the additive effect of a randomly sampled haploid genome from the reference m-ploid population. Now, in a mixed-ploidy system, it is not immediately clear how the $V_{x,m}$ should scale across ploidy levels, as this depends on the effects of polyploidization *per se* on additive allelic effects (Porturas *et al.* 2019; Clo and Kolář 2021; Clo 2022). If we assume equal equilibrium variances (and concomitantly segregation variances), the genetic variance for a haploid genome in tetraploids will be half that of diploids, entailing a reduction in the additive allelic effect per locus as ploidy level rises. On the other hand, if we assume equal allelic effects, and hence that the genetic variance associated with a hypothetical haploid genome copy is identical across ploidy levels, then the segregation and equilibrium variances in tetraploids will be twice those of diploids, and the associated phenotypic range will be doubled as well.

To ease interpretation, let us consider an underlying Mendelian system, consisting of n unlinked additive bi-allelic loci in Hardy-Weinberg and linkage equilibrium (HWLE). Assuming that the additive effects in tetraploids are homogeneously scaled by a factor β relative to those in diploids, and writing $V_{z,m}$, $V_{0,m}$ and $V_{x,m}$ for respectively the equilibrium genetic variance, the segregation variance and the genetic variance associated with a haploid genome in m-ploids, the following relationships hold:

$$\frac{V_{z,4}}{V_{z,2}} = \frac{2V_{0,4}}{2V_{0,2}} = \frac{4V_{x,4}}{2V_{x,2}} = \frac{2\sum_i^n (\beta a_i)^2 p_i(1-p_i)}{\sum_i^n a_i^2 p_i(1-p_i)} = 2\beta^2$$

So we see that assuming equal equilibrium phenotypic variances $V_{z,4}/V_{z,2} = 1$ entails that allelic effects are scaled by a factor $\beta = \sqrt{1/2}$. On the other hand, assuming equal allelic effects ($\beta = 1$) entails that the phenotypic variance at equilibrium in tetraploids is twice that of diploids, as we noted above. To study the effect of such assumptions, we introduce β_m as a parameter so that $V_{z,m} = \beta_m^2 V_{z,2}$, keeping in mind its interpretation as a scaler of allelic effects in the underlying Mendelian system, and setting $\beta_2 = 1$. The trait value distribution for offspring of ploidy level m_{ij} from parental pair (i, j) with ploidy levels m_i and m_j, formed through the union of a g_i-ploid gamete from parent i and a g_j-ploid gamete from parent j will then be

$$Z_{ij} \sim \text{Normal}\left(\beta_{m_{ij}}\left(\frac{g_i}{m_i}\frac{z_i}{\beta_{m_i}} + \frac{g_j}{m_j}\frac{z_j}{\beta_{m_j}}\right), V_{i,g_i} + V_{j,g_j}\right)$$

Where V_{i,g_i} is the gametic segregation variance associated with the formation of a g_i-ploid gamete by individual i. We now proceed to characterize the V_{i,g_i} and their evolution over time.

2.4.2 Unreduced gamete formation in diploids

The mechanisms of unreduced gamete formation do not necessarily lead to a faithful transmission of the complete diploid genome. Unreduced gametes are formed in two ways, depending on the meiotic abberation that leads to their origin: (1) first division restitution (FDR) of (2) second division restitution (SDR) (Bretagnolle and Thompson 1995). Consider a locus in a

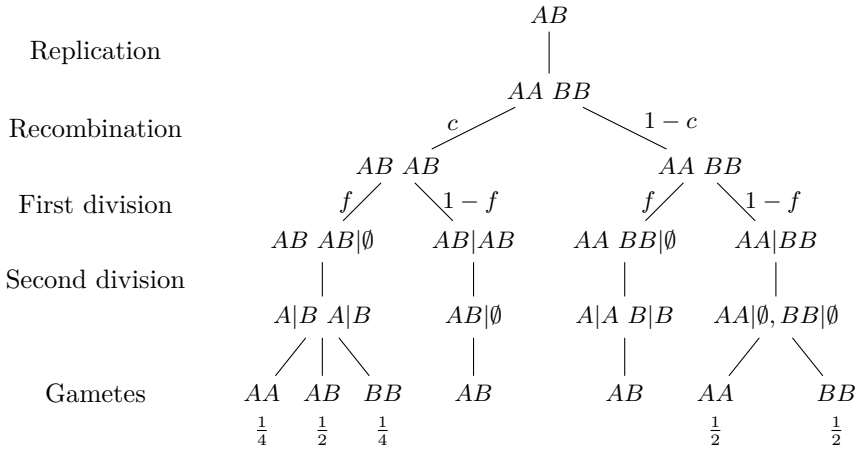

Figure 2: Schematic representation of the different pathways for unreduced gamete formation in diploids and their different outcomes.

diploid with two distinct genes A and a. Assume recombination happens with probability c and that conditional on unreduced gamete formation, formation is due to FDR with probability f while it is due to SDR with probability $1-f$. The different unreduced gametes that are formed are represented schematically in fig. 2. Writing the genotype at a locus in the diploid parent as X_1X_2, with allelic effects X_1 and X_2, the genotypic value of an unreduced gamete at this locus will be

$$Y = \begin{cases} 2X_1 & \text{w.p. } \alpha_2/2 = \frac{1}{4}fc + \frac{1}{2}(1-f)(1-c) \\ 2X_2 & \text{w.p. } \alpha_2/2 \\ X_1 + X_2 & \text{w.p. } 1-\alpha_2 \end{cases}$$

where $\alpha_2 = 1 - f - c + \frac{3}{2}cf$ is the probability that two copies of the same gene end up in a diploid gamete produced by a diploid individual. Conditional on the latter event, we get, by a similar reasoning as the one leading eq. 3, a segregation variance equal to

$$\begin{aligned} \mathbb{E}[\text{Var}[Y|X_1, X_2]] &= \text{Var}[Y] - \text{Var}[\mathbb{E}[Y|X_1, X_2]] \\ &= \text{Var}[2X] - \text{Var}\left[\frac{1}{2}(2X_1 + 2X_2)\right] \\ &= 2v_x \end{aligned}$$

Conditioning on the complementary event, all gametes have genetic value $Y = X_1 + X_2$, so that the segregation variance is 0. Summing across independent loci, we have $V_{i,2} = 2\alpha_2(1-F_i)V_x$ for a diploid individual i.

2.4.3 Meiosis in triploids

Triploids, when viable, may be important for the dynamics of mixed-ploidy populations due to the formation of a so-called triploid bridge. The formation of triploids presents no immediate issues, we simply need to track the segregation variance contributions from both donor gametes, and relate these to $V_{0,3}$. Sexual reproduction in triploids is however more

complicated. There are no known mechanisms to coordinate the assortment of chromosomes in for instance a haploid and diploid gamete, and meiosis, if it happens, usually results in aneuploid gametes (Ramsey and Schemske 1998). Experimental results indicate that, at least in yeast, triploids usually form trivalents and undergo recombination, after which each trivalent is randomly assorted in the daughter cells, some receiving one, others two copies of a given chromosome (Charles *et al.* 2010). In the absence of gametic nonreduction, the probability of obtaining euploid gametes (two diploid and two haploid gametes) from such a process is $(1/2)^n$, where n is the number of chromosomes. If the number of chromosomes is small this is not negligible, for instance in *A. thaliana* we would have $(1/2)^5 \approx 0.03$, which is of the same order as the unreduced gamete formation rate. Unreduced (triploid) gametes may also be produced and important for the dynamics of mixed-ploidy populations (Ramsey and Schemske 1998). However, they generate additional difficulty, since in order to compute the contributed variance under inbreeding, we would need an additional identity coefficient recording the probability that three genes are IBD at a locus. We will hence ignore the possibility of unreduced gamete production in triploids. Denoting by α_3 the probability that a diploid gamete produced by a triploid individual contains two copies of the same parental gene, we can use the same approach as above to derive $V_{i,1}$ and $V_{i,2}$ for a triploid individual i. We note that, on the supposition that diploid gametes are produced by random assortment of chromosomes in a haploid and diploid gametes, $\alpha_3 \leq 1/4$. The formulae for the different V_{i,g_i} are summarized in tbl. 1.

Table 1: Gametic segregation variance for different cytotype × gametic cytotype combinations. V_x refers to genetic variance associated with a haploid genome in the reference population of the cytotype to which the gamete will contribute. Under the assumptions outlined in the main text, we have the following constraints for the α_k: $\alpha_2 \leq 1$, $\alpha_3 \leq 1/4$ and $\alpha_4 \leq 1/6$.

cytotype	$V_{i,1}$ (haploid)	$V_{i,2}$ (diploid)
diploid ($m_i = 2$)	$\frac{1}{2}(1 - F_i)V_x$	$2\alpha_2(1 - F_i)V_x$
triploid ($m_i = 3$)	$\frac{2}{3}(1 - F_i)V_x$	$\frac{2}{3}(1 + 3\alpha_3)(1 - F_i)V_x$
tetraploid ($m_i = 4$)	–	$(1 + 2\alpha_4)(1 - F_i)V_x$

2.4.4 Recursions for inbreeding coefficients

In a mixed-ploidy system tracking inbreeding coefficients becomes only slightly more complicated. Recall that α_m is the probability that a diploid gamete from an m-ploid individual transmits two copies of the same gene at some locus. Let m_i denote the ploidy level of individual i. We still have the recursion shown in eq. 4, where now

$$F_k^* = F_k(1 - \alpha_{m_k}) + \alpha_{m_k}$$

For a triploid individual, where the parent contributing the $2n$ gamete is k, we have

$$F_i = \frac{1}{3}\left(F_k^* + 2\Phi_{kl}\right)$$

For Φ the recursion in eq. 5 remains valid, but where diagonal elements are now given by

$$\Phi_{ii} = \frac{1}{m_i}\left(1 + (m_i - 1)F_i\right)$$

and where the elements of the the pedigree matrix P now have entries in $\{0, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, 1\}$.

2.4.5 Population dynamics

What comes here depends on the results we will include.

The dynamics of the mixed-ploidy system (in the absence of selection on the trait) are determined by the rates at which different types of gametes are formed. We assume an individual of ploidy level k forms haploid and diploid gametes with proportions u_{k1} and u_{k2}, so that the (relative) fecundity of a k-ploid individual is $u_{k1} + u_{k2}$. Unless stated otherwise, we will assume

$$U = \begin{pmatrix} u_{21} & u_{22} \\ u_{31} & u_{32} \\ u_{41} & u_{42} \end{pmatrix} = \begin{pmatrix} 1-u & u \\ v & v \\ 0 & 1-u \end{pmatrix}$$

where u is referred to as the proportion of unreduced gametes, and $2v$ is the proportion of euploid gametes produced by a triploid individual. An analysis of a deterministic model for the cytotype dynamics (also analyzed in Felber 1991; Felber and Bever 1997) is included in sec. S2.1.

Table 2: Glossary for the notation used to define the mixed-ploidy infinitesimal model.

Symbol	Explanation
z_i	Trait value of individual i
Z_{ij}	Trait value of a random offspring from parental pair (i, j)
$V_{z,m}$	Equilibrium variance in an m-ploid population
$V_{0,m}$	Segregation variance in the m-ploid reference population
$V_{x,m}$	Equilibrium variance of a haploid genome in the m-ploid reference population
$V_{i,m}$	Gametic segregation variance in individual i
F_i	Inbreeding coefficient of individual i
Φ_{ij}	Coancestry coefficient of individuals i and j
m_i	Ploidy level of individual i
α_m	Probability that a diploid gamete from an m-ploid individual contains two copies of the same parental gene
β_m	Scaling factor for the allelic effects in ploidy level m
$u_{m,g}$	Proportion of g-ploid gametes produced by an m-ploid individual
u	Proportion of unreduced gametes in even-ploids
v	Proportion of haploid and diploid gametes in triploids

3 Results

We first verify that, in the absence of inbreeding, our implementation of the mixed-ploidy infinitesimal model yields the predicted equilibrium variance for each cytotype (fig. S1) and the expected equilibrium proportions (fig. S2). We next evaluate the accuracy of the infinitesimal model as an approximation to the evolution of a quantitative trait determined by L additive loci. We find that the infinitesimal model with inbreeding generally yields accurate predictions for the evolution of the genetic variance, even as the number of loci becomes fairly small (e.g. $L = 50$, fig. S3). For $L = 20$, substantial deviations from the infinitesimal approximation can be observed, although it may still be considered reasonable (fig. S3).

Although inbreeding is slower in autotetraploids than in diploids for the same population size, the tetraploid fraction of a diploid-dominated mixed-ploidy population will have an equal or higher average inbreeding coefficient (fig. S5). This is because in such a population, triploid and tetraploid individuals mostly arise from gametes formed by diploid individuals or polyploid individuals with very recent diploid ancestry (about $1 + u + 2v$ generations ago for tetraploids, and $1 + \frac{2}{3}(u + 2v)$ generations ago for triploids, see sec. S2.2), so that the polyploid subpopulations will show an average relatedness similar to the diploid population and not evolve as an isolated higher-ploidy population would. A nonzero probability of producing IBD diploid gametes (α) will then further increase the inbreeding coefficient in the tetraploid and triploid fraction of the population relative to their diploid progenitors (fig. S5). Therefore, as long as diploids dominate, the effect of harboring some of the gene pool in triploid and tetraploid individuals on the rate of inbreeding is negligible, and we find that the evolution of the inbreeding coefficient over time is well predicted by $1 - e^{-t/2N_e}$, where the inbreeding-effective population size is, to first order in u, given by $(1 - 2u)N$ (sec. S2.3, fig. S8).

4 Discussion

- It appears that many papers on diploid – tetraploid contact zones or mixed-ploidy populations fail to find significant numbers of triploids. Note though that such studies focus on established tetraploids, where the cytotypes may have diverged to some extent, whereas we focus on the case where a tetraploid cytotype has not yet established. The model is therefore more relevant for populations that are likely to be referred to as diploid. It remains to be seen to what extent such populations actually exhibit a mixed-ploidy equilibrium. See Suda and Herben (2013) and Herben *et al.* (2016).
- The issue with dominance. On the one hand, it is often said that polyploids accumulate a larger mutation load, and will hence show inbreeding depression, but this really only applies to already established polyploid populations. Similarly, the sort of effects of dominancestudied in Booker and Schrider during range expansions apply mostly to polyploids that have accumulated a large mutation load over a rather long time as a stable polyploid population. In our case, polyploids are virtually always derived very recently from diploid ancestors, so they will not have accumulated more deleterious recessives than their diploid counterparts. As inbreeding will be slower in tetraploids, we should hence have reduced inbreeding expression, as we have the same load as in diploids, but it is less expressed (at least when considering recessives). There should therefore be an immediate benefit to polyploidy (reduced inbreeding depression), which is however a long term disadvantagem because eventually the established polyploid will start to accumulate a higher load due to less efficient purging.

5 References

Arnold B., K. Bomblies, and J. Wakeley, 2012 Extending coalescent theory to autotetraploids. Genetics 192: 195–204.

Barton N. H., A. M. Etheridge, and A. Véber, 2017 The infinitesimal model: Definition, derivation, and implications. Theoretical population biology 118: 50–73.

Bretagnolle F., and J. D. and Thompson, 1995 Gametes with the somatic chromosome number: Mechanisms of their formation and role in the evolution of autopolyploid plants. New Phytologist 129: 1–22.

Charles J. S., M. L. Hamilton, and T. D. Petes, 2010 Meiotic chromosome segregation in triploid strains of saccharomyces cerevisiae. Genetics 186: 537–550.

Clo J., and F. Kolář, 2021 Short-and long-term consequences of genome doubling: A meta-analysis. American Journal of Botany 108: 2315–2322.

Clo J., 2022 The evolution of the additive variance of a trait under stabilizing selection after autopolyploidization. Journal of Evolutionary Biology 35: 891–897.

Felber F., 1991 Establishment of a tetraploid cytotype in a diploid population: Effect of relative fitness of the cytotypes. Journal of evolutionary biology 4: 195–207.

Felber F., and J. D. Bever, 1997 Effect of triploid fitness on the coexistence of diploids and tetraploids. Biological Journal of the Linnean Society 60: 95–106.

Herben T., P. Trávniček, and J. Chrtek, 2016 Reduced and unreduced gametes combine almost freely in a multiploidy system. Perspectives in Plant Ecology, Evolution and Systematics 18: 15–22.

Lynch M., and B. Walsh, 1998 *Genetics and analysis of quantitative traits*. Sinauer Sunderland, MA.

Moody M. E., L. Mueller, and D. Soltis, 1993 Genetic variation and random drift in autotetraploid populations. Genetics 134: 649–657.

Porturas L. D., T. J. Anneberg, A. E. Curé, S. Wang, D. M. Althoff, *et al.*, 2019 A meta-analysis of whole genome duplication and the effects on flowering traits in plants. American Journal of Botany 106: 469–476.

Ramsey J., and D. W. Schemske, 1998 Pathways, mechanisms, and rates of polyploid formation in flowering plants. Annual review of ecology and systematics 29: 467–501.

Rousset F., 2004 *Genetic structure and selection in subdivided populations*. Princeton University Press.

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.

Suda J., and T. Herben, 2013 Ploidy frequencies in plants with ploidy heterogeneity: Fitting a general gametic model to empirical population data. Proceedings of the Royal Society B: Biological Sciences 280: 20122387.

Turelli M., 2017 Commentary: Fisher's infinitesimal model: A story for the ages. Theoretical population biology 118: 46–49.

S1 Supplementary figures

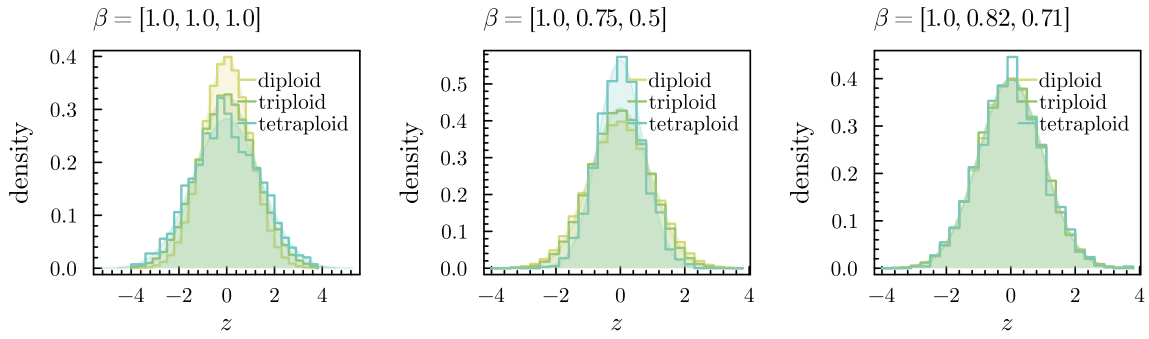

Figure S1: Equilibrium trait value (z) distribution in a mixed-ploidy population evolving according to the infinitesimal model without inbreeding for three different scaling assumptions of allelic effects across ploidy levels. Histograms (lines) show simulation results, whereas the filled Gaussian densities show the expected distribution at equilibrium Normal $\left(0, k\beta_k^2 V_x\right)$. In the $\beta = (1, 1, 1)$ case (left), allelic effects remain constant across ploidy levels, causing the equilibrium variance to be doubled in tetraploids relative to diploids. In the $\beta = (1, 3/4, 1/2)$ case (middle), allelic effects are scaled so that the phenotypic range is held constant (i.e. the maximum and mimimum phenotypes are teh same across ploidy levels). In the $\beta = (1, \sqrt{2/3}, \sqrt{1/2})$ case (right), allelic effects are scaled so that the variance in trait values at equilibrium is equal across ploidy levels.

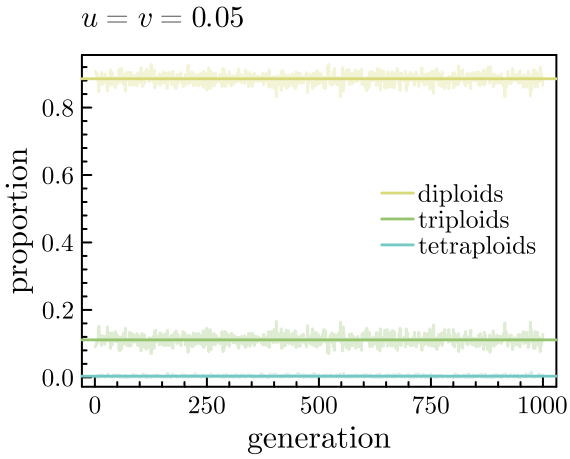

Figure S2: Equilibrium proportions of the different cytotypes in a mixed-ploidy population where $u = v = 0.05$, light lines show the observed proportions in a simulation of a Wright-Fisher mixed-ploidy infinitesimal model with $N = 500$. The horizontal lines show the deterministic predictions $(\tilde{g}_1^2, 2\tilde{g}_1\tilde{g}_2, \tilde{g}_2^2)$.

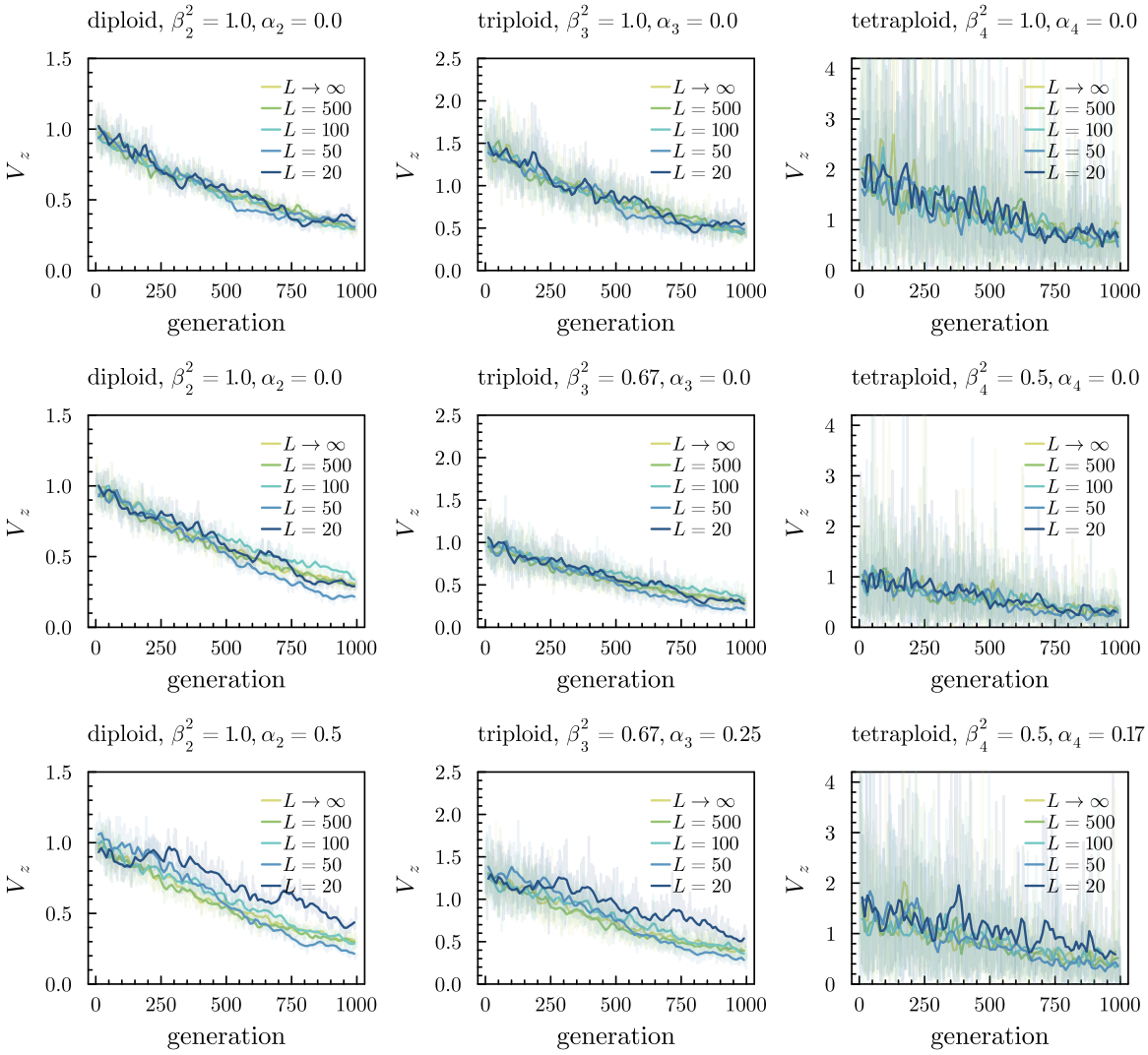

Figure S3: Comparison of the mixed-ploidy infinitesimal model with the L-locus model, for $L = 500, 100, 50$ and 20. The decline in the genetic variance V_z within each cytotype due to drift is shown. The transparent lines show the complete simulation, whereas the solid line shows the same data but smoothed in overlapping windows of 20 generations. We assume $N = 500, u = v = 0.08$ and no selection. In the top row where $\beta_k^2 = 1, \alpha_k = 0$, the equilibrium variance in the absence of inbreeding in triploids is 2/3 that of diploids, and in tetraploids it is twice that in diploids. In the middle row, $\beta_3^2 = 2/3$ and $\beta_4^2 = 1/2$, so that the equilibrium variance in the absence of inbreeding is equal across cytotypes. In the bottom row, $\alpha_2 = \alpha_3 = 1/2$ and $\alpha_4 = 1/6$, causing an immediate increase in the genetic variance in higher cytotypes, but also accelerated inbreeding.

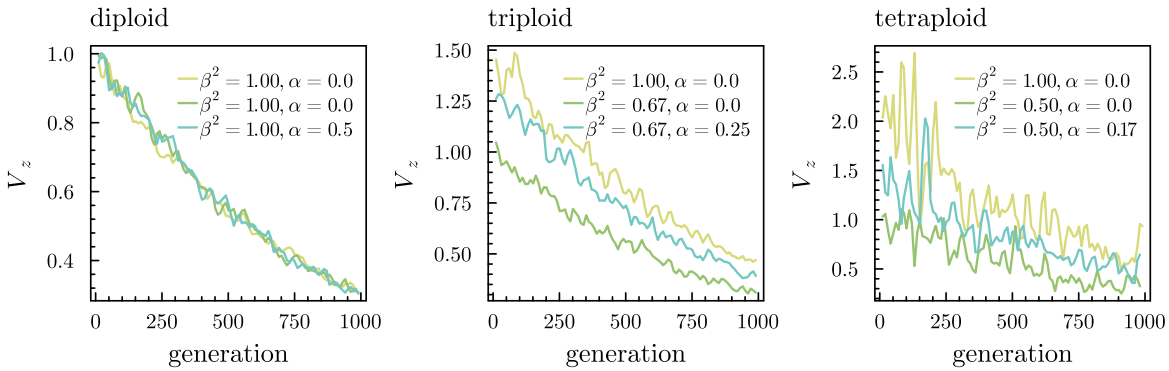

Figure S4: As in fig. S3 ($L \to \infty$ simulations) but highlighting the effect of β and α within ploidy levels. Only smoothed data is shown.

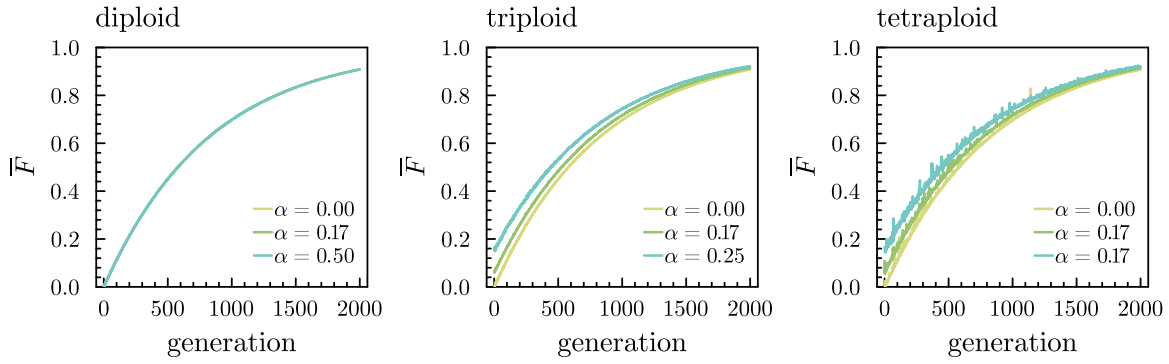

Figure S5: Average inbreeding coefficient $\bar{F}$ in each cytotype in a mixed-ploidy population for different values of α (we assume $\alpha_k = \alpha$, where α_k is the probability that a diploid gamete produced by a k-ploid cytotype contains two copies of the same parental gene at a random locus). We assume $N = 500, u = v = 0.08$ and no selection.

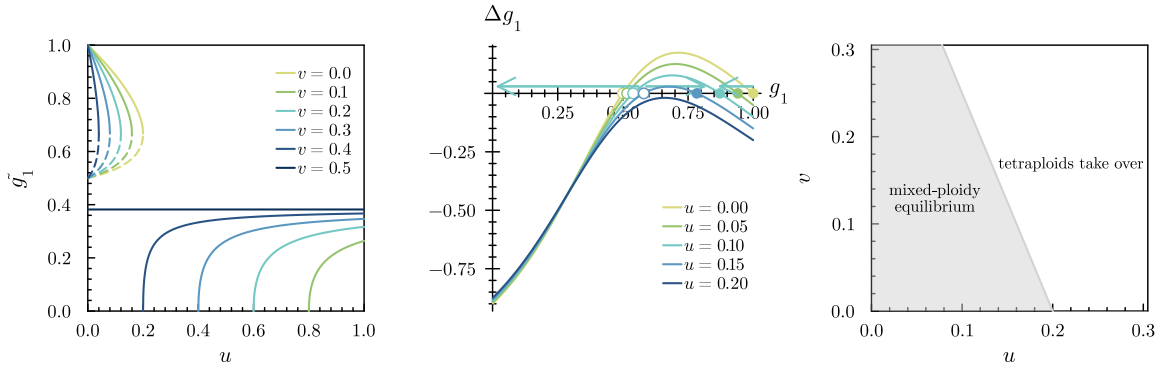

Figure S6: Deterministic mixed-ploidy equilibrium. The left plot shows the stable (solid lines) and unstable (dashed lines) equilibria for the proportion of haploid gametes in the gamete pool g_1 as a function of u for different values of v. The middle plot shows the relationship between $\Delta g_1 = g_1' - g_1$ and g_1. The zeros of this graph are the fixed points of the dynamical system and are indicated by the hollow (unstable equilibrium) and solid (stable equilibrium) dots. The rightmost plot shows the region of parameter space where a stable mixed-ploidy equilibrium exist.

S2 Appendix

S2.1 Deterministic mixed-ploidy model

Let g_k be the frequency of k-ploid gametes in the gamete pool, and let us consider only haploid and diploid gametes, so that $g_2 = 1 - g_1$. Diploids produce unreduced gametes with probability u and reduced ones with probability $1 - u$, triploids produce haploid and triploid gametes both with probability v, and tetraploids produce reduced diploid gametes with probability $(1 - u)$ (we assume they produce, just like diploids, a proportion u of unreduced gametes, but these are assumed not to lead to viable offspring and are ignored). We get after one generation of random mating

$$g_1' = \frac{(1-u)g_1^2 + 2vg_1g_2}{g_1^2 + 4vg_1g_2 + (1-u)g_2^2}.$$

We see that $g_1 = 0$ is always an equilibrium (no haploid gametes, tetraploids take over). Two more fixed points are obtained at

$$\tilde{g_1}, \tilde{g_1}' = \frac{3 - 3u - 6v \pm \sqrt{(u + 2v - 1)(5u + 2v - 1)}}{2(2 - u - 4v)} \tag{S1}$$

Of which the larger one, when it exists, corresponds to a stable equilibrium, and the smaller one to an unstable equilibrium. As there are no viability differences, the equilibrium cytotype frequencies can be readily obtained from these through the relations

$$\pi_2 = \tilde{g_1}^2 \qquad \pi_3 = 2\tilde{g_1}\tilde{g_2} \qquad \pi_4 = \tilde{g_2}^2 \tag{S2}$$

Assuming $v = O(u)$, we have to second order in u

$$\begin{aligned} \pi_2 &= 1 - 2u - 4uv - u^2 + O(u^3) \\ \pi_3 &= 2u + 4uv + O(u^3) \\ \pi_4 &= u^2 + O(u^3) \end{aligned} \tag{S3}$$

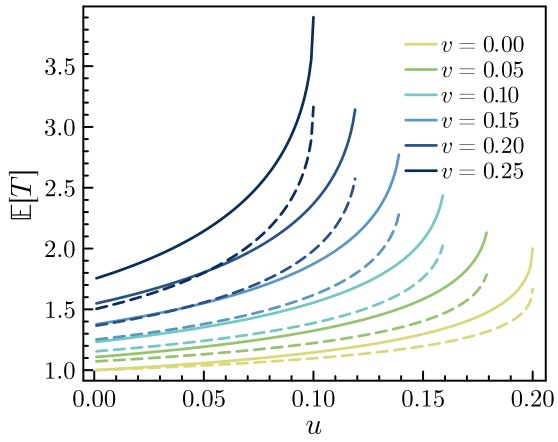

Figure S7: Expected time to diploid ancestry. The solid lines show $\mathbb{E}[T_4]$, i.e. the expected time since being inherited from a diploid ancestor for a random gene in a tetraploid individual at equilibrium, for different values of v (half the triploid fertility). The dashed lines show $\mathbb{E}[T_3]$, i.e. the same quantity for a gene sampled from a triploid. Note that $\mathbb{E}[T]$ blows up whenever u and v exceed their critical value for tetraploid establishment.

At the critical point where the stable equilibrium disappears, we have that $\Delta g_1 = \frac{d\Delta g_1}{dg_1} = 0$ (fig. S6, middle). We find that, in the region of parameter space that is biologically relevant (roughly $u < 0.1, v < 0.1$, say), the critical unreduced gamete formation rate u_c beyond which tetraploids take over can be expressed as a linear function of triploid fertility ($2v$):

$$u_c = \frac{1}{5}(1 - 2v)$$

(fig. S6, right). This shows that, for plausible parameter values, we can safely assume that an initially diploid population will evolve to a mixed-ploidy equilibrium.

S2.2 Expected time to diploid ancestry

Consider a gene sampled from a tetraploid individual in a mixed-ploidy population at equilibrium and not subjected to selection. Let T_4 denote the number of generations in the past until such a gene is found in a diploid ancestor, and let T_3 denote a similar random variable for a randomly sampled gene from a triploid in the same population. Assuming the different cytotypes are at their deterministic equilibrium frequencies π_2, π_3 and π_4 (see sec. S2.1, eq. S2), we have the recursive relations

$$\begin{aligned}
\mathbb{E}[T_4] &= \frac{1}{Z_2}\Big(\pi_2 u + (1 + \mathbb{E}[T_3])\pi_3 v + (1 + \mathbb{E}[T_4])\pi_4(1 - u)\Big) \\
\mathbb{E}[T_3] &= \frac{1}{3Z_1}\Big(\pi_2(1 - u) + (1 + \mathbb{E}[T_3])\pi_3 v\Big) \\
&\quad + \frac{2}{3Z_2}\Big(\pi_2 u + (1 + \mathbb{E}[T_3])\pi_3 v + (1 + \mathbb{E}[T_4])\pi_4(1 - u)\Big)
\end{aligned} \tag{S4}$$

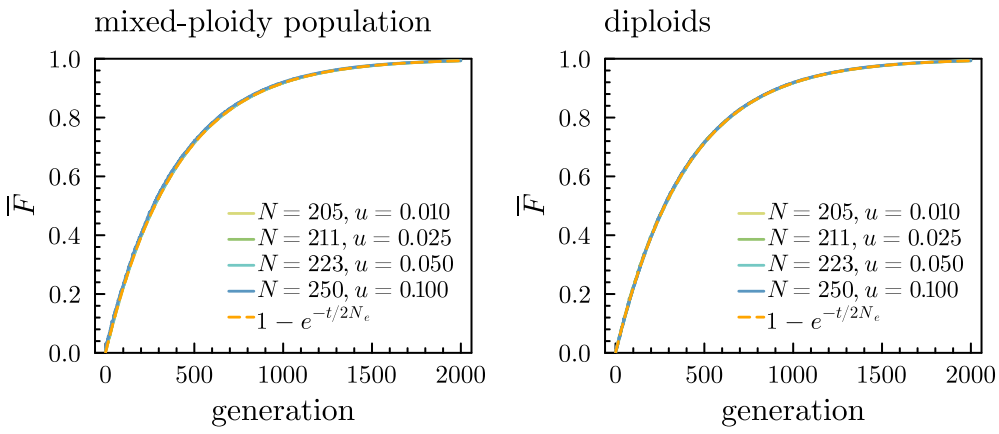

Figure S8: The evolution of $\bar{F}$ in the mixed-ploidy population and in the diploid subpopulation are shown for different values of u and associated values of N, keeping $N_e = (1 - 2u)N$ constant at 200. We assume $u = v$. All lines coincide almost completely and are indistinguishable from $1 - e^{-t/2N_e}$. Results are shown for $\alpha_k = 1/6$. As α_k decreases to 0, $\bar{F}$ in the mixed-ploidy population becomes completely indistinguishable from $\bar{F}$ in the diploid subpopulation.

where

$$Z_1 = \pi_2(1-u) + \pi_3 v$$
$$Z_2 = \pi_2 u + \pi_3 v + \pi_4(1-u)$$

(these expressions are straightforwardly modified when more general u_{ij} are assumed, see e.g. sec. S2.3, eq. S5). The system in eq. S4 can be solved to yield expressions for $\mathbb{E}[T_4]$ and $\mathbb{E}[T_3]$, which are however rather unwieldy. Again assuming $v = O(u)$, we obtain to first order in u

$$\mathbb{E}[T_4] = 1 + u + 2v + O(u^2)$$
$$\mathbb{E}[T_3] = 1 + \frac{2}{3}(u + 2v) + O(u^2)$$

Numerical examples are shown in (fig. S7). Clearly, for plausible parameter values, $\mathbb{E}[T]$ will be very close to 1. For instance, for $u = 0.05$ and $v = 0.05$ (which are already rather large values for these parameters), we would have $\mathbb{E}[T_3] \approx 1.13$ and $\mathbb{E}[T_4] \approx 1.19$.

S2.3 Effective population size of a mixed-ploidy deme

We use the approach outlined in Rousset (2004) (pp. 153, 157) to determine the effective size of a randomly mating mixed-ploidy population. Denote by $\nu_k(t)$ the probability that the ancestral lineage of a given gene in the present is found in a individual of ploidy level k t generations in the past, and let $\nu(t) = (\nu_2(t)\ \nu_3(t)\ \nu_4(t))$ be the corresponding row vector. Assuming the population is at cytotype equilibrium (eq. S2), we have

$$\begin{aligned}\nu(t+1) &= \nu(t)P \\ &= \nu(t)\begin{pmatrix} \frac{u_{21}}{Z_1}\pi_2 & \frac{u_{31}}{Z_1}\pi_3 & 0 \\ \left(\frac{u_{21}}{3Z_1} + \frac{2u_{22}}{3Z_2}\right)\pi_2 & \left(\frac{u31}{3Z_1} + \frac{2u_{32}}{3Z_2}\right)\pi_3 & \frac{2u_{42}}{3Z_2}\pi_4 \\ \frac{u_{22}}{Z_2}\pi_2 & \frac{u_{32}}{Z_2}\pi_3 & \frac{u_{42}}{Z_2}\pi_4 \end{pmatrix}\end{aligned} \tag{S5}$$

where we assume, as usual, that tetraploids do not produce haploid gametes ($u_{41} = 0$), and where

$$Z_1 = u_{21}\pi_2 + u_{31}\pi_3$$
$$Z_2 = u_{22}\pi_2 + u_{32}\pi_3 + u_{42}\pi_4$$

At stationarity, $\lim_{t\to\infty} \nu(t) = \nu$, and we have $\nu = \nu P$. Hence, the probability that the ancestral lineage of a given gene in the present is found in an individual of ploidy level k in an indefinite past is given by ν_k, where ν is the left eigenvector of P associated with the unit eigenvalue. The effective size of a mixed-ploidy population of size N can then be obtained as

$$N_e = N\left(\sum_k \frac{\nu_k^2}{\pi_k}\right)^{-1}$$

After plugging in π in accordance with eq. S2 and solving the eigenvalue problem, this yields an unwieldy expression in the u_{ij}. For our usual parameterization where $u_{21} = u_{42} = 1-u, u_{22} = u$ and $u_{31} = u_{32} = v$, and $v = O(u)$, we can find that

$$\begin{pmatrix}\nu_2\\ \nu_3\\ \nu_4\end{pmatrix} = \begin{pmatrix}1 - 2uv + O(u^3)\\ 2uv + O(u^3)\\ O(u^3)\end{pmatrix}$$

and

$$\frac{N_e}{N} = 1 - 2u + O(u^2)$$

which yields an excellent fit in simulations for plausible parameter values (fig. S8). When $v = 0$ and $u < u_c$ (see sec. S2.1), $N_e = \pi_2 N$, as in that case (i.e. when triploids are infertile) there can be no gene flow from tetraploids to diploids. Since we assume the cytotype composition to be constant, and polyploids are continually formed from diploids, no gene in a triploid or tetraploid will have any descendants in the distant future in this case, so that the effective size is just the diploid fraction of the population.