Supplement: Autopolyploid establishment through polygenic adaptation

Arthur Zwaenepoel*[1]

[1]Department of Biology, University of Antwerp, 2020 Antwerp, Belgium

Contents

*arthur.zwaenepoel@uantwerpen.be

S1 Supplementary Figures

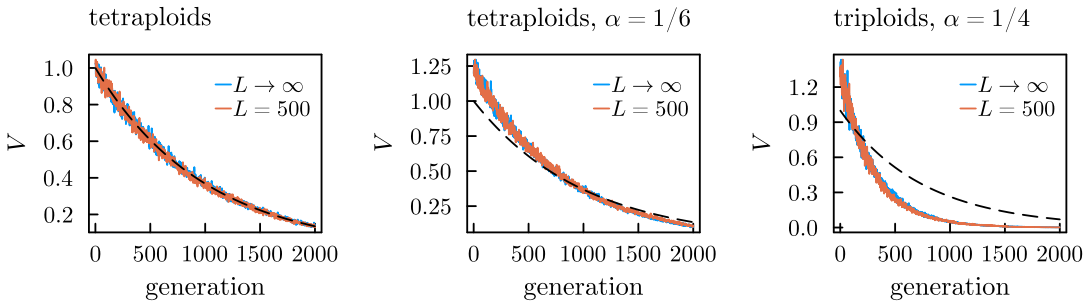

Figure S1: Validation of the autotetraploid and triploid infinitesimal model by a comparison against the discrete locus model with $L = 500$ additive loci, showing the average phenotypic variance in each generation averaged over 10 replicate simulations for both models. We assume the initial phenotypic variance to be one for all simulations, and all replicates are initialized randomly in accordance with this initial phenotypic variance. Allelic effects for the discrete locus model are sampled from a Gaussian with mean 0 and variance $1/2L$.

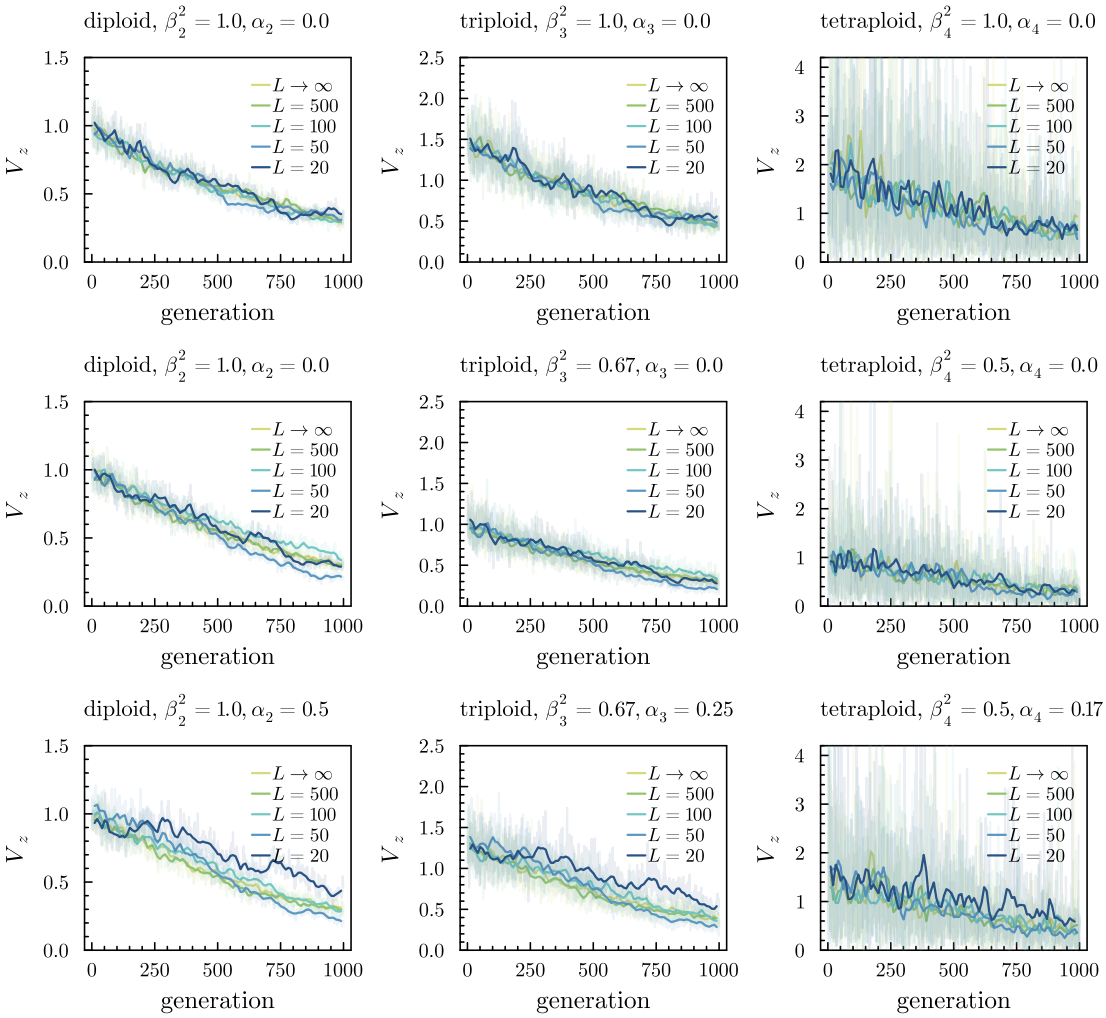

Figure S2: Comparison of the mixed-ploidy infinitesimal model with the L-locus model, for $L = 500, 100, 50$ and 20. The decline in the genetic variance V_z within each cytotype due to drift is shown. The transparent lines show the complete simulation, whereas the solid line shows the same data but smoothed 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.

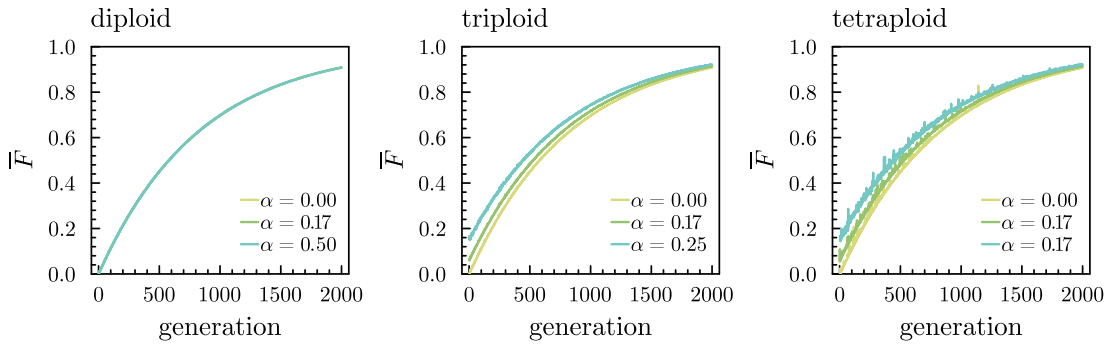

Figure S3: Average inbreeding coefficient $\bar{F}$ in each cytotype in a mixed-ploidy population for different values of α (we assume $\alpha_k = \alpha$, where α_k is the probability that a diploid gamete produced by a k-ploid cytotype contains two copies of the same parental gene at a random locus). We assume $N = 500, u = v = 0.08$ and no selection.

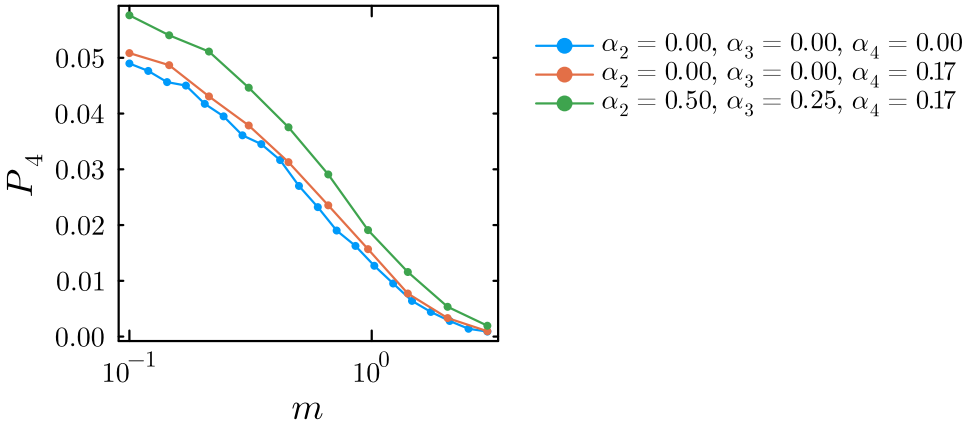

Figure S4: Nonzero α_k parameters increase the tetraploid establishment probability, but mainly due to α_2, which increases the segregation variance associated with the formation of diploid gametes by diploids. All other parameters and simulation details are as in fig. 3.

S2 Supplementary Information

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

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

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

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

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

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

At the critical point where the stable equilibrium disappears, we have that $\Delta g_1 = \frac{d\Delta g_1}{dg_1} = 0$ (fig. S5, 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. S5, right). This shows that, for plausible parameter values, we can safely assume that an initially diploid population will evolve to a mixed-ploidy equilibrium. A similar model was first analyzed in Felber and Bever (1997).

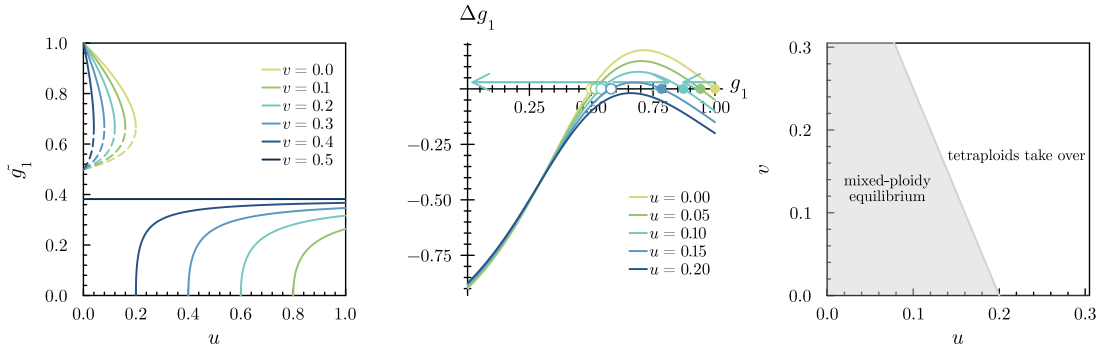

Figure S5: Deterministic mixed-ploidy equilibrium. The left plot shows the stable (solid lines) and unstable (dashed lines) equilibria for the proportion of haploid gametes in the gamete pool g_1 as a function of u for different values of v. The middle plot shows the relationship between $\Delta g_1 = g_1' - g_1$ 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.

S2.2 Stochastic mixed-ploidy model

For finite N, the basic mixed-ploidy model defines a Markov chain on the state space $[0..N] \times [0..N]$.

$$\begin{aligned} p_{ij,kl} &= \Pr\{N_2(t+1) = k, N_3(t+1) = l | N_2(t) = i, N_3(t) = j\} \\ &= \frac{N!}{k!l!(N-k-l)!} p_2^k p_3^l (1 - p_2 - p_3)^{N-k-l} \end{aligned} \tag{4}$$

where

$$p_2 = \left(\frac{i(1-u) + jv}{N(1-u) + (i+j)u + j(2v-1)} \right)^2 \tag{5}$$

$$p_3 = \frac{2(i(1-u) + jv)(N(1-u) + i(2u-1) + j(u+v-1))}{(N(1-u) + (i+j)u + j(2v-1))^2} \tag{6}$$

Associating a unique index with each pair (i, j) with $0 \leq i, j \leq N$, we can define a transition probability matrix P of dimensions $(N+1)^2 \times (N+1)^2$ for this Markov chain.

For nonzero u and v, the only absorbing state is the one where $N_2 = N_3 = 0$, i.e. the tetraploid cytotype fixes. All other states are transient, and hence tetraploid fixation occurs with probability one. The expected time until fixation may however be extremely long. Using standard theory for absorbing Markov chains, we can numerically compute the expected time until fixation $\mathbb{E}[T_{\text{fix}}]$ from the transition probability matrix. Calculations for the case where $u = v = 0.05$ (which are large parameter values conducive for tetraploid fixation) are shown in table 1. Clearly, tetraploid establishment by drift alone requires very small population sizes to occur at an appreciable rate. A similar model without triploids has been analyzed by Rausch and Morgan (2005).

Table 1: Expected number of generations until fixation of the tetraploid cytotype for different population sizes, assuming $u = v = 0.05$ and an initially diploid population.

N	10	20	30	40	50
$\mathbb{E}[T_{\text{fix}}]$	5.4×10^3	6.4×10^5	7.9×10^7	9.8×10^9	1.2×10^{12}

S2.3 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. 2), we have the recursive relations

$$\begin{aligned}
\mathbb{E}[T_4] &= \frac{1}{Z_2}\Big(\pi_2 u + (1 + \mathbb{E}[T_3])\pi_3 v + (1 + \mathbb{E}[T_4])\pi_4(1-u)\Big) \\
\mathbb{E}[T_3] &= \frac{1}{3Z_1}\Big(\pi_2(1-u) + (1 + \mathbb{E}[T_3])\pi_3 v\Big) \\
&\quad + \frac{2}{3Z_2}\Big(\pi_2 u + (1 + \mathbb{E}[T_3])\pi_3 v + (1 + \mathbb{E}[T_4])\pi_4(1-u)\Big)
\end{aligned} \tag{7}$$

where

$$\begin{aligned}
Z_1 &= \pi_2(1-u) + \pi_3 v \\
Z_2 &= \pi_2 u + \pi_3 v + \pi_4(1-u)
\end{aligned}$$

(these expressions are straightforwardly modified when more general u_{ij} are assumed, see e.g. sec. S2.4, eq. 8). The system in eq. 7 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

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

Numerical examples are shown in (fig. S6). 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 $\mathbb{E}[T_4] \approx 1.19$.

S2.4 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

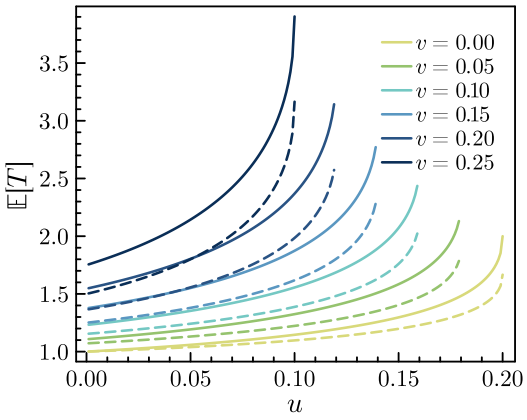

Figure S6: Expected time to diploid ancestry. The solid lines show $\mathbb{E}[T_4]$, i.e. the expected time since being inherited from a diploid ancestor for a random gene in a tetraploid individual at equilibrium, for different values of v (half the triploid fertility). The dashed lines show $\mathbb{E}[T_3]$, 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.

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. 2), we have

$$\begin{aligned}\nu(t+1) &= \nu(t)P \\ &= \nu(t)\begin{pmatrix} \frac{u_{21}}{Z_1}\pi_2 & \frac{u_{31}}{Z_1}\pi_3 & 0 \\ \left(\frac{u_{21}}{3Z_1} + \frac{2u_{22}}{3Z_2}\right)\pi_2 & \left(\frac{u31}{3Z_1} + \frac{2u_{32}}{3Z_2}\right)\pi_3 & \frac{2u_{42}}{3Z_2}\pi_4 \\ \frac{u_{22}}{Z_2}\pi_2 & \frac{u_{32}}{Z_2}\pi_3 & \frac{u_{42}}{Z_2}\pi_4 \end{pmatrix}\end{aligned} \tag{8}$$

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

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

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. 2 and solving the eigenvalue problem, this yields an unwieldy expression in the u_{ij}. For our usual parameterization where $u_{21} =$

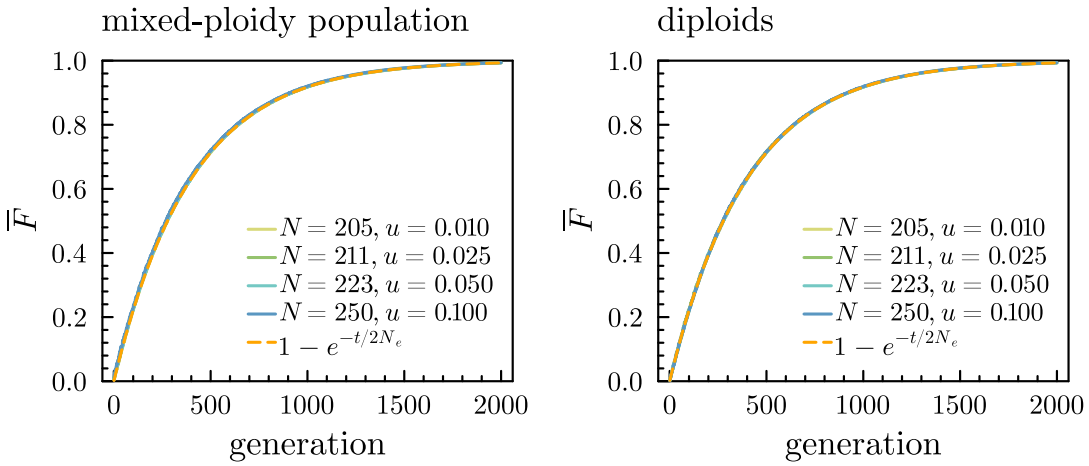

Figure S7: The evolution of $\bar{F}$ in the mixed-ploidy population and in the diploid subpopulation are shown for different values of u and associated values of N, keeping $N_e = (1-2u)N$ constant at 200. We assume $u = v$. All lines coincide almost completely and are indistinguishable from $1 - e^{-t/2N_e}$. 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.

$u_{42} = 1 - u, u_{22} = u$ 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. S7). 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.

S2.5 Inbreeding in the mixed-ploidy model

S2.5.1 Effect of inbreeding on segregation variance in autotetraploids

In polyploids, the inbreeding coefficient 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. Similar considerations apply to triploids if we only model haploid and diploid gametes. For higher gametic ploidy levels, one would need to track higher order identity coefficients in polyploids, which is intractable in general (Barton et al., 2023).

S2.5.2 Recursions for inbreeding coefficients in the mixed-ploidy model

Denoting the parents of individual i by k and l, the recursion for the inbreeding coefficients in an autotetraploid population is

$$F_i = \frac{1}{6}(F_k^* + F_l^* + 4\Phi_{kl}) \tag{9}$$

where $F_k^* = \alpha_4 + (1-\alpha_4)F_k$. The recursion 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 coancestry coefficient for individuals k and l).

There is little difficulty in extending the recursions for diploids (Barton et al., 2017) and autotetraploids (eq. (9)) to the mixed-ploidy case. Denoting the parents of individual i by k and l, the recursion for the inbreeding coefficients in the mixed-ploidy case becomes

$$\begin{aligned}
F_i &= \Phi_{kl} && \text{if } c_i = 2 \\
F_i &= \frac{1}{3}(F_k^* + 2\Phi_{kl}) && \text{if } c_i = 3, g_k = 2, g_l = 1 \\
F_i &= \frac{1}{3}(F_l^* + 2\Phi_{kl}) && \text{if } c_i = 3, g_k = 1, g_l = 2 \\
F_i &= \frac{1}{6}(F_k^* + F_l^* + 4\Phi_{kl}) && \text{if } c_i = 4
\end{aligned} \tag{10}$$

where $F_k^* = \alpha_{c_k} + (1 - \alpha_{c_k})F_k$, as in the autotetraploid model. The recursion for the coancestry coefficients in eq. (10) is given by

$$\begin{aligned}
\Phi_{ii} &= \frac{1}{c_i}(1 + (c_i - 1)F_i) \\
\Phi_{ij} &= \sum_k \sum_l P_{ik} P_{jl} \Phi_{kl} && i \neq j
\end{aligned} \tag{11}$$

where the sums are over individuals in the parental population, and where $P_{ik} \in \{0, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, 1\}$ is the probability that a gene copy in i is derived from parent k.

S2.6 Segregation variance expressions for the mixed-ploidy infinitesimal model

Consider a single locus in a population of k-ploids, and let $X_1, \ldots, X_k$ denote a random genotype at this locus, where X_i is the additive genetic value of the allele on homolog i. The variance in gametic values Y produced by some meiotic process can be decomposed as

$$\mathrm{Var}[Y] = \underbrace{\mathbb{E}\left[\mathrm{Var}[Y|X_1, \ldots, X_k]\right]}_{\text{segregation variance}} + \underbrace{\mathrm{Var}\left[\mathbb{E}[Y|X_1, \ldots, X_k]\right]}_{\text{population variation}} \tag{12}$$

We use this relationship to derive the segregation variance associated with haploid and diploid gamete production in the three cytotypes of the mixed-ploidy population.

We write $V_{S(k,l)}$ for the segregation variance associated with a k-ploid individual producing an l-ploid gamete, and we express this in terms of the variance $\mathcal{V}_k$ associated with a haploid genome in a (hypothetical) k-ploid reference population at HWLE. Using the scaling assumptions outlined in the main text, i.e. $\mathcal{V}_k = \beta_k^2 V$, we can then express all segregation variances in terms of the diploid segregation variance in the reference population V.

As an example, consider ordinary diploid meiosis at a single locus, where a diploid produces a haploid gamete. Y is a random haploid gamete sampled from the population, and $\mathrm{Var}[Y] = \mathrm{Var}[X] = v_x$, where X is a a randomly sampled gene from the population and v_x is the variance in additive genetic values at the locus. We have

$$\mathrm{Var}[\mathbb{E}[Y|X_1, X_2]] = \mathrm{Var}\left[\frac{X_1 + X_2}{2}\right] = \frac{v_x}{2}$$

Using eq. (12), we can then find the segregation variance

$$\mathbb{E}[\mathrm{Var}[Y|X_1, X_2]] = v_x - \frac{v_x}{2} = \frac{v_x}{2}$$

Summing over L independent loci, we find the gametic segregation variance in diploids producing haploid gametes (which is halve the zygotic segregation variance V, by definition) as

$$V_{S(2,1)} = \frac{V}{2} = \sum_{i=1}^{L} \frac{v_{x,i}}{2} = \frac{\mathcal{V}_2}{2}$$

and hence $V = \mathcal{V}_2$, where $\mathcal{V}_2$ is the genetic variance associated with a haploid genome in the diploid cytotype, as defined in the main text. This will hold in the infinitsimal limit, where L becomes very large and v_x smaller and smaller.

Meiosis in autotetraploids. When an autotetraploid forms quadrivalents during prophase I, a form of 'internal inbreeding' may occur as a result of the phenomenon called *double reduction* (see e.g. 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. S8. In the example

shown in fig. S8, 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 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_4, and summing over independent loci, we find that the gametic segregation variance in the presence of double reduction should be

$$V_{S(4,2)} = (1-\alpha_4)\mathcal{V}_4 + 3\alpha_4\mathcal{V}_4 = \mathcal{V}_4(1+2\alpha_4). \tag{13}$$

where, again, $\mathcal{V}_4$ is the genetic variance associated with a haploid genome in the (hypothetical) tetraploid reference population, as defined in the main text.

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; De Storme and Geelen, 2013). Consider a locus in a diploid with two distinct genes A and a. Assume recombination between the centromere and the locus 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. S9. Writing the genotype at a locus in the diploid

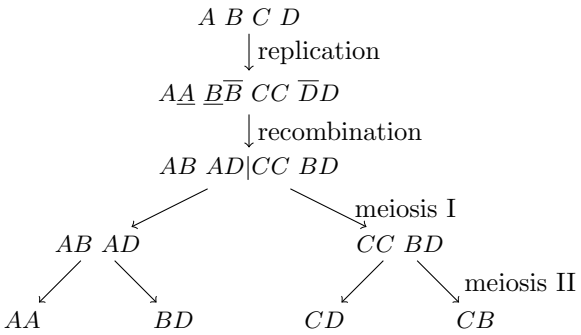

Figure S8: Schematic illustration of a meiotic division in an autotetraploid leading to double reduction at a locus with genotype $ABCD$. Two recombination events are assumed to occur at the locus (denoted by the bars).

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

$$Y = \begin{cases} 2X_1 & \text{w.pr. } \alpha_2/2 \\ 2X_2 & \text{w.pr. } \alpha_2/2 \\ X_1 + X_2 & \text{w.pr. } 1 - \alpha_2 \end{cases}$$

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 (see the diagram in fig. S9). Conditional on the latter event, we get

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

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 $V_{S(2,2)} = 2\alpha_2\mathcal{V}_2$.

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).

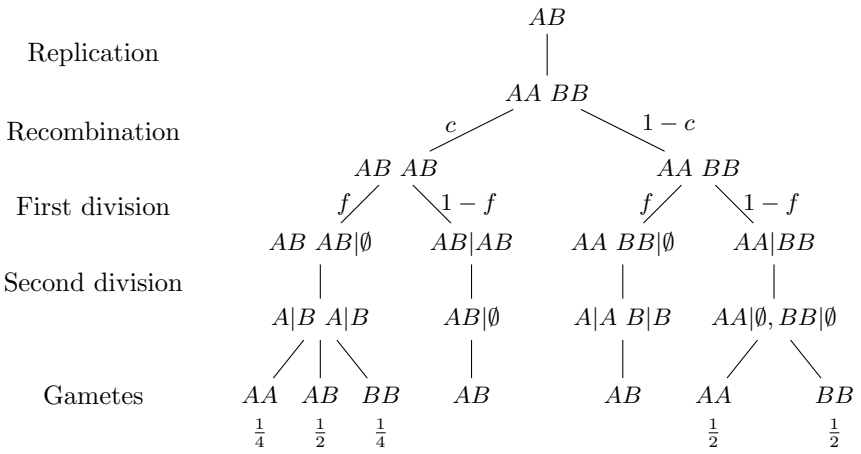

Figure S9: Schematic representation of the different pathways for unreduced gamete formation in diploids and their different outcomes.

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 *Arabidopsis 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. We note that, on the supposition that diploid gametes are produced by random assortment of chromosomes in a haploid and diploid gamete, $\alpha_3 \leq 1/4$.

When a triploid produces a haploid gamete, we get as segregation variance at a single locus

$$\begin{aligned}\mathbb{E}[\mathrm{Var}[Y|X_1, X_2, X_3]] &= \mathrm{Var}[Y] - \mathrm{Var}[\mathbb{E}[Y|X_1, X_2, X_3]] \\ &= v_x - \mathrm{Var}\left[\frac{1}{3}(X_1 + X_2 + X_3)\right] \\ &= v_x - \frac{1}{9}3v_x = \frac{2}{3}v_x\end{aligned}$$

Summing over many loci, we have $V_{S(3,1)} = \frac{2}{3}\mathcal{V}_3$. When a triploid produces a diploid gamete, we assume there is, as in diploids and tetraploids, a probability α_3 that the

same gene copy ends up twice in the gamete. Conditional on this happening, we have

$$\begin{aligned}\mathbb{E}[\mathrm{Var}[Y|X_1,X_2,X_3]] &= \mathrm{Var}[Y] - \mathrm{Var}[\mathbb{E}[Y|X_1,X_2,X_3]] \\ &= \mathrm{Var}[2X] - \mathrm{Var}\left[\frac{1}{3}(2X_1+2X_2+2X_3)\right] \\ &= 4V_x - \frac{4}{9}3V_x = \frac{8}{3}V_x\end{aligned}$$

Conditional on this *not* happening,

$$\begin{aligned}\mathbb{E}[\mathrm{Var}[Y|X_1,X_2,X_3]] &= \mathrm{Var}[Y] - \mathrm{Var}[\mathbb{E}[Y|X_1,X_2,X_3]] \\ &= 2V_x - \mathrm{Var}\left[\frac{1}{3}((X_1+X_2)+(X_1+X_3)+(X_2+X_3))\right] \\ &= 2V_x - \frac{4}{9}3V_x = \frac{2}{3}V_x\end{aligned}$$

Putting this together and summing over many loci, the segregation variance for a diploid gamete from a triploid individual would be

$$V_{S(3,2)} = \frac{8}{3}\mathcal{V}_3\alpha_3 + \frac{2}{3}\mathcal{V}_3(1-\alpha_3) = \frac{2}{3}\mathcal{V}_3(1+3\alpha_3)$$

The different expressions for the gametic segregation variance in the mixed-ploidy model, adjusted for inbreeding, are summarized in table 2 in the main text.

S2.7 Equilibrium trait value distribution for a large mixed-ploidy population

In our establishment model, we assume that migrant individuals are drawn randomly from a large (effectively infinite) non-inbred mixed-ploidy population at HWLE and cytotype equilibrium, with mean trait value zero. The equilibrium trait value distribution in a monocytotypic population under such conditions is straightforwardly obtained under the infinitesimal model as a Gaussian with variance $k\beta_k^2 V$ (see main text). However, the equilibrium trait value distributions for the different cytotypes in the mixed-ploidy model are more challenging to obtain.

Indeed, each cytotype produces haploid and diploid gametes with Gaussian gametic values, distributed according to eq. (6) in the main text. Conditional on the parental trait values, offspring trait values will be sums of independent Gaussians, and hence again Gaussian, distributed according to eq. (7) in the main text. Consider now, for instance, a triploid individual, there are six different parental cytotype combinations that may yield a triploid offspring, with relative frequencies proportional to the terms in the expansion of the following polynomial in π

$$2((1-u)\pi_2 + v\pi_3)(u\pi_2 + v\pi_3 + (1-u)\pi_4)$$

(where the term in π_2^2 is proportional to the frequency of triploids derived from two diploid parents, the term in $\pi_2\pi_3$ proportional to the frequency of triploids derived from

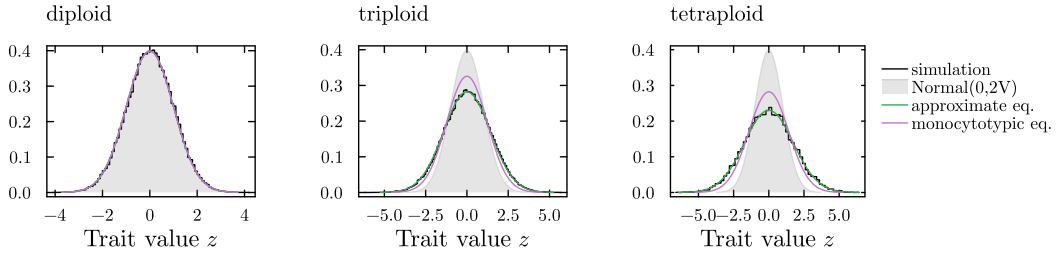

Figure S10: Equilibrium trait value distribution for diploids, triploids and tetraploids. The simulated equilibrium distribution, the small u approximation and the monocytotypic equilibrium distribution are shown for all three cytotypes. The diploid monocytotypic distribution $\mathcal{N}(0, 2V)$ is shown everywhere as a reference. Here we assumed $\beta_3 = \beta_4 = 1$ and $\alpha_2 = 1/2, \alpha_3 = 1/4$ and $\alpha_4 = 1/6$.

a diploid and a triploid parent, *etc.*). Each of these combinations of parental cytotypes potentially yields a different Gaussian distribution over offspring trait values. As a result, the triploid trait distribution in the next generation will be a *mixture* of Gaussians, with mixture proportions given by the relative frequencies in the expansion of the above polynomial. A similar reasoning applies to the diploid and tetraploid subpopulation. Proceeding to the next generation, the trait distributions will be mixtures mixtures of Gaussians, and so on for further generations.

We have not identified whether this process yields a tractable equilibrium distribution. However, for a predominantly diploid population with u sufficiently small, an accurate approximation is readily obtained by assuming that effectively all individuals in the population have diploid parents. In this case, still assuming the mean trait value to be zero, diploid trait values will be Gaussian with variance $2V$. Using eq. (7), the trait value distribution of newly formed triploids will be Gaussian with mean zero and variance

$$\begin{aligned}
\mathrm{Var}[Z_{ij}^{12}] &= \mathbb{E}[\mathrm{Var}[Z_{ij}^{12}|Z_i, Z_j]] + \mathrm{Var}[\mathbb{E}[Z_{ij}^{12}|Z_i, Z_j]] \\
&= \beta_3^2 \left(\frac{V}{2} + 2\alpha_2 V \right) + \mathrm{Var}\left[\beta_3 \left(\frac{Z_i}{2} + Z_j \right) \right] \\
&= 3\beta_3^2 V \left(1 + \frac{2}{3}\alpha_2 \right)
\end{aligned}$$

(where Z_{ij}^{12} denotes the trait value of a random offspring from individuals i and j, and where the superscript indicates that i contributes a haploid gamete, and j a diploid gamete – as in eq. (7)). Under the small u assumption, Z_i and Z_j are random draws from a Gaussian with mean zero and variance $2V$. The trait value distribution for tetraploids will be Gaussian with mean zero and variance

$$\mathrm{Var}[Z_{ij}^{22}] = 4\beta_4^2 V \left(1 + \alpha_2\right)$$

This approximation is illustrated in fig. S10, where, the fit to simulations is shown to be very accurate. Note that when the α_k parameters are set to zero (as in most of our results), the equilibrium variance for each cytotype in the mixed-ploidy population is just the equilibrium variance in an isolated monocytotypic population.

References

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

N. H. Barton, A. M. Etheridge, and A. Véber. The infinitesimal model with dominance. *Genetics*, 225(2):iyad133, 2023.

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

J. S. Charles, M. L. Hamilton, and T. D. Petes. Meiotic chromosome segregation in triploid strains of saccharomyces cerevisiae. *Genetics*, 186(2):537–550, 2010.

N. De Storme and D. Geelen. Sexual polyploidization in plants–cytological mechanisms and molecular regulation. *New Phytologist*, 198(3):670–684, 2013.

F. Felber and J. D. Bever. Effect of triploid fitness on the coexistence of diploids and tetraploids. *Biological Journal of the Linnean Society*, 60(1):95–106, 1997.

M. Lynch and B. Walsh. *Genetics and analysis of quantitative traits*, volume 1. Sinauer Sunderland, MA, 1998.

M. E. Moody, L. Mueller, and D. Soltis. Genetic variation and random drift in autotetraploid populations. *Genetics*, 134(2):649–657, 1993.

J. Ramsey and D. W. Schemske. Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annual review of ecology and systematics*, 29(1):467–501, 1998.

J. H. Rausch and M. T. Morgan. The effect of self-fertilization, inbreeding depression, and population size on autopolyploid establishmen. *Evolution*, 59(9):1867–1875, 2005.

F. Rousset. *Genetic structure and selection in subdivided populations*, volume 40. Princeton University Press, 2004.

M. Stift, C. Berenos, P. Kuperus, and P. H. van Tienderen. Segregation models for disomic, tetrasomic and intermediate inheritance in tetraploids: a general procedure applied to rorippa (yellow cress) microsatellite data. *Genetics*, 179(4):2113–2123, 2008.