Supplement: Autopolyploid establishment through polygenic adaptation

Arthur Zwaenepoel

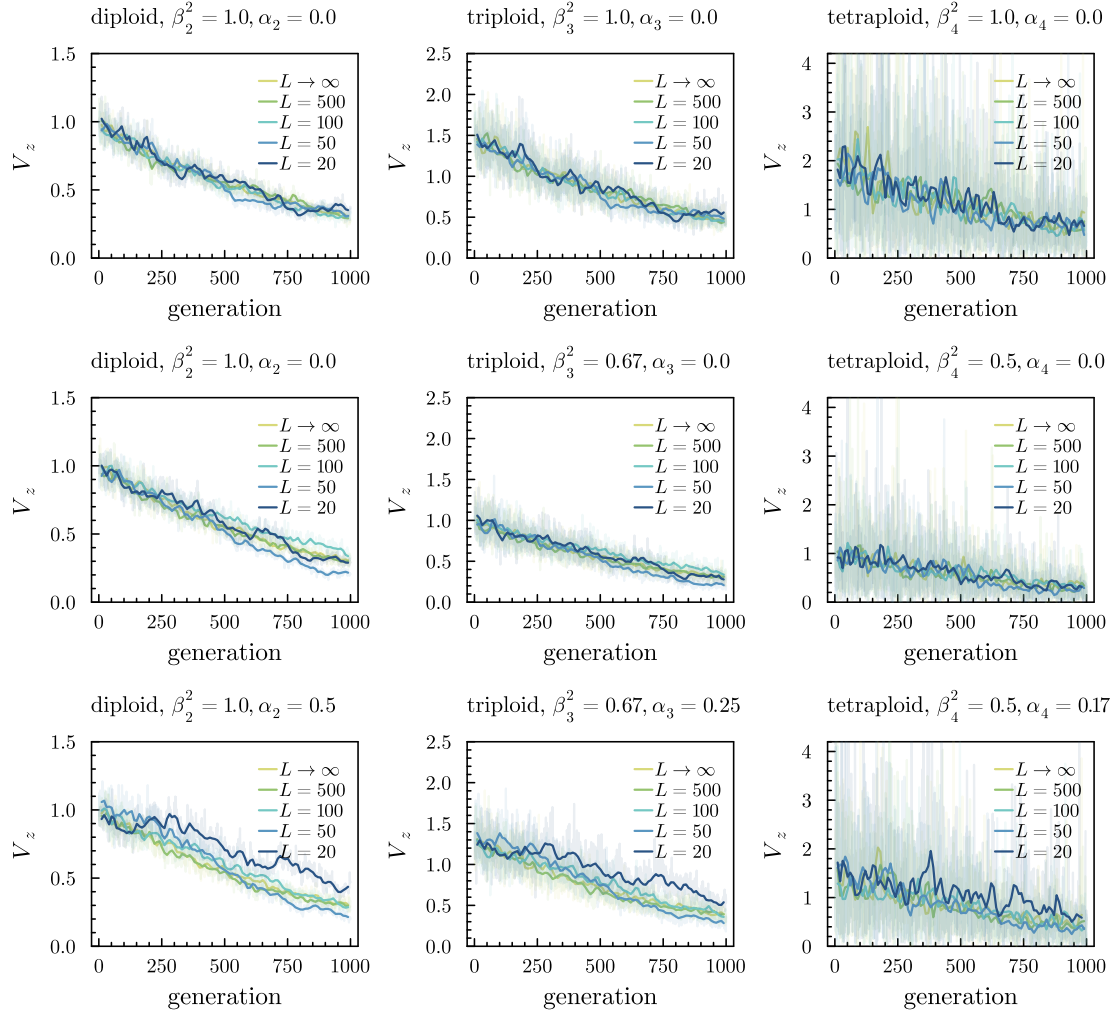# Contents

S1   Supplementary Figures

Figure S1: 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.
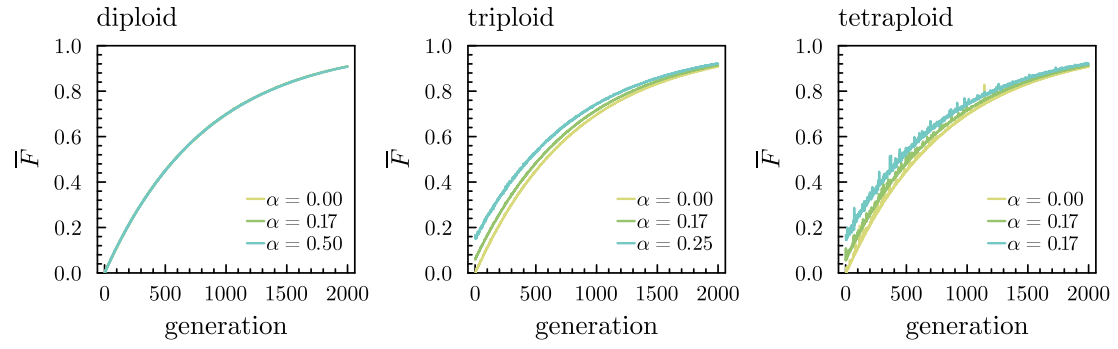
Figure S2: Average inbreeding coefficient $\bar{F}$ in each cytotype in a mixed-ploidy population for different values of α (we assume $\alpha_k = \alpha$, where α_k is the probability that a diploid gamete produced by a k-ploid cytotype contains two copies of the same parental gene at a random locus). We assume $N = 500, u = v = 0.08$ and no selection.

S2 Supplementary Information

S2.1 Inbreeding 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. For higher ploidy levels, one would need to track higher order identity coefficients.

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

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

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).

S2.2 Mixed-ploidy infinitesimal model

S2.2.1 Discrete locus model and variance scaling

Consider an L-locus additive model, with two alleles (0 and 1) at each locus. For a k-ploid individual, let $X_{i,j}$ be the allele at homolog j of locus i. We assume the trait value is determined by

$$z = \sum_{i=1}^{L} \sum_{j=1}^{k} a_{i,k} X_{i,j} \tag{2}$$

Where $a_{i,k}$ is the allelic effect of the 1 allele at locus i in k-ploids. The genetic variance at HWLE in k-ploids ($\tilde{V}_{z,k}$) will then be

$$\tilde{V}_{z,k} = k \sum_{i=1} a_{i,k}^2 p_i q_i = k V_{x,k} \tag{3}$$

where we refer to $V_{x,k}$ as the variance associated with a haploid genome in k-ploids at HWLE. Note that we also have $\tilde{V}_{z,2} = 2V_{x,k} = 2V$, where V is the segregation variance in the diploid population.

We now assume $a_{i,k} = \beta_k a_{i,2}$, i.e. allelic effects in k-ploids are as in diploids, but scaled by a factor β_k. This implies that

$$\frac{\tilde{V}_{z,k}}{\tilde{V}_{z,2}} = \frac{kV_{x,k}}{2V_{x,2}} = \frac{k}{2}\beta_k^2 \tag{4}$$

and hence also that $V_{x,k} = \beta_k^2 V_{x,2} = \beta_k^2 V$. These relations will also hold in the infinitesimal limit. Below, we derive the segregation variance expressions for the different meiotic processes in the mixed-ploidy models in terms of the $V_{x,k}$. Using the relationships outlined here, we will then be able to express all the required variances in terms of V.

Note that under the above assumptions, not only the variance is rescaled, but also the expected value of offspring within a family. Specifically, for a k-ploid offspring of parental pair i and j, where the gamete coming from i is g_i-ploid and the other gamete g_j-ploid, we have

$$\mathbb{E}[Z_{ij}] = \beta_k \left(\frac{g_i}{c_i} \frac{z_i}{\beta_{c_i}} + \frac{g_j}{c_j} \frac{z_j}{\beta_{c_j}} \right) \tag{5}$$

where c_i is the ploidy level of individual i. Here we rescale the parental trait values to the diploid scale by dividing by β_{c_i}, sum the expected rescaled gametic values, and rescale to the offspring ploidy level.

S2.2.2 Segregation variance

S2.3 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

$$\tilde{g}_1, \tilde{g}_1' = \frac{3 - 3u - 6v \pm \sqrt{(u + 2v - 1)(5u + 2v - 1)}}{2(2 - u - 4v)} \tag{6}$$
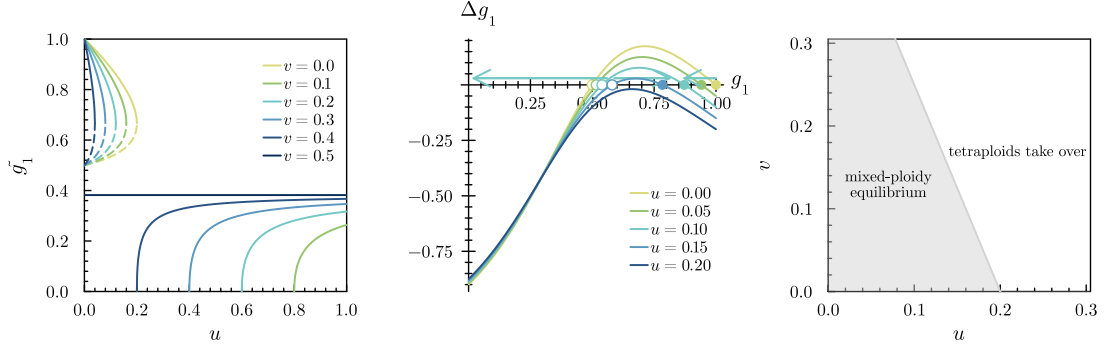
5

Figure S3: 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.

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

$$\pi_2 = \tilde{g}_1{}^2 \qquad\qquad \pi_3 = 2\tilde{g}_1\tilde{g}_2 \qquad\qquad \pi_4 = \tilde{g}_2{}^2 \qquad (7)$$

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

$$\begin{aligned} \pi_2 &= 1 - 2u - 4uv - u^2 + O(u^3) \\ \pi_3 &= 2u + 4uv + O(u^3) \\ \pi_4 &= u^2 + O(u^3) \end{aligned} \qquad (8)$$

At the critical point where the stable equilibrium disappears, we have that $\Delta g_1 = \frac{d\Delta g_1}{dg_1} = 0$ (fig. S3, 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. S3, 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].

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}

S2.4 Stochastic mixed-ploidy model

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

$$p_{ij,kl} = \Pr\{N_2(t+1) = k, N_3(t+1) = l | N_2(t) = i, N_3(t) = j\}$$
$$= \frac{N!}{k!l!(N-k-l)!} p_2^k p_3^l (1 - p_2 - p_3)^{N-k-l} \tag{9}$$

where

$$p_2 = \left(\frac{i(1-u) + jv}{N(1-u) + (i+j)u + j(2v-1)} \right)^2 \tag{10}$$

$$p_3 = \left(\frac{2(i(1-u) + jv)(N(1-u) + i(2u-1) + j(u+v-1))}{N(1-u) + (i+j)u + j(2v-1)} \right)^2 \tag{11}$$

Associating a unique index with each pair (i, j) with $0 \le i, j \le N$, 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].

S2.5 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. S2.3, eq. 7), we have the recursive relations

$$\mathbb{E}[T_4] = \frac{1}{Z_2}\Big(\pi_2 u + (1 + \mathbb{E}[T_3])\pi_3 v + (1 + \mathbb{E}[T_4])\pi_4(1 - u)\Big)$$

$$\mathbb{E}[T_3] = \frac{1}{3Z_1}\Big(\pi_2(1 - u) + (1 + \mathbb{E}[T_3])\pi_3 v\Big)$$
$$+ \frac{2}{3Z_2}\Big(\pi_2 u + (1 + \mathbb{E}[T_3])\pi_3 v + (1 + \mathbb{E}[T_4])\pi_4(1 - u)\Big) \qquad (12)$$

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. S2.6, eq. 13). The system in eq. 12 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. S4). 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 $\mathbb{E}[T_4] \approx 1.19$.

S2.6 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. 7), we have

$$\nu(t + 1) = \nu(t)P$$

$$= \nu(t)\begin{pmatrix} \frac{u_{21}}{Z_1}\pi_2 & \frac{u_{31}}{Z_1}\pi_3 & 0 \\ \left(\frac{u_{21}}{3Z_1} + \frac{2u_{22}}{3Z_2}\right)\pi_2 & \left(\frac{u_{31}}{3Z_1} + \frac{2u_{32}}{3Z_2}\right)\pi_3 & \frac{2u_{42}}{3Z_2}\pi_4 \\ \frac{u_{22}}{Z_2}\pi_2 & \frac{u_{32}}{Z_2}\pi_3 & \frac{u_{42}}{Z_2}\pi_4 \end{pmatrix} \qquad (13)$$

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

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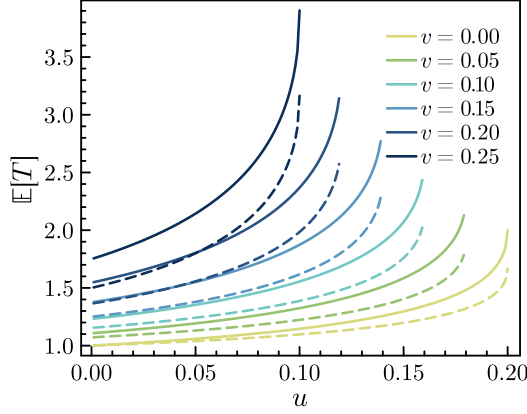$$Z_2 = u_{22}\pi_2 + u_{32}\pi_3 + u_{42}\pi_4$$

8

Figure S4: 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 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. 7 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. S5). When $v = 0$ and $u < u_c$ (see sec. S2.3), $N_e = \pi_2 N$, 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.
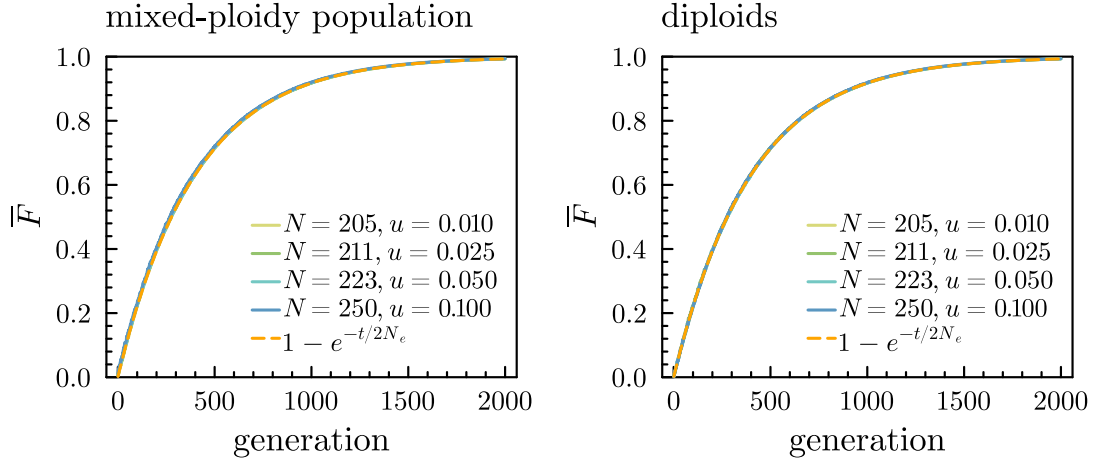
Figure S5: The evolution of $\bar{F}$ in the mixed-ploidy population and in the diploid subpopulation are shown for different values of u and associated values of N, keeping $N_e = (1 - 2u)N$ constant at 200. We assume $u = v$. All lines coincide almost completely and are indistinguishable from $1 - e^{-t/2N_e}$. 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.

References

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

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.