

The infinitesimal model for polyploid and mixed-ploidy populations

Arthur Zwaenepoel

Contents

The basic infinitesimal model	1
Phenotypic model definition	1
Inbreeding and the evolution of the segregation variance	2
Double reduction in autotetraploids	4
The infinitesimal model for mixed-ploidy populations	6
Scaling of genetic variance across ploidy levels	6
Unreduced gamete formation in diploids	8
Inbreeding coefficients in the mixed-ploidy system	9
Triploids	10
References	11

The basic infinitesimal model

Phenotypic model definition

The infinitesimal model assumes that offspring trait values Z of a parental pair with trait values z_i and z_j are normally distributed:

$$Z_{ij} \sim \mathcal{N}\left(\frac{z_i + z_j}{2}, V_{i,j}\right)$$

where $V_{i,j}$ is called the *segregation variance* and is determined by the hereditary process. The basic infinitesimal model can be derived as the limit of a model where a quantitative trait is controlled by a large number n of unlinked Mendelian loci with additive gene action, each of small effect $\sim O(1/\sqrt{n})$.

A slightly different, and perhaps more insightful, way to specify the same model is to write $Z_{ij} = X_i + X_j$, where X_i is the contribution of parent i to the genotypic value of the offspring and X_j the same for j . That is, X_i is the genotypic value of a gamete from i . For gametes produced by a normal meiotic division, we assume $X_i \sim \mathcal{N}(z_i/2, V_i)$, where V_i is the contribution to the segregation variance from i . Segregation occurs independently in both parents, contributing additively to the segregation variance $V_{i,j} = V_i + V_j$. When considered as the limit of a large number of unlinked additive Mendelian loci of small effect, we can partition the segregation variance in contributions from each locus

$$V_0 = \sum_{k=1}^n v_{0,k}$$

Where $v_{0,k}$ is the contribution to the segregation variance of the k th locus. Considering this viewpoint explicitly often helps in deriving properties of the infinitesimal model.

If we assume a population consisting of unrelated individuals with genetic variance V and segregation variance V_0 for all parental pairs, we find that under random mating the variance in the offspring generation is

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

So that at equilibrium ($V' = V$), the trait distribution for the population will be Gaussian with variance $2V_0$.

Inbreeding and the evolution of the segregation variance

Importantly, $V_{i,j}$ is not a function of z_i or z_j (although they may be correlated), but nevertheless evolves over time. For a finite population, inbreeding will lead to an increase in homozygosity and cause the segregation variance to decrease over time. Indeed, from the viewpoint of the Mendelian limit, it is clear that in the extreme case where an individual is completely homozygous for all loci affecting some trait, Mendelian segregation does not generate any variance, and all gametes of such an individual have, barring mutation, the same genotypic value. Importantly, while the basic phenotypic model (where offspring traits are distributed according to a Gaussian around the midparent value) holds for arbitrary ploidy levels, genetic drift – and consequently, the evolution of the segregation variance in finite populations – will differ for different ploidy levels.

The infinitesimal model for finite populations of haploid and diploid individuals is described in detail in Barton, Etheridge, and Véber (2017). We shall use a slightly different notation here. Let F_i be the inbreeding coefficient of individual i , i.e. the probability that two *distinct* genes sampled from individual i are identical by descent (IBD). Let, furthermore, Φ_{ij} be the coancestry coefficient of individuals i and j , or the probability that two genes sampled independently from individuals i and j are IBD. Note that for any ploidy level m , we have the relationship $\Phi_{ii} = (1 + (m - 1)F_i)/m$.

In haploids, the situation is somewhat different from other ploidy levels, owing due to the absence of reduced gametes in sexual reproduction. While we have no notion of homozygosity for an individual haploid individual, whenever alleles at some locus in a mating pair of individuals is IBD, Mendelian segregation during meiosis will fail to contribute to the segregation variance. It is easy to see that for a parental pair (i, j) , the segregation variance is reduced to

$$V_{ij} = V_0(1 - \Phi_{ij})$$

Now, for diploids and higher ploidy levels, the situation is different, since Mendelian segregation happens in the generation of gametes through meiosis. Segregation hence happens independently in the generation of the two gametes, and the variance is reduced to the degree that each parent is inbred. For diploids, we can again easily find the resulting segregation variance

$$V_{ij} = V_i + V_j = \frac{V_0}{2}(1 - F_i) + \frac{V_0}{2}(1 - F_j) = V_0\left(1 - \frac{F_i + F_j}{2}\right) \quad (1)$$

In higher ploidy levels, the situation gets somewhat more complicated, as there are different degrees of homozygosity. For instance, the different possible states of homozygosity in a tetraploid can be symbolically represented as $abcd$, $aabc$, $aabb$, $aaab$ and $aaaa$, and in general, the number of homozygosity states grows according to the partition function $(1, 2, 3, 5, 11, 15, 22, \dots)$. If we represent the probability of being in these five increasingly homozygous states as $\delta_1, \dots, \delta_5$, we find that the segregation variance is reduced by a factor

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

Note, furthermore, that the inbreeding coefficient in tetraploids is related to the homozygosity coefficients as

$$F_i = \frac{1}{6}\delta_2 + \frac{1}{3}\delta_3 + \frac{1}{2}\delta_4 + \delta_5 = 1 - \phi$$

So that the reduced segregation variance is, as in diploids, given by Equation 1. This is an important result, showing that we need not track the array of homozygosity

coefficients to compute the segregation variance contributed by a tetraploid individual, but only require its inbreeding coefficient.

When simulating the infinitesimal model, we shall hence need a way to efficiently track the inbreeding and coancestry coefficients during the simulation. The recursion for the inbreeding coefficients is

$$\begin{aligned} F_i &= \frac{1}{2}(1 + \Phi_{kl}) && \text{diploids} \\ F_i &= \frac{1}{6}(F_k + F_l + 4\Phi_{kl}) && \text{tetraploids} \end{aligned}$$

where k and l are the parents of i (note that $k = l$ is possible). The general formula for m -ploids, with m even, is

$$F_i = \binom{m}{2}^{-1} \left[\binom{m/2}{2} (F_k + F_l) + \left(\frac{m}{2}\right)^2 \Phi_{kl} \right]$$

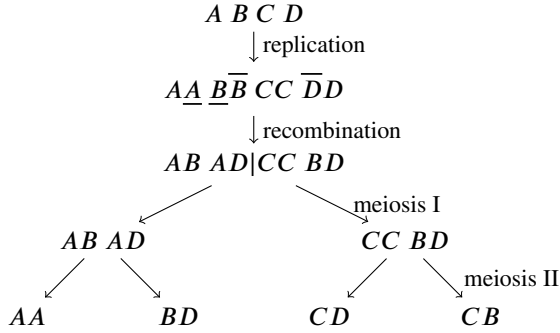
The recursion for the coancestry coefficients in m -ploids is given by

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

where the sums are over individuals in the parental population, and where $P_{ik} \in \{0, 1/2, 1\}$ 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 .

Double reduction in autotetraploids

When an autotetraploid forms tetravalents during prophase I, a form of internal inbreeding may occur as a result of the phenomenon called double reduction. Double reduction happens when, as a result of recombination, replicated copies on sister chromatids move to the same pole during anaphase I. Schematically, an example of double reduction for a genotype $ABCD$ could look like:



Where we have two recombination events involving the locus (denoted by the bars). One of the four generated gametes is AA , which is not possible in the case of bivalent meiosis, because in anaphase I paired chromosomes (involved in cross-overs) are separated in that case. 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, and should therefore be in part determined by the distance of the locus to the centromere. In the context of the infinitesimal model however, we may consider the probability of double reduction a parameter, α .

Clearly, in the presence of double reduction, such an $ABCD$ genotype would generate 10 distinct gametes:

$$\begin{array}{cccc}
 AA & \cdot & \cdot & \cdot \\
 AB & BB & \cdot & \cdot \\
 AC & BC & CC & \cdot \\
 AD & BD & CD & DD
 \end{array}$$

where each of the gametes on the diagonal is produced with probability $\alpha/4$ and the other six ‘normal’ gametes with probability $(1 - \alpha)/6$ each. For a random genotype $X_1 X_2 X_3 X_4$, we can find the expected segregation variance contributed by a locus when double reduction happens as follows. Let Y be the genotypic value of a gamete formed by double reduction, let G be the genotype, and let X denote a randomly sampled gene copy from the base population. We have $\text{var}[X] = v_0/2$

$$\begin{aligned}
 \mathbb{E}[\text{var}[Y|G]] &= \text{var}[Y] - \text{var}_G[\mathbb{E}[Y|G]] \\
 &= \text{var}[2X] - \text{var}\left[\frac{1}{4}(2X_1 + 2X_2 + 2X_3 + 2X_4)\right] \\
 &= 2v_0 - \frac{1}{2}v_0 \\
 &= \frac{3}{2}v_0
 \end{aligned}$$

Summing across independent loci, we find that segregation variance in the presence

of double reduction is increased by a factor $(1 + 2\alpha)$:

$$(1 - \alpha)\frac{V_0}{2} + \alpha\frac{3}{2}V_0 = (1 + 2\alpha)\frac{V_0}{2} \quad (2)$$

While double reduction increases the segregation variance in any given cross, in the long term it causes a decrease in the segregation variance through its effect on the rate of inbreeding. Indeed, double reduction leads to a kind of ‘internal inbreeding’ (Lynch and Walsh 1998), accelerating the decay of heterozygosity. While double reduction does not affect the recursions for the coancestry coefficients (due to the symmetry of the phenomenon), it does affect the inbreeding coefficient, as the following calculation shows

$$\begin{aligned} F_i &= \left(\frac{1}{6}F_k + \frac{1}{6}F_l + \frac{2}{3}\Phi_{kl} \right)(1 - \alpha)^2 \\ &\quad + \left(\frac{1}{6} + \frac{1}{6}F_l + \frac{2}{3}\Phi_{kl} \right)\alpha(1 - \alpha) \\ &\quad + \left(\frac{1}{6}F_k + \frac{1}{6} + \frac{2}{3}\Phi_{kl} \right)\alpha(1 - \alpha) \\ &\quad + \left(\frac{1}{6} + \frac{1}{6} + \frac{2}{3}\Phi_{kl} \right)\alpha^2 \\ &= \frac{1}{6} \left(2\alpha + (1 - \alpha)(F_k + F_l) + 4\Phi_{kl} \right) \end{aligned}$$

In fig. 1, simulations for a Wright-Fisher population model with a quantitative trait following the infinitesimal model are shown for diploids and autotetraploids, with and without double reduction.

The infinitesimal model for 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.

Scaling of genetic variance across ploidy levels

[should start from the variance per allele? at equilibrium V_0 for dups, $V_0/2$ for tets (since $2V_0 = \text{var}[Z] = \text{var}[X_1 + X_2 + X_3 + X_4] = 4\text{var}[X]$ in tet case?)]

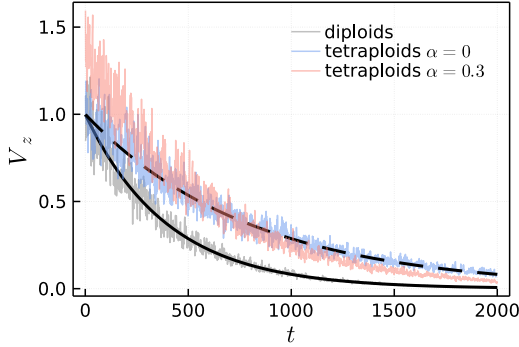


Figure 1: Decline of the phenotypic variance (V_z , here equal to the additive genetic variance) in Wright-Fisher populations ($N = 200$) simulated according to the infinitesimal model for different ploidy levels. The solid and dashed black lines show the (approximate for tetraploids) expected exponential decline of the variance for diploids and tetraploids respectively, given by $V_z(0)e^{-t/(mN)}$, where m is the ploidy level.

If we consider the infinitesimal model as the limit of a large number of additive Mendelian loci, the genotypic value of an individual of ploidy level m is defined as

$$z = \sum_{k=1}^n \sum_{l=1}^m a_{k,l}$$

where $a_{k,m}$ is the allelic effect of homolog l of locus k . In a mixed-ploidy system, it may seem natural to assume equal allelic effects in diploids and congeneric polyploids. However, clearly, if we make such an assumption, the phenotypic range of tetraploids will be double that of diploids. While the assumption of an additive genetic architecture is of course problematic in itself, this additional assumption may be particularly unjustified. We hence introduce an additional scaling parameter β_m for ploidy level m , such that the segregation variance $V_{0,m} = \beta_m V_{0,2}$.

We can provide an interpretation for this parameter in the context of the underlying hypothetical Mendelian system. The (additive) genetic variance for a genetic architecture consisting of n unlinked biallelic loci in an m -ploid population at Hardy-Weinberg and linkage equilibrium (HWLE) is given by

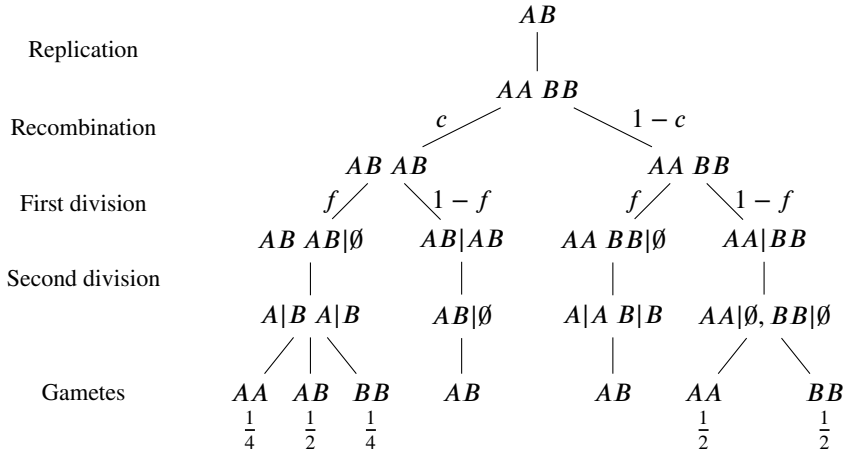
$$\tilde{V}_A = k \sum_{i=1}^n a_i^2 p_i (1 - p_i)$$

Which should equal the equilibrium prediction $2V_0$ of the infinitesimal model in the large n limit. If we rescale the segregation variance in m -ploids $V_{0,m}$ relative to that

in diploids so that $V_{0,m} = \beta_m V_{0,2}$, this can be interpreted as a scaling of $\sqrt{\beta_m}$ of the allelic effects of the underlying Mendelian system.

Unreduced gamete formation in diploids

Naively, one may think that an unreduced gamete contains the parental genome, and that as a result the segregation variance for a $2n \times 2n \rightarrow 4n$ cross would be zero. However, 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 aberration that leads to their origin: (1) first division restitution (FDR) of (2) second division restitution (SDR). Consider a locus in a 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 in the following diagram:



Clearly, the mechanism of unreduced gamete formation generates segregation variance, as not all random tetraploid offspring from a single diploid parental pair will receive the same pair of genes from each parent, depending on whether or not recombination has occurred and FDR rather than SDR generates the unreduced gamete.

If we consider the diploid genotype Aa depicted in the diagram above and consider X_A the number of A alleles transmitted, we can easily find

$$\text{var} X_A = \left(1 - c - f + \frac{3}{2}cf\right) := \xi$$

which is also the probability of transmitting two copies of the same allele. We shall denote this derived quantity as ξ .

Of course, if the locus is IBD with probability $F_{i,i}$, we have that the expected value of $\text{var } X_A$ is $\xi(1 - F_{i,i})$. Under the model delineated above with the scaling of allelic effects across ploidy levels, we find that the contribution of a single diploid parent i to the segregation variance among tetraploid offspring is

$$V_i = 2\beta_4 V_{0,2} \xi (1 - F_{i,i})$$

Notably, the mean offspring phenotype among tetraploid offspring from a diploid cross can no longer be the average of the two parental phenotypes if we assume the infinitesimal model as the limit of the Mendelian unlinked additive loci model. The expected contribution of a single parent i will be $X_i = \sqrt{\beta_4} z_i$ so that for a cross of two diploids via unreduced gametes

$$Z_{ij} \sim \mathcal{N}\left(\sqrt{\beta_4}(z_i + z_j), V_i + V_j\right)$$

Inbreeding coefficients in the mixed-ploidy system

In a mixed-ploidy system tracking inbreeding coefficients becomes slightly more complicated, as our recursions will differ whether some individual is derived from parents of the same cytotype or not.

Recall that ξ is the probability that an unreduced gamete transmits two copies of the same allele at some locus. Let m_k denote the ploidy level of individual k . We still have that for a tetraploid individual i

$$F'_{i,i} = \frac{1}{6} (F_{k,k}^* + F_{l,l}^* + 4F_{k,l})$$

where now

$$F_{k,k}^* = \begin{cases} F_{k,k} & \text{if } m_k = 4 \\ F_{k,k}(1 - \xi) + \xi & \text{if } m_k = 2 \end{cases}$$

For $F'_{i,j}$ we still have

$$F'_{i,j} = \sum_k \sum_l P_{i,k} P_{i,l} F_{k,l}^*$$

where now

$$F_{k,l}^* = \begin{cases} F_{k,l} & \text{if } k \neq l \\ \frac{1}{2}(1 + F_{k,k}) & \text{if } k = l, m_k = 2 \\ \frac{1}{4}(1 + 3F_{k,k}) & \text{if } k = l, m_k = 4 \end{cases}$$

(I think), so that we still have a matrix expression

$$F' = P \left(F + (I - m \text{diag} F) \right) P^T$$

still holds, but where m is now a vector where the k th element is $1/m_k$, recording the reciprocal of the ploidy levels in the parental generation.

Triploids

In the above we did not include triploid individuals in the population. Triploids may however 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 issues, we simply need to track the segregation variance contributions from both donor gametes, and relate these to $V_{0,3}$. The reproduction of triploids is however more complicated. We may assume that aneuploid gametes do not contribute to viable crossings and we need assumptions on what percentage of meioses render euploid ($1n, 2n$ or $3n$) gametes. We may treat these as parameters.

Experimental results indicate that, at least in yeast, triploid 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. 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 $(1/2)^5 \approx 0.03$. This is on the order of the unreduced gamete formation rate and – if we wish to incorporate triploids in the model – should not be ignored if n is sufficiently small.

The question remains what the segregation variance contributed by a haploid, diploid or triploid gamete to its offspring is.

References

- Barton, Nicholas H, Alison M Etheridge, and Amandine Véber. 2017. “The Infinitesimal Model: Definition, Derivation, and Implications.” *Theoretical Population Biology* 118: 50–73.
- Lynch, Michael, and Bruce Walsh. 1998. *Genetics and Analysis of Quantitative Traits*. Vol. 1. Sinauer Sunderland, MA.