The infinitesimal model for polyploid and mixed-ploidy populations

Arthur Zwaenepoel

Contents

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_{ij} 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 $V_{ij} = V_i + V_j$. 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 kth 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

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

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_{ij} 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 idividual 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

$$V_{ij} = V_i + V_j = \frac{V_0}{2}(1 - F_i) + \frac{V_0}{2}(1 - F_j) = V_0\Big(1 - \frac{F_i + F_j}{2}\Big) \tag{1}$$

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 homoygosity states grows according to the partition function $(1, 2, 3, 5, 11, 15, 22, \ldots)$. If we represent the probability of being in these five increasingly homozygous states as $\delta_1, \ldots, \delta_5$, we find that the segregation variance is reduced by a factor

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

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

$$F_i = \Phi_{kl} \qquad \text{diploids}$$
$$F_i = \frac{1}{6}(F_k + F_l + 4\Phi_{kl}) \qquad \text{tetraploids}$$

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

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

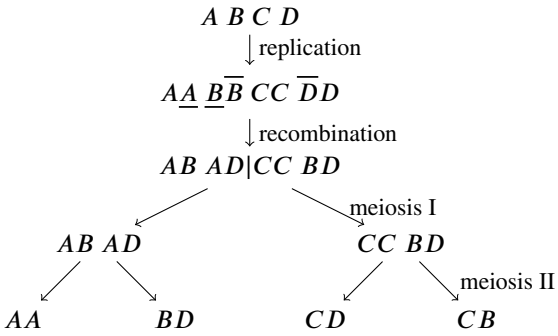
The recursion for the coancestry coefficients in m-ploids is given by

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

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 $X_1X_2X_3X_4$, 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 $\mathrm{var}[X] = v_0/2$

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

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

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

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

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, let $F_k^* = F_k(1-\alpha) + \alpha$ be the probability that a sampled gamete from tetraploid individual k contains IBD genes at a locus, the previous recursive relation for the inbreeding coefficient in tetraploids becomes:

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

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.

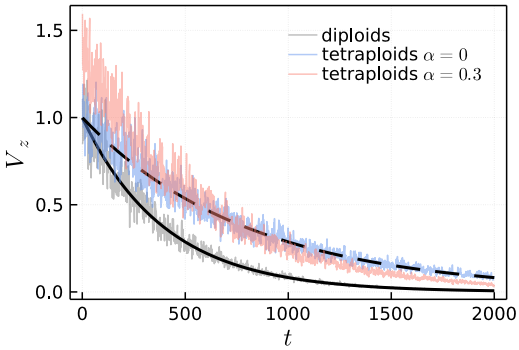


Figure 1: Decline of the phenotypic variance (V_z, here equal to the additive genetic variance) in Wright-Fisher populations ($N = 200$) simulated according to the infinitesimal model for different ploidy levels. The solid and dashed black lines show the (approximate for tetraploids) expected exponential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

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. Throughout the following paragraphs, we specialize to a diploid – autotetraploid complex, possibly allowing for triploid hybrids, where polyploids originate through the fusion of unreduced gametes.

Scaling of genetic variance across ploidy levels

The equilibrium phenotypic variance $V_z = \mathrm{var}[Z]$ under the infinitesimal model was derived above as $2V_0$, where V_0 is the segregation variance. This argument is independent of the ploidy level, as long as the population is of a single (and even-ploid) cytotype. In diploids, this entails that the variance of the additive effect X of a randomly sampled haploid genome from the base population is $V_x = \mathrm{var}[X] = V_0$, whereas for tetraploids this would be $V_x = V_0/2$. Indeed, in general, we have for m-ploids under infinitesimal assumptions that $2V_0 = V_z = \mathrm{var}[X_1 + \cdots + X_m] = mV_x$. When considering a mixed-ploidy system, we hence shall have to make additional assumptions on how the genetic variance scales across ploidy levels. If we assume equal equilibrium variances (and concomitantly segregation variances), the genetic variance for a haploid genome in tetraploids will be halve 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, the following relationships hold in the infinitesimal limit:

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

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 $\beta = 2^{-1/2}$. On the other hand, assuming equal allelic effects ($\beta = 1$) entails that the phenotypic variance 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 $V_{z,m} = \beta_4^2 V_{z,2}$, keeping in mind its interpretation as a scaler of allelic effects in the Mendelian system.

When two diploids with trait values z_i and z_j produce tetraploid offspring through unreduced gametes, we shall hence have the following offspring trait distribution

$$Z_{ij} \sim \mathcal{N}\left(\beta_4(z_i + z_j), \beta_4^2 V_{ij}\right)$$

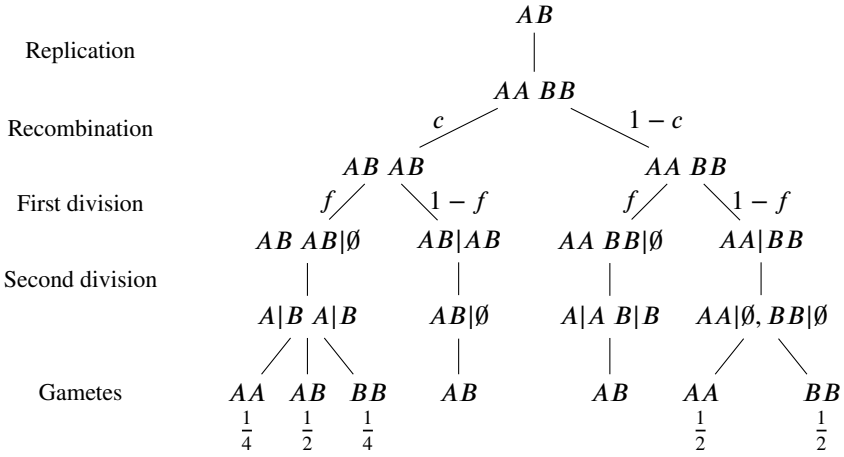
where the relevant V_{ij} of such a cross is derived in the following section. When a triploid is formed through the union of a reduced and unreduced gamete from i and j respectively, we similarly have

$$Z_{ij} \sim \mathcal{N}\left(\beta_3\left(\frac{z_i}{2} + z_j\right), \beta_3^2 V_{ij}\right)$$

It remains to be seen whether such assumptions can actually be motivated empirically.

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 abberation 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. 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 will be

$$Y = \begin{cases} 2X_1 & \text{w.p. } p_1 = \frac{1}{4}fc + \frac{1}{2}(1-f)(1-c) \\ 2X_2 & \text{w.p. } p_1 \\ X_1 + X_2 & \text{w.p. } p_2 = 1 - 2p_1 \end{cases}$$

and, conditional on X_1 and X_2 not being IBD, we find that, by the law of total variance, $\text{var}[Y] = 4p_1V_x$. Defining $\xi = 2p_1$, the segregation variance contributed by an unreduced gamete of individual i is hence

$$V_{i,22} = 2(1 - F_i)\xi V_x$$

where V_x will depend on the cytotype of the zygote to which this gamete (potentially) contributes (e.g. $\beta_4^2 V_{x,2}$ in the tetraploid case).

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 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. Meiosis, if it happens, usually results in aneuploid gametes, as there are no know mechanisms to coordinate the assortment of chromosomes in for instance a haploid and diploid gamete (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, Hamilton, and Petes 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 $(1/2)^5 \approx 0.03$. 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.

Assuming that aneuploid gametes do not contribute to viable gametes or crossings, we shall hence need to make certain assumptions on what percentage of meioses render euploid ($1n, 2n$ or $3n$) gametes. Triploid gametes 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.

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

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

$$F_i = \frac{1}{6}\left(F_k^* + F_l^* + 4\Phi_{k,l}\right)$$

where now, assuming no triploid gametes exist in the system,

$$F_k^* = \begin{cases} F_k(1-\xi) + \xi & \text{if } m_k \in \{2,3\} \\ F_k(1-\alpha) + \alpha & \text{if } m_k = 4 \end{cases}$$

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

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

For Φ the recursion above remains valid, but where diagonal elements are now given by

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

Segregation variances

gamete	$1x$	$2x$
cytotype		
$2n$	$(1-F)\frac{V_0}{2}$	$2(1-F)\xi V_0$
$3n$	$(1-F)\frac{4V_0}{9}$	$(1-F)(\frac{1}{3}+\xi)\frac{4V_0}{3}$
$4n$	·	$(1-F)(1+2\alpha)\frac{V_0}{2}$

References

Barton, Nicholas H, Alison M Etheridge, and Amandine Véber. 2017. "The Infinitesimal Model: Definition, Derivation, and Implications." *Theoretical Population Biology* 118: 50–73.

Charles, Jordan St, Monica L Hamilton, and Thomas D Petes. 2010. "Meiotic Chromosome Segregation in Triploid Strains of Saccharomyces Cerevisiae." *Genetics* 186 (2): 537–50.

Lynch, Michael, and Bruce Walsh. 1998. *Genetics and Analysis of Quantitative Traits*. Vol. 1. Sinauer Sunderland, MA.

Ramsey, Justin, and Douglas W Schemske. 1998. "Pathways, Mechanisms, and Rates of Polyploid Formation in Flowering Plants." *Annual Review of Ecology and Systematics*, 467–501.