

# **POLYSTAB**

**INDIVIDUAL-BASED MODELING TO STUDY POLYPLOID  
ESTABLISHMENT FROM AN ECO-EVO PERSPECTIVE**

Word count: 23054

**Michael Van de Voorde**

Student number: 00902198

Promoter(s): Prof. Dr. Yves Van de Peer

Supervisor: Arthur Zwaenepoel

A dissertation submitted to Ghent University in partial fulfilment of the requirements for the degree of Master of Science Bioinformatics: Systems Biology

Academic Year: 2020- 2021



## Preface

“Don't slide down the rabbit hole. The way down is a breeze but climbing back's a battle.”  
— Kate Morton, *The Clockmaker's Daughter*

This master dissertation has been an adventure through a landscape full of alluring rabbit holes. Sometimes it was a struggle trying to focus on the relevant aspects instead of getting lost in the vast space of interesting theory that exists on evolution and ecology. This resulted in having to rush the writing of this dissertation more than I would have liked. However, it is also exactly the interdisciplinarity that lies at the heart of this dissertation that made it interesting in the end. The combination of approaching biological problems from an integrated eco-evo perspective using (mathematical) modeling, of which I hope to illustrate the potential for studying the problem of polyploid establishment in this dissertation, and trying to implement these models myself in silico: especially in the beginning it was a steep wall to climb, but in the end intellectually stimulating all the more. And I daresay that I learned a lot from this experience.

Especially in the beginning of this dissertation I drifted regularly through unexplored territory. However, I didn't undertake this journey blindfolded. I want to thank my promoter prof. Yves Van de Peer but especially my supervisor Arthur Zwaenepoel for their invaluable help. Arthur proposed to me the idea for this dissertation that allowed me to wander through the glittering landscape of eco-evo modeling, while leaving room for a lot of personal freedom for personal exploration. I also thank him for introducing me to the programming language Julia and assisting me with the coding of the model, for providing many helpful insights in the results but most of all for the precious guidance and discussions that were always intellectually stimulating.

## **List of abbreviations**

**IBM:** Individual-based model

**HWLE:** Hardy-Weinberg and linkage equilibrium

**LOH:** Loss of heterozygosity

**MCE:** Minority cytotype exclusion

**OV:** Offspring viability

**UG:** Unreduced gamete(s)

## **Abstract**

Polyplloid organisms, characterized by having more than two complete sets of chromosomes, have been observed to be common in nature. However, theoretical models indicate that their establishment in natural populations is not straightforward since polyploids face considerable challenges for their establishment in an existing diploid population attributed to the effect of minority cytotype exclusion. Observations from polyplloid occurrence in nature indicate that spatio-temporal heterogeneity might play an important role in polyplloid establishment but current studies that implemented spatial heterogeneity to model polyplloid establishment are scarce and have important limitations. Likewise, the changes in key population genetic processes like genetic drift and quantitative genetic aspects of polyploids may also play an important role. An individual-based model was implemented inspired by some recent eco-evolutionary models that allows to study these different aspects in an integrated way. While this dissertation ended up more as an exploration of what individual-based models in an integrated eco-evolutionary framework can contribute to studying polyplloid establishment than as an in-depth analysis, some aspects of the implemented model are possibly a promising avenue for further research. One of the main observations from simulations is that the effects of ploidy level on drift are very pronounced. Genetic drift erodes genetic variance much less rapidly in tetraploids compared to diploids. However, it is important to note that in a constant environment a higher genetic variance also comes with a cost in the presence of stabilizing selection. Also unreduced gamete formation comes with a cost, a maladaptive load that could also be called cytotype load, was clearly present in the simulations. Lastly, an important effect of assortative mating on the probability of tetraploid establishment was shown. In extenso an interesting finding was that there seems to be a differential effect between self-fertilization and assortative mating attributed to their different mechanics for shielding from migration load. One-dimensional and two-dimensional habitats were briefly explored but no clear results can be drawn from those yet. Further expanding the current individual-based model in a more integrated multi-scale modeling approach could be worthwhile for further research.

## **Table of contents**

Preface .....	1
List of abbreviations.....	2
Abstract.....	3
Table of contents .....	4
1    Introduction .....	6
1.1    A general introduction on polyploid establishment .....	6
1.2    Some aspects of polyploid population genetics .....	10
1.3    Observations on the occurrence of polyploidy nature .....	12
1.4    Modeling polyploid establishment .....	13
1.5    Eco-evo modeling with spatial heterogeneity .....	15
1.6    Forward-time simulations and individual-based models.....	18
2    Research question.....	20
3    Material and methods .....	22
3.1    Model description.....	22
3.1.1    Genetic architecture .....	23
3.1.2    The genotype-phenotype map .....	23
3.1.3    The environment.....	25
3.1.4    The life cycle.....	27
3.1.5    Migration.....	29
3.2    Specificities for large scale simulations and data analysis.....	29
3.2.1    Establishment in a single mixed-ploidy deme.....	29
3.2.2    Establishment in a source-sink scenario .....	30
3.2.3    Establishment in a linear habitat .....	30
3.3    Code availability .....	30
3.4    Acknowledgements.....	31
4    Results and discussion .....	32
4.1    Single deme dynamics.....	32
4.1.1    Neutral single ploidy random mating population .....	32
4.1.2    Single ploidy population with stabilizing selection and density dependence .....	34
4.1.3    A note on the genotype-phenotype map .....	36
4.1.4    Mixed-ploidy populations in a single deme .....	40
4.1.5    A note on cytotype load.....	43
4.1.6    Effect of breeding system: Assortative mating and self-fertilization .....	45

4.1.7	Some observations from a single deme with changing environment .....	49
4.2	Multiple demes with migration .....	50
4.2.1	Mainland-island model .....	50
4.2.2	Migration with a single migrant.....	51
4.2.3	Continuous migration .....	53
4.2.4	Effect of breeding system: Assortative mating and self-fertilization .....	56
4.2.5	Effect of genotype-phenotype map.....	59
4.3	Linear habitat.....	61
4.4	Habitat in two dimensions .....	66
5	Conclusion.....	69
6	Bibliography .....	71

# 1 Introduction

## 1.1 A general introduction on polyploid establishment

Polyploid organisms, characterized by having more than two complete sets of chromosomes, have been observed to be common in nature and their abundance is likely underestimated due to the presence of cryptic polyploids that are difficult to recognize by morphological traits alone (Barker et al., 2016). Different mechanisms that result in polyploidy formation have been described (Ramsey et al., 1998)). Allopolyploids result from interspecific hybridization. These are often contrasted with autopolyploids. The focus of this dissertation lies on autopolyploid formation through the union of unreduced gametes resulting from gametic nonreduction.

One of the key questions that was phrased by Soltis et al. (2010) in their paper “What we still don’t know about polyploidy” was the following: “How do new polyploids establish and then persist in natural populations?”. It is this very question that lies at the origin of this dissertation. While polyploidy may indeed be prevalent in nature, especially in plant species, theoretical models indicate that their establishment in natural populations is not straightforward. It is important to keep in mind what is meant by establishment. Grisswold (2021) uses the following definition: “Establishment means a stable population is locally supported through successful reproduction and survivorship and does not rely on unreduced gamete formation or continued dispersal from another region”.

It was already shown by Levin (1985) that tetraploids face considerable challenges for their establishment in an existing diploid population. Under the assumptions of a large random mating population, emerging tetraploids are expected to be rapidly eliminated from a population. This can be explained by a frequency-dependent behavior called minority cytotype exclusion, an effect that arises from mating between different cytotypes in a mixed-ploidy population where the minority cytotype is at a disadvantage because it is more likely to be fertilized by gametes from the more common cytotype. This intercytotype mating results, in the case of a mixed population of diploids and tetraploids, in triploid offspring that is often assumed to be sterile or to have low viability.

To formalize the principle of minority cytotype exclusion, Levin (1995) formulated a simple but elegant mathematical model: assuming a large random mating mixed-ploidy population of diploids with initial frequency  $p$  and tetraploids with an initial frequency  $1 - p$ , it can be found that in the subsequent generation the frequency of diploids, denoted as  $p'$ , will become:

$$p' = \frac{p^2}{(p^2 + q^2)}$$

The change in frequency of diploids between two generations can thus be expressed as follows:

$$\Delta p = p' - p = \frac{(p^2 q - pq^2)}{(p^2 + q^2)}$$

This system has an unstable steady state for  $p = q = 1/2$ . If  $p > q$ , the frequency of diploids will increase to the steady state  $p = 1$  and will decrease to  $p = 0$  if  $p < q$ . Under this model it is impossible for an emerging tetraploid ( $q \approx 0$ ) to establish in a diploid population ( $p \approx 1$ ).

Felber (1991) expanded this mathematical model by Levin (1995) by incorporating the formation of unreduced gametes resulting in a continuous emergence of new polyploid organisms. This model allows for a quantitative analysis of how high the frequency of unreduced gamete formation theoretically needs to be to overcome minority cytotype exclusion as formulated by Levin (1985). Again, this model makes some important assumptions: it is limited to large random mating populations, the fertility and viability of both diploids and tetraploids are the same and triploids resulting from intercytotype matings are not viable.

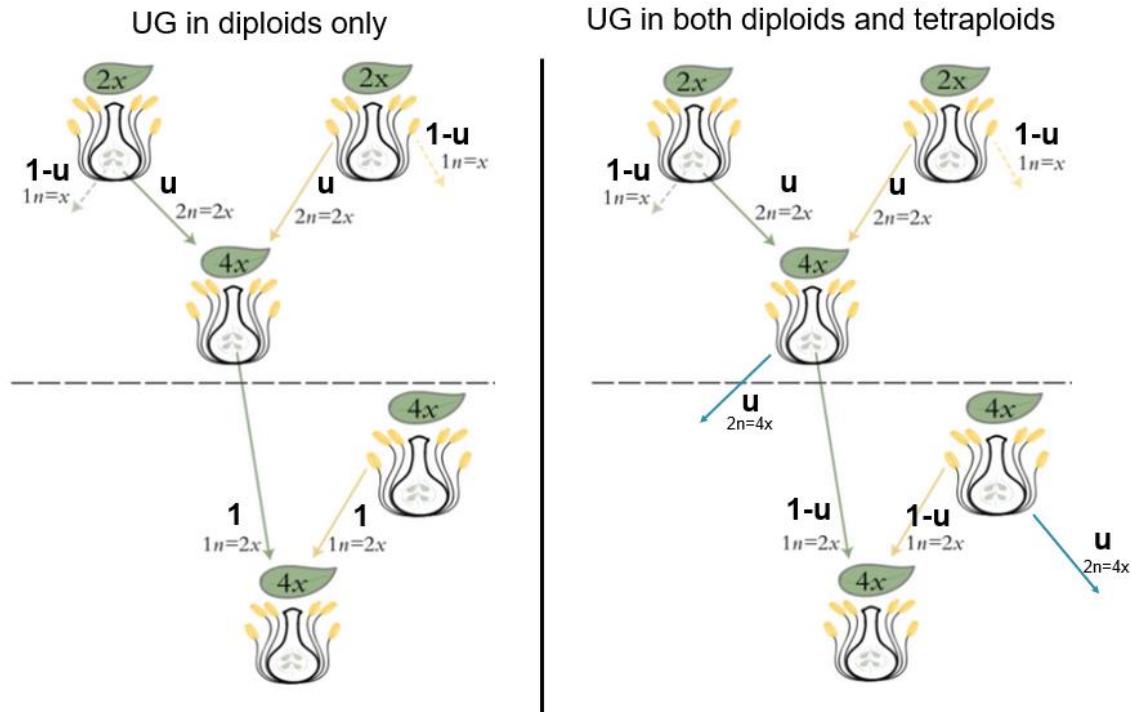
The model by Felber (1991) adds two additional parameters to the mixed-ploidy population model by Levin (1985), where  $p$  was the initial frequency of diploids and  $1 - p$  was the initial frequency of tetraploids: the parameter  $u$  denotes the frequency of unreduced gametes formed by diploid individuals and the parameter  $v$  for the frequency of unreduced gametes formed by tetraploids individuals. The recursions that follow from this model for diploids  $p'$  and tetraploids  $q'$  will be as follows:

$$p' = \frac{((1-u)p)^2}{c}$$

$$q' = \frac{(up + (1-v)q)^2}{c}$$

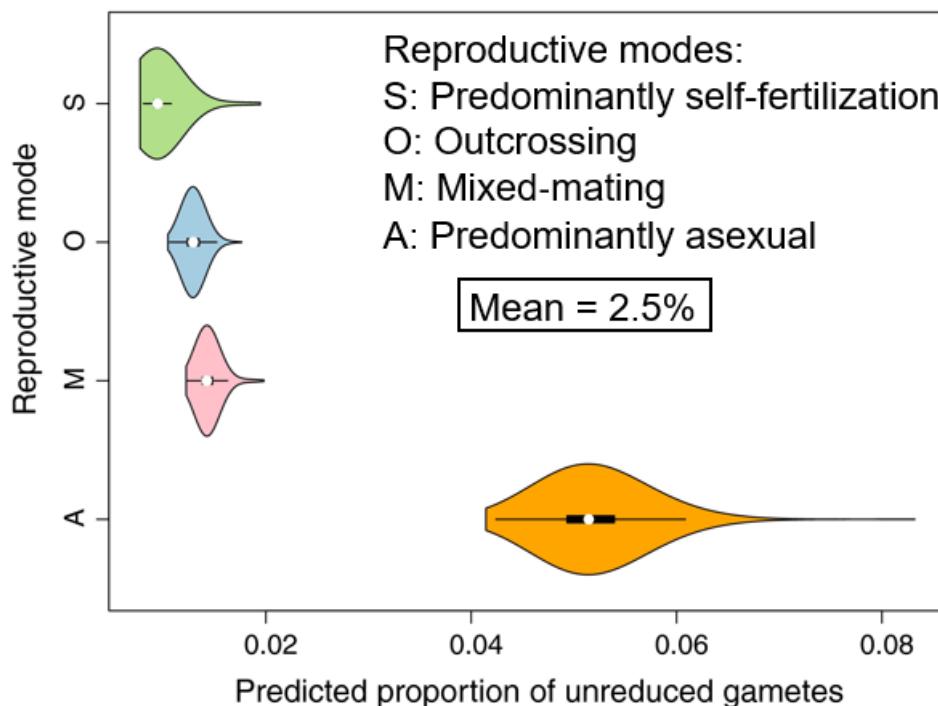
where  $c = ((1-u)p)^2 + (up + (1-v)q)^2$ .

Felber (1991) analyzed two specific cases where either  $v = 0$  (there is no unreduced gamete formation in tetraploids) or  $v = u$  (the frequency of unreduced gametes formed by tetraploids is equal to that in diploids). Assuming  $v = 0$  and  $p' - p \approx 0$ , the tetraploid population will start to increase, (i.e.  $\Delta q > 0$ ), independent of the initial frequencies, from values of  $u \approx 0.17$  and higher. Setting  $v = u$ , will increase this critical frequency of unreduced gametes to  $u \approx 0.2$ . Figure 1 shows a visual summary unreduced gamete formation as parametrized in the model by Felber (1991).



**Figure 1. Autotetraploid formation through the union of unreduced gametes resulting from gametic non-reduction.** This figure shows how (unreduced) gametes combine assuming triploids and higher ploidy are unviable. On the left the case is shown for the model by Felber (1991) where only diploids form unreduced gametes with a frequency denoted by parameter  $u$ . On the right the case is shown for the model by Felber (1991) where both diploids and tetraploids produce unreduced gametes. Adapted from Kreiner et al. (2017).

Figure 2 shows the frequency of unreduced pollen production found by Kreiner et al. (2017) using flow cytometry in over 24 species from the Brassicaceae family. They estimated the mean frequency unreduced gamete production overall to be around 2.5%. However, this result greatly depends on the mode of reproduction and is mostly found to be high for clonally reproducing species. A mean frequency of unreduced gamete formation of 0.05, and even lower if only taking sexually reproducing species into account, is quite a bit lower than the frequencies of unreduced gametes for establishment between 0.17 and 0.20 that followed from theoretical predictions by the model from Felber (1991) as discussed before.



**Figure 2. Reproductive mode was significantly related the frequency of production across species.** Results using flow cytometry in over for 24 species from the Brassicaceae family. The mean frequency of unreduced gamete production overall was estimated to be around 2.5%. However, this result greatly depends on the mode of reproduction and is mostly found to be high for clonally reproducing species, followed by selfing, mixed mating, and then outcrossing species. Adapted from Kreiner (2017).

In situations with unreduced gamete frequencies that approximate those that can be found in nature, it could be hypothesized that polyploids have a large competitive advantage in terms of fitness, enabling establishment despite minority cytotype exclusion. A lot of attention has been given to the importance of abiotic factors in polyploid establishment. This is, among other things, inspired by the observation that large polyploidization events tend to co-occur with more extreme environments or global scale climatic events and mass extinctions like the Cretaceous–Paleogene extinction event 66 million years ago (Moura et al., 2021; Van de Peer et al., 2021). Several hypotheses have been formulated in an attempt to explain these observations. First of all, stressful conditions have been shown to increase the rate of unreduced gamete formation directly resulting in higher probabilities of polyploidization events (Ramsey et al., 1998). In addition, polyploids are also hypothesized to have more adaptive potential to abiotic stress (drought, temperature, ...) through increased genetic variation resulting from genome duplication (Van de Peer et al., 2017; Levin, 2020; Van de Peer et al., 2021).

Recently also more thought has been given on the importance of biotic interactions in polyploid establishment. Segraves (2017) highlights the importance of studying polyploidy in the context of community ecology and the properties of ecological networks. For example interactions have been described between changes plant traits such as nectar production or flower size associated with polyploidy causing pollinator shifts (Van de Peer et al., 2017; Rezende et al., 2020). This could enhance polyploid establishment through reproductive isolation from

the founding diploid population. It is likely that to come to a cohesive theory of polyploid establishment a unified eco-evolutionary framework considering both biotic and abiotic factors on different scales (from gene to ecosystem) will have to be considered (Ramsey et al., 2014; Fox et al., 2020).

For completeness also the ratchet hypothesis is mentioned here. According to this hypothesis a neutral mechanism is able to (partly) explain the prevalence of polyploids (Meyers and Levin, 2006). Since polyploid species only produce more polyploids, whereas diploid species can produce both diploid and polyploid species through unreduced gamete formation, the ratchet hypothesis states that polyploidization is a one-way street from low to high ploidy levels. The abundance of polyploids is then merely considered a mechanistic result of unreduced gamete formation (Ramsey and Ramsey, 2014; Levin, 2020). However, return to diploid state (rediploidization) on longer time scales through loss of duplicated genes makes it possible to escape from this ratchet (Li et al., 2021).

## 1.2 Some aspects of polyploid population genetics

Autopolyplody through genome doubling has been described to have some profound effects on the level of population genetics of which the most interesting are probably related to quantitative genetics and genetic variance, although the number of publications on these aspects is scarce. This part tries to summarize some of their major findings.

Moody et al. (1991), already proposed a mathematical model that indicated that for autotetraploids, as direct effect of the increase in chromosome number (polysomy), heterozygosity is expected to degrade slower than for diploids in an idealized population (under the assumptions of large random mating populations with discrete and non-overlapping generations and without selection, mutation or migration). Using coalescent based theory Arnold et al. (2012) showed that in an idealized population (and without factoring in complicating mechanism like double reduction in tetraploids) the effective population size of a tetraploid population can be approximated as double to that of a diploid population. The effective population size  $N_e$  is the population size of idealized population would have in order to behave like an observed population of size  $N$ . As can be seen from the formulas for the expected loss of heterozygosity as shown below (Moody et al. (1991), Lynch and Walsh (1998)) (effective) population size  $N_e$  is an important determinator for the rate of erosion of heterozygosity and loss of heterozygosity is expected to be slower in tetraploids compared to diploids.

$$H_t = H_0 \left(1 - \frac{1}{N_e}\right)^t \quad (\text{haploid})$$

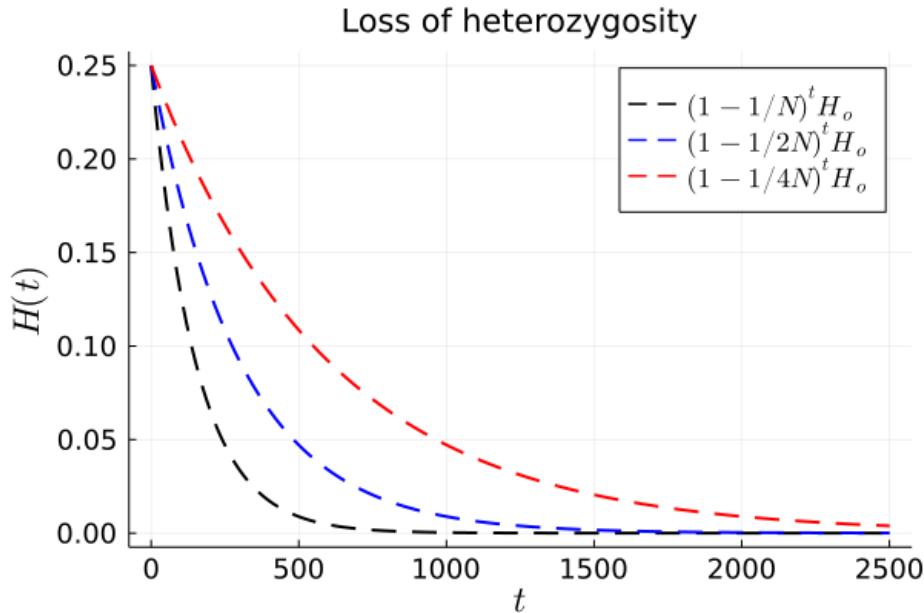
$$H_t = H_0 \left(1 - \frac{1}{2N_e}\right)^t \quad (\text{diploid})$$

$$H_t = H_0 \left(1 - \frac{1}{4N_e}\right)^t \quad (\text{tetraploid})$$

where  $H_0$  is the starting heterozygosity, assumed to be the same for all ploidy levels, and  $H_t$  is the heterozygosity after  $t$  generations. These can be written in a more general form as:

$$H_t = H_0 \left(1 - \frac{1}{kN_e}\right)^t$$

where  $k$  is the ploidy level and  $t$  is the number of generations. Though it remains illusive to apply this formula for ploidy levels higher than tetraploids. These theoretical predictions are visualized in figure 3.



**Figure 3. Loss of heterozygosity for different ploidy levels.** The curves show loss of heterozygosity for different ploidy levels (haploid in black, diploid in blue and tetraploid in red) in a neutral random mating population (without factoring in complicating mechanism like double reduction), where the effective population size of a tetraploid population is approximated as double to that of a diploid population.

Otto, S. P., & Whitton, J. (2000) hinted at another interesting effect in polyploids. Due to a higher number of chromosomes, for recessive traits longer masking of deleterious recessive mutations can be expected. Therefore, it is of paramount importance to take the genotype-phenotype map (in short this is a function that takes as input the genotype of an individual and returns as output its phenotypic value) or genetic model into account when studying polyploid establishment. The choice of additive, dominant or recessive gene models is expected to have important implications for example on how selection on a trait works in polyploids compared to diploids (Otto and Whitton, 2000; Griswold, 2021).

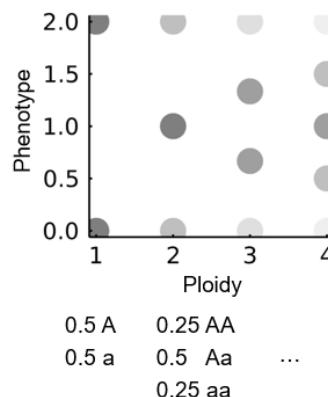
A last important aspect to be mentioned is that in autopolyploids the different chromosome copies are assumed to be sufficiently similar to form multivalent complexes. They tend to line up during meiosis, forming a tetravalent in the case of tetraploids. If a crossover occurs between replicated arms of two of the four chromosomes during meiosis, the two identical gene copies from the sister chromatids can segregate in the same gamete. A parent with a genotype  $A_1A_2A_3A_4$  may thus occasionally produce a gamete with the genotype  $A_1A_1$  (Lynch and Walsh, 1998; Dufresne et al., 2014).

Expected genotype frequencies for a bi-allelic locus in Hardy-Weinberg equilibrium are predicted by the formula (Dufresne et al., 2014):

$$(p + q)^k$$

where  $p$  and  $q$  are the frequencies of both allelic states and  $k$  is the ploidy level.

In the diploid case this gives  $p^2 + 2pq + q^2$ . In the tetraploid case this becomes  $p^4 + 4p^3q + 6p^2q^2 + 4qp^3 + q^4$ . Note that the frequency of extreme phenotype is smaller in tetraploids ( $p^4$ ) compared to diploids ( $p^2$ ) (since  $p < 1$ ). This is especially important for a genotype-phenotype map that assumes the phenotypic range of diploids and tetraploids is the same as shown for a single bi-allelic locus figure 4. The phenotypic values in tetraploids will tend to get squeezed between the largest and smallest phenotypic value, that are expected to be much less frequent in HWLE for tetraploids than for diploids.



**Figure 4. Example of genotypic frequencies in HWLE for different ploidy levels.** A bi-allelic locus is considered with a phenotypic effect size  $\alpha = 1$  in a diploid individual. The phenotypic range is assumed to be the same for all ploidy levels. Details on how the phenotypic value is calculated can be found in materials and methods. The shading of the dots corresponds to the relative abundance of the phenotype (higher frequencies are darker).

With double reduction the expected frequencies of homozygous genotypes increase causing a deviation from HWLE predictions (Dufresne et al. (2014)). Including double reduction decreases the effective population size and further erodes genetic variance, comparable to the effects of inbreeding. However, according to Arnold et al. (2012) double reduction occurs only rarely so it will not be taken into further consideration for this dissertation.

### 1.3 Observations on the occurrence of polyploidy nature

Some associations have been described between polyploids (mostly in plants) and their geographical distribution. An interesting observation is that polyploids do not seem to be evenly distributed geographically but appear to be distributed unevenly on a global scale. Using phylogenetic biogeography methods Rice et al. (2019) showed that polyploids are more often associated with lower temperature regions at higher latitudes (either through increased rate of unreduced gamete formation that is environmentally dependent (Ramsey et al., 1998) or

through higher adaptive potential). In the same vein, Baniaga et al. (2020) found niche divergence to be faster in polyploids compared to diploids and they hypothesize this to be due to the effect of polyploidization on genetic variance and selection. Husband et al. (2013) summarized some of the main observations concerning the distribution of polyploids in a species' range, most importantly being that polyploidy is possibly related to a larger range size and that polyploids are thought to be tolerant to more extreme conditions.

The hypothesis that polyploidy is positively correlated to a larger range size can in some cases be attributed to the adaptive potential of polyploids due to higher genetic variance (Otto and Whiton, 2000; Sheth, 2020). Genetic variance in polyploids allows for adaptation over a larger range relates back to slower loss of heterozygosity as explained in the previous part. Also, a larger phenotypic range due to a dosage effect could allow for colonization of more extreme phenotypes in polyploids (Otto and Whiton, 2000). Another possible factor is that polyploids are less vulnerable to gene flow from diploids, but this depends on the amount of intercytotypic gene flow (Husband et al., 2013). Lack of intercytotypic gene flow shields arising polyploids at range edges from maladaptive gene flow from the center in models where genetic variance is fixed (Kirckpatrick and Barton, 1997), tough gene flow also has a positive effect on genetic variance as has been shown in later models (Barton, 2001).

An important remark made by Rice et al. (2019) however is that there is also an association between climatic region and life cycle. There is for example a higher abundance of perennial plants at higher latitudes. Van Drunen et al. (2019) showed that perenniallity (typically plant species that live for multiple years) is evolutionary associated with polyploidy. Perenniallity, by increasing intracytotype mating through longer persistence, at higher latitudes could potentially enhance polyploid establishment through weakening of minority cytotype exclusion, in addition to stress centered hypotheses.

The same accounts for clonal reproductive systems: there is even a stronger correlation between the evolution of clonal reproductive systems in plants and polyploidy (Van Drunen et al., 2019). Clonal reproduction may facilitate polyploid establishment by bypassing the frequency dependent effect of minority cytotype exclusion altogether. Alternatively, as Kreiner et al. (2017) assume unreduced gamete formation in sexual reproducing species to be maladaptive. The frequency of unreduced gamete formation in clonal species has been observed to be very high (up to 80%) possibly due to the lack of selection pressure on unreduced gamete formation in asexual species. However, the focus of this dissertation will remain on sexual reproduction.

## 1.4 Modeling polyploid establishment

Observations from the previous part are mostly based on statistical association. Polyploid establishment is difficult to observe directly in nature. There are however several publications where the authors tried implementing mathematical models to learn something about important underlying mechanistical processes that could be of relevance to polyploid establishment.

The models by Levin (1975) and Felber (1991) were already referred to in the beginning of this introduction but are conceptually very important since they introduced the concept of

minority cytotype exclusion and the important dynamics resulting from unreduced gamete formation. In order to explain the possibility of successful establishment of polyploids in the face of minority cytotype exclusion, theoretical models were studied that do not make at least one of the following assumptions: large populations, random mating, equal survival of different cytotypes or inviable intermediate cytotypes (here restricted to triploids).

Niche differentiation is an example of assortative mating (mostly implemented as a higher probability for intracytotype matings) that has been found to be a key factor in polyploid establishment (Fowler & Levin (1984); Rodriguez (1996); Oswald & Nuismer (2011)). Niche differentiation evidently leads to less competition between polyploids and their diploid progenitors, shielding them from minority cytotype exclusion.

Spoelhof et al. (2020) modelled mixed-ploidy populations in a two-dimensional uniform space to study the effect of habitat shape on polyploid establishment. They hypothesized that spatial structure of a population through limited dispersal could result in non-random mating patterns that might show to be important in overcoming minority cytotype exclusion, i.e. through assortative mating in local clusters of isoploidy. This is similar to what was already shown by Baack (2005). They also assumed neutrality, absence of an inherent fitness difference between different ploidy levels. The models of Spoelhof et al. (2020) showed that habitats that approximate a linear shape (long but small) might enhance polyploid establishment, something the authors attribute to spatial blocking because of density dependence. This is reinforced through both life cycle aspects like increased self-fertilization or clonal reproduction and limited dispersal. An important limitation in the modeling approach of Spoelhof et al. (2020) is that they did not implicitly model the genetics, making it impossible to study important dynamics like spatial gene flow within and between cytotypes.

Griswold (2021) implemented a theoretical model for tetraploid establishment in a diploid population with spatial dynamics and incorporated the genetics of the system explicitly as a single bi-allelic locus with different possible modes of gene action (additive, dominant and recessive). The spatial aspect was modelled as a central core population with migration to a periphery with an offset in optimal phenotype. This approach allowed to study important mechanics resulting from a spatial context like (maladaptive) gene flow directly. Griswold (2021) concluded that polyploid establishment is prominent when the rate of gamete dispersal is high (i.e. in the presence of high levels of gene flow) and for recessive modes of gene action in combination with moderate to high rates of self-fertilization. This paper beautifully illustrates how the interaction of spatial dynamics, life cycle aspects and gene action can facilitate polyploid establishment. The proposed mechanism behind this is that self-fertilization shields tetraploids from minority cytotype exclusion and further decreases maladaptive gene flow from the center of the habitat in combination with inbreeding depression being less pronounced in tetraploids relative to diploids, especially with recessive gene action. This relates back to the slower loss of heterozygosity in tetraploids compared to diploids as was explained before.

## 1.5 Eco-evo modeling with spatial heterogeneity

As alluded to above spatio-temporal heterogeneity might play an important role in polyploid establishment but current studies that implemented a spatial dimension and/or spatial heterogeneity are scarce and have important limitations (Li et al., 2004; Baack, 2005; Spoelhof et al., 2020; Griswold, 2021). Likewise, the changes in key population genetic processes like genetic drift and quantitative genetic aspects may also play an important role, greatly depending on the gene mode (additive, dominant or recessive). Lastly, other life cycle and other ecological factors, such as assortative mating (e.g. due to pollinator shift), perennially, self-fertilization, etc. may further affect the probability of successful establishment. Implementing an integrative model of evolution in mixed-ploidy populations that captures all the biologically relevant element is a considerable challenge. This chapter will explore some recent more complex integrated spatial eco-evolutionary models, albeit limited to single ploidy populations, that try to integrate quantitative polygenic evolution in a spatial context.

Eco-evolutionary models try to joint model different aspects of both ecological and evolutionary processes combined with special attention to eco-evolutionary feedbacks that link both aspects since evolution can be rapid, especially on standing genetic variation (Govaert et al., 2019). Incorporating more realistic assumptions about the genetic architecture of traits and genotype-phenotype map could be important to model aspects possibly relevant in polyploid establishment, such as the evolution of genetic variation and phenotypic variance due to selection, migration and stochastic effects like random drift and demographic stochasticity (Szép et al., 2021). Concerning the genetic architecture of traits, one approach that models including eco-evolutionary feedback mechanisms over time can take is by using a quantitative genetics approach to study the capacity of a population to adapt to environmental change in a spatio-temporal context (Govaert et al., 2019).

Many traits show continuous variation instead of discrete phenotypic states. This continuous variation in phenotypes is (partly) determined by a continuous genetic variation instead of discrete genotypes. Studying the evolution of this genetic variation is at the heart of quantitative genetics. An important theoretical assumption often made in quantitative genetics for mathematical convenience is that phenotypes follow a normal distribution with mean  $\mu$  and variance  $\sigma^2$ . This can be augmented for because many traits in nature are approximated by a normal distribution (Lynch & Walsh, 1998; Rice (2004)). In this dissertation the focus will lie on additive genetic variance (effects caused by dominance and epistasis will not be studied).

Additive genetic variance is often defined as the total phenotypic variance caused by the additive effects of the genes of a certain polygenic trait (Rice, 2004) as shown by the following formula:

$$V_A = h^2 V_P$$

where  $V_A$  is the additive genetic variance,  $h$  is the heritability and  $V_p$  is the total phenotypic variance. Heritability here is defined as the regression of the offspring phenotype on the parent phenotype and indicates the strength of inheritance of a trait (narrow-sense heritability (Rice, 2004)).

In natural populations significant levels of additive genetic variance in traits can be observed (Brady et al., 2019). This standing genetic variation implies the potential for adaptation to be rapid and effective, e.g. in response to climate change (Cotto et al., 2018) and the importance of it in a spatio-temporal context will be discussed later on.

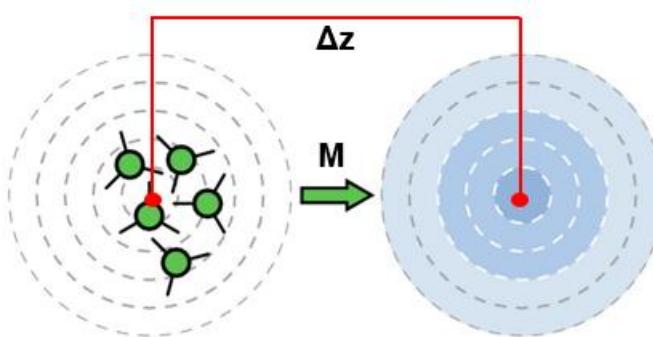
Maladaptation is an important concept throughout the rest of this dissertation, here defined as a mismatch between the phenotype of an individual  $z$  and the optimal phenotype of the environment  $\theta$  it resides in (Crespi, 2000). To illustrate this a bit more, the example of a trait under stabilizing selection can be given. The fitness of an individual under (Gaussian) stabilizing selection is given by

$$W(z) = \exp\left(\frac{-(z - \theta)^2}{2V_s}\right)$$

where  $z$  is the phenotype of the individual,  $\theta$  is the phenotypic optimum of the environment and  $V_s$  is the variance of stabilizing selection. It can readily be observed that the fitness of an individual decreases if its phenotype deviates more from the environmental optimum.

Important to note is that, when considering a quantitative trait with a population phenotypic mean and variance, both the mean and the variance of that trait can cause maladaptation. The phenotypic mean can be displaced from the optimum of the environment but also the phenotypic variance can deviate from the optimal variance. Note that the optimal variance under stabilizing selection alone is  $\sigma^2 = 0$  where the phenotype of all individuals perfectly matches the environment, but this is different in spatio-temporal settings (where the phenotypic optimum changes either through space or time) (Brady et al. (2019)).

Maladaptation can be attributed to different forms of genetic load (Brady et al., 2019). One type of load that is especially important in the presence of gene flow for spatially structured populations is migration load as shown in figure 5. Migration load can be conceptualized as the combination of two parameters: the difference in trait optima between source and sink  $\Delta z$  and migration rate  $M$ .



**Figure 5. A spatial setting with migration load.** Migration load can be conceptualized as the combination of two parameters: the difference in trait optima between source and sink  $\Delta z$  and migration rate  $M$ . Adapted from Brady et al. (2019).

As reviewed by Holt et al. (2001) several migration scenarios can be modeled reaching from a relatively simple back-hole sink model with one way migration from a source to a sink population to complex heterogeneous landscapes. Barton & Etheridge (2018) modeled migration in a source-sink with unidirectional migration from source to sink scenario using a quantitative genetics approach to study under which conditions (mostly in the presence of maladaptive gene flow in combination with stochastic effects like random drift) a population can establish itself in a new habitat. The source population is a large founder population that is maladapted to the new environment to which there is migration. The effect of drift and maladaptive gene flow are expected to be highest right after the founder event. If the growth rate is positive, trade-off between selection, here on a single polygenic trait, and gene flow, the effects of drift and maladaptive gene flow will decrease resulting in a positive feedback. In this setting different migration rates and difference in maladaptation are found to have an important effect on establishment. The time to establishment was found to increase with increasing migration load caused by a difference in phenotypic optima. For different migration rates some optimum is expected due to a trade-off between the maladaptive effects of migration and the positive effects of gene flow on genetic variance (Barton & Etheridge, 2018; Bridle et al., 2019).

Relating back to the observation that polyploids are associated with marginal habitats there could be interesting relationship between polyploid establishment and species range theory that studies important processes affecting range limits and the causes of range limits in nature by trying to combine features of environmental variation, evolution, and genetics (Connallon and Sgrò, 2018).

Limiting to studies on range expansion for quantitative traits under stabilizing selection, the important effect of gene flow on the adaptation to peripheral conditions of a habitat was already emphasized by García-Ramos and Kirkpatrick (1997). They specifically showed that gene flow due to migration from the center of a linear habitat to the periphery caused deviations in the mean phenotype of the peripheral populations from the local phenotypic optima. A limitation of their model however, is that it assumed a constant genetic variance over the whole habitat. Kirkpatrick-Barton model (1997) modelled joint changes in population size and the trait and came to the same results: using a model with a fixed genetic variance gene flow due to migration along a phenotypic gradient could have only maladaptive effects.

Barton (2001) developed a deterministic model that combines changes in population size, trait mean and allele frequencies, allowing genetic variance to evolve instead of treating it as constant over the whole linear habitat. Analysis of this model showed that positive effects of gene flow on additive genetic variance for a quantitative trait tend to overwhelm its negative effects of maladaptive gene flow, leading to the possibility of continued range expansion. Bridle et al. (2019) summarizes the effect of migratory gene flow in a heterogeneous environment to be twofold for quantitative traits: it has a tendency to reduce the mean fitness of the population but to increase local genetic variance which might increase adaptive potential. This is double-edged sword of migration is similar to what Barton & Etheridge (2018) showed for a source-sink scenario (see explanation there).

Polechová & Barton (2015) added stochasticity to this mixture through the inclusion of genetic drift and demographic stochasticity since stochastic effects might play an important role at the expansion front during range expansion where peripheral population sizes are smaller. One of the main results they found is that genetic drift erodes genetic variation beneath a level that is needed to overcome adaptation to a heterogeneous environment, limiting the continuous range expansion that was found using models without stochastic effects. The authors showed the following relationship holds:

$$B \geq 0.15N\sigma \sqrt{\frac{\alpha^2}{2V_s}}$$

where  $B$  is the environmental gradient,  $N$  is the population size,  $\sigma$  is the dispersal per generation and  $\alpha$  is the allelic effect size and  $V_s$  is the variance of stabilizing selection per locus.

Polechová (2018) extended the model from Polechová & Barton (2015) to a two-dimension habitat with a linear gradient. An important consequence of this extra dimension for migration is that This is because in two dimensions, dispersal mitigates the loss of diversity due to genetic drift more effectively, such that it becomes (almost) independent of selection.

To conclude, models for quantitative traits that also model allele frequencies and look at genetic variance show a link between quantitative genetic variation and adaptive potential. This genetic variation is possibly reduced near the edges of an expanding species range resulting locally in less adaptive potential (Pennington, 2021 but see Polechová & Barton, 2015 for mechanistic model). Additive genetic variance can be a proxy for the adaptive potential of a quantitative trait (cfr. changing environment, link with climate change, gradient, etc.). Inability to adapt can lead to population extinction (extinction vortex). What determines rate of adaptation? Standing genetic variation, life cycle (generation time, selfing related to inbreeding depression, selection pressure (difference in phenotypic optima), migration (gene flow) ...). Though high additive genetic variance does not necessarily correlate with adaptive potential (Hoffmann et al., 2017).

## 1.6 Forward-time simulations and individual-based models

The mostly used paradigms for simulating large scale genetic data are backward-time and forward-time simulations (Yuan et al., 2012). It is however out of the scope of this introduction to discuss these methods in detail. The focus will mostly lie on the applicability of these methods to study eco-evolutionary dynamics of mixed-ploidy populations in the context of polyploid establishment.

A first method that is often used for the simulation of genetic data are back-ward time or coalescent based simulations (Yuan et al., 2012). There is some limited work on coalescent models for polyploid populations, see for example Arnold et al. (2012), for an extension of coalescent theory to tetraploids or Monnahan et al. (2020) who studied the population genetic signals of hard sweeps in polyploids. However, as stated by Bradburd et al. (2019), coalescent based theory is lacking to quantitatively model more complex scenarios combining

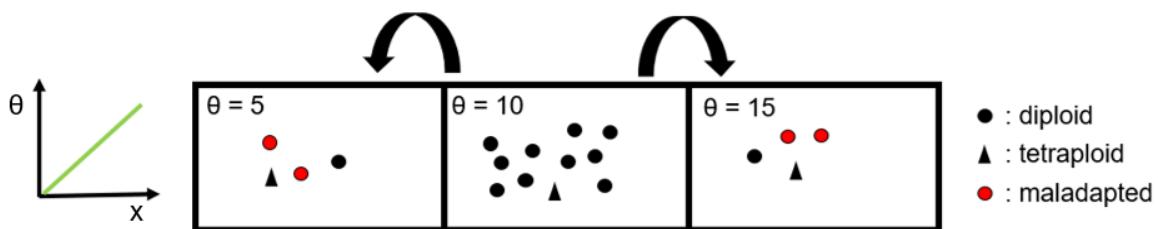
spatial dynamics, selection and life cycle aspects of which the effect on polyploid establishment are studied in this dissertation.

Forward-time simulations are an interesting alternative approach to tackle eco-evolutionary research questions in silico (Carcajal-Rodriguez, 2010). They allow simulating more complex scenarios than backward-time simulators but are generally slower, especially when simulating large populations, because they are often individual-based (Andrello et al., 2021). Individual-based models are a widely adapted bottoms-up modelling approach in eco-evo research in which a population can be modelled as a number of discrete interacting individuals. Each individual can be characterized by a set of state variables such as location and a genetic architecture that make it possible to study, amongst other things, the evolution of genetic variance that might be relevant for polyploid establishment (DeAngelis and Mooij, 2005; Bradburd et al., 2019). Such a modelling approach was for example implemented by Cotto et al. (2017) to study the role of adaptive potential in the response of alpine plants to climate warming. Several forward-time simulators exist such as SLiM 3 (Haller et al., 2019) and Nemo (Neuenschwander et al., 2019) but these are restricted to haploid and diploid organisms and none of these currently allow for simulations with mixed-ploidy populations for higher ploidy levels. Since individual-based models are likely the most suitable method at hand to study more complex scenarios with mixed-ploidy populations in the quest of trying to solve some of the fundamental questions related to polyploid establishment (being able to combine spatial modelling, a quantitative genetic background, selection, life-cycle aspects,...), it will be one of the aims of this dissertation to manually implement such a model.

## 2 Research question

The general aims of this dissertation could be broadly summarized as an attempt at combining aspects from species range theory, polyploid population genetics and existing polyploid establishment theory in an eco-evolutionary quantitative genetics framework with the main goal of tackling some important questions concerning polyploid establishment. A modelling approach will be used to study these questions, more specific through the implementation of an individual-based model that allows for simulations of mixed-ploidy populations at spatio-temporal scales. Since no current simulators exist that are practically suitable for this, (large parts of) the IBM will have to be coded manually and will need to be tested and validated.

One observation that was elaborated on in the introduction is that polyploid populations are sometimes associated with peripheral habitats of a species range. Figure 6 tries to illustrate the most important dynamics that play at the edge of a linear habitat as can be for example observed during species range expansions. An important aspect is the presence of maladaptive gene flow resulting from migration of individuals from a well-adapted, say diploid, central population to the periphery that is assumed to have a different phenotypic optimum. Second, due to the dynamics of range expansion, this edge population is expected to have a lower population size. Small population sizes are typically associated with more stochasticity, like for example genetic drift.



**Figure 6. Illustration of a spatial setting that summarizes some of the aspects hypothesized to be play a role in polyploid establishment.** This is an example of a three-deme landscape with a linear phenotypic gradient ( $\theta = 5x$ ) that causes spatial heterogeneity and migration from the center to the periphery. In the center of the habitat the diploid population is well adapted and high in population size making it unlikely for polyploids to establish due to MCE. During species range expansion however, the peripheral population is expected to be small and maladapted due to migration load from the center. It is hypothesized that under these peripheral conditions tetraploid establishment might be facilitated.

The allow formulating some concrete hypotheses the goal of this dissertation was narrowed down to studying polyploid establishment in a “neutral” framework, meaning that it is assumed that polyploids don’t have an inherent fitness advantage compared to diploids. Relating the observations from the previous paragraph back to the problem of polyploid establishment, some hypotheses can be formulated of why the specific conditions that exist at in peripheral habitats might enhance polyploid establishment in such a neutral framework. The first hypothesis is that, since polyploids that establish in marginal demes are semi-isolated from diploids, i.e. there is only direct intercytotype gene flow from diploids to polyploids through unreduced gametes, they would be less susceptible to maladaptive gene flow from

a central diploid population. Shielding polyploids from maladaptive gene flow could be further enhanced through life-cycle aspects like self-fertilization or forms of assortative mating. With low population sizes at the periphery, more demographic stochasticity, and less minority cytotype exclusion might be expected. Likely a trade-off will be present between the maladaptive effect of gene flow on diploids in the periphery and the frequency effect of minority cytotype exclusion (i.e. more migration results in higher migration load but also a higher frequency of diploids flooding peripheral polyploids).

Another aspect is that low population size at the periphery will likely result in higher genetic drift that erodes genetic variance. Based on the theoretical predictions for the loss of heterozygosity as explained in the introduction, it could be hypothesized that genetic variance erodes less in polyploids allowing for further range expansion than diploids. The genotype-phenotype map that is implemented, and the gene mode (additive, dominant, recessive) will likely also be important in this aspect since it directly influences the genetic variance.

Some of the initial hypotheses of why tetraploid establishment could be more prevalent in periphery can be summarized as follows:

1. Tetraploids in periphery are semi-isolated and less susceptible to maladaptive gene flow from the center.
2. Small population size results in more pronounced stochasticity (demographic stochasticity and genetic drift) and affects minority cytotype effect.
3. A difference in genetic variance of tetraploids compared to diploids (depending on the genotype-phenotype map) could influence local adaptive potential and evolution of genetic variance after a founder event.
4. There is an interaction with life cycle aspects (self-fertilization, assortative mating), type of selection and gene mode (additive, dominant, recessive).

Based on these hypotheses, the individual-based modelling approach allows investigating under what specific assumptions polyploids have a higher probability of establishment by studying the effect of some important parameters: the environmental gradient, migration rate, frequency of unreduced gamete formation, quantitative genetics and the genotype-phenotype map, assortative mating and self-fertilization.

### 3 Material and methods

#### 3.1 Model description

An individual-based model (IBM) was implemented in the programming language Julia (Bezanson et al., 2012) that allows forward genetic simulations of mixed-ploidy populations under varying conditions. Some basic aspects of this IBM were inspired by an eco-evolutionary model for the evolution of a species range in a haploid population by Polechová & Barton (2015) that allows for spatio-temporal modeling joint evolution of population size and allele frequencies. The parameters used in the current implementation of the IBM used for the simulations in this dissertation are summarized in table 1.

**Table 1. Overview of the parameters used in the model.**

Parameter	Level of implementation	Description
$k$	Agent	Ploidy level: {1,2,3,4}
$d$	Agent	Allelic effect scaler
$\alpha$	Deme	Allelic effect size
$L$	Deme	Number of loci per single genome copy
$K$	Deme	Local carrying capacity
$\theta$	Deme	Local phenotypic optimum
$rm$	Deme	Growth rate coefficient
$Vs$	Deme	Variance of stabilizing selection
$\mu$	Deme	Mutation rate
$V$	Deme	Matrix with offspring viability
$G$	Deme	Matrix with frequencies of (unreduced) gamete formation
$u$	Deme	Frequency of UG formation in diploid
$v$	Deme	Frequency of UG formation in tetraploid
$a$	Deme	Frequency of assortative mating
$s$	Deme	Frequency of self-fertilization
$\beta$	Deme (Island)	Strength of directional selection
$M$	Deme (Island)	Migration rate
$\sigma$	Habitat	Dispersal per generation
$b$	Habitat	Steepness of environmental gradient

Dm	Habitat	Number of demes
----	---------	-----------------

### 3.1.1 Genetic architecture

A fundamental aspect of the model is that it explicitly encodes the genome of each individual for a single additive quantitative trait. This allows for exact tracking of the evolution of allele frequencies through space and time (from which other statistics like phenotypic mean and variance can be calculated). Figure 7 shows such a single polygenic trait in a tetraploid individual that is determined by 9 additive unlinked loci. Loci are biallelic with an allelic effect size  $\alpha$  that is taken to be identical over all loci and for all individuals in a simulation.

<i>L</i> loci →								
↓ <i>K</i> chromosomes								
0	0	0	0	0	0	1	1	1
0	0	0	0	0	1	1	1	1
0	0	0	1	1	1	1	1	1
0	0	1	1	1	1	1	1	1

**Figure 7. Illustration of a genome of a single tetraploid individual.** This genome carries 4 chromosomes ( $k = 4$ ) with an allelic effect size  $\alpha = 1$  and number of loci  $L = 9$ .

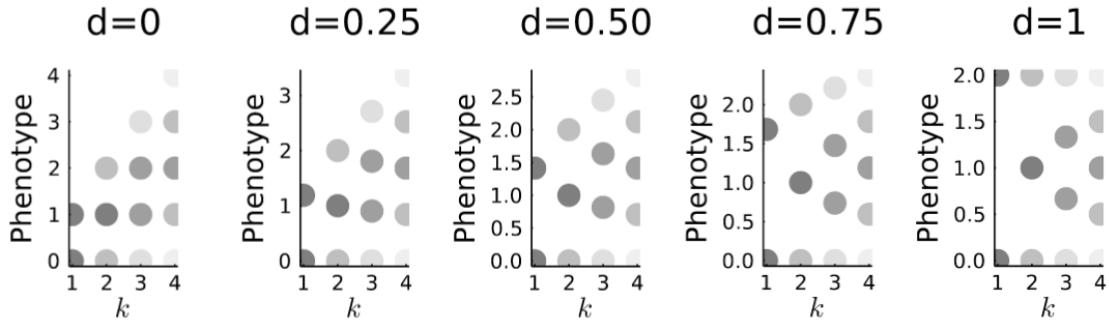
### 3.1.2 The genotype-phenotype map

How the phenotype of an individual is calculated and how this scales between different ploidy levels depends on the underlying genotype-phenotype map. In this model a parameter  $d$  was introduced to scale the phenotypes for different ploidy levels. The genotypic value  $G$  of an individual is then calculated using the following genotype-phenotype map:

$$G = k^{-d} \sum_{i=1}^k \sum_{j=1}^L \alpha_{i,j}$$

where  $k$  is the ploidy level,  $\alpha$  is the allelic effect size and  $L$  is the number of loci. As an example, the phenotype of the individual in the previous part (figure 7) would be simply calculated as:  $G = 4^{-1} \sum_{i=1}^4 \sum_{j=1}^9 1_{i,j} = 5$  (with  $d = 1$ ).

Figure 8 shows the implications of this map and the effect of the parameter  $d$  on the phenotypic mean and variance for a single bi-allelic locus.



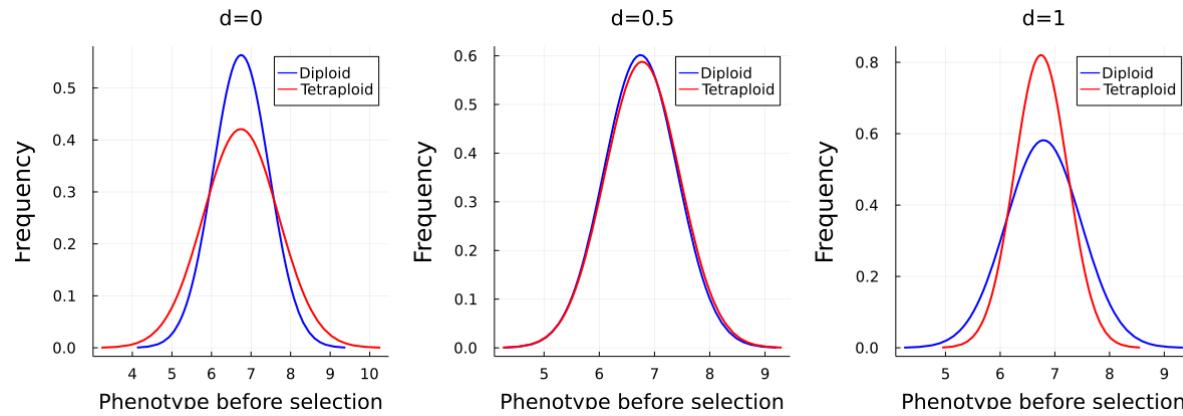
**Figure 8. Scaling of the phenotypic range for a single bi-allelic locus for different values of  $d$ .** The shading of the dots corresponds to the relative abundance of the phenotype (higher frequencies are darker). Note that polyploids do have a larger range of possible genotypic values despite the reduction in additive genetic variance for values of  $d < 1$ .

The additive genetic variance of a diploid population, as derived under conditions of HWLE, is defined as:

$$V_A = 2\alpha^2 \sum_{i=1}^L p_i(1 - p_i)$$

With this implementation additive genetic variance ( $V_A$ ) scales with a factor  $(k/2)^{(1-2d)}$ , where  $k$  is the ploidy level. Comparing the genetic variance of tetraploids to diploids at HWLE with this model for the extreme cases  $d = 0$  and  $d = 1$ , it can be found for  $d = 0$  that the genetic variance is double for tetraploids compared to diploids ( $V_A = k/2$ ), while for  $d = 1$  the genetic variance of tetraploids is only half that of diploids ( $V_A = \frac{k}{2}$ ).

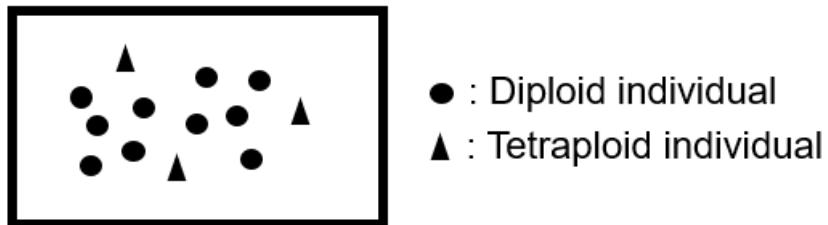
Populations were initiated with  $p = 0.5$ , the gene frequency that maximizes the starting mean heterozygosity and in extenso the additive genetic variance and with phenotypic mean equal to the phenotypic optimum of the starting deme. Figure 8 shows the effect of different values of parameter  $d$  on the phenotypic variance of a starting population under these conditions. Note that most simulations were done with  $d = 1$  for which the phenotypic range for all polyploidy levels is identical as shown in figure 9, unless explicitly stated otherwise. This implies that the phenotypic variance for an initialized tetraploid population is expected to be smaller, according to theoretical predictions more precisely only half the variance of a diploid population.



**Figure 9. Effect of genotype-phenotype map for different values of  $d$  on starting phenotypic variance.** Important to note here is how the phenotypic variance of tetraploids compared to diploids scales for different values of  $d$ . The following parameters were used for the simulations:  $p = 0.5$ ,  $\alpha = 0.25$ ,  $L = 100$ .

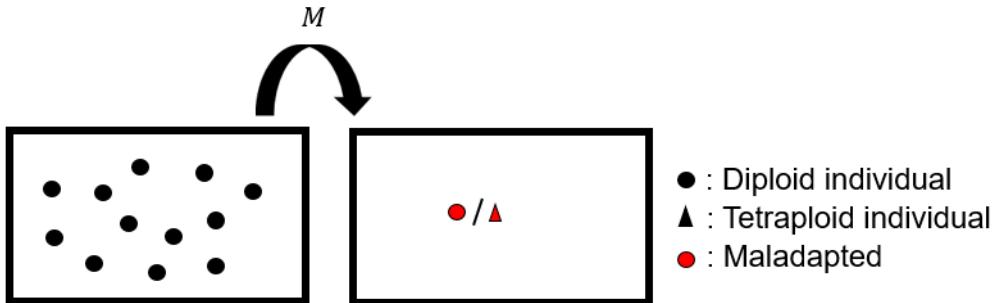
### 3.1.3 The environment

The central building block of the environment is termed a deme, which is populated by individual organisms. In such a deme individuals follow a general life cycle consisting out of selection, mutation, mating with recombination and dispersal that is described further on. Important parameters that are defined at the level of a deme are mentioned in table 1. Specifically worth mentioning here is each deme has a single phenotypic optimum. Figure 10 shows a simplified visualization of a mixed-ploidy deme with a population of co-existing diploids and tetraploids. Many simulations in this dissertation are limited to this single deme environment.

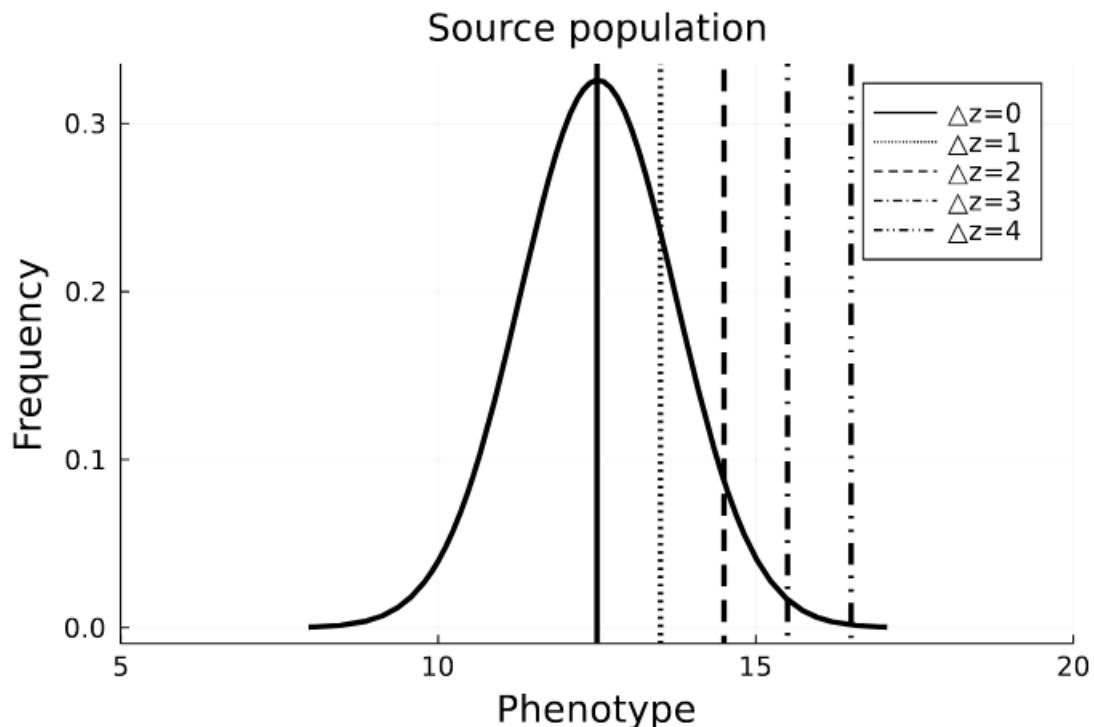


**Figure 10. Example of a mixed-ploidy deme as the central building block of the environment in the model.**

A second spatial scenario used throughout this dissertation is a source-sink scenario (or mainland-island model). Figure 11 shows that this environment exists out of two demes: one deme with an infinite population assumed to be in HWLE (source) and a second deme that is empty (sink). There is unidirectional migration with a migration rate  $M$  (explained further on) from the source to the sink and typically the phenotypic optimum of the source will be different from that of the sink, so migrants are considered to be maladapted. Figure 12 shows the phenotypic mean and variance of the founder population ( $\Delta z = 0$ ) and the differences in phenotypic optima between source and sink that were used for simulations (varying from  $\Delta z = 1$  to  $\Delta z = 4$ ).

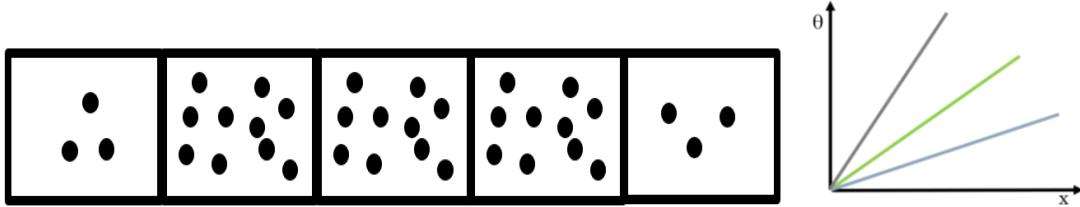


**Figure 11. Source-sink scenario with migration and a difference in phenotypic optima.** There is migration from an idealized founder population of diploids (left) to a sink with a different phenotypic optimum (right) with migration rate  $M$ .



**Figure 12. Difference in phenotypic optima between source and sink.** The differences in the phenotypic optimum the founder population (vertical full black line) and the different sink environments were taken to be constants since all simulations are started from the same conditions ( $p = 0.5$ ,  $\alpha = 0.3$ ,  $L = 50$ ,  $d = 1$ ,  $\beta = 0.2$ ). For more flexibility it would have been more appropriate to select the difference relative to the phenotypic variance of the source population as described in Barton & Etheridge (2015).

A third spatial scenario is a linear habitat with a linear phenotypic gradient and is shown in figure 13. The environment is a concatenation of multiple demes in one dimension. The phenotypic optimum  $\theta$  of the demes varies according to some linear function  $\theta = bx$ , where  $b$  is the steepness of the gradient and  $x$  is a discrete spatial coordinate ranging from 0 to the maximum number of demes in the habitat ( $Dm$ ). Simulations typically start from a population in the central deme with subsequent bidirectional migration. Details on the implementation of migration are described further on.



**Figure 13. One-dimensional habitat with a linear phenotypic gradient.**

Finally, some simulations in a two-dimensional habitat were briefly explored. Extending the model that was previously described for a linear gradient to a two-dimensional space is straightforward: the two-dimensional habitat is represented as an  $m \times n$  matrix of  $mn$  demes. Migration is so far only implemented as a random walk to neighboring demes with a probability  $\sigma$  and stops at the edges of the habitat.

### 3.1.4 The life cycle

The life cycle as it is currently implemented assumes that individuals are hermaphroditic with a lifespan of a single generation (i.e. age structure is not part of current model). The life cycle consists of selection, mating, mutation and migration. Time is measured in discrete non-overlapping generations.

For directional selection the fitness  $w(z)$  of an individual in a deme was calculated according to:

$$w(z) = \beta(z - \theta)$$

where  $\beta$  is the strength of directional selection,  $z$  is the phenotype of the individual and  $\theta$  is the phenotypic optimum of the deme.

The Malthusian fitness  $r(z, N)$  of an individual was calculated as described by Polechová & Barton (2015) by summing together a density dependent term  $r_e(N)$  and a term for balancing selection  $r_g(z)$ :

$$r(z, N) = r_e(N) + r_g(z)$$

$$r_e(N) = r_m \left(1 - \frac{N}{K}\right)$$

$$r_g(z) = \frac{-(z - \theta)^2}{2V_s}$$

where  $r_m$  is the growth rate coefficient,  $N$  is the local population size of the deme,  $K$  is the carrying capacity of the deme,  $z$  is the phenotype of the individual,  $\theta$  is the phenotypic optimum of the deme and  $V_s$  is the variance of stabilizing selection.

After calculating the fitness, every individual mates with other individuals drawn from the population in the same deme in proportion to their fitness for the total number of offspring

it produces. The number of offspring for each individual is drawn from a Poisson distribution with mean and variance equal to the exponential of the fitness of that individual:

$$\text{Number of offspring} \sim \text{Poisson}(W)$$

with  $W = \exp(r(z, N))$  in the case of Malthusian fitness or  $W = \exp(w(z))$  for directional selection. This aspect of the model introduces demographic stochasticity.

Note that some simulations also are done in a neutral random mating population (thus without selection). Here mating pairs are drawn at random from the population, each producing a single offspring, until a fixed population size is reached.

For the mating itself, randomly picked gametes from two individuals are merged following recombination and unreduced gamete formation. The model assumes free recombination between all loci. The matrix for the frequency of unreduced gamete formation  $G$  can be represented as follows:

$$G = \begin{pmatrix} u_{1,1} & 0 & 0 & 0 \\ u_{2,1} & u_{2,2} & 0 & 0 \\ u_{3,1} & u_{3,2} & u_{3,3} & 0 \\ u_{4,1} & u_{4,2} & u_{4,3} & u_{4,4} \end{pmatrix}$$

where  $u_{i,j}$  is the frequency of gametes with a ploidy level of  $j$  formed in an individual with a ploidy level  $i$ . For clarification, the unreduced gamete formation matrix for model by Felber (1991) as described in the introduction with unreduced gamete formation rate  $u$  in both diploids and tetraploids would be implemented as follows:

$$G = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1-u & u & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1-u & 0 & u \end{pmatrix}$$

In addition to the matrix for the frequencies of unreduced gamete formation  $G$ , another matrix  $V$  was built in the model to differentiate between offspring viability for different ploidy levels:

$$V = \begin{pmatrix} v_{1,1} & v_{1,2} & v_{1,3} & 0 \\ v_{2,1} & v_{2,2} & 0 & 0 \\ v_{3,1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

where  $v_{i,j}$  is the probability of survival if a gamete of ploidy level  $i$  is combined with a gamete of ploidy level  $j$ . The viability matrix used for most simulations throughout this dissertation is as follows:

$$V = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

The viability of diploids resulting from the combination of 2 haploid gametes ( $\nu_{1,1}$ ) and of tetraploids resulting from the combination of 2 diploid gametes ( $\nu_{2,2}$ ) is equal. The viability of triploid offspring in this matrix is 0. To incorporate formation of a triploid bridge in the model, different values of  $\nu_{1,2}$  and  $\nu_{2,1}$  could be chosen. Important to note is that ploidy levels higher than tetraploidy will always be treated as unviable in this model.

For some simulations assortative mating or self-fertilization (selfing) were added to the life cycle. Hereto two new parameters were introduced:  $a$  is the probability of assortative mating. Assortative mating is implemented so that during mate selection there is a probability of  $a$  or less to mate with an individual of the same ploidy level. Other types of assortative mating are possible, for example on phenotypic distance between individuals (Oswald & Nuismer, 2011). Analogous,  $s$  denotes the probability for self-fertilization.

Mutation rate in the simulations is set to be small, i.e. the probability of a mutation per locus per generation is set to  $10^{-6}$ , so that its contribution to genetic variance is negligible and genetic variance across the habitat is maintained by gene flow (Polechová & Barton (2015)). Mutation is symmetrical. For bi-allelic loci as implemented in the model (see genetic architecture), this comes down to equal probabilities to mutate between both alleles.

### 3.1.5 Migration

The absolute number of migrants in the source-sink model is randomly drawn from  $Poisson(M)$ , where  $M$  is the migration rate. The following values of  $M$  were used: 0.01, 0.1, 1, 10 (corresponding to 1,2,3,4 in the plots). The migrant itself is chosen at random from a source population following a normal distribution with phenotypic mean and variance as described earlier.

The current implementation of migration in the linear habitat is very basic, i.e. nearest neighbor migration where the parameter  $\sigma$  is simply the probability of moving to a neighbouring deme (the probabilities are symmetrical to the left or right). Boundaries are impermeable but these are never reached in the linear habitat since the number of starting demes  $Dm$  is always taken to be large enough. A discretized Gaussian dispersal kernel as described in Polechová & Barton (2015) could be implemented but this wasn't properly validated so no results are shown in this dissertation using it.

## 3.2 Specificities for large scale simulations and data analysis

Parameters and other details for single simulations were described under the respective plots (see Results and discussion). More extended simulations and subsequent data analysis are described below.

### 3.2.1 Establishment in a single mixed-ploidy deme

To study the effect of the parameter  $u$  on tetraploid establishment in a single mixed-ploidy deme, a simple grid search algorithm was implemented. An interval was specified from  $u = 0$  to  $u = 0.5$  over which values were chosen uniformly with steps of  $\Delta u = 0.025$ . For each value

of  $u$ , 20 independent simulations were run for 50 generations and the probability of tetraploid establishment was calculated as the ratio of the number of simulations where a majority of the individuals were tetraploids to the total number of simulations (thus unconditional on the population size). Subsequently a logistic curve was fitted to these probabilities of tetraploid establishment for all values of  $u$  with logit regression using the Julia package *GLM*. The value  $u_{crit}$  was introduced as the value of  $u$  that corresponds to the intercept of the fitted logistic curve with the horizontal line for which the probability of tetraploid establishment is 0.5. For different probabilities of assortative mating and self-fertilization, a linear model was fit on different values of  $u_{crit}$  using linear regression (ordinary least squares) from the Julia package *GLM*.

### 3.2.2 Establishment in a source-sink scenario

For more large-scale simulations to study establishment in the source-sink scenario, mean time to establishment was calculated as the number of generations before a population with size 100 or more was established. For computational reasons 10000 generations was taken as the upper limit. If this limit of 10000 generations was reached, the population was unable to establish. For each combination of migration rate ( $M = [0.01 \quad 0.1 \quad 1 \quad 10]$ ) and difference in phenotypic optima ( $\Delta z = [1 \quad 2 \quad 3 \quad 4]$ ) 20 independent simulations were run. The mean time to establishment and the standard error on the mean were reported for these 20 simulations assuming a normal distribution. Mean tetraploid establishment was defined as the ratio of simulations where a majority of the individuals were tetraploids to the total number of simulations with an established population. Mean and the standard error on the mean ( $\sqrt{p(1-p)/n}$ ) of tetraploid establishment were reported assuming they follow a binomial distribution.

### 3.2.3 Establishment in a linear habitat

For simulations to study the effect of the environmental gradient on tetraploid establishment a simple grid search algorithm was implemented. An interval was specified from  $b = 0.1$  to  $b = 1$  over which values were picked uniformly with steps of  $\Delta b = 0.1$ . For each value of  $b$ , 10 independent simulations were run for 250 generations and the probability of tetraploid establishment was calculated as the ratio of demes with tetraploids as majority to the total number of populated demes (thus unconditional on range size). Again mean and standard error on the mean over these 10 simulations were reported. Also the mean and standard error on the mean of the total number of populated demes over 10 simulations was reported and a linear curve was fitted for different values of  $b$  using linear regression (ordinary least squares) from the Julia package *GLM*.

## 3.3 Code availability

The Julia code for the individual-based model and simulations are available on GitHub (<https://github.com/MMichaelVdV/PolyStab>). The basic aspects of the model are available as a package called *PolyStab* with some basic documentation and tests, though there is still

some work to be done to make this user friendly. Simulations are available as Pluto notebooks with some basic annotation.

### 3.4 Acknowledgements

I gratefully thank my supervisor Arthur Zwaenepoel for the hypothesis-space that was the main driver for this dissertation, for introducing the concept of cytotype load and the plot on the evolution of mean fitness in that section (figure 29), for introducing me to Julia and assisting with the coding of the model, for providing many helpful comments on draft version of the manuscript and helpful insights in the results but most of all for the precious guidance and discussions that were always intellectually stimulating.

## 4 Results and discussion

### 4.1 Single deme dynamics

#### 4.1.1 Neutral single ploidy random mating population

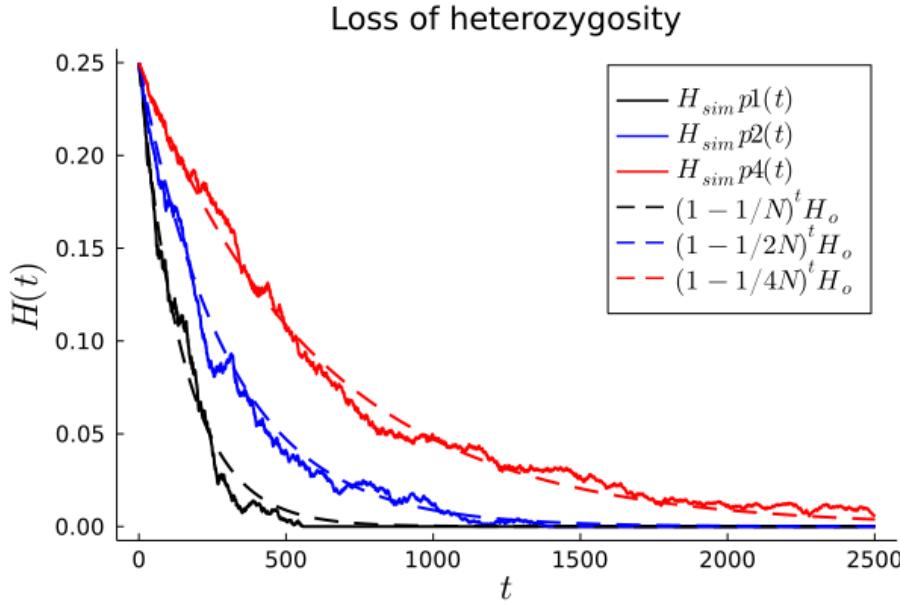
Some of the first simulations with the individual-based model were run in a single deme environment, that is without yet incorporating any spatial dynamics due to migration. The reasoning behind this was two-fold: firstly, to validate the correct implementation of the basic mechanisms of the model and secondly a lot of interesting insights could already potentially be gained on the quantitative genetics of polyploids by first studying the basic mechanics of, among other things, genetic drift and balancing selection without the complicating presence gene flow due to migration.

As described in the introduction, the effect of genetic drift might be expected to erode genetic variance slower in tetraploids compared to diploids. In fact, in the absence of double reduction, loss of heterozygosity in tetraploids can be approximated with the expression:

$$H_t = H_0 \left(1 - \frac{1}{kN_e}\right)^t$$

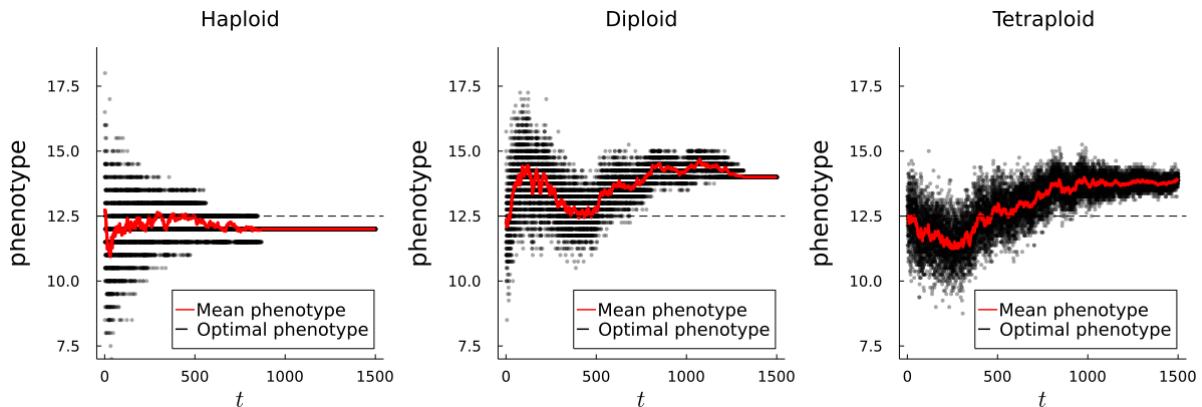
where  $k$  is the ploidy level and  $t$  is the number of generations.

To mainly observe the behavior of genetic drift for different ploidy levels, a finite neutral random mating deme, that is without any form of selection, and with fixed population size of 200 was simulated for 2500 generations. Figure 14 shows that simulations almost perfectly trace the theoretical predictions for the loss of heterozygosity.



**Figure 14. Loss of heterozygosity in a single ploidy neutral random mating deme.** The dashed lines show the theoretical predictions for the expected loss of heterozygosity in a haploid (black), diploid (blue) and tetraploid (red) population. Solid lines show the results of a single simulation for 2500 generations, showing that the expected loss of heterozygosity is well approximated. The following parameters were used for the simulations:  $K = 200$ ,  $p = 0.5$ ,  $\alpha = 0.5$ ,  $L = 50$ ,  $d = 1$ ,  $u = 0$ .

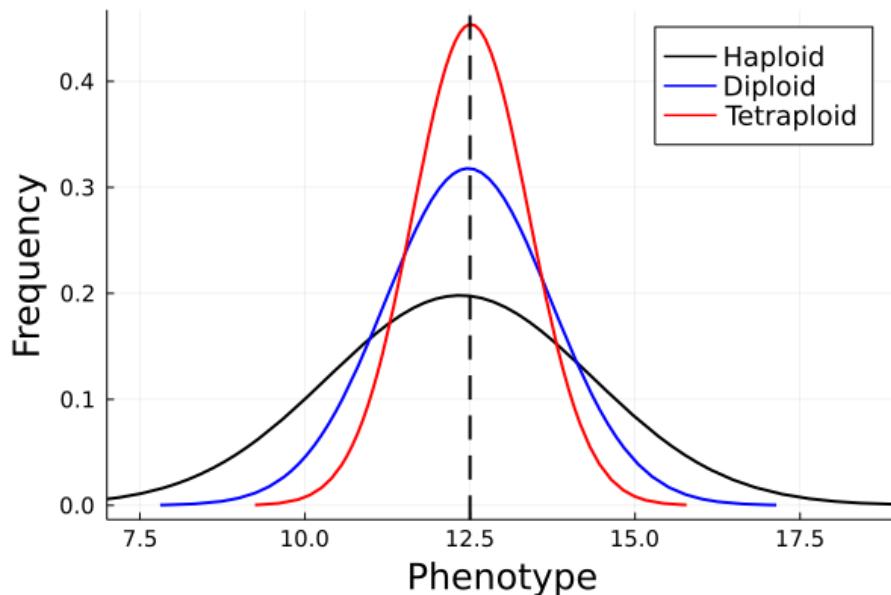
Figure 15 shows the evolution of the phenotypic mean and variance for each individual of the population (each represented by a black dot), and the mean phenotype of the population (red line) for 1500 generations. It can be observed that the phenotypic variance around the mean decreases for all ploidy levels but more rapidly so for lower ploidy levels. This agrees with the difference in loss rates of heterozygosity as observed above. The phenotypic mean can be observed to randomly drift through time. Since these simulations are in a neutral random mating deme, due to the absence of stabilizing selection it doesn't converge to any phenotypic optimum.



**Figure 15. Evolution of phenotypic mean and variance in a single ploidy neutral random mating deme.** The phenotype of each individual is represented by a black dot and the mean phenotype of the population is shown by the red line. These plots clearly show the differential loss of phenotypic variance over time due to drift for different ploidy levels. Since these are simulations in a deme without

stabilizing selection, the phenotypic mean will not converge to the optimum of the deme. The following parameters were used for the simulations:  $K = 200$ ,  $p = 0.5$ ,  $\alpha = 0.5$ ,  $L = 50$ ,  $d = 1$ ,  $u = 0$ .

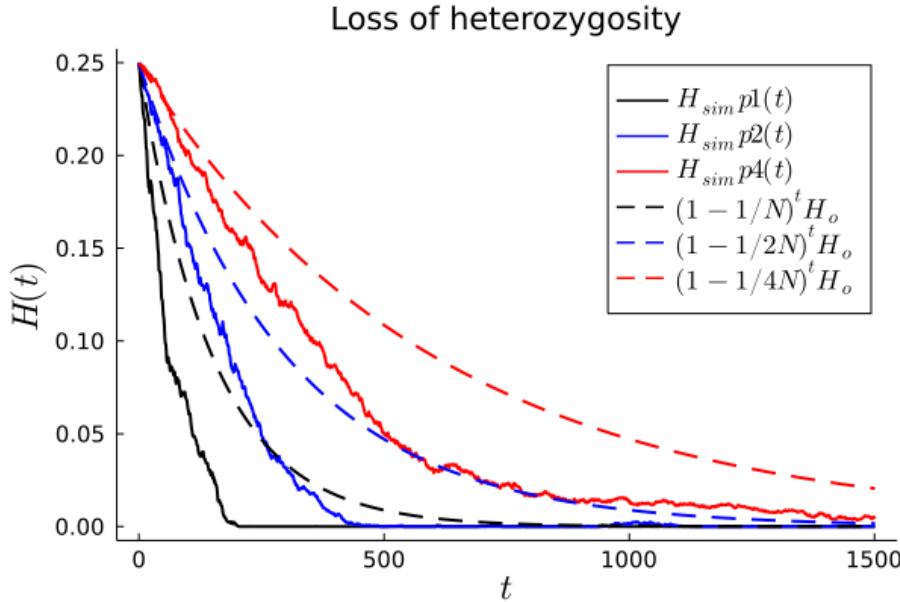
An important remark to be made concerning these simulations however, is that the starting phenotypic variance for different ploidy levels is not the same when using a genotype-phenotype map that assumes the same maximum phenotypic range over all ploidy levels ( $d = 1$ ). This is shown in figure 16. As explained in material and methods the expected phenotypic variance of a tetraploid population using in HWLE using this map is only half that of diploids, i.e. the phenotypic variance in the tetraploid population is still higher than in the diploid population after 1500 generations even though it only started with half of the genetic variance.



**Figure 16. Phenotypic mean and variance of starting population for different ploidy levels.** This plot shows the starting phenotypic mean and variance of different ploidy levels using a genotype-phenotype map that assumes the same maximum phenotypic range over all ploidy levels ( $d=1$ ). The starting phenotypic variance with this type of map decreases with increasing ploidy level. The following parameters were used for the simulations:  $K = 200$ ,  $p = 0.5$ ,  $\alpha = 0.5$ ,  $L = 50$ ,  $d = 1$ ,  $u = 0$ .

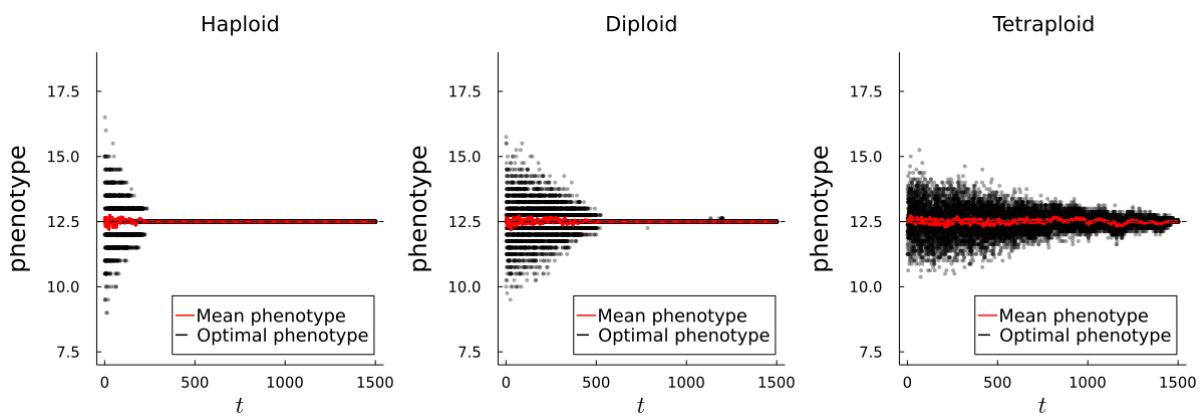
#### 4.1.2 Single ploidy population with stabilizing selection and density dependence

The preceding simulations with finite random mating populations in a single deme were subsequently extended with a form of density dependence and stabilizing selection (or together as Malthusian fitness as described in materials and methods). Figure 17 shows again a plot for the loss of heterozygosity, which shows that the addition of stabilizing selection further erodes genetic variance on top of the effect of genetic drift that was explained before. This effect of stabilizing selection has been described in the literature and can be expected (for example by Polechová (2018)).



**Figure 17. Loss of heterozygosity in a single ploidy deme with density dependence and stabilizing selection.** The dashed lines show the theoretical predictions for the expected loss of heterozygosity in a haploid (black), diploid (blue) and tetraploid (red) population. Solid lines show the results of a single simulation for 1500 generations, showing that the expected loss of heterozygosity is well approximated. The following parameters were used for the simulations:  $K = 200$ ,  $p = 0.5$ ,  $\alpha = 0.5$ ,  $L = 50$ ,  $d = 1$ ,  $u = 0$ ,  $r_m = 1.06$ ,  $V_s = 0.5$ .

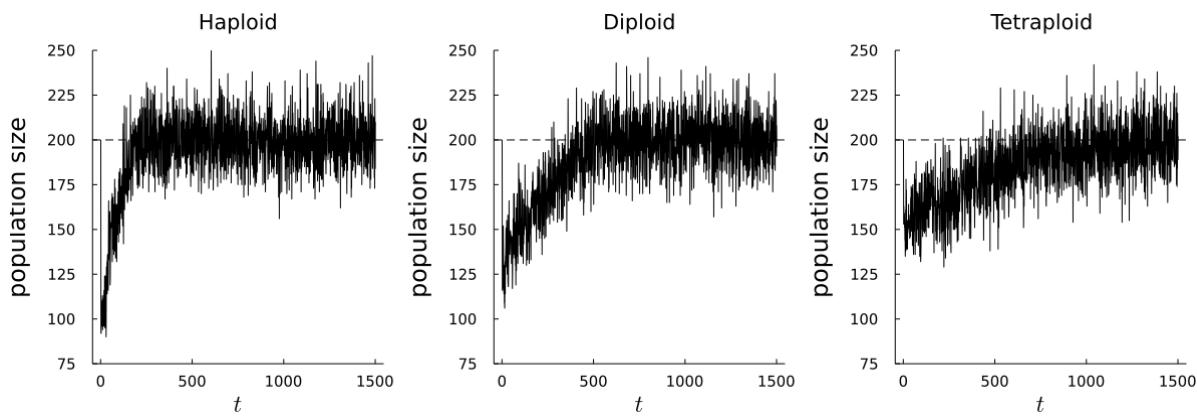
Figure 18 shows the evolution of the phenotype for each individual of the population (each represented by a black dot, and the mean phenotype of the population (red line) but now, through the effect of stabilizing selection, the mean phenotype of the population will converge to the phenotypic optimum of the deme. Again, it is obvious that phenotypic variance erodes less for higher ploidy levels.



**Figure 18. Evolution of phenotypic mean and variance in a single ploidy deme with density dependence and stabilizing selection.** The phenotype of each individual is represented by a black dot and the mean phenotype of the population is shown in red line. These plots clearly show the differential loss of phenotypic variance over time due to drift and stabilizing selection for different ploidy levels. Since these are simulations in a deme with stabilizing selection, the phenotypic mean will converge to the

optimum of the deme. The following parameters were used for the simulations:  $K = 200$ ,  $p = 0.5$ ,  $\alpha = 0.5$ ,  $L = 50$ ,  $d = 1$ ,  $u = 0$ ,  $r_m = 1.06$ ,  $V_s = 0.5$ .

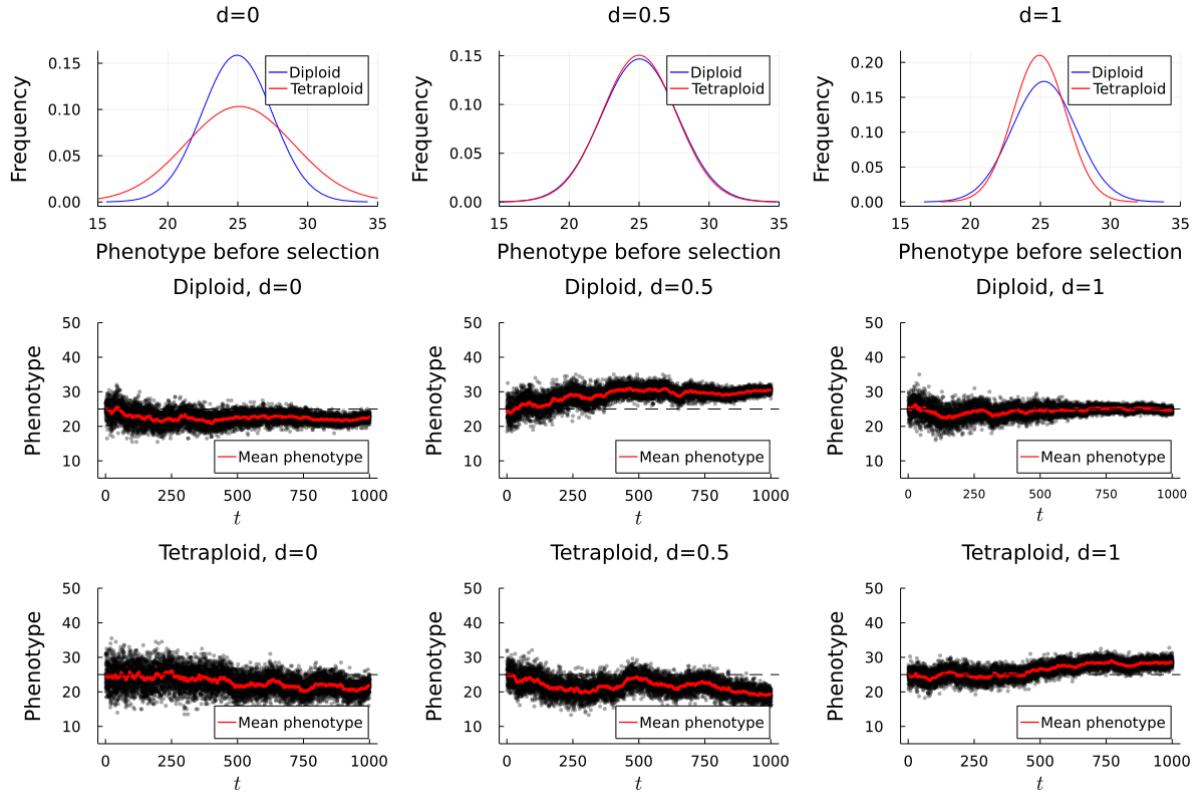
Figure 19 shows the population growth for different ploidy levels with density dependence and stabilizing selection. The population size can be seen to firstly decrease due to the presence of maladapted individuals in the population before it grows to the carrying capacity of the deme. The initial drop in population size is lower for higher ploidy levels. This is due to initiating populations with maximal phenotypic variance and with mean equal to the phenotypic optimum (as explained in materials and methods). The number of maladapted individuals greatly depends on the starting variance of the population (though it also depends on strength of selection that wasn't explored in depth in this dissertation). Since the starting phenotypic variance is the smallest in tetraploids as explained above, this drop in population size is also less pronounced than for smaller ploidy levels. This effect likely differs for values of  $d \neq 1$ .



**Figure 19. Population growth in a single ploidy deme with density dependence and stabilizing selection.** The black curve shows the evolution of population size for different ploidy levels. Population size will increase to carrying capacity  $k$  and fluctuate around the equilibrium because of demographic stochasticity. The following parameters were used for the simulations:  $K = 200$ ,  $p = 0.5$ ,  $\alpha = 0.5$ ,  $L = 50$ ,  $d = 1$ ,  $u = 0$ ,  $r_m = 1.06$ ,  $V_s = 0.5$ .

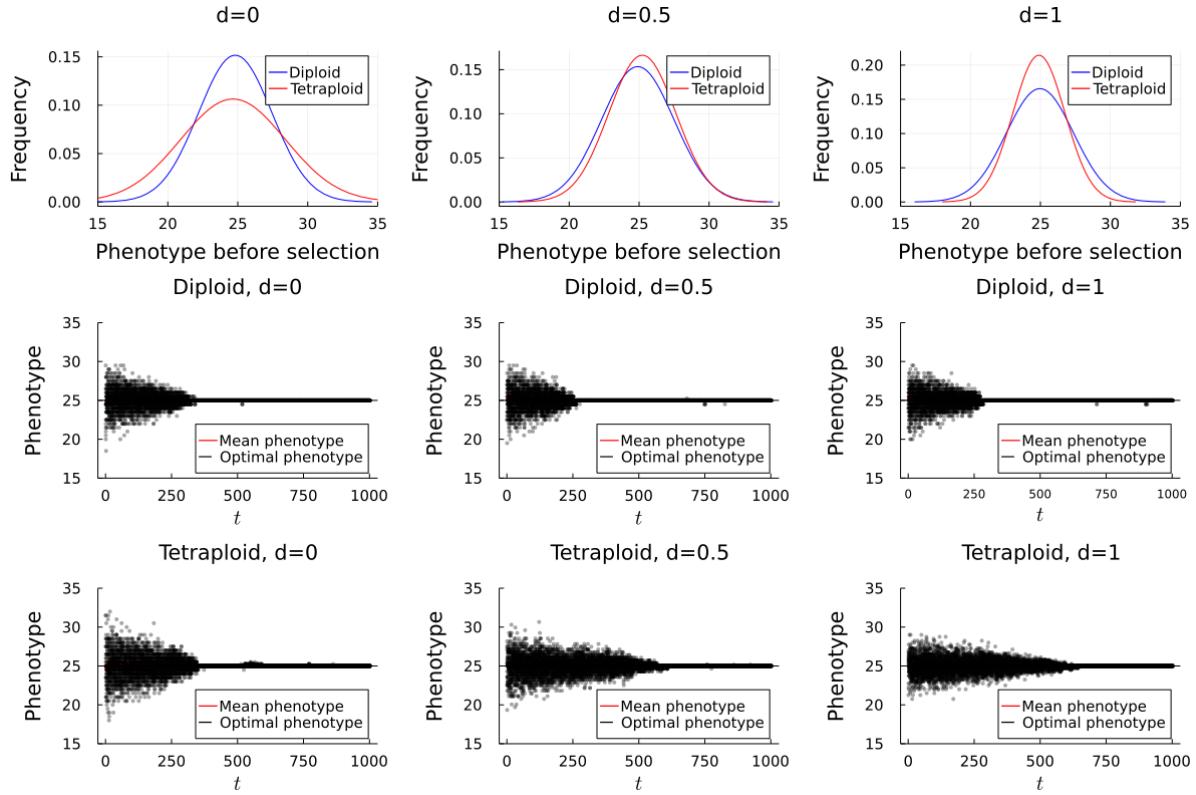
#### 4.1.3 A note on the genotype-phenotype map

Since it is unclear to what extent the difference in starting phenotypic mean and variance for different ploidy levels due to the genotype-phenotype map influence evolution of a population, some additional exploratory simulations were done using different values for  $d$ . Figure 20 shows the phenotypic mean and variance for a neutral random mating deme. These plots are more meant to be illustrative and need more in depth quantitative analysis, but it can be noted that the parameter  $d$  doesn't influence the starting phenotypic variation of diploids (as explained in material and methods). For  $d = 0.5$ , where the starting phenotypic variance of diploids and tetraploids is the same, it is also evident that variance degrades more rapidly in diploids. The genotype-phenotype map seems to have no effect on the loss of phenotypic variance, if only expected to be due to drift in a neutral population. Only the effect of ploidy level can again be noted (slower erosion in tetraploids).



**Figure 20. Effect of genotype-phenotype map for different values of  $d$  on starting phenotypic variance and erosion of phenotypic variance in a single ploidy neutral random mating deme.** The plots in the upper row show the starting phenotypic mean and variance of different ploidy levels for different values of  $d$  (see genotype-phenotype map in materials in methods for specifics). For a value of  $d = 0.5$ , the starting phenotypic variance of diploids and tetraploids will be equal. The middle and lower show the differential loss of phenotypic variance over time due to drift different ploidy levels and for different values of  $d$ . Since these are simulations in a deme without stabilizing selection, the phenotypic mean will not converge to the optimum of the deme. The phenotype of each individual is represented by a black dot and the mean phenotype of the population is shown in red. The following parameters were used for the simulations:  $K = 200$ ,  $p = 0.5$ ,  $\alpha = 0.5$ ,  $L = 100$ ,  $d = 1$ ,  $u = 0$ ,  $r_m = 1.06$ ,  $V_s = 0.5$ .

Whereas there was no effect to be expected between genotype-phenotype map and loss of phenotypic variance in neutral populations, figure 21 shows that with stabilizing selection this is no longer the case. Again, for diploids no effect is noticed on the loss of phenotypic variance since all populations start with the same phenotypic variance for all values of  $d$ . In tetraploids however, it can be noticed that if the starting phenotypic variance is increased, the rate of loss of variance also increases. This can likely be explained by the starting phenotypic variance becoming an important load under stabilizing selection since higher variance from the mean implicates that more maladapted individuals are present in the population (at least if only considering the single deme scenario without migration). This maladaptive effect of phenotypic variation will be elaborated on further as well but has also been explained by Lande & Shannon (1997).



**Figure 21. Effect of genotype-phenotype map for different values of  $d$  on starting phenotypic variance and erosion of phenotypic variance in a single ploidy deme with density dependence and stabilizing selection.** The plots in the upper row show the starting phenotypic mean and variance of different ploidy levels for different values of  $d$  (see genotype-phenotype map in materials in methods for specifics). For a value of  $d = 0.5$ , the starting phenotypic variance of diploids and tetraploids will be equal. The middle and lower show the differential loss of phenotypic variance over time due to drift and stabilizing selection for different ploidy levels and different values of  $d$ . Since these are simulations in a deme with stabilizing selection, the phenotypic mean will converge to the optimum of the deme. The phenotype of each individual is represented by a black dot and the mean phenotype of the population is shown in red. The following parameters were used for the simulations:  $K = 200$ ,  $p = 0.5$ ,  $\alpha = 0.5$ ,  $L = 100$ ,  $u = 0$ ,  $r_m = 1.06$ ,  $V_s = 0.5$ .

Figure 22 shows another important population dynamic at play, evident in tetraploids where the starting phenotypic variance changes with  $d$ . Due to the maladaptive load of phenotypic variance under stabilizing selection, population size after initiation crashes deeper for higher variance (lower values of  $d$  in the case of tetraploids) due to selective deaths but the population converges faster to carrying capacity afterwards due to the decreased maladaptive load that is caused by this (i.e. most maladapted phenotypes are weeded out resulting in a smaller population variance around the phenotypic optimum).

A mathematical model by Lande & Shannon (1997) already indicated that genetic variance for a quantitative trait in a constant environment under stabilizing selection and density dependence results in a genetic load, i.e. the mean fitness of the population is expected to be reduced for higher additive genetic variance due to the phenotype of more individuals deviating from the phenotypic optimum of the environment. This might have important implications when analyzing the population dynamics in a single deme with a constant environment, possibly

due to a difference in genetic variation for different ploidy levels (depending on the implementation of the genotype-phenotype map) in combination with the associated dynamics that influence the evolution of genetic variance (the rate of decrease of genetic variance due to drift that might for example be different for different ploidy levels as shown before).

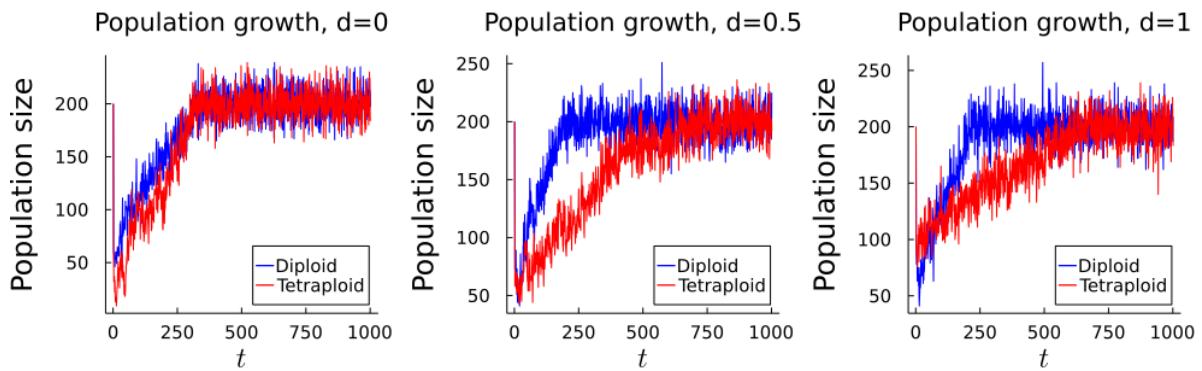
By looking at the population growth for  $d = 0.5$  in figure 19 where the starting phenotypic variance of diploids and tetraploids is the same, a tetraploid population seems to take longer before reaching carrying capacity. One of the explanations could be that the slower loss of genetic variance for higher ploidy levels results from the genetic load under stabilizing selection: because of the higher additive genetic variance less adapted individuals (of which the phenotype deviates more from the phenotypic optimum) in the population will persist longer which decreases the rate of population growth.

The rate of population growth might also be a meaningful approximation to mean fitness of that population (Lande & Shannon, 1997). The rate of population growth has been formulated by Polechová & Barton (2015) as follows:

$$r^* = r_m - \frac{V_G}{2V_s}$$

where  $r^*$  is the instantaneous growth rate of a population,  $r_m$  is a growth rate coefficient,  $V_G$  is the genetic variance and  $V_s$  is the variance of stabilizing selection.

This growth rate is a function of additive genetic variance, the evolution of which can differ for different ploidy levels. Another way to view this is again in terms of genetic load under stabilizing selection. With more persistent additive genetic variance in tetraploids there will be relatively more maladapted individuals in the population. Mean population fitness of tetraploids can thus expected to be temporarily lower compared to diploids under specific circumstances ( $d \geq 0.5$ , starting from maximum genetic variance conditions with  $p = 0.5$ ).



**Figure 22. Effect of genotype-phenotype map for different values of  $d$  on the rate of population growth in a single ploidy deme with density dependence and stabilizing selection.** The blue curve shows the evolution of diploid population size and the red curve shows the evolution of the tetraploid population size for each value of  $d$ . Population size will increase to carrying capacity  $K$  and fluctuate around the equilibrium due to the effect of demographic stochasticity. The following parameters were used for the simulations:  $K = 200$ ,  $p = 0.5$ ,  $\alpha = 0.5$ ,  $L = 50$ ,  $u = 0$ ,  $r_m = 1.06$ ,  $V_s = 0.5$ .

#### 4.1.4 Mixed-ploidy populations in a single deme

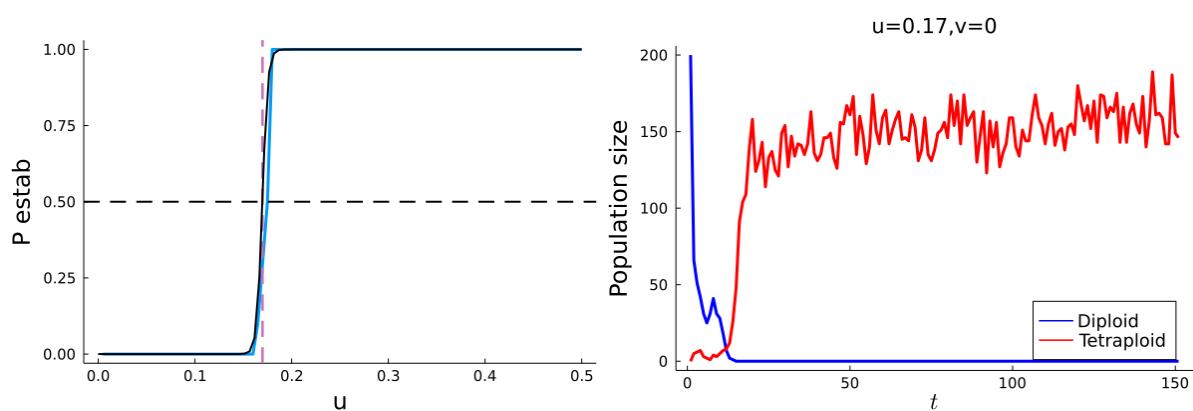
Moving on from studying some of the population genetics aspects of single ploidy populations, simulations were implemented for mixed-ploidy populations to investigate the effects some different conditions might have on polyploid establishment. The model by Felber (1991) that was explained in the introduction was used as an inspiration and as a reference for comparison of results.

The first case that was described by Felber (1991) can be summarized as follows: the fitness of diploids and tetraploids is considered to be the same and triploids are not viable. There is only unreduced gamete formation in diploids and unreduced gamete frequencies are described by the following matrix:

$$G = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1-u & u & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1-v & 0 & v \end{pmatrix}, \text{ with } v = 0$$

where  $u$  is the frequency of unreduced gamete formation in diploids and  $v$  the rate of unreduced gamete formation in tetraploids, i.e. for this model there are no unreduced gametes formed by tetraploids. Starting from a population of only diploid individuals, according to this model, tetraploids will exclude diploids if unreduced gamete formation is higher than  $\approx 0.17$ .

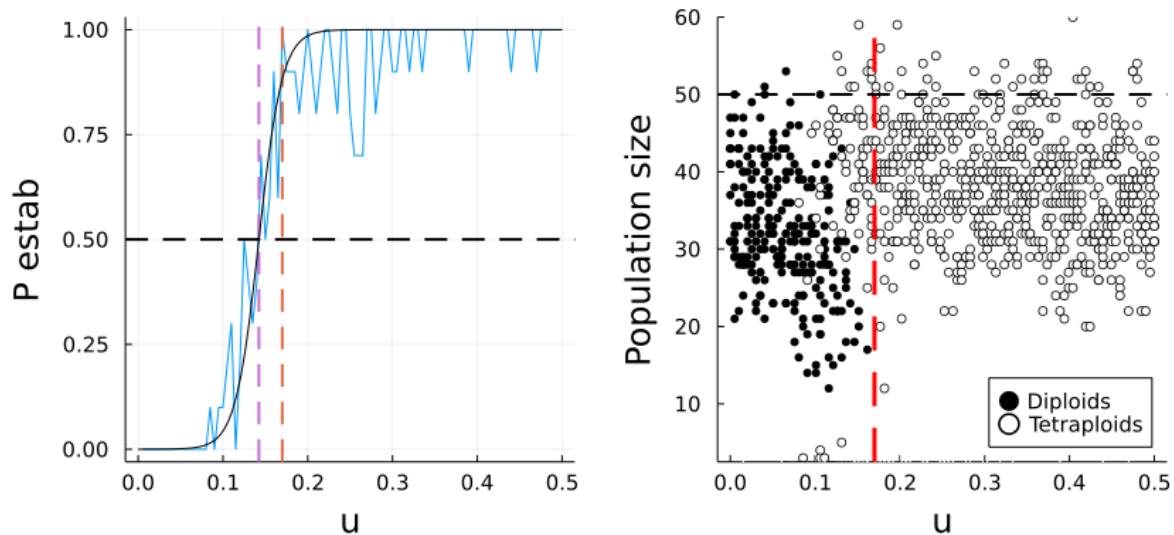
As a baseline for comparison, firstly simulations were done in a neutral random mating deme with a fixed population size that corresponds to the assumptions made by the model from Felber (1991). The plot on the left in figure 25 shows that the result of the model from Felber is very well approximated: from values for  $u = 0.17$  and higher, tetraploids can be seen establishing unequivocally. Subsequently mixed-ploidy simulations were extended with stabilizing selection and density dependence. The plot on the right in figure 23 illustrates the establishment of a tetraploid population for  $u = 0.17$ .



**Figure 23. Illustration of simulations for a single mixed-ploidy deme with unreduced gamete formation in diploids only ( $v = 0$ ).** On the left the result of a simulation with grid search is shown for a neutral random mating mixed-ploidy population. The blue line shows the simulated curve and the black curve on top is fitted with logistic regression. The dashed vertical line shows the critical value of  $u = 0.17$  resulting from the model by Felber (1991). On the right an example of tetraploid establishment is shown for  $u = 0.17$ . The blue curve shows the evolution of diploid population size and the

red curve shows the evolution of the tetraploid population size. The following parameters were used for the simulations:  $K = 200, p = 0.5, \alpha = 0.5, L = 50, d = 1, r_m = 1.06, V_s = 0.5$ .

Figure 24 shows the results from a grid search for different values of  $u$  (see materials and methods for technical details). The plot on the left shows in blue the probability of tetraploid establishment in a diploid population of size 50 with stabilizing selection and density dependence after a grid search over values of  $u$  ranging from 0 to 0.5. The black curve is fitted on top using logistic regression. The point where this fitted curve reaches a probability of tetraploid establishment of 50% is called  $u_{crit}$ . This value of  $u_{crit}$  approximates 0.17 in a neutral random mating population as shown in figure 25. After introducing stabilizing selection and density dependence this value of  $u_{crit}$  slightly shifts to the left (0.142). There is also some window at lower values of  $u$  (between  $u \approx 0.10$  and  $u \approx 0.14$ ) due to stochasticity where there some (rare) establishment of tetraploids is possible.



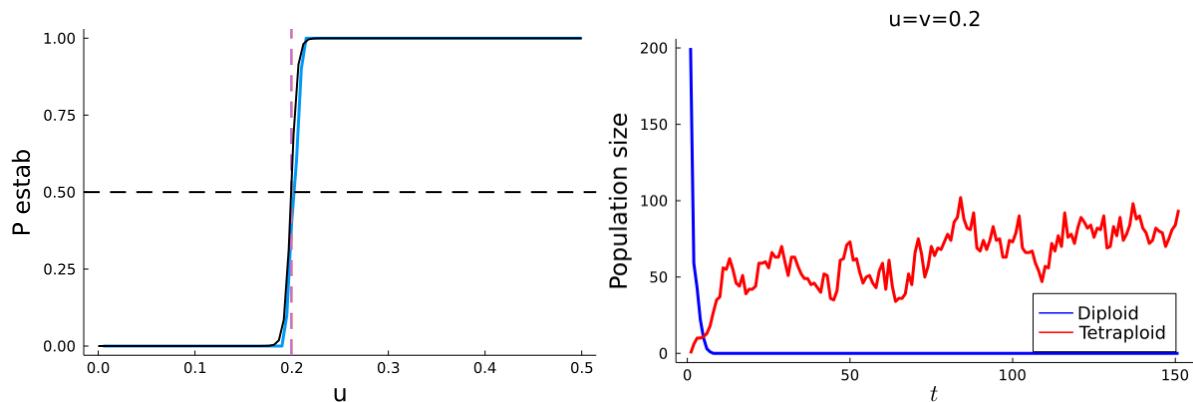
**Figure 24. Simulations in a single mixed-ploidy deme with density dependence and stabilizing selection and unreduced gamete formation in diploids only ( $\nu = 0$ ).** On the left the result of a simulation with grid search is shown for a mixed-ploidy population with stabilizing selection and density dependence. The blue line shows the simulated curve and the black curve on top is fitted with logistic regression. The dashed red vertical line shows the critical value of  $u = 0.17$  resulting from the model by Felber (1991). The dashed purple vertical line shows the value  $u_{crit} = 0.142$ . On the right both diploids and tetraploid population sizes are shown for each simulation in the grid search. The following parameters were used for the simulations:  $K = 50, p = 0.5, \alpha = 0.5, L = 50, d = 1, r_m = 1.06, V_s = 0.5$ .

The second case that was described by Felber (1991) can be summarized as follows: the fitness of diploids and tetraploids is considered to be the same and triploids are not viable. There is unreduced gamete formation in both diploids and tetraploids and unreduced gamete frequencies are described by the following matrix:

$$G = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1-u & u & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1-v & 0 & v \end{pmatrix}, \text{ with } u = v$$

where  $u$  is the frequency of unreduced gamete formation in diploids and  $v$  the rate of unreduced gamete formation in tetraploids. Starting from a population of only diploid individuals, according to the model by Felber (1991) tetraploids exclude diploids if unreduced gamete formation is higher than 0.2.

Again as a baseline for comparison, firstly simulations were done in a neutral random mating deme with a fixed population size that corresponds to the assumptions made by the model from Felber (1991). The plot on the left in figure 25 shows that the result of the model from Felber is very well approximated: from values for  $u = 0.20$  and higher, tetraploids can be seen establish unequivocally. Subsequently mixed-ploidy simulations were extended with stabilizing selection and density dependence. The plot on the right in figure 25 illustrates the establishment of a tetraploid population for  $u = 0.20$ .

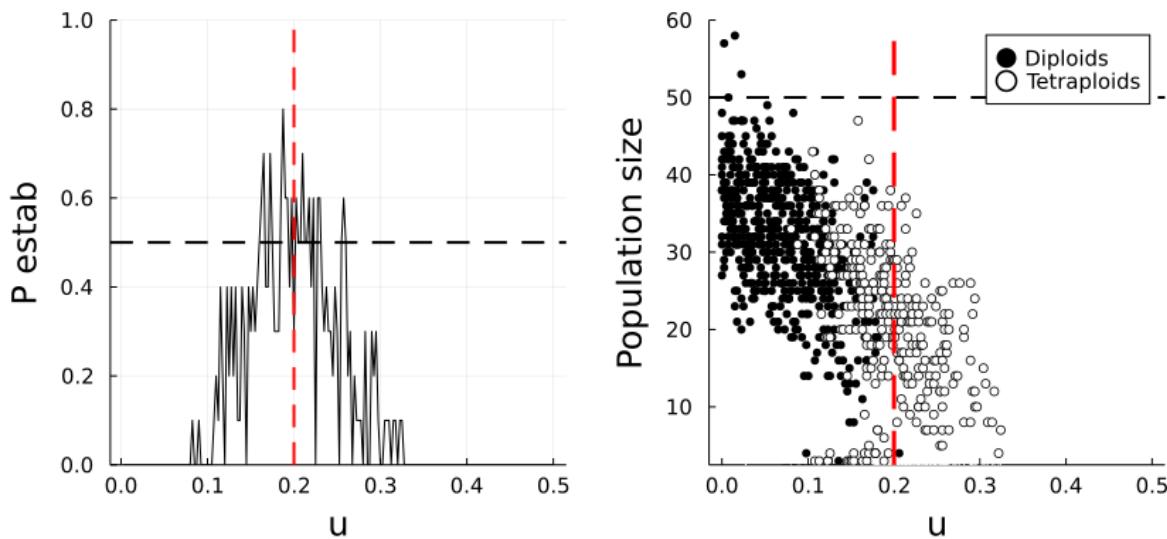


**Figure 25. Illustration of simulations for a single mixed-ploidy deme with unreduced gamete formation in both diploids and tetraploids ( $v = u$ ).** On the left the result of a simulation with grid search is shown for a neutral random mating mixed-ploidy population. The blue line shows the simulated curve and the black curve on top is fitted with logistic regression. The dashed vertical line shows the critical value of  $u = 0.20$  resulting from the model by Felber (1991). On the right an example of tetraploid establishment is shown for  $u = 0.20$ . The blue curve shows the evolution of diploid population size and the red curve shows the evolution of the tetraploid population size. The following parameters were used for the simulations:  $K = 200$ ,  $p = 0.5$ ,  $\alpha = 0.5$ ,  $L = 50$ ,  $d = 1$ ,  $r_m = 1.06$ ,  $V_s = 0.5$ .

Something else that can be noted from figure 25 on the right is that the population size barely grows to carrying capacity, see for example figure 19 for comparison where  $v = 0$ . This is possibly cause by increasing cytotype load for increasing values of  $u$ , something that will be explained more in depth later on.

Figure 26 shows the results from a grid search for different values of  $u = v$  (see materials and methods for technical details). The plot on the left shows in black the probability of tetraploid establishment in a diploid population of size 50 with stabilizing selection and density dependence after a grid search over values of  $u$  ranging from 0 to 0.5. In this case with unreduced gametes for both diploids and tetraploids the probability of tetraploid establishment seems to decrease again for high values of  $u$ . This is due to an artifact in the simulations that should

be accounted for. The probability of tetraploid establishment is calculated as the ratio of simulations where a majority of the individuals of the population are tetraploids after 50 generation to the total number of all simulations, unconditional on a certain population size. What can be seen from figure 28 on the right is that populations completely collapse for high levels of  $u$ . This is likely again due to cytotype load in the population. Since this result wasn't expected to be so drastically, it wasn't accounted for in these simulations. For future simulations it could be interesting to define a cut-off value for total population size.

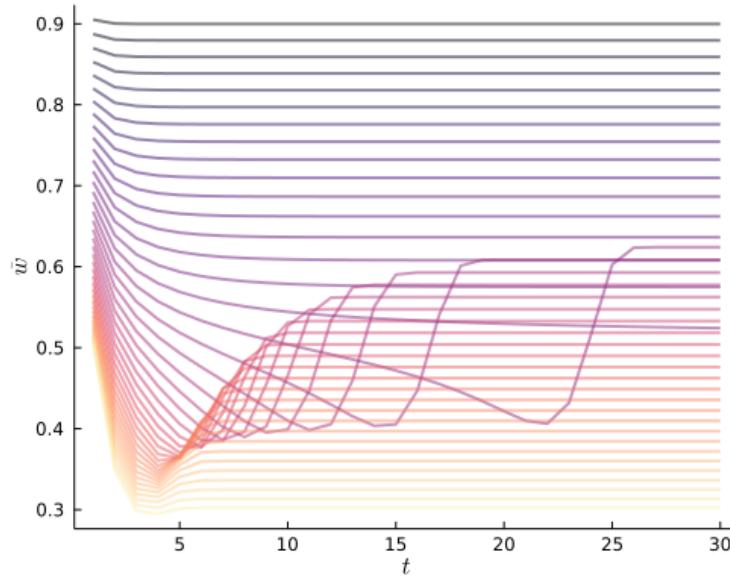


**Figure 26. Simulations in a single mixed-ploidy deme with density dependence and stabilizing selection and unreduced gamete formation in both diploids and tetraploids ( $v = u$ ).** On the left the result of a simulation with grid search is shown for a mixed-ploidy population with stabilizing selection and density dependence. The black line shows the simulated. The dashed red vertical line shows the critical value of  $u = 0.20$  resulting from the model by Felber (1991). On the right both diploids and tetraploid population sizes are shown for each simulation in the grid search. Population size can be seen to collapse for higher values of  $u$ . The following parameters were used for the simulations:  $K = 50$ ,  $p = 0.5$ ,  $\alpha = 0.5$ ,  $L = 50$ ,  $d = 1$ ,  $r_m = 1.06$ ,  $V_s = 0.5$ .

#### 4.1.5 A note on cytotype load

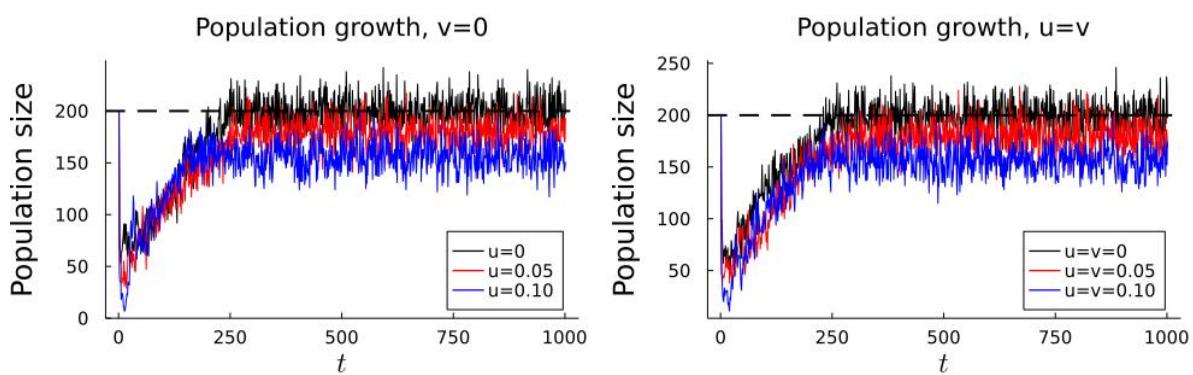
It has been hypothesized by Kreiner et al. (2017) that the formation of unreduced gametes in a population is maladaptive. This is also evident from the model by Felber (1997). In this model (no triploids, tetraploids and diploids equally fit, shared unreduced gamete formation rate, no higher polyploids) tetraploids take over irrespective of the initial frequencies as soon as  $u > 0.2$ . However, values of  $u > 0.2$  will cause a serious maladaptive load in the population since as at least a fraction  $(1-0.8^2)$ , in general number will be at least  $(1 - (1 - u)^2)$  of inviable offspring will be generated each generation due to the fact that unreduced gametes resulting from tetraploids never combine in viable offspring. Figure 27 shows how for this model by Felber (1997) during establishment, starting from a diploid population, the mean population fitness first shows a remarkable decrease for increasing values of  $u$  (from top to bottom). The formation of unreduced gametes leading to a certain frequency of inviable offspring causes a maladaptive load in a population that might be termed cytotype load. Note

that this term cytotype load doesn't explicitly exist yet in the literature and can be attributed to Arthur Zwaenepoel, the supervisor of this dissertation.

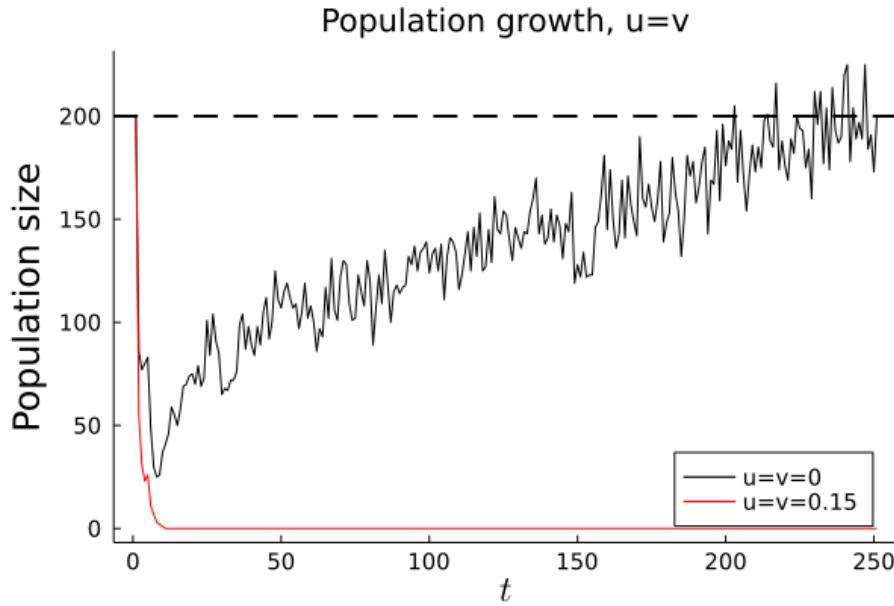


**Figure 27. Evolution of mean fitness for different values of  $u$ .** It can be noticed that during establishment, starting from a diploid population, the mean population fitness first decreases for increasing values of  $u$  from top to bottom.

A similar observation that can be made using individual-based simulations in a single deme diploid population with balancing selection and density dependence is that for increasing values of  $u$ , the equilibrium around which the population size will eventually fluctuate stays away further from the carrying capacity  $K$  of the deme. It could be hypothesized that this is also caused by a maladaptive cytotype load (figure 28). Figure 29 illustrates how a population can even go extinct due to cytotype load for high values of  $u$ .



**Figure 28. Population growth in a mixed-ploidy deme with density dependence and stabilizing selection for different values of  $u$  to show the effect of cytotype load.** Growth of population size is shown for the two different cases of the model by Felber (1991). In both cases population size will increase to carrying capacity  $k$  for  $u = 0$ . For higher values of  $u$  however, the equilibrium population size around which the population sizes fluctuate will be lower. The following parameters were used for the simulations:  $K = 200$ ,  $p = 0.5$ ,  $\alpha = 0.5$ ,  $L = 50$ ,  $d = 1$ ,  $u = 0.05$ ,  $r_m = 1.06$ ,  $V_s = 0.5$ .



**Figure 29. Illustration of putative cytotype load causing a population to collapse for high values of  $u$ .** The black curve shows the population growth of a population without cytotype load ( $u = 0$ ). The red curve shows how a population can go extinct due to cytotype load for high values of  $u$ . The following parameters were used for the simulations:  $K = 200$ ,  $p = 0.5$ ,  $\alpha = 0.5$ ,  $L = 50$ ,  $d = 1$ ,  $u = 0.05$ ,  $r_m = 1.06$ ,  $V_s = 0.5$ .

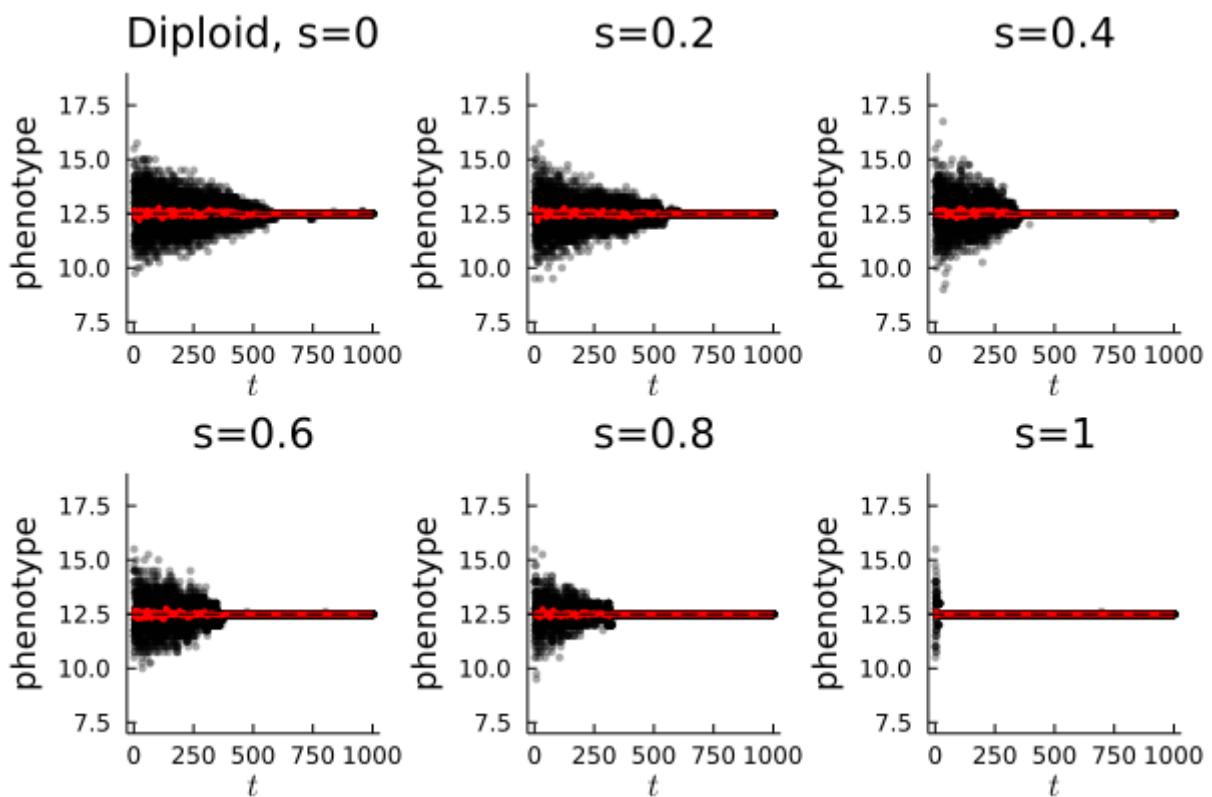
#### 4.1.6 Effect of breeding system: Assortative mating and self-fertilization

Results from mixed-ploidy single deme simulations showed that the value of  $u$  for which at least 50% tetraploid establishment can be expected ( $u_{crit}$ ) under stabilizing selection and density dependence was found to be 0.142 (albeit with a frame of smaller probabilities of establishment for as low as  $u \approx 0.10$ ) but this is still very high compared to frequencies of unreduced gamete formation found in nature with frequencies of around 0.025 or lower for sexually reproducing species (Kreiner, 2017).

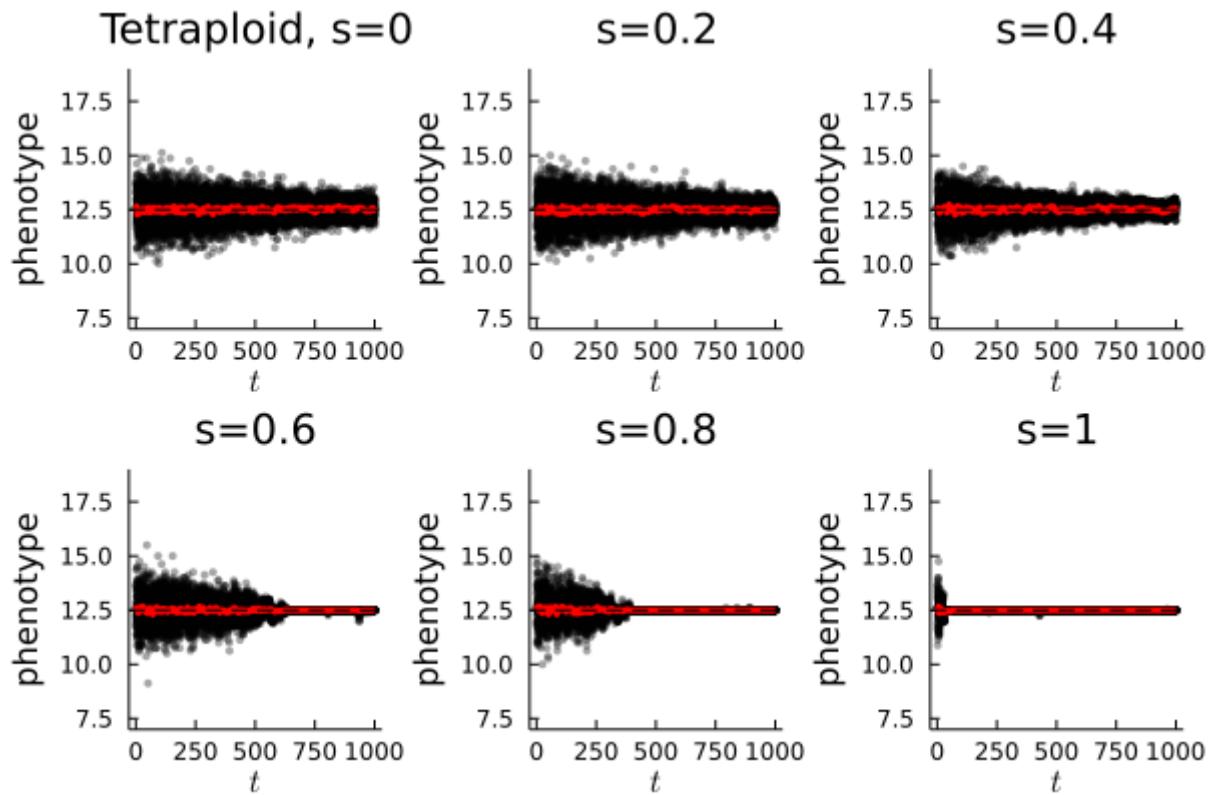
As laid out in the introduction, life cycle aspects potentially influence the dynamics of polyploid establishment in the introduction. Some basic simulation studies on tetraploid establishment already showed the effects of self-fertilization or assortative mating on polyploid establishment as mentioned in the introduction because self-fertilization and assortative mating (here defined as a higher probability of mating within a certain ploidy level (intracytotype mating) than between different ploidy levels) shield polyploids from the detrimental effects of minority cytotype exclusion. Assortative mating is used here as an abstraction: the underlying mechanics for why there is increased frequency of intracytotype mating are not specified but can be very diverse. It can be interpreted as a single parameter representative for spatial dynamics as in Baack (2005) where intracytotype mating have increased frequencies because individuals with the same ploidy level are clustered together in space. It could also be used to model other ecological relevant things like a difference in phenology or biotic interactions like pollinator shift. The interpretation of Self-fertilization is more straightforward. The inclusion hereof is inspired by the hypothesis that polyploids would be less susceptible to inbreeding depression relative to diploids creating a competitive advantage. Both assortative

mating and self-fertilization are hypothesized to decrease the impact from maladaptive gene flow due to migration and minority-cytotype exclusion (Grisswold, 2021).

Since self-fertilization causes a loss of heterozygosity due to inbreeding depression, firstly some simulations were run in single ploidy populations to study the effect of self-fertilization on evolution of phenotypic variance and the interaction with ploidy level. Figure 30 (diploids) and figure 31 (tetraploids) show that the effect of self-fertilization is very prominent: phenotypic variance decreases (much) more rapidly for higher levels of self-fertilization. The differential effect of loss of phenotypic variance due to the ploidy level as explained before is also visible.

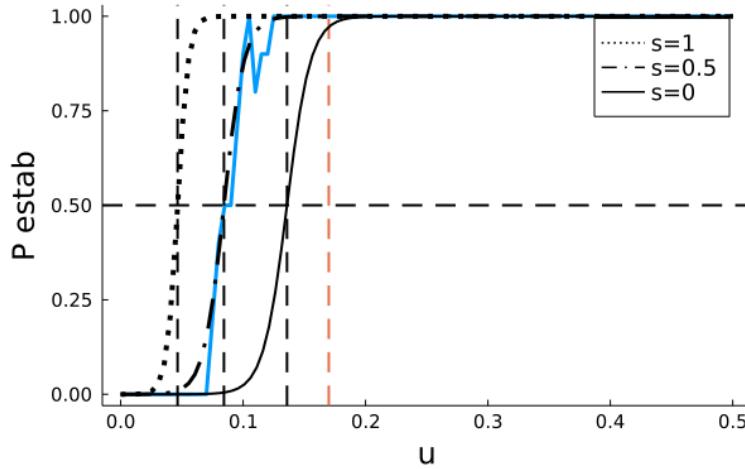


**Figure 30. The effect of self-fertilization on the evolution of phenotypic mean and variance in a diploid deme with density dependence and stabilizing selection.** The phenotype of each individual is represented by a black dot and the mean phenotype of the population is shown in red. These plots clearly show the differential loss of phenotypic variance over time due to drift and stabilizing selection for different self-fertilization rates in diploids. Since these are simulations in a deme with stabilizing selection, the phenotypic mean will converge to the optimum of the deme. The following parameters were used for the simulations:  $K = 200$ ,  $p = 0.5$ ,  $\alpha = 0.5$ ,  $L = 50$ ,  $d = 1$ ,  $r_m = 1.06$ ,  $V_s = 0.5$ .



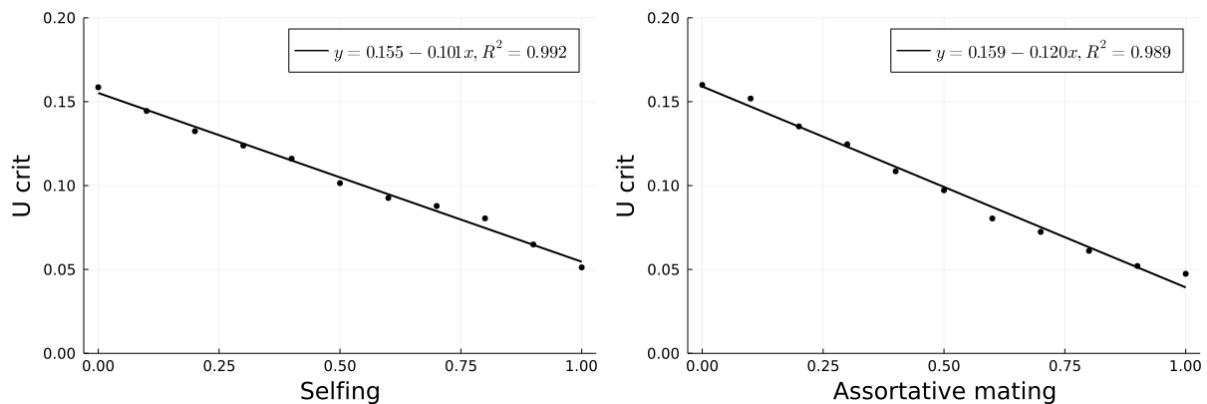
**Figure 31. The effect of self-fertilization on the evolution of phenotypic mean and variance in a tetraploid deme with density dependence and stabilizing selection.** The phenotype of each individual is represented by a black dot and the mean phenotype of the population is shown in red. These plots clearly show the differential loss of phenotypic variance over time due to drift and stabilizing selection for different self-fertilization rates in tetraploids. Since these are simulations in a deme with stabilizing selection, the phenotypic mean will converge to the optimum of the deme. The following parameters were used for the simulations:  $K = 200$ ,  $p = 0.5$ ,  $\alpha = 0.5$ ,  $L = 50$ ,  $d = 1$ ,  $u = 0$ ,  $r_m = 1.06$ ,  $V_s = 0.5$ .

Subsequently simulations were run for mixed-ploidy populations to study the effect of self-fertilization and assortative mating on polyploid establishment. Figure 32 illustrates how the value of  $u_{crit}$ , corresponding to the value of  $u$  for which 50% or more polyploid establishment can be observed, shifts to the left for increasing values of self-fertilization ( $u_{crit} = 0.16$  for  $s = 0$  is 0.16,  $u_{crit} = 0.097$  for  $s=0.5$  and as low as 0.047 for the extreme case of  $s=1$  (only self-fertilization occurs)).



**Figure 32. The effect of self-fertilization on tetraploid establishment in a single mixed-ploidy deme with density dependence and stabilizing selection.** On the left the result of a simulation with grid search is shown for a mixed-ploidy population with stabilizing selection and density dependence. The blue line shows the simulated curve for  $s = 0.5$  and the black curves on top are fitted with logistic regression for different values of  $s$ . The dashed red vertical line shows the critical value of  $u = 0.17$  resulting from the model by Felber (1991). The dashed black vertical lines show the values  $u_{crit}$  for different frequencies of self-fertilization. The following parameters were used for the simulations:  $K = 50$ ,  $p = 0.5$ ,  $\alpha = 0.5$ ,  $L = 50$ ,  $d = 1$ ,  $u = 0.05$ ,  $r_m = 1.06$ ,  $V_s = 0.5$ .

Figure 33 shows the effect of different degrees of self-fertilization and assortative mating on the probability of tetraploid establishment, here denoted by  $u_{crit}$ , corresponding to the value of  $u$  for which 50% or more polyploid establishment can be observed. There seems to be a clear effect of both self-fertilization and assortative mating that can be approximated by a linear relationship. Values of  $u_{crit}$  approach 0.05 for both the most extreme cases of assortative mating and self-fertilization.



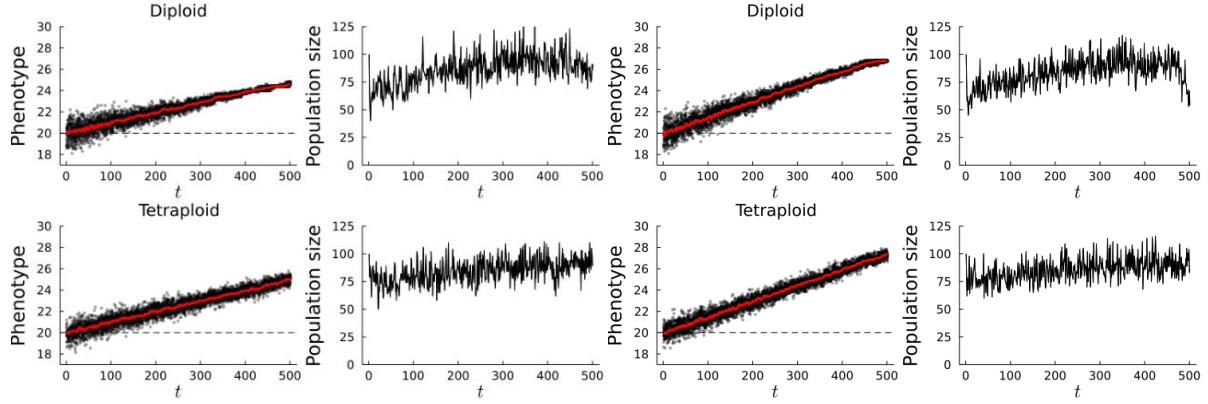
**Figure 33. Effect of self-fertilization and assortative mating on  $u_{crit}$  in a single mixed-ploidy deme with density dependence and stabilizing selection.** For different probabilities of self-fertilization (left) and assortative (mating) a grid search for parameter  $u$  was implemented and a value  $u_{crit}$  was calculated corresponding to the value of  $u$  for which at least 50% tetraploid establishment occurred (see materials and methods for details). The decrease in  $u_{crit}$  due to either increased self-fertilization or assortative mating seems to be approximated well by a linear relationship.

#### 4.1.7 Some observations from a single deme with changing environment

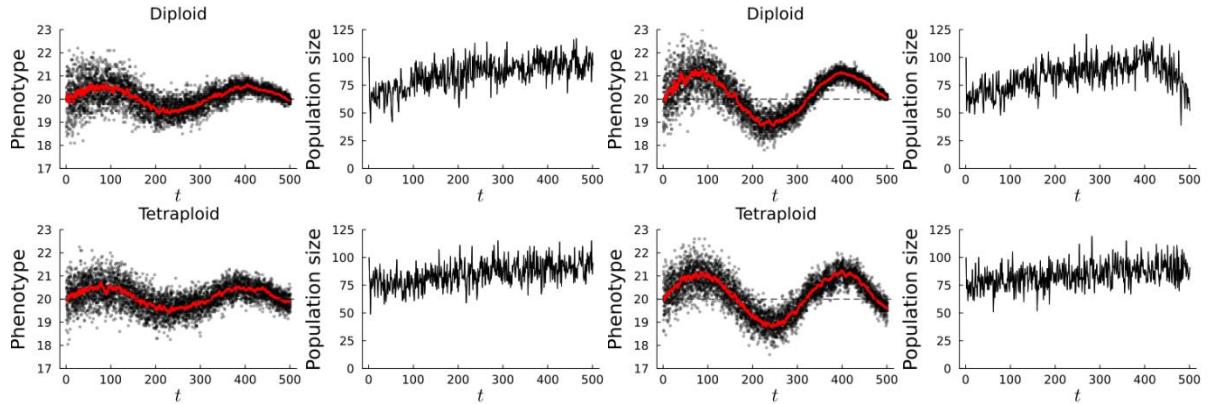
Based on previous findings a small sidestep was taken. In previous simulations genetic variation was shown to persist longer in tetraploids, something that could be especially important with unpredictable environmental change through time, e.g. in relation to mass extinction events (Cretaceous–Paleogene extinction event 66 million years ago (Van de Peer et al., 2021)) or adaptation to climate change. However, it was also shown that the load under stabilizing selection due to persisting genetic variance can be higher in tetraploids compared to diploids under specific assumption of the genotype-phenotype map. Lande & Shannon (1997) showed that depending on the pattern of environmental change, additive genetic variance can be either adaptive or maladaptive. Thus, a changing environment through time can be expected to influence the load under stabilizing selection seen in tetraploids, creating a window of opportunity where the positive effects of standing genetic variation outweigh the costs. It can be hypothesized that tetraploids will be more robust to environmental change than diploids, but that diploids will be more efficient in stable environments due to the cost of standing genetic variation being higher in tetraploids.

Lande & Shannon (1997) describe some different scenarios of environmental change through time, two of which were briefly explored. With directional change of the environment, the mean phenotype of the populations tracks the environmental gradient that changes linearly through time but with some lag, causing a load depending on the steepness of the gradient (termed ‘evolutionary load’ by Lande & Shannon (1997)). This load is inversely proportional to additive genetic variance, i.e. in this scenario genetic variance is expected to be adaptive. In a cyclic environment additive genetic variance reduces the evolutionary load by allowing the mean phenotype to track the optimum more closely but the benefit depends on frequency and amplitude of oscillations. Also random environmental changes through time are possible: the theoretical analysis of these is more complex and depends for example on the autocorrelation of environmental change.

Figure 34 and 35 show illustrative simulations in for both an environmental gradient and a cyclic environment. For small steepness of the gradient of small oscillations (simulations on the left), both diploid and tetraploid populations can be seen to track the phenotypic optimum. However, by increasing the steepness or by increasing the amplitude of oscillations (simulations on the right), at a certain point in time the diploid population fails to track the optimum and starts to collapse.



**Figure 34. The evolution of phenotypic mean and variance for a linearly changing phenotypic optimum through time.** The phenotypic gradient on the left is described by  $y = 20 + 0.01x$ . The phenotypic gradient on the right is described by  $y = 20 + 0.015x$ . For diploids the population size starts to collapse for the steeper gradient. The following parameters were used for the simulations:  $K = 100$ ,  $p = 0.5$ ,  $\alpha = 0.2$ ,  $L = 200$ ,  $d = 1$ ,  $r_m = 1.06$ ,  $V_s = 0.5$ .



**Figure 35. The evolution of phenotypic mean and variance for a sinusoidal changing phenotypic optimum through time.** The phenotypic gradient on the left is described by  $y = 20 + 0.5\sin\left(\frac{x}{50}\right)$ . The phenotypic gradient on the right is described by  $y = 20 + 1.10\sin\left(\frac{x}{50}\right)$ . For diploids the population size starts to collapse for the steeper gradient. The following parameters were used for the simulations:  $K = 100$ ,  $p = 0.5$ ,  $\alpha = 0.2$ ,  $L = 200$ ,  $d = 1$ ,  $r_m = 1.06$ ,  $V_s = 0.5$ .

## 4.2 Multiple demes with migration

After exploration of some important single deme dynamics, the transition was made to more complex spatial models with migration. A spatial setting allows studying some of the central aspects of the initial hypotheses that were formulated under aims.

### 4.2.1 Mainland-island model

First a spatial scenario was explored with migration in a single direction from a source population to a sink population. This spatial setting can also be interpreted as a mainland-island model. The source population is implemented as the mainland population with an infinite

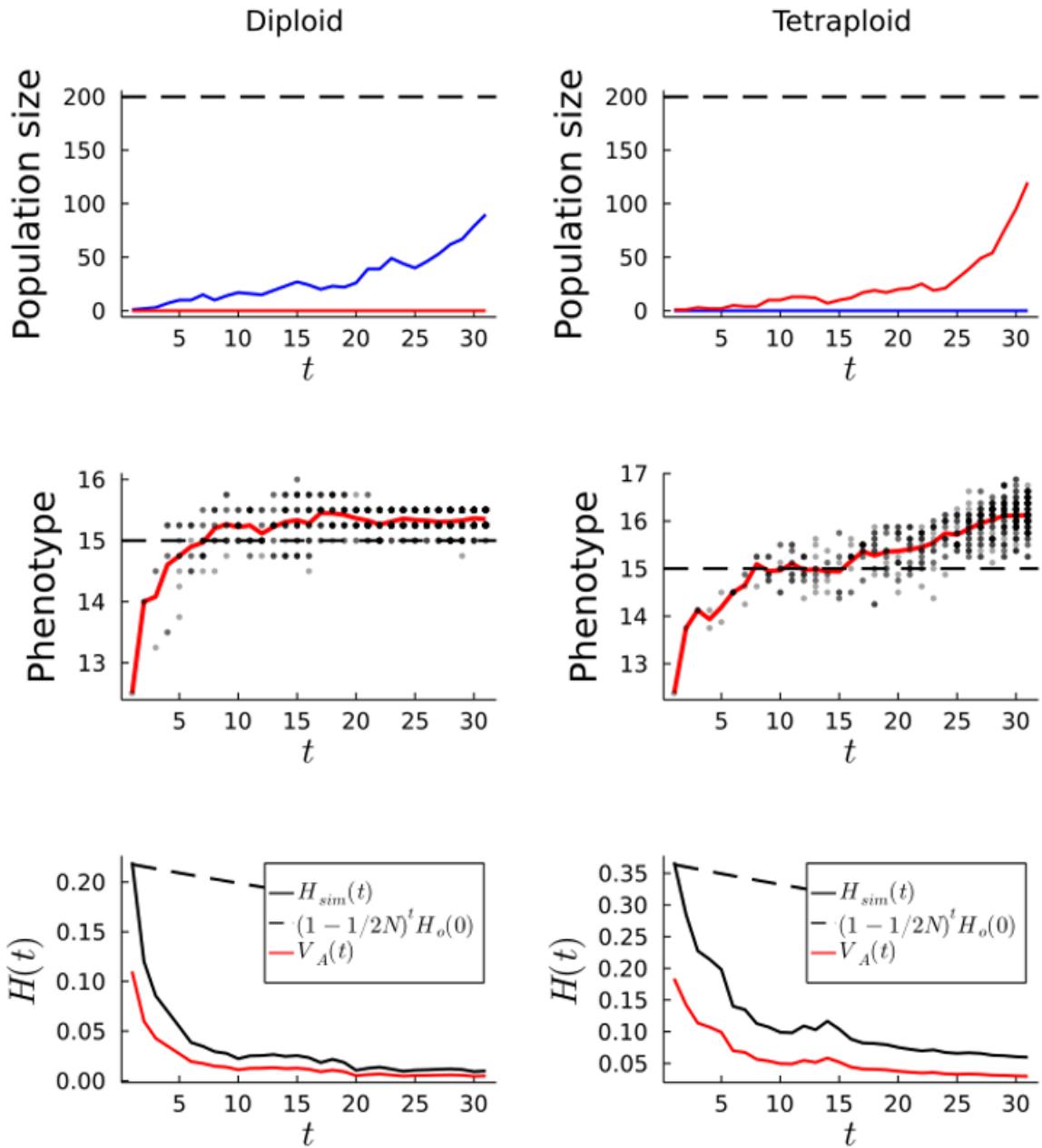
population size that is expected to be in HWLE. From this population individuals are chosen randomly to colonize an empty island. Two elements are important in the mainland-island model to quantitate maladaptive gene flow: a first important aspect is that the mean phenotype of the mainland population is maladapted to the optimal phenotype of the island, i.e. there is a difference in phenotypic optima of the source population and the island, and the second element is the migration rate.

#### 4.2.2 Migration with a single migrant

First simulations for the mainland-island model were done with a single migrant that is chosen randomly from the founder population. Figure 36 shows an example of a successful establishment event with a single migrant from either a diploid or a tetraploid founder population. Important to note is that these simulations (and all mainland-island simulations for that matter) are implemented with directional selection and without density dependence to study the dynamics of solely the first generations of establishment. These plots are more meant to be illustrative of aspects that could be explored by this type of model like population growth, evolution of genetic variance and loss of heterozygosity after a founder event.

A couple of observations can be made that correspond well to the finding by Barton & Etheridge (2018) for a diploid population: the mean phenotype of the population will increase rapidly in the first generations due to directional selection. When the mean phenotype of the population reaches the phenotypic optimum of the island, the population size will start to increase more rapidly. The heterozygosity will drop rapidly in the first generations due to high prevalence of inbreeding and drift because of the low population size. The effect of directional selection erodes heterozygosity even further, leading to a large deviation from the expected loss of heterozygosity for a neutral population. Something similar was evident from the simulations in a single deme with stabilizing selection.

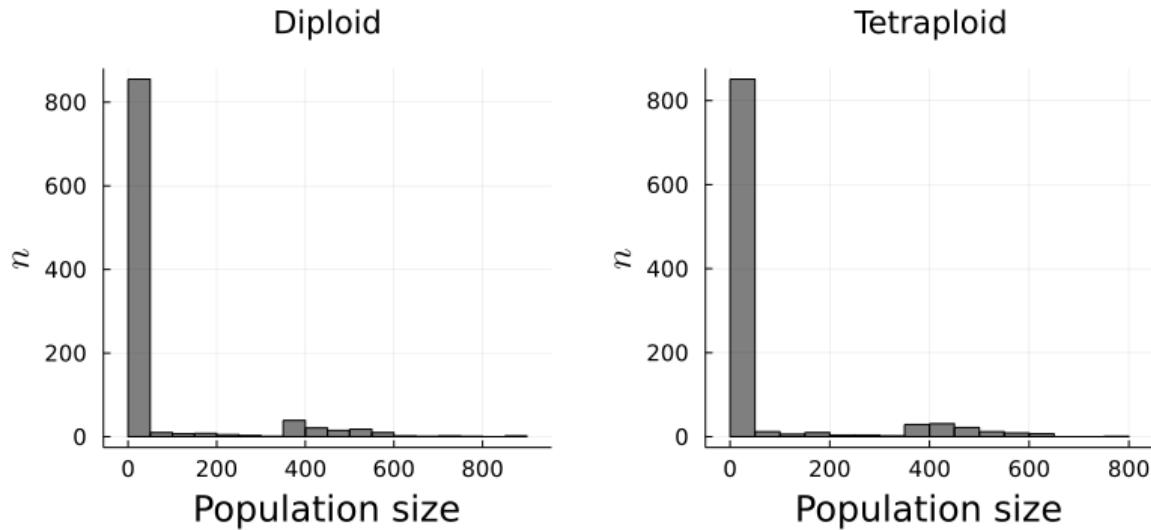
Note that while it looks like heterozygosity degrades slower in tetraploid and the phenotypic variance increases more, these simulations remain mostly illustrative. While the difference between source-sink phenotypic optima is the same for diploids and tetraploids, other aspects haven't been controlled for (the heterozygosity of the founder tetraploid is for example higher than for the diploid founder). More standardized large-scale simulations are necessary to explore these aspects.



**Figure 36. Examples of successful establishment of a single migrant from a diploid or tetraploid founder population.** Population growth, evolution of phenotypic variance and loss of heterozygosity after an establishment event are plotted. The difference in phenotypic optima between mainland and island  $\Delta z = 2.5$ . The following parameters were used for the simulations:  $p = 0.5$ ,  $\alpha = 0.5$ ,  $L = 50$ ,  $d = 1$ ,  $\beta = 0.2$ . These plots are greatly inspired by simulations from Barton & Etheridge (2018).

Refining and exploring these simulations more in depth could be interesting to get more insight into the evolution of genetic variance after founder events and comparing this between diploids and tetraploids. It could be hypothesized that there is a difference in establishment for diploids and tetraploids due to differences in the evolution of genetic variance. Some basic simulations were implemented to start exploring this, but these focus as yet mostly on probability of establishment and give no insight in the evolution of genetic variance. Figure 37 shows a comparison between the number of establishment events starting with a single migrant from either a diploid or tetraploid founder population with a difference in

mainland-island phenotypic optimum of  $\Delta z = 1$ . The probability of establishment for separate 1000 simulations was found to be very comparable between diploids (18% establishment) and tetraploids (18.2% establishment).

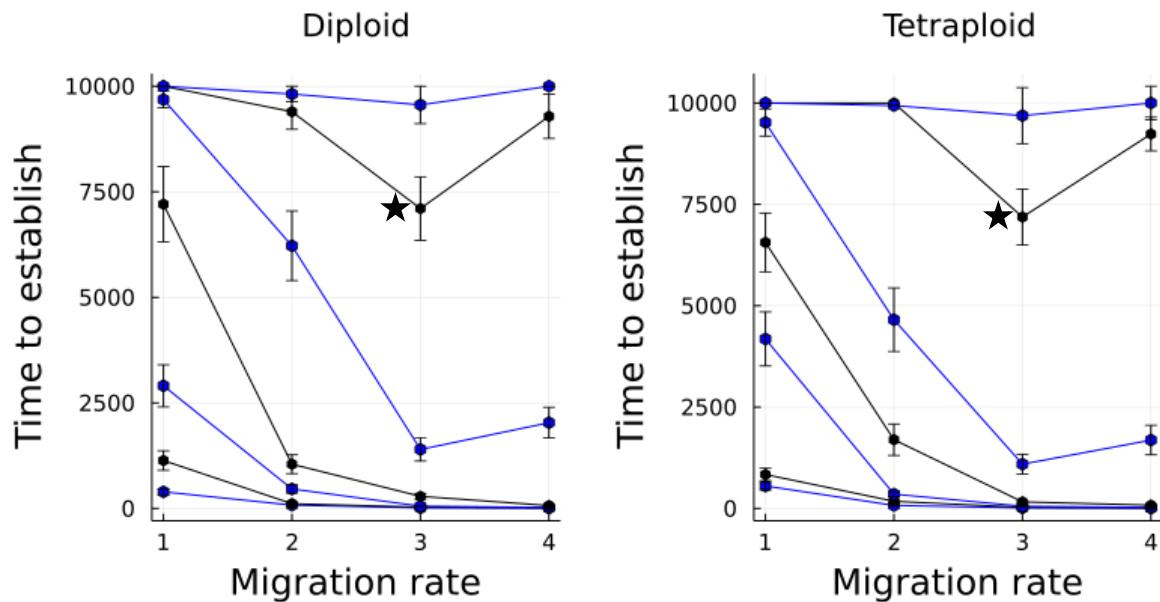


**Figure 37. Comparison of the amount of establishment events from a single migrant from a diploid versus a tetraploid founder population.** A population is considered to be established from population sizes of 100 or higher.  $n$  denotes the number of simulations. The following parameters were used for the simulations:  $p = 0.5$ ,  $\alpha = 0.3$ ,  $L = 50$ ,  $d = 1$ ,  $\beta = 0.2$ .

#### 4.2.3 Continuous migration

Although there can be probably already learned a lot from scenarios with a single migrant as discussed in the previous part, they were only briefly explored. Especially interesting are some aspects due to the founder event on an island that were also clear from the simulations, such as the high prevalence of inbreeding and drift found for low population sizes which might facilitate tetraploid establishment as hypothesized before. However, a transition was made to mainland-island models with continuous migration to also incorporate the continuous effect of maladaptive gene flow due to migration.

Some first exploratory simulations were done to compare the mean time to establishment with continuous migration starting from a diploid versus a tetraploid founder population for which the results are shown in figure 38. On the x-axis the migration rate is shown, increasing from left to right (see materials and methods on the specifics of how the number of migrants is calculated). The y-axis shows the time to establishment, where a longer time to establishment of a population corresponds to a lower probability of establishment. The different curves relate to a difference in phenotypic optima between mainland and island. This difference increases from bottom to top. From visual inspection of these plots there doesn't seem to be an inherent advantage for either ploidy level to establish on an island. This is comparable to the results found in the scenario with a single migrant.

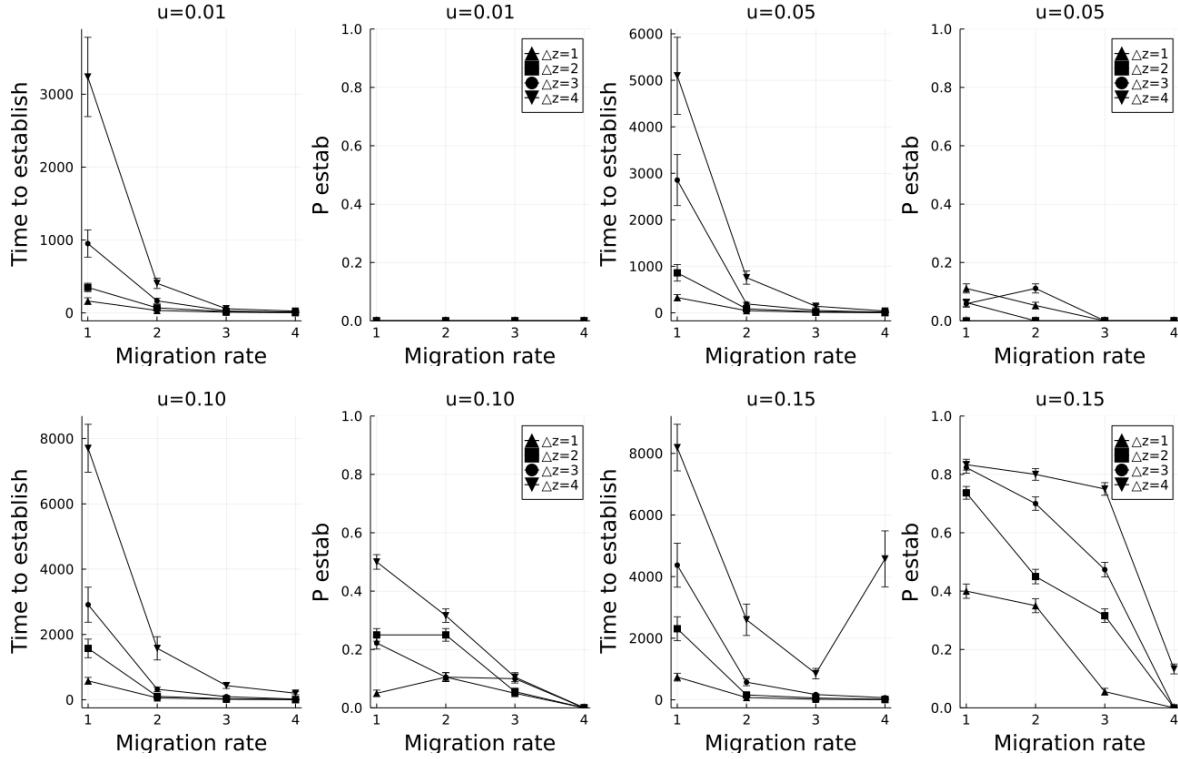


**Figure 38. Comparison of the mean time to establishment with continuous migration starting from a diploid versus a tetraploid founder population.** The points show the mean time to establishment of a population and error bars shows standard error on the mean. The black stars mark the optimum due to a trade-off between the maladaptive effects of migration and the positive effects of gene flow on genetic variance. The following parameters were used for the simulations:  $p = 0.5$ ,  $\alpha = 0.3$ ,  $L = 50$ ,  $d = 1$ ,  $\beta = 0.2$ .

The time to establishment is expected to increase with increasing migration load caused by a difference in phenotypic optima. For different migration rates some optimum is expected due to a trade-off between the maladaptive effects of migration and the positive effects of gene flow on genetic variance (Barton & Etheridge, 2018; Bridle et al., 2019). This pattern is indeed something that can be seen from these simulations as indicated by the black star.

The transition was made from simulations with continuous migration for single ploidy populations to mixed-ploidy populations by adding unreduced gamete formation to study the effect of maladaptive gene flow on polyploid establishment. Simulations were started with a diploid founder population with continuous migration to an island for different migration rates and differences in phenotypic optima.

Figure 39 shows the effect of the frequency of unreduced gamete formation in diploids, denoted by the parameter  $u$ , on the time to establishment of a population on the island (independent of ploidy level) and the probability of tetraploid establishment ( $P_{\text{estab}}$ ). As was expected in accordance with the model from Felber (1991) and previous findings for single deme simulations, the probability of tetraploid establishment was found to higher for higher values of  $u$ . However only for high ( $u = 0.10$ ) to very high ( $u = 0.15$ ) frequencies of unreduced gamete formation there is a significant probability of tetraploid establishment. For values of  $u$  that approximate more realistic values ( $u = 0.5$ ) these probabilities are only baseline but arguably not negligible.



**Figure 39. Effect of the frequency of unreduced gamete formation  $u$  on the mean time to establishment and the probability of tetraploid establishment.** For each combination of migration rate ( $M = [0.01 \quad 0.1 \quad 1 \quad 10]$ ) and difference in phenotypic optima ( $\Delta z = [1 \quad 2 \quad 3 \quad 4]$ ). Points show the mean time to establishment of a population (left) and probability of tetraploid establishment ( $P_{\text{estab}}$ ) on the right. Error bars show the standard error on the mean. Upper triangle:  $\Delta z = 1$ , square:  $\Delta z = 2$ , circle:  $\Delta z = 3$ , down triangle:  $\Delta z = 4$ . The following parameters were used for the simulations:  $p = 0.5$ ,  $\alpha = 0.30$ ,  $L = 50$ ,  $d = 1$ ,  $\beta = 0.2$ .

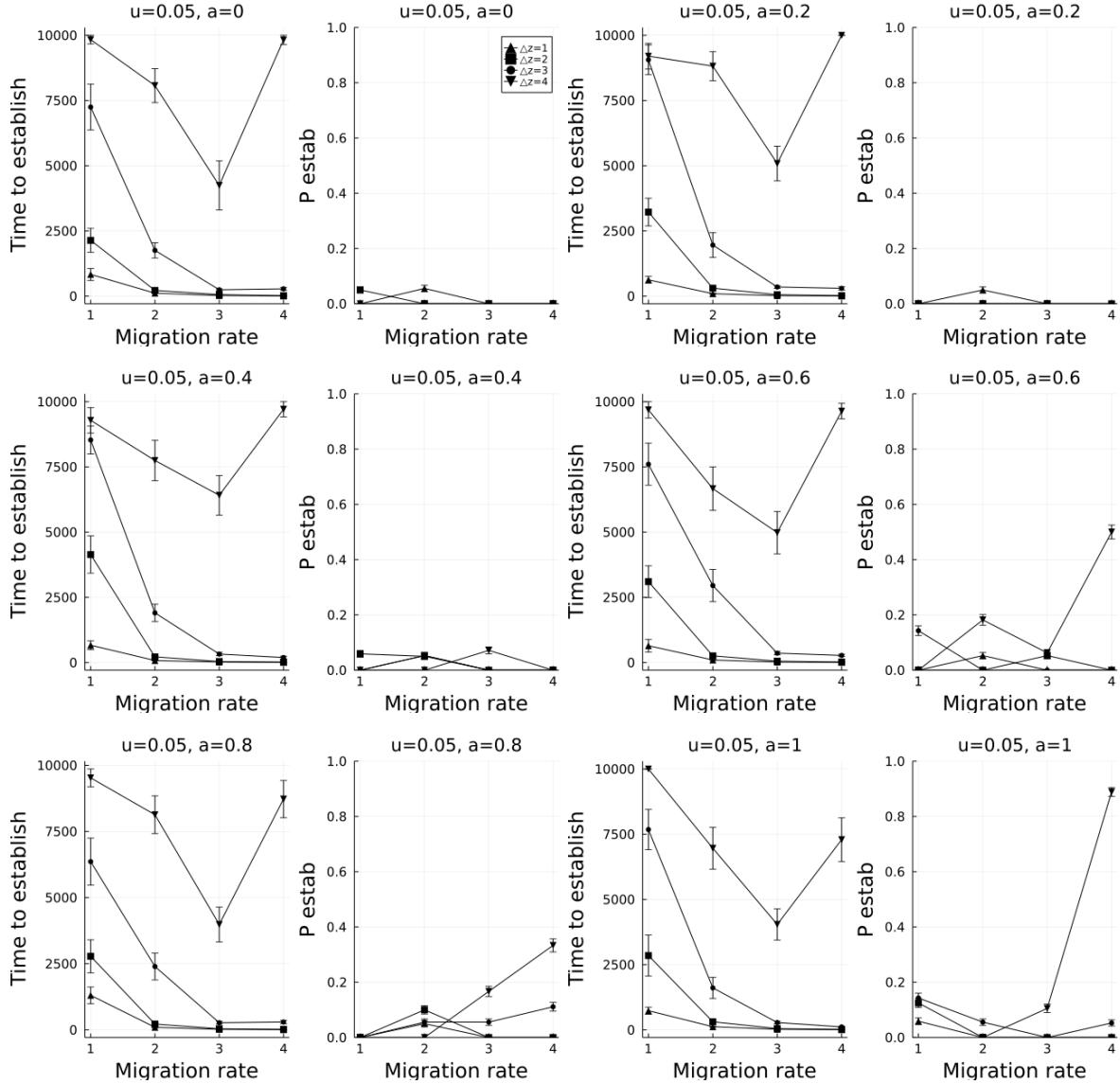
An interesting pattern that arises from these plots, however, is that there seems to be an interaction between migration rate and difference in phenotypic optima between mainland and island on the probability of tetraploid establishment. In general, probabilities of tetraploid establishment seem highest for low to medium migration rates and high differences in phenotypic optima but decrease rapidly for higher migration rates. Some mechanisms can be hypothesized to explain these findings: tetraploids are more shielded from maladaptive gene flow than diploids and maladaptive gene flow is expected to increase for a larger difference in phenotypic optima. The observation that tetraploid establishment decreases for higher levels of migration from the diploid founder population is likely due to MCE since the relative abundance of diploids on the island is expected increase for higher migration rates.

What can also be noted is that for higher values of the parameter  $u$ , the mean time to establishment significantly increases (note that the y-axis for time to establishment is scaled differently for different values of  $u$ ). Since the parameter  $u$  is the only thing that changes and it has been previously shown that there is no inherent difference in time to establishment between diploids and tetraploids, this effect is likely due to an increasing maladaptive cytotype load in the population for higher values of  $u$ .

#### 4.2.4 Effect of breeding system: Assortative mating and self-fertilization

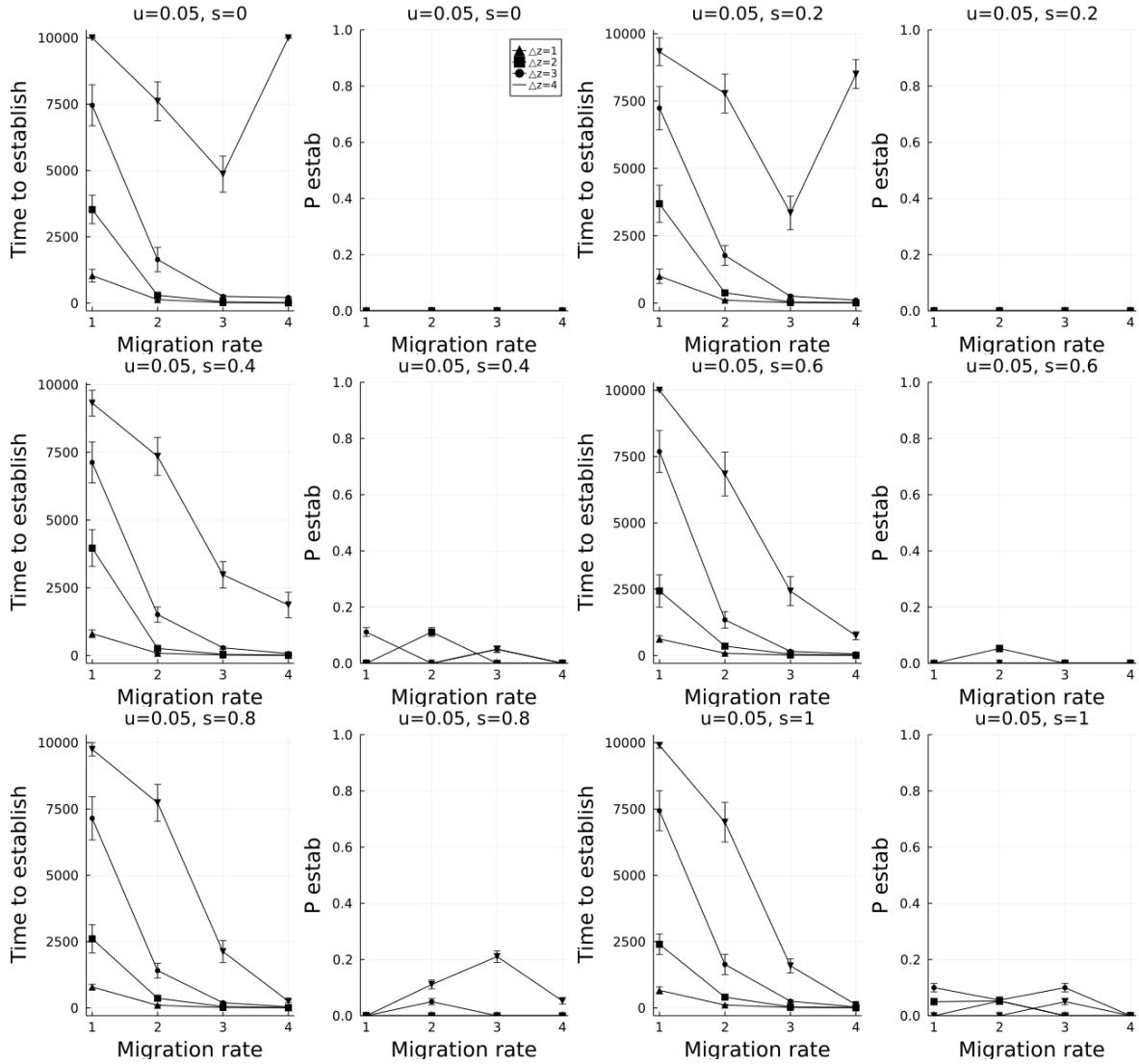
In the previous simulations tetraploid establishment was observed to decrease for higher migration rates. This was hypothesized to be due to MCE caused by an increasing abundance of migrating diploids. An implementation of assortative mating where there is a higher probability for intracytotype mating, as was explained more in detail for the single deme setting, is expected to shield tetraploids from MCE. For higher probabilities of assortative mating, more tetraploid establishment could be expected for higher migration rates than in the case without assortative mating, especially in combination with high differences in phenotypic optima between mainland and island.

Figure 40 shows time to establishment and probability of tetraploid establishment ( $P_{\text{estab}}$ ) for a fixed value of  $u = 0.05$  and different values of assortative mating  $\alpha$ . The pattern that arises here seems to be coherent with what was hypothesized in the previous paragraph: while in the absence of assortative mating or for low probabilities thereof, there is only a very low probability for tetraploid establishment similar to previous simulations, the probability of tetraploid establishment seems to rise significantly for values of  $\alpha$  and higher in the cases where diploids experience the highest maladaptive gene flow, i.e. where there is a combination of high migration rates and a high difference in phenotypic optima.



**Figure 40. Effect of assortative mating on the mean time to establishment and the probability of tetraploid establishment.** For each combination of migration rate ( $M = [0.01 \quad 0.1 \quad 1 \quad 10]$ ) and difference in phenotypic optima ( $\Delta z = [1 \quad 2 \quad 3 \quad 4]$ ). Points show the mean time to establishment of a population (left) and probability of tetraploid establishment ( $P_{\text{estab}}$ ) on the right. Error bars show the standard error on the mean. Upper triangle:  $\Delta z = 1$ , square:  $\Delta z = 2$ , circle:  $\Delta z = 3$ , down triangle:  $\Delta z = 4$ . The following parameters were used for the simulations:  $p = 0.5$ ,  $\alpha = 0.27$ ,  $L = 50$ ,  $d = 1$ ,  $\beta = 0.2$ .

Figure 41 shows time to establishment and probability of tetraploid establishment ( $P_{\text{estab}}$ ) for a fixed value of  $u = 0.05$  and different values of self-fertilization  $s$ . The effect of self-fertilization on tetraploid establishment seems to be almost negligible in this spatial setting. What can be observed however is an effect of self-fertilization on time to establishment of a population on the island: the time to establishment decreases especially for higher migration rates with higher values of  $s$ .



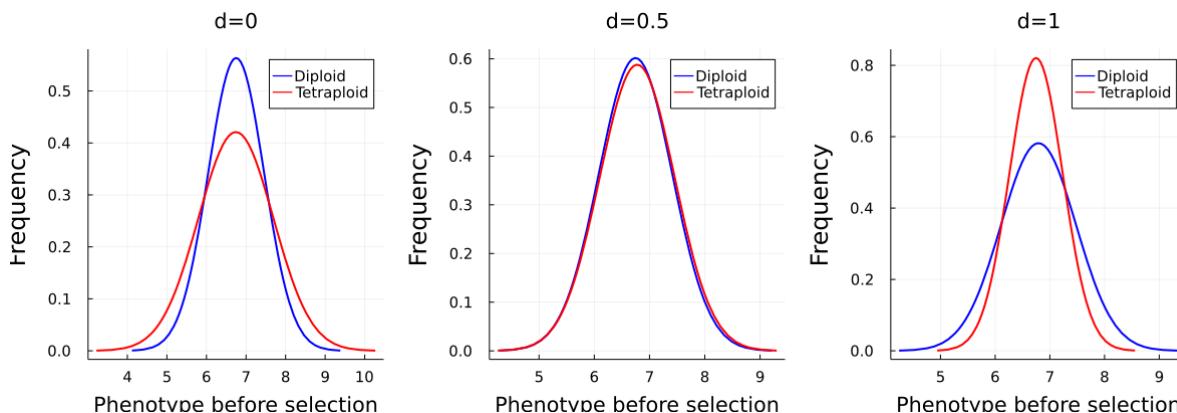
**Figure 41. Effect of self-fertilization on the mean time to establishment and the probability of tetraploid establishment.** For each combination of migration rate ( $M = [0.01 \quad 0.1 \quad 1 \quad 10]$ ) and difference in phenotypic optima ( $\Delta z = [1 \quad 2 \quad 3 \quad 4]$ ). Points show the mean time to establishment of a population (left) and probability of tetraploid establishment ( $P_{\text{estab}}$ ) on the right. Error bars show the standard error on the mean. Upper triangle:  $\Delta z = 1$ , square:  $\Delta z = 2$ , circle:  $\Delta z = 3$ , down triangle:  $\Delta z = 4$ . The following parameters were used for the simulations:  $p = 0.5$ ,  $\alpha = 0.27$ ,  $L = 50$ ,  $d = 1$ ,  $\beta = 0.2$ .

A clear difference can be noted between the effects of assortative mating and self-fertilization on both general establishment and the probabilities of tetraploid establishment. Something interesting to note is that there is a profound effect of self-fertilization rates on the time to establishment that wasn't present for assortative mating. Inbreeding might enhance establishment in the presence of directional selection by decreasing genetic load, comparable to the effect of genetic variance for stabilizing selection as explained before. However, most likely self-fertilization shields from migration load, an effect that seems to be very prominent for higher values of  $s$ , both in diploids and tetraploids. In the case of assortative mating there is an increased frequency of intracytotype mating but this doesn't influence the migration load in diploids. The effect of assortative mating is mostly due to shielding tetraploids from

migration load and MCE caused by diploids. This might explain why higher probabilities of tetraploid establishment are only found with assortative mating. The effect of self-fertilization might however be more pronounced with other gene modes, for example as found by Grisswold (2021) for recessive gene modes in the case of a single bi-allelic locus. The effect of inbreeding depression that arises for higher degrees of self-fertilization is less pronounced in tetraploids for recessive traits compared to diploids.

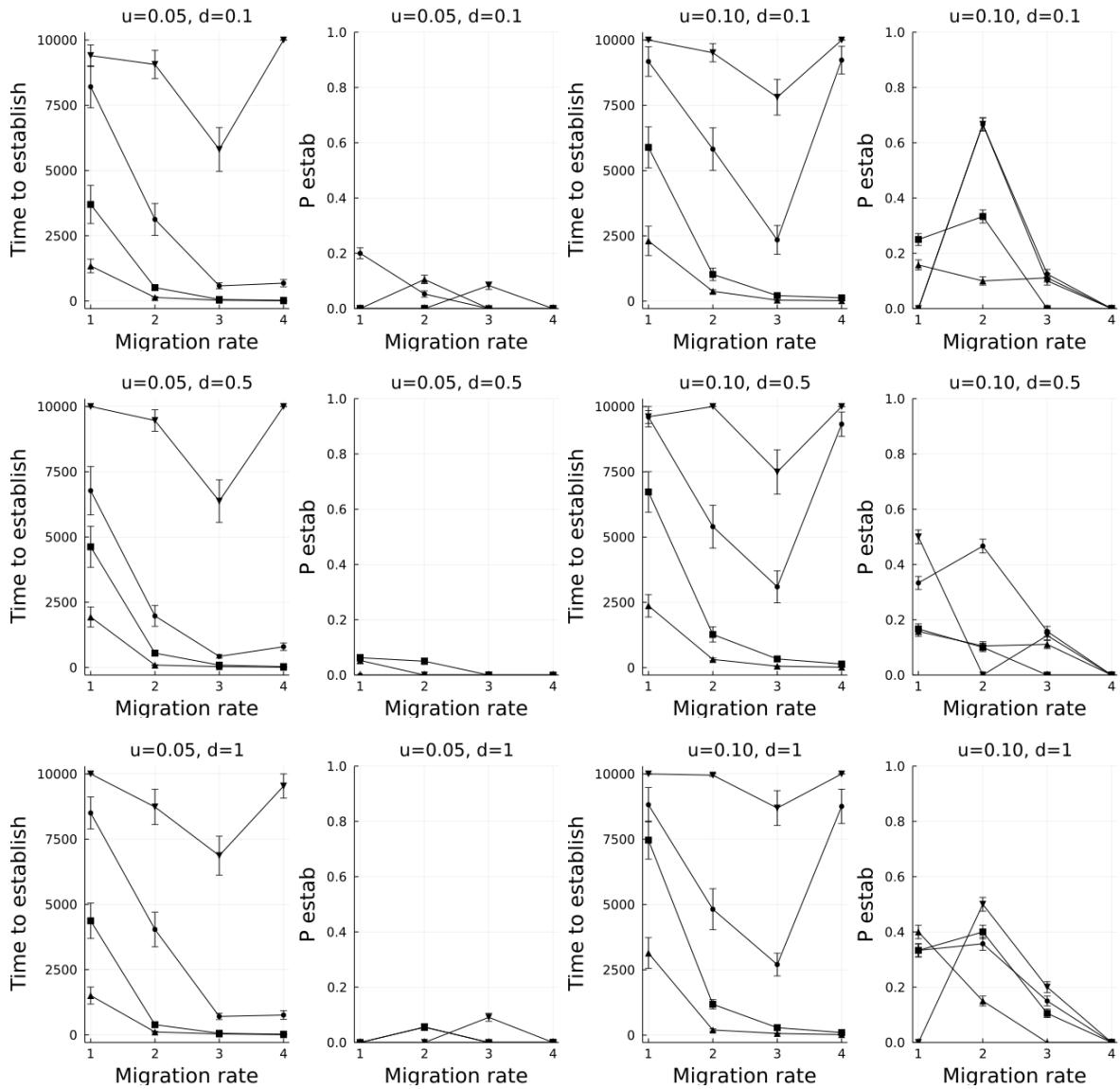
#### 4.2.5 Effect of genotype-phenotype map

There is also a possible effect of the genotype-phenotype map that hasn't been investigated. For the previous simulations a value of  $d = 1$  was used (the maximum phenotypic range of diploids and tetraploids is the same). Smaller values of  $d$  result in more extreme phenotypes occurring in tetraploids compared to diploids which might facilitate overcoming the phenotypic difference between source and sink (see figure 8 for the phenotypic range for different values of  $d$ ). The value of  $d$  also influences the starting phenotypic variation of the source population assumed to be in HWLE as shown in figure 42 that might have a similar effect on establishment (Barton & Etheridge (2018)).



**Figure 42. Effect of genotype-phenotype map for different values of  $d$  on the phenotypic mean and variance of the founder population.** The following parameters were used for the simulations:  $p = 0.5$ ,  $\alpha = 0.27$ ,  $L = 50$ .

Figure 43 shows the effect of the genotype-phenotype map for different values of  $d$ , again on the mean time to establishment and the probability of tetraploid establishment. Clearly there is an effect of the choice for the parameter  $u$  on tetraploid establishment but there doesn't seem to be a differential effect caused by altering the parameter  $d$ .



**Figure 43. Effect of genotype-phenotype map for different values of  $d$  on the mean time to establishment and the probability of tetraploid establishment.** For each combination of migration rate ( $M = [0.01 \quad 0.1 \quad 1 \quad 10]$ ) and difference in phenotypic optima ( $\Delta z = [1 \quad 2 \quad 3 \quad 4]$ ). Points show the mean time to establishment of a population (left) and probability of tetraploid establishment ( $P_{\text{estab}}$ ) on the right. Error bars show the standard error on the mean. Upper triangle:  $\Delta z = 1$ , square:  $\Delta z = 2$ , circle:  $\Delta z = 3$ , down triangle:  $\Delta z = 4$ . The following parameters were used for the simulations:  $p = 0.5$ ,  $\alpha = 0.27$ ,  $L = 50$ ,  $d = 1$ ,  $\beta = 0.2$ .

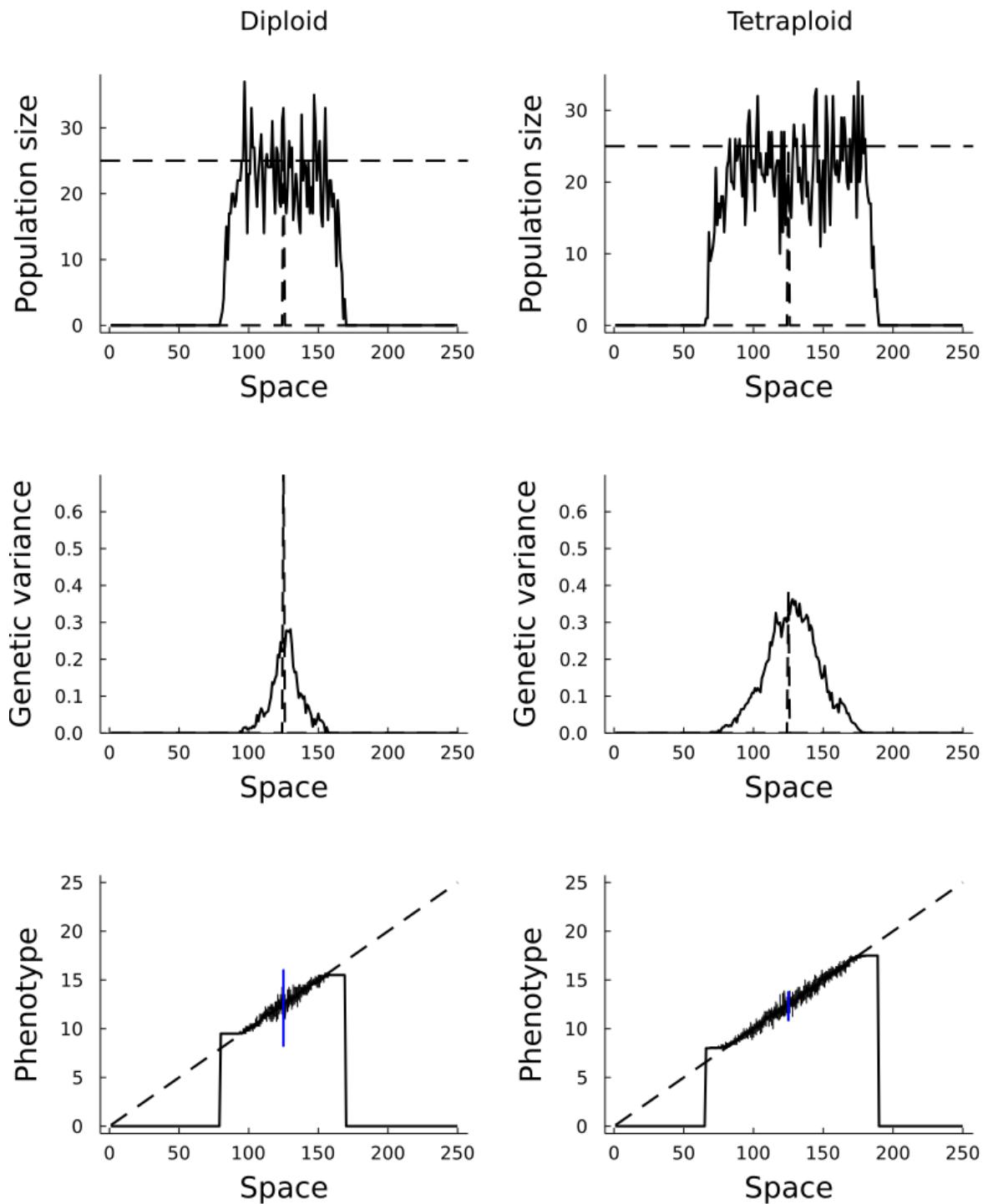
Since for diploids the genetic variance stays constant for all values of  $d$  it is currently not clear if the mean time to establishment is expected to stay the same for different values of  $d$  with only diploid migrants from the source population: the variance of the founder population in HWLE stays the same for different values of  $d$ . The effect of  $d$  on tetraploid establishment that is now not clearly visible might be more pronounced for different selection schemes however that are not discussed in this dissertation but Polechová & Barton (2015). Bell et al. (2021) describes hard selection as a type of selection where the fitness of an individual only depends on the difference between its phenotypic value and the environmental optimum under stabilizing selection. It is independent of the phenotypic composition of the rest of the

population. With soft selection on the other hand, the fitness of an individual is relative to the phenotypic value of other individuals in the population. Truncation selection can be thought of as an example of soft selection if implemented with a frequency-based cut-off. A certain fixed percentage of highest fittest individuals of a population is selected for mating. Soft selection might result in more pronounced effects on polyploid establishment in combination with phenotypic scaling in the context of the source-sink scenario if tetraploids have for example a higher phenotypic range compared to diploids ( $d < 1$ ).

### 4.3 Linear habitat

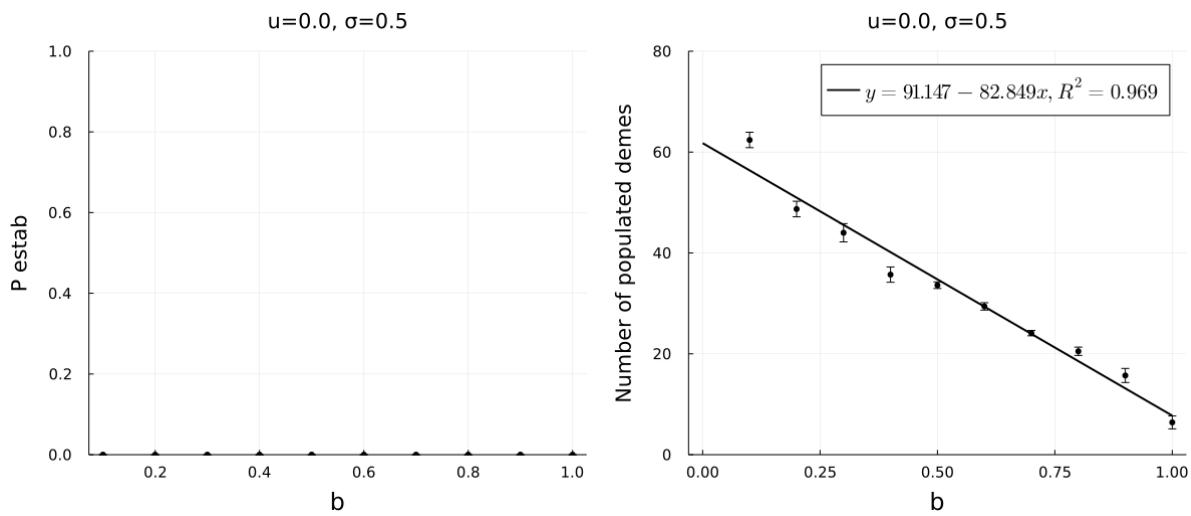
Moving onward from a mainland-island model, polyploid establishment can be explored in more complex spatial settings. These simulations are inspired by Polechová & Barton (2015), whose model was adapted and extended to mixed-ploidy populations. A population is started from the central deme and subsequently migrates outwards along a linear environmental gradient (see materials and methods for details).

First some preliminary simulations were run for single ploidy populations to compare the dynamics of species range evolution in diploids versus tetraploids. Figure 44 illustrates the results of a single simulation after 500 generations starting from a single populated deme in the center of the habitat. What can be noticed from the upper plots for the population size is that the range that is occupied after 500 generations by tetraploids is larger compared to diploids. Since the parameter  $d = 1$  for the genotype-phenotype map, the starting genetic variance of tetraploids is half to that of diploids. However, after 500 generations the genetic variance in tetraploids can be seen in the middle plots to be higher compared to diploids, indicating that, in agreement with previous findings, the eroding effect of genetic drift is again less pronounced in tetraploids. The bottom plots show how the phenotype of the individuals evolves and manages to track the linear gradient of the habitat.

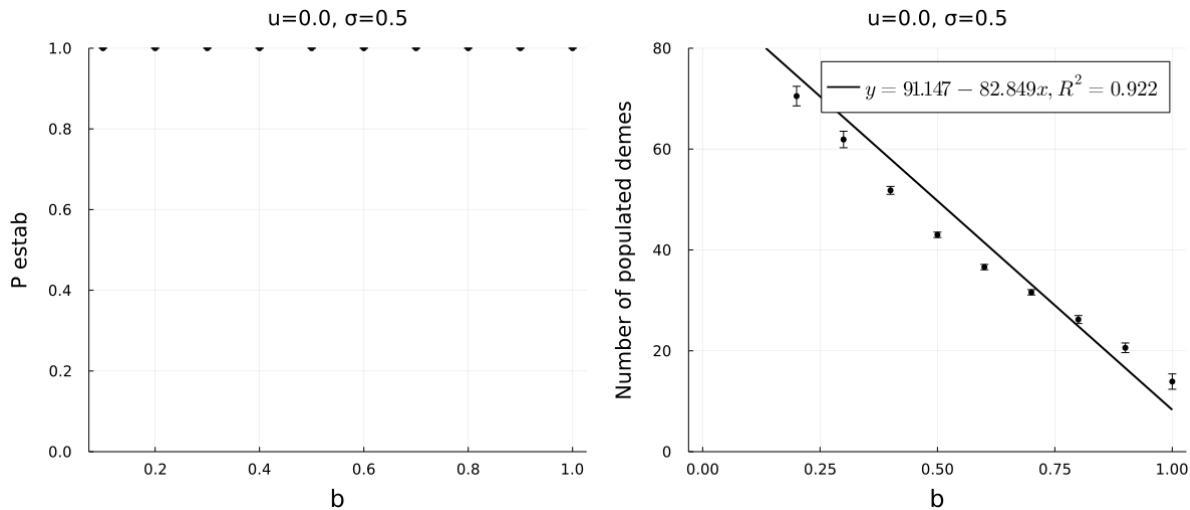


**Figure 44. Illustration of a single simulation with single ploidy populations after 500 generations for a linear habitat of 250 demes.** The simulation was initiated from a single deme in the center of the habitat with a diploid population of 25 individuals. It gives an overview of the model output. On top it shows the evolution of population size through space. The dashed vertical line shows the starting population size. In the middle the evolution of genetic variance can be observed with the dashed vertical line showing genetic variance of the starting population. At the bottom the phenotypic mean and the phenotype for all individuals are plotted. The starting phenotypic variance is shown in blue. The dashed diagonal line shows the phenotypic gradient. The following parameters were used for the simulations:  $K = 25$ ,  $p = 0.5$ ,  $\alpha = 0.5$ ,  $L = 50$ ,  $d = 1$ ,  $u = 0$ ,  $r_m = 1.06$ ,  $V_s = 1$ ,  $\mu = 10^{-6}$ ,  $\sigma = 0.5$ .

The difference in range expansion between diploids and tetraploids was analyzed more in depth, the results of which are shown in figure 45 and figure 46 for diploids and tetraploids respectively. The most important thing that comes to the eye is that for low values of  $b$  further range expansion can be seen for a starting tetraploid population. It could be hypothesized that this is due to the different in erosion of genetic variance by drift for different ploidy levels since genetic drift is a major determinant in species range expansions (Polechová & Barton, 2015). However, the relationship between the number of populated demes and  $b$  also decreases more steeply and for higher values of  $b$  (steeper environmental gradients) this effect seems to diminish.

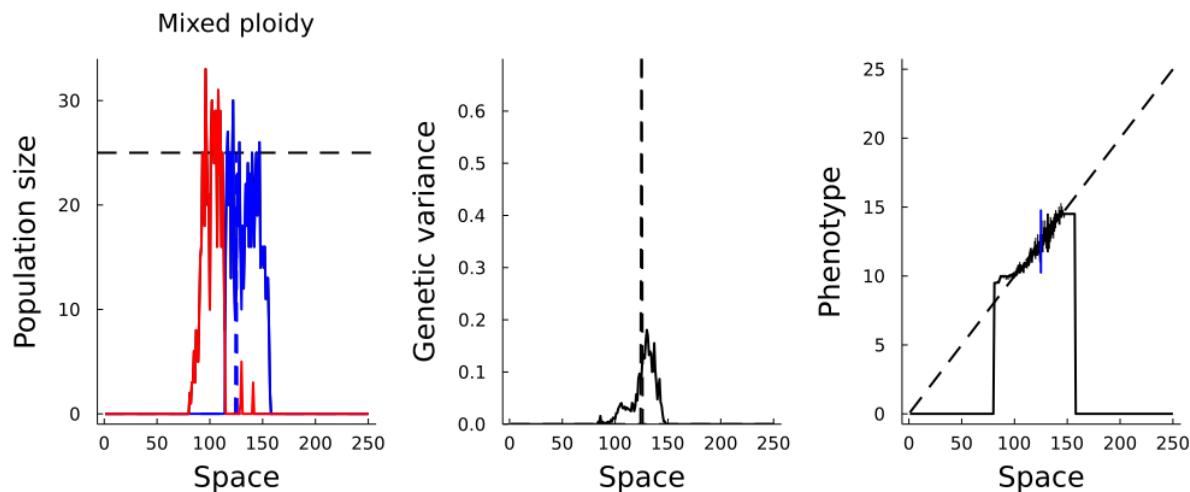


**Figure 45. Range expansion in a linear habitat starting from a central single ploidy diploid deme.** The left panel shows the probability of tetraploid establishment which is obviously zero in the case of a diploid only population without unreduced gamete formation and was more implemented to test the proper working of the model. The panel on the right shows the mean number of populated demes (10 simulations) for different values of  $b$  for which a linear curve was fitted. The following parameters were used for the simulations:  $K = 25$ ,  $p = 0.5$ ,  $\alpha = 0.5$ ,  $L = 50$ ,  $d = 1$ ,  $u = 0$ ,  $r_m = 1.06$ ,  $V_s = 1$ ,  $\mu = 10^{-6}$ ,  $\sigma = 0.5$ .



**Figure 46. Range expansion in a linear habitat starting from a central single ploidy tetraploid deme.** The left panel shows the probability of tetraploid establishment which is obviously one in the case of a tetraploid only population since all populations are expected to be tetraploid. This was more implemented to test the proper working of the model. The panel on the right shows the mean number of populated demes (10 simulations) for different values of  $b$  for which a linear curve was fitted. The following parameters were used for the simulations:  $K = 25$ ,  $p = 0.5$ ,  $\alpha = 0.5$ ,  $L = 50$ ,  $d = 1$ ,  $u = 0$ ,  $r_m = 1.06$ ,  $V_S = 1$ ,  $\mu = 10^{-6}$ ,  $\sigma = 0.5$ .

Figure 47 shows an illustration of a mixed-ploidy simulation where a diploid population (blue) starts in the central deme of a linear habitat. After 500 generations a tetraploid population (red) is established at the edge of this diploid population. The model also output other data that hasn't been fully analyzed yet in this dissertation but might lead to interesting insight: the evolution of the phenotypic mean and variance and the evolution of genetic variance along the linear gradient and how they play their role in polyploid establishment.

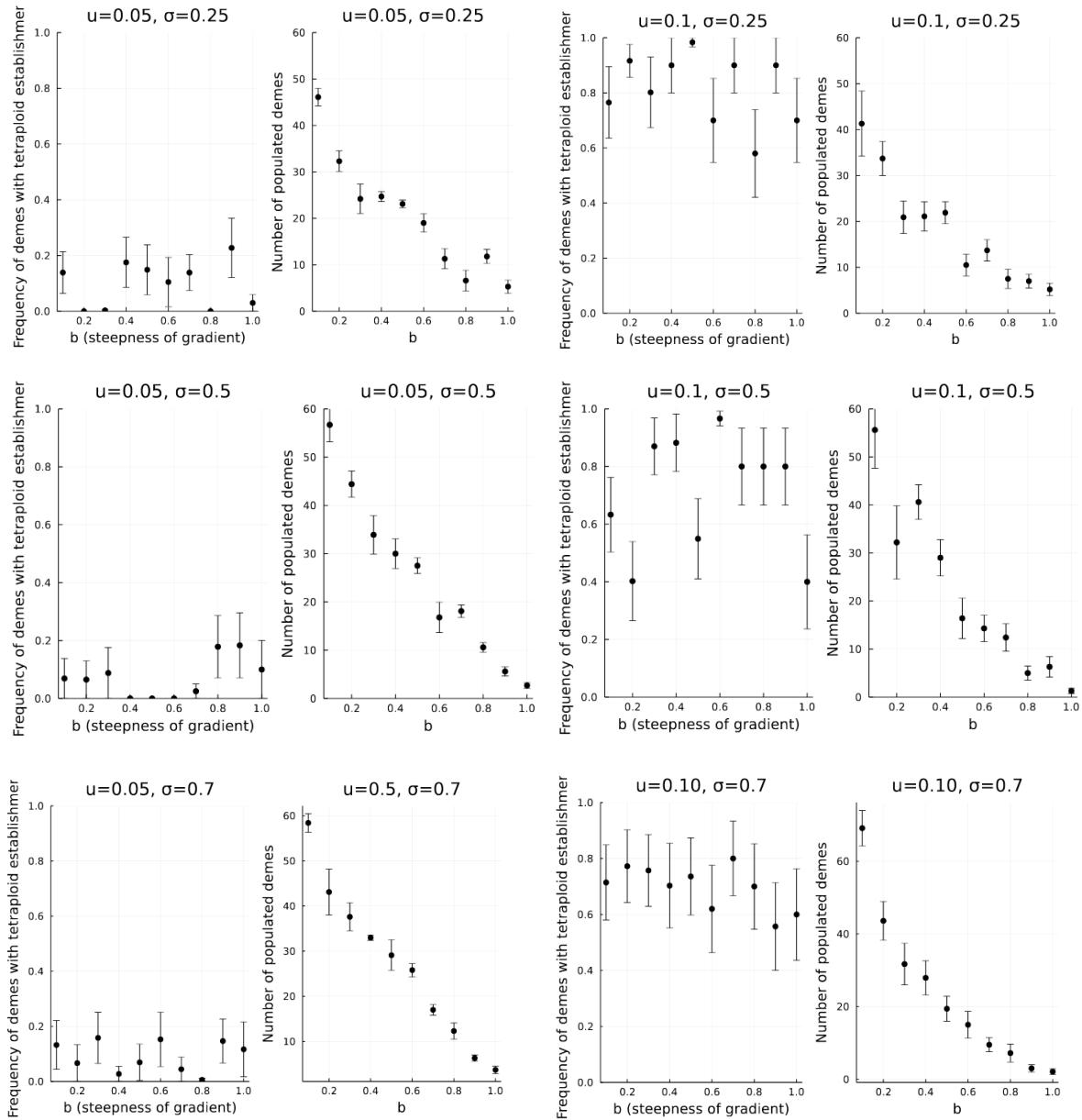


**Figure 47. Illustration of a single simulation with a mixed-ploidy population after 500 generations for a linear habitat of 250 demes.** The simulation was initiated from a single deme in the center of the habitat with a diploid population of 25 individuals. On the left the population sizes through space are shown for tetraploids (red) and diploids (blue). In the middle the evolution of genetic variance can be observed with the dashed vertical line showing genetic variance of the starting population. At the

bottom the phenotypic mean and the phenotype for all individuals are plotted. The starting phenotypic variance is shown in blue. The dashed diagonal line shows the phenotypic gradient. The following parameters were used for the simulations:  $K = 25$ ,  $p = 0.5$ ,  $\alpha = 0.5$ ,  $L = 50$ ,  $d = 1$ ,  $u = 0.07$ ,  $r_m = 1.06$ ,  $V_s = 1$ ,  $\mu = 10^{-6}$ ,  $\sigma = 0.5$ .

Finally, some simulations were done that relate back to the core hypothesis of this dissertation: that maladaptive gene flow plays an important role in polyploid establishment. Figure 48 shows the result of the simulations along a linear gradient with different steepness  $b$  ranging from 0.1 to 1 (increasing steepness of the gradient) and for different migration rates  $\sigma$ , comparable to the difference in phenotypic optima and migration rates that were part of the source-sink simulations. For every combination of parameters  $(u, \sigma)$  the ratio of demes that are populated by tetraploids to all populated demes is shown (left) versus the total number of populated demes (right). The Upper row shows results for the case with low frequency of unreduced gamete formation ( $u = 0.05$ ) and the bottom row for higher frequencies ( $u = 0.10$ ).

The only two things that seems to jump out right away are higher frequencies of demes populated by tetraploids for higher value of  $u$  and the difference in habitat range for different values of both  $b$  and  $\sigma$ , but there does not seem to be a clear interaction between migration load and polyploid establishment. Likely the effect of maladaptive gene flow will need to be reinforced by a similar effect as was seen for the source-sink model, either by implementing different levels of assortative mating or self-fertilization. There is also a possible effect of the genotype-phenotype map that hasn't been investigated. For these simulations a value of  $d = 1$  was used (the maximum phenotypic range of diploids and tetraploids is the same).

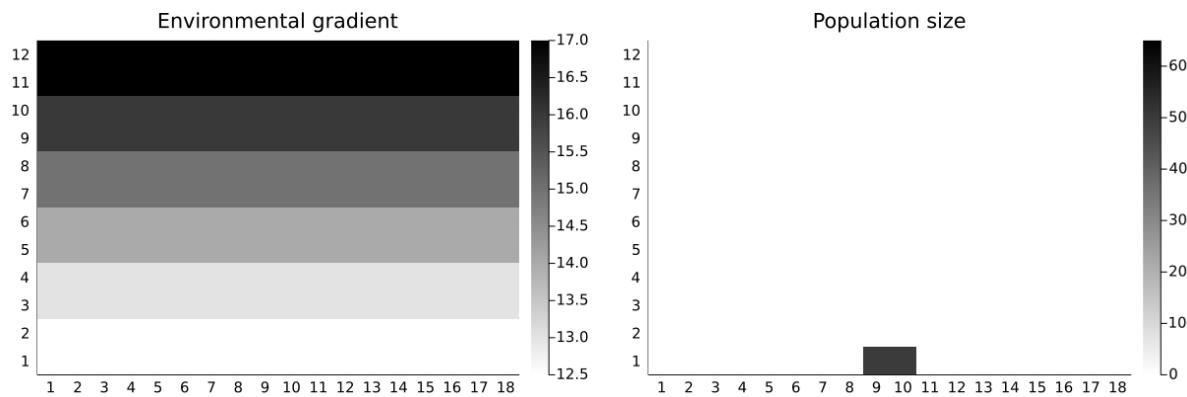


**Figure 48. Effect of the steepness of the environmental gradient  $b$  and migration rate  $\sigma$  on tetraploid establishment.** The following parameters were used for the simulations:  $K = 25$ ,  $p = 0.5$ ,  $\alpha = 0.5$ ,  $L = 50$ ,  $d = 1$ ,  $r_m = 1.06$ ,  $V_s = 1$ ,  $\mu = 10^{-6}$ .

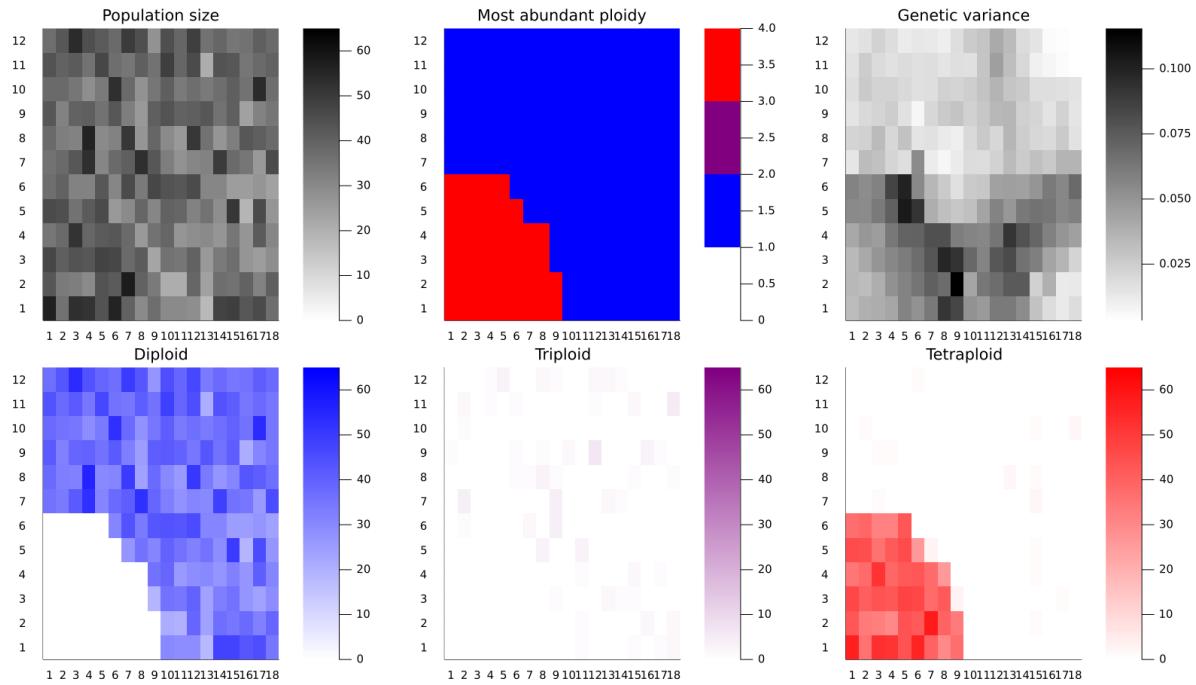
#### 4.4 Habitat in two dimensions

There are some indications that it might be worthwhile to investigate the effect of the dimensionality of a habitat on polyploid establishment. As described by Polechová (2018) the effects of dispersal in species range expansion might differ fundamentally for a two-dimensional habitat compared to a linear habitat: the effect of genetic drift becomes independent of the strength selection. As elaborated on in the introduction, Spoelhof et al. (2020) showed with a simple spatial model that habitats approximating a linear shape (long but small, approaching a one-dimensional species range) might enhance polyploid establishment compared to more square-shaped habitats (two-dimensional).

Extending the model that was previously described for a one-dimensional linear gradient to a two-dimensional space was straightforward: the two-dimensional habitat is represented as an  $m \times n$  matrix of  $mn$  demes. Migration was implemented as a random walk to a random neighboring deme with a probability  $\sigma$  and stops at the edges of the habitat. Figure 49 shows an exemplary start of a simulation. An example of a single simulation in a 2D habitat is shown in Figure 50 where a population is initiated in two demes at the center bottom and with a linear gradient that increases from top to bottom. The output of the model per generation is the total population size per deme, population size for each ploidy level, the most abundant ploidy level and the genetic variance per deme.



**Figure 49. Initiation of the simulation for a two-dimensional habitat.** On the left a linear phenotypic gradient is shown increasing from bottom to top by jumps in phenotypic optima of  $\Delta = 1$ . The plot on the right shows an example starting point of a simulation with two populated demes. The phenotypic mean of the starting population corresponds to the mean of these demes (see materials and methods for further details). From here the population will expand and attempt to climb up the phenotypic gradient.



**Figure 50. Illustration of a single simulation after 200 generations for a two-dimensional habitat (12x18 demes).** The simulation was initiated from two demes in the bottom center of the habitat with a diploid population of 50 individuals each. There is a linear phenotypic gradient increasing from bottom to top by jumps in phenotypic optima of  $\Delta = 1$ . A tetraploid population spanning several demes has established in the left bottom corner of the habitat (in red) at the edge of a diploid population (in blue). The following parameters were used for the simulation:  $K = 50$ ,  $p = 0.5$ ,  $\alpha = 0.5$ ,  $L = 50$ ,  $d = 1$ ,  $u = 0.05$ ,  $r_m = 1.06$ ,  $V_s = 0.5$ ,  $\mu = 10^{-6}$ ,  $\sigma = 0.1$ .

The plots shown here are more meant to be illustrative since the two-dimensional model hasn't been analyzed. From preliminary visual inspections however, tetraploids do seem to establish most frequently at the edge of a diploid expansion front. Sometimes it was also observed that tetraploids take over the whole habitat sometime after initial establishment. This can possibly be caused by a ratchet effect (as explained in the introduction; also Levin, 2006). Since ploidy levels higher than tetraploids are assumed to be inviable and there is no rediploidization in the model, once a tetraploid population is established it cannot return to a diploid state. It is also important to consider edge effects in these simulations since the space is limited; there is for example less maladaptive gene flow at the edges of the habitat, and it is yet unclear how effect like this impact polyploid establishment (Polechová & Barton, 2015).

## 5 Conclusion

While this dissertation ended up more as an exploration of what individual-based models in an integrated eco-evolutionary framework can contribute to studying polyploid establishment than as an in-depth analysis, some aspects of the implemented model are possibly a promising avenue for further research. Quantitative traits with an explicitly encoded genetic architecture and a flexible genotype-phenotype map for example have been illustrated throughout this dissertation as an interesting approach to study the evolution of genetic variance in mixed-ploidy systems through space and time.

However, there are already a few main results that might shed some light on the initially formulated hypotheses on conditions that might facilitate polyploid establishment. Small population size in a newly colonized habitat or at the periphery of a species range results in more pronounced stochasticity (demographic stochasticity and genetic drift). One of the main observations from simulations is that the effects of ploidy level on drift are very pronounced. Genetic drift erodes genetic variance much less rapidly in tetraploids compared to diploids. However, it is important to note that in a constant environment a higher genetic variance also comes with a cost in the presence of stabilizing selection. Another remark is that for single ploidy simulations to study these effects, populations always started with maximized genetic variance. Simulations under these conditions may shed insight on what evolutionary factors affect the persistence of a newly formed polyploid lineage if some genetic variance is already present in the population. Studying the effects of genetic load and rate of adaptation may be important for longer-term effects on the persistence and diversification of polyploids once formed and they are important. Studying why polyploids can persist on the term might be important though, since unreduced gamete formation comes with a cost, a maladaptive load that could be called cytotype load, that was clearly present in the simulations (Otto and Whitton, 2000).

The genetic variance following a founder event is expected to be much smaller. An important thing to note is that establishment of autopolyploid populations through fusion of unreduced gametes, due to the rarity of this event, will result in a profound genetic bottleneck and the evolution of genetic variation after such a founder event will typically depend multiple aspects. Some that come to mind are migration and the rate of unreduced gamete formation. Higher rates of unreduced gamete formation will contribute to intercytotypic gene flow (for example in a diploids-tetraploids mixed-ploidy population through fusion of unreduced gametes from diploids with reduced gametes from tetraploids) but also simply due to higher probability of multiple establishment events, increasing genetic variance of the autopolyploid population (Arnold et al., 2012). These aspects weren't studied in depth in this dissertation.

The ecological effects of polyploidization are more likely to determine establishment in the short-term (Otto and Whitton, 2000). Simulations in this dissertation clearly showed an important effect of assortative mating on the probability of tetraploid establishment. However, these aren't very surprising and were already shown in other modeling papers (for example Baack (2005)). An interesting finding however is that there seems to be a differential effect

between self-fertilization and assortative mating. There is a profound effect of self-fertilization rates on the time to establishment that wasn't present for assortative mating. Most likely self-fertilization shields from migration load, an effect that seems to be very prominent for higher values of  $s$ , both in diploids and tetraploids. In the case of assortative mating there is an increased frequency of intracytotype mating but this doesn't influence the migration load in diploids. The effect of assortative mating is mostly due to shielding tetraploids from migration load and MCE caused by diploids. This example shows the potential of the modeling approach in this dissertation, where different mechanisms come together in an integrated way: in this case the quantitative genetics through the effect of inbreeding on genetic variance, migration load due to the spatial heterogeneity and life cycle aspects.

Since the individual-model is implemented in a modular way, it can easily be extended. Some possible future implementations might include the possibility of a triploid bridge, where triploids were now considered to be unviable. This could have large impacts on the probabilities of tetraploid establishment (Ramsey et al., 1998). Making the frequency of unreduced gamete formation a trait on the individual level would make sense as there is a lot of individual difference observed in nature (Kreiner, 2017), where it is now implemented on the level of the deme, or in extenso an evolvable quantitative trait (that responds for example to external stress since this has been shown to increase unreduced gamete formation). The same goes for other aspects of the genetic architecture that are now implemented on deme level for simplicity: allelic effect size, number of loci, ... but could ideally be implemented on the level of individual agents. In this context also other models for gene mode could be explored (dominant, recessive, ... epistasis). Life cycle aspects that can be added are for example age structure to allow studying the effects of perennials. It would be interesting to pursue the further combination of all these aspects into an integrated model to allow for further insights in the eco-evolutionary dynamics behind polyploid establishment. As was explained in the introduction, understanding how interactions at different scales (from quantitative traits to gene regulatory networks or complex traits) affect processes at higher scales (up to species interaction networks) might be the way forward in eco-evolutionary research (Govaert et al., 2019) but also specifically applied to the problem of polyploid establishment. The individual-based model presented in this dissertation could be easily implemented in a multiscale modeling approach (especially due to the ease of integration between different Julia packages for which some frameworks already exist for multi-scale modeling but are also being developed, e.g. MultiScaleArrays.jl). However, it will also be of paramount importance to parametrize the model on real data. Now this dissertation lacks the use of empirical data, though mostly because it was meant to implement an interesting model and study some basic aspects of it, but this would be paramount for future research.

## 6 Bibliography

- Andrello, M., Noirot, C., Débarre, F., & Manel, S. (2021). MetaPopGen 2.0: A multilocus genetic simulator to model populations of large size.
- Arnold BJ, Bomblies K, Wakeley J (2012) Extending coalescent theory to autotetraploids. *Genetics* 192:195–204.
- Baack, E. J. (2005). To succeed globally, disperse locally: effects of local pollen and seed dispersal on tetraploid establishment. *Heredity*, 94(5), 538-546.
- Baniaga, A. E., Marx, H. E., Arrigo, N., & Barker, M. S. (2020). Polyploid plants have faster rates of multivariate niche differentiation than their diploid relatives. *Ecology letters*, 23(1), 68-78.
- Barker, M. S., Arrigo, N., Baniaga, A. E., Li, Z., & Levin, D. A. (2016). On the relative abundance of autopolyploids and allopolyploids. *New Phytologist*, 210(2), 391-398.
- Barton N (2001) Adaptation at the edge of a species' range. Integrating Ecology and Evolution in a Spatial Context, eds Silvertown J, Antonovics J (Blackwell, London), pp 365–392.
- Barton, N. H., & Etheridge, A. M. (2018). Establishment in a new habitat by polygenic adaptation. *Theoretical Population Biology*, 122, 110-127.
- Barton, N. H., Etheridge, A. M., & Véber, A. (2013). Modelling evolution in a spatial continuum. *Journal of Statistical Mechanics: Theory and Experiment*, 2013(01), P01002.
- Bell, D. A., Kovach, R. P., Robinson, Z. L., Whiteley, A. R., & Reed, T. E. (2021). The ecological causes and consequences of hard and soft selection. *Ecology Letters*, 24(7), 1505-1521.
- Bezanson, J., Karpinski, S., Shah, V. B., & Edelman, A. (2012). Julia: A fast dynamic language for technical computing. arXiv preprint arXiv:1209.5145.
- Bradbard, G. S., & Ralph, P. L. (2019). Spatial population genetics: it's about time. *Annual Review of Ecology, Evolution, and Systematics*, 50, 427-449.
- Brady, S. P., Bolnick, D. I., Angert, A. L., Gonzalez, A., Barrett, R. D., Crispo, E., ... & Hendry, A. P. (2019). Causes of maladaptation. *Evolutionary Applications*, 12(7), 1229-1242.
- Brady, S. P., Bolnick, D. I., Barrett, R. D., Chapman, L., Crispo, E., Derry, A. M., ... & Hendry, A. (2019). Understanding maladaptation by uniting ecological and evolutionary perspectives. *The American Naturalist*, 194(4), 495-515.
- Bridle, J. R., Kawata, M., & Butlin, R. K. (2019). Local adaptation stops where ecological gradients steepen or are interrupted. *Evolutionary Applications*, 12(7), 1449-1462.
- Carvajal-Rodríguez, A. (2010). Simulation of genes and genomes forward in time. *Current genomics*, 11(1), 58-61.
- Castro, M., Loureiro, J., Figueiredo, A., Serrano, M., Husband, B. C., & Castro, S. (2020). Different patterns of ecological divergence between two tetraploids and their diploid counterpart in a parapatric linear coastal distribution polyploid complex. *Frontiers in plant science*, 11, 315.
- Chevin, L. M., Cotto, O., & Ashander, J. (2017). Stochastic evolutionary demography under a fluctuating optimum phenotype. *The American Naturalist*, 190(6), 786-802.

- Chrtěk, J., Herben, T., Rosenbaumová, R., Münzbergová, Z., Dočkalová, Z., Zahradníček, J., ... & Trávníček, P. (2017). Cytotype coexistence in the field cannot be explained by inter-cytotype hybridization alone: linking experiments and computer simulations in the sexual species *Pilosella echioides* (Asteraceae). *BMC Evolutionary Biology*, 17(1), 1-14.
- Cotto, O., Wessely, J., Georges, D., Klonner, G., Schmid, M., Dullinger, S., ... & Guillaume, F. (2017). A dynamic eco-evolutionary model predicts slow response of alpine plants to climate warming. *Nature Communications*, 8(1), 1-9.
- Crespi, B. J. (2000). The evolution of maladaptation. *Heredity*, 84(6), 623-629.
- DeAngelis, D. L., & Mooij, W. M. (2005). Individual-based modeling of ecological and evolutionary processes. *Annu. Rev. Ecol. Evol. Syst.*, 36, 147-168.
- Felber, F. (1991). Establishment of a tetraploid cytotype in a diploid population: effect of relative fitness of the cytotypes. *Journal of evolutionary biology*, 4(2), 195-207.
- Fowler, N. L., & Levin, D. A. (1984). Ecological constraints on the establishment of a novel polyploid in competition with its diploid progenitor. *The American Naturalist*, 124(5), 703-711.
- Fowler, N. L., & Levin, D. A. (2016). Critical factors in the establishment of allopolyploids. *American Journal of Botany*, 103(7), 1236-1251.
- Fox, D. T., Soltis, D. E., Soltis, P. S., Ashman, T. L., & Van de Peer, Y. (2020). Polyploidy: a biological force from cells to ecosystems. *Trends in cell biology*.
- García-Ramos, G., & Kirkpatrick, M. (1997). Genetic models of adaptation and gene flow in peripheral populations. *Evolution*, 51(1), 21-28.
- Govaert, L., Fronhofer, E. A., Lion, S., Eizaguirre, C., Bonte, D., Egas, M., ... & Matthews, B. (2019). Eco-evolutionary feedbacks—Theoretical models and perspectives. *Functional Ecology*, 33(1), 13-30.
- Griswold, C. K. (2021). The effects of migration load, selfing, inbreeding depression, and the genetics of adaptation on autotetraploid versus diploid establishment in peripheral habitats. *Evolution*, 75(1), 39-55.
- Haller, B. C., & Messer, P. W. (2019). Evolutionary modeling in SLiM 3 for beginners. *Molecular biology and evolution*, 36(5), 1101-1109.
- Hoffmann, A. A., Sgrò, C. M., & Kristensen, T. N. (2017). Revisiting adaptive potential, population size, and conservation. *Trends in Ecology & Evolution*, 32(7), 506-517.
- Holt, R. D., & Barfield, M. (2011). Theoretical perspectives on the statics and dynamics of species' borders in patchy environments. *The American Naturalist*, 178(S1), S6-S25.
- Husband, B. C., Baldwin, S. J., & Suda, J. (2013). The incidence of polyploidy in natural plant populations: major patterns and evolutionary processes. In *Plant genome diversity volume 2* (pp. 255-276). Springer, Vienna.
- Kardos, M., Armstrong, E., Fitzpatrick, S. W., Hauser, S., Hedrick, P., Miller, J., ... & Funk, W. C. (2021). The crucial role of genome-wide genetic variation in conservation. *bioRxiv*.
- Kirkpatrick, M., & Barton, N. H. (1997). Evolution of a species' range. *The American Naturalist*, 150(1), 1-23.

- Kreiner, J. M., Kron, P., & Husband, B. C. (2017). Frequency and maintenance of unreduced gametes in natural plant populations: associations with reproductive mode, life history and genome size. *New Phytologist*, 214(2), 879-889.
- Lande, R., & Shannon, S. (1996). The role of genetic variation in adaptation and population persistence in a changing environment. *Evolution*, 434-437.
- Levin, D. A. (1975). Minority cytotype exclusion in local plant populations. *Taxon*, 24(1), 35-43.
- Levin, D. A. (2020). Has the polyploid wave ebbed? *Frontiers in plant science*, 11, 251.
- Li, B. H., Xu, X. M., & Ridout, M. S. (2004). Modelling the establishment and spread of autotetraploid plants in a spatially heterogeneous environment. *Journal of Evolutionary Biology*, 17(3), 562-573.
- Li, Z., McKibben, M. T., Finch, G. S., Blischak, P. D., Sutherland, B. L., & Barker, M. S. (2021). Patterns and processes of diploidization in land plants. *Annual Review of Plant Biology*, 72.
- Lynch, M., & Walsh, B. (1998). Genetics and analysis of quantitative traits.
- Meyers, L. A., & Levin, D. A. (2006). On the abundance of polyploids in flowering plants. *Evolution*, 60(6), 1198-1206.
- Monnahan, P., & Brandvain, Y. (2020). The effect of autoploidy on population genetic signals of hard sweeps. *Biology letters*, 16(2), 20190796.
- Moody M, Mueller L, Soltis DE (1993) Genetic-variation and random drift in autotetraploid populations. *Genetics* 134:649–657.
- Moura, R. F., Queiroga, D., Vilela, E., & Moraes, A. P. (2021). Polyploidy and high environmental tolerance increase the invasive success of plants. *Journal of plant research*, 134(1), 105-114.
- Neuenschwander, S., Michaud, F., & Goudet, J. (2019). QuantiNemo 2: A Swiss knife to simulate complex demographic and genetic scenarios, forward and backward in time. *Bioinformatics*, 35(5), 886-888.
- Oswald, B. P., & Nuismer, S. L. (2011). A unified model of autoploidy establishment and evolution. *The American Naturalist*, 178(6), 687-700.
- Oswald, B. P., & Nuismer, S. L. (2011). A unified model of autoploidy establishment and evolution. *The American Naturalist*, 178(6), 687-700.
- Otto, S. P., & Whitton, J. (2000). Polyploid incidence and evolution. *Annual review of genetics*, 34(1), 401-437.
- Otto, S. P., & Whitton, J. (2000). Polyploid incidence and evolution. *Annual review of genetics*, 34(1), 401-437.
- Pennington, L. K., Slatyer, R. A., Ruiz-Ramos, D. V., Veloz, S. D., & Sexton, J. P. (2021). How is adaptive potential distributed within species ranges?. *Evolution*.
- Polechová, J. (2018). Is the sky the limit? On the expansion threshold of a species' range. *PLoS biology*, 16(6), e2005372.
- Polechová, J., & Barton, N. H. (2015). Limits to adaptation along environmental gradients. *Proceedings of the National Academy of Sciences*, 112(20), 6401-6406.

- Ramsey, J., & Ramsey, T. S. (2014). Ecological studies of polyploidy in the 100 years following its discovery. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1648), 20130352.
- Ramsey, J., & Schemske, D. W. (1998). Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annual review of ecology and systematics*, 29(1), 467-501.
- Rausch, J. H., & Morgan, M. T. (2005). THE EFFECT OF SELF-FERTILIZATION, INBREEDING DEPRESSION, AND POPULATION SIZE ON AUTOPOLYPLOID ESTABLISHMENT. *Evolution*, 59(9), 1867-1875.
- Rezende, L., Suzigan, J., Amorim, F. W., & Moraes, A. P. (2020). Can plant hybridization and polyploidy lead to pollinator shift?. *Acta Botanica Brasilica*, 34, 229-242.
- Rice, A., Šmarda, P., Novosolov, M., Drori, M., Glick, L., Sabath, N., ... & Mayrose, I. (2019). The global biogeography of polyploid plants. *Nature Ecology & Evolution*, 3(2), 265-273.
- Rice, S. (2004). Evolutionary theory: mathematical and conceptual foundations.
- Rodríguez, D. J. (1996). A model for the establishment of polyploidy in plants: viable but infertile hybrids, iteroparity, and demographic stochasticity. *Journal of Theoretical Biology*, 180(3), 189-196.
- Sachdeva, H., Olusanya, O. O., & Barton, N. H. (2021). Genetic load and extinction in peripheral populations: the roles of migration, drift and demographic stochasticity. *bioRxiv*.
- Sheth, S. N., Morueta-Holme, N., & Angert, A. L. (2020). Determinants of geographic range size in plants. *New Phytologist*, 226(3), 650-665.
- Soltis, D. E., Buggs, R. J., Doyle, J. J., & Soltis, P. S. (2010). What we still don't know about polyploidy. *Taxon*, 59(5), 1387-1403.
- Soltis, D. E., Visger, C. J., Marchant, D. B., & Soltis, P. S. (2016). Polyploidy: pitfalls and paths to a paradigm. *American journal of botany*, 103(7), 1146-1166.
- Spoelhof, J. P., Soltis, D. E., & Soltis, P. S. (2020). Habitat Shape Affects Polyploid Establishment in a Spatial, Stochastic Model. *Frontiers in plant science*, 11, 1770.
- Svensson, E. I., & Connallon, T. (2019). How frequency-dependent selection affects population fitness, maladaptation and evolutionary rescue. *Evolutionary Applications*, 12(7), 1243-1258.
- Szép, E., Sachdeva, H., & Barton, N. H. (2021). Polygenic local adaptation in metapopulations: A stochastic eco-evolutionary model. *Evolution*, 75(5), 1030-1045.
- Thompson, K. A., Rieseberg, L. H., & Schlüter, D. (2018). Speciation and the city. *Trends in ecology & evolution*, 33(11), 815-826.
- Van de Peer, Y., Ashman, T. L., Soltis, P. S., & Soltis, D. E. (2021). Polyploidy: an evolutionary and ecological force in stressful times. *The Plant Cell*, 33(1), 11-26.
- Van de Peer, Y., Mizrahi, E., & Marchal, K. (2017). The evolutionary significance of polyploidy. *Nature Reviews Genetics*, 18(7), 411-424.
- Van Drunen, W. E., & Husband, B. C. (2019). Evolutionary associations between polyploidy, clonal reproduction, and perenniability in the angiosperms. *New Phytologist*, 224(3), 1266-1277.
- Yuan, X., Miller, D. J., Zhang, J., Herrington, D., & Wang, Y. (2012). An overview of population genetic data simulation. *Journal of Computational Biology*, 19(1), 42-54.

- Zenil-Ferguson, R., Burleigh, J. G., Freyman, W. A., Igić, B., Mayrose, I., & Goldberg, E. E. (2019). Interaction among ploidy, breeding system and lineage diversification. *New Phytologist*, 224(3), 1252-1265.