The evolution of polyploidy and the extreme cost of pollen-swamping.

Dissertation Thesis

Chintamani Rose McKeon (2417024)

Abstract

Amazing summary of project here. . .

Keywords

polyploidy; evolution; angiosperm; mimulus; guttatus; pollen-swamping; mate-choice;

1 Introduction

Polyploidy arises via mechanisms of whole-genome duplication (such as non-disjunction) and describes organisms with more than two sets of chromosomes. No polyploid mammal or bird species have been found, as the condition appears to be fatal in most animals (Svartman et al. 2005; Wertheim et al. 2013). However, polyploidy has been observed in amphibians, fish, fungi, reptiles, and plants. Amongst these, polyploidy is exceptionally well tolerated in plants. In fact, genome-doubling is a ubiquitous feature in the lineages of almost all flowering plants

(angiosperms) and has occurred even more recently (post genera formation) for 35% of all vascular plants (Jiao et al. 2011; Wood et al. 2009).

1.1 Autopolyploidy vs allopolyploidy

Genome-doubling in plants occurs either intraspecifically (autopolyploidy) due to unreduced gametes or via merging of genomes during hybridisation (allopolyploidy). The relative abundance and rates of the appearance of each type have long been debated (Darlington 1937; Stebbins 1947), but improved methods of determining ploidy levels (flow cytometry) have led to a much clearer understanding of this topic in recent years as both cytological and genetic data have accumulated (Doležel et al. 2007). In the past, autopolyploids were not recognised as distinct species but categorised instead as cytotypes of their progenitors which lead to a gross underestimation of their prevalence (Soltis et al. 2007). Although the estimated rate of autotetraploid formation–10^{-5}–is usually higher than that of allotetraploid formation (Ramsey and Schemske 1998), it is now thought that there are roughly equal autopolyploid and allopolyploid taxa (Barker et al. 2016). Tropical and woody taxa, however, were underrepresented in the cytological and genomic data that informed this estimate, so that could change if reviewed when these taxa are better represented.

1.2 The role of genome-doubling in plant diversification

Because of the ubiquity of polyploidisation events in plant lineages, genome-doubling is thought to be fundamental in the diversification of plants. Furthermore, these doubling events have been found to coincide specifically with speciation at much larger rates than previously estimated; 15% in angiosperms and 31% in ferns (Otto and Whitton 2000; Wood et al. 2009). Whole-genome duplication has even been

linked with major biological innovations that lead to diversification of the seed plant (spermatophyte) and angiosperm phyla (Jiao et al. 2011). Seeds and flowers both being crucial innovations that allowed plants to transition from an aquatic life-cycle to a terrestrial one by removing their dependence on water for reproduction. These adaptations are broadly considered as one of the most significant in the history of our planet as they enabled the 'greening' of earth. It follows, therefore, that there has been a substantial drive to understand the evolution of polyploidy and the mechanisms by which genome-doubling has been able to drive such adaptations. As a result of these efforts, polyploidisation has been shown to drastically alter phenotypes, breeding system, and physiology within just a few generations due to the massive genomic alterations that newly formed polyploids (neopolyploids) rapidly undergo (Adams and Wendel 2005; Levin 2002; Soltis et al. 2014b). Typically, polyploids tend to have larger flowers, seeds and stomata as well as being more robust all-round (Ramsey and Schemske 2002). Because of these attributes, genome-doubling has been crucial in the domestication of many crop plants, for instance; wheat, maize, sugarcane, coffee, cotton and tobacco (Dubcovsky and Dvorak 2007; Otto and Whitton 2000).

Talk about invasive potential?

Despite being broadly considered as one of the main driving forces in angiosperm evolution (Barker et al. 2016; Otto 2007; Soltis et al. 2014b), polyploids have been shown to contribute far less to diversification than their diploid counterparts (Mayrose et al. 2011). While genome-doubling significantly increases speciation rates of diploids, new polyploid lines do not appear to speciate further by that same mechanism and so speciate at a slower rate. Furthermore, polyploid extinction rates have been found to be far higher than those of diploids which further reduces their overall speciation rate relative to diploids (Arrigo and Barker 2012; Mayrose

et al. 2011). While the validity of these findings has been the topic of some debate (Mayrose et al. 2015; Soltis et al. 2014a), the results seem to support the long-standing evolutionary 'dead-end' hypothesis, which had previously lost popularity.

If this is the case after all, why is the success of polyploids so rare given the body of evidence that touts their rapid adaptability? Furthermore, under what conditions does polyploidy become advantageous and actually succeed? The discussion surrounding the advantages and disadvantages of polyploidy is well-reviewed in Comai2005, Ramsey and Ramsey (2014), and Otto (2007). However, even after a century of research, the factors that drive the success of polyploid establishment in the face of high extinction rates are still unclear. If polyploids tend to become extinct at the establishment phase due to reproductive disadvantages such as triploid sterility, or limited mate-choice (Arrigo and Barker 2012), then it follows that traits which increase reproductive assurance should lead to successfully evading extinction; perhaps why the polyploids we do see in nature exhibit traits like increased size, flower size, and number. As well as increasing potential fecundity, these traits should increase reproductive assurance by increasing the probability of geitonogamous self-pollination between flowers on the same individual (de Jong et al. 1993; Karron et al. 2004).

However, even in the best-studied systems there has–so far–been a lack of continuity between the genomic data and its ecological context, which has hindered our ability to fully understand the drivers in the evolution of this extreme mutation (Soltis et al. 2016). In order to elucidate these factors and test the theory presented, much work is being done to find and describe wild neopolyploid populations (Ferrero et al. 2015; Schlaepfer et al. 2008; Simón-Porcar et al. 2017). Discovering these recent genome-doubling events in the wild means that the processes of adaptation that leads to evolutionary success or failure of polyploidy can

be better examined, without the accumulation of genetic change clouding the picture. In addition, wild study populations will better inform us of the ecological factors which are integral to understanding evolutionary processes.

1.3 Mate choice limitation

Neopolyploids are thought to be mate-limited, either by diploid pollen-swamping (due to gamete incompatibility) or triploid sterility. It has been widely suggested that triploids tend not to be fertile (**?**); however this has recently been brought into dispute (**?**). However, if all the pollen that a neopolyploid receives is from an incompatible diploid donor, then it clearly figures that fitness will be severely reduced due to an incapacity to produce viable seeds. Furthermore, pollen-swamping has been shown to significantly disrupt species distribution in mixed diploid-hexaploid angiosperm populations of *Mercurialis annua* where it gained diploids such a competitive advantage that it leads to hexaploid displacement in multiple regions (Buggs and Pannell 2006).

> Needs more content. Totally a bit flat on the heart of the matter. Move some of the selfing stuff from the discussion here? Or introduce the discussion subheadings, so the traits which might give reproductive assurance are all covered early on?

1.4 Theoretical advances

> Talk about the modelling done so far. See Oswald and Nuismer (2007) and definitely Oswald and Nuismer (2011) which looks pretty similar. Need to delve into all of this more. Like a brief history of the modelling work that has lead up to this work would be nice. I think.

1.5 Aims

Using an individual-based modelling approach, we aim to lay the groundwork necessary for further investigations by examining the basic response of relative polyploid and diploid frequencies to changes in the rate of genome-duplication, under the assumption that mate-choice for neopolyploids is limited via pollen-swamping. We aim to quantify the theoretical cost of pollen-swamping by testing how high the rate of whole-genome duplication would need to be in order to overcome the strength of negative selection which it creates. We asked:

- Will the proportion of polyploids exceed the rate of whole-genome duplication? And, if so at what rate does this occur?
- Will increasing the duplication rate lead to the fixation of polyploidy?

In addition, we aim to discover if established and pioneering populations react in the same way to these increased duplication rates.

1.6 The study system

The model that follows was designed to simulate the life cycle of the yellow monkeyflower *Erythranthe gutatta* (previously: *Mimulus guttatus*) as laid out by Peterson et al. (2016) and detailed in figure 1. *E. guttata* was chosen due to ancient and recent genome-doubling that has resulted in both autopolyploids and allopolyploids being well documented in the species complex of this popular study organism (Beardsley et al. 2004; Buggs 2012; Simón-Porcar et al. 2017). The experimental design will begin with a cohort of diploid individuals that give rise to autotetraploid offspring; a scenario which loosely reflects the neo-autopolyploidisation recently discovered to be responsible for a mixed diploid-tetraploid population of *E. guttata* in the Shetland Isles (Simón-Porcar et al. 2017). All individuals will be hermaphrodites (in possession of perfect flowers

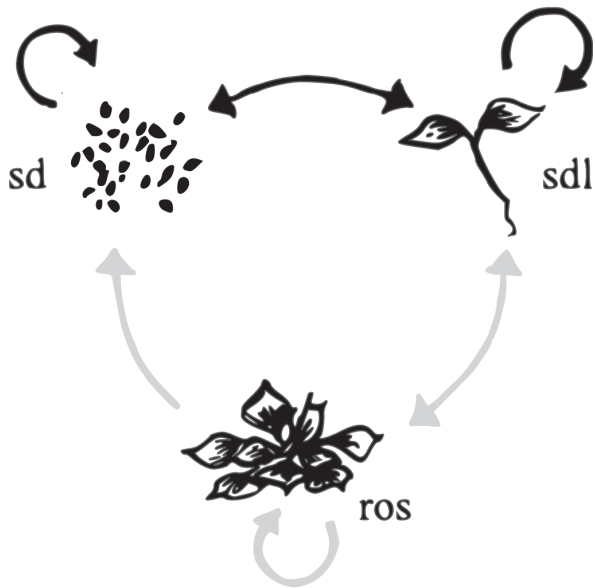

Figure 1. Life cycle of *Erythranthe guttata* (previously: *Mimulus guttatus*) taken from Peterson et al. (2016). Seedlings and rosettes are both sexually reproductive flowering stages. Annual populations follow the black arrows, whereas perennials also follow grey arrows to form rosettes (asexually produced clones).

that provide both male and female function), as is most common for *E. guttata* (Wise et al. 2011) and, indeed, all angiosperms. Seed dormancy will be disabled to remove any lag in the effects of changing the rate of genome duplication; this also fits with the lack of dormancy described for *E. guttata* (Willis 1993).

2 Methods

An Individual-based model was created, from scratch, in R (R Core Team 2019), as an open-source R package {*sploidy*}, and is available along with scripts and data on Github (McKeon 2020).

2.1 Assumptions

To test our hypotheses, the model needed to make particular assumptions about the reproductive interactions between individuals, as well as how particular parameters (like genome-duplication rate) were applied mechanistically. To see if the rate of duplication could overcome the negative effects of pollen-swamping, mate choice had to be limited in a way that represented failed fertilisation attempts. This was achieved using a simple matching rule based on parental ploidy level. In full, the model assumptions were as follows:

- All individuals began as diploid.
- Polyploids arose at a set rate of genome-duplication during all sexually reproductive transitions.
- Diploids and polyploids could both outcross, but never with each other. In addition, polyploids could not outcross with other polyploids of differing ploidy level, i.e.:- a tetraploid was not compatible with an octoploid.
- All individuals could self-fertilise.
- Offspring produced had the same ploidy level as their parents; two diploids produced a diploid, and so on.
- When genome-duplication occurred, the ploidy level of an individual was doubled; in this case, two diploid parents would produce a tetraploid.
- All plants were hermaphrodite; mothers and fathers were chosen via random sampling from the same pool of individuals.
- Pollen range spanned the entire landscape.
- Seed dispersal spanned the entire landscape.

Transition probabilities passed to the model defined how many individuals made it between life stages at any given generation. The *rbinom* function in base R was used throughout–where transitions

reduced cohort size–to choose individuals that transitioned successfully. This allowed for a natural amount of stochasticity around the transition probabilities. Where transition values increased cohort size, the function *sample_n* from the package {*Dplyr*} (Wickham et al. 2020) was used, instead, with replacement.

2.2 Simulation strategy

Rounds of simulations were run using the model function *sploidy* in order to home-in on the parameter ranges to include in the experiment design. These initial simulations were run for 1000 generations on a 100x100 landscape grid (to allow the default carrying capacity to maintain the adult population size at around one hundred thousand). The starting population contained sixty diploid individuals (twenty of each life stage), to simulate a pioneering species entering an empty space and the **ploidy_rate** parameter was varied between zero and 0.5. Seed dormancy was disabled by setting **seed_longevity** to zero as was clonal reproduction, via the transition matrix provided to the **trans** parameter. Transition parameters–the transition matrix (**trans**), germination rate (**G**), and seed survival rate (**D**)–were taken from known demographic data described in more detail below.

2.3 Transition data

Data for an *M. guttatus* population that had the most stable properties were used; minimal growth that could be kept in check with carrying capacity and low elasticity. Peterson et al. (2016) described the formula for computing transition data as laid out in figure 2 and provided the values for a low-elevation perennial *M. guttatus* population as shown in table 1. The seed survival value (D) they presented was taken from Elderd and Doak (2006). These data produced the transition matrix

$$
\begin{array}{c} \\ \text{Seed}_{t+1} \\ \text{Seedling}_{t+1} \\ \text{Rosette}_{t+1} \end{array}
\begin{array}{c}
\begin{array}{ccc} \text{Seed}_t & \text{Seedling}_t & \text{Rosette}_t \end{array} \\
\begin{pmatrix} \text{D}(1-\text{G}) & \text{FOA}(1-\text{G}) & \text{FOA}(1-\text{G}) \\ \text{DG} & \text{FOAG} & \text{FOAG} \\ 0 & \text{SR} & \text{SR} \end{pmatrix}
\end{array}
$$

Figure 2. How the transition matrix (table 2) was calculated from the data provided in table 1, according to this formula taken from Peterson et al. (2016).

Table 1. Components of Low-elevation perennial *Erythranthe guttata* (previously: *Mimulus guttatus*) life-cycle transition. Seed survival (D) is a general value taken from Elderd and Doak (2006). All other components taken from Peterson et al. (2016) who recorded values specific to that *E. guttata* population in the year 2013 (N = 77). G = germination rate, O = ovule number per flower, F = flower production, S = winter survival, R = rosette production, A = proportional recruitment success of ovules relative to rosettes.

D	G	O	F	S	R	A
0.534	0.652	494	5.34	0	0	0.00067

Table 2. Transition matrix used for the model, based on the formula shown in 2 and the data from figure 1.

	Seed (t)	Seedling (t)	Rosette (t)
Seed (t+1)	0.186	0.615	0.615
Seedling (t+1)	0.348	1.152	1.152
Rosette (t+1)	0.000	0.000	0.000

(figure 2) provided to the model along with the values of G (0.652) and D (0.534) shown in figure 1.

2.4 Varying the rate of genome-duplication

In order to quantify how the rate of genome-duplication affected fixation of polyploid genotypes, a randomly sampled range of polyploidisation rates were chosen from an exponential distribution of 100000 random uniform numbers ranging between 0.001 and 0.010. Using an exponential distribution gave higher resolution around the lower parameter rates which we expected to be most interesting.

Diagnostic plotting of these data revealed that increased rates needed to be examined. Further simulations were then run using randomly sampled rates from 100000 random uniform numbers between 0.010 and 0.500 in

an attempt to determine how high the mutation rate needed to be in order for polyploids to break free from the inherent selection against them due to mate-choice limitation.

3 Results

The pattern of population growth over time when no whole-genome duplication occurred; whatever the starting population size, the population quickly stabilised (in under fifty generations) around the carrying capacity (mean frequency = 1.00, sd = 0.00163) with seed counts at just under three-quarters of the adult population size (mean frequency = 0.729, sd = 0.00210).

Irrelevant?

*3.1 Did the proportion of polyploids exceed the rate of whole-genome duplication?***

At duplication rates between 0.001 and 0.01, we saw–and demonstrate in figure 3–that polyploid frequencies spread evenly around around the rate of duplication (mean relative frequency at generation 1000 = 1.00, sd = 0.0617, CI = 0.881 to 1.12, N = 375). Out of these 375 simulations, almost half (46.9%) of the polyploid population frequencies fell just below the duplication rate and the other half (53.1%) made it slightly over. However, once the rate of duplication increased past 0.01, the polyploids that appeared began to consistently escape the negative selection caused by pollen-swamping. Between duplication rates of 0.01 and 0.171, only one polyploid population in 194 (0.5%) ended the simulation with a frequency below it's duplication rate; the overall trend was one of increased population growth for polyploids (mean relative frequency = 1.23, sd = 0.122, CI = 0.988 to 1.46, N = 194). The maximum relative frequency of polyploids observed was 1.39, seen in 2 simulations with duplication rates of 0.167 and 0.170.

Should I fit a model for the relative frequencies?

Out of 411 simulations that had a duplication rate over 0.171, only 15 (3.65%) had survivors at 1000 generations. Figure 4 shows the difference in population growth patterns. Below the threshold, all 569 simulations survived to 1000 generations, and the majority (98.8%) reached stability; seven simulations with a duplication rate that rounded to 0.171 were less predictable and showed patterns of booms and busts. Figure 5 shows the frequency of diploids and polyploids reached for all populations that survived to 1000 generations with whole-genome duplication enabled at random rates between 0.001 and 0.25. As the rate of whole-genome duplication was increased, maximum polyploid frequencies rose while diploid frequencies fell until a threshold rate was reached; once the duplication rate exceeded 0.171 neither diploids or polyploids survived well.

Should I fit a model for extinction?

3.2 Did increasing the duplication rate lead to the fixation of polyploidy?

As seen in figures 5 and 4, out of 994 simulations with whole-genome duplication enabled, polyploidy never made it to fixation. In fact, polyploid populations never represented more than a quarter of the landscape; the maximum polyploid frequency of surviving simulations at generation 1000 was 0.236. Increasing the rate of whole-genome duplication, alone, was not sufficient for polyploids to dominate diploids on the landscape, let alone outcompete them and take over.

3.3 Did established, and pioneering populations react in the same way to increased duplication rates?

No matter whether simulations began with small or large numbers of diploids, the overall patterns observed were the same; polyploid growth

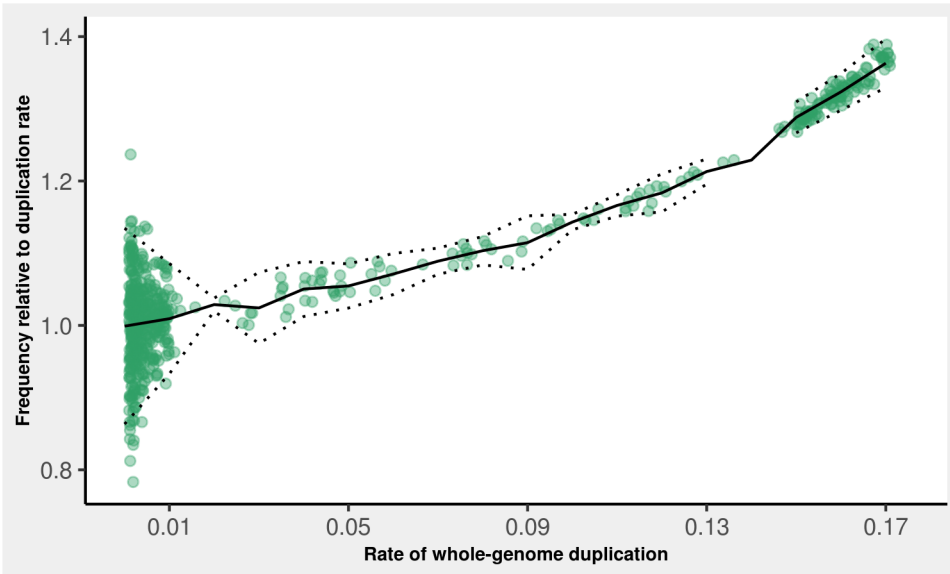

Figure 3. Escaping the cost of pollen-swamping. Points represent simulations where the rate of duplication was between 0.001 and 0.171 (N = 569). Each point (green) shows the the frequency of polyploids at the end of the simulation (generation 1000) relative to the rate of whole-genome-duplication set (frequency/rate). The solid black line represents the mean relative frequency for duplicatin rates rounded to 2 decimal places, and the dotted black lines represent the 95% confidence envelope around the mean.

and stability below a rate of 0.171 or decline and ultimate extinction at rates over that threshold. The main difference starting population size made, was to the length of time it took for the diploids and polyploids to reach their stable frequency range, or to reach extinction (see figure 4). For extremely low starting population sizes, some extinction was observed at duplication rates below 0.171. Extremely high starting population sizes did not appear to buffer against this effect, though in some cases they did delay it. Figure 6 shows the sudden mass of extinction events that occurred as the rate of duplication was increased; outliers had particularly low or high starting population sizes.

I think I need to pull some actual numbers out of the data here.

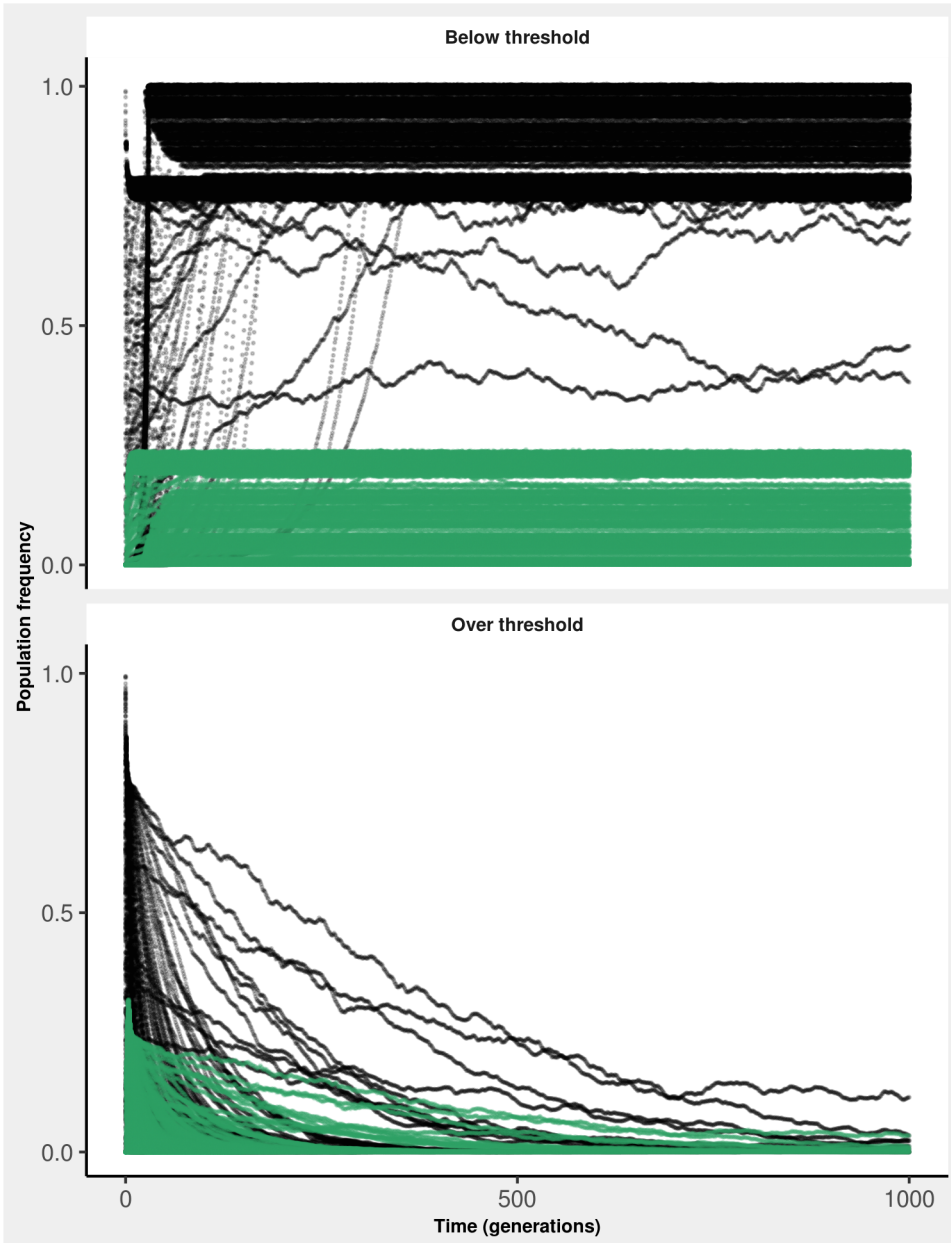

Figure 4. Threshold point of reproductive interference occurs over a whole-genome duplication rate of 0.171. Panels show whole-genome duplication rates of simulations where duplication rates were between 0.001 and 0.171 (below threshold N = 583) and where rates were between 0.172 and 0.25 (over threshold N = 411). Points represent the adult population proportion over time, coloured by ploidy type; diploid (black) and polyploid (green).

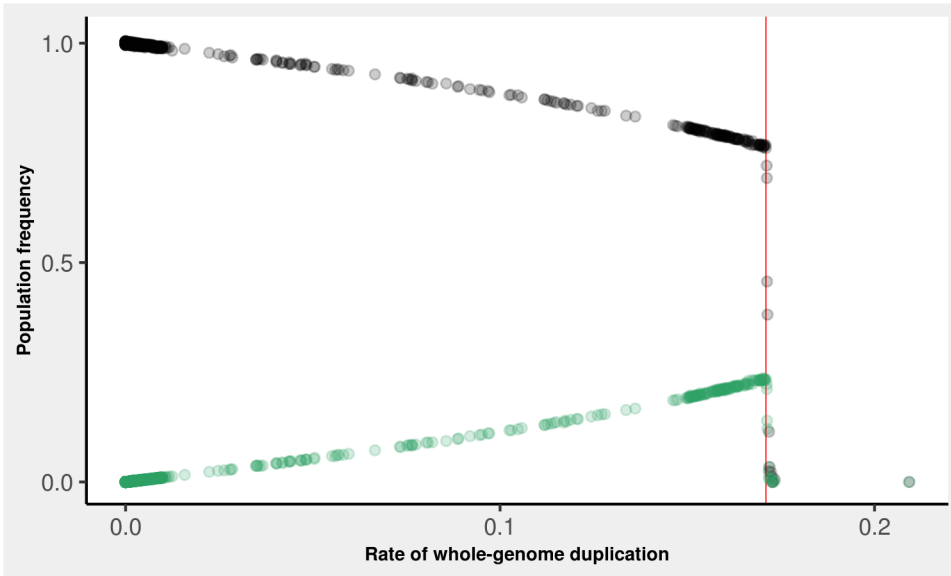

Figure 5. The relationship between the rate of whole-genome duplication and freqency of diploid (black) vs polyploid (green) individuals. Each point represents the final population proportion reached for every simulation that reached 1000 generations with survivors (N = 788 simulations). Vertical red line marks a duplication rate of 0.171.

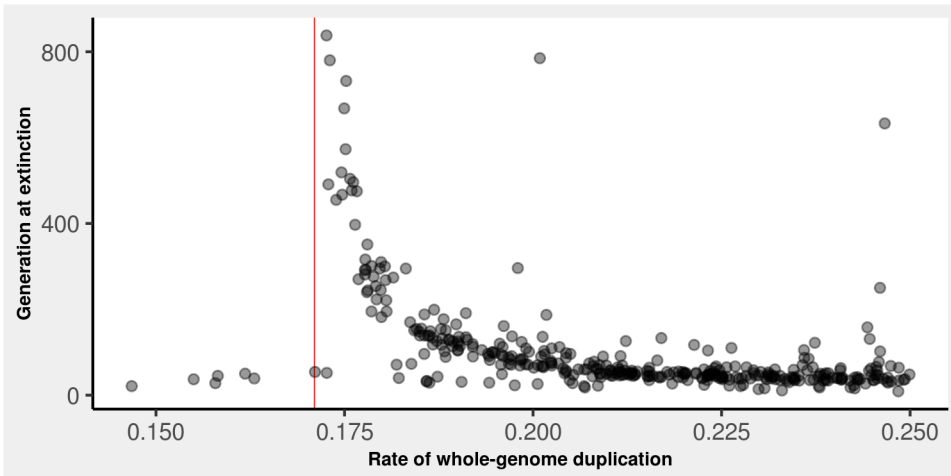

Figure 6. Extinction events as a functon of whole-genome duplication rate. Each point represents a simulation that went extinct before one thousand generations had passed (N = 369 simulations). Vertical red line marks a duplication rate of 0.171.

4 Discussion

We have shown here how the extreme cost of pollen-swamping selects against the evolution of polyploidy. Polyploidy did not rise to fixation in a single simulation because of the dramatic effect these mating rules had on their ability to produce viable seeds. We have also shown how increasing the rate of whole-genome duplication caused reproductive interference across the entire landscape, swamping diploids and polyploids alike with incompatible pollen. The resulting reduction in fecundity doomed both diploids and polyploids to extinction. In effect, because of the model assumptions, increasing the rate of whole-genome duplication increased the diversity of mating types in the system. When no pollen specificity mechanisms existed to sort the mating pool so that these types could find each other effectively, the whole population was doomed to failure once enough different types were in the system. Therefore, it follows that this mate-choice limiting mutation must have a part to play in the evolution of beneficial traits that assist in compatibility assortment, such as; flower timing, pollinator specificity, and selfing. Could these mechanisms of mating assortment allow the population to escape extinction?

> Are there reasons to think that pollen swamping will lead to dead ends under some conditions, but diversification (or at least not dead ends) under others?

Here we explore how mechanisms for compatibility assorting traits would work with the current model. We consider how these biological observations could be boiled down into a computational form, and we follow that, in each case, by discussing how these new conditions may be expected to change the pattern of results.

4.1 Delayed flowering

Delayed flowering is often observed in autopolyploids due to their increased cell size and longer development times (**?**). But rather than a cost that causes mate-limitation–as discussed in **?**–this could be thought of more as a strategy to escape the reproductive interference caused by diploid pollen swamping. If the model presented here were expanded so that the parental pool–which is sampled randomly to create seeds–was further subdivided depending on ploidy level, then fertilisation attempts would only be carried out between compatible individuals. This subdivision would rudimentarily simulate the segregation we see in mating pools caused by phenological differences like delayed flowering. Furthermore, the interference we see caused by the rate of whole-genome duplication should vanish under these conditions.

However, not all polyploids exhibit delayed flowering; for instance, many allopolyploid hybrids have intermediate phenotypes rather than enlarged ones (**?**). To create some more natural variation, we could give individuals an inheritable trait value that denotes flowering time instead; a random uniform number between one and six perhaps for the months of spring and summer, with the lower values set to be more common by some probability. We could assume an equal flowering time of a single month, and subset the population to create parental mating pools by rounding these numbers to no decimal places. The trait values would be subject to mutation by a set rate during sexual reproduction, and polyploids could be gifted a higher chance of mutation to a value closer to six. This probability could depend on the number of sets of chromosomes that specific individual has, so for instance: an octoploid would have a higher chance of flowering in August than a tetraploid, but many polyploids would continue to flower at the same time as their parents. Under these conditions, we might see delayed flowering evolve in polyploid lineages, naturally assorting the mating pools and

overcoming the pollen-swamping interference so that we see diversification rather than dead ends.

Relate this to study system

4.2 Pollinator specificity

Many flowering plants adjust their flower morphology, colour and nectar in order to attract very specific pollinators (**?**). This acts to target pollen to specific compatible flower types in a complex and fascinating manner that relies on symbiosis with entirely separate species. This too could be a mechanism by which polyploids can escape pollen-swamping and increase their chances to reproduce in the face of extreme mate-limitation. Rather than add in explicit pollinators, this could be modelled in a similar way to the delayed flowering time. A random trait value for pollinator preference would be used to subset the population into mating pools. Therefore, the addition of a single sorting trait value would, in fact, test both the delayed flowering and pollinator specificity strategies, because they boil down to the same core process.

Relate this to study system: see Wise et al. (2011) for pollinator behaviour info

4.3 Selfing

Self-incompatibility is commonly found in angiosperms (**?**) and limits sexual reproduction via selfing. Various mechanisms control this by limiting the amount of self-pollination that can occur either by morphology, genetics or flower timing (**?**). It has been hypothesised that reversal of self-incompatibility may be a route by which polyploidy evolves as it would remove constraints imposed by mate-limitation (**?**). In addition, multiple gene copies are thought to buffer against the deleterious effects of inbreeding depression (**?**). While a general pattern

of association between polyploidy and selfing has not been found across a wide range of taxa (Mable 2004), there has been some evidence for increased selfing in angiosperm tetraploids (Cook and Soltis 2000; Husband and Schemske 1997).

To investigate this mechanism further, we would need to restrict self-fertilisation in the model (under current assumptions, all individuals can self-fertilise). This could be done very simplistically by removing attempts where parental identifiers match. However, this approach would reduce the seed output as those attempts are removed and add a negative selection pressure to the system [think more about this, is that what we want?].

Perhaps a better method would be to elaborate the flower timing trait described above in order to allow the evolution of monoecy; separate male and female flowers which are freed to bloom at different times. However, this mechanism may not be such a good fit for the mating system actually observed in *E. guttata*. It would be relatively simple to add a second flower timing trait value so that the first defines the maternal mating pool, and the second the paternal. The model would need to begin with selfing disabled, i.e.:- the maternal and paternal flower timing trait values of any individual would need to differ by at least one whole number. This could be enforced by choosing maternal numbers at random and then resampling a paternal value until one that does not match is found. It would be important to make sure there was some overlap by using random numbers so that some mating was definitely able to occur; if all maternal flowers bloomed in month 1, and all paternal in month 6, then the population would not survive longer than one generation. Allowing these random values to mutate as before would enable us to see if selfing naturally evolves, and if so, whether it occurs more in polyploid or diploid lineages.

I think the above might need scrapping.

While hermaphroditic flowers are the most commonly found in *E. guttata* (Wise et al. 2011), male sterility genes have also been described (Willis 1999). The variation found was considered ample enough to lead to the evolution of gynodioecy (females plus hermaphrodites), an intermediate step in the evolution of dioecy (separate male and female individuals) (Charlesworth 2006). Investigation of potential constraints in the evolution of gynodioecy (Wise et al. 2011) demonstrated that females had a reproductive advantage due to increased flower number, as well as ovule number per flower. In addition, the potential for this advantage to be even greater in the face of hermaphroditic inbreeding depression was raised; the theory being that females will necessarily outcross. However, the strength of these advantages appears only to keep the genes for male sterility at low frequencies wild *E. guttata* populations, as no fully gynodioecious population has been described.

> I've gone off on a tangent... what is the selfing mechanism in *E. guttata*? Check over the more recent papers Matty gave me about estimating outcrossing and selfing rates... pretty sure there isn't a decent method for this yet as that was part of the project, because the stats so far are better suited to obligate outcrossers. Do we even know how much selfing wild *E. guttata* already do?

They were mostly outcrossing with fluctuating partial selfing in two populations from Oregon (Willis 1993).

> Are these methods outdated or still valid?

> Also... what about geitonogamy? See de Jong et al. (1993) and Karron et al. (2004)... adaptations that prevent selfing within a flower or by flower timing don't stop the transfer of pollen between individual flowers on the same plant, or a

genetically identical one. It's common and increases with plants size. Selfing is getting complicated. Individual flowers aren't represented in the model. But if polyploids are bigger with more flowers, then it follows that more selfing would occur due to this mechanism as shown in the floral size comparisons of Karron et al. (2004). Maaaaaybe this is why Mable (2004) did not find an association between polyploids and selfing across wider taxa when there is evidence it occurs in angiosperms–check which specific taxa that study attempted to generalise those findings to and make sure they were not mostly angiosperm.

Where diploid *M. annua* were seen to displace hexaploid relatives in multiple locations via pollen-swamping, selfing only minorly allowed the hexaploids to mitigate these effects. Escaping pollen-swamping via selfing was density-dependent and slight (Buggs and Pannell 2006).

4.4 Model limitations

The mechanism of genome-duplication modelled did not differentiate between diploid and polyploid progenitors; for every viable seed created there was a chance that its genome would double regardless of its ploidy level (which was the same as its parents). This, therefore, could be improved to better reflect the fact that polyploids have not been found to speciate by further genome-doubling at the same rate as diploids (Mayrose et al. 2011; **?**). If the duplication rate were lowered in this manner for polyploids, it would alleviate some mate-limitation by reducing the diversity of ploidy level that arises in the system (i.e.:- octoploids, hexadecaploids, and so on). Under the current assumptions, individuals cannot mate successfully with those of differing ploidy level. Therefore, due to tetraploids producing more tetraploids, there would be

an increased probability that their gametes might meet and become viable seeds; especially in simulations with particularly high duplication rates.

The modelling method used to increase cohort size (when transition probabilities were greater than 1) lacked some natural variation due to the sampling process used. The transition matrix used to parameterise the simulations presented here increased the cohort size of seedlings at every generation based on the previous generations seedling number. Some stochasticity around seedling number in every generation was maintained, however, by the transitions that controlled fecundity and germination; these decreased cohort size and so used a better sampling method to maintain variation in the actual number sampled based on the given probability. The impact this programming decision made on the overall predictions of the model is likely negligible.

> Should I even bother mentioning this? I am not sure I have managed to describe it a way that makes sense anyway.

Genetically, all neopolyploids that avoid extinction enter a period of diploidisation, by which duplicated genes are lost discriminantly via various mechanisms over many generations until a state of diploidy has returned (Comai 2005). The model presented here included no rate or mechanism for this process.

> Not really sure that matters? Probably not relevant. It is cool, in any case.

5 Conclusion

The cost of pollen-swamping is extreme…

> Say more.

References

Adams K.L., Wendel J.F. (2005) Polyploidy and genome evolution in plants. *Current Opinion in Plant Biology*, 8 (2), pp. 135–141.

Arrigo N., Barker M.S. (2012) Rarely successful polyploids and their legacy in plant genomes. *Current Opinion in Plant Biology*, 15 (2), pp. 140–146.

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, pp. 391–398.

Beardsley P.M., Schoenig S.E., Whittall J.B., Olmstead R.G. (2004) Patterns of evolution in western North American Mimulus (Phrymaceae). *American Journal of Botany*, 91 (3), pp. 474–489.

Buggs R.J.A. (2012) Monkeying around with ploidy. *Molecular Ecology*, 21 (21), pp. 5159–5161.

Buggs R.J., Pannell J.R. (2006) Rapid Displacement of a Monoecious Plant Lineage Is Due to Pollen Swamping by a Dioecious Relative. *Current Biology*, 16 (10), pp. 996–1000.

Charlesworth D. (2006) Evolution of Plant Breeding Systems. *Current Biology*, 16, pp. R726–R735.

Comai L. (2005) The advantages and disadvantages of being polyploid. *Nature Reviews Genetics*, 6 (11), pp. 836–846.

Cook L.M., Soltis P.S. (2000) Mating systems of diploid and allotetraploid populations of Tragopogon (Asteraceae). II. Artificial populations. *Heredity*, 84 (4), pp. 410–415.

Darlington C.D. (1937) *Recent advances in cytology*. 2d ed. ed. Philadelphia: P. Blakiston's Son & Co. Inc.

Doležel J., Greilhuber J., Suda J. (2007) Estimation of nuclear DNA content in plants using flow cytometry. *Nature Protocols*, 2 (9), pp. 2233–2244.

Dubcovsky J., Dvorak J. (2007) *Genome plasticity a key factor in the success of polyploid wheat under domestication*. American Association for the Advancement of Science.

Elderd B.D., Doak D.F. (2006) Comparing the direct and community-mediated effects of disturbance on plant population dynamics: flooding, herbivory and Mimulus guttatus. *Journal of Ecology*, 94 (3), pp. 656–669.

Ferrero V., Barrett S.C., Castro S., Caldeirinha P., Navarro L., Loureiro J., Rodríguez-Echeverría S. (2015) Invasion genetics of the Bermuda buttercup (Oxalis pes-caprae): Complex intercontinental patterns of genetic diversity, polyploidy and heterostyly characterize both native and introduced populations. *Molecular Ecology*, 24 (9), pp. 2143–2155.

Husband B.C., Schemske D.W. (1997) The Effect of Inbreeding in Diploid and Tetraploid Populations of Epilobium angustifolium (Onagraceae): Implications for the Genetic Basis of Inbreeding Depression. *Evolution*, 51 (3), p. 737.

Jiao Y., Wickett N.J., Ayyampalayam S., Chanderbali A.S., Landherr L., Ralph P.E., Tomsho L.P., Hu Y., Liang H., Soltis P.S., Soltis D.E., Clifton S.W., Schlarbaum S.E., Schuster S.C., Ma H., Leebens-Mack J., Depamphilis C.W. (2011) Ancestral polyploidy in seed plants and angiosperms. *Nature*, 473 (7345), pp. 97–100.

de Jong T.J., Waser N.M., Klinkhamer P.G. (1993) *Geitonogamy: The neglected side of selfing*. Elsevier Current Trends.

Karron J.D., Mitchell R.J., Holmquist K.G., Bell J.M., Funk B. (2004) The influence of floral display size on selfing rates in Mimulus ringens. *Heredity*, 92 (3), pp. 242–248.

Levin D. (2002) *The role of chromosomal change in plant evolution*. Oxford: Oxford University Press.

Mable B.K. (2004) Polyploidy and self-compatibility: is there an association? *New Phytologist*, 162 (3), pp. 803–811.

Mayrose I., Zhan S.H., Rothfels C.J., Arrigo N., Barker M.S., Rieseberg L.H., Otto S.P. (2015) Methods for studying polyploid diversification and the dead end hypothesis: a reply to Soltis et al . (2014). *New Phytologist*, 206 (1), pp. 27–35.

Mayrose I., Zhan S.H., Rothfels C.J., Magnuson-Ford K., Barker M.S., Rieseberg L.H., Otto S.P. (2011) Recently Formed Polyploid Plants Diversify at Lower Rates. *Science*, 333 (6047), p. 1257.

McKeon R. (2020) *sploidy: An individual-based model of polyploid evolution that demonstrates the extreme cost of pollen-swamping.* R package version 0.1.9.

552 Oswald B.P., Nuismer S.L. (2007) Neopolyploidy and pathogen resistance. *Proceedings*
553 *of the Royal Society B: Biological Sciences*, 274 (1624), pp. 2393–2397.

554 Oswald B.P., Nuismer S.L. (2011) Unified model of autopolyploid establishment and
555 evolution. *American Naturalist*, 178 (6), pp. 687–700.

556 Otto S.P. (2007) The evolutionary consequences of polyploidy. *Cell*, 131 (3), pp.
557 452–62.

558 Otto S.P., Whitton J. (2000) Polyploid Incidence and Evolution. *Annual Review of*
559 *Genetics*, 34 (1), pp. 401–437.

560 Peterson M.L., Kay K.M., Angert A.L. (2016) The scale of local adaptation in Mimulus
561 guttatus: comparing life history races, ecotypes, and populations. *New Phytologist*,
562 211 (1), pp. 345–356.

563 R Core Team (2019) *R: A Language and Environment for Statistical Computing*. R
564 Foundation for Statistical Computing Vienna, Austria.

565 Ramsey J., Ramsey T.S. (2014) Ecological studies of polyploidy in the 100 years
566 following its discovery. *Phil Trans R Soc B*, 369.

567 Ramsey J., Schemske D.W. (1998) Pathways, mechanisms, and rates of polyploid
568 formation in flowering plants. *Annual Review of Ecology and Systematics*, 29 (1), pp.
569 467–501.

570 Ramsey J., Schemske D.W. (2002) Neopolyploidy in Flowering Plants. *Annual Review*
571 *of Ecology and Systematics*, 33 (1), pp. 589–639.

572 Schlaepfer D.R., Edwards P.J., Semple J.C., Billeter R. (2008) Cytogeography of
573 Solidago gigantea (Asteraceae) and its invasive ploidy level. *Journal of*
574 *Biogeography*, 35 (11), pp. 2119–2127.

575 Simón-Porcar V.I., Silva J.L., Meeus S., Higgins J.D., Vallejo-Marín M. (2017) Recent
576 autopolyploidization in a naturalized population of Mimulus guttatus (Phrymaceae).
577 *Botanical Journal of the Linnean Society*, 185 (2), pp. 189–207.

578 Soltis D.E., Segovia-Salcedo M.C., Jordon-Thaden I., Majure L., Miles N.M.,
579 Mavrodiev E.V., Mei W., Cortez M.B., Soltis P.S., Gitzendanner M.A. (2014a) Are
580 polyploids really evolutionary dead-ends (again)? A critical reappraisal of Mayrose et
581 al . (2011). *New Phytologist*, 202 (4), pp. 1105–1117.

582 Soltis D.E., Soltis P.S., Schemske D.W., Hancock J.F., Thompson J.N., Husband B.C.,
583 Judd W.S. (2007) Autoploidy in angiosperms: have we grossly underestimated the
584 number of species? *Taxon*, 56 (1), pp. 13–30.

585 Soltis D.E., Visger C.J., Marchant D.B., Soltis P.S. (2016) Polyploidy: Pitfalls and paths
586 to a paradigm. *American Journal of Botany*, 103 (7), pp. 1146–1166.

587 Soltis D.E., Visger C.J., Soltis P.S. (2014b) The polyploidy revolution then...and now:
588 Stebbins revisited. *American Journal of Botany*, 101 (7), pp. 1057–1078.

589 Stebbins G.L. (1947) Types of Polyploids: Their Classification and Significance. In:
590 M. Demerec, ed., *Advances in Genetics*. Academic Press. pp. 403–429.

591 Svartman M., Stone G., Stanyon R. (2005) Molecular cytogenetics discards polyploidy
592 in mammals. *Genomics*, 85, pp. 425–430.

593 Wertheim B., Beukeboom L.W., Van De Zande L. (2013) Polyploidy in animals: Effects
594 of gene expression on sex determination, evolution and ecology. *Cytogenetic and*
595 *Genome Research*, 140 (2-4), pp. 256–269.

596 Wickham H., François R., Henry L., Müller K. (2020) *dplyr: A Grammar of Data*
597 *Manipulation*. R package version 0.8.5.

598 Willis J.H. (1993) Partial self-fertilization and inbreeding depression in two populations
599 of mimulus guttatus. *Heredity*, 71 (2), pp. 145–154.

600 Willis J.H. (1999) The contribution of male-sterility mutations to inbreeding depression
601 in Mimulus guttatus. *Heredity*, 83 (3), pp. 337–346.

602 Wise M.J., Vu J.V., Carr D.E. (2011) Potential ecological constraints on the evolution of
603 gynodioecy in Mimulus guttatus: Relative fecundity and pollinator behavior in a
604 mixed-sex population. *International Journal of Plant Sciences*, 172 (2), pp. 199–210.

605 Wood T.E., Takebayashi N., Barker M.S., Mayrose I., Greenspoon P.B., Rieseberg L.H.
606 (2009) The frequency of polyploid speciation in vascular plants. *Proceedings of the*
607 *National Academy of Sciences of the United States of America*, 106 (33), pp.
608 13875–9.