
1

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

Dissertation Thesis

2 Chintamani Rose McKeon (2417024)

Abstract

3 Amazing summary of project here...

Keywords

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

5

1 Introduction

6 Would some images of genome duplication be helpful?

7 Whole-genome duplication (polyploidisation) is an extreme mutation
8 which occurs due to errors during cell division; such as non-disjunction
9 where sister chromatids are not separated during anaphase. Failed
10 disjunction leads to all the genetic material ending up in one of the two
11 daughter cells; essentially doubling the genome within that cell. When
12 this occurs during gametogenesis, it can lead to whole organisms with

more than two complete sets of chromosomes (polyploids). According to ?, the estimated rate of polyploidisation is one order of magnitude higher than that of standard mutation (10^{-5}). In humans, it coincides with around 5% of miscarriages (Creasy et al. 1976). Moreover, while there have been some live birth cases of triploid human infants, they usually die within hours. In irregular cases, they survive for several months, but none have survived more than six years (Hashimoto et al. 2018).

Newer data for miscarriage rate?

No polyploid mammal species has ever been found (Svartman et al. 2005), and the same is also usually true for birds (?). However, while in these taxa, genome-doubling is not well tolerated, for amphibians, fish, fungi, reptiles, and plants, the story is quite different. Amongst these, polyploidy is exceptionally well tolerated in plants. Jiao et al. (2011) found that genome-doubling is a ubiquitous feature in the lineages of almost all flowering plants (angiosperms), and Wood et al. (2009) also found that it has occurred even more recently (post genera formation) for 35% of all vascular plants. Plant surveys of natural populations over the last century further support this, reporting mean polyploid frequencies of 25% (?).

The Role of Genome-Doubling in Plant Diversification

Genome-doubling is thought to be fundamental in the diversification of plant species; having been found to coincide with 15% of angiosperm speciation events, and 31% in ferns (Wood et al. 2009). Furthermore, Jiao et al. (2011) showed that whole-genome duplication could even be linked with major biological innovations that have lead to diversification of the plant phyla. Their phylogenetic analysis brought to light two ancient groups of duplication events: the first grouping of duplications (around 319 millennia ago) influenced the diversification of regulatory genes that

were integral to seed development and the more recent grouping (around 192 millennia ago) affected genes that enabled flower development. The authors appropriately conclude that these events must have played a significant role in the rise of seed plants, as well as in the later diversification of angiosperms. Seeds and flowers were both crucial innovations that allowed plants to transition from an aquatic life-cycle to a terrestrial one by removing their dependence on water for reproduction — an event, broadly considered as one of the most significant in the history of our planet.

The full story, however, is not quite that clear-cut. According to [Mayrose et al. \(2011\)](#), genome-doubling significantly increases speciation rates of diploids, but new polyploid lines do not further speciate by that same mechanism. So, polyploid speciation rates are far smaller in comparison. Furthermore, they found polyploid extinction rates to be far higher than those of diploids because polyploids are clumped around the tips of phylogenies. This Likelihood-based analysis of vascular plants provided clear, and quantitative, support for the traditionally popular view that polyploidy most often leads to evolutionary dead ends.

First proposed by? Stebbins 71?

So, what is it then that allows this extreme mutation to succeed and become fixated? What scenarios and mechanisms combine to allow rare polyploid lines to succeed? Are the majority of the polyploids we see in natural populations doomed to extinction? Furthermore, can we predict which conditions will lead to their survival?

Despite having been studied for over a century, the factors that drive the success of polyploid establishment in the face of high extinction rates are still unclear. [Arrigo and Barker \(2012\)](#) conclude that polyploids tend to become extinct at the establishment phase due to reproductive

70 disadvantages such as triploid sterility, or limited mate-choice; the latter
71 occurring via diploid pollen-swamping, or delayed flowering. On the
72 other hand, polyploids are frequently linked with distinct traits such as
73 “gigas effects”, which describes their increased size and plant organs.
74 Additionally, benefits of polyploidy can include enhanced capabilities for
75 buffering of deleterious mutation (due to increased heterozygosity),
76 hybrid vigour (heterosis), and a reversal of selfing inhibition ([Woodhouse et al. 2009](#); [Ramsey and Ramsey 2014](#)). These traits are thought to
77 overcome the reproductive disadvantages of polyploidy and, in those rare
78 cases, make this mutation key to the invasive and adaptive potential of
79 plants.
80

81 *The cost of being polyploid (limited mate-choice)*

82 *Diploid Pollen-Swamping*

83 *Delayed flowering*

84 Delayed flowering (cost) is associated with gigas-effects of
85 increased size (benefit). TRADE-OFF.

86 *Mimulus study system*

87 Polyploidisation events have been well documented in *Mimulus guttatus*
88 (?).

89 ? details 3-stage *M. guttatus* life-cycle and transition matrices.

90 *Aims*

91 Here we aimed to quantify the cost of pollen-swamping by testing how
92 high the rate of whole-genome duplication would need to be in order to
93 overcome the strength of negative selection. We asked:

- 94 • Will increasing the rate of whole-genome duplication lead to the
95 fixation of polyploidy?

- 96 ● And, if so, at what threshold point does this occur?

97 In addition, we aimed to discover how varying the starting population
98 size affected these results:

- 99 ● Does an established population react in the same way to increased
100 duplication rates as a pioneering population entering an empty
101 landscape?

102 **2 Methods**

103 An individual-based modelling approach was used to examine these
104 aims. The model was created, from scratch, in R ([R Core Team 2019](#)) to
105 simulate the life cycle of the yellow monkey flower (*Mimulus guttatus*) as
106 laid out by ? and detailed in figure 1. The *M. guttatus* complex was a
107 suitable choice due to whole-genome duplication being well documented
108 throughout (?), and data being readily available.

knitr:::include_graphics ("_images/m-guttatus-lifecycle-pe

109 *The model*

110 The model was created as an open source R package *{sploidy}*, and is
111 available along with the simulation scripts used for this work and the raw
112 data, on Github ([McKeon 2020](#)).

113 **Assumptions** To test our hypotheses the model needed to make
114 particular assumptions about the reproductive interactions between
115 individuals, as well as how particular parameters (like whole genome
116 duplication rate) were applied mechanistically. To see if the rate of
117 duplication could overcome the effects of limited mate-choice, mate
118 choice had to first be limited in some manner. This was achieved using a
119 simple matching rule that defined whether attempts to fertilise were

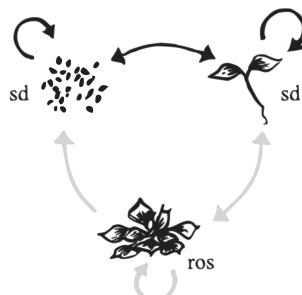


Fig. 2 Life cycle graph for *Mimulus guttatus* based on a yearly census at the start of the growing season. Annual populations (black arrows only) can exist as seeds (sd) in the seed bank or newly germinated seedlings (sdl), whereas perennial populations (black and gray arrows) can also exist as vegetative rosettes (ros) that have successfully overwintered. Either seedlings or rosettes can flower to produce seeds or seedlings, or survive and grow to produce rosettes, the following year. Vital rate parameters are given in matrix 1.

Figure 1. Life cycle of *Mimulus guttatus* taken from ?. Seedlings and rosettes are sexually reproductive flowering stages. Rosettes are asexually produced clones.

120 successful, based on parental ploidy level. Simulating, therefore, the cost
 121 of pollen-swamping.

122 The model assumptions were as follows:

- 123 • All individuals began as diploid.
- 124 • Polyploids arose at a set rate of whole-genome duplication during
 125 all sexually reproductive transitions.
- 126 • Diploids and polyploids could both out-cross, but never with each
 127 other. In addition, polyploids could not outcross with other
 128 polyploids of differing ploidy level, ie:- a tetraploid was not
 129 compatible with an octoploid.
- 130 • All individuals could self-fertilise.
- 131 • Offspring produced had the same ploidy level as their parents; two
 132 diploids produced a diploid, and so on.
- 133 • When whole-genome duplication occurred, the ploidy level of an
 134 individual was doubled; in this case two diploid parents would
 135 produce a tetraploid.

- 136 ● All plants were hermaphrodite; mothers and fathers were chosen via
137 random sampling from the same pool of individuals.
138 ● Pollen range spanned the entire landscape.
139 ● Seed dispersal spanned the entire landscape.

140 The *rbinom* function in base R was used throughout the model—where
141 transitions reduced cohort size—to choose individuals that transitioned
142 successfully. This allowed for a natural amount of stochasticity around
143 the transition probabilities, similar to that which we would expect in
144 nature. Where transition values increased cohort size, the function
145 *sample_n* from the package *{Dplyr}* (Wickham et al. 2020) was used,
146 instead, with replacement.

147 **Simulation strategy**

148 Rounds of simulations were run using the model function *sploidy* in
149 order to home-in on the parameter ranges to include in the experiment
150 design. These initial simulations were run for 1000 generations on a
151 100x100 landscape grid (to allow the default carrying capacity to
152 maintain the adult population size at around one hundred thousand). The
153 starting population contained sixty diploid individuals (twenty of each
154 life stage), to simulate a pioneering species entering an empty space and
155 the **ploidy_rate** parameter was varied between zero and 0.5. Seed
156 dormancy was disabled by setting **seed_longevity** to zero. As was clonal
157 reproduction, via the transition matrix provided to the **trans** parameter.
158 Transition parameters—the transition matrix (**trans**), germination rate (**G**),
159 and seed survival rate (**D**)—were taken from known demographic data (??).

160 **Transition data** Data for an *M. guttatus* population that had the most
161 stable properties were used; minimal growth that could be kept in check
162 with carrying capacity and low elasticity. ? described the formula for
163 computing transition data as laid out in figure 2 and provided the values

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

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

Table 1. Components of Low-elevation perennial **Mimulus guttatus** life-cycle transition. Seed survival (D) is a general value taken from ?. All other components taken from ? who recorded values specific to that **M. guttatus** 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

for a low-elevation perennial *M. guttatus* population as shown in table 1. These data produced the transition matrix (figure 2) provided to the model, along with the values of G (0.652) and D (0.534) shown in figure 1.

```
knitr:::include_graphics("_images/m-guttatus-transformula")
```

Varying the rate of whole-genome duplication In order to quantify how the rate of whole-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.

175 Diagnostic plotting of these data revealed that increased rates needed to
176 be examined. Further simulations were then run using randomly sampled
177 rates from 100000 random uniform numbers between 0.010 and 0.500 in
178 an attempt to determine how high the mutation rate needed to be in order
179 for polyploids to break free from the inherent selection against them due
180 to mate-choice limitation.

181 *Germination rate benefit* Further simulations which inferred a
182 germination benefit to polyploids were run using the same randomly
183 sampled rates of polyploidisation as above. A randomly sampled range of
184 germination rate modifiers were chosen from 100000 random uniform
185 numbers between 1 and 1.5, where the modifier was used to multiply the
186 germination rate ($G = 0.652$). The max value of 1.5 would cause
187 polyploid seeds to germinate at very close to complete success ($G =$
188 0.978), simulating the observation that polyploid seeds can have
189 increased germination success due to germination over a wider
190 temperature range (?).

191 3 Results

```
#library(spliody)
#count_combine("data/results", "thesis/_data/results")
counts <- readRDS("_data/results.rds") %>%
  dplyr::select(-diploid_seeds, -polyploid_seeds, -rosette)
  tidyverse::gather("subset", "count", seeds:total)

ploidylvl_colours = c("#000000", "#30a068")
ploidy_gradient = c("#10442a", "#3bbf7d")
```

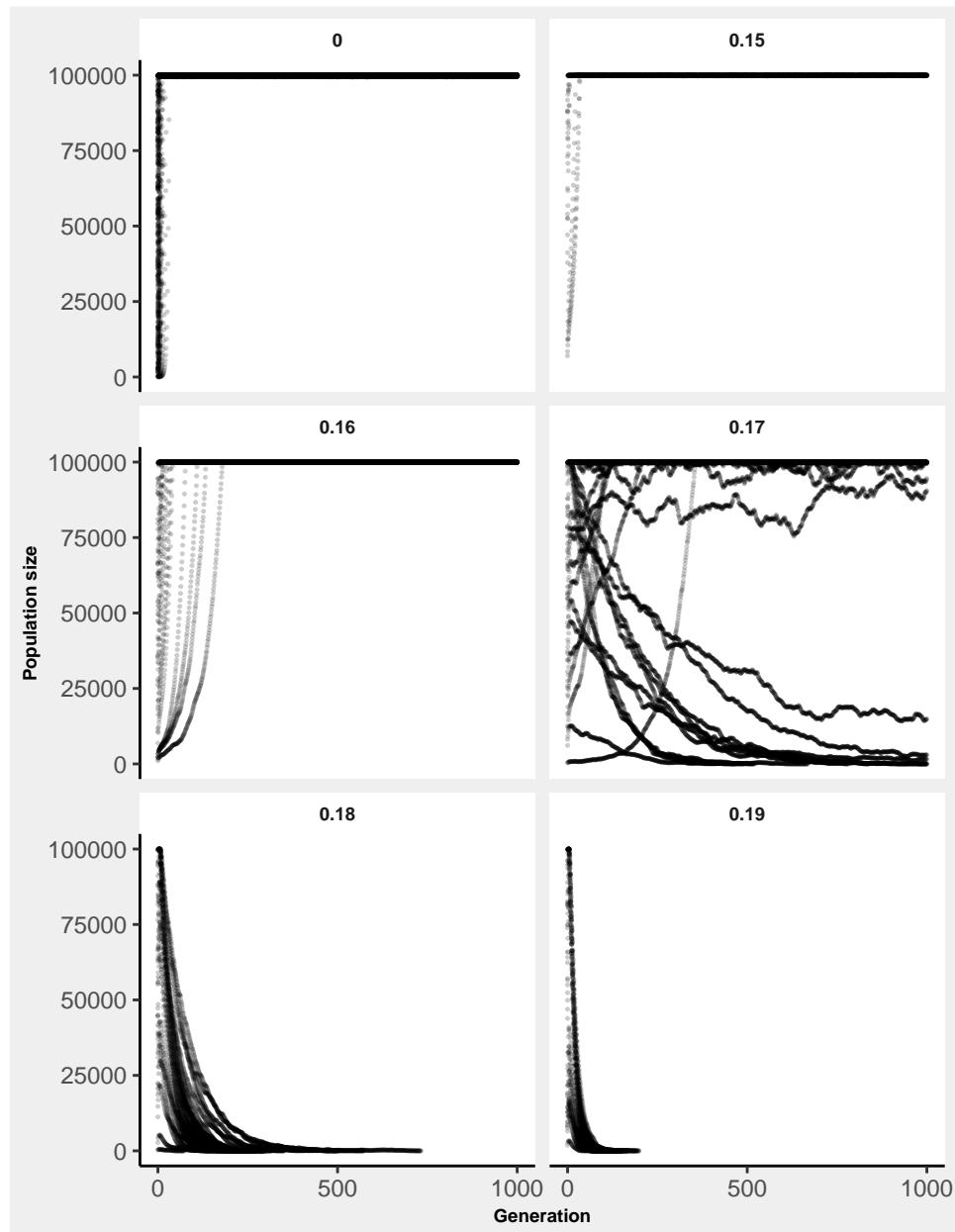


Figure 3. Total adult population size over time (generations). Each panel represents simulations run at varying rates of whole-genome duplication (rounded to 2 significant figures); 0 shows the expected results without mutation while panels 0.15 through to 0.19 show how the pattern of population growth and decline begins to change within this range of rates.

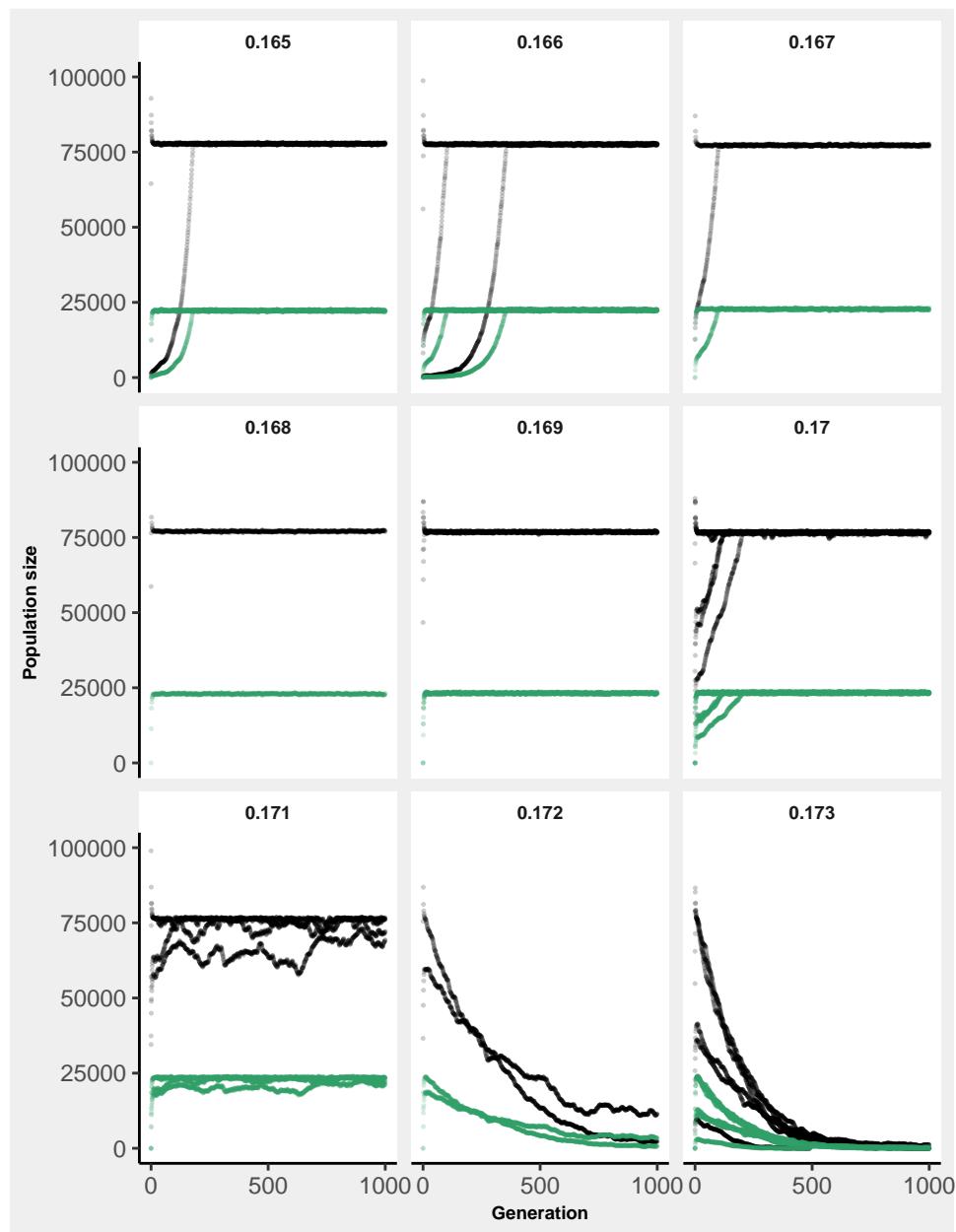


Figure 4. Adult population size over time (generations), split by diploid (black) and polypliod (green). This set of panels shows panel 4 of figure 3 in higher resolution of whole-genome duplication rates between 0.165 and 0.174 (when rounded to 3 significant figures).

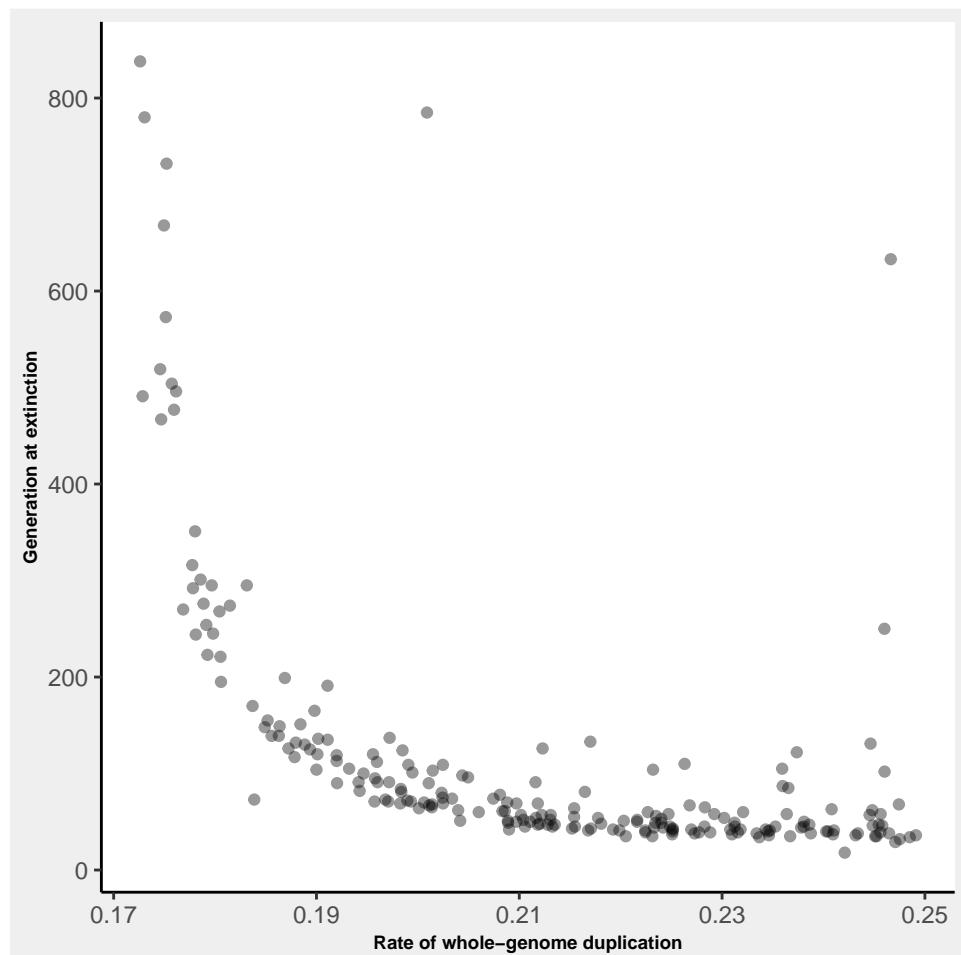


Figure 5. How the rate of whole-genome duplication affected extinction events. Each point represents a simulation that went extinct before one thousand generations had passed.

192 4 Discussion

193 Does limited mate-choice set the scene for the evolution of beneficial
194 traits like selfing?

195 **5 Conclusion**

196 Does limited mate-choice set the scene for the evolution of beneficial
197 traits like selfing?

198 A bit more manageable for a dissertation would be to try to flesh out
199 one or two of these mechanisms and ask some questions about the
200 conditions that it specifically is predicted to contribute substantially to
201 diversification of polyploids – e.g., are there reasons to think that pollen
202 swamping will lead to dead ends under some conditions, but
203 diversification (or at least not dead ends) under others?

204 **References**

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

207 Creasy M.R., Crolla J.A., Alberman E.D. (1976) A cytogenetic study of human
208 spontaneous abortions using banding techniques. *Human Genetics*, 31 (2), pp.
209 177–196.

210 Hashimoto S., Igarashi L., Kobayashi M., Tsutsumi H. (2018) Long-term extrauterine
211 survival in a triploid infant: A review of the clinical features of live-born infants with
212 triploidy. *Hong Kong Journal of Paediatrics*, 23 (4), pp. 298–301.

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

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

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

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

- 224 Ramsey J., Ramsey T.S. (2014) Ecological studies of polyploidy in the 100 years
225 following its discovery. *Phil Trans R Soc B*, 369.
- 226 Svartman M., Stone G., Stanyon R. (2005) Molecular cytogenetics discards polyploidy
227 in mammals. *Genomics*, 85, pp. 425–430.
- 228 Wickham H., François R., Henry L., Müller K. (2020) *dplyr: A Grammar of Data*
229 *Manipulation*. R package version 0.8.5.
- 230 Wood T.E., Takebayashi N., Barker M.S., Mayrose I., Greenspoon P.B., Rieseberg L.H.
231 (2009) The frequency of polyploid speciation in vascular plants. *Proceedings of the*
232 *National Academy of Sciences of the United States of America*, 106 (33), pp.
233 13875–9.
- 234 Woodhouse M., Burkart-Waco D., Comai L. (2009) Polyploidy. *Nature Education*,
235 2 (1), p. 1.