Electronic supplementary material

Polyploid establishment and evolution: A neutral speciation model for the eco-evolutionary dynamics of mixed-ploidy systems

Felipe Kauai[1,2], Frederik Mortier[1,2], Silvija Milosavljevic[1,2], Yves Van de Peer[2,3,4*], Dries Bonte[1*]

[1]Department of Biology, Terrestrial Ecology Unit, Ghent University, Karel Lodewijk Ledeganckstraat 35, BE-9000 Ghent, Belgium

[2]Department of Plant Biotechnology and Bioinformatics, Ghent University, VIB - UGent Center for Plant Systems Biology, B-9052 Ghent, Belgium

[3]Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Pretoria, South Africa

[4]College of Horticulture, Nanjing Agricultural University, Nanjing, China

*yves.vandepeer@psb.vib-ugent.be

*Dries.Bonte@UGent.be

Contents

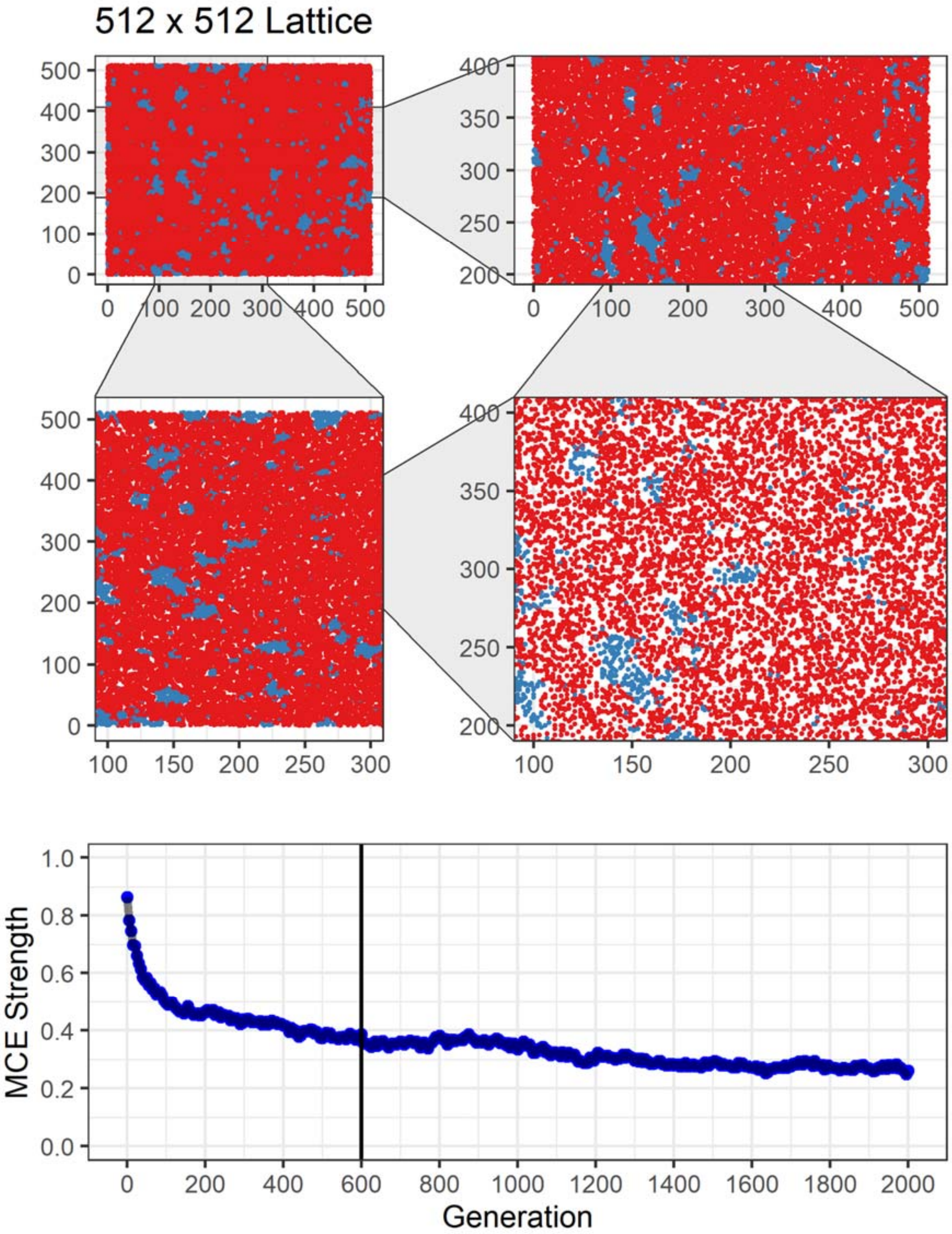

Figure S1. **Dynamics of mixed ploidy populations on a square lattice of size 512x512 cells.** Upper panel: screenshot of the system at generation 600, with a population size of 40000 individuals. Blue dots represent polyploids and red dots diploids. Upper left corner is a screenshot of the entire space, whereas the three remaining lattices represent different zooms of the whole space. Lower panel: time evolution of MCE Strength. Vertical black line represents the place where the screenshot was taken. Reduced gamete frequency $\varphi = 0.05$ and Reduced Fertility $\varepsilon = 0.12$. Notice that the dynamics is the same as in the system presented in the main text.

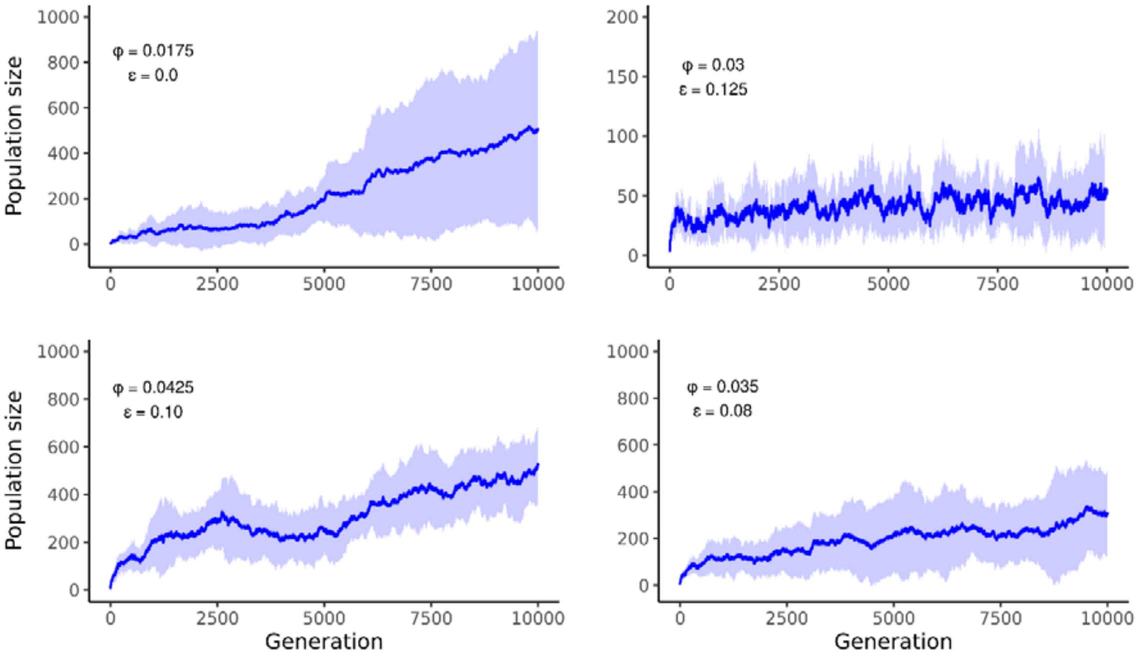

Figure S2. **Polyploid population size for different combinations of φ and ε.** Different dynamics are observed for distinct combinations of (φ, ε), with the fully neutral scenario $\varepsilon = 0.0$ in a). In b) it can be observed that polyploids never establish in the system, as at every generation their population size is roughly φS, i.e., polyploids are the product of the merging of unreduced gametes from diploid parents and do not reproduce among themselves. In c) and d) a slow increase in polyploid population size is observed towards the end of the simulation.

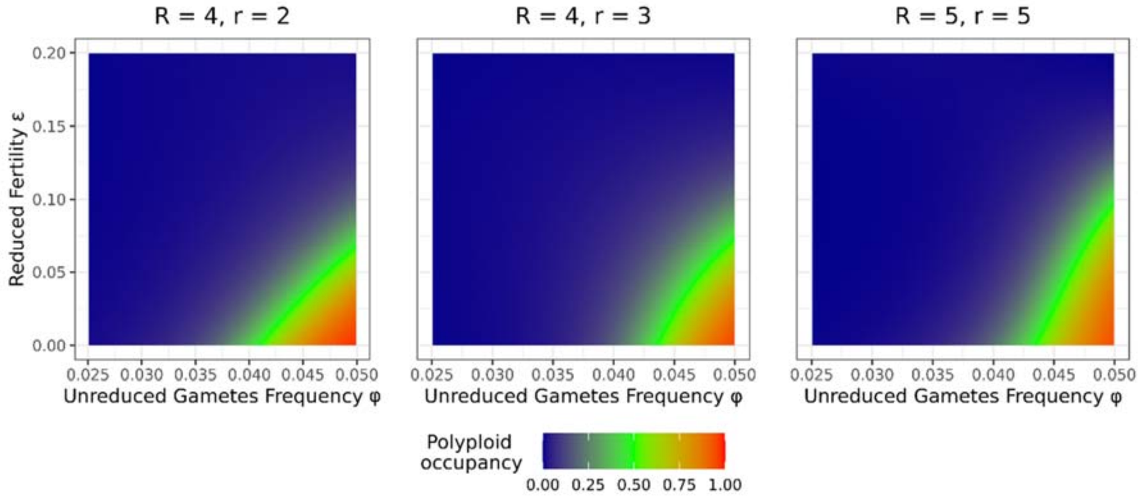

Figure S3. **Polyploid occupancy for various mating and dispersal radii**. The same pattern displayed in the main text can be obtained by varying R and r. Notice, however, that for polyploid occupancy to reach 50% increasing values of φ are necessary for increasing mating and dispersal radii (compare with main text). This is due to the fact the number of potential mates is clearly proportional to R^2 and thus polyploid formation must increase accordingly to offset MCE. By the same token, ε must increase to avoid polyploid dominance as φ increases.

Text S1. Model dynamics for single-ploidy systems.

To gain a better understanding of speciation events in diploids and tetraploids here we consider single-ploidy populations. In a diploid-only system, the average number of species starts to stabilize around generation 6000, and by the end of the simulation the average value is 4.5 (Figure S4A). On the other hand, for a tetraploid-only system, the average number of species stabilizes sooner and has an average of 1.9 by the end of the run. Thus, the number of species emerging in mixed-ploidy populations is considerably higher than in systems with a single cytotype. This higher diversity is attributed to the iterative aggregation of polyploids in different regions of space, which leads to the spatial segregation of both diploid and tetraploid subpopulations. This induces allopatric speciation on an otherwise connected population. Such a scenario produces differential evolution among subpopulations by the continuous sampling of genotypes containing information that do not get transferred among spatially segregated individuals of the system.

The number of species produced in tetraploid-only populations is much lower compared to diploid-only populations. The increased size of the genome in tetraploids leads to a higher number of genome combinations that allow individuals to become compatible mates and therefore maintain gene flow. For example, consider a sequence of 10 nucleotides and a genetic similarity requirement for compatibility of 90%. Then, there are ${}_{10}C_1$ ways that two sequences can differ and still be compatible. If, however, the sequence is doubled, i.e., has 20 nucleotides, then there are ${}_{20}C_2$ ways that these sequences are still compatible, that is, 19 times more combinations. This can be seen by measuring the genetic similarity among individuals as a function of their distances in space (Figure S4B).Tetraploid populations are much more connected throughout space, requiring greater genomic divergences to take place in order to break the network into disjoint clusters.

Nevertheless, the higher mutational load in tetraploids, measured as the expected number of mutations $\pi = c\mu|B|$, where c is the ploidy level, increases the genetic diversity of tetraploid populations much more than in diploid populations. By computing the average genetic similarity between any two individuals in a system where no spatial or genetic restrictions are imposed to mating, the average GS in the population with respect to time is well described by an exponential decay function, with tetraploids reaching much lower values by the end of the simulation (Figure S4C). As no sampling of specific genetic signatures are in place, which hold subpopulations of similar individuals together, the higher mutational load leads to fast divergence among individuals. This is further driven by the recombination rate during meiosis, which has a more pervasive effect on the evolution of higher ploidy genomes (Figure S4D). Notice that by setting the recombination rate to zero, the average genetic similarity difference between cytotypes is only due to mutations, and this difference increases as the recombination rate increases, but no discernable increase can be seen for recombination rates greater than 1%.

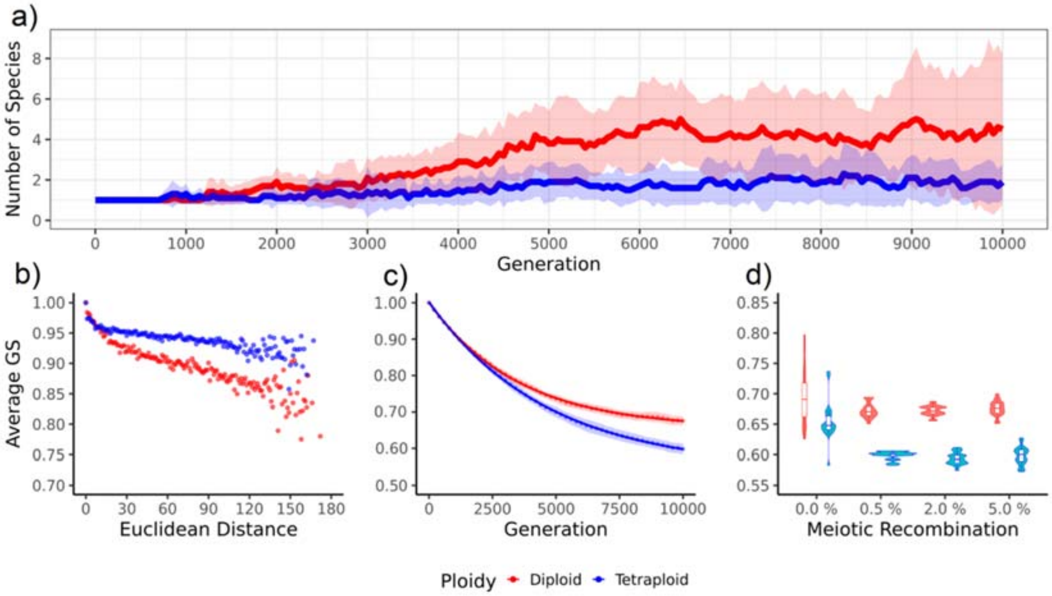

Figure S4**. Speciation dynamics in single-ploidy populations.** (a) Number of species per generation for diploid (red) and polyploid (blue) single cytotype populations. (b) Average genetic similarity between any two individuals as a function of the Euclidean distance in the lattice. (c) Average genetic similarity as a function of time between any two individuals in a population evolving with no space restrictions or GS_{min}, and (d) average genetic similarity computed at generation 10,000 as a function of recombination rate during meiosis, expressed as percentage values.