Sympatric speciation as explanation for protractedness

Kees W. Wesselink
Supervisors: Richèl J. C. Bilderbeek & Rampal S. Etienne
April 07, 2017

Abstract

In nature there are a lot of different species which all evolved from one single species. This research focuses on one mechanism of this diversification, sympatric speciation. The Protracted Birth-Death model of Etienne and Rosindell (2012) has no mechanism on individual level, but assumes no gene flow between incipient and good species, thus hinting that the protractedness is caused by allopatric speciation. In this research I tried to explain the protractedness of this model with sympatric speciation. This research focuses on another individual-based model of sympatric speciation of van Doorn and Weissing (2001). I found difficulties to get sympatric speciation in the simulation with this model. I suggest using a higher number of individuals for future simulations, because simulations with a higher number of individuals lead to more stable branching of the sexual traits in van Doorn´s model.

Inhoud

Introduction

When we look around we see a lot of different species. They all evolved from one species. This process is called speciation. One of the most well-known geographic modes of speciation is allopatric speciation (Mayr, 1942).

During allopatric speciation the original population is geographically separated after which the two new populations become genotypic diverse, by means of for example genetic drift (Mayr, 1942). When this physical barrier would be removed and the two distinct populations are no longer able to mate a new species is born.

Another less studied form of speciation is sympatric speciation (Dieckmann & Doebeli, 1999). Individuals of the same population are no longer geographically isolated. Instead, the speciation happens within the same population.

There are several mechanisms for this to happen. One mechanism is disruptive selection (Dieckmann & Doebeli, 1999). Disruptive selection is when, in a given trait, selection favors extreme values over intermediate values (see Figure 1). One example of disruptive selection is when under influence of, for example, competition, a fitness optimum might become a fitness minimum. This happens when a population arrives at the resource and fitness optimum and not all resources are used (see figure 1B). Offspring that makes use of the unused resources have a higher fitness. The population then moves away from the resource optimum to the new fitness optimum after which two new populations come into existence that both occupy a different ecological niche (van Doorn & Weissing, 2001).

Speciation is often researched with theoretical models (van Doorn et al., 1998; Dieckmann & Doebeli, 1999; Drossel & McKane, 2000). These models are used to answer questions about single aspects of speciation. Therefore these models are extremely simplified compared to speciation in nature. Sometimes there's not even a mechanism implemented on individual level.

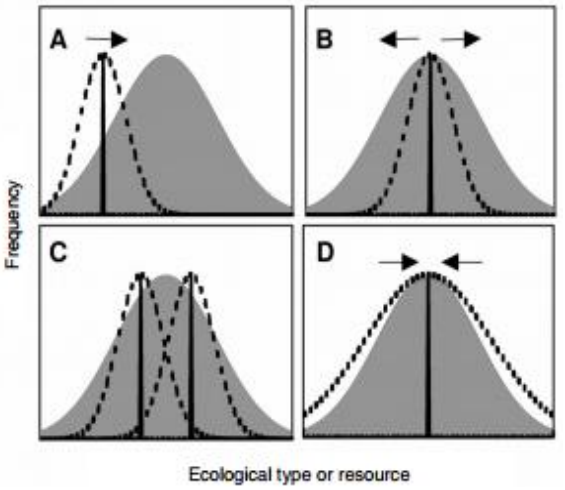

Figure 1: Disruptive selection (van Doorn & Weissing, 2001) (A) A population (solid line) not on the peak of a resource distribution (grey area) will evolve towards this resource optimum (directional selection). **(B)** An ecological specialist (width of resource utilization (dashed line) is smaller than width of resource distribution) will experience disruptive selection at the peak of the resource distribution. **(C)** Evolutionary branching occurs, the population becomes dimorphic for ecological type. **(D)** An ecological generalist (width of resource utilization is bigger than width of resource distribution) will remain at the peak of the resource distribution (van Doorn *et al.*, 2001)

This is the case with the Protracted Birth-Death (PBD) model of Etienne and Rosindell (2012). In their paper they explicitly mention they never assumed a mechanism for the speciation. The model looks similar to allopatric speciation, because it is assumed there is no gene flow between species (Etienne & Rosindell, 2012). In the model new species are not "recognized" as good species immediately and are called incipient. After a while, the incipient species will be recognized and are good species from then on. This is what makes the model protracted.

Since it's only an assumption that there is no gene flow between the incipient and good species, without an actual implementation of an individual based mechanism, I hypothesize that limited gene

flow in this model is possible and could actually be an explanation for this protractedness of the model.

Therefore my research question will be:

Can sympatric speciation be an explanation for protractedness in the Protracted Birth-Death model?

The Models

Sympatric speciation model

To answer my research question this research will compare the PBD model of Etienne and Rosindell (2012) with a model that explains sympatric speciation.

A model that does exactly this is the model of van Doorn and Weissing (2001). I picked this model, because this model combined both an ecological trait and two sexual traits. This is unlike other models which only looked at either speciation on an ecological or a sexual trait. They tried to explain when and how sympatric speciation occurred.

In the model all individuals have three heritable traits: an ecological trait, a male trait and a female preference for said male trait. At the beginning of every iteration the mortality rate (m_i) of all individuals will be calculated with equation 1.

$$m_i = \gamma \frac{\sum_k g_c(x_i - x_k)}{g_K(x_i)}$$

Equation 1: Mortality rate (van Doorn & Weissing, 2001)

The model uses a lot of Gaussian distributions which result in values between 0 and 1. All Gaussian distributions have a mean value μ and a variance σ^2. The Gaussian distributions in this model have a mean value μ of 0 and the deviation σ to calculate the variance is settable. For every deviation σ there is a parameter found in table 1.

Parameter	Biological interpretation
b	birth rate
ε, η	cost of mate choice (ε = η / number of males)
σ_k	width of the ecological resource distribution
σ_c	width of the resource utilization function
σ_e	specificity of mate choice with respect to ecological type
σ_m	specificity of mate choice with respect to mating type
σ_s	strength of viability selection on male mating type
σ_v	width of distribution of mutation sizes
σ_p	width of female mating type distribution
μ	Population average female mating type

Table 1: Parameters model van Doorn (van Doorn & Weissing, 2001)

In the simulation, m_i will usually have values between 0 (this individual will not die) and 1 (this individual will die), but it's theoretically possible to have a higher value (the individual will die). In this equation, g_c is a function that calculates the competition between two individuals (a value between 0 and 1 picked from a Gaussian distribution). The function $g_k(x_i)$is a value on a Gaussian distribution for the fitness value of individual i with ecotype x.

The inverse of this function (1 – m_i) gives the survival rate. The survival rate is the chance of an individual to survive. When an individual survives (determined by chance and the survival rate), the model picks a second individual at random. To pick another individual the attractiveness between the first individual and all other individuals in the population is calculated. The individual with the highest attractiveness has the highest chance to be picked.

$$a_{ij} = g_m(p_i - q_j)\, g_e(x_i - x_j)$$

Equation 2: Attractiveness between individuals i and j (van Doorn & Weissing, 2001)

The attractiveness is calculated according to equation 2. Where a_{ij} is the attractiveness between two individuals i and j. The attractiveness has a value between 0 and 1 since it's the product of the two values g_m and g_e which have a value between 0 and 1 for the attractiveness with respect to the difference in mating type and the difference in ecological type respectively. The chance that a certain individual i will mate with individual j, α_{ij} follows from the attractiveness a_{ij} as seen in equation 3.

$$\alpha_{ij} = \frac{a_{ij}}{\eta + \sum_{\text{males } k} a_{ik}}$$

Equation 3: chance that individual i will mate with individual j (van Doorn & Weissing, 2001)

In equation 3 η is a value that shows the cost of searching for a partner. As a bigger cost η makes searching for a partner more costly, it results in a smaller chance of mating between individual i and j α_{ij} (see equation 3).

A fertilized female (each female can be fertilized only once) will produce b children and will then die. When the model did the previous steps for every individual the new population (surviving individuals that did not mate and children) will be the next generation.

Protracted Birth-Death (PBD) model

The PBD model of Etienne and Rosindell (2012) is an extension of the Birth-Death (BD) model, a stochastic model of Kendall (1948). In the BD model new species are born (without mechanism) at a constant rate (the birth rate b). The species in the model will also go extinct at a constant rate (the extinction rate μ). The lineages-through-time (LTT) plots of phylogenetic trees we get from real DNA alignments often show a slowdown in the speciation rate closer to the present. The BD model however misses this (see figure 2 dashed line) and instead has a upward turn towards the present, called the pull of the present (Etienne & Rosindell, 2012; Nee et al., 1994).

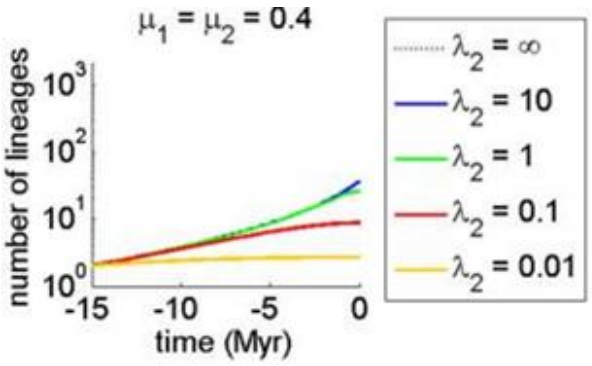

Figure 2: nLTT (Janzen *et al.*, 2005) plot of the PBD model with an extinction rate of 0.4. The dashed line goes through the blue line. An infinite speciation-completion rate means the model acts the same as the BD model (species are born as good species). A lower value of λ means it takes longer for incipient species to become good species. (Etienne & Rosindell, 2012)

The PBD resolves this (see figure 2) by adding protractedness which means that it takes some time for newly born species to be recognized as good species. Until they become good species they are called incipient. The model has 5 parameters (see figure 3 and table 2). One of the parameters is the speciation-completion rate λ. It has values from 0 (speciation never completes) to infinite (speciation completes instantly).

Parameter	Interpretation
b_g	Rate of good species giving birth to new species
b_i	Rate of incipient species giving birth to new species
μ_g	Extinction rate of good species
μ_i	Extinction rate of incipient species
λ	Speciation-completion rate

Table 2: Parameters PBD model (Etienne and Rosindell, 2012)

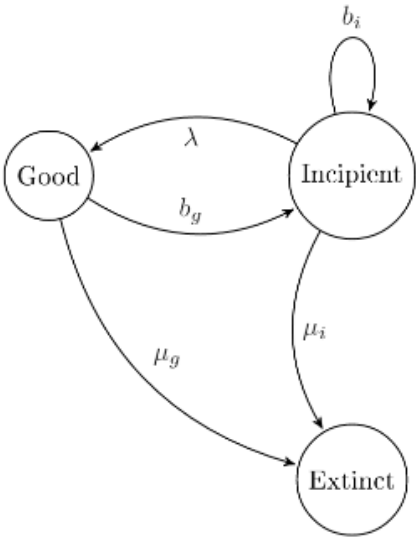

Figure 3: Protracted Birth-Death model (Bilderbeek & Etienne, 2017)
A graph showing the parameters of the PBD model of Etienne and Rosindell (2012); b_g and b_i: speciation-initiating rates for good and incipient species. μ_g and μ_i: extinction rates for good and incipient species. λ: speciation-completion rate.

Incipient species that have a good parent species that is still alive do not show up in the phylogeny, since they are not recognized as good species yet, . This explains the curves becoming less steep closer to the present, since the chance of a species being incipient is higher closer to the present.

The PBD model shows great similarities with allopatric speciation. With allopatric speciation there is no gene flow between the two populations that will become species. The PBD model assumes no gene flow between two or more (incipient) species. There is, however, as mentioned before, no individual based mechanism included that proves this assumption.

Differences between PBD and the sympatric model

The models are in almost all aspects opposites. The sympatric model of van Doorn has an individual-based mechanism, while PBD does not. PBD shows great similarities with allopatric speciation, while the model of van Doorn is sympatric.

This research

Right now it is uncertain if the protractedness in the PBD model can be explained by sympatric speciation. The original model does not assume an individual based mechanism and even looks more like an allopatric model, because there it is assumed there is no gene flow between species. A lot of research on sympatric speciation has focused on unique scenario's for which sympatric speciation will occur and not on the similarities with allopatric speciation (Gavrilets, 2003).

In this research I planned to look at the results of both models to see if they gave comparable outputs and to find an explanation for the protractedness in the PBD model. However, I found that I could not get the model of van Doorn to get speciation as shown in his paper of 2001. That's why the main focus of this report is on explaining the absence of speciation in this model.

Hypothesis

I expect the phylogenetic trees of the sympatric model of van Doorn to be highly likely in the PBD. I think the incipient part of the lineages in the PBD model can be seen as species with a reduced gene flow between them. This is similar to the reduced gene flow found in sympatric speciation (Drossel and McKane, 2000). If there is still some gene flow we fail to recognize the species as good species, but as soon as the gene flow is small enough (as in sympatric speciation), we can recognize these species as good species. Which is the same as the protractedness in the PBD model. That's why I expect that sympatric speciation can explain the protractedness in the PBD model.

Method

Both models use vastly different parameters (see table 1 and 2). The parameters of the PBD model will be estimated by a maximum-likelihood estimation (Scholz, 2006) on the trees of the model of van Doorn to make the phylogenetic trees of both

models comparable. The maximum-likelihood estimation will give an estimation of the parameters for the PBD that are the most likely to give the tree that was used as an input. Using those parameters in the PBD should give trees that are comparable to the trees of the model of van Doorn.

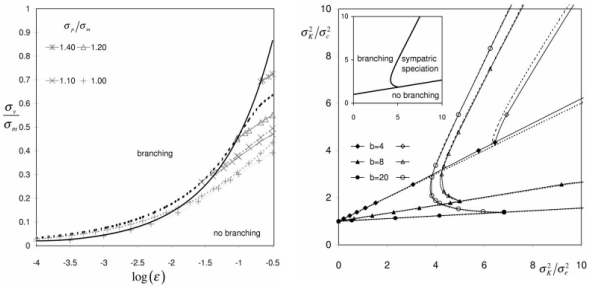

Figure 4: Parameters spaces of the SES model (van Doorn & Weissing, 2001)
(left panel) The parameter space for which branching of the sexual types will (not) occur. **(right panel)** The parameter space for which branching / sympatric speciation of the ecological type will occur.

The parameters for the model of van Doorn are found in the paper of van Doorn and Weissing (2001). In their paper they used an analytical approach to find parameters for which branching and / or sympatric speciation would occur (see figure 4). Individuals will be haploid and will have a single locus (van Doorn & Weissing, 2011).

The phylogenies of both models will be compared using nLTT plots (Janzen *et al.*, 2005) . I will look at the squared error of PBD phylogenies and PBD phylogenies from estimated parameters (from a maximum-likelihood estimation). I will compare this squared erro with the squared error of the sympatric model's phylogenies and PBD phylogenies from estimated parameters from the sympatric trees.

If the second squared errors (of the sympatric trees) are smaller than the first squared errors (the control group) I can conclude that sympatric speciation can be used as an explanation for the protractedness.

Results

With the initial parameters from the paper of van Doorn and Weissing (2001) the simulations did not show speciation, while van Doorn and Weissing did show it in their paper (see figure 5). The population actually had around 5,000 individuals with the given parameters so I increased constant γ to get it back to 1,000 individuals. This should increase the mortality rate, as seen in equation 1.

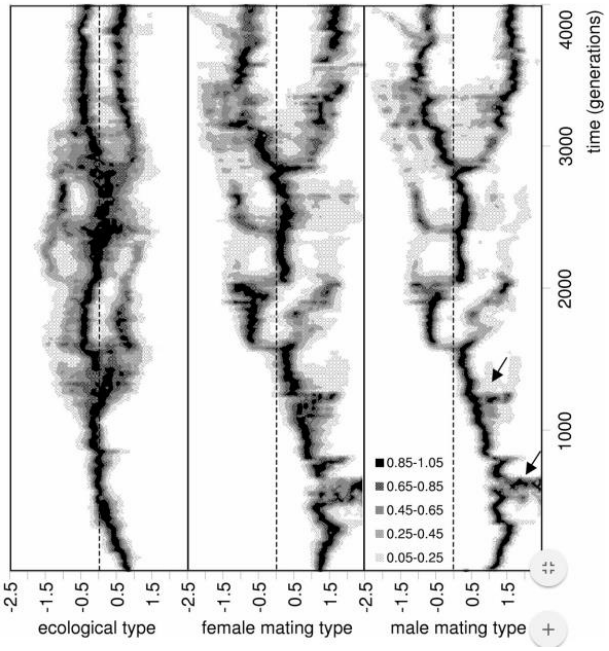

Figure 5: Speciation in the model of van Doorn. The three panels show the distribution of ecological and mating types (greyscale). Several times, polymorphisms of mating types originate (arrows), but these are unstable. Parameters were: $\sigma_k = 1.2$, $\sigma_c = 0.4$, $\sigma_e = 0.6$, $\sigma_m = 0.2$, $\sigma_v = 0.02$, $\sigma_s = 1.0$, $\eta = 1.0$, b 4.0. Parameters $\gamma = 5 \cdot 10^{-4}$ should keep the population sizes in the simulation close to about 1,000 individuals, but kept the populations sizes to about 5,000 individuals.

With these parameters I wasn't able to get speciation, mostly because there was no branching of the two sexual traits (see figure 6). Also the ecological trait started branching immediately when the simulation started. Parameter settings with a lower value for σ_m, to increase the chance of sexual type branching (see figure 4, left panel), did not result in branching.

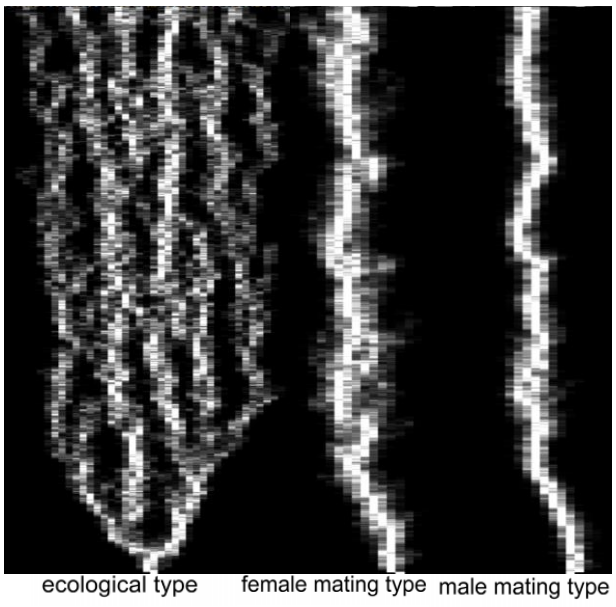

Figure 6: No speciation with same parameters as figure 5. This simulation ran for 900 generations. $\gamma = 5 \cdot 10^{-3}$ to stay close to about 1,000 individuals per generation. The whiteness of each pixel indicates the relative number of individuals with that specific trait. The most common trait was set to white. Black pixels mean there are no individuals with that trait.

The correlation for the simulation in figure 6 was still high (around either 0.9 or -0.9), even though there was no speciation.

One simulation (see figure 7) got to three species in 900 generations. After around 10 generations more the species merged back together to form one species. This was not enough to run a (useful) maximum likelihood estimation for the PBD on.

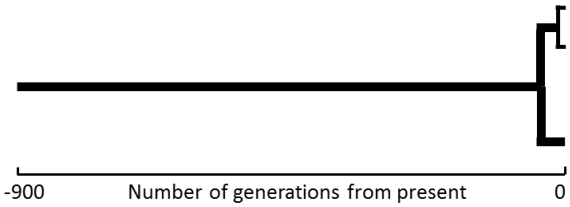

Figure 7: Phylogenetic tree of 1 simulation. Parameters were: $\sigma_k = 1.2$, $\sigma_c = 0.6$, $\sigma_e = 0.5$, $\sigma_m = 0.05$, $\sigma_v = 0.03$, $\sigma_s = 1.0$, $\eta = 1.0$, b 4.0. Parameters $\gamma = 1.75 \cdot 10^{-3}$. The simulation ran for 900 generations and had 2 speciation events after 869 generations and after 899 generations.

In the end I found a simulation that shows branching of the mating types (see figure 8), but this simulation was not tested for speciation. For this simulation parameters were kept exactly the same as in the paper of van Doorn and Weissing, which resulted, as expected, in more individuals, but also mating type branching.

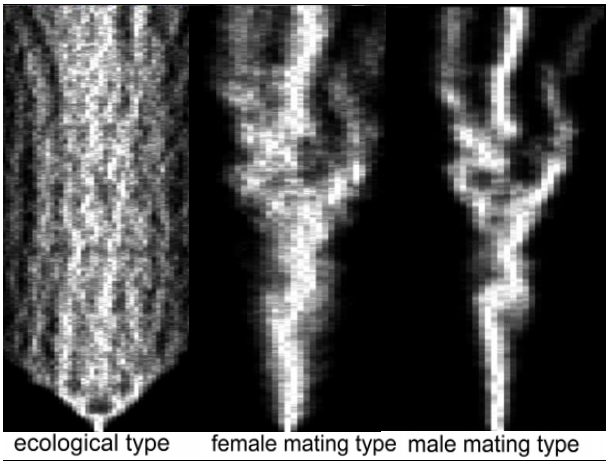

Figure 8: Branching of the mating type. Parameters were exactly as in figure 5. The whiteness of each pixel indicates the relative number of individuals with that specific trait. The most common trait was set to white. Black pixels mean there are no individuals with that trait

Discussion

Although all the parameters that were used were inside the parameter space van Doorn and Weissing pointed out as parameter space suitable for branching or speciation (see figure 4) I couldn't find speciation. I tried extremely small values of σ_m, down to $5 \cdot 10^{-3}$ which is four times as small as σ_v, the parameter for the width of the distribution of mutation sizes. Which means that an individual is able to mutate four times as much as the specificity of the mate choice of another individual.

So my definition of speciation might differ from the definition of van Doorn and Weissing. Their definition of speciation stated that speciation happened when there was a linkage disequilibrium between the ecological trait and the sexual trait. It has to form two distinct groups that are ecologically differentiated and reproductively isolated (van Doorn & Weissing, 2001). This is a definition that does not differ from mine.

Another interesting aspect of the results was the high correlation between the ecological trait and the sexual traits, but also the female preference and the male trait. As mentioned in the results it could go up as high as 0.9 without speciation. I had no way of showing what the correlation looked like except for this value.

This might be explained by a potential incorrect intuition of correlation. In figure 9 I tried to demonstrate this by showing my intuition of the correlation between female mating type and ecological type in the upper panel, and what might be happening in the lower panel. It is not clear which of the two is the case in the simulations.

With all these uncertainties and the low number of results it's impossible for me to give a conclusive answer on the research question. Although the latest result (see figure 8) suggests it's still possible to get speciation with this model.

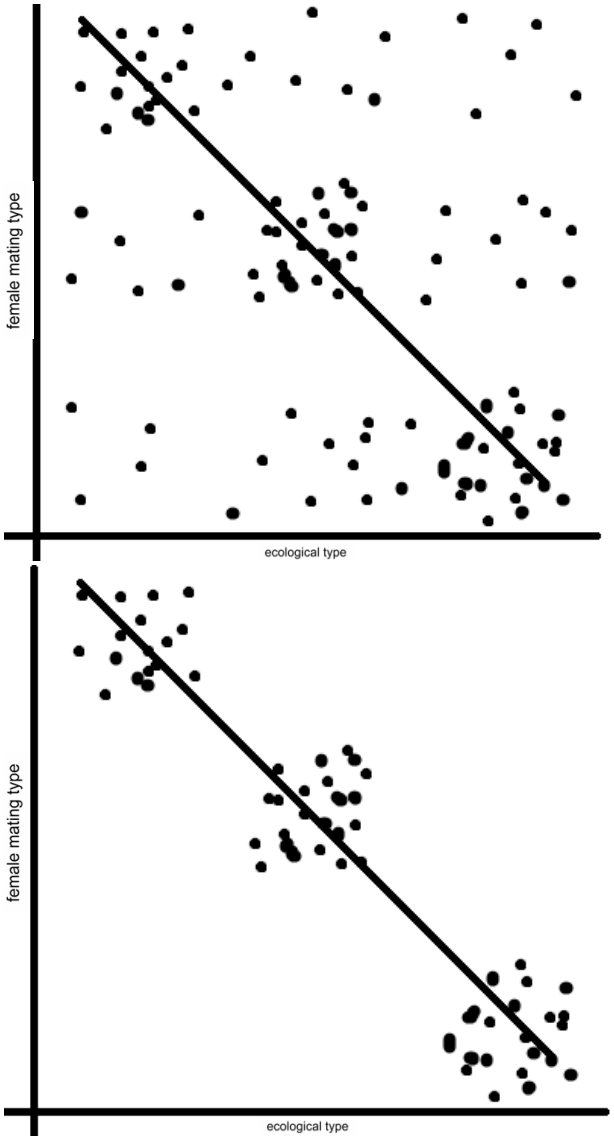

Figure 9: Potential correlation between female mating type and ecological type. Each dot is an individual. **(Upper panel)** First intuition **(Lower panel)** real situation.

The biggest difference between the latest result and the results before that, is that there is a lower number of individuals in the simulations without branching of the mating types. This suggests that increasing this number, by making the constant γ lower, might increase the chance of speciation.

Future research

Because the last result showed potential for actual speciation with this model I suggest some future research. To get speciation you could increase the number of individuals per generation which will probably have a higher chance of speciation. If this does not work I suggest looking at the correlation and try to find a real linkage disequilibrium.

Either way the latest results looks promising for this model to be used to simulate sympatric speciation on individual level.

References

- Bilderbeek J.C. & Etienne R.S. (2017) "*Should protracted speciation be incorporated in phylogenetic tree construction methods?"* Manuscript in preparation.

- Dieckmann U. & Doebeli M. (1999). "*On the Origin of Speciation by Sympatric Speciation"* Nature, **400**, 354-357

- Drossel B. & McKane A. (2000). *"Competitive speciation in quantitative genetic models"* J. Theor. Biol. **204:** 467-478

- Etienne R.S. & Rosindell J. (2012). *"Prolonging the past Counteracts the pull of the present: Protracted speciation can explain observed slowdowns in diversivication"* Syst. Biol. **61**, 204-213

- Janzen T., Höhna S. & Etienne R.S. (2005). "*Approximate Bayesian Computation of diversification rates for molecular phylogenies: introducing a new efficient summary statistic, the nLTT"* Methods in Ecology and Evolution **6**, 566-575

- Kendall D.G. (1948). *"On some modes of population growth giving rise to R.A. Fisher's logarithmic series distribution"* Biometrika **35**, 6-15

- Mayr E. (1942). "*Systematics and the Origin of Species from the Viewpoint of a Zoologist*" New York: Columbia University Press.

- Nee S., May R.M., Harvey P.H. (1994). "*The reconstructed evolutionary process.*" Philos. Trans. R. Soc. Lond. B. **344**, 305–311.

- Scholz F.W. (2006). *"Maximum Likelihood Estimation"* Encyclopedia of Statistical Sciences **7**.

- Van Doorn G.S., Noest A.J. & Hogeweg P. (1998). *"Sympatric speciation and extinction driven by environment dependent sexual selection"* Proc. R. Soc. Lond. B **265**, 1915-1919

- Van Doorn G. S. & Weissing F. J. (2001) "*Ecological versus Sexual Selection models of Sympatric Speciation: a Synthesis"* Selection **2**, 17-40