Is or should sympatric speciation be implemented in the Protracted Birth-Death model?

Kees W. Wesselink
Supervisors: Richèl J.C. Bilderbeek & Rampal S. Etienne
September 05, 2016

Abstract

In nature there are a lot of different species which all evolved from one single species. This research focuses on the similarities between two mechanisms for this diversification, sympatric and allopatric speciation. In this research I will look at two models (The Protracted Birth-Death model (Etienne & Rosindell, 2012) and a Sexual and Ecological Selection model (van Doorn & Weissing, 2001)) and look for similarities in the phylogenetic trees created by them. Both outputs are compared using an nLTT statistic (Janzen *et al.,* 2005). In this research I will find something and conclude something, of course I will discuss my own research.

Introduction

When we look around we see a lot of different species. They all evolved from one species. The exact mechanism is still not completely known, but many theories exist. One of the most well-known theories is allopatric speciation (Mayr, 1942). During allopatric speciation the original population is geographically separated after which the two new populations become genotypic divers. This happens because of selection or genetic drift (Mayr, 1942). If the physical barrier is then removed it's possible that individuals of the two distinct populations are no longer able to mate. A new species is born.

Another less researched 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 ways for this to happen. One way is disruptive selection (Dieckmann & Doebeli, 1999). Under influence of for example competition a fitness optimum becomes a fitness minimum. Then individuals from the population move away
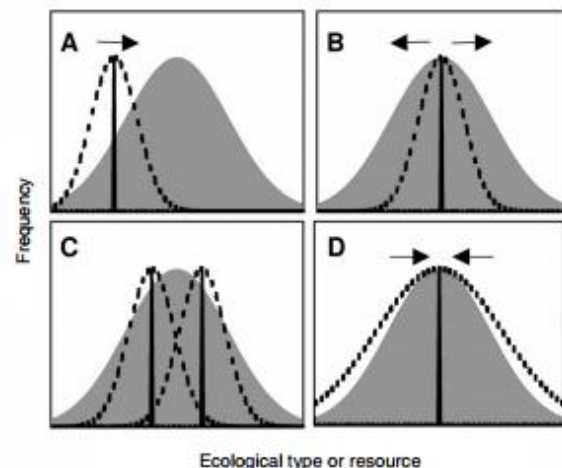

Figure 1: Disruptive selection (van Doorn & Weissing, 2001) (A) A population (solid line) not on the peak of a resource distribution (grey area) will evolve towards this resource optimum (directional selection). **(B)** An ecological specialist (width of resource utilization (dashed line) is smaller 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)

from the resource optimum in different ways (see fig. 1B) after which two new populations come in to existence that both occupy a different ecological niche (van Doorn & Weissing, 2001).

Speciation is often researched with theoretical models (Kawecki, 1997; Payne & Krakauer, 1997, van Doorn *et al.*, 1998; Dieckmann & Doebeli, 1999; Higashi *et al.,* 1999; Kondrashov & Kondrashov, 1999; Drossel & McKane, 2000). These models are used to answer questions about single aspects of speciation to get as close to reality as possible. Diversification as a whole is better understood with the improvement and implementation of each model. This way the process of going from one species to the billions of species we have right now will be unraveled. These models are of course extremely simplified compared to real life. Sometimes there's not even a mechanism implemented. 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, but the model looks similar to allopatric speciation (Etienne & Rosindell, 2012). The question this research tries to answer is if the PBD model (without a mechanism) also works for species born from sympatric speciation and (if this is not the case) if it should be implemented. Therefore my research question will be:

Is or should sympatric speciation be implemented in the Protracted Birth-Death model?

The Models

Sexual and Ecological (SES) model

To answer this 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 Ecological and Sexual selection model of van Doorn and Weissing (2001). In this model van Doorn and Weissing tried to combine older models for either ecological

selection or sexual selection. 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 of all individuals will be calculated according to the following formula:

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

In this formula m_i is the mortality rate of individual I, γ is a constant to scale the carrying capacity of the system. $\sum_k gc(xi - xk)$ is the cumulative competition and $g_k(x_i)$ is the fitness value of individual I with ecotype x (this value lies on the Gaussian distribution g_k with mean 0 and standard deviation σ_k). The reserve of this function $(1 - m)$ gives the survival rate. The survival rate is the chance to survive of an individual.

When an individual survives the model picks a second individual. 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. The attractiveness is calculated with a value on the Gaussian curve g_m (mean 0 and standard deviation of σ_m) dependent on the difference in female preference and male trait of the two individuals and a value on the Gaussian curve g_e (mean 0 and standard deviation of σ_e) dependent on the difference in ecological trait of the two individuals. The two values are multiplied and result in a chance α_{ij}, the chance that the two chosen individuals I and J will mate.

Individual I has a chance of η to not mate. η has a higher value if searching for a partner has a high cost. A fertilized female 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 of Kendal (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 and instead has a upward turn towards the present, called the pull of the present (Etienne & Rosindell, 2012; Nee *et al.*, 1994). The
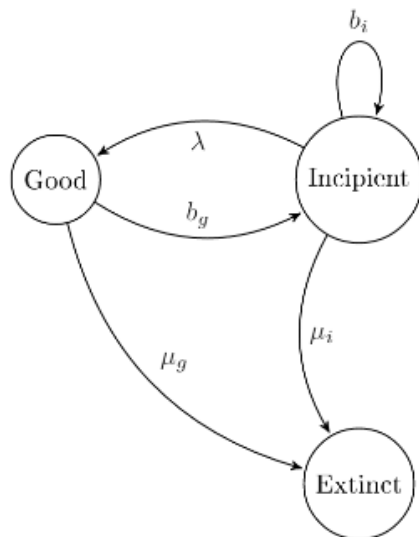
Figure 2: Protracted Birth-Death model (Bilderbeek & Etienne, 2017)
A graph showing the parameters of the PBD model of Etienne and Rosindell (2012); b_g and b_i: speciation-initiating rates for good and incipient species. μ_g and μ_i: extinction rates for good and incipient species. λ: speciation-completion rate.

PBD model resolves this by making the simple assumption that speciation takes time. The model differentiates between good and incipient species. Both good and incipient species initiate new incipient species with speciation-initiation rate b_g and b_i respectively. Both species also go extinct with the constant rate μ_g and μ_i. So far not much has changed compared to the BD model. Lastly they added a speciation-completion rate λ. This is a rate for which incipient species become good

species (see fig. 2). When set to infinity the PBD model will act exactly like the BD model. At the present only good species or the incipient species whose parent species are extinct will show up in the phylogeny and thus be counted in the LTT plot (see fig. 3). This way incipient species are species that already exist, but are not recognized as good species yet. Right now there are no further differences between incipient and good species. Both kinds of species have the same birth and extinction rate. The PBD model assumes there is no gene flow between two different (incipient) species, thus showing great similarities with allopatric speciation. There is however as mentioned before no mechanism included in the model.
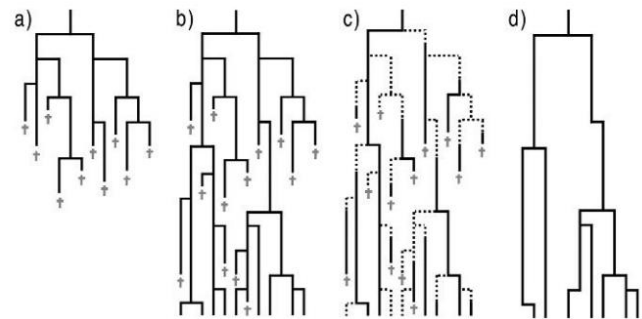
Figure 3: The birth-death model with and without protracted speciation (Etienne & Rosindell, 2012)
(A) A birth-death process that went extinct before the present.
(B) A birth-death process that survives up to the present.
(C) Birth-death process with protracted speciation, solid lines are good species, dotted lines are incipient species.
(D) Phylogeny of the process in panel c. Only good lineages or incipient lineages whose parents are extinct before the present show up.

Differences between PBD and SES

Both models are in almost all aspects opposites. SES has a mechanism on individual level, while PBD does not. PBD shows great similarities with allopatric speciation while SES is sympatric. In the SES model there is (a small) gene flow between good and "incipient" species, even though the chance is small. If this chance would be too big a new "incipient" species would flow right back to the old good species before branching could occur.

This research

Right now it is uncertain if the PBD model works for sympatric speciation, because the original model does not assume a mechanism. This research will compare the results of both models to see if and where there are differences. There is not a lot of research on the outcomes of sympatric speciation therefore it's useful to see how many differences there are compared to the outcomes of allopatric speciation. This research will alter both models to make sure they will give comparable outputs. The SES model was made in 2001 and includes many "unnecessary" methods to get its results.

Novelty

The PBD and the SES models haven't been compared to each other before. A lot of research on sympatric speciation is focused on unique scenario's for which sympatric speciation will occur (Gavrilets, 2003). Comparisons with allopatric speciation have been made, but most research focusses on the differences. This research will look at the similarities between both kinds of speciation and will find if they differ as much as often is claimed.

Hypothesis

The PBD will probably not differ a lot from the SES model, because the incipient part of the lineages in the PBD model can be seen as a reduced gene flow between the incipient species and the parent good species, similar to the reduced gene flow found in sympatric speciation, instead of no gene flow when assuming the PBD is allopatric. That means that the PBD will be usable for sympatric speciation.

Method

Both models use vastly different parameters. To make the phylogenetic trees of both models comparable a maximum-likelihood estimation (Scholz, 2006) will be used on the trees of the SES model to get parameters of the PBD. Using those parameters in the PBD should give trees that are comparable to the trees of the SES model.

The parameters for the SES model 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 fig. 4)

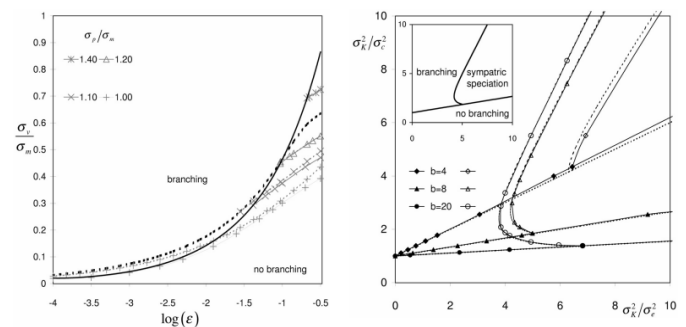Individuals will be haploid and will have a single locus (van Doorn & Weissing, 2011).

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
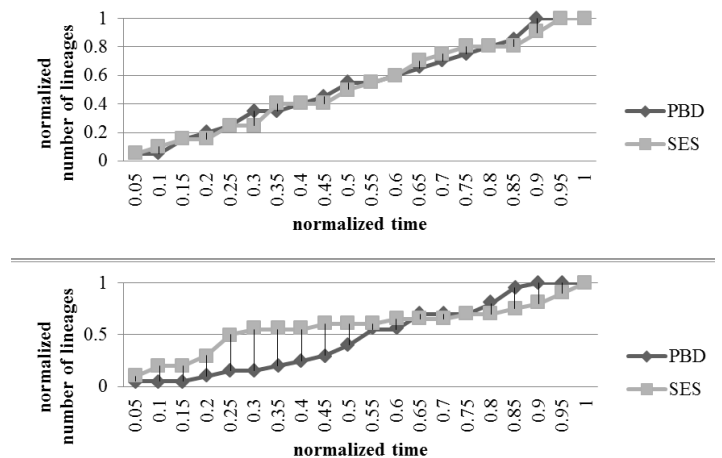

Figure 5: nLTT plots
Possible outcomes of the nLTT **(top panel)** very little difference between two trees gives a small area between the two plots. **(bottom panel)** more differences gives more area between the two plots.

A normalized lineage-through-time plot will be made of the outcomes of both models. The area between both plots gives the differences between both phylogenetic trees (see fig. 5) (Janzen *et al.*, 2015).

The SES model has to be changed to give nLTT plots. Right now SES gives normalized number of

all ecotypes, female preferences and male traits through time. A lineage can be made by calculating the attractivenesses between all individuals. Between the individuals with a big enough chance of mating (5%) will be drawn a line. All individuals connected with a line will be clustered. Both the drawing of lines and the clustering is done using the Boost Library (Siek, 2002). The different clusters (or groups) that are formed will each be a new lineage / species.

Peripherals

Software

All C++ code can be found on
https://github.com/richelbilderbeek/pbdmms

(GitHub, 2016) where we host all versions of the models.

All code is checked for code coverage by CodeCov (CodeCov, 2016; Del Frate *et al.,* 1995) and style and compile errors by Travis (Travis, 2016)

Hardware

The models were ran on the Peregrine HPC cluster of the RUG ("Peregrine HPC cluster", 2016)

Results

No results yet, but see hypothesis.

Conclusion & Discussion

No conclusion and discussion yet, but see hypothesis

References

- (2016). GitHub http://github.com
- (2016). Travis CI. http://travis-ci.org
- (2016). CodeCov http://codecov.io
- Bilderbeek J.C. & Etienne R.S. (2017) *"Should protracted speciation be incorporated in phylogenetic tree construction methods?"* Manuscript in preparation.
- Del Frate F., Garg P., Mathur A.P. and Pasquini A. (1995) *"On the correlation between code coverage and software reliability"* In *Software Reliability Engineering (1995). Proceedings., Sixth International Symposium on,* 124-132. IEEE, 1995.
- 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
- Higashi M., Takimoto G. & Yamamura N. (1999). *"Sympatric speciation by sexual selection"* Nature **402**. 523-526
- 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
- Kawecki T. J. (1997). *"Sympatric speciation via habitat specialization driven by deleterious mutations"* Evolution **51**, 1751-1763
- Kendall D.G. (1948). *"On some modes of population growth giving rise to R.A. Fisher's logarithmic series distribution"* Biometrika **35**, 6-15
- Kondrashov A.S. & Kondrashov F.A. (1999). *"Interactions among quantitative traits in the course of sympatric speciation"* Nature **400**. 351-354
- 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.
- Payne R.J.H. & Krakauer D.C. (1997). *"Sexual selection, space,, and speciation"* . Evolution **51**, 1-9
- *"Peregrine HPC cluster"* (2016, december 22). Retrieved from http://www.rug.nl/society-business/centre-for-information-technology/research/services/hpc/facilities/peregrine-hpc-cluster/ [In Dutch]
- Scholz F.W. (2006). *"Maximum Likelihood Estimation"* Encyclopedia of Statistical Sciences **7**.
- Siek J.G., Lee L. and Lumsdaine A. (2002). "*The Boost Graph Library: Used Guide and Reference manual*" Upper Saddle River, NJ: Pearson Education, Inc.
- 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