

The differences in phylogenies of the BDM-model and the PBD-model

Jorik de Boer, September 2016

Abstract

The speciation process leading to modern biodiversity could have been affected by time dependent diversification processes. Multiple diversification models have been created to model speciation processes and to make corresponding phylogenetic trees. The protracted birth death model (PBD) tried to implement the concept that the speciation process takes time in the standard birth-death model. In LTT plots of phylogenies a decrease of lineages is visible towards the present. The PBD-model gives a possible explanation for the decline of lineages near the present, however the model lacks the implementation of a mechanistic process for speciation. In the Bateson-Dobzhansky-Müller (BDM) model genetic divergence leads to speciation. For this project I will create phylogenetic trees with the BDM-model and make a comparison between phylogenetic trees from the PBD-model and phylogenetic trees from the BDM-model. What are the differences in phylogenies based on the protracted birth-death model and the Bateson-Dobzhansky-Müller model? Phylogenies can be compared with summary statistics, for example the gamma summary statistic and the normalized lineages through time summary statistic.

Introduction

Biodiversity as it can be seen nowadays, with its huge range of species, raises questions about the process of speciation which has led to present day biodiversity. Which speciation processes did contribute to modern day biodiversity and how does time affect these processes? Multiple diversification models have been created to simulate phylogenies of species. For my research project I will work with two different speciation models: the protracted birth-death model by Etienne and Rosindell (2012) and the Bateson-Dobzhansky-Müller model by Bateson (1909), Dobzhansky (1936) and Müller. First I will explain the principles and background of both speciation models. Thereafter I will discuss the main question of the research project and the methods I will use. In the end I will mention assumptions and possible weaknesses that could play a role in the research process.

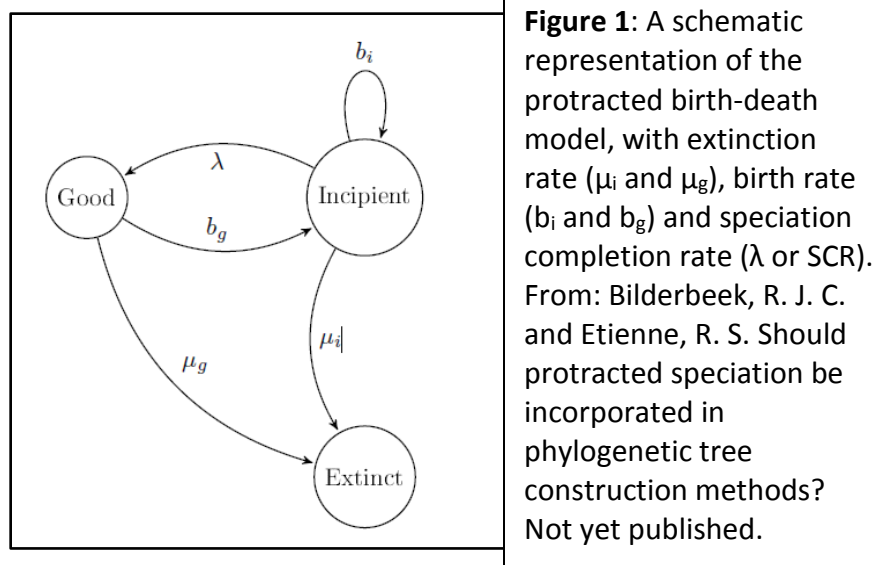
BD-Model

The standard birth-death model (BD) allows for extinction and birth of species (Referentie). A plot of the number of lineages versus time (lineages through-time (LTT) plot) of the BD-model shows a constant or accelerating increase in the number of lineages towards the present. This increase in the number of lineages towards the present is also called “pull of the present”. The pull of the present is caused by the circumstance that more recent originated species had lower chance to become extinct in contrast to older species. In LTT plots of phylogenies a decrease of lineages is visible towards the present, inconsistent with the BD-model simulations.

PBD-Model

The protracted birth-death model (PBD, also called protracted speciation model) (Etienne and Rosindell, 2012) tries to explain the slowdown towards the present, as seen in phylogenetic trees. The PBD-model is based on the BD-model. The PBD-model tried to implement the concept that the speciation process takes time in the BD-model. The PBD-model divided the speciation process into an incipient species stage and a good species stage (Figure 1). New incipient species can arise from both good species and incipient species. If the speciation

completion rate (SCR) in the protracted speciation model is close to infinity, the model is the same as the birth-death model if the extinction rates and birth rates have the same values.



BDM-Model

The Bateson-Dobzhansky-Müller (BDM) model is a mechanistic model, based on individuals with a certain genotype. The model is centered around the idea that two alleles at two different loci are incompatible (Figure 2). This incompatibility can be expressed in a reduced fitness component, for example fertility, viability, sexual attractiveness (also described in terms of epistatic interactions, lack of functions in hybrids and lack of matching between maternal and paternal genes) (Gavrilets, 2004). Examples of genes involved in incompatibilities of genotypes as in the BDM-model, like sterility of hybrids, can be found in fish species and fruit flies (Orr and Presgraves, 2000). The BDM-model can explain ridges of high-fitness genotypes in the simplest example of a “holey fitness landscape”. The model can evolve to strong reproductive isolation (RI) without crossing “fitness valleys”. In the BDM-model species can maintain high fitness during the speciation process, while in other speciation models species need to cross fitness valleys in a vulnerable state of low fitness. One of the biggest differences in compare with the PDB-model is the possibility of backward mutations. With backward mutations, so called good species could return to their former incipient state and vice versa.

The BDM-model can be used in a one-population scenario and a two-population scenario and for both haploid and diploid populations. The one-population scenario includes two substitutions evolving in a reproductive isolated state from the original population. The two-population scenario includes two populations including two substitutions at two different loci, evolving in a reproductive isolated state. For all spatial modes of speciation the BDM-model can be used. The BDM-model can be used in the scenario of no selection for local adaptation (mutation and random drift only) and in the scenario with selection for local adaptation. In a situation with selection for local adaptation, one can distinguish a scenario with uniform selection and a scenario with heterogeneous selection. Another specific characteristic that can be implemented in the BDM-model is different mutation rates per loci.

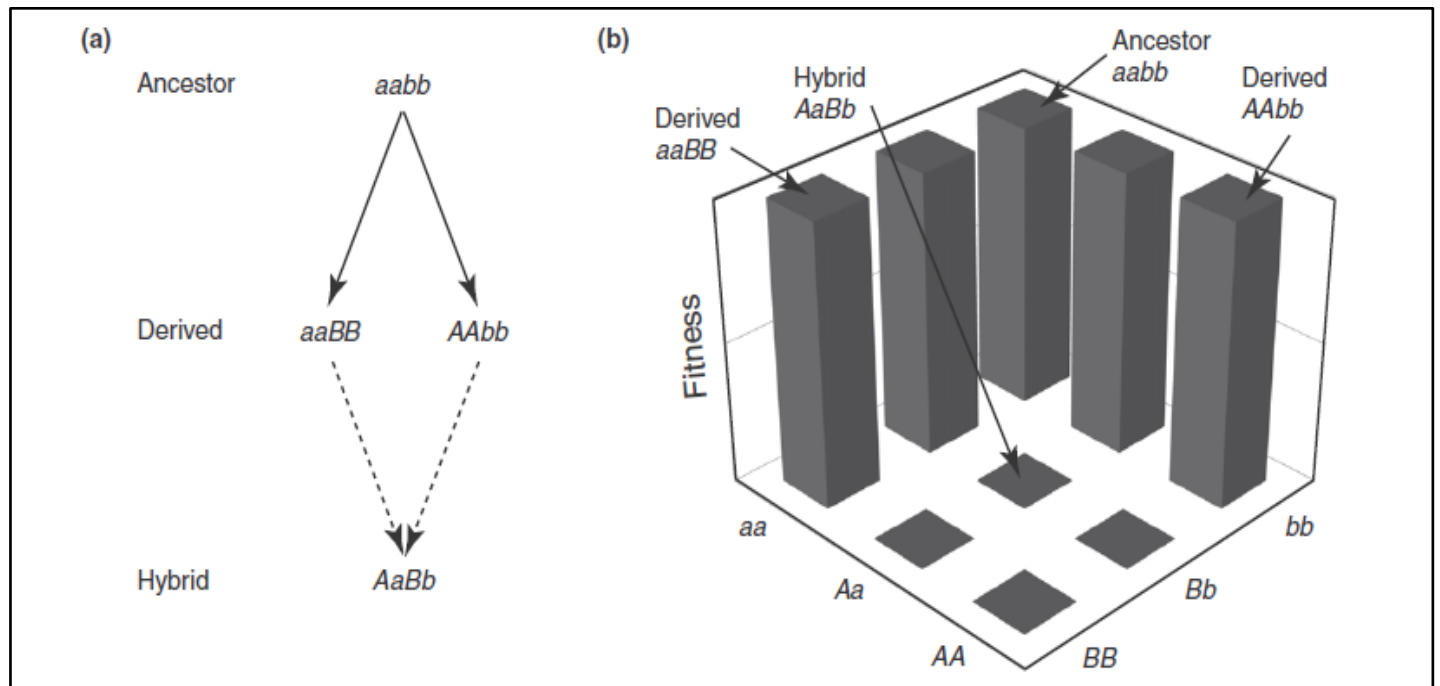


Figure 2: The BDM-model: (a) A schematic representation of a speciation process within a diploid species where loci A and B are incompatible. The ancestral population (fixed for alleles a and b) evolves in two derived populations with fixed allele B in the first population and allele A in the second population. The hybrid descendant of the derived populations faces the consequences of the incompatibility of alleles A and B. (b) A fitness landscape based on the BDM-model with similar genotypes as in the schematic representation. From: Cutter, A. D. 2012. The polymorphic prelude to Bateson–Dobzhansky–Muller incompatibilities. *Trend in Ecology and Evolution* **27**: 209–218.

The evolutionary dynamics of the BDM-model by large population sizes are characterized by a rapid movement towards a ridge of high fitness genotypes with afterwards only slow changes by weak factors. The population dynamics are characterized for both haploid and diploid populations by a short phase with rapid evolution towards one of the limiting values of allele frequencies and a second extended phase with slow evolution along the ridge of high-fitness genotypes. The speciation modelling process with the BDM-model in the chapter five from “Fitness landscapes and the origin of species” uses a “weak mutation-weak migration approximation” neglecting within-population variation. This approximation was made so the only role of mutation and migration in the modelling of allopatric and parapatric speciation was the introduction of new alleles (Gravilets, 2004).

Research question

For both the PBD-model and the BDM-model the speciation characteristics (the probability of speciation, the average duration of speciation and the average waiting time to speciation) are important variables to gain insight in the probability and duration of the speciation process. For my research project I will create phylogenies with the BDM-model. With these new phylogenies from the BDM-model and the existing phylogenies from the PBD-model I will try to make a comparison between the two different models, which has never been done before. What are the differences in phylogenies based on the protracted birth-death model (PBD) and the Bateson-Dobzhansky-Müller model (BDM)? The possible differences and similarities between the phylogenies of both models can give more information about the usability of a mechanistic model including time consuming speciation to explain the decrease of lineages in LTT plots of phylogenies towards

the present. With the phylogenetic trees of the BDM-model the duration of speciation can be examined with a genetic based mechanistic model.

Summary statistics

There are different possibilities to compare phylogenies with the help of summary statistics. Examples of summary statistics are the gamma (γ) summary statistic, the normalized lineages through time (nLTT) summary statistic, tree size and the Phylogenetic Diversity metric (PD) (Janzen *et al.*, 2012).

The gamma summary statistic determines the location of the phylogeny's internal nodes in compare to its tips and root. Gamma gives information if the phylogeny branches in an early period (so called tippy, $\gamma > 0$) or in the end of the timespan (so called stemmy, $\gamma < 0$) in compare with the expected branches from the pure-birth model. The gamma summary statistic is not recommended for time-dependent models, because of the underestimation of the initial speciation rate (Janzen *et al.*, 2012).

The nLTT plot gives the number of lineages in a phylogeny over time. The nLTT summary statistic is used to (Janzen *et al.*, 2012) compare phylogenetic trees. Tree size gives information about the number of tips in a phylogeny. The Phylogenetic Diversity metric (PD) gives the sum of all branch lengths in the tree, divided by the number of branches.

Methods

Parameters

The PBD-model includes five parameters, as can be seen in Table 1. The parameters of the BDM-model, partly described by Gravilets (2004), can be seen in Table 2.

Symbol	Description
μ_i	Extinction rate incipient species
μ_g	Extinction rate good species
b_i	Speciation initiation rate / birth rate incipient species
b_g	Speciation initiation rate / birth rate good species
λ	Speciation completion rate (SCR)

Table 1: Parameter descriptions PBD-model.

Symbol	Description
u	The probability of speciation
μ	Extinction rate
b	Speciation initiation rate / birth rate
m	Migration
T	Waiting time to speciation
t	Average duration of speciation
η	Gene flow factor
K_d	Rate of fixation of deleterious/disadvantageous alleles
K_a	Rate of fixation of advantageous alleles

Table 2: Parameter descriptions BDM-model.

Software

I will use the software program Qt-creator to produce the simulations of phylogenies of the BDM-model.

Discussion

The mechanistic BDM-model I will use can be characterized by a more individual based perspective on speciation models. Another example of a mechanistic individual based speciation model is the ecologic and sexual selection model, which will be discussed by Kees Wesselink in a parallel project. The BDM-model tries to implement the more individual based perspective with the use of a genetic background of genotypes to explain the speciation process. I expect that the BDM-model will give a more realistic view on the speciation process in compare with the PBD-model, because of the genotypic mechanisms which can be found in *Drosophila* and fish species. These examples from species with an inviability of hybrids show the mechanisms of the BDM model in a real biological context. The BDM-model does still include the same idea of a gradual mechanism of speciation as seen in the PBD-model. I will start to create the BDM-model in an allopatric scenario without migration and without advantages or disadvantages for certain alleles. The mutation rates will be the same for different alleles in the model. In this way I start with a relative simple scenario that perhaps can be expanded in the later stages of the project. The implementation of backward mutation is expected to increase the number of species, because extinct ancestor species could reappear in the phylogenies by backward mutation. Other differences in phylogenies of the PBD-model and the BDM-model are not expected.

References

Literature

1. Etienne, R. S. and Rosindell, J. 2012. Prolonging the past counteracts the pull of the present: protracted speciation can explain observed slowdowns in diversification. *Systematic Biology* **61**: 204-213.
2. Bateson, W. 1909. Heredity and variation in modern lights. *Darwin and modern science*, Cambridge University Press: 85-101.
3. Dobzhansky, T. G. 1935. Studies on hybrid sterility. II. Localization of sterility factors in *drosophila pseudoobscura* hybrids. *Genetics* **21**: 113-135.
4. Müller (1940, 1942)
5. Bilderbeek, R. J. C. and Etienne, R. S. Should protracted speciation be incorporated in phylogenetic tree construction methods? *Not yet published*.
6. Gavrillets, S. 2004. Fitness landscapes and the origin of species. *Princeton University Press*.
7. Orr, H. A. and Presgraves D. C. 2000. Speciation by postzygotic isolation:

forces, genes and molecules. *Bioessays* **22**: 1085-1094.

8. Janzen, T. *et al.* 2015. Approximate Bayesian Computation of diversification rates from molecular phylogenies: introducing a new efficient summary statistic, the nLTT. *Methods in ecology and evolution* **6**, **5**: 566-575.

9. Cutter, A. D. 2012. The polymorphic prelude to Bateson–Dobzhansky–Muller incompatibilities. *Trend in Ecology and Evolution* **27**: 209-218.

Figures

1. Bilderbeek, R. J. C. and Etienne, R. S. Should protracted speciation be incorporated in phylogenetic tree construction methods? Not yet published.

2. Cutter, A. D. 2012. The polymorphic prelude to Bateson–Dobzhansky–Muller incompatibilities. *Trend in Ecology and Evolution* **27**: 209-218.