Studentnumber: 2176661   Rijksuniversiteit Groningen
Master Ecology and Evolution   Theoretical Research in Evolutionary Life Sciences
Supervisor: Richèl Bilderbeek

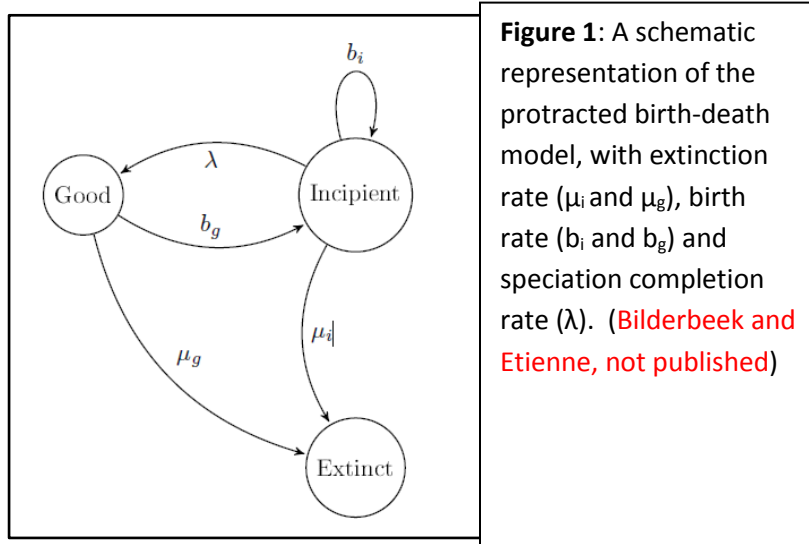# The differences in phylogenies of the BDM-model and the PBD-model
Jorik de Boer, September 2016

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? To gain more knowledge about the speciation processes and possible temporal patterns of diversification multiple models have been created to simulate phylogenies of species. For my research project I will work with two different models: the protracted birth-death model by Etienne and Rossendell (2012) and the Bateson-Dobzhansky-Müller model by Bateson (1909), Dobzhanksy (1936, 1937) and Müller (1940, 1942). First I will discuss the principles and background of both models. Thereafter I will explain the main question of the research project, my hypotheses and the methods I will use. In the end I will discuss the assumptions and possible weaknesses that could play a role in the research process.

PBD-Model

The protracted birth-death model (PBD, also called protracted speciation model) tries to explain the slowdown towards the present, as seen in phylogenetic trees. The PBD-model is based on the standard birth-death model (Etienne and Rossendell, 2012). The standard birth-death model allows for extinction and birth of species. The standard birth-death model shows in a plot of the number of lineages versus time (lineages through-time (LTT) plot) 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". In LTT plots of phylogenies a decrease of lineages is visible, in contrast to the standard birth-death model simulations. The protracted birth-death model tried to implement the concept that the speciation process takes time in the standard birth-death model. The model divided the speciation process in an incipient species stage and a good species stage (Figure 1). New incipient species can arise from both good species and incipient species. The protracted birth-death model can predict more imbalanced phylogenies in compare with the standard birth-death model. In the specific situation that the speciation completion rate (SCR) in the protracted speciation model is limitless ($\lambda = \infty$), the model is the same as the birth-death model with comparable parameters (extinction rates and birth rates).

Figure 1: A schematic representation of the protracted birth-death model, with extinction rate (μ_i and μ_g), birth rate (b_i and b_g) and speciation completion rate (λ). (Bilderbeek and Etienne, not published)

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 (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). Examples of incompatibilities of genotypes such as in the BDM model have been found in nature. Examples of genes involved in incompatibilities, like sterility of hybrids, can be found in fish species, fruit flies *D. simulan*s and *D. melanogaster*. The BDM-model can explain ridges of high-fitness genotypes, a so called "holey fitness landscape", as seen in Figure 2b. The model can evolve to strong reproductive isolation (RI) without crossing "fitness valleys". The model can be used with selection against individuals or with selection against pairs of genotypes.

The BDM-model can be used in a one-population scenario and a two-population scenario and for haploid and diploid populations. The one-population scenario includes two substitutions evolving in a reproductive isolated (RI) state from the original population. The two-population scenario includes two populations including two substitutions at two different loci, evolving in a reproductive isolated (RI) state. For both allopatric and parapatric (or sympatric) speciation the BDM model can be used. The model makes a distinction between selection for local adaptation (mutation and random drift only) and selection for local adaptation with spatially uniform selection or spatially heterogeneous selection. Other specific characteristics that can be implemented are backward mutations and different mutation rates per loci.
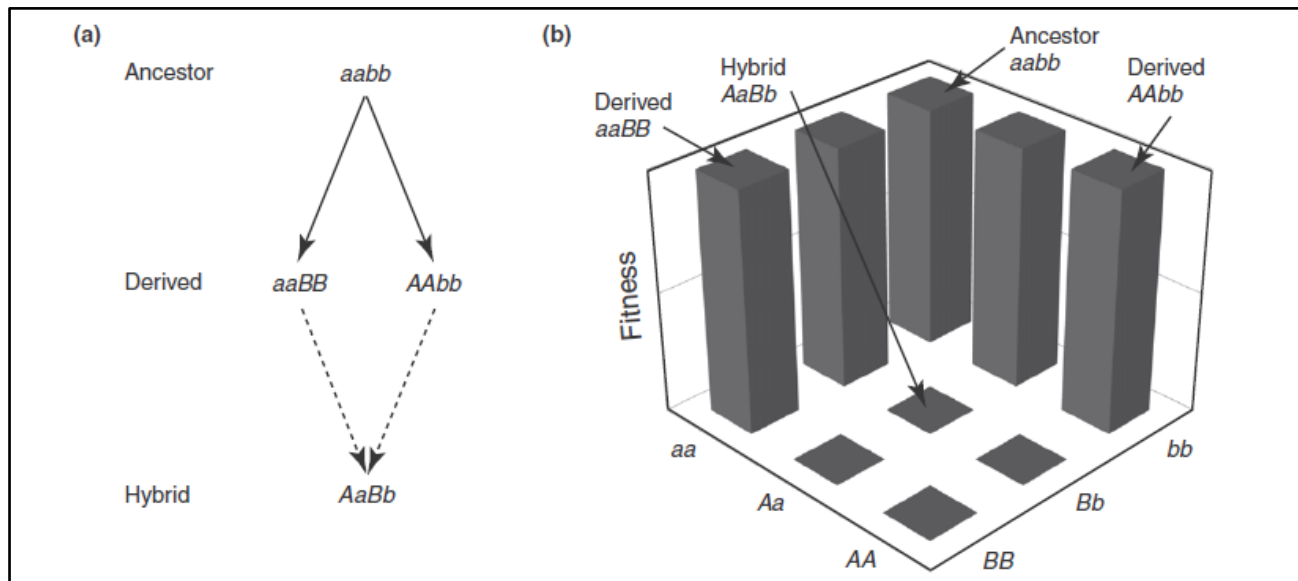

Figure 2: The BDM model: (a) A schematic representation of a speciation process within a diploid species where loci A and B are incompatible. The ancestral population (fixed for alleles a and b) evolves in two derived population with fixed allele B in the first population and allele A in the second population. The 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. (Cutter, 2012)

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.

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 try to 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 is not been done before. What are the differences in phylogenies based on the protracted birth-death model (PBD) and the Bateson-Dobzhansky-Müller model (BDM)?

Summary statistics
There are different possibilities to compare phylogenies with the help of summary statistics. Examples of summary statistics are the Gamma summary statistic, the normalized lineages through time (nLTT) summary statistic, tree size and the Phylogenetic Diversity metric (PD). The Gamma summary statistic determines the location of the phylogeny's internal nodes in compare to its tips and root. The Gamma value 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 statistics is strongly influenced by the mutation rate. The speciation completion rate (SCR) appears to have less impact on the Gamma summary statistics. The nLTT summary statistic gives the number of lineages in a phylogeny over time. The nLTT summary statistic is according to (Janzen *et al.*, 2012) more suitable for comparing phylogenetic trees than characterize a single phylogenetic tree. 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, as described by Gravilets (2004), can be seen in Table 2. By modeling the allopatric and parapatric speciation Gravilets (2004) used the weak mutation-weak migration approximation neglecting within-population variation.

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
m	Migration
μ	Extinction
T	Waiting time to speciation
t	Average duration of speciation
η	Gene flow factor
K_d	Rate of fixation of deleterious alleles
K_a	Rate of fixation of advantageous alleles

Table 2: Parameter descriptions BDM model.

Software
The software programs and websites I will use to produce the simulations of phylogenies of the BDM model include the application software Qt writer version 5.7.0, GitHub (http://github.com/JorikdeBoer), Travis CI and CodeCov.

Discussion
Hypotheses
Aannames & simplificaties
More individual based
Trying to create bridge between mechanistic models like BDM model and PBD model
Link with slowdown in the rate of increase of number of lineages in the phylogeny
Mechanistic model v/ Kees

References

Literature

1 Etienne and Rossendell (2012)
2 Bateson (1909)
3 Dobzhanksy (1936,1937)
4 Müller (1940, 1942)
5 Bilderbeek and Etienne, not published
6 Janzen *et al.*, 2012
7
8
9

Figures

Bilderbeek and Etienne, not published
Cutter, 2012