# Introduction

Richèl J.C. Bilderbeek[1]

[1]Groningen Institute for Evolutionary Life Sciences, University of Groningen, Groningen, The Netherlands

November 12, 2019

# 1

# INTRODUCTION

Once upon a time, there was the first living organism, the First Universal Common Ancestor (FUCA). We do not know when it lived.

One unknown day, FUCA speciated, resulting in two species. This event doubled the biodiversity on Earth. The two species, which we will call species A and B are sister species. We do not know what caused the speciation.

Both species A had their unknown histories: they speciated themselves, and they and/or their descendants went extinct. Extinction is a common event. Let's assume A and/or its clade went exctint and that species B created a sister species C. Species B and C will give rise to all biodiversity. This ancestor of species B and C is called the Last Universal Common Ancestor and lived around [then].

The biodiversity derived from LUCA is important to us humans, apart from that is has created us. A review paper (Cardinale et al.) shows that biodiversity unsually improves ecosystem services that are positive for human well-being.

Speciation is the process that increases biological diversity and is therefore worth to study. We can study the mechanism ('what causes a speciation event?') or we can study the patterns of many of such events ('is speciation rate constant through time?')

The mechanism of a speciation event has many facets. For more than half a century ago, it was hypothesized that speciation is caused by geographical isolation (e.g. Mayr, 1942) or due to ecological factors (e.g. Lack, 1947). [other proximate causes of speciation]

Instead of look at the mechanism behind each speciation event, we can also look at patterns of speciation events through evolutionary time. How do the rates at which speciation and extinction take place through time change? Is there an upper limit on the number of species? [ADD FOSSILS TO ]

The field of phylogenetics is the research discipline that infers phylogenies from heritable data (DNA, RNA, protein) and optionally adds morphological data and/or fossils. It strives to produce the most accurate phylogenies, regarding topology and speciation times.

[Example with viruses?]

Phylogenetics has taken a huge flight, due to the massively increased computational power and techniques. Milestones are Felsensteyn's work in 1980, PHYLIP, the first software package for phylogenetic analysis (and is still in development!). The Metropolis-Hasting algorithm allowed for efficient MCMC sampling, allowing Bayesian phylogenetics to thrive. Contemporay Bayesian phylogenetics tools are BEAST, BEAST2, MrBayes and RevBayes.

A clear example of what phylogenetics can do nowadays, is the Tree Of Life, which uses 3,083 genomes of 2,596 amino-acid positions to create one big phylogeny of all (sequenced) life on Earth, which took 3,840 computational hours on a modern supercomputer [Hug et al., 2016].

To create the Tree Of Life from a protein alignment, there have been many assumptions made, regarding the evolution of a protein sequence, the rate(s) at which this happens and the rate(s) at which a branching/speciation event takes place. For example, the amino acids are assumed to follow the LG [Si Quang Le and Olivier Gascuel, An Improved General Amino Acid Replacement Matrix] which uses empirical transition rates.

For speciation, the most interesting part of the phylogenetic assumptions is the tree prior. A tree prior is the Bayesian form of the knowledge about the speciation process.

[BE MORE EXPLICIT HERE] [THE IMPORTANCE OF PHYLOGENTIC TREES]

The most basic speciation model is the Yule model [Yule, 19..] which assumes that speciation is constant and there is no extinction. [Research on fossils with Yule model would be fun]. The Yule model predict that the number of extant species grows exponentially through time.

The Birth-Death model [Nee et al., 1994] is an extension of the Yule that allows for a constant extinction rate. If the speciation rate exceeds the extinction rate, also the BD model predicts that the number of extant species grows exponentially through time. If the extinction rate exceeds the speciation rate, the number of lineages is expected to decline exponentially. The latter is biologically irrelevant.

It is clear that an exponential growth in the expected number of lineages is biologically nonsense. To state the obvious: a finite area (Earth) results in a finite number of species. Applying the BD model to molecular data already shows that it does not always hold, see figure [below]

One extension of the BD model that offers a biological explanation is the diversity-dependent model [Etienne and Haegeman, ?2012] in which the speciation rate decreases up until a certain carrying capacity is reached. This carrying capacity amounts to the number of niches in an environment.

The are multiple other extensions of the BD model, filling in different biological aspects that are (purposefully) lacking. The time-dependent BD model [?Lovelocke], for example, assumes that speciation and extinction rates are time-dependent. The explanation for this would be that speciation is mostly driven by abiotic factors, like for example, temperature, which is known to fluctuate.

When constructing a phylogeny from Lake Tangyanika [or some other adaptive radiation place], we can find evidence that one clade has a sudden higher speciation rate than another. One explanation for this, is to assume that a certain key innovation (e.g. jaw/opsin) causes a change in speciation rate. This idea is incorporated by the trait-dependent BD model [Maddison].

Another facet of speciation is that speciation takes time. It takes time for the build-up of reproductive isolations and it takes time for us humans to recognize the two-not-one species. A biological example is from [Fennesy, 201?] in which some new giraffe species have been discovered by sequencing part of their DNA. Although these new species have been 'discovered' recently, the had been no gene flow between species for already two million years.

Using the BD model in species that are slow to speciate, will cause an underestimation of the number of lineages in the present (as in the giraffes), in effect possibly giving the illusion that speciation slows down, where in reality it does not. Note that also the time-dependent and diversity-dependent speciation models also offer an explanation of this knowledge.

The speciation model that encorporates the fact that speciation takes time is the PBD model [Etienne and Rosindell, 201?]. This extension of the BD model adds that each species has a state: a 'good' species is a species that is recognized as such, where an 'incipient' species is not yet recognized yet. It takes time for an incipient species to become a good species. This time is called the speciation time [or those other times Rampal uses].

Another facet of speciation uncaptured by the BD model is the effect of a geographical isolation. When a habitat (lake or mountain range) gets separated, this will have an effect on both isolated communities. One can imagine that this triggers a speciation event in multiple species of both communities at the same time. This is posed as one alternative explanation for the high speciation rate in lake Tangyanika, where the water level rises and falls with ice ages, triggering co-occurring speciation events each change.

Where the BD model allows for exactly one speciation model at one timepoint, the Multiple-Birth Death (MBD) model allows for co-occurring speciation events [Laudanno, 201?]. For systems that are likely to have such co-occurence, we can improve our inference.

[WHAT DO PEOPLE WANT TO KNOW]

## 1.1. PHOTO ATTRIBUTION

Figure **??**, the left image, Preserved specimen of chalumnae by Alberto Fernandez Fernandez is licensed under CC BY-SA 3.0. the image at the right, Akha couple in northern Thailand by Weltenbummler84 is licensed under CC BY-SA 2.0 DE.

For figure **??**, the selection of these eight images was done by OneZoom. These images, from top-left to bottom-row, row-first: Nile Crocodile by Marco Schmidt is licensed under CC-BY-SA 3.0. American Alligator by NASA Kennedy. False Gavial by Yinan Chen is marked as being in the public domain. Common Caiman by Michael Wolf is licensed under CC-BY 2.5. African Dwarf Crocodile by Staycoolandbegood is marked as being in the public domain. Smooth-fronted Caiman by Whaldener Endo is licensed under CC-BY-SA 4.0 Phillipines Crocodile by Vanderploeg is marked as being in the public domain. Board-snouted Crocodile by Hadonos is Marked as being in the public domain.