**Differences in phylogenetic trees of penguin taxa by FBD model and clock models**

**Abstract**

* Coming soon -

**Introduction**

In this report, The time-calibrated trees of penguin taxa were estimated under the FBD model, strict clock model, and relaxed clock model. The data used is a morphological matrix associated with age information. The phylogenetic data derived from osteological characters consisting of 14 species of living genera, five fossil species, and 3 outgroups. The fossilized birth-death (FBD) model describes the stochastic processes of speciation and extinction rate. It is also used to describe the distribution of fossils and the sampling of fossil and extant species (Gavryushkina and Zhang, 2020). The idea of a molecular clock has been essential to understanding evolution for almost half a century (Zuckerkandl and Pauling, 1962). In this study, we chose two molecular clock models which are a strict clock model and a relaxed clock model. A strict clock model assumes that the evolutionary rate remains constant for every branch in a phylogenetic tree (Ferreira and Suchard, 2008). Meanwhile, a relaxed clock model assumes that each branch of a phylogenetic tree has its own evolutionary rate (Drummond et al., 2006). Since these models have the same purpose but different approaches, I proposed to see different results in penguins' time trees for different models.

**Method**

**Data**

**Morphological data**

The morphological data consist of 22 species including 14 representative species of living genera, 5 fossil species of Spheniscidae, and three outgroups. The data was derived by using 44 morphological characters. *Diomedea exulans*, *Fregata magnificens*, and *Gavia immer* were used as outgroups. According to the hypothesis proposed by Van Tuinen et al., 2001 and Mayr, 2004, *Gavia immer* was used as a root species (Hospitaleche, 2007).

**Age information**

* Coming soon -

**Analysis**

**Tree model**

The tree model uses to describe how trees grow over time. For a tree model, the FBD model was applied. There are three parameters which are speciation rate, extinction rate, and fossil sampling rate along with fossil age were used for the model. These variables were assumed as a constant rate over time and exponential prior were applied for each parameter.

These parameters are stochastic variables derived from exponential distributions with rates. We set each parameter for exponential distribution with *δ* = 10. Following, the moves were specified by setting lambda = 0.5, Tune = true for that MCMC looks for the optimum size for the move during the initial step. Subsequently, rho probability which is a constant variable was specified as 1 since all samples are extant penguins.

Next step, the origin time parameters were specified by using the age of the oldest penguin fossil as a minimum (20.0) and the age of the outgroup fossil that was used as a root species (35.0). Eventually, the moves on the tree topology and branch lengths and model were specified. The clade constraints were created to monitor specific clades as I want to focus on extant species.

The deterministic variables used in this model are turnover rate, diversification rate, and number of ancestors' samples. Due to the uncertainty of the fossil age, I used a loop to go through each dataset. To know the interval of uncertainty, the uniform distribution between minimum and maximum bounds was assigned.

**Clock models**

The clock models use to describe how evolutionary rates vary across the tree. All parameters and substitution models were set as the same since we want to focus on the results of using different clock models which are a strict clock model and a relaxed clock model.

1. **The strict clock model**

The strict clock model uses to describe the rate of character change over time by assuming that the evolutionary rate remains constant for all branches in our tree. Furthermore, it is also assumed that all lineages share the same rate. Since the rate of evolution remains unknown, an exponential prior was used for a strict clock model.

1. **The relaxed clock model**

The relaxed clock model uses to describe the rate of character change over time by assuming that the evolutionary rate varies for each branch and is drawn independently from the underlying distribution. The uncorrelated lognormal model was applied since it is simple and widely used. The relaxed clock model assumes each branch has an independent rate drawing from a lognormal distribution. In this model, the stochastic variables which are the mean and the variance of the lognormal distribution are defined by the estimation of the data.

**Substitution models**

The substitution model uses to describe how sites evolve over time. The Mk substitution model is a generalization of the Jukes-Cantor model which assumes that the rate of change between character states is equal for all branches as well as the state frequencies. For the substitution model used, the Mk substitution model was used to describe the probability of transitioning from one character state to another.

The fnJC function was used to define a Q matrix in which our data contain 3 character states of the morphological character matrix. Furthermore, the stochastic node was defined as standard and clamped with morphological data.

**Convergence**

**ESS values**

Posterior

Likelihood

Prior

Diversification rate

Extinction rate

Speciation rate

**Histogram**

A group of blue and green lines

Description automatically generated with medium confidence

Figure 1. The distributions of each model. a) FBD model, b) Strict clock model, c) Relaxed clock model