# A Robust Estimation of the Relationship between Size and Trophic Level in Ray-Finned Fish

**Rojan Karakaya**

Supervisor : Krzysztof Bartoszek
Examiner : Josef Wilzén

External supervisor : John Clarke

**Abstract**

There is an emerging area of research into the drivers of fish size evolution. One prong of this is the study of how food web position and body size co-evolves.

This study attempts to add to previous research on this topic by investigating the matter in a way that is robust to phylogenetic covariance in the observed data. Previous research, which either avoids the issue of phylogenetic relatedness or controls for it only in rudimentary ways, has had varying non-conclusive results for body length and mass, but fairly strong results for a positive relationship between jaw size and position on the food web.

The results from this study show that the food web position has a stronger connection to the logarithmic transformation of the maxillary length than it does the logarithmic transformation of the body length. The results vary over different branches of the phylogenetic tree, however, as the best-fitting model is an an eclectic model of multiple regimes each adapted to their own branch of the phylogeny.

Another possible source of residual covariance arises from erroneous estimations of trophic level, a measure of how many steps removed a species's diet is from consumption of primary producers, is explored and a secondary investigation sheds some light on how to counteract this problem in some limited circumstances. A solution for a very limited form of this problem is demonstrated, but no general solution is found. Ways of avoiding the influence of this unpredictable issue in future work is discussed.

# Acknowledgments

# Contents

# List of Figures

# List of Tables

# Glossary

| | |
|---|---|
| **size** | In this text, the word is used synonymously with "length", except for when used in the term "effect size" |
| **trait** | An inherited value of a biological species |
| **allometry** | The relationship between the size of species and some other trait (morphological or otherwise) |
| **phylogeny** | A description of the relatedness of different species |
| **dendrogram** | A tree chart, used here as a way to illustrate phylogeny |
| **ultrametricity** | Ulrametric trees are phylogenetic trees where all tips (i.e. species) are equally distant from the root |
| **actinopterygia** | Ray-finned fishes; a large and diverse family of bony fish |
| **maxilla** | Upper jaw |
| **lineage** | In this work, we denote as "lineage" a continuous line through a phylogenetic tree to a single tip, and which does not cross the same point in time twice. This is identical to Simpson's definition if we assume that species can only descend from single ancestor species [26] |
| **food web** | A directed graph where species is a vertex, and edges between vertices indicate a consumer relationship |
| **cartilaginous fish** | Fish that are not bony, e.g. sharks |
| **species** | A term the exact meaning of which is debated among biologists, with no real consensus; for the purposes of this work, it is defined as in 2 |
| **specific** | In this text the word can be used either in its vernacular sense, or to mean "relating to species" (e.g. "inter- and intraspecific variance") |
| **biomagnification** | If a chemical compound in an environment is absorbed by creatures in that environment and cannot be excreted or broken down by biological processes within those creatures, then the concentrations of that compound will accumulate in species that are higher in the food web. This process is called biomagnification |
| **evolutionary regime** | See 3.4. Note that we will sometimes use "regime $n$" interchangeably with "the sub-model for regime $n$" as a shorthand |
| **caudal fin** | The tail fin of a fish |

# Chapter 1
# Introduction

## 1.1   Motivation

Fish populations around the world are under pressure from multiple sources of ecosystem stress, relating to e.g. overfishing or global warming and other environmental changes. Overfishing tends to affect larger size predator fish to a greater degree than it does smaller size fish, although the relative abundance of predator to prey fish might mediate this effect [22]. Global warming directly affects the extent and temperature of many inland lakes (and this effect will affect different lakes differently depending on their characteristics [28], and it affects the circulation of nutrients and the acidification of oceans [9]. These effects have a differential impact on fish depending on their size and food web locus.

   Because of these stressors, it is reasonable to believe that a great number of ecosystems and food webs around the world will change drastically in the near future, and there is good reason to investigate if and how these changes might impact the species constituting those ecosystems in the long term. There is good research [27] to show that the sizes of commercial fish species may have shrunk over the recent decades as an evolutionary adaptation to size regulations within the fishing industry.

   Food web structure, and the positions of species within food webs, is broadly understood to be an important influence on the evolution of species, with e.g. dentation and gastric structures differing in near-predictable ways between carnivorous and herbivorous species of animals. While there has long been a presumed positive relationship between size and trophic level for fish, this relationship is not obvious.

### Why would we expect a relationship between size and their position in the food web?

The common feeding mode of most carnivorous fish require that their jaws be either broad enough or distensible enough to fit the entirety of their prey (or at least some axis of their prey) in their mouths, which implies that there should be a clear positive relationship between size (especially jaw size) and trophic level. On the other hand, the supply of energy and biomass tends to be greater in species that are closer in the food web to primary producers. Species that are distant from primary producers therefore would consequently constitute a smaller amount of total biomass, which combined with demands for genetic diversity in breeding groups would imply smaller individuals.

   Theoretical arguments thus point in both the positive and negative direction, and a proper empirical investigation is necessary.

### How can we operationalize the idea of a "position in a food web"?

Apart from the broader trophic guild classifications such as "producer", "primary consumer" etc., it is not a straightforward task to compare the food web loci of two species in different food webs. A well-utilized measure in studies of biomagnification is that of the trophic level,

a weighted average of the trophic level of the food items of a species plus one (with plants and detritus having a trophic level of one by definition). While this measure is comparable over different food webs, and at least in theory easy to calculate, it does suffer from the fact that mismeasurements in consumed food item proportions for even one species are liable to propagate to species that directly or indirectly feed off of it, due to the concept's recursive nature. See Chapter 4 for an extended discussion on this issue.

**Statistical research with biological species as points of data**

A recurring assumption in statistical methods is that of independence of observations. For an investigation into the traits of various living species this assumption can not be assumed to hold other than in certain edge cases, as most traits are at least somewhat inherited, and closely related species will therefore be expected to be more similar to one another than distant relatives are. Furthermore, as there are no absolute measures of similarity, the use of the term "similarity" will depend on the specific set of species that are under investigation, analogously to the relativeness of the term "closely related".

The seminal paper dealing with this issue was written by Felsenstein [10], who proposed a solution based on a regression on the differences between pairs of closely related species (since these differences are independent if we assume that each species' incremental evolution is independent of that of other species). The specific algorithm suggested, however, implicitly assumed that the traits evolve as a Brownian motion process. This method was refined into the Phylogenetic Generalized Least Squares model [12] and the Phylogenetic Generalized Linear Mixed Model [14], both of which are designed to (within the frame of their assumptions) find linear relations for the observed data that are corrected for phylogenetic covariance.

Mitov, Bartoszek and Stadler (2019) [21] and its predecessors (e.g. [2] and [6]) model the evolutionary process as a stochastic process. This allows for the testing of a far greater number of hypotheses, such as the long-term evolutionary trajectory of young phylogenetic groups. One of the specific contributions of [21] is the ability to fit mixed models, e.g. ones where the stochastic process of different branches can differ if the difference improves some information criterion.

**Previous Research**

Keppeler, Montaña and Winemiller [16] use two methods, one a Bayesian phylogenetic linear mixed model over interspecific data with a rudimentary control for phylogeny, and the second a Bayesian generalized linear mixed model based on intraspecific data (as a means of getting around issues of phylogenetic covariance). The interspecific investigation showed no significant relationship for body size and trophic level, whereas the intraspecific investigation showed a positive relationship only within some of the investigated species. While this second intraspecific investigation is interesting, it does not really indicate anything about the evolutionary relationship between the two traits.

Keller Kopf et al. [17] use a conditional inference random forest model and a Bayesian linear mixed model to predict the trophic level of non-marine fish based on various morphological traits. They found jaw length (specifically the ratio of maxillary length over head depth) to have the largest positive effect size in the Bayesian model and the greatest influence in the random forest model. This investigation did not include body length as a factor, but did have maximum body mass instead; this variable was not found to have a significant effect in either model, even when mediated by type of aquatic environment. This study did not correct for phylogenetic relatedness in either model.

Romanuk, Hayward and Hutchings [25] use an ANCOVA over different orders of fish (including cartilaginous fish). This study implements a rudimentary control for phylogeny by aggregating data by phylogenetic rank, and then investigating those aggregations. It finds

that in some orders, particularly those whose constituent species are small in size, body length is positively related to trophic level. In none of the investigated orders was trophic level significantly negatively related to body length.

## 1.2  Aim

The primary aim of this thesis is to investigate the relationship between body and jaw size (operationalized here as body length and jaw length) and trophic level in actinopterygian fish, in a manner that is robust to phylogenetic covariance.

A secondary aim for this thesis is to investigate the robustness of the concept of trophic level.

## 1.3  Research questions

1. How does trophic level vary evolutionarily with body and jaw length?

2. How robust are estimates of trophic level to mismeasurements?

## 1.4  Delimitations

This study focuses on actinopterygian fish[1] that are not solely marine (i.e. that spend at least some of their life in freshwater aquatic environments) and whose mouths are not sucker-mouths (see 2.2).

---

[1]This limitation is mostly for reasons of data availability. Actinopterygia is a large group, however, and comprises the vast majority of fish species.

# Chapter 2
# Data

For this investigation, I collated species trait data from FishBase [11] and FISHMORPH [4] with a timed phylogenetic relatedness graph from Fish Tree of Life [7] (for a minimalist example of such a graph as well as its Newick encoding, see Figure 2.1).
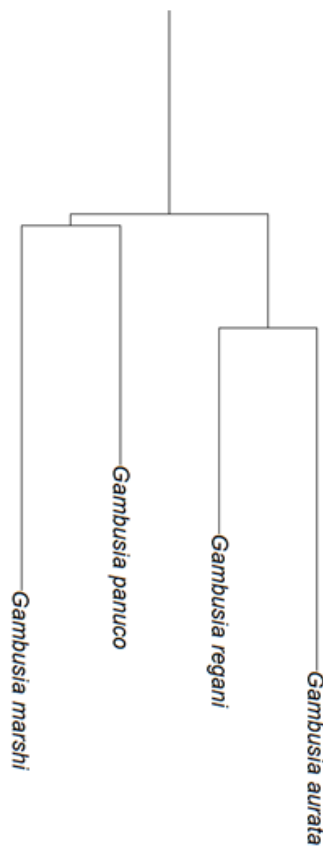


Figure 2.1: Example of a phylogenetic tree

The corresponding Newick format encoding of this tree is ((Gambusia marshi:32, Gambusia panuco:21):1, (Gambusia regani:18, Gambusia aurata:30):10):19

The data in FishBase has been collated from a great number of papers, all dealing with either currently living species of fish or species that have died out at some point since research on them had last been conducted. In practice this means that all the fish species in the data set have been extant at at least some point in the last fifty years. Because the source papers vary widely in research intent and methodology, the data can be presumed to be fairly noisy. Most importantly for our investigation, maximum body length is provided as standard lengths for

some species and as total length for others (see Figure 2.2 for a visual presentation of the difference between these two measurement types). For the purposes of this investigation I have proceeded without taking this difference into account. This can be justified either on the basis that the choice to use standard or total length was an independent random bernoulli trial for each species, or that the expected difference between the two measures is 0 (neither of these justifications are correct, however - see Section 6.2 for a discussion about this).

The morphometric data in FISHMORPH has been collected from images (photographs and scientific drawings) of freshwater fish from the scientific literature. As these images do not, as a rule, contain absolute measurements, the morphometries are denoted as ratios of one another (e.g. the BEl variable in Table 2.1, which is the ratio between body length and depth). The species delineations used in FISHMORPH are identical to those of FishBase.

Fish Tree of Life has phylogenies for actinopterygian fish calculated from genetic sequence comparisons. The species in these phylogenies are denoted with scientific names. In many cases these names overlap with the scientific names in FishBase and FISHMORPH, but in other cases they do not. To what extent this is because of different naming conventions as opposed to different species definitions is not clear. In this investigation I have taken as agnostic of a view on this issue as I have been able to by only including those species that have the same scientific names in both FishBase and Fish Tree of Life. An exact description of the matching algorithm can be found in Section 2.2.

For discussions and differing definitions of the species concept, see e.g. [13], [26] or [18].

Since all of the fish species under study have been alive within the last few decades, we will treat the phylogenetic tree as if it is an ultrametric tree.

In the final dataset used for this thesis there were 1551 observations of fish species with three variables each: trophic level, log maximum jaw length and log maximum body length. The logarithmic transformation has been implemented on both the maximum body length and jaw length variables, as the raw distributions of the length measurements are log-normal in nature (see Figure 2.3 for histograms of the distributions, and Figure 2.4 for a scatterplot that illustrates the relationships among the transformed variables).

## 2.1 Raw Data

From FishBase, I collected data on species' maximum body length (Length) from the Species table, and two types of trophic level estimates (FoodTroph and DietTroph) from the Ecology table. While there is a third type of estimate as well, it is an imputation based on species' body size and would thus be inappropriate for this investigation. This collection was done through the R package `rfishbase` [3].

From FISHMORPH, I collected the variables MBl, BEl, RMl and BLs, which are described in Table 2.1.

| Variable name | Description |
|---|---|
| MBl | Maximum body length in centimetres according to FishBase |
| BEl | Body elongation measured as the ratio between body length and body depth |
| RMl | Relative maxillary length measured as the ratio between maxillary jaw length and head depth |
| BLs | Body lateral shape measured as the ratio between head depth and body depth |

Table 2.1: Variables extracted from FISHMORPH

From Fish Tree of Life, I gathered the timed phylogenetic relatedness graph "actinopt_12k_treePL", which is in the Newick format and contains a phylogeny for actinopterygian fish calculated based on genetic evidence [7].

## 2.2 Secondary Data

With the above-described FISHMORPH variables, I calculated the jaw length (or more accurately: the maxillary length) for each species as

$$JawLength = RMI \cdot \frac{BLs}{BEI} \cdot MBI$$

17 species had jaw length values of 0; a closer investigation into these showed that they are armored catfish with ventrally oriented suckermouths, and how to define their upper jaw seems to be conceptually unclear. In order to avoid any confusion, these species were excluded entirely.
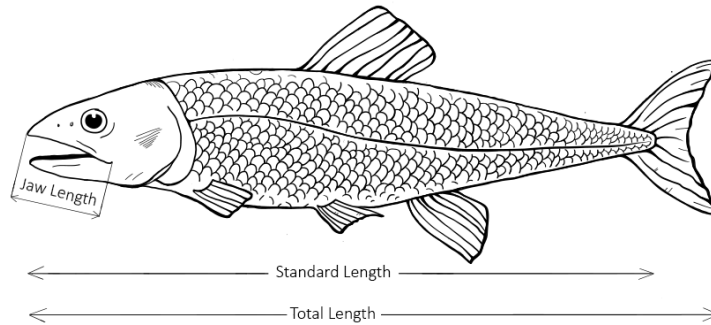
Figure 2.2: Simplified fish body diagram illustrating the three relevant measurements

Note that Jaw Length measures the maxilla and that the length of the caudal fin is what constitutes the difference between the total Length and standard Length

### Selection of Trophic Level Variables from FishBase

For the trophic level, I chose to use DietTroph wherever possible, and where it was missing but a FoodTroph estimate was available, I used that instead. The reason for this is that DietTroph is estimated directly from food proportions, whereas FoodTroph is based on a simulated estimate of the food proportions from food item rankings [23].

Out of FishBase's 34721 species, 11986 have a trophic level from one of the two subtypes described above. Species with imputed trophic levels or without any estimate were removed from the dataset.

As can be seen in Figure 2.3, the distribution of trophic levels is not normally distributed, nor is it entirely symmetric. 2 is the lowest value possible for the trophic level of an animal, and we see quite a large concentration of species just by that lower limit.

### Matching Species from Fish Tree of Life with those from FishBase

Finding the correspondence between the FishBase species IDs (called SpecCode) and the FISHMORPH IDs was trivial, as FISHMORPH uses the same species codes as FishBase. Correspondence between the FishBase SpecCode and Fish Tree of Life species names was established in the following way:

1. From the FishBase table Species the variables SpecCode, Genus and Species were extracted and a new temporary table was set up with the variables SpecCode and Genus_Species, with the latter consisting of the Genus and Species strings concatenated with an underscore between, to match the format of the species names in Fish Tree of Life

2. R:s "match"-function was run over the two lists of names to find matching pairs

3. Names in Fish Tree of Life that had no correspondents in FishBase had their tips removed from the phylogenetic tree

4. The remaining tips were matched with their corresponding species names' SpecCode values in FishBase

The number of species that both have available trophic level data from FishBase and are matchable to species on the Fish Tree of Life tree is 6079. After the removal from this group of the species without FISHMORPH-data, exactly 1568 species remained. From these the above mentioned 17 species with ventrally oriented suckermouths were removed, for a final total of 1551 species. Their last common ancestor population is estimated by Fish Tree of Life to have lived 369 million years ago.
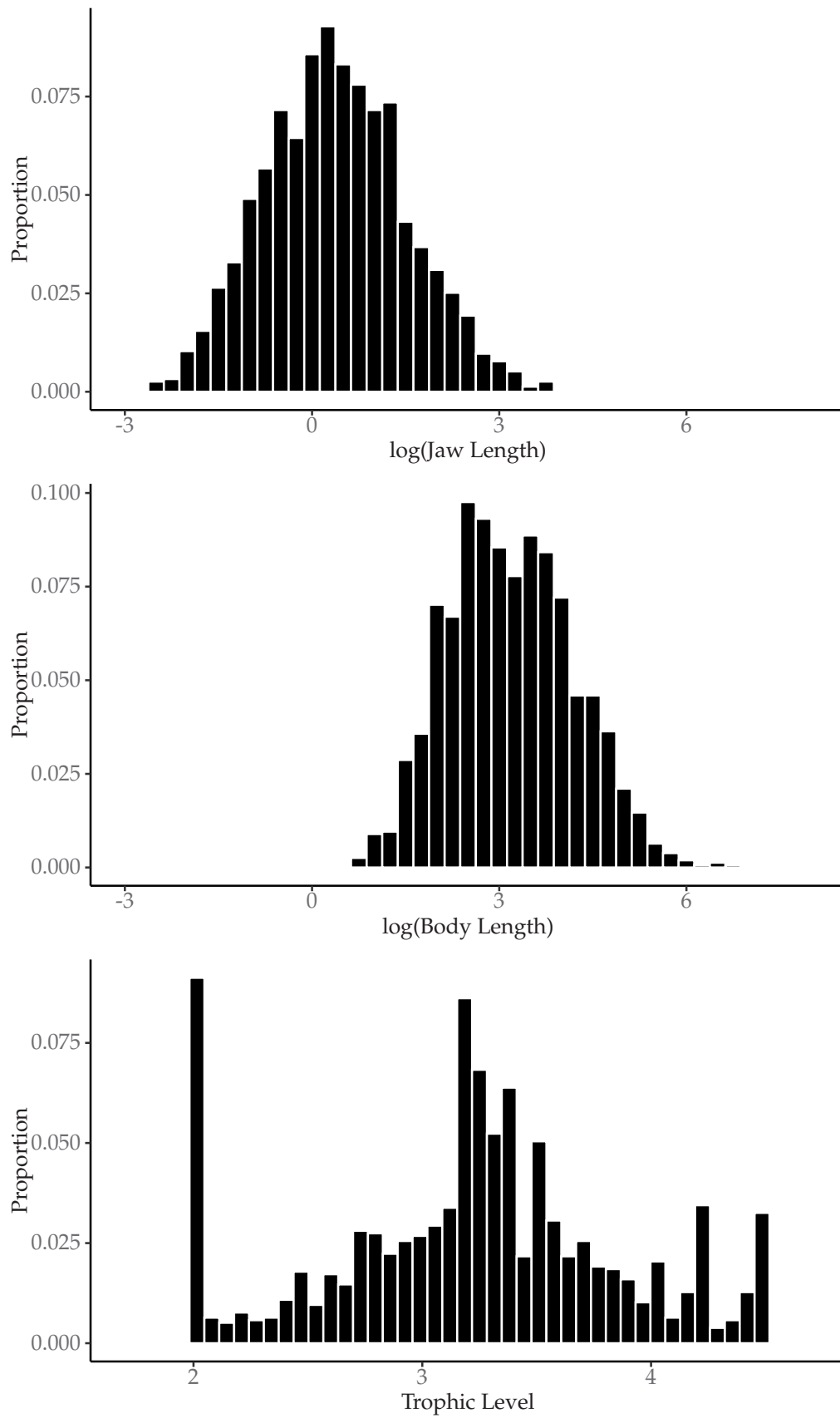
Figure 2.3: Distributions for the logarithimic transformations of jaw length, body length and the non-transformed trophic level
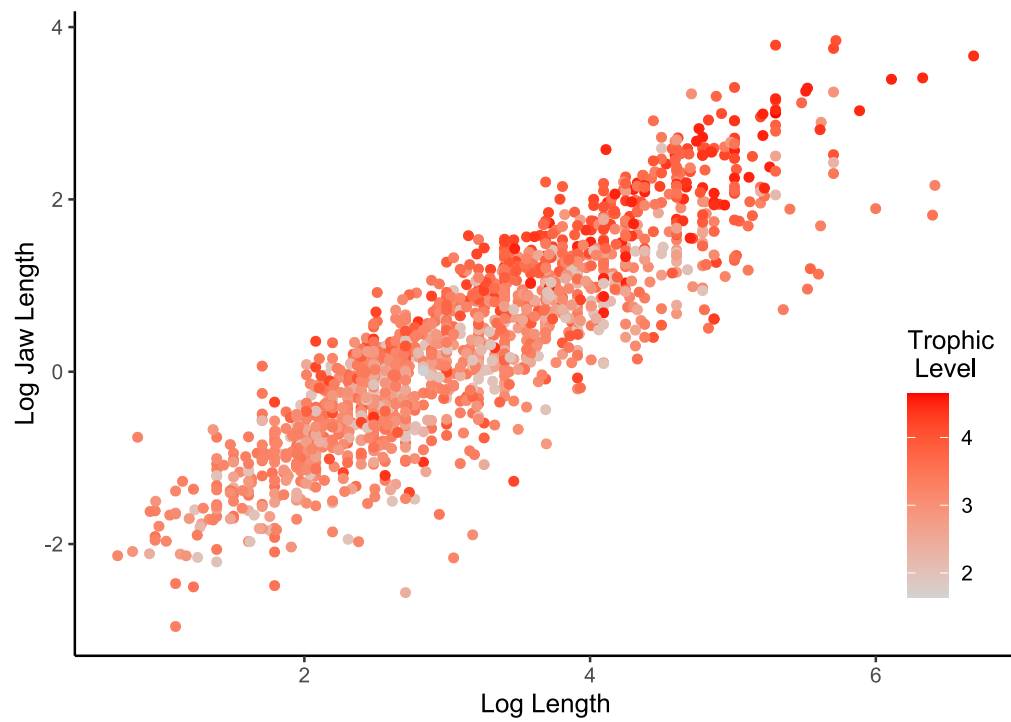
Figure 2.4: Scatterplot of species' log jaw length and log body length with trophic level illustrated by color gradient

# Chapter 3
# Theory

## 3.1 Handling phylogenetic covariance

Let us define first the standard Ordinary Least Squares (OLS) linear regression model as

$$\vec{y} = \mathbf{X}\vec{b} + \vec{\varepsilon}$$

where $\varepsilon \sim N(\vec{0}, \mathbf{V})$ and $\mathbf{V} = \sigma^2 \mathbf{I}$. In other words, we assume that the residual errors are uncorrelated to one another. Where data are correlated, this assumption is broken and the covariance matrix $\mathbf{V}$ has to be corrected for in some way. If the covariance $\mathbf{V}$ for the residuals is known it can be used to find covariance-corrected model parameter estimates as [12]

$$\hat{\vec{b}} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \vec{y} \tag{3.1}$$

With phylogenetic data with known phylogenies, we can, for a given random diffusion model[1] for the investigated traits, calculate what $\mathbf{V}$ should look like. We can also, after having calculated the best-fitting (adjusted for number of parameters) diffusion model, skip equation 3.1 entirely and instead use the diffusion models themselves to draw inferences not just about the current relationship between the traits, but also about the evolutionary trajectory of their relationship; see section 3.5.

For first degree Markovian processes, we can calculate the (expected) covariance matrix from timed phylogenetic trees, but the exact formulae used will differ based on the model. The next two sections will introduce the two main models used in this investigation.

## 3.2 The Brownian Motion process

The Brownian motion (BM) process is the simplest diffusion model and can be described in the differential form

$$dZ(t) = \sigma dW(t)$$

where $W(t)$ is the Wiener process, a continuous generalization of a discrete process where the displacement in each time-step is distributed as a standard random normal. In other words, for the continuous Wiener process we have $W(t_2) - W(t_1) \sim N(0, t_2 - t_1)$ for all values $t_1 \leqslant t_2$. The unidimensional Brownian motion process is thus nothing more than a scaled Wiener process.

The multivariate version can be written as

$$d\vec{Z}(t) = \Sigma d\vec{W}(t)$$

where $\vec{W}(t)$ is a vector of independent Wiener processes, and $\Sigma$ is the covariance matrix for the process. The correlation of a single trait of two species that have evolved under a

---

[1] In some of the biological literature the term "drift model" is used instead.

BM process (for an ultrametric tree) is simply the amount of time for which they share an ancestry divided by the total amount of time from the present to the last common ancestor of all species in the dataset. The covariance is that correlation times $\sigma$.

For the multivariate case, the covariance of traits $k$ and $l$ of species $i$ and $j$ is

$$Cov[Z_{ik}, Z_{jl}] = L_{ij}\Sigma_{kl}$$

where $L_{ij}$ is the amount of time from the start of the tree to the last common ancestor node of species $i$ and $j$ ([8], Equation 4). Let us note that this implies that the stochastic processes for $Z_{ik}$ and $Z_{jl}$ are independent of one another from the moment that their lineages diverge.

## 3.3 The Ornstein-Uhlenbeck process

The Ornstein-Uhlenbeck (OU) process is typically written as a stochastic differential equation of the form

$$dZ(t) = -\alpha(Z(t) - \psi(t))dt + \sigma dW(t)$$

This can be understood as a random process with a long-term mean $\psi$. $\alpha$ determines the speed of the regression to that long-term mean, and $\sigma dW$ represent the infinitesimal increment of a Wiener process with standard deviation $\sigma$.

The multivariate generalization of the OU process is

$$d\vec{Z}_t = -\mathbf{F}(\vec{Z}(t) - \vec{\Psi}(t))dt + \mathbf{\Sigma}d\vec{W}(t)$$

with $\mathbf{F}$ and $\mathbf{\Sigma}$ being the matrix analogues of $\alpha$ and $\sigma$, respectively. The solution to this SDE is

$$\vec{Z}(t) = e^{-\mathbf{F}t}\vec{Z}(0) + \int_0^t e^{-\mathbf{F}(t-\nu)}\mathbf{F}\vec{\Psi}(t)d\nu + \int_0^t e^{-\mathbf{F}(t-\nu)}\mathbf{\Sigma}d\vec{W}(t) \tag{3.2}$$

which will be normally distributed for all $t$ [2].

The covariance matrix for pairs of individual lineages for this process is

$$Cov[\vec{Z}_i, \vec{Z}_j] = e^{-\mathbf{F}(L_i - L_{ij})}\left(\int_0^{L_{ij}} e^{-\mathbf{F}\nu}\mathbf{\Sigma}\mathbf{\Sigma}^T e^{-\mathbf{F}^T\nu}d\nu\right)e^{-\mathbf{F}^T(L_j - L_{ij})}$$

where $L_i$ and $L_j$ are the lengths of time from the start of the tree to tips $i$ and $j$, respectively. For ultrametric trees, these lengths will be equal. For further information, see appendix B of [2]. Let us note that here, too, the stochastic processes for the vectors $\vec{Z}_i$ and $\vec{Z}_j$ must be independent from the point of divergence on.

**The half-life of an OU-process**

The deterministic second term of the right hand side of Equation 3.2 decreases exponentially, and the rate of decrease is a function of the eigenvalues of $\mathbf{F}$ (as the exponentiation implies a diagonalization of $\mathbf{F}$). For the $i$:th eigenvector, the associated half-life can be defined as the time that it takes for its distance from the optimum to halve [2]. We can express it as

$$0.5 = e^{-h_i Real(\lambda_i)} \implies h_i = \frac{log(2)}{Real(\lambda_i)} \tag{3.3}$$

where $h_i$ and $\lambda_i$ are the half-life and eigenvalue of eigenvector $i$.

## 3.4 Mixed models and the Mixed Gaussian Phylogenetic Model

A mixed model[2] is one that combines several diffusion models. These sub-models apply to one contiguous group of nodes in the phylogenetic relatedness graph each, and for any node only one sub-model applies. The contiguity of nodes that share the same sub-model will henceforth be called evolutionary regimes, or just regimes.

The theoretical reasoning behind mixed models is that, for large and diverse phylogenetic trees, structural or other forms of paradigmal changes may have occurred in sub-lineages which makes it so that they are ill-described by the stochastic process that best describes their ancestral nodes. This is in contrast with global models, which instead fit a single diffusion model on the entire phylogeny.

A Mixed Gaussian phylogenetic model (MGPM) is a mixed model where each of the sub-models are from a family of gaussian stochastic processes (that are collectively referred to as $GL_{inv}$-processes). For details about this family of models, see [21]. For the purposes of this investigation, I will emphasize two aspects of this family: the first is that the traits of lineages tracing from some common starting point will at any point $t$ in the future be distributed as a multivariate normal distribution, and the second is that any pair of lineages, once branched off, are independent of one another. We have seen in Section 3.3 and Section 3.2 that both of these aspects are correct in the case of the BM and OU models.

Formulae for maximum likelihood estimates of the parameters for $GL_{inv}$-processes are derived in [20].

MGPMs have as many parameters as the sum of the number of parameters of their constituent sub-models, and are therefore prone to overfitting if the number of parameters is not properly penalized during model selection. The Akaike Information Criterion (see Section 3.6 for the exact formula used) was used in this investigation to prevent overfitting.

## 3.5 The Concept of Regression in the Phylogenetic Setting

With the multivariate normal distribution $\vec{Z}(t)$, we can calculate the conditional relationship $\mathbb{E}[Z_1(t)|Z_2(t) = z_2, Z_3(t) = z_3, ...]$ which must be a linear relationship with regard to the predictors with normally distributed residuals. Because of this, we can introduce the concept of the evolutionary regression, first described in [12]. This can be viewed as the expected standard regression line for a group of species that all started at the same point $\vec{Z}(0)$ and have developed independently for a duration $t$.

If we have $t$ approach infinity, the distribution of $\vec{Z}(t)$ will reach an equilibrium distribution either around its optimum $\vec{\Psi}$ (assuming that this optimum is constant), or around a non-optimal specific point. [2] calls these optimal and limiting regressions, respectively. In this work I will simply call the linear relationship as $t$ approaches infinity the asymptotic regression, without regard as to whether this value is centered on $\vec{Psi}$ or not. For BM processes, since there is no optimum, the distribution will be centered on the starting point.

While the derivation of the aforementioned linear conditional relationships can be quite analytically tricky, it is quite easy to simulate these stochastic processes. We will therefore illustrate them with standard linear regressions on independent realizations of these processes for specific values of $t$ (in other words, through simulated species that are all products of a radiation from the common ancestor of the evolutionary regime which corresponds to that stochastic process).

## 3.6 Evaluation Metrics on Phylogenetic Models

Because of the highly nested structure of datasets that have biological species as their observations, there are difficulties in constructing valid absolute measures of model fit. Whereas a

---

[2]Not to be confused with any member of the family of linear mixed models

traditional ordinary least squares regression would give you a measure of explained variance (under the assumption of uncorrelated observations), no such measure exists for phylogenetic methods.

Furthermore there is no obvious null hypothesis, and any proposed one would necessarily be easily rejected by even a marginally complex model. In this investigation I shall instead compare a mixed model to some reasonable single-regime models, and choose the best among them based on their AIC scores.

### The Akaike's Information Criterion

The Akaike's Information Criterion (AIC) is used as a relative metric of model propriety; i.e. for a set of candidate models for a dataset the one with the lowest AIC score is considered to be the most appropriate in the sense of best balancing goodness-of-fit and model complexity. For a model with $k$ adjustable parameters and with likelihood $L$ over a given data set, it is calculated as [1]:

$$AIC = 2k - 2ln(L)$$

### The Akaike's information criterion with correction for small sample size (AICc)

While the AIC does punish overly complex models, there is a wide consensus that it does not do so enough in situations where the data set is small in relation to the number of parameters (for an explanation and a list of publications arguing this, see [15]). The AICc criterion exists for such situations where the AIC may be biased toward more complex models, and its formula is [15]

$$AICc = AIC + 2\frac{k(k+1)}{n-k-1}$$

Note that if $n$ is very large relative to $k$, the second term will be close to 0 and therefore the entire expression would be well approximated by the AIC. Note also that this expression expressly requires $n$ as an argument.

## 3.7 The PCMFit package

The PCMFit package for R [21] fits maximum likelihood estimates for single stochastic process models over given phylogenetic trees and can, by way of a recursive clade partitioning (RCP) algorithm, fit an MGPM model (see B for a detailed pseudocode description of the algorithm).

In order to reduce the search space of the RCP algorithm, I set the minimum number of tips per regime to 20. This is a fairly low number from a performance perspective, since there are more than 1500 species in the data set.

### Submodel specifications

In order to fit a MGPM model, one must first specify what types of non-mixed models it can be comprised of. For this investigation, we have chosen six whose forms are outlined in Table 3.1. Note that these are ordered by increasing complexity; for our investigation with three traits, the number of parameters go from 3 to 18 in steps of 3.

| Model name | General type | Shape of $F$ matrix | Shape of $\Sigma$ matrix |
|---|---|---|---|
| A | Brownian motion | - | Diagonal |
| B | Brownian motion | - | Symmetric |
| C | Ornstein-Uhlenbeck | Diagonal | Diagonal |
| D | Ornstein-Uhlenbeck | Diagonal | Symmetric |
| E | Ornstein-Uhlenbeck | Symmetric | Symmetric |
| F | Ornstein-Uhlenbeck | Fully asymmetric | Symmetric |

Table 3.1: Table of non-mixed model shapes

# Chapter 4
# Trophic Level

One can, simplistically, say that a species's trophic level corresponds to its position in the food web. The idea is correct, but the formal definition is recursive and therefore highly dependent on the trophic level estimates of other species in the same food web.

In theory, the trophic level of species $i$ can be thought of as the average number of other entities (either animals or plants) that a unit of energy consumed by species $i$ has previously been consumed by. It is calculated recursively for each node of a food web graph as a function of that node's neighbours' trophic levels. The specific formula used for the trophic level of species $i$ in an ecosystem of $n$ different animal species is [25]

$$T_i = 1 + \sum_{j=1}^{n+1} p_{ij}\, T_j$$

where $p_{ij}$ is the proportion of species $i$'s diet that is made up of species $j$. Note the extra index that stands for plants and detritus, which definitionally have a trophic level of 1 and which we model here as a species which does not consume anything.

If we consider a situation where food proportion percentages have been mismeasured for any one node of the food web graph, then we can intuitively understand that the errors in trophic level estimates caused are propagated to other "downstream" parts of the food web, possibly in ways that amplify the error, and that these estimate errors ought to be correlated for the "downstream" species. For the data used in this investigation, most estimates of trophic level are not based on direct diet item proportions, but rather on estimates simulated from the relative diet proportion ranking of food items. This means that we ought to expect our trophic level data to be extremely noisy.

It is easy to come up with worst-case scenarios where a measurement error can end up having disastrously distorting effects. Imagine e.g. a species of fish which only eats plants, with the occasional act of cannibalism. Let us say that the real long-term ratio of plants to cannibalism for this species is something like 90 to 10. We can then calculate its real trophic level as

$$T = 1 + 0.9 \cdot 1 \;+\; 0.1 \cdot T \implies (1 - 0.1) \cdot T = 1.9 \implies T \approx 2.11$$

Now imagine instead that this species was observed under a period of ecosystem duress, and that the ratio during this short period was instead 50/50. We would then calculate the trophic level as being 3. Furthermore this error would be spread out to estimates for other species that directly or indirectly feed on the species that was misobserved, even if their food proportions are completely accurately measured.

The phylogenetic regression used in the main analysis of this thesis assumes that any correlation in the residuals is phylogenetic in origin. Given that we clearly must have some level of food web error correlation, it is of interest to investigate how it might affect our results and, if possible, how to mitigate the effects of that correlation.

## 4.1 Expressing the trophic level formula with linear algebra

Say that we have $n$ different species (excepting plants and detritus) in an ecosystem. We summarize their trophic levels in the vector $\vec{T}$:

$$\vec{T} = \begin{pmatrix} T_n \\ T_{n-1} \\ \vdots \\ T_1 \\ 1 \end{pmatrix}$$

where the bottom 1 is the only initially known value (which is the trophic level for plants and detritus). Let us also define the $n + 1 \times n + 1$ matrix $P$ as the matrix where each element $p_{ij}$ is the proportion of species $i$'s diet that is made up of species $j$. The bottom row of this matrix consists of all 0s (as plants and detritus do not eat anything). We can now write our trophic level formula as

$$\vec{T} = \mathbf{P}\,\vec{T} + \vec{1} \implies \vec{0} = (\mathbf{P} - \mathbf{I})\vec{T} + \vec{1}$$

This can, if we assume that $\mathbf{P} - \mathbf{I}$ is invertible, be further simplified to

$$\vec{T} = -(\mathbf{P} - \mathbf{I})^{-1}\vec{1} \tag{4.1}$$

Intuitively, we know that the key aspect of whether or not $\mathbf{P} - \mathbf{I}$ is invertible (or in other words, if trophic level is calculable) is if all species are connected (either directly or indirectly) to the bedrock plant/detritus "species". Algebraically, we can tell from Equation 4.1 that the determinant of $\mathbf{P} - \mathbf{I}$ needs to be something other than 0. It is easy to show for the case where $n = 2$, where the corresponding $\mathbf{P} - \mathbf{I}$ matrix is $3 \times 3$, that any cycle in the graph that is unconnected to the plant/detritus "species", whether it be pure cannibalism or two species eating each other but not the plant and detritus "species", or any combination of those two, will cause the determinant to be 0. With a bit more work this can be proven for the general case of $n$ species as well, using Leibniz's formula for the determinant of an arbitrary size matrix.

## 4.2 The impact of a disturbance in P

Imagine that we have $\mathbf{P}^* = \mathbf{P} + \mathbf{K}$, where $\mathbf{K}$ is the $n + 1 \times n + 1$ 0-matrix with the exception of some row $r$. We can think of $\mathbf{P}^*$ as a matrix where the diet proportions of species $r$ have been mismeasured. We wish to study the difference vector

$$\vec{\varepsilon} = \vec{T^*} - \vec{T}$$

With Equation 4.1 and the expression for $\mathbf{P}^*$, we can rewrite this as

$$\vec{\varepsilon} = ((\mathbf{P} - \mathbf{I})^{-1} - (\mathbf{P}^* - \mathbf{I})^{-1})\vec{1} = ((\mathbf{P} - \mathbf{I})^{-1} - (\mathbf{P} + \mathbf{K} - \mathbf{I})^{-1})\vec{1}$$

Let us note that $K$ has rank 1 (as there is only one row that is non-zero). In order to further simplify this expression we will use the Sherman-Morrison formula, which states that

$$(\mathbf{A} + \vec{u}\vec{v}^T)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\vec{u}\vec{v}^T\mathbf{A}^{-1}}{1 + \vec{v}^T\mathbf{A}^{-1}\vec{u}}$$

Since $K$ has rank 1, we can rewrite it as $K = uv^\mathsf{T}$, where $u$ is the zero vector with the exception of the $r$:th index, which we will set to 1, and $v^\mathsf{T}$ is identical to the non-zero row in $K$. We get

$$\vec{\varepsilon} = ((\mathbf{P} - \mathbf{I})^{-1} - ((\mathbf{P} - \mathbf{I}) + \mathbf{K})^{-1})\vec{1} \tag{4.2}$$

$$\vec{\varepsilon} = \left( (\mathbf{P} - \mathbf{I})^{-1} - (\mathbf{P} - \mathbf{I})^{-1} + \frac{(\mathbf{P} - \mathbf{I})^{-1}\vec{u}\vec{v}^{\mathsf{T}}(\mathbf{P} - \mathbf{I})^{-1}}{1 + \vec{v}^{\mathsf{T}}(\mathbf{P} - \mathbf{I})^{-1}\vec{u}} \right)\vec{1} =$$

$$\vec{\varepsilon} = \frac{(\mathbf{P} - \mathbf{I})^{-1}\vec{u}\vec{v}^{\mathsf{T}}(\mathbf{P} - \mathbf{I})^{-1}\vec{1}}{1 + \vec{v}^{\mathsf{T}}(\mathbf{P} - \mathbf{I})^{-1}\vec{u}} \tag{4.3}$$

The correlation matrix of the difference vector is

$$Cov[\vec{\varepsilon}, \vec{\varepsilon}] = E[(\vec{\varepsilon} - E[\vec{\varepsilon}])(\vec{\varepsilon} - E[\vec{\varepsilon}])^{\mathsf{T}}]$$

Remember that the vector $v$ is the only one where we have any randomness; i.e.

$$E[\vec{\varepsilon}] = \frac{(\mathbf{P} - \mathbf{I})^{-1}\vec{u}\, E[\vec{v}^{\mathsf{T}}]\, (\mathbf{P} - \mathbf{I})^{-1}\vec{1}}{1 + \vec{v}^{\mathsf{T}}(\mathbf{P} - \mathbf{I})^{-1}\vec{u}}$$

Assuming that the possible disturbances on $v$ are unbiased[1], $E[\vec{\varepsilon}] = \vec{0}$. The expression for the covariance thus simplifies to

$$Cov[\vec{\varepsilon}, \vec{\varepsilon}] = E[\vec{\varepsilon}\vec{\varepsilon}^{\mathsf{T}}]$$

We have that the product is

$$\vec{\varepsilon}\vec{\varepsilon}^{\mathsf{T}} = \frac{(\mathbf{P} - \mathbf{I})^{-1}\vec{u}\vec{v}^{\mathsf{T}}(\mathbf{P} - \mathbf{I})^{-1}\vec{1}\vec{1}^{\mathsf{T}}(\mathbf{P} - \mathbf{I})^{-\mathsf{T}}\vec{v}\vec{u}^{\mathsf{T}}(\mathbf{P} - \mathbf{I})^{-\mathsf{T}}}{(1 + \vec{v}^{\mathsf{T}}(\mathbf{P} - \mathbf{I})^{-1}\vec{u})^2} \tag{4.4}$$

Note that the product of the factors from $v^{\mathsf{T}}$ to $\vec{v}$ evaluate to a scalar. Let's call that scalar $R$, and further note that in simplifying in this way, this scalar and the denominator are the sole random variables in the entire expression. Thus the resulting covariance matrix will have each of its elements be a scalar multiple of $R$ divided by the denominator. Finally, let us also note that $R$ is essentially the dot product of a (non-complex) linear transformation of the (non-complex) $\vec{v}$ on itself, and will therefore be non-negative. Similarly the denominator, also, must be non-negative.

Remember also, from Equation 4.1, that $(\mathbf{P} - \mathbf{I})^{-1}\vec{1} = -\vec{T}$. This means that $R = v^{\mathsf{T}}(-1)^2\vec{T}\vec{T}^{\mathsf{T}}v = v^{\mathsf{T}}\vec{T}\vec{T}^{\mathsf{T}}v$. If we describe $v$ as $v = (v_1, v_2, ..., v_{n+1})^{\mathsf{T}}$, then we have

$$R = \sum_{i=1}^{n+1}\sum_{j=1}^{n+1} T_j T_i v_i v_j$$

Let us now note that $uu^{\mathsf{T}}$ is the 0-matrix with the exception of the $r, r$th element, which is 1. This means that the factors $uu^{\mathsf{T}}(\mathbf{P} - \mathbf{I})^{-\mathsf{T}}$ will evaluate to a matrix that is all zeros with the exception of the $r$:th row, which will be the transpose of the $r$:th column of $(\mathbf{P} - \mathbf{I})^{-1}$. For the sake of simplicity we will refactor the numerator of 4.4 to $R\mathbf{M}$, where the elements $M_{ij}$ of $\mathbf{M}$ are

$$M_{ij} = [(\mathbf{P} - \mathbf{I})^{-1}]_{i,r}[(\mathbf{P} - \mathbf{I})^{-1}]_{j,r}$$

The matrix $M$ will be proportional to the covariance matrix, with the proportionality constant equal to $c/(1 + v^{\mathsf{T}}(\mathbf{P} - \mathbf{I})^{-1}u)^2$. We note the trivial fact that, generally, $M_{ij} \neq M_{kl}$ when $(i, j) \neq (k, l)$. We also can not assume that off-diagonal elements will be 0.

---

[1]This is only true if we assume that the uncertainty is intensive, i.e. only in the proportions, and not in the support/edges. For those situations 0-valued $p_{ij}$:s will have positive non-zero expectation in $p_{ij}^*$, and all non-zero elements on the same row will have a negative expectation to compensate.

While mismeasurements[2] for one species differentially affect the variance of the trophic levels of other species, that differential is linear and independent of the exact form of the mismeasurement, as $\mathbf{M}$ does not depend on $v$. When it comes to the actual variance introduced by the factor $c/(1 + v^\mathsf{T}(\mathbf{P} - \mathbf{I})^{-1}u)^2$, further study is necessary to determine its asymptotic behaviour. While it is tempting to mistakenly conclude that the denominator should grow faster than the numerator since they are both quadratic and the denominator has an added positive constant term, the non-constant term can actually be negative as well, and should it be close to -1 the full random factor will likely be quite large.

## 4.3 The impact of multiple disturbances in P

We can try and generalize this result for $K$s of ranks up to $n$. In order to do this, we need to use an extended form of the Sherman-Morrison formula in 4.2. This extension[3] uses the fact that a rank $n_r$ matrix can be rewritten as a sum of $n_r$ rank 1 matrices, i.e. $\mathbf{K} = \sum_{i=1}^{n_r} u_i v_i^\mathsf{T}$:

$$\mathbf{C}_{k+1}^{-1} = \mathbf{C}_k^{-1} - g_k \mathbf{C}_k^{-1} u_i v_i^\mathsf{T} \mathbf{C}_k^{-1} \tag{4.5}$$

where $\mathbf{C}_k = (\mathbf{P} - \mathbf{I}) + \sum_{i=1}^{k} \vec{u}_i \vec{v}_i^\mathsf{T}$ and $g_k = \frac{1}{1 + \vec{v}^\mathsf{T} \mathbf{C}_k^{-1} \vec{u}}$.

If we denote the number of terms in $\mathbf{C}_k^{-1}$ as $n_k$, we have that $n_{k+1} = n_k(1 + n_k)$. This means that the number of terms even for relatively small values of $n_r$ will be unmanageable in a manual calculation. However, by considering what the indices of the leftmost $\vec{u}_i$s are in each term, we can draw some conclusions as to the form of $\vec{\varepsilon}\vec{\varepsilon}^\mathsf{T}$. It is not too difficult to see that for every $i \in \{1, 2, ..., n_r\}$, there will be at least one term in Equation 4.5 for which the leftmost $\vec{u}$-factor is $\vec{u}_i$. Since this sum of terms will be multiplied by itself, it also means that, for $\vec{\varepsilon}\vec{\varepsilon}^\mathsf{T}$, we will have a sum of $n_r^2$ terms each consisting of one distinct random scalar factor and one distinct non-random matrix.

This means, essentially, that the error covariance matrix for a scenario where there might be disturbances in food item proportion for all species will have every single one of its entries be random. Unlike in the situation where we have a disturbance in the food item proportions of a single species, these random variables will not be fully correlated with one another.

---

[2]Specifically intensive mismeasurements, i.e. mismeasurements that only affect non-zero $P_{ij}$:s

[3]First published by [19]

# Chapter 5
# Results

## 5.1 Model Comparison

In the table below we compare global versions of the six models outlined in Table 3.1 with the multi-regime MGPM model ("Mixed"). By every metric, the Mixed model performs best. It does consist of far more parameters than any of the other models, however it still has lower AIC and AICc scores than the other models. Note that there are an additional three parameters for each model compared to in Table 3.1; these are for estimates of the starting point of the ancestral root node of the phylogeny.

| Model | Number of Parameters | AIC | AICc | Log Likelihood |
|-------|----------------------|-------|-------|----------------|
| A     | 6                    | 15038 | 15038 | -7512          |
| B     | 9                    | 13206 | 13206 | -6593          |
| C     | 12                   | 11351 | 11351 | -5662          |
| D     | 15                   | 9241  | 9241  | -4604          |
| E     | 18                   | 8986  | 8986  | -4474          |
| F     | 21                   | 8992  | 8992  | -4474          |
| Mixed | 312                  | 6736  | 6894  | -3013          |

Table 5.1: Table of AIC, AICc scores and log-likelihoods of the six baseline models and the Mixed model

## 5.2 Overview of Mixed model regimes

The following table presents the sub-models for each regime in the mixed model, with maximum likelihood estimates for the parameters.

| Regime | Type | X0 | $\mathbf{F}$ | $\Psi$ | $\Sigma$ |
|---|---|---|---|---|---|
| :global: | NA | $\begin{bmatrix} 3.01 \\ 2.85 \\ 0.53 \end{bmatrix}$ | | | |
| 1 | E | | $\begin{bmatrix} 1.48 & -0.12 & -0.19 \\ -0.12 & 0.02 & 0.01 \\ -0.19 & 0.01 & 0.04 \end{bmatrix}$ | $\begin{bmatrix} 3.30 \\ 3.15 \\ 0.29 \end{bmatrix}$ | $\begin{bmatrix} 0.44 & 0.01 & 0.01 \\ 0.01 & 0.01 & 0.01 \\ 0.01 & 0.01 & 0.02 \end{bmatrix}$ |
| 2 | D | | $\begin{bmatrix} 1.33 & 0.00 & 0.00 \\ 0.00 & 2.51 & 0.00 \\ 0.00 & 0.00 & 2.12 \end{bmatrix}$ | $\begin{bmatrix} 3.11 \\ 3.45 \\ -0.02 \end{bmatrix}$ | $\begin{bmatrix} 0.88 & 0.26 & 0.31 \\ 0.26 & 4.76 & 4.58 \\ 0.31 & 4.58 & 5.50 \end{bmatrix}$ |
| 3 | E | | $\begin{bmatrix} 4.24 & 1.82 & -0.49 \\ 1.82 & 2.50 & 0.47 \\ -0.49 & 0.47 & 3.34 \end{bmatrix}$ | $\begin{bmatrix} 3.27 \\ 3.24 \\ 0.74 \end{bmatrix}$ | $\begin{bmatrix} 2.78 & 2.32 & 2.60 \\ 2.32 & 5.59 & 6.95 \\ 2.60 & 6.95 & 8.74 \end{bmatrix}$ |
| 4 | E | | $\begin{bmatrix} 1.76 & 0.81 & -1.25 \\ 0.81 & 1.81 & -1.36 \\ -1.25 & -1.36 & 1.42 \end{bmatrix}$ | $\begin{bmatrix} 3.66 \\ 4.20 \\ 1.50 \end{bmatrix}$ | $\begin{bmatrix} 0.34 & 0.09 & -0.24 \\ 0.09 & 0.12 & -0.04 \\ -0.24 & -0.04 & 0.38 \end{bmatrix}$ |
| 5 | F | | $\begin{bmatrix} 3.69 & -1.46 & 1.63 \\ 0.10 & 1.74 & -1.69 \\ 1.41 & -3.11 & 3.12 \end{bmatrix}$ | $\begin{bmatrix} 2.94 \\ 2.67 \\ -0.28 \end{bmatrix}$ | $\begin{bmatrix} 2.93 & -0.02 & 1.32 \\ -0.02 & 0.14 & -0.16 \\ 1.32 & -0.16 & 0.88 \end{bmatrix}$ |
| 6 | E | | $\begin{bmatrix} 1.42 & -0.07 & -0.14 \\ -0.07 & 0.01 & 0.01 \\ -0.14 & 0.01 & 0.02 \end{bmatrix}$ | $\begin{bmatrix} 3.35 \\ 3.01 \\ 0.34 \end{bmatrix}$ | $\begin{bmatrix} 0.83 & 0.01 & 0.01 \\ 0.01 & 0.02 & 0.02 \\ 0.01 & 0.02 & 0.02 \end{bmatrix}$ |
| 7 | D | | $\begin{bmatrix} 0.07 & 0.00 & 0.00 \\ 0.00 & 0.08 & 0.00 \\ 0.00 & 0.00 & 0.05 \end{bmatrix}$ | $\begin{bmatrix} 3.04 \\ 2.01 \\ -0.80 \end{bmatrix}$ | $\begin{bmatrix} 0.02 & -0.01 & -0.01 \\ -0.01 & 0.05 & 0.04 \\ -0.01 & 0.04 & 0.04 \end{bmatrix}$ |
| 8 | F | | $\begin{bmatrix} 0.02 & 0.02 & -0.02 \\ 0.01 & 0.02 & -0.01 \\ -0.02 & -0.02 & 0.03 \end{bmatrix}$ | $\begin{bmatrix} 3.95 \\ 2.42 \\ 0.30 \end{bmatrix}$ | $\begin{bmatrix} 0.02 & 0.00 & -0.00 \\ 0.00 & 0.02 & 0.02 \\ -0.00 & 0.02 & 0.03 \end{bmatrix}$ |
| 9 | F | | $\begin{bmatrix} 0.51 & 0.50 & -0.48 \\ 0.06 & 0.17 & -0.13 \\ -0.07 & -0.07 & 0.09 \end{bmatrix}$ | $\begin{bmatrix} 3.59 \\ 3.95 \\ 0.77 \end{bmatrix}$ | $\begin{bmatrix} 0.24 & 0.05 & -0.02 \\ 0.05 & 0.06 & 0.04 \\ -0.02 & 0.04 & 0.08 \end{bmatrix}$ |
| 10 | E | | $\begin{bmatrix} 0.47 & -0.04 & 0.02 \\ -0.04 & 0.01 & -0.00 \\ 0.02 & -0.00 & 0.00 \end{bmatrix}$ | $\begin{bmatrix} 2.13 \\ 3.87 \\ 1.44 \end{bmatrix}$ | $\begin{bmatrix} 0.04 & 0.00 & 0.00 \\ 0.00 & 0.02 & 0.02 \\ 0.00 & 0.02 & 0.03 \end{bmatrix}$ |
| 11 | F | | $\begin{bmatrix} 0.02 & 0.17 & -0.14 \\ 0.02 & 0.17 & -0.13 \\ 0.00 & 0.02 & -0.01 \end{bmatrix}$ | $\begin{bmatrix} 4.75 \\ 3.17 \\ 0.92 \end{bmatrix}$ | $\begin{bmatrix} 0.01 & 0.01 & 0.00 \\ 0.01 & 0.03 & 0.02 \\ 0.00 & 0.02 & 0.02 \end{bmatrix}$ |
| 12 | E | | $\begin{bmatrix} 0.05 & -0.01 & -0.00 \\ -0.01 & 0.01 & -0.00 \\ -0.00 & -0.00 & 0.00 \end{bmatrix}$ | $\begin{bmatrix} 3.20 \\ 2.11 \\ 0.31 \end{bmatrix}$ | $\begin{bmatrix} 0.01 & -0.01 & -0.01 \\ -0.01 & 0.01 & 0.01 \\ -0.01 & 0.01 & 0.01 \end{bmatrix}$ |
| 13 | B | | | | $\begin{bmatrix} 0.01 & 0.00 & -0.00 \\ 0.00 & 0.01 & 0.00 \\ -0.00 & 0.00 & 0.01 \end{bmatrix}$ |

| Regime | Type | X0 | F | Ψ | Σ |
|---|---|---|---|---|---|
| 14 | D | | $\begin{bmatrix} 0.04 & 0.00 & 0.00 \\ 0.00 & 0.03 & 0.00 \\ 0.00 & 0.00 & 0.04 \end{bmatrix}$ | $\begin{bmatrix} 2.88 \\ 1.88 \\ -1.14 \end{bmatrix}$ | $\begin{bmatrix} 0.02 & 0.00 & 0.00 \\ 0.00 & 0.02 & 0.01 \\ 0.00 & 0.01 & 0.02 \end{bmatrix}$ |
| 15 | E | | $\begin{bmatrix} 1.15 & 0.21 & -0.72 \\ 0.21 & 1.50 & -0.66 \\ -0.72 & -0.66 & 1.92 \end{bmatrix}$ | $\begin{bmatrix} 2.90 \\ 2.90 \\ 0.38 \end{bmatrix}$ | $\begin{bmatrix} 1.16 & -0.30 & -0.48 \\ -0.30 & 0.67 & 0.84 \\ -0.48 & 0.84 & 1.35 \end{bmatrix}$ |
| 16 | D | | $\begin{bmatrix} 1.39 & 0.00 & 0.00 \\ 0.00 & 0.18 & 0.00 \\ 0.00 & 0.00 & 0.17 \end{bmatrix}$ | $\begin{bmatrix} 3.31 \\ 2.83 \\ 0.38 \end{bmatrix}$ | $\begin{bmatrix} 0.15 & 0.09 & 0.13 \\ 0.09 & 0.10 & 0.15 \\ 0.13 & 0.15 & 0.24 \end{bmatrix}$ |
| 17 | B | | | | $\begin{bmatrix} 0.01 & -0.00 & -0.00 \\ -0.00 & 0.02 & 0.01 \\ -0.00 & 0.01 & 0.02 \end{bmatrix}$ |
| 18 | F | | $\begin{bmatrix} 2.67 & -1.75 & -0.79 \\ -3.50 & 3.33 & 1.49 \\ -6.42 & 1.30 & 7.05 \end{bmatrix}$ | $\begin{bmatrix} 3.18 \\ 2.83 \\ 0.44 \end{bmatrix}$ | $\begin{bmatrix} 2.10 & -1.20 & -1.80 \\ -1.20 & 1.42 & 1.97 \\ -1.80 & 1.97 & 2.96 \end{bmatrix}$ |
| 19 | E | | $\begin{bmatrix} 0.14 & 0.01 & -0.05 \\ 0.01 & 0.10 & -0.02 \\ -0.05 & -0.02 & 0.07 \end{bmatrix}$ | $\begin{bmatrix} 2.73 \\ 2.29 \\ -0.84 \end{bmatrix}$ | $\begin{bmatrix} 0.03 & -0.00 & -0.01 \\ -0.00 & 0.02 & 0.02 \\ -0.01 & 0.02 & 0.02 \end{bmatrix}$ |
| 20 | D | | $\begin{bmatrix} 0.42 & 0.00 & 0.00 \\ 0.00 & 0.56 & 0.00 \\ 0.00 & 0.00 & 0.61 \end{bmatrix}$ | $\begin{bmatrix} 3.47 \\ 2.42 \\ 0.05 \end{bmatrix}$ | $\begin{bmatrix} 0.24 & 0.12 & 0.19 \\ 0.12 & 0.24 & 0.24 \\ 0.19 & 0.24 & 0.40 \end{bmatrix}$ |
| 21 | D | | $\begin{bmatrix} 0.00 & 0.00 & 0.00 \\ 0.00 & 0.07 & 0.00 \\ 0.00 & 0.00 & 0.13 \end{bmatrix}$ | $\begin{bmatrix} 3.23 \\ 2.91 \\ 0.25 \end{bmatrix}$ | $\begin{bmatrix} 0.03 & 0.01 & 0.02 \\ 0.01 & 0.04 & 0.05 \\ 0.02 & 0.05 & 0.07 \end{bmatrix}$ |
| 22 | E | | $\begin{bmatrix} 3.88 & -0.43 & -2.77 \\ -0.43 & 8.93 & 0.80 \\ -2.77 & 0.80 & 7.79 \end{bmatrix}$ | $\begin{bmatrix} 3.57 \\ 3.06 \\ 0.69 \end{bmatrix}$ | $\begin{bmatrix} 2.01 & 0.70 & 0.21 \\ 0.70 & 2.41 & 2.67 \\ 0.21 & 2.67 & 3.33 \end{bmatrix}$ |

Table 5.2: Table of maximum-likelihood estimates for the parameters of the sub-models of the Mixed model

## 5.3 Linear relationships

Table 5.3 shows the results of linear regressions for each regime on a million simulated data points independently drawn from the asymptotic distribution of that regime. These data points have been allowed to "evolve" in accordance with their diffusion models for 10 billion years, which is about 27 times as long as the length of the phylogeny itself[1].

In these regressions, trophic level is the dependent variable and log length and log jaw length the independent variables. The third and fourth columns from the left are the coefficients for the two independent variables. The variables have been centered to have mean 0 and scaled to have standard deviation 1 before the linear fitting.

Figures 5.1 and 5.2 plot the marginal linear relations, i.e. the linear regressions of trophic level on each of the size variables without correction for the correlation between the size variables, unlike in the multivariate regressions in Table 5.3 which show the linear relationships given by the conditional asymptotic distributions. They also show the 70% contour lines for the marginal asymptotic distributions with respect to each of the size variables.

---

[1]As the examiner of this thesis Josef Wilzén pointed out, this is also several times longer than the history of life on Earth. The point here is not, however, to make a far-flung prediction, but rather to examine what stationary distributions these diffusion processes are moving to

| Regime | Intercept | log(Length) | log(Jaw Length) | $R^2$ | Collinearity |
|--------|-----------|-------------|-----------------|-------|--------------|
| 1  | 3.3 | 0.09  | 0.24  | 0.43 | 0.91 |
| 2  | 3.1 | 0.00  | 0.08  | 0.02 | 0.89 |
| 3  | 3.3 | -0.78 | 0.92  | 0.34 | 0.94 |
| 4  | 3.7 | 0.19  | 0.18  | 0.53 | 0.93 |
| 5  | 2.9 | -1.02 | 0.99  | 0.15 | 0.97 |
| 6  | 3.4 | 0.03  | 0.17  | 0.12 | 0.87 |
| 7  | 3.0 | -0.01 | -0.12 | 0.12 | 0.83 |
| 8  | 4.0 | -1.33 | 2.23  | 0.83 | 0.80 |
| 9  | 3.6 | -0.16 | 0.32  | 0.11 | 0.94 |
| 10 | 2.2 | 0.12  | -0.85 | 0.94 | 0.82 |
| 11 | 4.8 | -1.03 | 1.00  | 0.13 | 0.95 |
| 12 | 3.3 | -0.25 | 1.89  | 0.96 | 0.99 |
| 13 | 3.0 | 7.91  | -5.72 | 0.44 | 0.57 |
| 14 | 2.9 | 0.00  | 0.12  | 0.05 | 0.75 |
| 15 | 2.9 | -0.77 | 0.76  | 0.24 | 0.89 |
| 16 | 3.3 | 0.13  | -0.02 | 0.22 | 0.94 |
| 17 | 3.0 | -0.07 | -1.51 | 0.02 | 0.80 |
| 18 | 3.2 | -0.12 | 0.56  | 0.33 | 0.93 |
| 19 | 2.7 | -0.25 | 0.29  | 0.22 | 0.82 |
| 20 | 3.5 | 0.01  | 0.32  | 0.36 | 0.79 |
| 21 | 3.0 | 0.46  | -0.22 | 0.00 | 0.86 |
| 22 | 3.6 | -0.17 | 0.49  | 0.31 | 0.90 |

Table 5.3: An overview of the properties and parameters of the conditional asymptotic distribution lines of each regime

Collinearity measures the Pearson correlation between the simulated independent variables.

## 5.4 Half lives

Below are the half lives for each eigenvector of each regime, as calculated from equation 3.3. The time units are given as percentages of the phylogeny age (which is 369 million years). The BM-regimes have no half-lives, as they do not have optima - theirs are given as Not Applicable (NA). Certain OU-regimes have only two non-zero eigenvalues, and for those the half life of the final eigenvector is also given as NA.

   Of particular note here is that both of the regimes for which the fitted sub-model had a $R^2$ over 0.9 (regimes 10 and 12; this can be read from Table 5.3) have only two half-lives. This means that the deterministic part of their diffusion processes converges towards a line, rather than a point. Depending upon orientations of the two eigenvectors, this can lead to a very strong linear relationship. The orientation is, key however - the sub-model for regime 21, for instance, also lacks a third half-life and yet has an $R^2$ of exactly 0. In calculating the eigenvectors for the **F** matrix, I find that the two eigenvectors with non-zero lambdas correspond exactly to log length-axis and the log jaw length-axis, respectively, and thus that no convergence happens along the trophic level-axis in the trait space.

## 5.5 Regime Phylogeny and Trait distribution overview

Figure 5.3 shows an overview of the phylogenetic relationships of the regimes found with the recursive clade partition algorithm. All the tips have been collapsed to a single tip for each regime, and each regime has an associated color. By tracing these colors "root-ward" along

| Regime | 1st eigenvector | 2nd eigenvector | 3d eigenvector |
|--------|----------------:|----------------:|---------------:|
| 1      | 0.12            | 13.79           | 43.25          |
| 2      | 0.07            | 0.09            | 0.14           |
| 3      | 0.03            | 0.05            | 0.16           |
| 4      | 0.05            | 0.19            | 2.75           |
| 5      | 0.03            | 0.06            | -0.64          |
| 6      | 0.13            | 20.28           | 23.67          |
| 7      | 2.22            | 2.63            | 3.48           |
| 8      | 3.36            | 3.36            | -3.91          |
| 9      | 0.48            | 0.48            | -15.20         |
| 10     | 0.39            | 15.42           | NA             |
| 11     | 0.56            | -1.93           | -3.78          |
| 12     | 3.62            | 34.12           | NA             |
| 13     | NA              | NA              | NA             |
| 14     | 4.29            | 5.04            | 6.55           |
| 15     | 0.07            | 0.16            | 0.28           |
| 16     | 0.14            | 1.07            | 1.09           |
| 17     | NA              | NA              | NA             |
| 18     | 0.02            | 0.05            | 0.12           |
| 19     | 1.10            | 1.87            | 4.62           |
| 20     | 0.31            | 0.34            | 0.45           |
| 21     | 1.45            | 2.62            | NA             |
| 22     | 0.02            | 0.02            | 0.08           |

Table 5.4: Table of half-lives along each of the three eigenvectors of the sub-models of the Mixed model in percentages of the total phylogeny age

the tree to the earliest point where they appear, we can find the starting time of each regime, as well as their ancestor regimes.

Figure 5.4 shows the observed distribution of each regime along the three variables that are under investigation. While there are a few regimes that seem to span the most of the range of all three variables, most are concentrated to small ranges in some or all of the variables. The main reason to check this plot is to see if there is a large mismatch between the range for the variables in which the extant species appear, and the ranges spanned by the stationary distributions of Figures 5.1 and 5.2.

Figure 5.1: Marginal linear dependence and 70% contour lines for the asymptotic distributions of the first 11 regimes
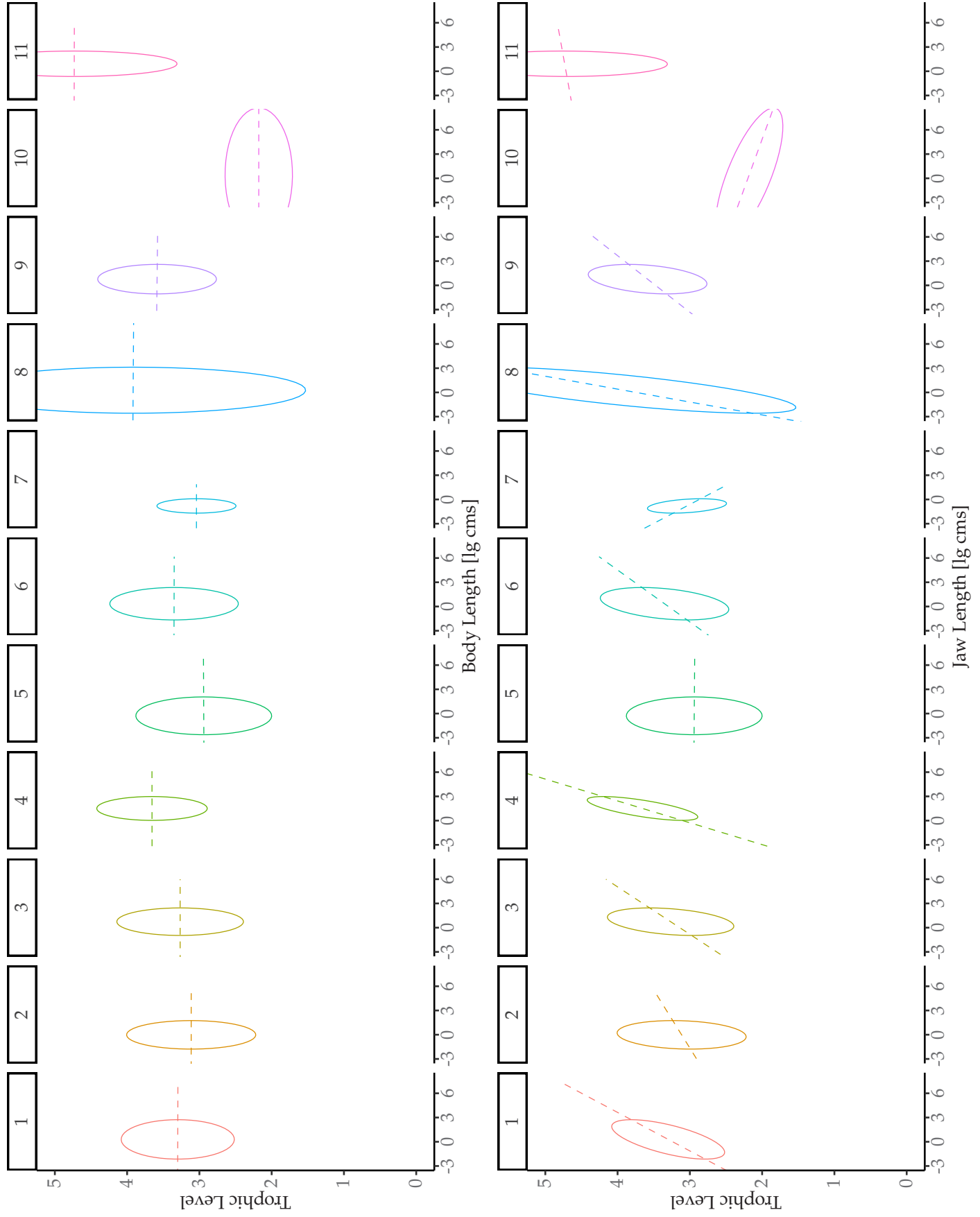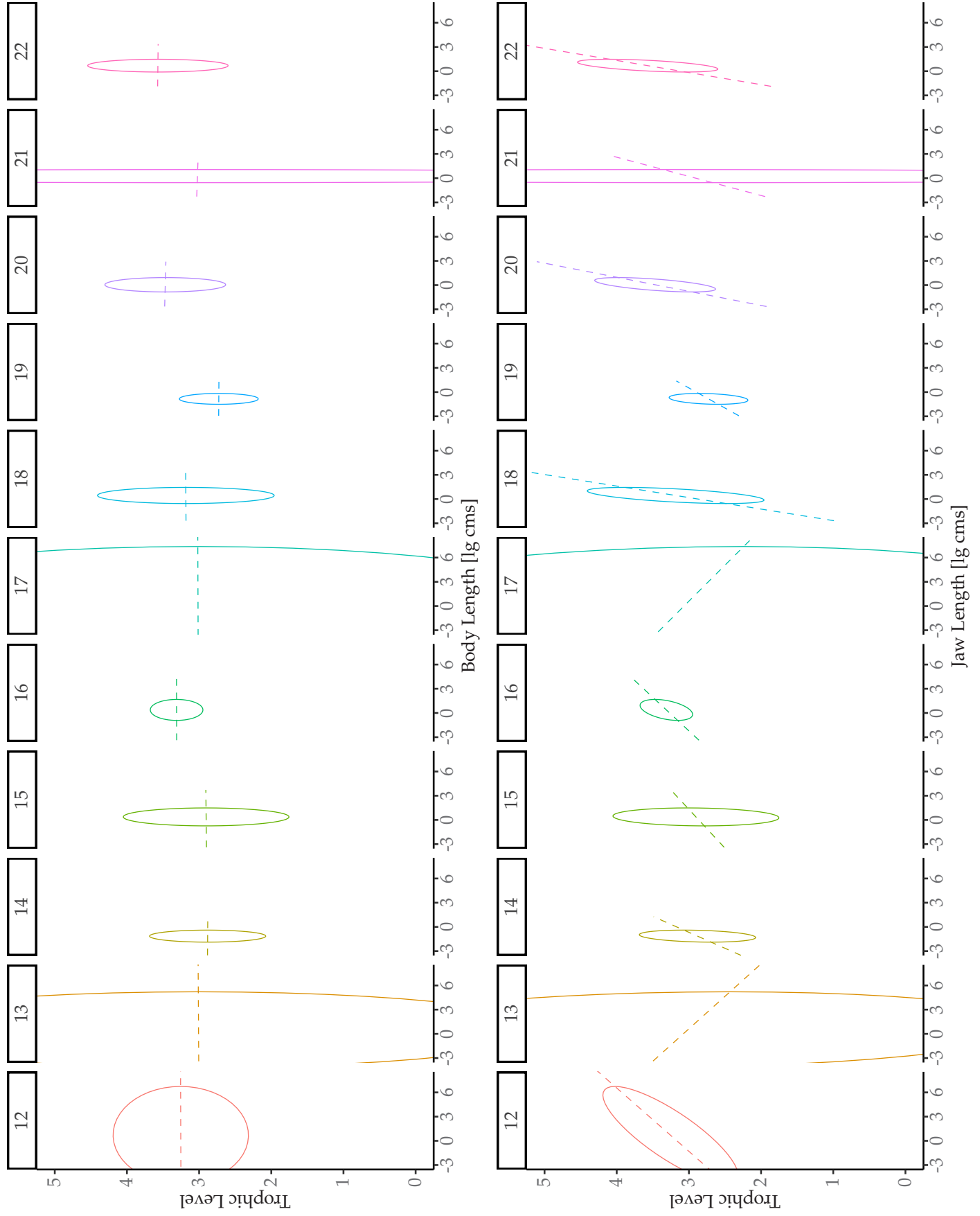
Figure 5.2: Marginal linear dependence and 70% contour lines for the asymptotic distributions of the last 11 regimes
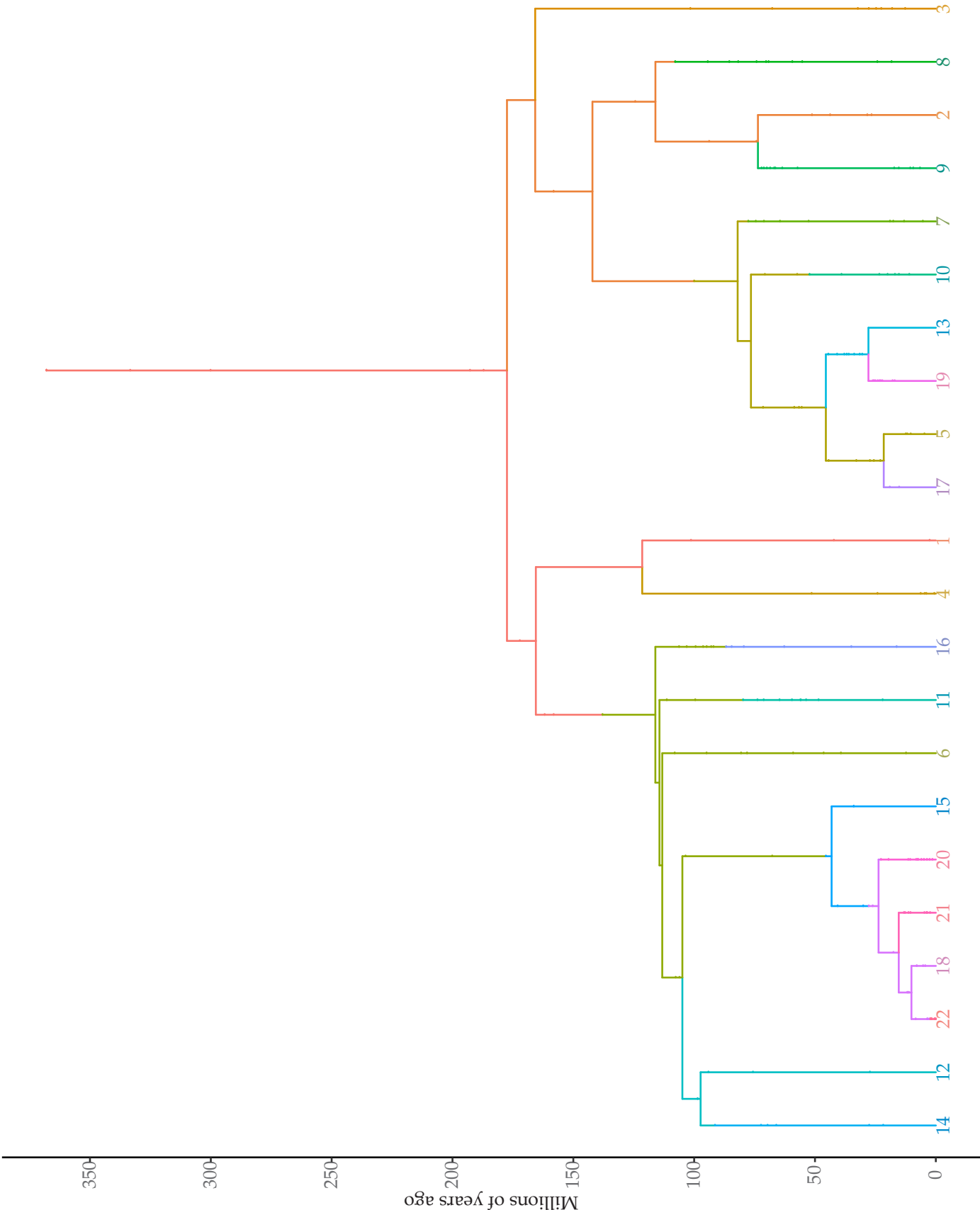
Figure 5.3: Simplified tree plot of evolutionary regimes

This plot collapses all the species into a single tip per regime. For the full plot, see A.1
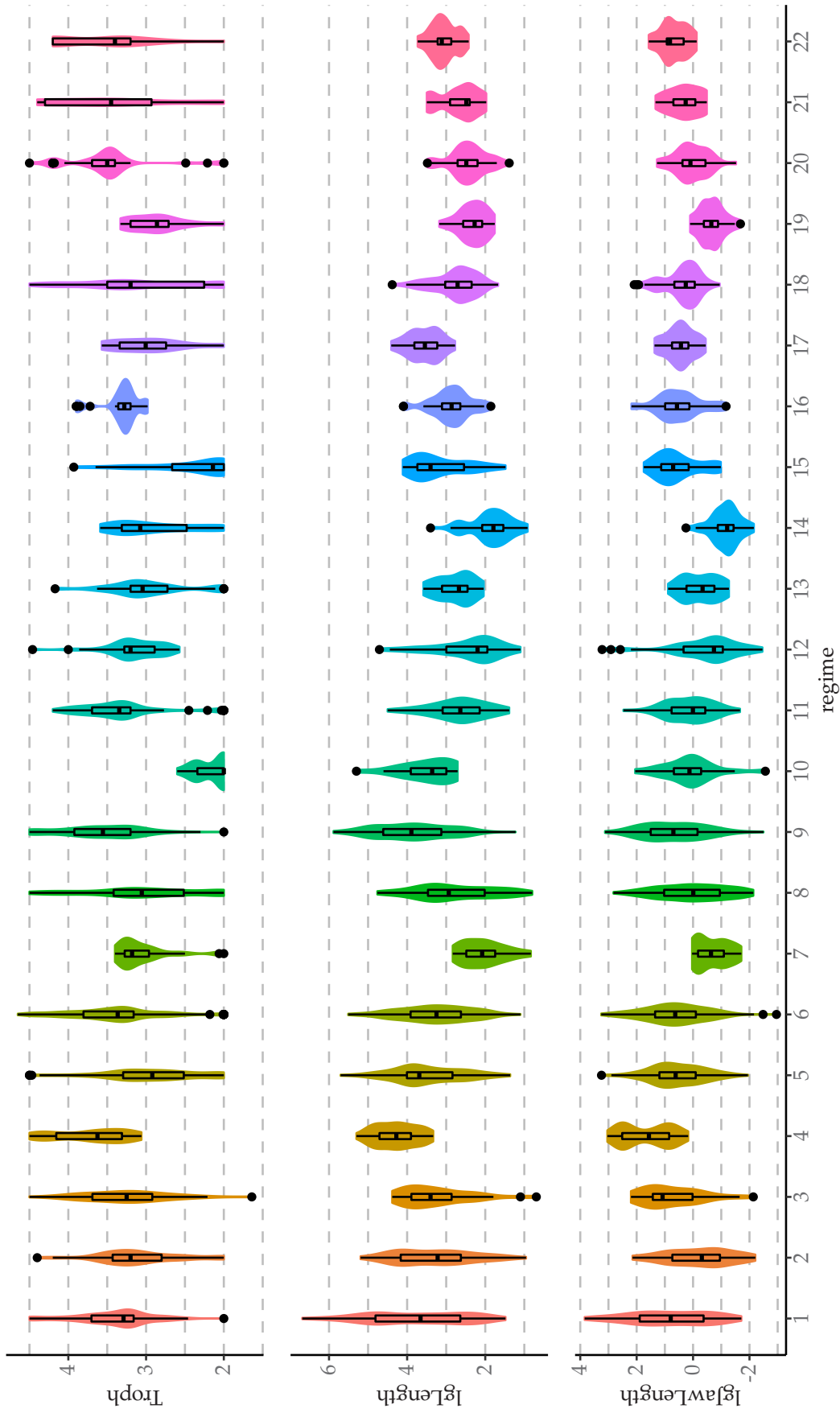
Figure 5.4: Violin box plots of the three variables (rows) for each regime (columns)

# Chapter 6
# Discussion

## 6.1 Results

The AIC scores are overwhelmingly in favor of the mixed model over the global-regime models. This is not too surprising as the phylogeny under investigation connects a large and diverse group of fish, and clearly that diversity is also a factor when it comes to the relationship between trophic level and size.

### Linear relationships

An inspection of the conditional linear relationships in Section 5.3 reveals that for all but three regimes jaw and body size explain at least 10% of the variance in trophic level, and for four regimes these two variables explain at least 50% of the variance. For the 19 regimes for which the size variables explain more than 10% of the variance, jaw size seems to be more positively related to trophic level than body size. Jaw size has a positive linear coefficient for 15 regimes as compared to 7 for body size. Two regimes, 10 and 13, stand out from the rest in that they have considerable positive effect sizes for body size while also having considerable negative effect sizes for jaw size. These two regimes consist of two closely related families, Cyprinidae for regime 10 and Leuciscidae for regime 13, where higher trophic level is most commonly associated with insectivory rather than piscivory [5].

Regime 7, which consists of species of the Danionidae family, has a negative relationship between either type of size and trophic level. Regime 16 (which has members from a wide range of families) has a negative coefficient for the effect of jaw size as well. For these two regimes, however, the effect size is relatively small.

Because of the high positive levels of collinearity in the size variables, there may be some value in comparing the insights from the discussion of the conditional linear relationships with the results for the marginal linear relationships in 5.1 and 5.2. These marginal lines show no real effect for body length, but does show some noticeable slopes for the jaw length. Together with what we noted for the conditional lines, this implies that, for most regimes, the effect of body size only has predictive value when taken into consideration together with jaw size.

### Negative half lives and eigenvalues

The Table 3.3 has several negative half-lives; Equation 3.3 implies that this must be the result of a negative eigenvalue. [2] mentions the possibility of negative eigenvalues and negative half-lives, and explains it as possibly being because of character displacement in groups of geographically overlapping species. In this particular case, it might also possibly be because of character displacement in groups where most species are part of the same food web (whether or not they are actually geographically overlapping).

While character displacement is a tempting explanation for negative eigenvalues, it does stand in fundamental opposition to the model assumption of independence of stochastic pro-

cesses. Another possible explanation is that the corresponding eigenvector of the negative eigenvalue is an evolutionarily beneficial line, but that the optimum itself is a maladaptive point on that line which exerts a repellent influence on the evolution of nearby species.

**What can we say about causal relationships between the investigated variables?**

While the methods used in this investigation do not lend themselves to good explanations of biomechanical chicken-or-egg type discussion, there are some things that can be about the co-evolution of traits.

Of the models described in 3.1, only the last two have non-zero elements in the off-diagonals of the $\mathbf{F}$ matrix. This means that they are the only sub-model types that could have eigenvectors that overlap in some way. When the off-diagonals are symmetric, the eigenvectors will be orthogonal to one another and as long as they are not perfectly aligned with the coordinate system of the trait space these eigenvectors will each give rise to linear relationships between the traits. For the model that has fully assymetric off-diagonal elements, the eigenvectors will not generally be orthogonal and so there, too, linear relationships must arise.

In other words, for our two most complex models we can expect non-trivial relationships between the investigated traits. This could be due to certain trait configurations being more synergistic, or body-structurally more robust. I will not dwell further on this issue as it is beyond the scope of this thesis.

## 6.2 Method

**Data**

The algorithm used in this investigation is quite time-consuming, and there were several early mistakes in the preparation of the data set which, because of time constraints, could not be fixed later on.

In the preprocessing of the data for this study, I made no attempt to try and correct for differences in total and standard length within the data. The reasoning for this was that I presumed that the measurement types would have been chosen in ways that were appropriate for the body type of each species. While this may or may not be true, I have come to realize that if it were the case, then that would most likely mean that there would be an additional (and unnecessary) source of phylogenetic correlation in body length. This additional "false" phylogenetic covariance is human-imposed and very likely a confounding variable in our model.

Another issue with the data is the use of maximum measured body length for a species as its body length measure. The expected value of an extreme value distribution must necessarily be a function of the number of observed individual specimens of that species. It is likely that the number of observed specimens for each species is also phylogenetically correlated for a number of reasons; e.g. because they are confined to a geographical area that gets a high or low amount of attention from researchers, or because they are alike in how rare or common they are. Bias-free estimates could be calculated with estimates of the number of specimens measured for each species - this was not done in this investigation.

One small but egregious mistake in the data is that one species, *Limnothrissa miodon*, has a reported trophic level value of 1.64. A TL value lower than 2 is impossible for a living animal, and this must be the result of an error in FishBase. It is unlikely that this mistake has had any major impact on the results of this study, however.

**The choice of information criterion**

In this investigation, much like in [21] which developed the model that is used here, I have used the AIC criterion for model selection. It is well known that AIC tends to favor overfitted models when the ratio of parameters to data points is high [15], and for this reason I also present the AICc scores in 5.1. While the AIC and AICc values differ for the Mixed model, their relative rankings do not change and the choice of optimal model is the same no matter which information criterion one chooses.

However, in the calculations of the AICc values in Table 5.1 I have used $n = 1551$, i.e. the number of species in the data set. It is not entirely clear that this is what the value of $n$ in Equation 3.6 ought to be in our case since our data points are not independent. This is potentially a serious problem for this investigation.

## 6.3 Trophic Level

Trophic level, while heritable in some sense and closely related to the morphology of species' teeth and digestive systems, is not itself like morphological traits. It is an ecological fact rather than a physical one, and as such it is not inherent to an individual species but rather something which arises in the interplay of species in a food web. This entanglement means that no species's TL value can be estimated in isolation, and therefore covariances in estimates can be highly complex. It also means that the true TL values will change as ecosystems change, even if no morphological change occurs in species within those ecosystems.

The results in Chapter 4 show that the error covariance matrix for $n$ species in a food web will have a sum of $n_r^2$ terms each consisting of one distinct random scalar factor and one distinct non-random matrix. In a worst case scenario this means that the error covariance matrix will have every single one of its entries be independently random. The unfortunate implication of this is that one can not generally correct for "food web covariance" in trophic level estimation. It is therefore a real possibility that the relationships that we have found in this study are confounded by the way TL estimation errors are propagated throughout food web graphs.

While this conclusion may seem bleak, future studies could get around this issue in a number of ways. One way would be to disaggregate what aspect of trophic level is actually of interest, and then try to isolate it; one could imagine, for instance, an investigation similar to this one which rephrases the primary research goal as being "Are fish bigger than their prey?". Another way of getting around this issue is to get TL estimates that are derived from measures of biomagnifying chemicals for each individual species; see e.g. [24] for a description of how.

# Chapter 7
# Conclusion

In broad strokes, our results show that higher trophic level in fish that are mainly mainly feed on other fish is positively related to greater jaw size. These jaw sizes are in turn associated with greater body sizes, but the effect of body size on trophic level (when disaggregated from the effect of jaw size) is a lot weaker than that of jaw size, and sometimes even negative. This relationship is flipped for groups where high trophic levels is mainly associated with the consumption of insects - their jaws evolve instead to be smaller.

The separate investigation into the robustness of the measure of trophic level with regards to measurement errors has shown severe practical problems with the concept. In the general case any mismeasurement in diet proportions for any species can cause knock-on effects for any higher trophic level species in the food chain. If the exact source and magnitude of the perturbation is not known it can not be corrected for, and thus we must assume that there is an invisible and yet confounding source of food web covariance.

# Chapter 8
# Ethical Considerations

In the process of writing this thesis no personal data was collected or analysed in any way. The results herein contained are unlikely to be used for any malign purpose.

The process of calculating regime partitions is highly power intensive - a back of the envelope calculation indicates that the amount of power consumed for the calculations undertaken for this investigation would be equivalent to burning about 15 liters of gasoline. This is approximately the same amount as is needed to travel from Linköping to Stockholm in a car of average fuel efficiency.

I argue that this energy expenditure is justifiable given the societal value of this research.

# Bibliography

[1]     Hirotugu Akaike. "A new look at the statistical model identification". In: *IEEE Transactions on Automatic Control* 19.6 (1974), pp. 716–723. DOI: `10.1109/TAC.1974.1100705`.

[2]     Krzysztof Bartoszek, Jason Pienaar, Petter Mostad, Staffan Andersson, and Thomas F. Hansen. "A phylogenetic comparative method for studying multivariate adaptation". In: *Journal of Theoretical Biology* 314 (2012), pp. 204–215. ISSN: 0022-5193. DOI: `doi.org/10.1016/j.jtbi.2012.08.005`. URL: `https://www.sciencedirect.com/science/article/pii/S0022519312003918`.

[3]     Carl Boettiger, Duncan Temple Lang, and Peter Wainwright. "rfishbase: exploring, manipulating and visualizing FishBase data from R". In: *Journal of Fish Biology* (Nov. 2012). DOI: `doi.org/10.1111/j.1095-8649.2012.03464.x`.

[4]     Sébastien Brosse, Nicolas Charpin, Guohuan Su, Aurèle Toussaint, Guido A. Herrera-R, Pablo A. Tedesco, and Sébastien Villéger. "FISHMORPH: A global database on morphological traits of freshwater fishes". In: *Global Ecology and Biogeography* 30.12 (2021), pp. 2330–2336. DOI: `doi.org/10.1111/geb.13395`. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1111/geb.13395`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1111/geb.13395`.

[5]     Edward Burress, Jordan Holcomb, Karine Bonato, and Jonathan Armbruster. "Body size is negatively correlated with trophic position among cyprinids". In: *Royal Society Open Science* 3 (May 2016), p. 150652. DOI: `doi.org/10.1098/rsos.150652`.

[6]     Marguerite Butler and Aaron A. King. "Phylogenetic Comparative Analysis: A Modeling Approach for Adaptive Evolution". In: *The American Naturalist* 164 (Dec. 2004), pp. 683–695. DOI: `doi.org/10.1086/426002`.

[7]     Jonathan Chang, Daniel L. Rabosky, Stephen A. Smith, and Michael E. Alfaro. "An r package and online resource for macroevolutionary studies using the ray-finned fish tree of life". In: *Methods in Ecology and Evolution* 10.7 (2019), pp. 1118–1124. DOI: `doi.org/10.1111/2041-210X.13182`. eprint: `https://besjournals.onlinelibrary.wiley.com/doi/pdf/10.1111/2041-210X.13182`. URL: `https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.13182`.

[8]     Julien Clavel, Gilles Escarguel, and Gildas Merceron. "mvmorph: an r package for fitting multivariate evolutionary models to morphometric data". In: *Methods in Ecology and Evolution* 6.11 (2015), pp. 1311–1319. DOI: `doi.org/10.1111/2041-210X.12420`. eprint: `https://besjournals.onlinelibrary.wiley.com/doi/pdf/10.1111/2041-210X.12420`. URL: `https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.12420`.

[9]     Scott Doney, Mary Ruckelshaus, J. Duffy, James Barry, Francis Chan, Chad English, Heather Galindo, Jacqueline Grebmeier, Anne Hollowed, Nancy Knowlton, Jeffrey Polovina, N. Rabalais, William Sydeman, and Lynne Talley. "Climate Change Impacts

on Marine Ecosystems". In: *Annual Review of Marine Science* 4 (Jan. 2012), pp. 11–37. DOI: `doi.org/10.1146/annurev-marine-041911-111611`.

[10] Joseph Felsenstein. "Phylogenies and the Comparative Method". In: *The American Naturalist* 125.1 (1985), pp. 1–15. ISSN: 00030147, 15375323. URL: `http://www.jstor.org/stable/2461605`.

[11] Rainer M. Froese and Daniel V. Pauly (Editors). "FishBase". In: 2021. URL: `https://www.fishbase.org`.

[12] Thomas Hansen, Jason Pienaar, and Steven Orzack. "A Comparative Method for Studying Adaptation to a Randomly Evolving Environment". In: *Evolution; International Journal of Organic Evolution* 62 (July 2008), pp. 1965–1977. DOI: `doi.org/10.1111/j.1558-5646.2008.00412.x`.

[13] Bernhard Hausdorf. "Progress toward a General Species Concept". In: *Evolution* 65.4 (2011), pp. 923–931. DOI: `doi.org/10.1111/j.1558-5646.2011.01231.x`. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1558-5646.2011.01231.x`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1558-5646.2011.01231.x`.

[14] Anthony R. Ives and Matthew R. Helmus. "Generalized linear mixed models for phylogenetic analyses of community structure". In: *Ecological Monographs* 81.3 (2011), pp. 511–525. ISSN: 00129615. URL: `http://www.jstor.org/stable/23047059`.

[15] David Anderson Kenneth P. Burnham. *Model Selection and Multimodel Inference: A Practical Information-theoretic Approach*. 2nd ed. Springer, 2002, pp. 323–326. ISBN: 9780387953649,0387953647.

[16] Friedrich W. Keppeler, Carmen G. Montaña, and Kirk O. Winemiller. "The relationship between trophic level and body size in fishes depends on functional traits". In: *Ecological Monographs* 90.4 (2020), e01415. DOI: `doi.org/10.1002/ecm.1415`. eprint: `https://esajournals.onlinelibrary.wiley.com/doi/pdf/10.1002/ecm.1415`. URL: `https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1002/ecm.1415`.

[17] R. Keller Kopf, Jian Yen, Dale Nimmo, Sebastien Brosse, and Sébastien Villéger. "Global patterns and predictors of trophic position, body size and jaw size in fishes". In: *Global Ecology and Biogeography* 30 (Nov. 2020). DOI: `doi.org/10.1111/geb.13227`.

[18] Ernst Mayr. "What Is a Species, and What Is Not?" In: *Philosophy of Science* 63.2 (1996), pp. 262–277. ISSN: 00318248, 1539767X. URL: `http://www.jstor.org/stable/188473`.

[19] Kenneth S. Miller. "On the Inverse of the Sum of Matrices". In: *Mathematics Magazine* 54.2 (1981), pp. 67–72. ISSN: 0025570X, 19300980. URL: `http://www.jstor.org/stable/2690437`.

[20] Venelin Mitov, Krzysztof Bartoszek, Georgios Asimomitis, and Tanja Stadler. "Fast likelihood calculation for multivariate Gaussian phylogenetic models with shifts". In: *Theoretical Population Biology* 131 (2020), pp. 66–78. ISSN: 0040-5809. DOI: `https://doi.org/10.1016/j.tpb.2019.11.005`. URL: `https://www.sciencedirect.com/science/article/pii/S0040580919301947`.

[21] Venelin Mitov, Krzysztof Bartoszek, and Tanja Stadler. "Automatic generation of evolutionary hypotheses using mixed Gaussian phylogenetic models". In: *Proceedings of the National Academy of Sciences of the United States of America* 116 (34 2019), pp. 16921–16926. URL: `https://www.pnas.org/lookup/doi/10.1073/pnas.1813823116`.

[22] DAniel V. Pauly, Villy Christensen, Johanne Dalsgaard, Rainer M. Froese, and F.C. Torres. "Fishing Down Marine Food Webs". In: *Science (New York, N.Y.)* 279 (Mar. 1998), pp. 860–863. DOI: `doi.org/10.1126/science.279.5352.860`.

[23]   Daniel V. Pauly and Rainer M. Froese. *FishBase Food Items Table*. 2021. URL: `https://www.fishbase.se/manual/English/fishbasethe_food_items_table.htm` (visited on 05/21/2022).

[24]   David M. Post. "Using Stable Isotopes to Estimate Trophic Position: MODELS, METHODS, AND ASSUMPTIONS". In: *Ecology* 83.3 (2002), pp. 703–718. DOI: `doi.org/10.1890/0012-9658(2002)083[0703:USITET]2.0.CO;2`. eprint: `https://esajournals.onlinelibrary.wiley.com/doi/pdf/10.1890/0012-9658%282002%29083%5B0703%3AUSITET%5D2.0.CO%3B2`. URL: `https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1890/0012-9658%282002%29083%5B0703%3AUSITET%5D2.0.CO%3B2`.

[25]   Tamara Romanuk, April Hayward, and Jeffrey Hutchings. "Trophic level scales positively with body size in fishes". In: *Global Ecology and Biogeography* 20 (Mar. 2011), pp. 231–240. DOI: `doi.org/10.2307/41058238`.

[26]   George G. Simpson. "The Species Concept". In: *Evolution* 5.4 (1951), pp. 285–298. ISSN: 00143820, 15585646. URL: `http://www.jstor.org/stable/2405675`.

[27]   Silva Uusi-Heikkilä, Andrew R. Whiteley, Anna Kuparinen, Shuichi Matsumura, Paul A. Venturelli, Christian Wolter, Jon Slate, Craig R. Primmer, Thomas Meinelt, Shaun S. Killen, David Bierbach, Giovanni Polverino, Arne Ludwig, and Robert Arlinghaus. "The evolutionary legacy of size-selective harvesting extends from genes to populations". In: *Evolutionary Applications* 8.6 (2015), pp. 597–620. DOI: `doi.org/10.1111/eva.12268`. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1111/eva.12268`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1111/eva.12268`.

[28]   Wayne A. Wurtsbaugh, Craig Miller, Sarah Elizabeth Null, R. Justin DeRose, Peter R. Wilcock, Maura Hahnenberger, Frank P. Howe, and Johnnie N. Moore. "Decline of the world's saline lakes". In: *Nature Geoscience* 10 (2017), pp. 816–821.

# Appendix A
# Full phylogeny of all species with regime partition

Figure A.1: Dendrogram of all species with evolutionary regimes denoted by color

# Appendix B
# The Recursive Clade Partition Algorithm

The following is a pseudocode description of the RCP algorithm taken from page 5 of the supplemental material for [21].

**Input:**
$\mathcal{T}$: a timed tree with $M$ nodes of which $N$ are tips;
$\mathcal{X} \in (\mathbb{R} \cup \{NA, NaN\})_{k \times N}$: data for $k$ traits associated with the tips, missing values or non-existing traits allowed;
$\mathcal{M} \subset \mathcal{G}_k$, $|\mathcal{M}| < \infty$: a finite set of k-variate Gaussian phylogenetic models;
$MLE$: $\bigcup_{i \in \{0, N+1, \ldots, M-1\}} S_i(\mathcal{T}_i, \mathcal{M}) \longrightarrow \{< \ell\ell^*, \Theta^* >\}$: a maximum likelihood estimator getting as input a MGPM on (a subtree of) $\mathcal{T}$ and returning the corresponding maximum likelihood, $\ell\ell^*$, and point estimate, $\Theta^*$, of the parameters contained in $S$;
$SCORE$: $\{< S, \ell\ell >\} \longrightarrow \mathbb{R}$: a scoring function, penalizing a maximum likelihood value $\ell\ell$ based on the complexity (e.g. degrees of freedom) of $S$;
**Output:** A quasi-optimal MGPM, $S^* \in \{\mathcal{S}(\mathcal{T}, \mathcal{M})\}$, with respect to $SCORE$.

**Data:**
$TableFits$: a table with columns $tree$, $model$, $\Theta$ and $q$, containing the tree, the MGPM, the parameter-values and the penalized score for all MLEs produced during the search;
$QueuePartitionRoots$: a first-in-first-served list (queue) of the nodes used as clade-partition roots during the search;
$S^*$: the current MGPM on $\mathcal{T}$ with best score;

```
1   // Step 1. Initialization. Fit each individual model to each clade in T.
2   foreach i ∈ {0, N + 1, ..., M − 1} do
3       foreach m ∈ M do
4           S_{i,m} ← {< i, m >};
5           < ℓℓ*_{i,m}, Θ*_{i,m} > ← MLE(S_{i,m}; T_i, X_i, M);
6           q*_{i,m} ← SCORE(S_{i,m}, ℓℓ*_{i,m});
7           Add to TableFits ⟨tree = T_i, model = S_{i,m}, Θ = Θ*_{i,m}, q = q*_{i,m}⟩;

8   // Step 2. Recursive clade-partition search for the optimal MGPM on T.
9   // Step 2.1. Initialize QueuePartitionRoots with root-node and the best individual model fit to T found in TableFits.
10  Add to QueuePartitionRoots < 0 >;
11  S* ← {model in TableFits with the best score on the whole tree};
    // Main loop
12  while QueuePartitionRoots is not empty do
13      // Step 2.2. Get the node at the head of the queue: this node is the partition root for the iteration.
14      j ← PopFrontElement(QueuePartitionRoots);
15      // Step 2.3. Extract the subtree of T containing all tips descending from j without an intermediate node from S* on their path to j.
16      T'_j ← ExtractClade(T, j);
17      foreach l ∈ Nodes(S*)\{j} do
18          if l ∈ Nodes(T'_j) then
19              T'_j ← RemoveClade(T'_j, l);

20      PartitionNodes ← Nodes(T'_j);
21      // Step 2.4. Make a list of all shift configurations including Nodes(S*) and a node from PartitionNodes.
22      P ← φ;
23      foreach p ∈ PartitionNodes do
24          P ← P ⋃ {Nodes(S*) ∪ {p}};

25      // Step 2.5. Restrict the sets of candidate models (the pseudo-code below implements Heuristic B.2)
26      foreach l ∈ S* ∪ PartitionNodes\{j} do
27          M_l ← {best model fit to clade l} ∪ {model assigned to l in S*};

28      M_j ← M;
29      // Step 2.6. MLE fits to all shift configurations in P and possible model mappings using M_l, l ∈ P
30      foreach Sp ∈ P do
31          foreach Sm ∈ ∏_{l∈Sp} M_l do
32              S ← {< Sp, Sm >};
33              < ℓℓ*_S, Θ* > ← MLE(S; T, X, M);
34              q* ← SCORE(S, ℓℓ*_S);
35              Add to TableFits ⟨tree = T, model = S, Θ = Θ*, q = q*⟩;

36      // Step 2.7. If step 2.6 has found a fit with a better score, update S* and add its nodes to the queue.
37      if TableFits[tree == T, Min(q)] < SCORE(S*, ℓℓ_{S*}) then
38          S* ← BestModel(TableFits[tree == T]);
39          Add to QueuePartitionRoots Nodes(S*);

40  // Step 3. Round robin search for the optimal model type assignment to Nodes(S*).
41  // This step is optional and can be repeated a user-specified number of times.
42  foreach i ∈ Nodes(S*) do
43      foreach m ∈ M do
44          S*' ← S*\{< i, · >} ∪ {< i, m >};
45          < ℓℓ*', Θ*' > ← MLE(S*'; T, X, M);
46          q*' ← SCORE(S*', ℓℓ*');
47          if q*' < SCORE(S*, ℓℓ*) then
48              S* ← S*';

49  return S*;
```

# Appendix C
# Species sorted by regime

Eleginus gracilis
Forbesichthys agassizii
Aphredoderus sayanus
Galaxias eldoni
Galaxias pedderensis
Galaxias paucispondylus
Galaxias auratus
Galaxias occidentalis
Galaxias parvus
Neochanna burrowsius
Galaxiella munda
Neosalanx pseudotaihuensis
Salanx ariakensis
Plecoglossus altivelis altivelis
Osmerus eperlanus
Mallotus villosus
Hypomesus nipponensis
Prototroctes maraena
Esox masquinongy
Esox americanus vermiculatus
Umbra pygmaea
Gnathonemus petersii
Marcusenius cyprinoides
Pollimyrus isidori isidori
Mormyrus rume rume
Petrocephalus bovei bovei
Papyrocranus afer
Chitala chitala

Microgadus tomcod
Typhlichthys subterraneus
Percopsis omiscomaycus
Galaxias brevipinnis
Galaxias depressiceps
Galaxias prognathus
Galaxias maculatus
Paragalaxias mesotes
Galaxias olidus
Neochanna cleaveri
Galaxiella nigrostriata
Neosalanx anderssoni
Salangichthys microdon
Spirinchus thaleichthys
Osmerus mordax dentex
Hypomesus transpacificus
Retropinna retropinna
Esox lucius
Esox niger
Dallia pectoralis
Umbra krameri
Marcusenius macrolepidotus
Hippopotamyrus ansorgii
Pollimyrus adspersus
Brienomyrus brachyistius
Petrocephalus catostoma
Xenomystus nigri
Notopterus notopterus

Lota lota
Chologaster cornuta
Galaxias anomalus
Galaxias johnstoni
Galaxias vulgaris
Galaxias truttaceus
Galaxias rostratus
Paragalaxias dissimilis
Galaxias platei
Galaxias zebratus
Galaxiella pusilla
Salanx prognathus
Plecoglossus altivelis ryukyuensis
Thaleichthys pacificus
Osmerus mordax mordax
Hypomesus olidus
Retropinna semoni
Esox reichertii
Esox americanus americanus
Umbra limi
Lepidogalaxias salamandroides
Marcusenius mento
Hyperopisus bebe
Pollimyrus castelnaui
Mormyrops anguilloides
Gymnarchus niloticus
Chitala ornata
Scleropages jardinii

| | | |
|---|---|---|
| Scleropages leichardti | Osteoglossum bicirrhosum | Osteoglossum ferreirai |
| Scleropages formosus | Arapaima gigas | Heterotis niloticus |
| Hiodon tergisus | Hiodon alosoides | Anguilla bicolor bicolor |
| Anguilla bicolor pacifica | Anguilla obscura | Anguilla marmorata |
| Anguilla reinhardtii | Anguilla rostrata | Anguilla anguilla |
| Anguilla japonica | Anguilla dieffenbachii | Anguilla australis schmidtii |
| Anguilla australis australis | Elops saurus | Elops machnata |
| Elops hawaiensis | Megalops atlanticus | Megalops cyprinoides |
| Lepisosteus osseus | Lepisosteus oculatus | Lepisosteus platyrhincus |
| Atractosteus spatula | Atractosteus tropicus | Amia calva |
| Acipenser fulvescens | Acipenser persicus | Acipenser ruthenus |
| Acipenser stellatus | Acipenser nudiventris | Huso huso |
| Acipenser transmontanus | Acipenser medirostris | Huso dauricus |
| Acipenser oxyrinchus | Acipenser sturio | Polyodon spathula |
| Psephurus gladius | Polypterus delhezi | Polypterus senegalus |
| Polypterus palmas | Erpetoichthys calabaricus | Polypterus bichir |
| Polypterus endlicherii | Polypterus retropinni | |

Table C.1: The scientific names of species belonging to regime 1

| | | |
|---|---|---|
| Nemacheilus triangularis | Nemacheilus pallidus | Acanthopsoides hapalias |
| Acanthopsoides gracilentus | Acanthopsoides molobrion | Acantopsis dialuzona |
| Yasuhikotakia morleti | Yasuhikotakia modesta | Yasuhikotakia eos |
| Yasuhikotakia lecontei | Syncrossus beauforti | Syncrossus helodes |
| Syncrossus hymenophysa | Ambastaia sidthimunki | Leptobotia elongata |
| Gyrinocheilus aymonieri | Erimyzon sucetta | Erimyzon oblongus |
| Minytrema melanops | Moxostoma valenciennesi | Moxostoma anisurum |
| Moxostoma macrolepidotum | Moxostoma hubbsi | Moxostoma carinatum |
| Moxostoma erythrurum | Thoburnia rhothoeca | Ictiobus bubalus |
| Ictiobus cyprinellus | Carpiodes cyprinus | Carpiodes carpio |
| Platydoras costatus | Pterodoras granulosus | Oxydoras niger |
| Acanthodoras cataphractus | Ageneiosus ucayalensis | Ageneiosus inermis |
| Trachelyopterus galeatus | Epapterus dispilurus | Anadoras weddellii |
| Tatia intermedia | Helogenes marmoratus | Corydoras splendens |
| Callichthys callichthys | Ochmacanthus alternus | Vandellia cirrhosa |
| Trichomycterus areolatus | Trichomycterus corduvensis | Trichomycterus diabolus |
| Distichodus fasciolatus | Distichodus mossambicus | Distichodus sexfasciatus |
| Distichodus engycephalus | Distichodus rostratus | Nannocharax procatopus |
| Neolebias ansorgii | Citharinus citharus | Gymnotus carapo |
| Gymnotus ucamara | Gymnotus arapaima | Gymnotus mamiraua |
| Gymnotus curupira | Gymnotus jonasi | Apteronotus albifrons |
| Hypopomus artedi | Sternopygus macrurus | Eigenmannia trilineata |
| Chanos chanos | | |

Table C.2: The scientific names of species belonging to regime 2

| | | |
|---|---|---|
| Ethmalosa fimbriata | Nematalosa nasus | Nematalosa erebi |
| Tenualosa ilisha | Hilsa kelee | Dorosoma anale |
| Dorosoma cepedianum | Dorosoma petenense | Pellonula leonensis |
| Microthrissa congica | Limnothrissa miodon | Stolothrissa tanganicae |
| Potamothrissa obtusirostris | Sierrathrissa leonensis | Rhinosardinia amazonica |
| Platanichthys platana | Alosa immaculata | Alosa alosa |
| Alosa fallax | Alosa alabamae | Alosa sapidissima |
| Alosa pseudoharengus | Alosa mediocris | Alosa aestivalis |
| Alosa chrysochloris | Pellona flavipinnis | Pellona castelnaeana |
| Tenualosa toli | Clupea pallasii pallasii | Anchoa parva |
| Lycengraulis grossidens | Lycengraulis batesii | Pterengraulis atherinoides |
| Anchoviella lepidentostole | Amazonsprattus scintilla | Anchoviella guianensis |
| Lycothrissa crocodilus | Coilia dussumieri | Coilia nasus |
| Coilia lindmani | Coilia mystus | Setipinna phasa |
| Setipinna tenuifilis | | |

Table C.3: The scientific names of species belonging to regime 3

| | | |
|---|---|---|
| Oncorhynchus mykiss | Oncorhynchus clarkii | Oncorhynchus kisutch |
| Oncorhynchus tshawytscha | Oncorhynchus gorbuscha | Oncorhynchus keta |
| Oncorhynchus nerka | Oncorhynchus masou masou | Oncorhynchus masou formosanus |
| Salmo marmoratus | Salmo carpio | Salmo ohridanus |
| Salmo trutta | Salmo salar | Salvelinus confluentus |
| Salvelinus alpinus alpinus | Salvelinus malma | Salvelinus namaycush |
| Salvelinus fontinalis | Salvelinus leucomaenis pluvius | Salvelinus leucomaenis leucomaenis |
| Salvelinus leucomaenis imbrius | Brachymystax lenok | Hucho hucho |
| Hucho taimen | Hucho bleekeri | Thymallus arcticus |
| Thymallus thymallus | Thymallus grubii flavomaculatus | Thymallus grubii grubii |
| Coregonus oxyrinchus | Coregonus pidschian | Coregonus lavaretus |
| Coregonus clupeaformis | Coregonus ussuriensis | Coregonus chadary |
| Coregonus nasus | Coregonus albula | Coregonus peled |
| Coregonus sardinella | Coregonus zenithicus | Coregonus kiyi |
| Coregonus nigripinnis | Coregonus hoyi | Coregonus autumnalis |
| Coregonus pollan | Coregonus laurettae | Coregonus huntsmani |
| Stenodus leucichthys | Prosopium williamsoni | Prosopium cylindraceum |
| Prosopium coulterii | | |

Table C.4: The scientific names of species belonging to regime 4

Alburnus belvica  
Leucaspius delineatus  
Vimba vimba  
Leuciscus leuciscus  
Notemigonus crysoleucas  
Pelecus cultratus  
Chrosomus oreas  
Rhynchocypris oxycephalus  
Tribolodon brandtii  
Semotilus corporalis  
Margariscus margarita  
Gobio soldatovi  
Pseudorasbora parva  
Rhodeus amarus  
Parabramis pekinensis  
Hypophthalmichthys nobilis  
Ctenopharyngodon idella  
Paralaubuca barroni  
Oxygaster anomalura  
Zacco platypus  
Capoeta capoeta sevangi  
Luciobarbus bocagei  
Luciobarbus guiraonis  
Barbus cyclolepis  
Barbus haasi  
Schizothorax esocinus  
Gymnodiptychus dybowskii  
Labeobarbus gorguari  
Labeobarbus surkis  
Labeobarbus acutirostris  
Labeobarbus johnstonii  
Tor tambroides  
Tor tor  
Hypselobarbus micropogon  
Barbodes carnaticus  
Pseudobarbus asper  
Pethia conchonius  
Dawkinsia filamentosa  
Puntius brevis  
Hampala macrolepidota  
Hypsibarbus malcolmi  
Scaphognathops stejnegeri  
Puntioplites falcifer  
Puntioplites bulu  
Albulichthys albuloides  
Cyclocheilichthys apogon  
Carassius auratus  
Cyprinus carpio  
Psilorhynchus homaloptera  
Engraulicypris sardella  
Raiamas guttatus  
Luciosoma setigerum  

Alburnus alburnus  
Anaecypris hispanica  
Ballerus ballerus  
Leuciscus waleckii  
Pelasgus stymphalicus  
Acrocheilus alutaceus  
Chrosomus cumberlandensis  
Rhynchocypris percnurus  
Plagopterus argentissimus  
Hemitremia flammea  
Tanichthys albonubes  
Gobio gobio  
Hemibarbus maculatus  
Rhodeus sericeus  
Hemiculter leucisculus  
Hypophthalmichthys molitrix  
Paralaubuca riveroi  
Parachela oxygastroides  
Macrochirichthys macrochirus  
Leptobarbus hoevenii  
Capoeta capoeta capoeta  
Luciobarbus sclateri  
Luciobarbus graellsii  
Barbus plebejus  
Barbus meridionalis  
Schizothorax curvifrons  
Diptychus maculatus  
Labeobarbus dainellii  
Labeobarbus longissimus  
Labeobarbus aeneus  
Neolissochilus hexagonolepis  
Tor douronensis  
Tor putitora  
Hypselobarbus curmuca  
Pseudobarbus quathlambae  
Systomus sarana  
Pethia phutunio  
Puntius dorsalis  
Puntius bimaculatus  
Mystacoleucus marginatus  
Barbonymus gonionotus  
Discherodontus ashmeadi  
Puntioplites proctozystron  
Amblyrhynchichthys truncatus  
Balantiocheilos melanopterus  
Anematichthys repasson  
Carassius gibelio  
Catlocarpio siamensis  
Opsaridium zambezense  
Rastrineobola argentea  
Securicula gora  
Opsarius koratensi  

Alburnus chalcoides  
Vimba melanops  
Leuciscus idus  
Leuciscus aspius  
Pachychilon pictum  
Ptychocheilus oregonensis  
Chrosomus eos  
Pseudaspius leptocephalus  
Semotilus atromaculatus  
Couesius plumbeus  
Tinca tinca  
Gobiocypris rarus  
Acheilognathus majusculus  
Chanodichthys erythropterus  
Hemiculter bleekeri  
Mylopharyngodon piceus  
Paralaubuca typus  
Parachela maculicauda  
Opsariichthys uncirostris  
Capoeta damascina  
Luciobarbus comizo  
Luciobarbus microcephalus  
Luciobarbus graecus  
Barbus barbus  
Schizothorax richardsonii  
Schizothorax biddulphi  
Spinibarbus sinensis  
Labeobarbus crassibarbis  
Labeobarbus platydorsus  
Labeobarbus marequensis  
Neolissochilus stracheyi  
Tor sinensis  
Tor khudree  
Hypselobarbus jerdoni  
Pseudobarbus burgi  
Pethia ticto  
Pethia gelius  
Puntius chola  
Puntius vittatus  
Sikukia gudgeri  
Scaphognathops bandanensis  
Barbonymus altus  
Puntioplites waandersi  
Cyclocheilichthys enoplos  
Cosmochilus harmandi  
Procypris rabaudi  
Carassius carassius  
Probarbus jullieni  
Chelaethiops bibie  
Opsarius pulchellus  
Luciosoma bleekeri  

Table C.5: The scientific names of species belonging to regime 5

Tylochromis polylepis
Theraps irregularis
Thorichthys meeki
Amphilophus labiatus
Amphilophus lyonsi
Parachromis motaguensis
Cryptoheros spilurus
Cryptoheros panamensis
Australoheros facetus
Symphysodon aequifasciatus
Heros efasciatus
Mesonauta insignis
Hoplarchus psittacus
Aequidens patricki
Krobia guianensis
Andinoacara pulcher
Nannacara anomala
Crenicichla lepidota
Satanoperca leucosticta
Geophagus brasiliensis
Cichla orinocensis
Cichla monoculus
Heterochromis multidens
Salaria fluviatilis
Ellochelon vaigiensis
Myxus elongatus
Mugil cephalus
Ambassis buruensis
Ambassis macleayi
Parambassis ranga
Platichthys flesus
Centropomus undecimalis
Caranx sexfasciatus
Eleutheronema tetradactylum
Lates calcarifer
Betta splendens
Macropodus opercularis
Trichopodus pectoralis
Osphronemus exodon
Sandelia capensis
Channa orientalis
Channa melasoma
Channa maculata
Channa micropeltes
Pristolepis fasciata
Mastacembelus frenatus
Macrognathus siamensis
Mastacembelus erythrotaenia
Macrognathus pancalus
Perccottus glenii
Kurtus gulliveri
Scomberomorus sinensis
Etheostoma exile
Etheostoma olmstedi
Etheostoma rubrum

Tylochromis lateralis
Herichthys cyanoguttatus
Tomocichla tuba
Amphilophus citrinellus
Cichlasoma trimaculatum
Parachromis friedrichsthalii
Hypsophrys nicaraguensis
Amatitlania nigrofasciata
Caquetaia myersi
Symphysodon discus
Uaru amphiacanthoides
Hypselecara temporalis
Cichlasoma bimaculatum
Aequidens tetramerus
Andinoacara rivulatus
Acaronia vultuosa
Cleithracara maronii
Acarichthys heckelii
Satanoperca pappaterra
Astronotus ocellatus
Cichla intermedia
Chaetobranchus flavescens
Polycentrus schomburgkii
Cymatogaster aggregata
Aldrichetta forsteri
Mugil curema
Ambassis miops
Ambassis ambassis
Ambassis agrammus
Ambassis agassizii
Centropomus ensiferus
Centropomus parallelus
Polydactylus macrochir
Lates niloticus
Toxotes jaculatrix
Trichopsis pumila
Pseudosphromenus cupanus
Sphaerichthys osphromenoides
Ctenopoma petherici
Anabas testudineus
Channa punctata
Channa striata
Channa argus argus
Badis badis
Mastacembelus moorii
Mastacembelus shiranus
Macrognathus circumcinctus
Mastacembelus favus
Monopterus albus
Protogobius attiti
Syngnathus abaster
Etheostoma flabellare
Etheostoma blennioides
Etheostoma spectabile
Etheostoma jordani

Paraneetroplus bulleri
Herichthys minckleyi
Rocio octofasciata
Archocentrus centrarchus
Parachromis loisellei
Parachromis managuensis
Hypsophrys nematopus
Petenia splendida
Caquetaia kraussii
Heros severus
Mesonauta festivus
Hypselecara coryphaenoides
Cichlasoma amazonarum
Aequidens tubicen
Andinoacara coeruleopunctatus
Acaronia nassa
Crenicichla britskii
Guianacara owroewefi
Geophagus steindachneri
Cichla ocellaris
Cichla temensis
Chromidotilapia guntheri
Monocirrhus polyacanthus
Chelon labrosus
Joturus pichardi
Mugil liza
Ambassis gymnocephalus
Ambassis interrupta
Parambassis wolffii
Denariusa australis
Centropomus pectinatus
Caranx latus
Polynemus paradiseus
Lates microlepis
Toxotes chatareus
Trichopsis vittata
Trichopodus trichopterus
Osphronemus goramy
Ctenopoma muriei
Channa gachua
Channa lucius
Channa marulius
Channa asiatica
Nandus nandus
Mastacembelus cunningtoni
Macrognathus aral
Macrognathus maculatus
Mastacembelus armatus
Indostomus paradoxus
Glossamia aprion
Microphis brachyurus
Etheostoma caeruleum
Etheostoma nigrum
Etheostoma microperca
Percina maculata

Percina notogramma
Percina sciera
Percina shumardi
Percina copelandi
Percina caprodes
Ammocrypta pellucida
Zingel asper
Zingel zingel
Romanichthys valsanicola
Zingel streber
Sander lucioperca
Sander volgensis
Sander vitreus
Sander canadensis
Perca flavescens
Perca schrenkii
Perca fluviatilis
Gymnocephalus schraetser
Gymnocephalus acerina
Gymnocephalus baloni
Percarina demidoffii
Gymnocephalus cernua
Lepomis gibbosus
Lepomis megalotis
Lepomis auritus
Lepomis macrochirus
Lepomis humilis
Lepomis cyanellus
Lepomis gulosus
Micropterus punctulatus
Micropterus dolomieu
Micropterus salmoides
Micropterus notius
Ambloplites rupestris
Pomoxis annularis
Pomoxis nigromaculatus
Enneacanthus gloriosus
Enneacanthus obesus
Centrarchus macropterus
Elassoma evergladei
Siniperca chuatsi
Siniperca roulei
Coreoperca kawamebari
Nannoperca obscura
Nannoperca oxleyana
Nannoperca vittata
Nannatherina balstoni
Bostockia porosa
Macquaria ambigua
Macquaria australasica
Percilia irwini
Percichthys trucha
Maccullochella peelii
Maccullochella macquariensis
Gadopsis marmoratus
Gadopsis bispinosus
Syncomistes rastellus
Hephaestus epirrhinos
Hephaestus fuliginosus
Pingalla gilberti
Bidyanus bidyanus
Amniataba percoides
Hannia greenwayi
Amniataba caudavittata
Mesopristes argenteus
Rhynchopelates oxyrhynchus
Mesopristes cancellatus
Terapon jarbua
Kuhlia rupestris
Lateolabrax japonicus
Ammodytes americanus
Eucinostomus gula
Eucinostomus argenteus
Eucinostomus melanopterus
Gerres cinereus
Gerres filamentosus
Eugerres plumieri
Morone americana
Morone saxatilis
Morone chrysops
Dicentrarchus labrax
Pomadasys crocro
Pomadasys argenteus
Plectorhinchus gibbosus
Lutjanus argentimaculatus
Lutjanus jocu
Lutjanus griseus
Monodactylus argenteus
Acanthopagrus butcheri
Acanthopagrus berda
Lagodon rhomboides
Johnius borneensis
Cynoscion acoupa
Cynoscion steindachneri
Aplodinotus grunniens
Bairdiella chrysoura
Plagioscion montei
Plagioscion surinamensis
Plagioscion auratus
Drepane punctata
Scatophagus argus
Scatophagus tetracanthus
Carinotetraodon lorteti
Auriglobus modestus
Auriglobus nefastus
Tetraodon lineatus
Chelonodon patoca
Colomesus psittacus
Colomesus asellus

Table C.6: The scientific names of species belonging to regime 6

Devario malabaricus
Devario devario
Devario aequipinnatus
Devario laoensis
Devario kakhienensis
Devario browni
Rasbora paucisqualis
Danio rerio
Danio albolineatus
Danio kerri
Esomus metallicus
Esomus longimanus
Rasbora hobelmani
Rasbora paviana
Rasbora caudimaculata
Trigonostigma heteromorpha
Brevibora dorsiocellata
Rasbora trilineata
Rasbosoma spilocerca
Rasbora borapetensis
Rasbora dusonensis
Rasbora aurotaenia
Rasbora argyrotaenia
Rasbora einthovenii
Boraras urophthalmoides
Trigonopoma pauciperforatum
Rasbora daniconius
Amblypharyngodon melettinus

Table C.7: The scientific names of species belonging to regime 7

Odontostilbe fugitiva
Saccoderma melanostigma
Aphyocharax anisitsi
Roeboides descalvadensis
Bryconamericus iheringii
Mimagoniates rheocharis
Astyanax mexicanus
Oligosarcus pintoi
Astyanax magdalenae
Gymnocorymbus ternetzi
Hemigrammus marginatus
Hyphessobrycon megalopterus
Paracheirodon innesi
Hollandichthys multifasciatus
Astyanax scabripinnis
Acestrorhynchus microlepis
Bryconops melanurus
Brycon hilarii
Brycon pesu
Carnegiella marthae
Triportheus angulatus
Hoplias microlepis
Erythrinus erythrinus
Nannostomus beckfordi
Steindachnerina insculpta
Cyphocharax magdalenae
Prochilodus rubrotaeniatus
Semaprochilodus insignis
Leporinus fasciatus
Leporinus lacustris
Anostomus ternetzi
Serrasalmus irritans
Serrasalmus maculatus
Serrasalmus spilopleura
Pristobrycon calmoni
Piaractus mesopotamicus
Mylossoma aureum
Psectrogaster amazonica
Cynodon gibbus
Caenotropus labyrinthicus
Hydrocynus brevis
Alestes dentex
Brycinus lateralis
Phenacogrammus interruptus
Hepsetus odoe
Crenuchus spilurus

Serrapinnus notomelas
Pseudocheirodon terrabae
Roeboides bouchellei
Roeboexodon guyanensis
Bryconamericus stramineus
Mimagoniates microlepis
Astyanax bimaculatus
Oligosarcus hepsetus
Nematobrycon palmeri
Pristella maxillaris
Moenkhausia intermedia
Hyphessobrycon eques
Moenkhausia sanctaefilomenae
Hyphessobrycon flammeus
Acestrorhynchus altus
Acestrorhynchus heterolepis
Bryconops affinis
Brycon amazonicus
Salminus brasiliensis
Gasteropelecus sternicla
Triportheus albus
Hoplias aimara
Hoplerythrinus unitaeniatus
Ctenolucius hujeta
Potamorhina latior
Prochilodus nigricans
Prochilodus reticulatus
Schizodon nasutus
Leporinus octofasciatus
Leporinus striatus
Serrasalmus rhombeus
Serrasalmus elongatus
Serrasalmus medinai
Catoprion mento
Myloplus rubripinnis
Piaractus brachypomus
Mylossoma duriventre
Psectrogaster rutiloides
Rhaphiodon vulpinus
Hydrocynus forskahlii
Alestes macrophthalmus
Micralestes acutidens
Brycinus nurse
Brycinus longipinnis
Arnoldichthys spilopterus

Odontostilbe dialeptura
Nanocheirodon insignis
Roeboides myersii
Exodon paradoxus
Knodus moenkhausii
Astyanax aeneus
Oligosarcus jenynsii
Ctenobrycon hauxwellianus
Carlana eigenmanni
Hemigrammus erythrozonus
Hyphessobrycon erythrostigma
Paracheirodon axelrodi
Thayeria boehlkei
Hyphessobrycon herbertaxelrodi
Acestrorhynchus falcatus
Acestrorhynchus falcirostris
Brycon melanopterus
Henochilus wheatlandii
Carnegiella strigata
Gasteropelecus maculatus
Chalceus macrolepidotus
Hoplias malabaricus
Nannostomus espei
Cyphocharax voga
Potamorhina altamazonica
Prochilodus lineatus
Semaprochilodus taeniurus
Schizodon fasciatus
Leporinus friderici
Anostomus anostomus
Serrasalmus marginatus
Pygocentrus nattereri
Pygocentrus cariba
Pristobrycon striolatus
Ossubtus xinguense
Colossoma macropomum
Hemiodus immaculatus
Hydrolycus scomberoides
Parodon hilarii
Hydrocynus vittatus
Alestes baremoze
Hemigrammopetersius barnardi
Brycinus imberi
Brycinus macrolepidotus
Characidium fasciatum

Table C.8: The scientific names of species belonging to regime 8

Synodontis victoriae
Synodontis clarias
Synodontis leopardinus
Chrysichthys nigrodigitatus
Clarotes laticeps
Phyllonemus typus
Imparfinis minutus
Parailia congica
Parauchenoglanis monkei
Phractocephalus hemioliopterus
Luciopimelodus pati
Hypophthalmus edentatus
Brachyplatystoma juruense
Sorubim lima
Pseudoplatystoma corruscans
Platysilurus malarmo
Microglanis iheringi
Pimelodella cristata
Nemapteryx augusta
Cochlefelis danielsi
Tachysurus sinensis
Cephalocassis borneensis
Netuma thalassina
Sciades couma
Cathorops fuerthii
Noturus insignis
Pylodictis olivaris
Ameiurus melas
Ameiurus catus
Pangasius larnaudii
Pseudolais pleurotaenia
Pangasianodon hypophthalmus
Pangasianodon gigas
Plotosus canius
Glyptothorax fuscus
Bagarius suchus
Eutropiichthys vacha
Laides longibarbis
Pseudomystus siamensis
Mystus tengara
Mystus mysticetus
Mystus bocourti
Hemibagrus nemurus
Kryptopterus bicirrhis
Phalacronotus apogon
Wallago attu
Ompok malabaricus
Chaca bankanensis
Clarias gabonensis
Clarias theodorae
Bathyclarias worthingtoni
Heterobranchus longifilis
Clarias batrachus
Rita rita

Synodontis zambezensis
Synodontis alberti
Synodontis afrofischeri
Chrysichthys helicophagus
Auchenoglanis biscutatus
Lacantunia enigmatica
Cetopsorhamdia iheringi
Schilbe intermedius
Pimelodus blochii
Megalonema platycephalum
Pinirampus pirinampu
Hypophthalmus marginatus
Brachyplatystoma rousseauxii
Pseudoplatystoma fasciatum
Sorubimichthys planiceps
Leiarius marmoratus
Rhamdia quelen
Neoarius utarus
Neoarius graeffei
Neoarius leptaspis
Arius maculatus
Osteogeneiosus militaris
Netuma bilineata
Aspistor quadriscutis
Noturus stigmosus
Noturus flavus
Ictalurus punctatus
Ameiurus nebulosus
Pangasius conchophilus
Pangasius krempfi
Pseudolais micronemus
Pangasius sanitwongsei
Porochilus rendahli
Glyptothorax lampris
Bagarius bagarius
Amblyceps mangois
Silonia silondia
Tachysurus fulvidraco
Bagrichthys macropterus
Mystus multiradiatus
Mystus nigriceps
Bagrus bajad
Hemibagrus wyckioides
Kryptopterus macrocephalus
Kryptopterus limpok
Wallago leerii
Silurus asotus
Chaca chaca
Clarias angolensis
Clarias stappersii
Bathyclarias nyasensis
Clarias dussumieri
Clarias teijsmanni

Synodontis frontosus
Synodontis nebulosus
Malapterurus electricus
Chrysichthys auratus
Auchenoglanis occidentalis
Imparfinis mirini
Parailia pellucida
Schilbe mystus
Pimelodus maculatus
Calophysus macropterus
Pimelodina flavipinnis
Brachyplatystoma platynemum
Brachyplatystoma filamentosum
Pseudoplatystoma tigrinum
Zungaro zungaro
Batrochoglanis raninus
Gladioglanis machadoi
Cinetodus froggatti
Nedystoma dayi
Plicofollis dussumieri
Ketengus typus
Hemiarius stormii
Ariopsis seemanni
Aspistor luniscutis
Noturus miurus
Noturus gyrinus
Ictalurus furcatus
Ameiurus natalis
Pangasius bocourti
Pangasius macronema
Helicophagus waandersii
Pangasius pangasius
Tandanus tandanus
Glyptothorax laosensis
Bagarius yarrelli
Acrochordonichthys rugosus
Clupisoma garua
Pseudomystus stenomus
Mystus vittatus
Mystus atrifasciatus
Mystus singaringan
Sperata aor
Horabagrus brachysoma
Phalacronotus bleekeri
Hemisilurus mekongensis
Ompok bimaculatus
Silurus glanis
Clarias agboyiensis
Clarias werneri
Clarias gariepinus
Clarias ngamensis
Clarias macrocephalus
Heteropneustes fossilis

Table C.9: The scientific names of species belonging to regime 9

Akrokolioplax bicornis        Garra cambodgiensis          Osteochilus vittatus
Osteochilus microcephalus     Osteochilus lini             Osteochilus waandersii
Labiobarbus siamensis         Labiobarbus leptocheilus     Thynnichthys thynnoides
Cirrhinus molitorella         Lobocheilos melanotaenia     Henicorhynchus siamensis
Crossocheilus reticulatus     Crossocheilus oblongus       Labeo forskalii
Labeo parvus                  Labeo altivelis              Labeo rohita
Labeo dussumieri              Cirrhinus cirrhosus          Cirrhinus microlepis
Bangana dero

Table C.10: The scientific names of species belonging to regime 10

Ponticola kessleri            Ponticola gorlap             Babka gymnotrachelus
Proterorhinus marmoratus      Neogobius fluviatilis        Benthophilus stellatus
Caspiosoma caspium            Padogobius nigricans         Padogobius bonelli
Glossogobius celebius         Knipowitschia panizzae       Knipowitschia caucasica
Pomatoschistus microps        Bathygobius fuscus           Glossogobius giuris
Glossogobius aureus           Stiphodon elegans            Sicyopterus lagocephalus
Cotylopus acutipinnis         Awaous guamensis             Awaous banana
Stenogobius genivittatus      Ctenogobius boleosoma        Oligolepis acutipennis
Caragobius urolepis           Boleophthalmus boddarti      Periophthalmus argentilineatus
Periophthalmus barbarus       Gymnogobius castaneus        Gymnogobius urotaenia
Mugilogobius chulae           Tridentiger obscurus         Acanthogobius lactipes
Acanthogobius flavimanus      Pseudogobiopsis oligactis    Redigobius bikolanus
Rhinogobius giurinus          Oxyeleotris lineolata        Oxyeleotris marmorata
Bostrychus sinensis           Ophiocara porocephala        Butis butis
Butis koilomatodon            Eleotris picta               Eleotris pisonis
Eleotris fusca                Eleotris amblyopsis          Gobiomorphus cotidianus
Gobiomorphus breviceps        Gobiomorphus hubbsi          Philypnodon grandiceps
Gobiomorphus australis        Gobiomorus dormitor          Hemieleotris latifasciata
Dormitator latifrons          Dormitator maculatus         Hypseleotris compressa
Hypseleotris klunzingeri      Hypseleotris galii           Hypseleotris cyprinoides
Mogurnda adspersa             Mogurnda mogurnd

Table C.11: The scientific names of species belonging to regime 11

Pterolebias longipinnis
Cynodonichthys isthmensis
Cynodonichthys uroflammeus
Laimosemion xiphidius
Austrolebias bellottii
Fundulopanchax filamentosus
Epiplatys sexfasciatus sexfasciatus
Hyporhamphus quoyi
Hyporhamphus sajori
Strongylura timucu
Xenentodon cancila
Oryzias mekongensis
Oryzias sinensis
Melanotaenia gracilis
Melanotaenia nigrans
Melanotaenia splendida inornata
Glossolepis multisquamata
Pseudomugil signifer
Atherina boyeri
Craterocephalus stercusmuscarum stercusmuscarum
Craterocephalus marjoriae
Labidesthes sicculus
Atherinella chagresi
Odontesthes bonariensis
Odontesthes mauleanu

Anablepsoides hartii
Cynodonichthys hildebrandi
Laimosemion agilae
Kryptolebias marmoratus
Austrolebias nigripinnis
Callopanchax occidentalis
Aplocheilus panchax
Hyporhamphus limbatus
Strongylura marina
Strongylura krefftii
Dermogenys pusilla
Oryzias latipes
Melanotaenia splendida rubrostriata
Melanotaenia exquisita
Melanotaenia splendida splendida
Melanotaenia splendida tatei
Rhadinocentrus ornatus
Cairnsichthys rhombosomoides
Craterocephalus stercusmuscarum fulvus
Craterocephalus stramineus
Leptatherina wallacei
Atherinella hubbsi
Atherinella milleri
Odontesthes regia

Table C.12: The scientific names of species belonging to regime 12

Notropis hudsonius
Macrhybopsis storeriana
Nocomis micropogon
Campostoma oligolepis
Rhinichthys umatilla
Rhinichthys atratulus
Clinostomus elongatus
Phoxinus phoxinus

Erimystax x-punctatus
Platygobio gracilis
Nocomis biguttatus
Rhinichthys cataractae
Rhinichthys osculus
Exoglossum maxillingua
Clinostomus funduloides

Macrhybopsis aestivalis
Dionda episcopa
Campostoma anomalum
Rhinichthys falcatus
Rhinichthys obtusus
Richardsonius balteatus
Mylocheilus caurinus

Table C.13: The scientific names of species belonging to regime 13

Gambusia holbrooki          Gambusia affinis              Gambusia nicaraguensis
Gambusia sexradiata         Xiphophorus maculatus         Xiphophorus hellerii
Priapella compressa         Poeciliopsis turrubarensis    Priapichthys annectens
Brachyrhaphis cascajalensis Brachyrhaphis parismina       Xenophallus umbratilis
Phallichthys quadripunctatus Phallichthys tico            Brachyrhaphis roseni
Brachyrhaphis terrabensis   Brachyrhaphis rhabdophora     Brachyrhaphis holdridgei
Phallichthys amates         Alfaro cultratus              Heterandria formosa
Poecilia mexicana           Poecilia sphenops             Poecilia velifera
Poecilia latipinna          Poecilia caucana              Poecilia vivipara
Poecilia reticulata         Anableps anableps             Oxyzygonectes dovii
Cyprinodon nevadensis nevadensis  Jordanella floridae     Floridichthys carpio
Fundulus heteroclitus heroclitus  Fundulus grandis        Fundulus diaphanus diaphanus
Fundulus notatus            Allotoca diazi                Skiffia francesae
Xenoophorus captivus        Characodon lateralis          Aphanius vladykovi
Aphanius fasciatus          Aphanius iberus               Aphanius dispar dispar
Aphanius dispar richardsoni Aphanius mento                Valencia hispanica

Table C.14: The scientific names of species belonging to regime 14


Coptodon rendalli           Coptodon guineensis           Coptodon louka
Gobiocichla ethelwynnae     Gobiocichla wonderi           Tilapia sparrmanii
Tilapia ruweti              Oreochromis karongae          Oreochromis squamipinnis
Oreochromis shiranus        Alcolapia alcalica            Oreochromis variabilis
Oreochromis urolepis        Oreochromis schwebischi       Oreochromis tanganicae
Oreochromis andersonii      Oreochromis mossambicus       Oreochromis macrochir
Oreochromis leucostictus    Oreochromis esculentus        Oreochromis niloticus
Oreochromis aureus          Sarotherodon galilaeus        Sarotherodon melanotheron
Iranocichla hormuzensis     Hemichromis letourneuxi       Hemichromis bimaculatus
Hemichromis elongatus       Hemichromis fasciatus         Nanochromis parilus
Pelvicachromis pulcher

Table C.15: The scientific names of species belonging to regime 15


Myoxocephalus scorpius      Myoxocephalus thompsonii      Myoxocephalus quadricornis
Limnocottus godlewskii      Limnocottus bergianus         Batrachocottus multiradiatus
Cottocomephorus grewingkii  Cottocomephorus inermis       Leocottus kesslerii
Procottus jeittelesii       Paracottus knerii             Comephorus baikalensis
Batrachocottus baicalensis  Cottus confusus               Cottus bairdii
Cottus rhotheus             Cottus cognatus               Cottus gobio
Cottus ricei                Cottus asper                  Cottus aleuticus
Cottus poecilopus           Trachidermus fasciatus        Pungitius pungitius
Culaea inconstans           Apeltes quadracus             Gasterosteus wheatlandi
Gasterosteus aculeatus      Notesthes robusta             Pseudaphritis urvillii

Table C.16: The scientific names of species belonging to regime 16

Achondrostoma arcasii / Pseudochondrostoma polylepis / Pseudochondrostoma willkommii
Parachondrostoma toxostoma / Iberochondrostoma lemmingii / Chondrostoma nasus
Chondrostoma prespense / Protochondrostoma genei / Chondrostoma soetta
Telestes souffia / Alburnoides bipunctatus / Rutilus pigus
Rutilus rutilus / Abramis brama / Blicca bjoerkna
Rutilus frisii / Squalius pyrenaicus / Squalius carolitertii
Squalius lucumonis / Squalius svallize / Squalius cephalus
Petroleuciscus borysthenicus

Table C.17: The scientific names of species belonging to regime 17

Mchenga eucinostomus / Copadichromis jacksoni / Ctenopharynx pictus
Otopharynx lithobates / Lethrinops furcifer / Labeotropheus fuelleborni
Maylandia aurora / Aulonocara hansbaenschi / Lethrinops microdon
Lethrinops gossei / Lethrinops longipinnis / Maylandia xanstomachus
Cynotilapia afra / Melanochromis vermivorus / Melanochromis auratus
Aulonocara baenschi / Maylandia zebra / Alticorpus peterdaviesi
Aulonocara stuartgranti / Placidochromis milomo / Sciaenochromis fryeri
Cyathochromis obliquidens / Petrotilapia genalutea / Diplotaxodon greenwoodi
Diplotaxodon macrops / Diplotaxodon limnothrissa / Rhamphochromis esox
Rhamphochromis longiceps / Rhamphochromis macrophthalmus / Astatotilapia calliptera
Haplochromis riponianus / Haplochromis phytophagus / Haplochromis simpsoni
Haplochromis rufocaudalis / Haplochromis degeni / Haplochromis brownae
Haplochromis xenognathus / Haplochromis laparogramma / Haplochromis obliquidens
Haplochromis nubilus / Haplochromis nyererei / Haplochromis aeneocolor
Haplochromis macropsoides / Haplochromis petronius / Astatotilapia burtoni
Petrochromis fasciolatus / Petrochromis famula / Petrochromis macrognathus
Petrochromis polyodon / Petrochromis trewavasae / Simochromis diagramma
Lobochilotes labiatus / Pseudosimochromis curvifrons / Limnotilapia dardennii
Gnathochromis pfefferi / Tropheus moorii / Astatoreochromis alluaudi
Sargochromis giardi / Serranochromis macrocephalus / Serranochromis altus
Serranochromis angusticeps / Serranochromis robustus / Pseudocrenilabrus multicolor
Pseudocrenilabrus philander / Ophthalmotilapia nasuta / Ophthalmotilapia boops
Cyathopharynx furcifer / Ophthalmotilapia ventralis / Cardiopharynx schoutedeni
Lestradea stappersii / Ectodus descampsii / Callochromis melanostigma
Callochromis pleurospilus / Xenotilapia boulengeri / Xenotilapia ochrogenys
Xenotilapia flavipinnis / Xenotilapia rotundiventralis / Xenotilapia caudafasciata
Xenotilapia longispinis / Grammatotria lemairii / Tanganicodus irsacae
Eretmodus cyanostictus / Spathodus erythrodon / Bathybates leo
Bathybates vittatus / Bathybates ferox / Bathybates hornii
Bathybates fasciatus / Bathybates graueri / Bathybates minor
Hemibates stenosoma / Boulengerochromis microlepi

Table C.18: The scientific names of species belonging to regime 18

Cyprinella lutrensis        Cyprinella spiloptera          Cyprinella venusta
Cyprinella analostana       Opsopoeodus emiliae emiliae    Pimephales notatus
Pimephales promelas         Cyprinella callistia           Notropis blennius
Notropis texanus            Notropis anogenus              Notropis boops
Notropis heterodon          Hybognathus argyritis          Hybognathus regius
Hybognathus hankinsoni      Notropis stramineus            Notropis procne
Notropis dorsalis           Notropis rubellus              Notropis atherinoides
Notropis buchanani          Notropis volucellus            Notropis photogenis
Luxilus cornutus            Luxilus chrysocephalus         Notropis bifrenatus
Lythrurus ardens            Lythrurus umbratili

Table C.19: The scientific names of species belonging to regime 19

Lepidiolamprologus attenuatus    Lepidiolamprologus elongatus    Lepidiolamprologus profundicola
Lepidiolamprologus kendalli      Neolamprologus fasciatus        Neolamprologus wauthioni
Altolamprologus compressiceps    Lamprologus lemairii            Neolamprologus caudopunctatus
Neolamprologus brevis            Neolamprologus multifasciatus   Lamprologus signatus
Neolamprologus cylindricus       Neolamprologus mustax           Neolamprologus niger
Neolamprologus brichardi         Telmatochromis bifrenatus       Telmatochromis dhonti
Neolamprologus savoryi           Julidochromis ornatus           Julidochromis marlieri
Telmatochromis temporalis        Neolamprologus christyi         Neolamprologus petricola
Julidochromis transcriptus       Chalinochromis brichardi        Chalinochromis popelini
Julidochromis dickfeldi          Neolamprologus tetracanthus     Neolamprologus modestus
Lepidiolamprologus cunningtoni   Lamprologus callipterus         Neolamprologus leleupi
Neolamprologus toae              Neolamprologus mondabu          Neolamprologus furcifer
Neolamprologus tretocephalus     Xenotilapia bathyphila          Labidochromis vellicans

Table C.20: The scientific names of species belonging to regime 20

Plecodus straeleni          Perissodus microlepis          Perissodus eccentricus
Plecodus multidentatus      Haplotaxodon microlepis        Plecodus paradoxus
Plecodus elaviae            Xenochromis hecqui             Cyprichromis microlepidotus
Cyprichromis leptosoma      Paracyprichromis brieni        Paracyprichromis nigripinnis
Benthochromis melanoides    Benthochromis tricoti          Orthochromis mazimeroensis
Orthochromis mosoensis      Orthochromis kasuluensis       Orthochromis uvinzae
Reganochromis calliurus     Triglachromis otostigma        Cyphotilapia frontosa

Table C.21: The scientific names of species belonging to regime 21

Dimidiochromis compressiceps    Copadichromis borleyi            Lethrinops parvidens
Protomelas similis              Dimidiochromis kiwinge           Nimbochromis fuscotaeniatus
Copadichromis virginalis        Copadichromis mbenjii            Champsochromis spilorhynchus
Buccochromis lepturus           Lethrinops auritus               Fossorochromis rostratus
Cheilochromis euchilus          Chilotilapia rhoadesii           Placidochromis johnstoni
Mylochromis mola                Copadichromis quadrimaculatus    Protomelas taeniolatus
Nimbochromis venustus           Nimbochromis linni               Nimbochromis livingstonii
Trematocranus placodon          Hemitilapia oxyrhyncha           Nimbochromis polystigma
Buccochromis heterotaenia       Aristochromis christyi           Taeniolethrinops furcicauda
Taeniolethrinops praeorbitalis  Mylochromis epichorialis         Tramitichromis variabilis
Placidochromis electra          Otopharynx brooksi               Stigmatochromis woodi
Tyrannochromis macrostoma       Docimodus evelyna

Table C.22: The scientific names of species belonging to regime 22