Computer Lab 3
Bioinformatics

Linköpings Universitet, IDA, Statistik

2023 XII 01

Kurskod och namn:	732A51 Bioinformatics
Datum:	2023 XI 27—2023 XII 03 (lab session 1 XII 2023 SU10/11)
Delmomentsansvarig:	Krzysztof Bartoszek, Hao Chi Kinag
Instruktioner:	This computer laboratory is part of the examination for the Bioinformatics course
	Create a group report, on the solutions to the lab as a **.PDF** file.
	Be concise and do not include unnecessary printouts and figures produced by the software and not required in the assignments.
	All R code should be included as an appendix into your report.
	In the report reference **ALL** consulted sources and disclose **ALL** collaborations.
	The report should be handed in via LISAM
	(or alternatively in case of problems e–mailed to hao.chi.kiang@liu.se),
	by **23:59 3 December 2023** at latest.
	Notice there is a deadline for corrections **23:59 31 January 2024** and
	a final deadline of **23:59 29 February 2024** after which
	no submissions nor corrections will be considered and you will have to
	redo the missing labs at the next course opportunity.
	The report has to be written in English.
	Some questions are marked with an asterisk.
	From these **at least one** must be done, according to your choice.

Question 1

Using the script `http://ape-package.ird.fr/APER/APER2/SylviaWarblers.R` obtain the *Sylvia* warblers phylogeny (the script saves in in the file `sylvia_nj_k80.tre`). The geographical range data can be found in `http://ape-package.ird.fr/APER/APER2/sylvia_data.txt` and in the script is referenced as `DF$geo.range`. Notice that one tip is removed due to missing data

```
tr <- drop.tip(tr, "Chamaea_fasciata")
```

and the data has to be ordered by the tips of the phylogeny

```
DF <- sylvia.eco[tr$tip.label, ]
```

WARNING: Running the script bare might result in errors and very long running times. **Choose only the lines that you actually need!**

Question 1.1

Explain all the steps in the script required to obtain the phylogeny and trait data.

Question 1.2*

Analyze the discrete (`type=discrete`) geographical range variable (`DF$geo.range`) using `ape::ace`. Consider different models (parameter `model`). Report on the results and interpret the estimated rates and their standard errors.

Question 2

Install the `ade4` R package. Included with it you will find the carnivores dataset, `data(carni70)`.

Question 2.1

Explore the data set and report what can be found in it. Provide some plots.

Question 2.2*

Analyze the two traits (size and range) with a number of different phylogenetic comparative models. Look at the R packages `ape`, `mvMORPH`, `mvSLOUCH`, `ouch` and `slouch` (the last one is NOT on CRAN and can be downloaded from `https://github.com/kopperud/slouch`). Analyze the data under the following models

1. Both traits evolve as independent Brownian motions.

2. The traits evolve as a correlated Brownian motion.

3. Both traits evolve as independent Ornstein–Uhlenbeck processes.

4. The traits evolve as a bivariate Ornstein–Uhlenbeck process (use `mvMORPH` or `mvSLOUCH` but be careful and check under what assumptions the estimation is done).

5. `size` evolves as a Brownian motion and `range` as an Ornstein–Uhlenbeck process adapting to it (use `slouch` or `mvSLOUCH` and be careful about column order).

Compare the five models and say which one seems to fit better to the data. Summarize (in your own words, **NOT** just provide the printout, as it can be very long) the output under each model. If for some reason the running time will be prohibitive (more than 12 hours) report how long you ran the code and skip the model. Report on any problems you experienced in your study.

WARNING It is possible to have very long running times for correlated OUOU and OUBM models (the last two), especially on slow computers. Give yourself plenty of reserve time. If you do not manage to complete the last two models in time for submission, submit without them, reporting (with e.g. screenshots) how much time the estimation procedures were running for. Then, you will complete this part on a later mutually agreed on date.