**Computing Assignment – Large datasets, focus on online access and management in R**

Rob Salguero-Gómez – [rob.salguero@biology.ox.ac.uk](mailto:rob.salguero@biology.ox.ac.uk)                February 2025

Demonstrator: Oliver Spacey [–oliver.spacey@linacre.ox.ac.uk](mailto:–oliver.spacey@linacre.ox.ac.uk)

Sarah Bull – [sarah.bull@ballio.ox.ac.uk](mailto:sarah.bull@ballio.ox.ac.uk)

**Introduction**

The objective of this assignment is to help you gain more confidence and independence in accessing, wrangling, and exploring biological patterns across different online open-access repositories. You will learn how to download species demographic data from COMADRE ([www.compadre-db.org](http://www.compadre-db.org)), access IUCN RedList species status (<https://www.iucnredlist.org>), phylogenetic data from the open tree of life (<https://tree.opentreeoflife.org/>) and plot relationships between various biological traits on a phylogenetic context.

I have provided an R Markdown file that walks you through the whole exercise. This includes step-by-step guidance for each section. This script makes no assumptions about your level of R programming expertise, nor biological knowledge on specialised courses in year 3 in your curriculum.

The R MarkDown for the practical is: \*\*\*Large datasets, focus on online access and management in R.Rmd\*\*\*, and you will also need the phylogenetic tree \*\*\* COMPADRE-COMADRE\_Phylo\_June\_16\_2019.tre\*\*\*. In addition, if the IUCN API token does not work, please use the file IUCN\_comadre\_compadre.csv, which contains the IUCN Red List status of species in the COMPADRE & COMADRE databases.

Work through this and make sure you understand all the steps, ask a demonstrator if you have any questions. Then you are ready to tackle the assignment.

**Assignment**

Choose a different taxonomic subset of the animal (COMADRE; e.g. reptiles or birds, instead of mammals) or plant (COMPADRE database; e.g: pine trees, or Cactacea). Whichever group you choose, please make sure that you choose a group for which there are more than 10 unique species available. Further subset your data to make sure that the population data come from the wild (not captive), and are not constrained geographically (anywhere around the globe). For each task below, provide a brief comment (no more than 150 words) summarising the steps used to generate each the plot(s) and describing what the results show.

1. Show a histogram of the population growth rate, and a separate histogram of the generation time of your subset of species. Describe the patterns. Do they make sense to you?

2. Fit a model to explore whether the generation time and population growth rate of your species’ populations are correlated. Would you have expected them to be correlated? Why?

3. Using the IUCN API token provided (or via your own), or via the IUCN csv file provided explore, using statistical models, whether generation time, on the one hand, and population growth rate, on the other hand, predict the Red List conservation status of your species subset.

4. Plot the values of generation time and of population growth rate of your species on the phylogeny provided to you (COMPADRE-COMADRE\_Phylo\_June\_16\_2019.tre). Is there any visual evidence of phylogenetic inertia in those traits? Why?

5. Create a new variable called “Population\_performance”. Assign “Decline” or “Increase” to those species whose unique population growth rate is <1 or >1, respectively. This variable is no longer continuous (like population growth rate, lambda was in your original R MarkDown). Explore in this handy website (http://blog.phytools.org ) a way to plot that new trait on your phylogenetic tree. Then add another column to the phylogeny in such a way that you show the conservation status of those species (collected in previous steps from the IUCN). See picture below for reference/inspiration.

**Diagram

Description automatically generated**

**Submission requirements**

Write your report using R Markdown and submit both your .Rmd file containing the code and the .html produced by knitting the document. Anyone unsure about how to produce a report using R Markdown should run through the material on canvas from last term.

Aim to write your code so that it would run for a new pair of species simply by changing the input data at the start, *i.e.* using general rather than specific coding.

Please note: You should not be collaborating on your assignments, and the examiners will check code for evidence of plagiarism.

The marking scheme for computer assignments will be weighted equally between how well your code runs, how explicitly commented it is, the quality of the presentation of the report, and the understanding shown in your answers.