|  |  |
| --- | --- |
| **What’s in a name?**  **Taxonomy and nomenclature for naming biodiversity** | **A nuthatch perching on a tree CC BY SA**  CC BY SA: hedera.baltica |

## Objectives

Upon completion of this module, each student should be able to:

* Identify the components of a Latin binomial
* Differentiate between the taxonomic levels of family, genus, and species
* Interpret taxonomic synonyms, orthographic/spelling errors, and taxonomic uncertainty
* Compare taxonomic diversity between datasets
* Inspect objects and extract information from data frames in R
* Edit R scripts by substituting functions and objects into existing code
* Retrieve taxonomic information from a database using R

**Introduction**

Studies in ecology and evolution depend on the accurate and consistent naming of species. While this statement may seem obvious, the importance of names in biological sciences is often overlooked. Ecologists use species names to identify the organisms that are interacting, but do they consider the species concepts underlying the species in the communities they study? When trying to understand how distributions of species have changed over time, do evolutionary biologists remember to consider data records that are stored under old synonyms? Good science depends on scientists understanding taxonomy and nomenclature to both understand what an organism is and what groups it belongs to, and to be able to retrieve relevant information for that organism. **Taxonomy** is the study of grouping organisms based on a certain set of characteristics, while **Nomenclature** is the system of naming these groups.

Names are the foundation for every study in ecology and evolution and reconciling these names is a crucial step before undertaking a study. However, the importance of the well-established field of taxonomy is underappreciated and both taxonomy and nomenclature have recently been neglected in undergraduate education. In contrast, the surge in courses for teaching computational skills, including computer programming, illustrates the demand and utility of coding skills, regardless of the students’ career plans or degrees. In this module, the two fields are linked to provide students with two important but very different skills: coding and taxonomy.

In this module “What’s in a name?,” students are walked through how to use and understand code in the R programming language while downloading, manipulating, and interpreting taxonomic information from the Taxonomic Name Resolution Service. Assessment questions are provided throughout to gauge student learning and ability to apply their coding skills. At the end of this module, students will have a list of reconciled species names which may be carried through the next steps of a phylogenetic analysis, if completing the rest of this course, or other scientific study if continuing on their own.

For more background on Taxonomy and Nomenclature, please refer to the Background.pdf document in the R Project (Instructions folder).

Some of the instructions in this document are redundant with the instructions in the module itself. For clarity, this document provides an overview of the module and contains the summative assessment after completion of the module, while the instructions within the module provide step by step directions on how to complete the module.

**Activities**

Duration: ~3 hrs

**Introduction to R**

Open the R Project “Module 1: What’s in a name?” using R Studio (which should have been previously installed) or R Studio Cloud. Within the Instructions folder, open the PDF titled “Introduction\_to\_R.pdf” and within the Scripts folder, open the R script titled “Introduction\_to\_R.R”. Proceed through the instructions listed in the PDF by running the lines of code in your R script (instructions for how to do so are included in the PDF).

Assessment: Does your output look like the output in the PDF document? If not, why not?

*Tip: Until you are more comfortable with coding, do not make changes to the original lines of code provided in the script. Copy and paste the line to a new line and make changes there. Remember to add comments to your code using the # symbol so you know what each line does.*

**Taxonomy**

Within the same R Project (“Module 1: What’s in a name?”), within the Instructions folder, open the PDF titled “Taxonomy.pdf” and within the Scripts folder, open the R script titled “Taxonomy.R”. Proceed through the instructions listed in the PDF by running the lines of code in your R script (you should know how to do this from the **Introduction to R** activity). All activities and formative assessments are contained within the R Project, but the activities are summarized below.

In this module, you will:

Import a list of species into R from a csv file and generate objects containing the species or genus names of interest:

* There are five groups of taxa belonging to vertebrates (2), invertebrates (1), and plants (2). For this module, proceed with all five groups. For future modules, you will choose a subset of these groups for analysis.

Retrieve information from an online database using a specific function:

* The Taxonomic Name Resolution Service can be accessed by the function *tnrs\_match\_names()* in R or at <http://tnrs.iplantcollaborative.org/>

Inspect the objects to retrieve information about them:

* We can get the number of rows and columns in a table and retrieve specific rows or columns from the table in two different ways

Retrieve taxonomic information at different levels for specific inquiry names

* There are two functions that can do this in the module: *downstream()* which retrieves lower taxonomic levels to the level specified by *downto*, and *children()* which retrieves just the immediate lower taxonomic level (e.g., species if the input name is a genus, and genus if the input name is a family, as in the example in the code)

Manipulate an object to use only a portion of it:

* Subsetting an object can be done in a number of ways, including using the symbol “-“ to remove values specified

Retrieve species names for a genus:

* Here, the genus *Sitta* was listed without including the species belonging to this genus. The output here is a list of species belonging to that genus.
* The same process is applied to the conifer genera.

Manipulate dataframes to store certain columns:

* The structure of the names retrieved for conifers needs to be manipulated to match the structure of the other lists of species.
* Here, the function *c()* is used to make a single list composed of a single column from five different dataframes composed of each genus

Write objects to an output file:

* To save these data for future analyses, we write the names to a text file as a list, and save the output from the TNRS which includes OTT\_IDs, unique identifiers which are linked to taxa in the Open Tree of Life.

**References for Module:**

Acer rubrum example file: GBIF.org (30 September 2020) GBIF Occurrence Download <https://urldefense.proofpoint.com/v2/url?u=https-3A__doi.org_10.15468_dl.fjphs7&d=DwICAg&c=sJ6xIWYx-zLMB3EPkvcnVg&r=zYKA9Xw4U6lYd0Pw8jLsRA&m=yzr8InMQufdy8oPlm845f9KQDYYy7bzx9aQE7Aphm_s&s=o499rIbJnBqK7LaQbsDmT5owvBrrky6F1VTCsQq_fbY&e=>

Casimirri, G. “Problems with integrating Traditional Ecological Knowledge into contemporary resource management.” XII World Forestry Congress, 2003, Quebec City, Canada.

James R.N. Glasier, John H. Acorn, Scott E. Nielsen and Heather Proctor. 2013. Ants (Hymenoptera: Formicidae) of Alberta: A key to species based primarily on the worker caste. Canadian Journal of Arthropod Identification 22, 1-104 (doi:10.3752/cjai.2013.22).

“Learn about the Frogs of British Columbia.” *FrogWatch*, NatureWatch, 12 Nov. 2014, www.naturewatch.ca/frogwatch/british-columbia/.

Turner, Nancy. “Maple Trees in Canada.” *The Canadian Encyclopedia*, 7 Apr. 2009, www.thecanadianencyclopedia.ca/en/article/maple.

Turner, Nancy, Burton, Carla, and Van Eijk, Jan. “Plants in language and classification among BC First Nations.” *BC Studies no. 170*, Autumn 2013.

Photo in header:

hedera.baltica. “Nuthatch” Photo. Flickr. CC BY SA. URL: <https://www.flickr.com/photos/hedera_baltica/48634416618/> Accessed 10 Dec 2020.

**Additional Resources:**

International Code of Nomenclature for Algae, Fungi and Plants <https://www.iapt-taxon.org/nomen/main.php>

International Commission on Zoological Nomenclature <https://www.iczn.org/>

Taxonomic Name Resolution Service <http://tnrs.iplantcollaborative.org/>

Vancouver Island University Taxonomy Resources <https://library.viu.ca/c.php?g=188912&p=1247781>

**Instructor Support Materials:**

1. Instructions for teachers on how to navigate the module are included in the pdf “Guide for Instructors” located in the Instructions folder
2. Answer keys for formative and summative assessments are provided in the R Project for teachers in Admin/Answers/
3. Data files are included in the Output/Intermediate\_output/ folder for R objects at every stage of the analysis and the expected saved output files are in the Output/ folder (for the teacher’s version of the R Project)
4. Input data files are located in the Data/ folder
5. Students should receive the Student Version of the R Project because the Instructor Version contains answer keys
6. This Module should take between 2-3 hours, depending on comfort level of students and teacher, and whether the summative assessment is conducted during class or as a take-home activity

**Assessments**

For a summative assessment, have the students pick one of these two options:

1. Modify the Taxonomy.R script included for the Module to resolve taxonomic name issues for a list of species of their choice:
   1. Create a text file containing a list of names with one name per line
   2. Read the text file in using R code
   3. Run the tnrs function to resolve names
   4. List the names that were identified as synonyms (report the answer)
2. Modify the Taxonomy.R script included for the Module to retrieve names at a different taxonomic level than an input query
   1. Identify a higher-level taxonomic group to query
   2. Write the line of code to retrieve names at a lower taxonomic level
   3. Use a function to count how many names there are at the lower taxonomic level. How many are there? (report the answer)
   4. Save the output names to a file using R code

*Tip: Remember that Google (and StackOverflow) are your friends when it comes to coding. Also, borrow lines of code from what is available and modify it to work for your specific question.*