***README: This includes R script and datasets to reproduce the analyses from:***

**Gaoue, O. G., J. K. Moutouama , M A. Coe, M. O. Bond, E. Green, N. B. Sero, B. S. Bezeng, and K. Yessoufou 2021. Methodological advances for hypothesis-driven ethnobotany. *Biological Reviews*, in press.**

**1. R Scripts**

This includes 6 independent R scripts developed to show how to develop and implement advanced statistical models in ethnobiology.

**Code R1**. R script for generalized linear model and comparison with generalized linear mixed effect models.

**Code R2**. R script for using piecewiseSEM to develop structural equation model of the influence of socio-demographic traits and urbanization on local people knowledge of plants.

**Code R3**. R script for phylogenetic general least square (PGLS) testing the effect of controlling for species phylogenetic relatedness on the effects of species preference and therapeutic redundancy on use pressure. This script calls the “PhyloMaker.R” function which is included.

**Code R4**. R script for phylogenetic signal test for species use pressure, preference and therapeutic redundancy across plant species used by local people.

**Code R5**. R script for social network analysis using exponential random graph modeling (ERGM) to investigating the role of homophily in shaping knowledge distribution.

**Code R6**. R script to implement train/test split and cross-validation for ethnobiological data.

**2. Datasets**

To use these R scripts, you need the following datasets

**Dataset 1:** Data on the effect of socio-demographic traits and urbanization on local people knowledge of plants (data\_1\_ethno.csv). This data includes 144 observations for 9 variables including “nb\_species”, the number of plants freelisted by an individual. This dataset is used for analyses in Code R1 and R2.

**Dataset 2:** This dataset is used for analyses in Code R3 and R4.

* Data on the list of plant species and their family to build the phylogeny (data\_2\_phylo.csv)
* Phylogeny which is pruned from the general phylogeny (data\_3\_shipibo\_phylo.nex)
* Ethnobiological data for which one which to conduct PGLS (data\_4\_shipibo\_redundancy.csv).

To build the phylogeny used for the PGLS source data and function from Qian & Jin (2016). This includes

* The general phylogeny (Qian\_PhytoPhylo.tre)
* The nodes (Qian\_nodes.csv)
* The R function (Qian\_S.phyloMaker.R).

**Dataset 3:** This dataset is used for Code S4

To test for the phylogenetic signal you need

* Data on the phylogeny constructed for the plant list, here as an R script (data\_5\_shipibo\_tree.txt)
* Ethnobiological data that includes a column of the trait(s) or variable(s) for which one wants to test for signal (data\_4\_shipibo\_redundancy.csv)
* Ethnobiological data for the NRI and NTI tests (data\_6\_ethno\_shipibo.txt)

**Dataset 4**:  Data available at: Bond, Matthew; Gaoue, Orou (2020), Adjacency matrices and nodal attributes for prestige and homophily predict network structure for social learning of medicinal plant knowledge, Dryad, Dataset, https://datadryad.org/stash/dataset/doi:10.5061/dryad.cfxpnvx3q

* Directed and asymmetric adjacency matrix of incoming knowledge sharing (data\_7\_adj\_knowledge.csv)
* Undirected and symmetric adjacency matrix of spouses (data\_8\_adj\_married.csv)
* Node attributes (data\_9\_node\_attributes.csv)

**Dataset 5**:

* Census income data set (adult.csv) used here to illustrate how test/split analysis can be done in ethnobiology. This data is not included here but publicly available at: https://archive.ics.uci.edu/ml/datasets/census+income.