# CorrelatedGeneFamilies

**This is a repository for article "The nonrandom evolution of gene families" for *American Journal of Botany* - On the Nature of Things.**

by Claudio Casola and A. Michelle Lawing

This repository contains four elements, three folders and a ReadMe document, which are described below.

1. **ReadMe.doc**: This file is the read me document that describes the contents of the repository.
2. **data**: This folder includes data files used for the three examples presented in the manuscript.

* There are three files in this folder that contain the number of genes within each gene family and for each species selected to be included in our three examples. The gene family ID is listed in the first column of each of the files followed by a column for each species. Plant gene families were retrieved from Phytozome, v.11 (https://phytozome.jgi.doe.gov/pz/portal.html). Saccharomycotina (our fungi example) gene families were retrieved from MycoCosm (https://genome.jgi.doe.gov/programs/fungi/index.jsf).
  + rWGD\_8ssp\_angiosperms\_9655families.csv
  + aWGD\_8ssp\_angiosperms\_9603families.csv
  + 10spp\_fungi\_12107.csv
* There are three files that contain the phylogenetic tree for species within each of the examples. The flowering plant phylogeny was based on the tree topology available through Phytozome (phytozome.jgi.doe.gov/pz/phytoweb/C3BF350E58BF3CCC0E109FC9467CDE31.cache.html). Nodes were dated using TimeTree (www.timetree.org). The fungi phylogeny was modified from the tree published in Shen, X.X. et al. 2016. Reconstructing the Backbone of the Saccharomycotina Yeast Phylogeny Using Genome-Scale Data. G3 6:3927-3939. (file: 1233AA\_concatenation\_ML.tre). The original tree was pruned using the Wasabi platform (http://wasabiapp.org/).
  + rWGD\_8ssp\_angiosperms\_9655families\_tree.txt
  + aWGD\_8ssp\_angiosperms\_9603families\_tree.txt
  + 10species\_fungi\_tree.txt

1. **code**: This folder contains one script file (script.R) used to quantify correlation coefficients, conduct permutation tests, save results and derived data, and make figures. To run each of the three examples, the comments (#) should be added or removed, as per the instructions in the file. We uploaded results files for the two plant examples (see results folder) and permutations for the fungi example should be run to obtain the results (we will upload the result example shortly).
2. **results**: This folder contains saved result files for the two plant examples, so a user can directly load in the results that we obtained without having to wait for the permutations to run, which can take hours to days, depending on processor speed. The code to implement the permutations is saved and available to the user in the code folder in the script.R file.

* There are two files that contain a data matrix of Spearman R correlation coefficients.
  + rWGD\_8ssp\_angiosperms\_9655families\_0426\_dr
  + aWGD\_8ssp\_angiosperms\_9603families\_0426\_dr.Rdata
* There are two files that contain a data matrix of p-values associated with the Spearman R correlation coefficients, which are calculated using the permutation procedure.
  + rWGD\_8ssp\_angiosperms\_9655families\_0426\_dr\_p.Rdata
  + aWGD\_8ssp\_angiosperms\_9603families\_0426\_dr\_p.Rdata
* There are two histogram figures saved in this folder that correspond to Figure 2 panels (A) and (B).
  + rWGD\_8ssp\_angiosperms\_9655families\_0426\_histograms.pdf
  + aWGD\_8ssp\_angiosperms\_9603families\_0426\_histograms.pdf