toothmorph documentation

Elizabeth Sibert

May 23, 2018

This is documentation for the functions and code written and used for tooth morphology analysis, used in the manuscript "Two pulses of morphological diversification in Pacific pelagic fish following the Cretaceous-Paleogene Mass Extinction" by Elizabeth Sibert, Matt Friedman, Pincelli Hull, Gene Hunt, and Richard Norris. The full code, data, and R Workspaces with the directory structure can be downloaded here.

This documentation includes information about the scripts and functions, as well as occasional work-flow/examples. This set of code is what was used in the analysis of the Sibert et al manuscript, however it has been updated and generalized, and will be publishd as a standalone R package called "ichthyoliths" (in development as of 5/23/2018). Extensive documentation and updated code is available at https://github.com/esibert/ichthyoliths. Please note that this analysis can be carried out using the (in development) ichthyoliths package as well. The toothmorph repository and documentation is mostly for reviewers who are curious about the data and want to understand and replicate the analyses we did more fully. If you have come to this site wanting to do your own analyses, we HIGHLY recommend that you go to the ichthyoliths R package homepage and work from there. It is a more flexible and cleaner set of code, and has updated ichthyolith morphological information to include a wider variety of teeth observed.

Due to GitHub file-size restrictions, only the R code and this readme.pdf are included in this repository. To download the entire set of data, code, etc., please go to http://elizabethsibert.com/toothmorph.zip and download the zipped directory that includes code as well as input and output CSV files, and RData workspaces discussed here.

The code included in this repository includes:

- toothmorph_functions.R: a collection of (most of) the functions written specifically for this analysis. Some of these, which contain particularly important notes, are discussed in detail below, particularly the distances.clust() function.

- toothmorph_setup.R: the code used to import the morphological data and calculate disparity between the fossils. This script can also calculate NMDS for each dataset and generate the list of NMDS outputs. Do *not* "source" this code, as it will make your computer run the distances.clust function on the data, which can take days, depending on the size of your dataset - use with caution! This file produces distmat objects for disparity analysis as well as range chart ("strat column") objects for making range charts. The workspace of this script is found in 'RData/toothmorph_setup.RData'.

- toothmorph_nmds.R: code used to calculate NMDS ordinations for each of the datasets, for NMDS with 2, 3, 4, and 5 dimensions (respectively). This file also includes a lot of loops for generating different variations of visualization for each of the datasets and ordinations. The final figure code is included in the toothmorph_figures.R file, and is not in this document. The workspace from this script is found in 'RData/toothmorph_nmds.RData'.

- toothmorph_foote.R: code used to calculate evolutionary rates, using Foote (2000) Boundary Crossers estimates for origination and extinction. Output/workspace from this script is found in 'RData/toothmorph_rates.RData'. Note that toothmorph_rates.RData is a significantly cleaned up version of the other toothmorph_*.RData workspaces, as it doesn't include the large morphometric matrices.

- toothmorph_mark.R: code used to calculate Capture-Mark-Recapture based evolutionary rates. Note that the MARKPATH variable will likely be different on your computer, you should set it to be wherever the mark.exe file is saved. Additionally, output from MARK runs are saved into a directory 'mark', because they otherwise fill up the root directory. The mark directory is included in this git

repo so results can be called easily from within the script. The workspace for this script is found in 'RData/toothmorph_mark.RData'.

- toothmorph_figures.R: code used to make the final manuscript figures.

The Data (csv) files included in this R repository are:

- pairwisedist_*.csv: output from the distpairs function, saved for easy reading back into R. Read into the workspace in toothmorph_setup.R script.

- distmat_*.csv: output from the distpairs –> distmat functions, saved for easy reading back into R. Read into the workspace in toothmorph_setup.R script.

- Morph3.0_596_compiled_*.csv: The original morphological input matrices that include all image objects included in the study. Select rows were removed in the .lib and .con csv files as discussed in the supplementary information and tables S1 and S2.

- 596_IAR_IMAGES.csv: a CSV file containing counts of tooth and denticle abundance converted into accumulation rate. This data was previously published in Sibert and Norris 2015, PNAS, and is used for plotting tooth accumulation on the range charts.

- Morphotypes.csv: master character/codes for all identified morphotypes in the dataset

- Trait*.csv: distance matrices for each character coded in the dataset.

Note that throughout the code and within the input csv files, "lib" refers to a high-reworking scenario and "con" refers to a low-reworking scenario; "orig" refers to the original dataset.

The full analysis requires use of the following libraries:

```
library(doParallel)
library(stratigraph)
library(vegan)
library(viridis)
library(RMark)
library(Hmisc)
```