**Data processing and diversity analyses**

A Small-Scale Analysis of Elevational Species Richness and Beta Diversity Patterns of Arthropods on an Oceanic Island (Terceira, Azores)

Jan Peter Reinier de Vries, Emiel van Loon and Paulo A. V. Borges

Analyses are performed in the following order:

* 1: DataPreparation.R
* 2: TransectAnalyses.R
* 3: TotalAnalyses.R
* 4: run\_coverage\_by\_rarefaction.R
* 5: AlphaVisualisation.R
* 6: BetaVisualisation.R

1: DataPreparation.R

* *Input*: raw data (rawdata.csv) presenting the abundance counts per sampling site from autumn 2014 up to and including summer 2018.
* Only adult records are selected.
* 90-day sampling equivalents are derived.
* 3-month and 1-month sampling schemes are merged.
* Site or elevation band numbers and elevations are added.
* *Output*: Transect data, only summer records (Transectdata.csv) and all sites data over all seasons combined in elevation bands (Totaldata.csv).

2: TransectAnalyses.R

* *Input*: Transect data (Transectdata.csv), total abundance counts from the excluded orders Hymenoptera, Lepidoptera, Diptera) (Ordercounts\_HyLeDi.csv), raw data (rawdata.csv) in order to derive the total abundance counts for the included orders (adults only but no 90-day equivalent)
* Derive the total abundance counts per order for the transect sites over 2015 in order to assess the representativeness of included orders.
* Create a matrix table (TransectTable) of all species (abundance data) on the transect.
* Derive the elevational ranges of all species on the transect (Rangestats.csv).
* Derive the total species list on the transect (Specieslist.csv).
* Derive matrix tables per subgroup.
* Prepare matrix tables per elevation per subgroup: These tables are used as input for EstimateS to derive diversity indices, and to derive sampling coverage per elevation.
* Calculate the mean richness per sample for each subgroup (RichnessMean.csv).
* Derive the incidence-based and abundance-based beta-diversity (total beta-diversity plus its replacement and richness difference components) per subgroup.
* *Outputs*: order counts (OrderList1.csv, OrderList2.csv), species elevational ranges (Rangestats.csv), transect species list (Specieslist.csv), input files for EstimateS, mean richness per sample (RichnessMean.csv), beta-diversity tables (BetaInc\_Total.csv, BetaInc\_repl.csv, BetaInc\_rich.csv, BetaAbu\_Total.csv, BetaAbu\_repl.csv, BetaAbu\_rich.csv).

3: TotalAnalyses.R

* *Input*: Total data (Totaldata.csv).
* Create a matrix table (TotalTable) of all species (abundance data) on the elevation gradient.
* Derive matrix tables per subgroup.
* Run the script ‘run\_coverage\_by\_rarefaction.R’ to derive sampling coverage over rarified samples (15 samples per elevation band).
* Prepare matrix tables per elevation per subgroup, which are used as input for EstimateS to derive diversity indices.
* Calculate the mean richness per sample for each subgroup (RichnessMean\_Total.csv).
* *Outputs*: input files for EstimateS, mean richness per sample (RichnessMean\_Total.csv).

4: run\_coverage\_by\_rarefaction.R

This script presents two functions, that are used to calculate sampling coverage over rarified samples (15 samples per elevation band) in ‘TotalAnalyses.R’:

* Function ‘cosumsort’ calculates sums the abundance counts from all samples, sorts species by total (i.e. summed) abundance, and removes empty columns (zeros).
* Function ‘calc\_coverage’ calculates sampling coverage for each elevation band, rarified to a common sampling size of 15 samples.

5: AlphaVisualisation.R

* *Inputs*: elevational range data (Rangestats.csv), transect alpha-diversity indices (EstimateS\_Transect.csv), transect mean richness per sample (RichnessMean.csv), all sites alpha-diversity indices (EstimateS\_Total.csv), all sites mean richness per sample (RichnessMean\_Total.csv).
* Derive elevational range data and alpha-diversity data per subgroup.
* Derive linear models of alpha-diversity patterns.
* Visualize elevational range data.
* Visualize transect alpha-diversity patterns.
* Visualize all sites alpha-diversity patterns.
* Visualize Hill number patterns (q2-q4) over the elevation transect.
* *Outputs*: Alpha-diversity parameters over the elevation transect (Alphadiv\_Transect.csv) and over all sites combined in elevation bands (Alphadiv\_Total.csv), elevational ranges figure (rangefig.tiff), transect alpha-diversity figures (alphadiv.tiff, alphadiv\_Jack1.tiff, alphadiv\_Mean.tiff, Orders\_Indigenous.tiff), all sites alpha-diversity figures (alphadiv\_Total.tiff, alphadiv\_Total\_Jack1.tiff, alphadiv\_Total\_Mean.tiff), Hill number figures (H\_Shannon.tiff, H\_Simpson.tiff, H\_Bergerparker.tiff).

6: BetaVisualisation.R

* *Inputs*: Beta-diversity data (BetaInc\_Total.csv, BetaInc\_repl.csv, BetaInc\_rich.csv, BetaAbu\_Total.csv, BetaAbu\_repl.csv, BetaAbu\_rich.csv).
* Derive Multiple Regression by distance Matrices (MRM) models of beta-diversity distance decay.
* Visualize beta-diversity patterns.
* *Outputs*: Beta-diversity figures (betadiv\_all.tiff, betadiv\_order.tiff, betadiv\_status.tiff).