# Biodiverse Calculations and Indices

Shawn W Laffan

2026-02-06

# Table of contents

# 1 Overview

This is a listing of the indices available in Biodiverse, ordered by the calculations used to generate them. It is generated from the system metadata and contains all the information visible in the GUI, plus some additional details.

Most of the headings are self-explanatory. For the others:

- The *Subroutine* is the name of the subroutine used to call the function if you are using Biodiverse through a script.
- The *Index* is the name of the index in the SPATIAL_RESULTS list, or if it is its own list then this will be its name. These lists can contain a variety of values, but are usually lists of labels with some value, for example the weights used in an endemism calculation. The names of such lists typically end in "LIST", "ARRAY", "HASH", "LABELS" or "STATS".
- *Grouping?* states whether or not the index can be used to define the grouping for a cluster or region grower analysis. A blank value means it cannot be used for either.
- The *Minimum number of neighbour sets* dictates whether or not a calculation or index will be run. If you specify only one neighbour set then all those calculations that require two sets will be dropped from the analysis. (This is always the case for calculations applied to cluster nodes as there is only one neighbour set, defined by the set of groups linked to the terminal nodes below a cluster node). Note that many of the calculations lump neighbour sets 1 and 2 together. See the SpatialConditions page for more details on neighbour sets.

Note that some calculations can provide different numbers of indices depending on the nature of the BaseData set used. This currently applies to the hierarchically partitioned endemism calculations (both central and whole).

For space reasons, columns are not shown if all cells are empty.

# 2 Table of calculations and indices

*Generated GMT Sat Feb 7 06:48:46 2026, Biodiverse version 5.0.*

  - Element arrays
  - Element counts
  - Element lists
  - Label counts
  - Label counts not in sample
  - Local range lists
  - Local range summary statistics
  - Non-empty element counts
  - Rank relative sample counts per label
  - Redundancy
  - Richness
  - Sample count lists
  - Sample count quantiles
  - Sample count summary stats
- Matrix
  - Compare dissimilarity matrix values
  - Matrix statistics
  - Rao's quadratic entropy, matrix weighted
- Numeric Labels
  - Numeric label data
  - Numeric label dissimilarity
  - Numeric label harmonic and geometric means
  - Numeric label quantiles
  - Numeric label statistics
  - Numeric labels Gi* statistic
- PhyloCom Indices
  - NRI and NTI expected values
  - NRI and NTI, abundance weighted
  - NRI and NTI, local range weighted
  - NRI and NTI, unweighted
  - Net VPD expected values
  - Net variance of pair-wise phylogenetic distances, unweighted
  - Phylogenetic and Nearest taxon distances, abundance weighted
  - Phylogenetic and Nearest taxon distances, local range weighted
  - Phylogenetic and Nearest taxon distances, unweighted
- Phylogenetic Endemism Indices
  - Corrected weighted phylogenetic endemism
  - Corrected weighted phylogenetic endemism, central variant
  - Corrected weighted phylogenetic rarity

  - PD-Endemism
  - PE clade contributions
  - PE clade loss
  - PE clade loss (ancestral component)
  - Phylogenetic Endemism
  - Phylogenetic Endemism central
  - Phylogenetic Endemism central lists
  - Phylogenetic Endemism lists
  - Phylogenetic Endemism single
  - RWiBaLD
- Phylogenetic Indices
  - Count labels on tree
  - Evolutionary distinctiveness
  - Evolutionary distinctiveness per site
  - Evolutionary distinctiveness per terminal taxon per site
  - Labels not on tree
  - Labels on tree
  - Last shared ancestor properties
  - PD clade contributions
  - PD clade loss
  - PD clade loss (ancestral component)
  - Phylogenetic Abundance
  - Phylogenetic Diversity
  - Phylogenetic Diversity (local)
  - Phylogenetic Diversity node list
  - Phylogenetic Diversity terminal node count
  - Phylogenetic Diversity terminal node list
- Phylogenetic Indices (relative)
  - Labels not on trimmed tree
  - Labels on trimmed tree
  - Relative Phylogenetic Diversity, type 1
  - Relative Phylogenetic Diversity, type 2
  - Relative Phylogenetic Endemism, central
  - Relative Phylogenetic Endemism, type 1
  - Relative Phylogenetic Endemism, type 2
- Phylogenetic Turnover
  - Phylo Jaccard
  - Phylo Range weighted Turnover
  - Phylo S2
  - Phylo Sorenson

    - Phylogenetic ABC
- Rarity
    - Rarity central
    - Rarity central lists
    - Rarity whole
    - Rarity whole lists
- Richness estimators
    - ACE
    - Chao1
    - Chao2
    - Hurlbert richness estimation
    - ICE
- Taxonomic Dissimilarity and Comparison
    - Beta diversity
    - Bray-Curtis non-metric
    - Bray-Curtis non-metric, group count normalised
    - Jaccard
    - Kulczynski 2
    - Nestedness-resultant
    - Range weighted Sorenson
    - Rao's quadratic entropy, taxonomically weighted
    - S2
    - Simpson and Shannon
    - Sorenson

## 2.1 Element Properties

### 2.1.1 Group property Gi* statistics

**Description:** List of Getis-Ord Gi* statistics for each group property across both neighbour sets

**Subroutine:** calc_gpprop_gistar

**Reference:** Getis and Ord (1992) Geographical Analysis

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| GPPROP_GISTAR_LIST | List of Gi* scores | 1 |

### 2.1.2 Group property data

**Description:** Lists of the groups and their property values used in the group properties calculations. Returns one list for each property, so if your data have properties named 'GPROP1' and 'GPROP2' then it will return two lists named 'GPPROP_STATS_GPROP1_DATA' and 'GPPROP_STATS_GPROP2_DATA', respectively.

**Subroutine:** calc_gpprop_lists

**Indices:**

- Data set dependent

### 2.1.3 Group property hashes

**Description:** Hashes of the groups and their property values used in the group properties calculations. Hash keys are the property values, hash values are the property value frequencies. Returns one list for each property, so if your data have properties named 'GPROP1' and 'GPROP2' then it will return two lists named 'GPPROP_STATS_GPROP1_HASH' and 'GPPROP_STATS_GPROP2_HASH', respectively.

**Subroutine:** calc_gpprop_hashes

**Indices:**

- Data set dependent

### 2.1.4 Group property quantiles

**Description:** Quantiles for each group property across both neighbour sets

**Subroutine:** calc_gpprop_quantiles

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| GPPROP_QUANTILE_LIST | List of quantiles for the label properties (05 10 20 30 40 50 60 70 80 90 95) | 1 |

### 2.1.5 Group property summary stats

**Description:** List of summary statistics for each group property across both neighbour sets

**Subroutine:** calc_gpprop_stats

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| GPPROP_STATS_LIST | List of summary statistics (count mean min max median sum sd iqr) | 1 |

### 2.1.6 Label property Gi* statistics

**Description:** List of Getis-Ord Gi* statistic for each label property across both neighbour sets

**Subroutine:** calc_lbprop_gistar

**Reference:** Getis and Ord (1992) Geographical Analysis

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| LBPROP_GISTAR_LIST | List of Gi* scores | 1 |

### 2.1.7 Label property Gi* statistics (local range weighted)

**Description:** List of Getis-Ord Gi* statistic values for each label property across both neighbour sets (local range weighted)

**Subroutine:** calc_lbprop_gistar_abc2

**Reference:** Getis and Ord (1992) Geographical Analysis

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| LBPROP_GISTAR_LIST_ABC2 | List of Gi* scores | 1 |

### 2.1.8 Label property data

**Description:** Lists of the labels and their property values used in the label properties calculations. Returns one list for each property, so if your data have properties named 'PROP1' and 'PROP2' then it will return two lists named 'LBPROP_STATS_PROP1_DATA' and 'LBPROP_STATS_PROP1_DATA', respectively.

**Subroutine:** calc_lbprop_data

**Indices:**

- Data set dependent

### 2.1.9 Label property hashes

**Description:** Hashes of the labels and their property values used in the label properties calculations. Hash keys are the property values, hash values are the property value frequencies. Returns one hash for each property, so if your data have properties named 'PROP1' and 'PROP2' then it will return two lists named 'LBPROP_STATS_PROP1_HASH' and 'LBPROP_STATS_PROP2_HASH', respectively.

**Subroutine:** calc_lbprop_hashes

**Indices:**

- Data set dependent

### 2.1.10 Label property hashes (local range weighted)

**Description:** Hashes of the labels and their property values used in the local range weighted label properties calculations. Hash keys are the property values, hash values are the property value frequencies. Returns one hash for each property, so if your data have properties named 'PROP1' and 'PROP2' then it will return two lists named 'LBPROP_STATS_PROP1_HASH2' and 'LBPROP_STATS_PROP2_HASH2', respectively.

**Subroutine:** calc_lbprop_hashes_abc2

**Indices:**

- Data set dependent

### 2.1.11 Label property lists

**Description:** Lists of the labels and their property values within the neighbour sets. Returns one list for each property, so if your data have properties named 'PROP1' and 'PROP2' then it will return two lists named 'LBPROP_LIST_PROP1' and 'LBPROP_LIST_PROP2', respectively.

**Subroutine:** calc_lbprop_lists

**Indices:**

- Data set dependent

### 2.1.12 Label property quantiles

**Description:** List of quantiles for each label property across both neighbour sets

**Subroutine:** calc_lbprop_quantiles

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| LBPROP_QUANTILES | List of quantiles for the label properties: (05 10 20 30 40 50 60 70 80 90 95) | 1 |

### 2.1.13 Label property quantiles (local range weighted)

**Description:** List of quantiles for each label property across both neighbour sets (local range weighted)

**Subroutine:** calc_lbprop_quantiles_abc2

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| LBPROP_QUANTILES_ABC2 | List of quantiles for the label properties: (05 10 20 30 40 50 60 70 80 90 95) | 1 |

### 2.1.14 Label property summary stats

**Description:** List of summary statistics for each label property across both neighbour sets

**Subroutine:** calc_lbprop_stats

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| LBPROP_STATS | List of summary statistics (count mean min max median sum skewness kurtosis sd iqr) | 1 |

### 2.1.15 Label property summary stats (local range weighted)

**Description:** List of summary statistics for each label property across both neighbour sets, weighted by local ranges

**Subroutine:** calc_lbprop_stats_abc2

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| LBPROP_STATS_ABC2 | List of summary statistics (count mean min max median sum skewness kurtosis sd iqr) | 1 |

## 2.2 Endemism

### 2.2.1 Absolute endemism

**Description:** Absolute endemism scores.

**Subroutine:** calc_endemism_absolute

**Indices:**

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* |
|---|---|---|---|
| END_ABS1 | Count of labels entirely endemic to neighbour set 1 | | 1 |
| END_ABS1_P | Proportion of labels entirely endemic to neighbour set 1 | | 1 |
| END_ABS2 | Count of labels entirely endemic to neighbour set 2 | | 1 |
| END_ABS2_P | Proportion of labels entirely endemic to neighbour set 2 | | 1 |
| END_ABS_ALL | Count of labels entirely endemic to neighbour sets 1 and 2 combined | region grower | 1 |
| END_ABS_ALL_P | Proportion of labels entirely endemic to neighbour sets 1 and 2 combined | | 1 |

### 2.2.2 Absolute endemism lists

**Description:** Lists underlying the absolute endemism scores.

**Subroutine:** calc_endemism_absolute_lists

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| END_ABS1_LIST | List of labels entirely endemic to neighbour set 1 | 1 |
| END_ABS2_LIST | List of labels entirely endemic to neighbour set 1 | 1 |
| END_ABS_ALL_LIST | List of labels entirely endemic to neighbour sets 1 and 2 combined | 1 |

### 2.2.3 Endemism central

**Description:** Calculate endemism for labels only in neighbour set 1, but with local ranges calculated using both neighbour sets

**Subroutine:** calc_endemism_central

**Reference:** Crisp et al. (2001) J Biogeog; Laffan and Crisp (2003) J Biogeog

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* | *Formula* | *Reference* |
|---|---|---|---|---|
| ENDC_CWE | Corrected weighted endemism | 1 | $= \frac{ENDC\_WE}{ENDC\_RICHNESS}$ | |
| ENDC_RICH-NESS | Richness used in ENDC_CWE (same as index RICH-NESS_SET1) | 1 | | |
| ENDC_SINGLE | Endemism unweighted by the number of neighbours. Counts each label only once, regardless of how many groups in the neighbourhood it is found in. Useful if your data have sampling biases and best applied with a small window. | 1 | $= \sum_{t \in T} \frac{1}{R_t}$ where $t$ is a label (taxon) in the set of labels (taxa) $T$ in neighbour set 1, and $R_t$ is the global range of label $t$ across the data set (the number of groups it is found in, unless the range is specified at import). | Slatyer et al. (2007) J. Biogeog |

| *Index* | *Description* | *Minimum number of neighbour sets* | *Formula* | *Reference* |
|---|---|---|---|---|
| ENDC_WE | Weighted endemism | 1 | $= \sum_{t \in T} \frac{r_t}{R_t}$ where $t$ is a label (taxon) in the set of labels (taxa) $T$ in neighbour set 1, $r_t$ is the local range (the number of elements containing label $t$ within neighbour sets 1 & 2, this is also its value in list ABC2_LABELS_ALL), and $R_t$ is the global range of label $t$ across the data set (the number of groups it is found in, unless the range is specified at import). | |

### 2.2.4 Endemism central hierarchical partition

**Description:** Partition the endemism central results based on the taxonomic hierarchy inferred from the label axes. (Level 0 is the highest).

**Subroutine:** calc_endemism_central_hier_part

**Reference:** Laffan et al. (2013) J Biogeog

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| ENDC_HPART_0 | List of the proportional contribution of labels to the endemism central calculations, hierarchical level 0 | 1 |
| ENDC_HPART_1 | List of the proportional contribution of labels to the endemism central calculations, hierarchical level 1 | 1 |
| ENDC_HPART_C_0 | List of the proportional count of labels to the endemism central calculations (equivalent to richness per hierarchical grouping), hierarchical level 0 | 1 |
| ENDC_HPART_C_1 | List of the proportional count of labels to the endemism central calculations (equivalent to richness per hierarchical grouping), hierarchical level 1 | 1 |
| ENDC_HPART_E_0 | List of the expected proportional contribution of labels to the endemism central calculations (richness per hierarchical grouping divided by overall richness), hierarchical level 0 | 1 |
| ENDC_HPART_E_1 | List of the expected proportional contribution of labels to the endemism central calculations (richness per hierarchical grouping divided by overall richness), hierarchical level 1 | 1 |

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| ENDC_HPART_OME_0 | List of the observed minus expected proportional contribution of labels to the endemism central calculations , hierarchical level 0 | 1 |
| ENDC_HPART_OME_1 | List of the observed minus expected proportional contribution of labels to the endemism central calculations , hierarchical level 1 | 1 |

### 2.2.5 Endemism central lists

**Description:** Lists used in endemism central calculations

**Subroutine:** calc_endemism_central_lists

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| ENDC_RANGELIST | List of ranges for each label used in the endemism central calculations | 1 |
| ENDC_WTLIST | List of weights for each label used in the endemism central calculations | 1 |

### 2.2.6 Endemism central normalised

**Description:** Normalise the WE and CWE scores by the neighbourhood size. (The number of groups used to determine the local ranges).

**Subroutine:** calc_endemism_central_normalised

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* | *Formula* |
|---|---|---|---|
| ENDC_CWE_NORM | Corrected weighted endemism normalised by groups | 1 | $= \frac{ENDC\_CWE}{EL\_COUNT\_ALL}$ |
| ENDC_WE_NORM | Weighted endemism normalised by groups | 1 | $= \frac{ENDC\_WE}{EL\_COUNT\_ALL}$ |

### 2.2.7 Endemism whole

**Description:** Calculate endemism using all labels found in both neighbour sets

**Subroutine:** calc_endemism_whole

**Indices:**

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* | *Formula* | *Reference* |
|---|---|---|---|---|---|
| ENDW_CWE | Corrected weighted endemism | | 1 | $= \frac{ENDW\_WE}{ENDW\_RICHNESS}$ | |
| ENDW_RICHNESS | Richness used in ENDW_CWE (same as index RICHNESS_ALL) | region grower | 1 | | |

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* | *Formula* | *Reference* |
|---|---|---|---|---|---|
| ENDW_SINGLE | Endemism unweighted by the number of neighbours. Counts each label only once, regardless of how many groups in the neighbourhood it is found in. Useful if your data have sampling biases and best applied with a small window. | region grower | 1 | $= \sum_{t \in T} \frac{1}{R_t}$ where $t$ is a label (taxon) in the set of labels (taxa) $T$ across neighbour sets 1 & 2, and $R_t$ is the global range of label $t$ across the data set (the number of groups it is found in, unless the range is specified at import). | Slatyer et al. (2007) J. Biogeog |

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* | *Formula* | *Reference* |
|---|---|---|---|---|---|
| ENDW_WE | Weighted endemism | region grower | 1 | $= \sum_{t \in T} \frac{r_t}{R_t}$ where $t$ is a label (taxon) in the set of labels (taxa) $T$ across both neighbour sets, $r_t$ is the local range (the number of elements containing label $t$ within neighbour sets 1 & 2, this is also its value in list ABC2_LABELS_ALL), and $R_t$ is the global range of label $t$ across the data set (the number of groups it is found in, unless the range is specified at import). | |

### 2.2.8 Endemism whole hierarchical partition

**Description:** Partition the endemism whole results based on the taxonomic hierarchy inferred from the label axes. (Level 0 is the highest).

**Subroutine:** calc_endemism_whole_hier_part

**Reference:** Laffan et al. (2013) J Biogeog

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| ENDW_HPART_0 | List of the proportional contribution of labels to the endemism whole calculations, hierarchical level 0 | 1 |
| ENDW_HPART_1 | List of the proportional contribution of labels to the endemism whole calculations, hierarchical level 1 | 1 |
| ENDW_HPART_C_0 | List of the proportional count of labels to the endemism whole calculations (equivalent to richness per hierarchical grouping), hierarchical level 0 | 1 |
| ENDW_HPART_C_1 | List of the proportional count of labels to the endemism whole calculations (equivalent to richness per hierarchical grouping), hierarchical level 1 | 1 |
| ENDW_HPART_E_0 | List of the expected proportional contribution of labels to the endemism whole calculations (richness per hierarchical grouping divided by overall richness), hierarchical level 0 | 1 |
| ENDW_HPART_E_1 | List of the expected proportional contribution of labels to the endemism whole calculations (richness per hierarchical grouping divided by overall richness), hierarchical level 1 | 1 |

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| ENDW_HPART_OME_0 | List of the observed minus expected proportional contribution of labels to the endemism whole calculations , hierarchical level 0 | 1 |
| ENDW_HPART_OME_1 | List of the observed minus expected proportional contribution of labels to the endemism whole calculations , hierarchical level 1 | 1 |

### 2.2.9 Endemism whole lists

**Description:** Lists used in the endemism whole calculations

**Subroutine:** calc_endemism_whole_lists

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| ENDW_RANGELIST | List of ranges for each label used in the endemism whole calculations | 1 |
| ENDW_WTLIST | List of weights for each label used in the endemism whole calculations | 1 |

### 2.2.10 Endemism whole normalised

**Description:** Normalise the WE and CWE scores by the neighbourhood size. (The number of groups used to determine the local ranges).

**Subroutine:** calc_endemism_whole_normalised

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* | *Formula* |
|---|---|---|---|
| ENDW_CWE_NORM | Corrected weighted endemism normalised by groups | 1 | $= \frac{ENDW\_CWE}{EL\_COUNT\_ALL}$ |
| ENDW_WE_NORM | Weighted endemism normalised by groups | 1 | $= \frac{ENDW\_WE}{EL\_COUNT\_ALL}$ |

## 2.3 Hierarchical Labels

### 2.3.1 Ratios of hierarchical labels

**Description:** Analyse the diversity of labels using their hierarchical levels. The A, B and C scores are the same as in the Label Counts analysis (calc_label_counts) but calculated for each hierarchical level, e.g. for three axes one could have A0 as the Family level, A1 for the Family:Genus level, and A2 for the Family:Genus:Species level. The number of list elements generated depends on how many axes are used in the labels. Axes are order from zero as the highest level in the hierarchy, so index 0 is the top level of the hierarchy.

Note that this calculation produces lists since version 4.99_002 so one can no longer use the SUMRAT indices for clustering. This can be re-enabled if there is a need.

**Subroutine:** calc_hierarchical_label_ratios

**Reference:** Jones and Laffan (2008) Trans Philol Soc

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| HIER_A | A score for each level | 2 |
| HIER_ARAT | Ratio of A scores between adjacent levels | 2 |
| HIER_ASUM | Sum of shared label sample counts | 2 |
| HIER_ASUMRAT | 1 - Ratio of shared label sample counts between adjacent levels | 2 |
| HIER_B | B score for each level | 2 |
| HIER_BRAT | Ratio of B scores between adjacent levels | 2 |
| HIER_C | C score for each level | 2 |

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| HIER_CRAT | Ratio of C scores between adjacent levels | 2 |

## 2.4 Lists and Counts

### 2.4.1 Element arrays

**Description:** Arrays of elements used in neighbour sets 1 and 2. These form the basis for all the spatial calculations.

**Subroutine:** calc_element_lists_used_as_arrays

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| EL_ARRAY_ALL | Array of elements in both neighbour sets | 2 |
| EL_ARRAY_SET1 | Array of elements in neighbour set 1 | 1 |
| EL_ARRAY_SET2 | Array of elements in neighbour set 2 | 2 |

### 2.4.2 Element counts

**Description:** Counts of elements used in neighbour sets 1 and 2.

**Subroutine:** calc_elements_used

**Indices:**

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* |
|---|---|---|---|
| EL_COUNT_ALL | Count of elements in both neighbour sets | region grower | 1 |
| EL_COUNT_SET1 | Count of elements in neighbour set 1 | | 1 |
| EL_COUNT_SET2 | Count of elements in neighbour set 2 | | 2 |

### 2.4.3 Element lists

**Description:** [DEPRECATED] Lists of elements used in neighbour sets 1 and 2. These form the basis for all the spatial calculations. The return types are inconsistent. New code should use calc_element_lists_used_as_arrays

**Subroutine:** calc_element_lists_used

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| EL_LIST_ALL | List of elements in both neighour sets | 2 |
| EL_LIST_SET1 | List of elements in neighbour set 1 | 1 |
| EL_LIST_SET2 | List of elements in neighbour set 2 | 2 |

### 2.4.4 Label counts

**Description:** Counts of labels in neighbour sets 1 and 2. These form the basis for the Taxonomic Dissimilarity and Comparison indices.

**Subroutine:** calc_abc_counts

**Indices:**

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* |
|---|---|---|---|
| ABC_A | Count of labels common to both neighbour sets | region grower | 2 |
| ABC_ABC | Total label count across both neighbour sets (same as RICHNESS_ALL) | region grower | 2 |
| ABC_B | Count of labels unique to neighbour set 1 | | 2 |
| ABC_C | Count of labels unique to neighbour set 2 | | 2 |

### 2.4.5 Label counts not in sample

**Description:** Count of basedata labels not in either neighbour set (shared absence) Used in some of the dissimilarity metrics.

**Subroutine:** calc_d

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| ABC_D | Count of labels not in either neighbour set (D score) | 1 |

### 2.4.6 Local range lists

**Description:** Lists of labels with their local ranges as values. The local ranges are the number of elements in which each label is found in each neighour set.

**Subroutine:** calc_local_range_lists

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| ABC2_LABELS_ALL | List of labels in both neighbour sets | 2 |
| ABC2_LABELS_SET1 | List of labels in neighbour set 1 | 1 |
| ABC2_LABELS_SET2 | List of labels in neighbour set 2 | 2 |

### 2.4.7 Local range summary statistics

**Description:** Summary stats of the local ranges within neighour sets.

**Subroutine:** calc_local_range_stats

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| ABC2_MEAN_ALL | Mean label range in both element sets | 1 |
| ABC2_MEAN_SET1 | Mean label range in neighbour set 1 | 1 |
| ABC2_MEAN_SET2 | Mean label range in neighbour set 2 | 2 |

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| ABC2_SD_ALL | Standard deviation of label ranges in both element sets | 2 |
| ABC2_SD_SET1 | Standard deviation of label ranges in neighbour set 1 | 1 |
| ABC2_SD_SET2 | Standard deviation of label ranges in neighbour set 2 | 2 |

### 2.4.8 Non-empty element counts

**Description:** Counts of non-empty elements in neighbour sets 1 and 2.

**Subroutine:** calc_nonempty_elements_used

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| EL_COUNT_NONEMPTY_ALL | Count of non-empty elements in both neighbour sets | 1 |
| EL_COUNT_NONEMPTY_SET1 | Count of non-empty elements in neighbour set 1 | 1 |
| EL_COUNT_NONEMPTY_SET2 | Count of non-empty elements in neighbour set 2 | 2 |

### 2.4.9 Rank relative sample counts per label

**Description:** Find the per-group percentile rank of all labels across both neighbour sets, relative to the processing group. An absence is treated as a sample count of zero.

**Subroutine:** calc_label_count_quantile_position

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| LABEL_COUNT_RANK_PCT | List of percentile ranks for each label's sample count | 1 |

### 2.4.10 Redundancy

**Description:** Ratio of labels to samples. Values close to 1 are well sampled while zero means there is no redundancy in the sampling

**Subroutine:** calc_redundancy

**Reference:** Garcillan et al. (2003) J Veget. Sci

**Formula:** $= 1 - \frac{richness}{sum\ of\ the\ sample\ counts}$

**Indices:**

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* | *Formula* |
|---|---|---|---|---|
| REDUNDANCY_ALL | for both neighbour sets | region grower | 1 | $= 1 - \frac{RICHNESS\_ALL}{ABC3\_SUM\_ALL}$ |
| REDUNDANCY_SET1 | for neighour set 1 | | 1 | $= 1 - \frac{RICHNESS\_SET1}{ABC3\_SUM\_SET1}$ |
| REDUNDANCY_SET2 | for neighour set 2 | | 2 | $= 1 - \frac{RICHNESS\_SET2}{ABC3\_SUM\_SET2}$ |

### 2.4.11 Richness

**Description:** Count the number of labels in the neighbour sets

**Subroutine:** calc_richness

**Indices:**

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* |
|---|---|---|---|
| RICHNESS_ALL | for both sets of neighbours | region grower | 1 |
| RICHNESS_SET1 | for neighbour set 1 | | 1 |
| RICHNESS_SET2 | for neighbour set 2 | | 2 |

### 2.4.12 Sample count lists

**Description:** Lists of sample counts for each label within the neighbour sets. These form the basis of the sample indices.

**Subroutine:** calc_local_sample_count_lists

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| ABC3_LABELS_ALL | List of labels in both neighbour sets with their sample counts as the values. | 2 |
| ABC3_LABELS_SET1 | List of labels in neighbour set 1. Values are the sample counts. | 1 |
| ABC3_LABELS_SET2 | List of labels in neighbour set 2. Values are the sample counts. | 2 |

### 2.4.13 Sample count quantiles

**Description:** Quantiles of the sample counts across the neighbour sets.

**Subroutine:** calc_local_sample_count_quantiles

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| ABC3_QUANTILES_ALL | List of quantiles for both neighbour sets | 2 |
| ABC3_QUANTILES_SET1 | List of quantiles for neighbour set 1 | 1 |
| ABC3_QUANTILES_SET2 | List of quantiles for neighbour set 2 | 2 |

### 2.4.14 Sample count summary stats

**Description:** Summary stats of the sample counts across the neighbour sets.

**Subroutine:** calc_local_sample_count_stats

**Indices:**

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* |
|---|---|---|---|
| ABC3_MEAN_ALL | Mean of label sample counts across both element sets. | | 2 |
| ABC3_MEAN_SET1 | Mean of label sample counts in neighbour set1. | | 1 |
| ABC3_MEAN_SET2 | Mean of label sample counts in neighbour set 2. | | 2 |
| ABC3_SD_ALL | Standard deviation of label sample counts in both element sets. | | 2 |
| ABC3_SD_SET1 | Standard deviation of sample counts in neighbour set 1. | | 1 |
| ABC3_SD_SET2 | Standard deviation of label sample counts in neighbour set 2. | | 2 |
| ABC3_SUM_ALL | Sum of the label sample counts across both neighbour sets. | region grower | 2 |
| ABC3_SUM_SET1 | Sum of the label sample counts across both neighbour sets. | | 1 |
| ABC3_SUM_SET2 | Sum of the label sample counts in neighbour set2. | | 2 |

## 2.5 Matrix

### 2.5.1 Compare dissimilarity matrix values

**Description:** Compare the set of labels in one neighbour set with those in another using their matrix values. Labels not in the matrix are ignored. (This calculation assumes a matrix of dissimilarities and uses 0 as identical, so take care).

**Subroutine:** calc_compare_dissim_matrix_values

**Indices:**

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* |
|---|---|---|---|
| MXD_COUNT | Count of comparisons used. | | 2 |
| MXD_LIST1 | List of the labels used from neighbour set 1 (those in the matrix). The list values are the number of times each label was used in the calculations. This will always be 1 for labels in neighbour set 1. | | 2 |
| MXD_LIST2 | List of the labels used from neighbour set 2 (those in the matrix). The list values are the number of times each label was used in the calculations. This will equal the number of labels used from neighbour set 1. | | 2 |
| MXD_MEAN | Mean dissimilarity of labels in set 1 to those in set 2. | | 2 |
| MXD_VARIANCE | Variance of the dissimilarity values, set 1 vs set 2. | cluster metric | 2 |

### 2.5.2 Matrix statistics

**Description:** Calculate summary statistics of matrix elements in the selected matrix for labels found across both neighbour sets. Labels not in the matrix are ignored.

**Subroutine:** calc_matrix_stats

**Indices:**

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* |
|---|---|---|---|
| MX_KURT | Kurtosis | | 1 |
| MX_LABELS | List of the matrix labels in the neighbour sets | | 1 |
| MX_MAXVALUE | Maximum value | region grower | 1 |
| MX_MEAN | Mean | | 1 |
| MX_MEDIAN | Median | | 1 |
| MX_MINVALUE | Minimum value | | 1 |
| MX_N | Number of samples (matrix elements, not labels) | | 1 |
| MX_PCT05 | 5th percentile value | | 1 |
| MX_PCT25 | First quartile (25th percentile) | | 1 |
| MX_PCT75 | Third quartile (75th percentile) | | 1 |
| MX_PCT95 | 95th percentile value | | 1 |
| MX_RANGE | Range (max-min) | | 1 |
| MX_SD | Standard deviation | | 1 |
| MX_SKEW | Skewness | | 1 |
| MX_VALUES | List of the matrix values | | 1 |

### 2.5.3 Rao's quadratic entropy, matrix weighted

**Description:** Calculate Rao's quadratic entropy for a matrix weights scheme. BaseData labels not in the matrix are ignored

**Subroutine:** calc_mx_rao_qe

**Formula:** $= \sum_{i \in L} \sum_{j \in L} d_{ij} p_i p_j$ where $p_i$ and $p_j$ are the sample counts for the i'th and j'th labels, $d_{ij}$ is the matrix value for the pair of labels $ij$ and $L$ is the set of labels across both neighbour sets that occur in the matrix.

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| MX_RAO_QE | Matrix weighted quadratic entropy | 1 |

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| MX_RAO_TLABELS | List of labels and values used in the MX_RAO_QE calculations | 1 |
| MX_RAO_TN | Count of comparisons used to calculate MX_RAO_QE | 1 |

## 2.6 Numeric Labels

### 2.6.1 Numeric label data

**Description:** The underlying data used for the numeric labels stats, as an array. For the hash form, use the ABC3_LABELS_ALL index from the 'Sample count lists' calculation.

**Subroutine:** calc_numeric_label_data

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| NUM_DATA_ARRAY | Numeric label data in array form. Multiple occurrences are repeated based on their sample counts. | 1 |

### 2.6.2 Numeric label dissimilarity

**Description:** Compare the set of numeric labels in one neighbour set with those in another.

**Subroutine:** calc_numeric_label_dissimilarity

**Indices:**

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* | *Formula* |
|---|---|---|---|---|
| NUMD_ABSMEAN | Mean absolute dissimilarity of labels in set 1 to those in set 2. | cluster metric | 2 | $= \frac{\sum_{l_{1i} \in L_1} \sum_{l_{2j} \in L_2} abs(l_{1i} - l_{2j})(w_{1i} \times w_{2j})}{n_1 \times n_2}$ where $L1$ and $L2$ are the labels in neighbour sets 1 and 2 respectively, and $n1$ and $n2$ are the sample counts in neighbour sets 1 and 2 |
| NUMD_COUNT | Count of comparisons used. | | 2 | $= n1 * n2$ where values are as for $NUMD\_ABSMEAN$ |
| NUMD_VARIANCE | Variance of the dissimilarity values (mean squared deviation), set 1 vs set 2. | cluster metric | 2 | $= \frac{\sum_{l_{1i} \in L_1} \sum_{l_{2j} \in L_2} (l_{1i} - l_{2j})^2 (w_{1i} \times w_{2j})}{n_1 \times n_2}$ where values are as for $NUMD\_ABSMEAN$ |

### 2.6.3 Numeric label harmonic and geometric means

**Description:** Calculate geometric and harmonic means for a set of numeric labels.

**Subroutine:** calc_numeric_label_other_means

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| NUM_GMEAN | Geometric mean | 1 |
| NUM_HMEAN | Harmonic mean | 1 |

### 2.6.4 Numeric label quantiles

**Description:** Calculate quantiles from a set of numeric labels. Weights by samples so multiple occurrences are accounted for.

**Subroutine:** calc_numeric_label_quantiles

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| NUM_Q005 | 5th percentile | 1 |
| NUM_Q010 | 10th percentile | 1 |
| NUM_Q015 | 15th percentile | 1 |
| NUM_Q020 | 20th percentile | 1 |
| NUM_Q025 | 25th percentile | 1 |
| NUM_Q030 | 30th percentile | 1 |
| NUM_Q035 | 35th percentile | 1 |
| NUM_Q040 | 40th percentile | 1 |
| NUM_Q045 | 45th percentile | 1 |
| NUM_Q050 | 50th percentile | 1 |
| NUM_Q055 | 55th percentile | 1 |
| NUM_Q060 | 60th percentile | 1 |
| NUM_Q065 | 65th percentile | 1 |
| NUM_Q070 | 70th percentile | 1 |
| NUM_Q075 | 75th percentile | 1 |
| NUM_Q080 | 80th percentile | 1 |
| NUM_Q085 | 85th percentile | 1 |
| NUM_Q090 | 90th percentile | 1 |
| NUM_Q095 | 95th percentile | 1 |

### 2.6.5 Numeric label statistics

**Description:** Calculate summary statistics from a set of numeric labels. Weights by samples so multiple occurrences are accounted for.

**Subroutine:** calc_numeric_label_stats

**Indices:**

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* |
|---|---|---|---|
| NUM_CV | Coefficient of variation (NUM_SD / NUM_MEAN) | | 1 |
| NUM_KURT | Kurtosis | | 1 |
| NUM_MAX | Maximum value (100th quantile) | region grower | 1 |

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* |
|---|---|---|---|
| NUM_MEAN | Mean | | 1 |
| NUM_MIN | Minimum value (zero quantile) | | 1 |
| NUM_N | Number of samples | region grower | 1 |
| NUM_RANGE | Range (max - min) | | 1 |
| NUM_SD | Standard deviation | | 1 |
| NUM_SKEW | Skewness | | 1 |

### 2.6.6 Numeric labels Gi* statistic

**Description:** Getis-Ord Gi* statistic for numeric labels across both neighbour sets

**Subroutine:** calc_num_labels_gistar

**Reference:** Getis and Ord (1992) Geographical Analysis

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| NUM_GISTAR | List of Gi* scores | 1 |

## 2.7 PhyloCom Indices

### 2.7.1 NRI and NTI expected values

**Description:** Expected values used in the NRI and NTI calculations. Derived using a null model without resampling where each label has an equal probability of being selected (a null model of even distrbution). The expected mean and SD are the same for each unique number of labels across all neighbour sets. This means if you have three neighbour sets, each with three labels, then the expected values will be identical for each, even if the labels are completely different.

**Subroutine:** calc_nri_nti_expected_values

**Reference:** Webb et al. (2008) https://doi.org/10.1093/bioinformatics/btn358, Tsirogiannis et al. (2012)

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| PHYLO_NRI_NTI_SAMPLE_N | Number of random resamples used | 1 |
| PHYLO_NRI_SAMPLE_MEAN | Expected mean of pair-wise distances | 1 |
| PHYLO_NRI_SAMPLE_SD | Expected standard deviation of pair-wise distances | 1 |
| PHYLO_NTI_SAMPLE_MEAN | Expected mean of nearest taxon distances | 1 |
| PHYLO_NTI_SAMPLE_SD | Expected standard deviation of nearest taxon distances | 1 |

### 2.7.2 NRI and NTI, abundance weighted

**Description:** NRI and NTI for the set of labels on the tree in the sample. This version is -1* the Phylocom implementation, so values >0 have longer branches than expected. Abundance weighted.

**Subroutine:** calc_nri_nti3

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| PHYLO_NRI3 | Net Relatedness Index, abundance weighted | 1 |
| PHYLO_NTI3 | Nearest Taxon Index, abundance weighted | 1 |

### 2.7.3 NRI and NTI, local range weighted

**Description:** NRI and NTI for the set of labels on the tree in the sample. This version is -1* the Phylocom implementation, so values >0 have longer branches than expected. Local range weighted.

**Subroutine:** calc_nri_nti2

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| PHYLO_NRI2 | Net Relatedness Index, local range weighted | 1 |
| PHYLO_NTI2 | Nearest Taxon Index, local range weighted | 1 |

### 2.7.4 NRI and NTI, unweighted

**Description:** NRI and NTI for the set of labels on the tree in the sample. This version is -1* the Phylocom implementation, so values >0 have longer branches than expected. Not weighted by sample counts, so each label counts once only.

**Subroutine:** calc_nri_nti1

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* | *Formula* |
|---|---|---|---|
| PHYLO_NRI1 | Net Relatedness Index, unweighted | 1 | $NRI = \frac{MPD_{obs} - mean(MPD_{rand})}{sd(MPD_{rand})}$ |
| PHYLO_NTI1 | Nearest Taxon Index, unweighted | 1 | $NTI = \frac{MNTD_{obs} - mean(MNTD_{rand})}{sd(MNTD_{rand})}$ |

### 2.7.5 Net VPD expected values

**Description:** Expected values for VPD, analogous to the NRI/NTI results

**Subroutine:** calc_vpd_expected_values

**Reference:** Warwick & Clarke (2001)

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| PHYLO_NET_VPD_SAMPLE_MEAN | Expected mean of pair-wise variance (VPD) | 1 |
| PHYLO_NET_VPD_SAMPLE_N | Number of random resamples used to calculate expected pair-wise variance scores(will equal PHYLO_NRI_NTI_SAMPLE_N for non-ultrametric trees) | 1 |
| PHYLO_NET_VPD_SAMPLE_SD | Expected standard deviation of pair-wise variance (VPD) | 1 |

### 2.7.6 Net variance of pair-wise phylogenetic distances, unweighted

**Description:** Z-score of VPD calculated using NRI/NTI resampling Not weighted by sample counts, so each label counts once only.

**Subroutine:** calc_net_vpd

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| PHYLO_NET_VPD | Net variance of pair-wise phylogenetic distances, unweighted | 1 |

### 2.7.7 Phylogenetic and Nearest taxon distances, abundance weighted

**Description:** Distance stats from each label to the nearest label along the tree. Compares with all other labels across both neighbour sets. Weighted by sample counts (which currently must be integers)

**Subroutine:** calc_phylo_mpd_mntd3

**Reference:** Webb et al. (2008)

**Indices:**

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* | *Formula* |
|---|---|---|---|---|
| PMPD3_MAX | Maximum of pairwise phylogenetic distances | | 1 | |

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* | *Formula* |
|---|---|---|---|---|
| PMPD3_MEAN | Mean of pairwise phylogenetic distances | | 1 | $MPD = \frac{\sum_{t_i=1}^{n_t-1} \sum_{t_j=1}^{n_t} d_{t_i \leftrightarrow t_j}}{(n_t-1)^2}, i \neq j$ where $d_{t_i \leftrightarrow t_j} = \sum_{b \in B_{t_i \leftrightarrow t_j}} L_b$ is the sum of the branch lengths along the path connecting $t_i$ and $t_j$ such that $L_b$ is the length of each branch in the set of branches $B$ |
| PMPD3_MIN | Minimum of pairwise phylogenetic distances | | 1 | |
| PMPD3_N | Count of pairwise phylogenetic distances | | 1 | |
| PMPD3_RMSD | Root mean squared pairwise phylogenetic distances | | 1 | |

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* | *Formula* |
|---|---|---|---|---|
| PMPD3_VARIANCE | Variance of pairwise phylogenetic distances, similar to Clarke and Warwick (2001; http://dx.doi.org/10.3354/meps216265) but uses tip-to-tip distances instead of tip to most recent common ancestor. | | 1 | |
| PNTD3_MAX | Maximum of nearest taxon distances | region grower | 1 | |
| PNTD3_MEAN | Mean of nearest taxon distances | | 1 | |
| PNTD3_MIN | Minimum of nearest taxon distances | | 1 | |
| PNTD3_N | Count of nearest taxon distances | | 1 | |
| PNTD3_RMSD | Root mean squared nearest taxon distances | | 1 | |
| PNTD3_VARIANCE | Variance of nearest taxon distances | | 1 | |

### 2.7.8 Phylogenetic and Nearest taxon distances, local range weighted

**Description:** Distance stats from each label to the nearest label along the tree. Compares with all other labels across both neighbour sets. Weighted by sample counts

**Subroutine:** calc_phylo_mpd_mntd2

**Reference:** Webb et al. (2008)

**Indices:**

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* | *Formula* |
|---|---|---|---|---|
| PMPD2_MAX | Maximum of pairwise phylogenetic distances | | 1 | |
| PMPD2_MEAN | Mean of pairwise phylogenetic distances | | 1 | $MPD = \frac{\sum_{t_i=1}^{n_t-1} \sum_{t_j=1}^{n_t} d_{t_i \leftrightarrow t_j}}{(n_t-1)^2}, i \neq j$ where $d_{t_i \leftrightarrow t_j} = \sum_{b \in B_{t_i \leftrightarrow t_j}} L_b$ is the sum of the branch lengths along the path connecting $t_i$ and $t_j$ such that $L_b$ is the length of each branch in the set of branches $B$ |
| PMPD2_MIN | Minimum of pairwise phylogenetic distances | | 1 | |
| PMPD2_N | Count of pairwise phylogenetic distances | | 1 | |
| PMPD2_RMSD | Root mean squared pairwise phylogenetic distances | | 1 | |

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* | *Formula* |
|---|---|---|---|---|
| PMPD2_VARIANCE | Variance of pairwise phylogenetic distances, similar to Clarke and Warwick (2001; http://dx.doi.org/10.3354/meps216265) but uses tip-to-tip distances instead of tip to most recent common ancestor. | | 1 | |
| PNTD2_MAX | Maximum of nearest taxon distances | region grower | 1 | |
| PNTD2_MEAN | Mean of nearest taxon distances | | 1 | |
| PNTD2_MIN | Minimum of nearest taxon distances | | 1 | |
| PNTD2_N | Count of nearest taxon distances | | 1 | |
| PNTD2_RMSD | Root mean squared nearest taxon distances | | 1 | |
| PNTD2_VARIANCE | Variance of nearest taxon distances | | 1 | |

### 2.7.9 Phylogenetic and Nearest taxon distances, unweighted

**Description:** Distance stats from each label to the nearest label along the tree. Compares with all other labels across both neighbour sets.

**Subroutine:** calc_phylo_mpd_mntd1

**Reference:** Webb et al. (2008)

**Indices:**

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* | *Formula* |
|---|---|---|---|---|
| PMPD1_MAX | Maximum of pairwise phylogenetic distances | | 1 | |
| PMPD1_MEAN | Mean of pairwise phylogenetic distances | | 1 | $MPD = \frac{\sum_{t_i=1}^{n_t-1} \sum_{t_j=1}^{n_t} d_{t_i \leftrightarrow t_j}}{(n_t-1)^2}, i \neq j$ where $d_{t_i \leftrightarrow t_j} = \sum_{b \in B_{t_i \leftrightarrow t_j}} L_b$ is the sum of the branch lengths along the path connecting $t_i$ and $t_j$ such that $L_b$ is the length of each branch in the set of branches $B$ |
| PMPD1_MIN | Minimum of pairwise phylogenetic distances | | 1 | |
| PMPD1_N | Count of pairwise phylogenetic distances | | 1 | |
| PMPD1_RMSD | Root mean squared pairwise phylogenetic distances | | 1 | |

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* | *Formula* |
|---|---|---|---|---|
| PMPD1_VARIANCE | Variance of pairwise phylogenetic distances, similar to Clarke and Warwick (2001; http://dx.doi.org/10.3354/meps216265) but uses tip-to-tip distances instead of tip to most recent common ancestor. | | 1 | |
| PNTD1_MAX | Maximum of nearest taxon distances | region grower | 1 | |
| PNTD1_MEAN | Mean of nearest taxon distances | | 1 | |
| PNTD1_MIN | Minimum of nearest taxon distances | | 1 | |
| PNTD1_N | Count of nearest taxon distances | | 1 | |
| PNTD1_RMSD | Root mean squared nearest taxon distances | | 1 | |
| PNTD1_VARIANCE | Variance of nearest taxon distances | | 1 | |

## 2.8 Phylogenetic Endemism Indices

### 2.8.1 Corrected weighted phylogenetic endemism

**Description:** What proportion of the PD is range-restricted to this neighbour set?

**Subroutine:** calc_phylo_corrected_weighted_endemism

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* | *Formula* | *Reference* |
|---|---|---|---|---|
| PE_CWE | Corrected weighted endemism. This is the phylogenetic analogue of corrected weighted endemism. | 1 | $PE\_WE/PD$ | |

### 2.8.2 Corrected weighted phylogenetic endemism, central variant

**Description:** What proportion of the PD in neighbour set 1 is range-restricted to neighbour sets 1 and 2?

**Subroutine:** calc_pe_central_cwe

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| PEC_CWE | Corrected weighted phylogenetic endemism, central variant | 1 |
| PEC_CWE_PD | PD used in the PEC_CWE index. | 1 |

### 2.8.3 Corrected weighted phylogenetic rarity

**Description:** What proportion of the PD is abundance-restricted to this neighbour set?

**Subroutine:** calc_phylo_corrected_weighted_rarity

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* | *Formula* | *Reference* |
|---|---|---|---|---|
| PHYLO_RARITY_CWR | Corrected weighted phylogenetic rarity. This is the phylogenetic rarity analogue of corrected weighted endemism. | 1 | $AED_T/PD$ | |

### 2.8.4 PD-Endemism

**Description:** Absolute endemism analogue of PE. It is the sum of the branch lengths restricted to the neighbour sets.

**Subroutine:** calc_pd_endemism

**Reference:** See Faith (2004) Cons Biol.

**Indices:**

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* |
|---|---|---|---|
| PD_ENDEMISM | Phylogenetic Diversity Endemism | region grower | 1 |
| PD_ENDEMISM_P | Phylogenetic Diversity Endemism, as a proportion of the whole tree | region grower | 1 |
| PD_ENDEMISM_WTS | Phylogenetic Diversity Endemism weights per node found only in the neighbour set | | 1 |

### 2.8.5 PE clade contributions

**Description:** Contribution of each node and its descendents to the Phylogenetic endemism (PE) calculation.

**Subroutine:** calc_pe_clade_contributions

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| PE_CLADE_CONTR | List of node (clade) contributions to the PE calculation | 1 |
| PE_CLADE_CONTR_P | List of node (clade) contributions to the PE calculation, proportional to the entire tree | 1 |
| PE_CLADE_SCORE | List of PE scores for each node (clade), being the sum of all descendent PE weights | 1 |

### 2.8.6 PE clade loss

**Description:** How much of the PE would be lost if a clade were to be removed? Calculates the clade PE below the last ancestral node in the neighbour set which would still be in the neighbour set.

**Subroutine:** calc_pe_clade_loss

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| PE_CLADE_LOSS_CONTR | List of the proportion of the PE score which would be lost if each clade were removed. | 1 |
| PE_CLADE_LOSS_CONTR_P | As per PE_CLADE_LOSS but proportional to the entire tree | 1 |
| PE_CLADE_LOSS_SCORE | List of how much PE would be lost if each clade were removed. | 1 |

### 2.8.7 PE clade loss (ancestral component)

**Description:** How much of the PE clade loss is due to the ancestral branches? The score is zero when there is no ancestral loss.

**Subroutine:** calc_pe_clade_loss_ancestral

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| PE_CLADE_LOSS_ANC | List of how much ancestral PE would be lost if each clade were removed. The value is 0 when no ancestral PE is lost. | 1 |
| PE_CLADE_LOSS_ANC_P | List of the proportion of the clade's PE loss that is due to the ancestral branches. | 1 |

### 2.8.8 Phylogenetic Endemism

**Description:** Phylogenetic endemism (PE). Uses labels across both neighbourhoods and trims the tree to exclude labels not in the BaseData object.

**Subroutine:** calc_pe

**Reference:** Rosauer et al (2009) Mol. Ecol; Laity et al. (2015); Laffan et al. (2016)

**Formula:** $PE = \sum_{\lambda \in \Lambda} L_\lambda \frac{r_\lambda}{R_\lambda}$ where $\Lambda$ is the set of branches found across neighbour sets 1 and 2, $L_\lambda$ is the length of branch $\lambda$ , $r_\lambda$ is the local range of branch $\lambda$ (the number of groups in neighbour sets 1 and 2 containing it), and $R_\lambda$ is the global range of branch $\lambda$ (the number of groups across the entire data set containing it).

**Indices:**

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* | *Formula* |
|---|---|---|---|---|
| PE_WE | Phylogenetic endemism | region grower | 1 | |

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* | *Formula* |
|---|---|---|---|---|
| PE_WE_P | Phylogenetic weighted endemism as a proportion of the total tree length | region grower | 1 | $PE\_WE/L$ where L is the sum of all branch lengths in the trimmed tree |

### 2.8.9 Phylogenetic Endemism central

**Description:** A variant of Phylogenetic endemism (PE) that uses labels from neighbour set 1 but local ranges from across both neighbour sets 1 and 2. Identical to PE if only one neighbour set is specified.

**Subroutine:** calc_pe_central

**Reference:** Rosauer et al (2009) Mol. Ecol

**Formula:** $PEC = \sum_{\lambda \in \Lambda} L_\lambda \frac{r_\lambda}{R_\lambda}$ where $\Lambda$ is the set of branches found across neighbour set 1 only, $L_\lambda$ is the length of branch $\lambda$ , $r_\lambda$ is the local range of branch $\lambda$ (the number of groups in neighbour sets 1 and 2 containing it), and $R_\lambda$ is the global range of branch $\lambda$ (the number of groups across the entire data set containing it).

**Indices:**

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* |
|---|---|---|---|
| PEC_WE | Phylogenetic endemism, central variant | region grower | 1 |
| PEC_WE_P | Phylogenetic weighted endemism as a proportion of the total tree length, central variant | region grower | 1 |

### 2.8.10 Phylogenetic Endemism central lists

**Description:** Lists underlying the phylogenetic endemism central indices. Uses labels from neighbour set one but local ranges from across both neighbour sets.

**Subroutine:** calc_pe_central_lists

**Reference:** Rosauer et al (2009) Mol. Ecol

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| PEC_LOCAL_RANGELIST | Phylogenetic endemism local range lists, central variant | 1 |
| PEC_RANGELIST | Phylogenetic endemism global range lists, central variant | 1 |
| PEC_WTLIST | Phylogenetic endemism weights, central variant | 1 |

### 2.8.11 Phylogenetic Endemism lists

**Description:** Lists used in the Phylogenetic endemism (PE) calculations.

**Subroutine:** calc_pe_lists

**Reference:** Rosauer et al (2009) Mol. Ecol

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| PE_LOCAL_RANGELIST | Local node ranges used in PE calculations (number of groups in which a node is found) | 1 |
| PE_RANGELIST | Node ranges used in PE calculations | 1 |
| PE_WTLIST | Node weights used in PE calculations | 1 |

### 2.8.12 Phylogenetic Endemism single

**Description:** PE scores, but not weighted by local ranges. This is the strict interpretation of the formula given in Rosauer et al. (2009), although the approach has always been implemented as the fraction of each branch's geographic range that is found in the sample window (see formula for PE_WE).

**Subroutine:** calc_pe_single

**Reference:** Rosauer et al (2009) Mol. Ecol

**Indices:**

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* |
|---|---|---|---|
| PE_WE_SINGLE | Phylogenetic endemism unweighted by the number of neighbours. Counts each label only once, regardless of how many groups in the neighbourhood it is found in. Useful if your data have sampling biases. Better with small sample windows. | region grower | 1 |
| PE_WE_SINGLE_P | Phylogenetic endemism unweighted by the number of neighbours as a proportion of the total tree length. Counts each label only once, regardless of how many groups in the neighbourhood it is found. Useful if your data have sampling biases. | region grower | 1 |

### 2.8.13 RWiBaLD

**Description:** Range weighted branch length differences. Values are spatially constant, only the subsets change

**Subroutine:** calc_rwibald

**Reference:** Mishler et al. (in review)

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| RWIBALD_CODES | RWiBaLD codes, 1=palaeo, 2=neo, 3=meso | 1 |
| RWIBALD_CODE_COUNTS | Counts of branches in each RWiBaLD category | 1 |
| RWIBALD_DIFFS | RWiBaLD scores (continuous differences) | 1 |
| RWIBALD_METADATA | General metadata for the RWiBaLD calculations | 1 |
| RWIBALD_RR_DIFFS | RWiBaLD scores for the range restricted subset (continuous differences) | 1 |

## 2.9 Phylogenetic Indices

### 2.9.1 Count labels on tree

**Description:** Count the number of labels that are on the tree

**Subroutine:** calc_count_labels_on_tree

**Indices:**

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* |
|---|---|---|---|
| PHYLO_LABELS_ON_TREE_COUNT | The number of labels that are found on the tree, across both neighbour sets | region grower | 1 |

### 2.9.2 Evolutionary distinctiveness

**Description:** Evolutionary distinctiveness metrics (AED, ED, ES) Label values are constant for all neighbourhoods in which each label is found.

**Subroutine:** calc_phylo_aed

**Reference:** Cadotte & Davies (2010)

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* | *Reference* |
|---|---|---|---|
| PHYLO_AED_LIST | Abundance weighted ED per terminal label | 1 | Cadotte & Davies (2010) |
| PHYLO_ED_LIST | "Fair proportion" partitioning of PD per terminal label | 1 | Isaac et al. (2007) |
| PHYLO_ES_LIST | Equal splits partitioning of PD per terminal label | 1 | Redding & Mooers (2006) |

### 2.9.3 Evolutionary distinctiveness per site

**Description:** Site level evolutionary distinctiveness

**Subroutine:** calc_phylo_aed_t

**Reference:** Cadotte & Davies (2010)

**Indices:**

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* | *Reference* |
|---|---|---|---|---|
| PHYLO_AED_T | Abundance weighted ED_t (sum of values in PHYLO_AED_LIST times their abundances). This is equivalent to a phylogenetic rarity score (see phylogenetic endemism) | region grower | 1 | Cadotte & Davies (2010) |

### 2.9.4 Evolutionary distinctiveness per terminal taxon per site

**Description:** Site level evolutionary distinctiveness per terminal taxon

**Subroutine:** calc_phylo_aed_t_wtlists

**Reference:** Cadotte & Davies (2010)

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* | *Reference* |
|---|---|---|---|
| PHYLO_AED_T_WTLISTS | Abundance weighted ED per terminal taxon (the AED score of each taxon multiplied by its abundance in the sample) | 1 | Cadotte & Davies (2010) |
| PHYLO_AED_T_WTLISTS_P | Proportional contribution of each terminal taxon to the AED_T score | 1 | Cadotte & Davies (2010) |

### 2.9.5 Labels not on tree

**Description:** Create a hash of the labels that are not on the tree

**Subroutine:** calc_labels_not_on_tree

**Indices:**

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* |
|---|---|---|---|
| PHYLO_LABELS_NOT_ON_TREE | A hash of labels that are not found on the tree, across both neighbour sets | | 1 |
| PHYLO_LABELS_NOT_ON_TREE_N | Number of labels not on the tree | region grower | 1 |
| PHYLO_LABELS_NOT_ON_TREE_P | Proportion of labels not on the tree | | 1 |

### 2.9.6 Labels on tree

**Description:** Create a hash of the labels that are on the tree

**Subroutine:** calc_labels_on_tree

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| PHYLO_LABELS_ON_TREE | A hash of labels that are found on the tree, across both neighbour sets | 1 |

### 2.9.7 Last shared ancestor properties

**Description:** Properties of the last shared ancestor of an assemblage. Uses labels in both neighbourhoods.

**Subroutine:** calc_last_shared_ancestor_props

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| LAST_SHARED_ANCESTOR_DEPTH | Depth of last shared ancestor from the root. The root has a depth of zero. | 1 |
| LAST_SHARED_ANCESTOR_DIST_TO_ROOT | Distance along the tree from the last shared ancestor to the root. Includes the shared ancestor's length. | 1 |
| LAST_SHARED_ANCESTOR_DIST_TO_TIP | Distance along the tree from the last shared ancestor to the furthest tip in the sample. This is calculated from the point at which the lineages merge, which is the branch end further from the root | 1 |
| LAST_SHARED_ANCESTOR_LENGTH | Branch length of last shared ancestor | 1 |

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| LAST_SHARED_ANCESTOR_POS_REL | Relative position of the last shared ancestor. Value is the fraction of the distance from the root to the furthest terminal.This uses the point at which the lineages merge, and is the branch end further from the root | 1 |

### 2.9.8 PD clade contributions

**Description:** Contribution of each node and its descendents to the Phylogenetic diversity (PD) calculation.

**Subroutine:** calc_pd_clade_contributions

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| PD_CLADE_CONTR | List of node (clade) contributions to the PD calculation | 1 |
| PD_CLADE_CONTR_P | List of node (clade) contributions to the PD calculation, proportional to the entire tree | 1 |
| PD_CLADE_SCORE | List of PD scores for each node (clade), being the sum of all descendent branch lengths | 1 |

### 2.9.9 PD clade loss

**Description:** How much of the PD would be lost if a clade were to be removed? Calculates the clade PD below the last ancestral node in the neighbour set which would still be in the neighbour set.

**Subroutine:** calc_pd_clade_loss

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| PD_CLADE_LOSS_CONTR | List of the proportion of the PD score which would be lost if each clade were removed. | 1 |
| PD_CLADE_LOSS_CONTR_P | As per PD_CLADE_LOSS but proportional to the entire tree | 1 |
| PD_CLADE_LOSS_SCORE | List of how much PD would be lost if each clade were removed. | 1 |

### 2.9.10 PD clade loss (ancestral component)

**Description:** How much of the PD clade loss is due to the ancestral branches? The score is zero when there is no ancestral loss.

**Subroutine:** calc_pd_clade_loss_ancestral

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| PD_CLADE_LOSS_ANC | List of how much ancestral PE would be lost if each clade were removed. The value is 0 when no ancestral PD is lost. | 1 |
| PD_CLADE_LOSS_ANC_P | List of the proportion of the clade's PD loss that is due to the ancestral branches. | 1 |

### 2.9.11 Phylogenetic Abundance

**Description:** Phylogenetic abundance based on branch lengths back to the root of the tree. Uses labels in both neighbourhoods.

**Subroutine:** calc_phylo_abundance

**Indices:**

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* | *Formula* | *Reference* |
|---|---|---|---|---|---|
| PHYLO_ABUNDANCE | Phylogenetic abundance | region grower | 1 | $= \sum_{c \in C} A \times L_c$ where $C$ is the set of branches in the minimum spanning path joining the labels in both neighbour sets to the root of the tree, $c$ is a branch (a single segment between two nodes) in the spanning path $C$ , and $L_c$ is the length of branch $c$ , and $A$ is the abundance of that branch (the sum of its descendant label abundances). | |
| PHYLO_ABUNDANCE_BRANCH_HASH | Phylogenetic abundance per branch | | 1 | | |

### 2.9.12 Phylogenetic Diversity

**Description:** Phylogenetic diversity (PD) based on branch lengths back to the root of the tree. Uses labels in both neighbourhoods.

**Subroutine:** calc_pd

**Indices:**

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* | *Formula* | *Reference* |
|---|---|---|---|---|---|
| PD | Phylogenetic diversity | region grower | 1 | $= \sum_{c \in C} L_c$ where $C$ is the set of branches in the minimum spanning path joining the labels in both neighbour sets to the root of the tree, $c$ is a branch (a single segment between two nodes) in the spanning path $C$ , and $L_c$ is the length of branch $c$ . | Faith (1992) Biol. Cons |

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* | *Formula* | *Reference* |
|---|---|---|---|---|---|
| PD_P | Phylogenetic diversity as a proportion of total tree length | region grower | 1 | $= \frac{PD}{\sum_{c \in C} L_c}$ where terms are the same as for PD, but $c$ , $C$ and $L_c$ are calculated for all nodes in the tree. | |
| PD_P_per_taxon | Phylogenetic diversity per taxon as a proportion of total tree length | | 1 | $= \frac{PD\_P}{RICHNESS\_ALL}$ | |
| PD_per_taxon | Phylogenetic diversity per taxon | | 1 | $= \frac{PD}{RICHNESS\_ALL}$ | |

### 2.9.13 Phylogenetic Diversity (local)

**Description:** Phylogenetic diversity (PD) based on branch lengths back to the last shared ancestor. Uses labels in both neighbourhoods.

**Subroutine:** calc_pd_local

**Indices:**

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* | *Formula* |
|---|---|---|---|---|
| PD_LOCAL | Phylogenetic diversity calculated to last shared ancestor | region grower | 1 | $= \sum_{c \in C} L_c$ where $C$ is the set of branches in the minimum spanning path joining the labels in both neighbour sets to the last shared ancestor, $c$ is a branch (a single segment between two nodes) in the spanning path $C$ , and $L_c$ is the length of branch $c$ . |
| PD_LOCAL_P | Phylogenetic diversity as a proportion of total tree length | region grower | 1 | $= \frac{PD}{\sum_{c \in C} L_c}$ where terms are the same as for PD, but $c$ , $C$ and $L_c$ are calculated for all nodes in the tree. |

### 2.9.14 Phylogenetic Diversity node list

**Description:** Phylogenetic diversity (PD) nodes used.

**Subroutine:** calc_pd_node_list

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| PD_INCLUDED_NODE_LIST | List of tree nodes included in the PD calculations | 1 |

### 2.9.15 Phylogenetic Diversity terminal node count

**Description:** Number of terminal nodes in neighbour sets 1 and 2.

**Subroutine:** calc_pd_terminal_node_count

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| PD_INCLUDED_TERMINAL_NODE_COUNT | Count of tree terminal nodes included in the PD calculations | 1 |

### 2.9.16 Phylogenetic Diversity terminal node list

**Description:** Phylogenetic diversity (PD) terminal nodes used.

**Subroutine:** calc_pd_terminal_node_list

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| PD_INCLUDED_TERMINAL_NODE_LIST | List of tree terminal nodes included in the PD calculations | 1 |

## 2.10 Phylogenetic Indices (relative)

### 2.10.1 Labels not on trimmed tree

**Description:** Create a hash of the labels that are not on the trimmed tree

**Subroutine:** calc_labels_not_on_trimmed_tree

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| PHYLO_LABELS_NOT_ON_TRIMMED_TREE | A hash of labels that are not found on the tree after it has been trimmed to the basedata, across both neighbour sets | 1 |
| PHYLO_LABELS_NOT_ON_TRIMMED_TREE_N | Number of labels not on the trimmed tree | 1 |
| PHYLO_LABELS_NOT_ON_TRIMMED_TREE_P | Proportion of labels not on the trimmed tree | 1 |

### 2.10.2 Labels on trimmed tree

**Description:** Create a hash of the labels that are on the trimmed tree

**Subroutine:** calc_labels_on_trimmed_tree

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| PHYLO_LABELS_ON_TRIMMED_TREE | A hash of labels that are found on the tree after it has been trimmed to match the basedata, across both neighbour sets | 1 |

### 2.10.3 Relative Phylogenetic Diversity, type 1

**Description:** Relative Phylogenetic Diversity type 1 (RPD1). The ratio of the tree's PD to a null model of PD evenly distributed across terminals and where ancestral nodes are collapsed to zero length.You probably want to use RPD2 instead as it uses the tree's topology.

**Subroutine:** calc_phylo_rpd1

**Indices:**

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* | *Formula* |
|---|---|---|---|---|
| PHYLO_RPD1 | RPD1 | | 1 | |
| PHYLO_RPD_DIFF1 | How much more or less PD is there than expected, in original tree units. | | 1 | $= tree\_length \times (PD\_P - PHYLO\_RPD\_NULL1)$ |
| PHYLO_RPD_NULL1 | Null model score used as the denominator in the RPD1 calculations | region grower | 1 | |

### 2.10.4 Relative Phylogenetic Diversity, type 2

**Description:** Relative Phylogenetic Diversity (RPD), type 2. The ratio of the tree's PD to a null model of PD evenly distributed across all nodes (all branches are of equal length).

**Subroutine:** calc_phylo_rpd2

**Reference:** Mishler et al. (2014)

**Indices:**

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* | *Formula* |
|---|---|---|---|---|
| PHYLO_RPD2 | RPD2 | | 1 | |
| PHYLO_RPD_DIFF2 | How much more or less PD is there than expected, in original tree units. | | 1 | $= tree\_length \times (PD\_P - PHYLO\_RPD\_NULL2)$ |
| PHYLO_RPD_NULL2 | Null model score used as the denominator in the RPD2 calculations | region grower | 1 | |

### 2.10.5 Relative Phylogenetic Endemism, central

**Description:** Relative Phylogenetic Endemism (RPE). The ratio of the tree's PE to a null model where PE is calculated using a tree where all branches are of equal length. Same as RPE2 except it only uses the branches in the first neighbour set when more than one is set.

**Subroutine:** calc_phylo_rpe_central

**Reference:** Mishler et al. (2014)

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* | *Formula* |
|---|---|---|---|
| PHYLO_RPEC | Relative Phylogenetic Endemism score, central | 1 | |
| PHYLO_RPE_DIFFC | How much more or less PE is there than expected, in original tree units. | 1 | $= tree\_length \times (PE\_WEC\_P - PHYLO\_RPE\_NULLC)$ |
| PHYLO_RPE_NULLC | Null score used as the denominator in the PHYLO_RPEC calculations | 1 | |

### 2.10.6 Relative Phylogenetic Endemism, type 1

**Description:** Relative Phylogenetic Endemism, type 1 (RPE1). The ratio of the tree's PE to a null model of PD evenly distributed across terminals, but with the same range per terminal and where ancestral nodes are of zero length (as per RPD1).You probably want to use RPE2 instead as it uses the tree's topology.

**Subroutine:** calc_phylo_rpe1

**Indices:**

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* | *Formula* |
|---|---|---|---|---|
| PHYLO_RPE1 | Relative Phylogenetic Endemism score | | 1 | |

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* | *Formula* |
|---|---|---|---|---|
| PHYLO_RPE_DIFF1 | How much more or less PE is there than expected, in original tree units. | | 1 | $= tree\_length \times (PE\_WE\_P - PHYLO\_RPE\_NULL1)$ |
| PHYLO_RPE_NULL1 | Null score used as the denominator in the RPE calculations | region grower | 1 | |

### 2.10.7 Relative Phylogenetic Endemism, type 2

**Description:** Relative Phylogenetic Endemism (RPE). The ratio of the tree's PE to a null model where PE is calculated using a tree where all non-zero branches are of equal length.

**Subroutine:** calc_phylo_rpe2

**Reference:** Mishler et al. (2014)

**Indices:**

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* | *Formula* |
|---|---|---|---|---|
| PHYLO_RPE2 | Relative Phylogenetic Endemism score, type 2 | | 1 | |
| PHYLO_RPE_DIFF2 | How much more or less PE is there than expected, in original tree units. | | 1 | $= tree\_length \times (PE\_WE\_P - PHYLO\_RPE\_NULL2)$ |

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* | *Formula* |
|---|---|---|---|---|
| PHYLO_RPE_NULL2 | Null2 score used as the denominator in the RPE2 calculations | region grower | 1 | |

## 2.11 Phylogenetic Turnover

### 2.11.1 Phylo Jaccard

**Description:** Jaccard phylogenetic dissimilarity between two sets of taxa, represented by spanning sets of branches

**Subroutine:** calc_phylo_jaccard

**Reference:** Lozupone and Knight (2005)

**Indices:**

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* | *Formula* |
|---|---|---|---|---|
| PHYLO_JACCARD | Phylo Jaccard score | cluster metric | 2 | $= 1 - (A/(A + B + C))$ where A is the length of shared branches, and B and C are the length of branches found only in neighbour sets 1 and 2 |

### 2.11.2 Phylo Range weighted Turnover

**Description:** Phylo Range weighted Turnover

**Subroutine:** calc_phylo_rw_turnover

**Reference:** Laffan et al. (2016)

**Indices:**

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* |
|---|---|---|---|
| PHYLO_RW_TURNOVER | Range weighted turnover | cluster metric | 2 |
| PHYLO_RW_TURNOVER_A | Range weighted turnover, shared component | region grower | 2 |
| PHYLO_RW_TURNOVER_B | Range weighted turnover, component found only in nbr set 1 | | 2 |
| PHYLO_RW_TURNOVER_C | Range weighted turnover, component found only in nbr set 2 | | 2 |

### 2.11.3 Phylo S2

**Description:** S2 phylogenetic dissimilarity between two sets of taxa, represented by spanning sets of branches

**Subroutine:** calc_phylo_s2

**Indices:**

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* | *Formula* |
|---|---|---|---|---|
| PHYLO_S2 | Phylo S2 score | cluster metric | 2 | $= 1 - (A/(A + min(B, C)))$ where A is the sum of shared branch lengths, and B and C are the sum of branch lengths foundonly in neighbour sets 1 and 2 |

### 2.11.4 Phylo Sorenson

**Description:** Sorenson phylogenetic dissimilarity between two sets of taxa, represented by spanning sets of branches

**Subroutine:** calc_phylo_sorenson

**Reference:** Bryant et al. (2008)

**Indices:**

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* | *Formula* |
|---|---|---|---|---|
| PHYLO_SORENSON | Phylo Sorenson score | cluster metric | 2 | $1 - (2A/(2A + B + C))$ where A is the length of shared branches, and B and C are the length of branches found only in neighbour sets 1 and 2 |

### 2.11.5 Phylogenetic ABC

**Description:** Calculate the shared and not shared branch lengths between two sets of labels

**Subroutine:** calc_phylo_abc

**Indices:**

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* |
|---|---|---|---|
| PHYLO_A | Sum of branch lengths shared by labels in nbr sets 1 and 2 | region grower | 2 |
| PHYLO_ABC | Sum of branch lengths associated with labels in nbr sets 1 and 2 | region grower | 2 |
| PHYLO_B | Sum of branch lengths unique to labels in nbr set 1 | | 2 |
| PHYLO_C | Sum of branch lengths unique to labels in nbr set 2 | | 2 |

## 2.12 Rarity

### 2.12.1 Rarity central

**Description:** Calculate rarity for species only in neighbour set 1, but with local sample counts calculated from both neighbour sets. Uses the same algorithm as the endemism indices but weights by sample counts instead of by groups occupied.

**Subroutine:** calc_rarity_central

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* | *Formula* |
|---|---|---|---|
| RAREC_CWE | Corrected weighted rarity | 1 | $= \frac{RAREC\_WE}{RAREC\_RICHNESS}$ |

| *Index* | *Description* | *Minimum number of neighbour sets* | *Formula* |
|---|---|---|---|
| RAREC_RICHNESS | Richness used in RAREC_CWE (same as index RICHNESS_SET1). | 1 | |
| RAREC_WE | Weighted rarity | 1 | $= \sum_{t \in T} \frac{s_t}{S_t}$ where $t$ is a label (taxon) in the set of labels (taxa) $T$ across neighbour set 1, $s_t$ is sum of the sample counts for $t$ across the elements in neighbour sets 1 & 2 (its value in list ABC3_LABELS_ALL), and $S_t$ is the total number of samples across the data set for label $t$ (unless the total sample count is specified at import). |

### 2.12.2 Rarity central lists

**Description:** Lists used in rarity central calculations

**Subroutine:** calc_rarity_central_lists

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| RAREC_RANGELIST | List of ranges for each label used in the rarity central calculations | 1 |
| RAREC_WTLIST | List of weights for each label used in therarity central calculations | 1 |

### 2.12.3 Rarity whole

**Description:** Calculate rarity using all species in both neighbour sets. Uses the same algorithm as the endemism indices but weights by sample counts instead of by groups occupied.

**Subroutine:** calc_rarity_whole

**Indices:**

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* | *Formula* |
|---|---|---|---|---|
| RAREW_CWE | Corrected weighted rarity | | 1 | $= \frac{RAREW\_WE}{RAREW\_RICHNESS}$ |
| RAREW_RICHNESS | Richness used in RAREW_CWE (same as index RICHNESS_ALL). | region grower | 1 | |
| RAREW_WE | Weighted rarity | region grower | 1 | $= \sum_{t \in T} \frac{s_t}{S_t}$ where $t$ is a label (taxon) in the set of labels (taxa) $T$ across both neighbour sets, $s_t$ is sum of the sample counts for $t$ across the elements in neighbour sets 1 & 2 (its value in list ABC3_LABELS_ALL), and $S_t$ is the total number of samples across the data set for label $t$ (unless the total sample count is specified at import). |

### 2.12.4 Rarity whole lists

**Description:** Lists used in rarity whole calculations

**Subroutine:** calc_rarity_whole_lists

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| RAREW_RANGELIST | List of ranges for each label used in the rarity whole calculations | 1 |
| RAREW_WTLIST | List of weights for each label used in therarity whole calculations | 1 |

## 2.13 Richness estimators

### 2.13.1 ACE

**Description:** Abundance Coverage-based Estimator of species richness

**Subroutine:** calc_ace

**Reference:** Chao and Lee (1992)

**Indices:**

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* |
|---|---|---|---|
| ACE_CI_LOWER | ACE lower confidence interval estimate | region grower | 1 |
| ACE_CI_UPPER | ACE upper confidence interval estimate | region grower | 1 |
| ACE_ESTIMATE | ACE score | region grower | 1 |
| ACE_ESTIMATE_USED_CHAO | Set to 1 when ACE cannot be calculated and so Chao1 estimate is used | | 1 |
| ACE_INFREQUENT_COUNT | Count of infrequent species | region grower | 1 |

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* |
|---|---|---|---|
| ACE_SE | ACE standard error | | 1 |
| ACE_UNDETECTED | Estimated number of undetected species | region grower | 1 |
| ACE_VARIANCE | ACE variance | | 1 |

### 2.13.2 Chao1

**Description:** Chao1 species richness estimator (abundance based)

**Subroutine:** calc_chao1

**Indices:**

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* | *Reference* |
|---|---|---|---|---|
| CHAO1_CI_LOWER | Lower confidence interval for the Chao1 estimate | region grower | 1 | |
| CHAO1_CI_UPPER | Upper confidence interval for the Chao1 estimate | region grower | 1 | |
| CHAO1_ESTIMATE | Chao1 index | region grower | 1 | NEEDED |
| CHAO1_F1_COUNT | Number of singletons in the sample | region grower | 1 | |
| CHAO1_F2_COUNT | Number of doubletons in the sample | region grower | 1 | |

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* | *Reference* |
|---|---|---|---|---|
| CHAO1_META | Metadata indicating which formulae were used in the calculations. Numbers refer to EstimateS equations at https://www.robertk-colwell.org/media_files/63 | | 1 | |
| CHAO1_SE | Standard error of the Chao1 estimator [= sqrt(variance)] | region grower | 1 | |
| CHAO1_UNDETECTED | Estimated number of undetected species | region grower | 1 | |
| CHAO1_VARIANCE | Variance of the Chao1 estimator | region grower | 1 | |

### 2.13.3 Chao2

**Description:** Chao2 species richness estimator (incidence based)

**Subroutine:** calc_chao2

**Indices:**

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* | *Reference* |
|---|---|---|---|---|
| CHAO2_CI_LOWER | Lower confidence interval for the Chao2 estimate | region grower | 1 | |

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* | *Reference* |
|---|---|---|---|---|
| CHAO2_CI_UPPER | Upper confidence interval for the Chao2 estimate | region grower | 1 | |
| CHAO2_ESTIMATE | Chao2 index | region grower | 1 | NEEDED |
| CHAO2_META | Metadata indicating which formulae were used in the calculations. Numbers refer to EstimateS equations at https://www.robertk-colwell.org/media_files/63 | | 1 | |
| CHAO2_Q1_COUNT | Number of uniques in the sample | region grower | 1 | |
| CHAO2_Q2_COUNT | Number of duplicates in the sample | region grower | 1 | |
| CHAO2_SE | Standard error of the Chao2 estimator [= sqrt (variance)] | region grower | 1 | |
| CHAO2_UNDETECTED | Estimated number of undetected species | region grower | 1 | |
| CHAO2_VARIANCE | Variance of the Chao2 estimator | region grower | 1 | |

### 2.13.4 Hurlbert richness estimation

**Description:** Hurlbert estimated species richness scores for given number of samples.

**Subroutine:** calc_hurlbert_es

**Reference:** Hurlbert, S.H. (1971)

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| HURLBERT_ES | List of Hurlbert estimated species richness scores for given number of samples | 1 |

### 2.13.5 ICE

**Description:** Incidence Coverage-based Estimator of species richness

**Subroutine:** calc_ice

**Reference:** Gotelli and Chao (2013)

**Indices:**

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* |
|---|---|---|---|
| ICE_CI_LOWER | ICE lower confidence interval estimate | region grower | 1 |
| ICE_CI_UPPER | ICE upper confidence interval estimate | region grower | 1 |
| ICE_ESTIMATE | ICE score | region grower | 1 |
| ICE_ESTIMATE_USED_CHAO | Set to 1 when ICE cannot be calculated and so Chao2 estimate is used | | 1 |
| ICE_INFREQUENT_COUNT | Count of infrequent species | region grower | 1 |
| ICE_SE | ICE standard error | | 1 |
| ICE_UNDETECTED | Estimated number of undetected species | region grower | 1 |
| ICE_VARIANCE | ICE variance | | 1 |

## 2.14 Taxonomic Dissimilarity and Comparison

### 2.14.1 Beta diversity

**Description:** Beta diversity between neighbour sets 1 and 2.

**Subroutine:** calc_beta_diversity

**Indices:**

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* | *Formula* |
|---|---|---|---|---|
| BETA_2 | The other beta | cluster metric | 2 | $= \frac{A+B+C}{max((A+B),(A+C))} - 1$ where $A$ is the count of labels found in both neighbour sets, $B$ is the count unique to neighbour set 1, and $C$ is the count unique to neighbour set 2. Use the Label counts calculation to derive these directly. |

### 2.14.2 Bray-Curtis non-metric

**Description:** Bray-Curtis dissimilarity between two sets of labels. Reduces to the Sorenson metric for binary data (where sample counts are 1 or 0).

**Subroutine:** calc_bray_curtis

**Formula:** $= 1 - \frac{2W}{A+B}$ where $A$ is the sum of the sample counts in neighbour set 1, $B$ is the sum of sample counts in neighbour set 2, and $W = \sum_{i=1}^{n} min(sample\_count\_label_{i_{set1}}, sample\_count\_label_{i_{set2}})$ (meaning it sums the minimum of the sample counts for each of the $n$ labels across the two neighbour sets),

**Indices:**

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* |
|---|---|---|---|
| BC_A | The A factor used in calculations (see formula) | | 2 |
| BC_B | The B factor used in calculations (see formula) | | 2 |
| BC_W | The W factor used in calculations (see formula) | region grower | 2 |
| BRAY_CURTIS | Bray Curtis dissimilarity | cluster metric | 2 |

### 2.14.3 Bray-Curtis non-metric, group count normalised

**Description:** Bray-Curtis dissimilarity between two neighbourhoods, where the counts in each neighbourhood are divided by the number of groups in each neighbourhood to correct for unbalanced sizes.

**Subroutine:** calc_bray_curtis_norm_by_gp_counts

**Formula:** $= 1 - \frac{2W}{A+B}$ where $A$ is the sum of the sample counts in neighbour set 1 normalised (divided) by the number of groups, $B$ is the sum of the sample counts in neighbour set 2 normalised by the number of groups, and $W = \sum_{i=1}^{n} min(sample\_count\_label_{i_{set1}}, sample\_count\_label_{i_{set2}})$ (meaning it sums the minimum of the normalised sample counts for each of the $n$ labels across the two neighbour sets),

**Indices:**

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* |
|---|---|---|---|
| BCN_A | The A factor used in calculations (see formula) | | 2 |
| BCN_B | The B factor used in calculations (see formula) | | 2 |
| BCN_W | The W factor used in calculations (see formula) | region grower | 2 |

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* |
|---|---|---|---|
| BRAY_CURTIS_NORM | Bray Curtis dissimilarity normalised by groups | cluster metric | 2 |

### 2.14.4 Jaccard

**Description:** Jaccard dissimilarity between the labels in neighbour sets 1 and 2.

**Subroutine:** calc_jaccard

**Formula:** $= 1 - \frac{A}{A+B+C}$ where $A$ is the count of labels found in both neighbour sets, $B$ is the count unique to neighbour set 1, and $C$ is the count unique to neighbour set 2. Use the Label counts calculation to derive these directly.

**Indices:**

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* |
|---|---|---|---|
| JACCARD | Jaccard value, 0 is identical, 1 is completely dissimilar | cluster metric | 2 |

### 2.14.5 Kulczynski 2

**Description:** Kulczynski 2 dissimilarity between two sets of labels.

**Subroutine:** calc_kulczynski2

**Formula:** $= 1 - 0.5 \times (\frac{A}{A+B} + \frac{A}{A+C})$ where $A$ is the count of labels found in both neighbour sets, $B$ is the count unique to neighbour set 1, and $C$ is the count unique to neighbour set 2. Use the Label counts calculation to derive these directly.

**Indices:**

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* |
|---|---|---|---|
| KULCZYNSKI2 | Kulczynski 2 index | cluster metric | 2 |

### 2.14.6 Nestedness-resultant

**Description:** Nestedness-resultant index between the labels in neighbour sets 1 and 2.

**Subroutine:** calc_nestedness_resultant

**Reference:** Baselga (2010) Glob Ecol Biogeog.

**Formula:** $= \frac{|B-C|}{2A+B+C} \times \frac{A}{A+min(B,C)} = SORENSON - S2$ where $A$ is the count of labels found in both neighbour sets, $B$ is the count unique to neighbour set 1, and $C$ is the count unique to neighbour set 2. Use the Label counts calculation to derive these directly.

**Indices:**

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* |
|---|---|---|---|
| NEST_RESULTANT | Nestedness-resultant index | cluster metric | 2 |

### 2.14.7 Range weighted Sorenson

**Description:** Range weighted Sorenson

**Subroutine:** calc_rw_turnover

**Reference:** Laffan et al. (2016)

**Indices:**

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* |
|---|---|---|---|
| RW_TURNOVER | Range weighted turnover | cluster metric | 2 |
| RW_TURNOVER_A | Range weighted turnover, shared component | | 2 |
| RW_TURNOVER_B | Range weighted turnover, component found only in nbr set 1 | | 2 |
| RW_TURNOVER_C | Range weighted turnover, component found only in nbr set 2 | | 2 |

### 2.14.8 Rao's quadratic entropy, taxonomically weighted

**Description:** Calculate Rao's quadratic entropy for a taxonomic weights scheme. Should collapse to be the Simpson index for presence/absence data.

**Subroutine:** calc_tx_rao_qe

**Formula:** $= \sum_{i \in L} \sum_{j \in L} d_{ij} p_i p_j$ where $p_i$ and $p_j$ are the sample counts for the i'th and j'th labels, $d_{ij}$ is a value of zero if $i = j$ , and a value of 1 otherwise. $L$ is the set of labels across both neighbour sets.

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* |
|---|---|---|
| TX_RAO_QE | Taxonomically weighted quadratic entropy | 1 |
| TX_RAO_TLABELS | List of labels and values used in the TX_RAO_QE calculations | 1 |
| TX_RAO_TN | Count of comparisons used to calculate TX_RAO_QE | 1 |

### 2.14.9 S2

**Description:** S2 dissimilarity between two sets of labels

**Subroutine:** calc_s2

**Reference:** Lennon et al. (2001) J Animal Ecol.

**Formula:** $= 1 - \frac{A}{A+min(B,C)}$ where $A$ is the count of labels found in both neighbour sets, $B$ is the count unique to neighbour set 1, and $C$ is the count unique to neighbour set 2. Use the Label counts calculation to derive these directly.

**Indices:**

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* |
|---|---|---|---|
| S2 | S2 dissimilarity index | cluster metric | 2 |

### 2.14.10 Simpson and Shannon

**Description:** Simpson and Shannon diversity metrics using samples from all neighbourhoods.

**Subroutine:** calc_simpson_shannon

**Formula:** For each index formula, $p_i$ is the number of samples of the i'th label as a proportion of the total number of samples $n$ in the neighbourhoods.

**Indices:**

| *Index* | *Description* | *Minimum number of neighbour sets* | *Formula* |
|---|---|---|---|
| SHANNON_E | Shannon's evenness (H / HMAX) | 1 | $Evenness = \frac{H}{HMAX}$ |
| SHANNON_H | Shannon's H | 1 | $H = -\sum_{i=1}^{n}(p_i \cdot ln(p_i))$ |
| SHANNON_HMAX | maximum possible value of Shannon's H | 1 | $HMAX = ln(richness)$ |
| SIMPSON_D | Simpson's D. A score of zero is more similar. | 1 | $D = 1 - \sum_{i=1}^{n} p_i^2$ |

### 2.14.11 Sorenson

**Description:** Sorenson dissimilarity between two sets of labels. It is the complement of the (unimplemented) Czechanowski index, and numerically the same as Whittaker's beta.

**Subroutine:** calc_sorenson

**Formula:** $= 1 - \frac{2A}{2A+B+C}$ where $A$ is the count of labels found in both neighbour sets, $B$ is the count unique to neighbour set 1, and $C$ is the count unique to neighbour set 2. Use the Label counts calculation to derive these directly.

**Indices:**

| *Index* | *Description* | *Grouping metric?* | *Minimum number of neighbour sets* |
|---|---|---|---|
| SORENSON | Sorenson index | cluster metric | 2 |