Co-occurrence based metric of species habitat specialization

Author: David Zelený (zeleny.david@gmail.com)

Calculation of species habitat specialization measure from co-occurrence species data (see Theoretical background for details).



Download JUICE-R script:
generalists-specialists_v7.0.r
(last update: 12.2.2015)

Major updates

Version 7.0

- The whole algorithm was packed into 'genspe' library, which is available in github repository and can be installed when running current JUICE-R script. All updates are newly build into this library - check the version of the library written in the head of tcltk wizard.

Version 6.0

- Added new metrics to calculate habitat specialization (according to Manthey & Fridley 2009).
- New version can calculate specialization for the species with the same species name occurring in different layers (e.g. trees vs seedlings).
- Changed the import of species values back to JUICE - newly via clipboard.
- Added Library admin - in case that some of the required libraries are not installed in current R version, at the beginning the small lib-admin window opens and offers installation of these libraries.
- Previous version (v 5.5 and lower) had a bug in calculating the values using metric according to Botta-Dukat (2012) - it calculated the specialization metric even for species with lower than set up occurrence frequencies (the frequency of the species for which the metric was being calculated was taken from beals-smoothed matrix of species pool instead of the original data).

Technical requirements

- JUICE version: 7.0.96 or higher
- R version: 3.0.2 or higher
- R libraries: parallel, betapart, genspe

How to use

1. Download the JUICE-R script from this website.
2. For instruction how to use JUICE-R functions, refer to How to use JUICE-R function section of this website. Mostly you need to link JUICE program with R program.
3. To run this specific JUICE-R script for calculation of habitat specialists, you need to install several R libraries in R. After launching the script, it will automatically check for installed libraries, and if some of them is missing, it will open **Library administration** wizard, where you can install missing libraries (internet connection is needed)[1)]''. The package 'parallel', which is also needed, comes already with basic R installation (since version 2.14.) and doesn't need to be installed.)). You need to do this **only once** for given R version (if you install new R version, you need to install these libraries again).
4. Open your species data in JUICE. In JUICE menu, go to Analysis > JUICE-R functions (or **CTRL + W**), and in newly opened wizard click on **Append R script**. Select the file with the JUICE-R script downloaded from this website to your computer. You do not need to do any other selections in this wizard. Click **Run**.
5. After launching JUICE-R application, you should be able to see the wizard (Fig. 1 below).
6. At this moment, your species data are already imported in R. Select the analysis you want to proceed and click the **Calculate** button. You will see the progress bar with information about calculation progress, and information message at the end of the calculation.
7. The resulting files are stored in the R program folder (the same where the Rgui.exe file is located - you used this folder to link JUICE and R in JUICE Options menu). If you are not sure which folder is it, you can find it in JUICE menu File > Options > External Program Paths > R-PROJECT.

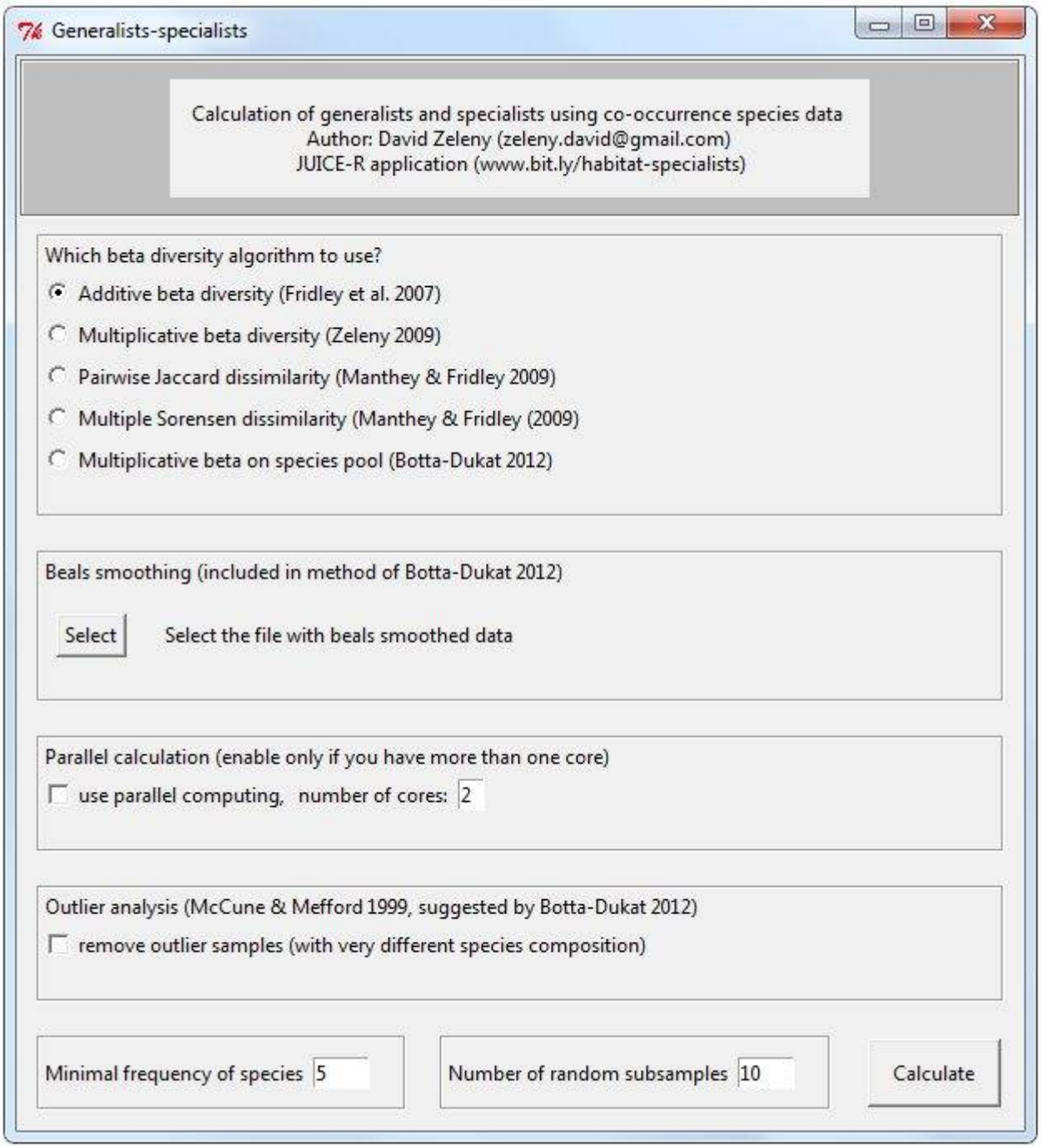


Fig. 1. Interface of the JUICE-R application calculating species habitat specialization measure.

Fix Me!

Options

Which beta diversity algorithm to use?

- **Additive beta diversity**. This is the original algorithm published by Fridley et al. (2007), in which beta diversity among samples containing given species is calculated by additive beta diversity measure.
- **Multiplicative beta diversity**. Uses the multiplicative Whittaker's measure of beta diversity instead of the original additive measure, as suggested by Zelený (2009).

- **Multiplicative beta on species pool**. Algorithm suggested by Botta-Dukát (2012), calculating the beta diversity using species pool matrix instead of the original species data matrix. Species pool matrix is calculated using Beals smoothing method (invented by Ewald 2002). While the previous multiplicative beta diversity method gives unbiased results only in case of not-saturated communities, this method should give unbiased results also in case of saturated communities. See Zelený (2009) and Botta-Dukát (2012) for detail discussion of this saturated/not-saturated communities issue.
- **Pairwise Jaccard dissimilarity** and **Multiple Sorensen dissimilarity** based on Manthey & Fridley (2009). Authors suggested that neither the original additive algorithm (introduced by Fridley et al. 2007), neither the modified version using the multiplicative beta diversity (Zeleny 2009) is the best solution, and introduce other alternatives, using pairwise or multiple site beta diversity algorithm. Pairwise Jaccard is based on calculating mean of Jaccard dissimilarities among all pairs of samples in each subset, while multiple Sorensen is using multiple-site Sorensen algorithm introduced by Baselga et al. (2007)[2)].

Beals smoothing

This option can be used if you calculate **Multiplicative beta on species pool**. Calculation of species pool using Beals smoothing can be time consuming in case of large datasets, and you may do this only once in case that you will recalculate the analysis again in future. First, run the **Multiplicative beta on species pool** analysis WITHOUT selecting the Beals smoothed species pool file. This will create the file `beals-data.txt` file in the R working directory, which is basically the matrix of samples x species containing presences of species in species pool of all samples. In the next run of the same analysis, you can select this file and save time for it's calculation. Warning: the file must have the same dimension as the original data! [3)]

Parallel calculation

If your computer has more than one core, you can speed up the calculation by running parallel processes on more than one core. Note, however, that you are still limited by the RAM memory of your computer, and that the whole original dataset is being loaded to each core separately. If your dataset is large, RAM may easily get saturated if you run the calculation on too many cores - solution is to choose less cores.

In parallel mode, you will not see the progress bar - to check for the progress, search for *GS-progress.txt* file in working R directory and open it for view, e.g. using Notepad[4)]. You may be prompted by firewall to allow the parallel process run on your computer - you need to allow it to run it in parallel.

Outlier analysis

This was suggested by Botta-Dukát (2012) as a way to avoid the situation, when given species will become generalists only because it occurs in a few samples with very different species composition. This method, used by McCune & Mefford in PC-ORD program, calculates the Euclidean distances among all pairs of samples containing particular species, and removes those samples with Euclidean distance to any other sample higher than mean of all distances among samples + 2 S.D. See Botta-Dukát (2012), page 205, for details.

Minimal frequency of species

Habitat specialization will be calculated for species occurring in number of samples equal or higher than minimal frequency threshold. Beta diversity is always calculated from the same number of samples for all species - this number is equal to this Minimal frequency threshold. Note that all species in the table are used to calculate beta diversity metric.

Number of random subsamples

Calculation of beta diversity is done from fixed number of randomly selected samples containing given species (this number is equal with Minimal frequency of species). Number of random subsamples speficies how many times this random selection will be repeated (this is important for calculating the mean and confidence intervals of habitat specialization metric).

Output

The program generates several output files, which can be found in the R working directory [5)]:

- **theta-out.txt** - the main output file (of tab delimited format), containing complete results, separated into several columns [6)], from which **the most important is the column *GS* with species habitat specialization measure (theta value)**:
  - *sci.name* - species scientific names, including the information about the layer (after the "_" sign);
  - *full.sci.name* - the full species name (the same as names in original JUICE file);
  - *layer* - information about species layer;
  - *local.avgS* - mean local species richness of samples containing given species;
  - *occur.freq* - total number of plots, in which given species occur;
  - *meanco* - mean number of species in table of *n* selected plots, from which beta diversity is calculated;
  - *meanco.sd* - s.d. of number of species in table of *n* selected pltos;
  - *meanco.u* - 97.5% confidence limit of number of species in table of *n* selected plots;
  - *meanco.l* - 2.5% confidence limit of number of species in table of *n* selected plots;
  - *GS* - mean theta value (measure of habitat specialization);
  - *GS.sd* - s.d. of theta value, calculated from all replications.
- **theta_import.species.data.via.clipboard.txt** - this is simple output file (of tab-delimited format), which can be used for importing the habitat specialization values back to JUICE (see below).
- **beals-data.txt** - this file (of tab delimited format) contains result of Beals smoothed estimate of species pool for each sample; it's a sample x species matrix with presences and absences of each species in the species pool of given plot. It's a by-product of the *Multiplicative beta on species pool* algorithm and once calculated, it can be used to speed up the recalculation.
- **Table.txt**, **basic.r**, **basic.out**, **Final.r** - these files, located in the same folder, are produced as a side effect of JUICE and R activity - they contain data exported from JUICE to R, and the R script with it's output. These files are not intended for direct use.

Import of calculated species specialization values back to JUICE

To import the calculated theta values back to JUICE, you can use the file **theta_import.species.data.txt**, which has format appropriate for import of species data into JUICE. In JUICE menu, go to ***File > Import > Species Data***, and selected the file **theta_import.species.data.txt**. In the newly open import wizard, you may need to modify the *First character* and the *Last character* values in *Parameters for species selection* and *Species information sections* [7)]. Select *Import data for* all *species* and click *Continue* and *Ok*.

Note that not all species have assigned specialization value, as this was not calculated for species with frequency lower than the selected threshold.

About the script

Original script was part of the supplementary materials of Fridley et al. (2007) and Zelený (2009). Current script is completely rewritten in order to enhance the efficiency of calculation, needed for larger datasets. Script performance will depend heavily on the technical parameters of the computer (mainly processor speed and size of memory). The script has an option to run parallel sessions, which greatly speeds up the calculation in case that the computer has more than one core.

Theoretical background

Fridley et al. (2007) introduced a novel technique to estimate the measure of species habitat specialization, based on analysis of co-occurrence data extracted from large vegetation data sets. The theory is simple and straightforward: for species occupying many different habitats – generalists – the rate of species turnover among plots in which they occur will be relatively high, while for species restricted to specific habitats – specialists – the species turnover rate will be relatively low, simply because they consistently occur with a limited number of other species. A continuous metric of habitat specialization proposed by Fridley et al. (2007) is called 'theta' (θ) and its calculation is based on a measure of beta diversity among the plots with given species. The θ value should be an estimate of species niche width. However, given that real vegetation data are used to calculate θ, the results will reveal realized, not fundamental, species niche, and their validity will be limited to the data set used for analysis. The main advantage of this method is that there is no need for information about the ecological gradient and species position along this gradient. Instead, only a sufficiently large data set of vegetation plots and an algorithm written in R by Fridley et al. (2007) is required.

Zelený (2009) found that the original algorithm, which was using additive metric of beta diversity, suffers from dependence on the size of the species pool of particular vegetation - the species occurring in more species rich vegetation tends to be more generalists than species occurring in species poor types. Zelený (2009) suggested modification of the original algorithm, which is based on replacing original additive measure of beta diversity by multiplicative one (Whittaker's beta). In response, Manthey and Fridley (2009) tested various beta diversity metrics in order to give a guideline to the selection of the appropriate one, and introduced the use of two other beta diversity metrics as alternatives to additive and multiplicative one (namely pair-wise Jaccard and multisite Sorensen, according to Baselga et al. 2007).

Botta-Dukát (2012) pointed out that the method of Fridley et al. (2007) and Zelený (2009) works well only in case of not-saturated communities, and suggested an algorithm which gives unbiased estimates even in case of saturated communities. This algorithm calculates theta values not from the original sample-species matrix, but from species pool of each plot. This species pool is estimated by Beals smoothing (algorithm proposed by Ewald 2002 and modified by Münzbergová & Herben 2004).

References

- Baselga A., Jiménez-Valverde A. & Niccolini G. (2007): A multiple-site similarity measure independent of richness. *Biology Letters*, 3: 642-645.
- Baselga A., Orme D., Villeger S., Bortoli J. & Leprieur F. (2013): betapart: Partitioning beta diversity into turnover and nestedness components. R package version 1.3. http://CRAN.R-project.org/package=betapart
- Botta-Dukát Z. (2012): Co-occurrence-based measure of species' habitat specialization: robust, unbiased estimation in saturated communities. *Journal of Vegetation Science*, 23: 201-207.
- Ewald J. (2002): A probabilistic approach to estimating species pools from large compositional matrices. *Journal of Vegetation Science*, 13: 191-198.
- Fridley J.D., Vandermast D.B., Kuppinger D.M., Manthey M. & Peet R.K. (2007): Co-occurrence based assessment of habitat generalists and specialists: a new approach for the measurement of niche width. *Journal of Ecology*, 95: 707-722.
- Manthey M. & Fridley J.D. (2009): Beta diversity metrics and the estimation of niche width via species co-occurrence data: reply to Zeleny. *Journal of Ecology*, 97: 18-22.
- McCune B. & Mefford M.J. (1999): *PC-ORD: multivariate analysis of ecological data. User's guide*. MjM Software, Gleneden Beach, OR, US.
- Münzbergová Z. & Herben T. (2004): Identification of suitable unoccupied habitats in metapopulation studies using co-occurrence of species. *Oikos*, 105: 408-414.
- Zelený D. (2009): Co-occurrence based assessment of species habitat specialization is affected by the size of species pool: reply to Fridley et al. (2007). *Journal of Ecology*, 97: 10-17.

1)
If this method is not successful for some reason, you need to install the libraries manually: open R program (e.g. from Windows Start menu, and make sure you open the same version which is linked to JUICE) and paste the following script to R console: ''install.packages (c('betapart'
2)
Technically, both algorithms are calculated by R packages `betapart` written by Baselga et al., namely by function `beta.pair` for pairwise and `beta.multi` for multisite dissimilarity.
3)
It should have the same number of samples and species as has the JUICE file. It means that if you deleted some species or samples in JUICE and want to recalculate the analysis, you can't use this file - follow the error message
4)
Individual cores are simultaneously saving the species number for which they just finished calculating the metric.
5)
the same directory where Rgui.exe file is located, usually c:\Program Files\R\R-x.xx.x\bin\x64 in 64-bit computers or c:\Program Files\R\R-x.xx.x\bin\i386 in 32-bit machines, where R-x.xx.x is actually used R version, e.g. R-3.0.2
6)
the column names follow the original naming used in the script of J. Fridley from the paper Fridley et al. 2007, available as the Appendix S2
7)

This setting will be most probably as followings: *Parameters for species selection* > *First Character*: 1, *Last Character*: 52; *Species information* > *First character*: 53, *Last character*: 68. In some cases, however, this numbers need to be changed.

From:
https://davidzeleny.net/juice-r/ - **JUICE-R**

Permanent link:
https://davidzeleny.net/juice-r/doku.php/scripts:generalists-specialists

Last update: **2015/07/31 16:54**