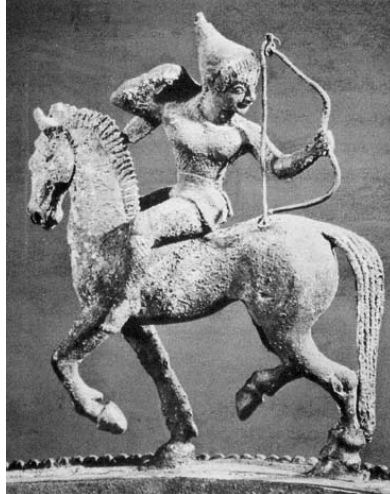


Parthy



*A test of neutrality using species abundance evenness,
and parameter inference by Approximate Bayesian
Computation*

<http://www.edb.ups-tlse.fr/equipe1/tetame.htm>

Franck Jabot
Jérôme Chave

*Laboratoire Evolution et Diversité Biologique UMR 5174
CNRS/UPS*

Université Paul Sabatier, Toulouse, France

September 30th, 2009
Version 1.0

Introduction

Hubbell's neutral theory of biodiversity (Hubbell, 2001) offers a simple explanation of species coexistence as a result of stochastic processes of birth, death, immigration and speciation. Species abundances are determined by two parameters only: θ , that governs the appearance of new species through speciation in the regional species pool, and m that governs immigration from the regional pool into local communities. This model is formally analogous to a continent-island model of population genetics (Wright 1931).

The freeware **Parthy** uses two approaches to directly assess the hypothesis of neutrality in a species abundance dataset. The first neutrality test is based on the comparison of the evenness of a community with the evenness of simulated neutral communities of same size and species richness. This is a simple test analogous to Watterson's heterozygosity test in population genetics. The program also returns the p -value of the neutrality test. For more conservative p -values, a protocol is provided in the section "Conservative p -values" (see Fig. S2 in Jabot and Chave *in press*).

Since such a test offers no mechanistic basis for interpretation, we also developed a different approach. We extend the neutral model by adding a third parameter δ . Under the model, when δ is negative, rare species have a competitive advantage over abundant ones, while when δ is positive, abundant species have an advantage. In this latter case, we propose to interpret δ as a measure of environmental filtering. **Parthy** computes the d parameter from species abundance data by Approximate Bayesian Computation (ABC). Raw ABC tables are produced, and these may easily be manipulated using a script written in the R software (<http://www.r-project.org/>), so as to produce δ estimates and graphs of the approximate posterior distributions.

Hardware and program installation

Parthy is written in C++. It has been compiled using the Linux-emulated environment cygwin on Windows and its C++ GNU compiler g++. It has been tested on the Windows XP operating software, but is expected to run on most environments.

To install **Parthy**, download the distribution into the same directory as your datafile, and unzip the folder. This folder contains the executable (.exe file) and example files to ensure the proper functionality of the program.

Running the program

1- Formatting your data

Species abundance data should be stored in a file with the ".txt" extension (simple text file). One abundance value should be reported by line, and the abundance vector does not need to be sorted. For instance, a typical datafile has the following structure

```
abundance_species_1  
abundance_species_2  
...
```

abundance_species_n

For instance, if your sample consists of (3 willows, 5 oaks, 1 eucalyptus), then the input datafile reads :

```
3
5
1
```

You may want to analyze several species surveys at once. This is an option of the Parthy program. In this case, the several censuses should simply be formatted as vectors like in the example above and stacked one after the other with a '&' symbol separating two consecutive census vectors.

Example: If your dataset consists of one census of (3 willows, 5 oaks, 1 eucalyptus), one census of (1 willows, 15 oaks), and one census of (2 willows, 2 oaks, 3 poplars), then the input datafile should read:

```
3
5
1
&
1
15
&
2
2
3
```

NB: make sure that the last line has a carriage return.

2- Setting up the ABC inference (prior distributions, number of simulations, etc...)

ABC inference (Approximate Bayesian Computing) is based on the sampling from a prior distribution for the parameters. Thus, the prior information should be provided to **Parthy**, and documented in the file "prior-distributions_ABC.txt". This file must be placed in the same directory as the **Parthy** executable and the data file. The different fields correspond to:

- *ln(theta)_min*: lower bound of the uniform prior distribution for the logarithm of the fundamental biodiversity number, parameter $\ln(\theta)$.
- *ln(theta)_max*: upper bound of the uniform prior distribution for parameter $\ln(\theta)$.
- *ln(I)_min*: lower bound of the uniform prior distribution for parameter $\ln(I)$. Parameter I is the scaled immigration rate (cf Etienne 2005). It is connected to parameter m by the relationship: $m=I/(I+J-1)$ where J is the size of the local community.
- *ln(I)_max*: upper bound of the uniform prior distribution for parameter $\ln(I)$.
- *delta_min*: lower bound of the uniform prior distribution for parameter δ .
- *delta_max*: upper bound of the uniform prior distribution for parameter δ .
- *J_largecom*: size of the large neutral community used as a source of immigrants for the local community following a non-neutral dynamics (see Jabot and Chave *in press*).

If the indicated value is less than three times the size J of the local community, **Parthy** enforces the value $J_largecom = 3 \times J$.

- *nb_simul_ABC*: number of ABC runs. In the example, it is set to 2000 for computing speed, but in real applications, it should be equal to 50,000 or more, so as to obtain a sufficiently smooth approximate posterior distribution (Jabot and Chave *in press*).
- *percentsimul*: percentage of simulations used to compute the approximate posterior distribution. For instance, when this percentage equals 0.001, only the 0.1% simulations, the closest to the empirical data, are retained.
- *nb_simul_neutrality-test*: number of neutral communities simulated for the neutrality test.

3- Running the program

Double-click on the **Parthy** executable file (delta-inference.exe) and answer the questions in the command-line window. The keys '0', 'n', or 'no' may be entered for 'no', and '1', 'y', or 'yes' for 'yes'. You may want to first test the program with the dataset provided in the distribution (test.txt). The typical result produced on our machine is provided in the Box below.

```
ESTIMATING PARAMETER DELTA FOR THE DISPERSAL-LIMITED NON-
NEUTRAL MODEL BY ABC
This program estimates the parameter delta of deviation from
neutrality using the approach of Jabot and Chave (submitted).
For more details, see the manual.
Options for answering yes/no questions: 1, y, or yes for
'yes'; 0, n, or no for 'no'.

Please enter the data file name (without '.txt')
```

test

```
Input file: test.txt
Reading the file stats ...
Number of samples: 2
In sample 1, number of individuals: 409, number of species:
84, Shannon's index: 3.9018
Do you want to perform the neutrality test in your sample(s)?
```

yes

```
Do you want to infer Delta value(s) in your sample(s)?
```

yes

```
NEUTRALITY TEST ...
SAMPLE 1: Number of individuals: 409
Composition of neutral simulations:
501 simulations with parameter Theta=339.752 and parameter
I=38.6749
```

```

499 simulations with parameter Theta=339.752 and parameter
I=38.6749
Neutral Simulations running...
100 200 300 400 500 600 700 800 900 1000

DELTA INFERENCE ...
100 200 300 400 500 600 700 800 900 1000 1100 1200 1300 1400
1500 1600 1700 1800 1900 2000

```

At the end of the simulation, the command line window closes automatically. The output files are:

- an ABC table for the first sample, named “ABC_J409_S84.txt”.
- This table contains the parameter values used for simulating the different communities, together with the summary statistics of these communities (species richness S , and Shannon’s index H).
- a file “test_out.txt” containing the summary statistics, and the filename of the ABC table corresponding to each sample.
- a file “test_outR.txt” which can be copied/pasted into the R software so as to compute the mode of the approximate posterior distributions for parameter δ , and to output graphically this posterior distribution. The estimates for parameter δ are written (by this R script) in the file “test_out2.txt”.
- an ABC table for neutral simulation fitting the species richness of the first sample, named “ABC_neutral_J409_S84.txt”.
- a file “test_test-neutrality_out.txt”, where the p -value of the neutrality test can be found.

Post-processing with the R software

We suggest you to use the freely available R statistical software to compute approximate posterior distributions and their mode from ABC tables (to install R, and for more information, please check <http://www.r-project.org/>). In R, make sure that your working directory is the one where you saved your data file (go to menu ‘File’, then ‘Change the directory’). The R commands for computing approximate posterior distributions and plotting them are provided in the output file “name_of_file_outR.txt” (in the example, this is the file test_outR.txt).

Troubleshooting

For problems with the R software, an online help is available on the R website <http://www.r-project.org/>

Any feedback on the software is most welcome (franck.jabot@m4x.org).

References

Etienne R.S., 2005. A new sampling formula for neutral biodiversity. *Ecology Letters*, **8**: 253-260.

Hubbell S.P., 2001. *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press, Princeton, NJ.

Jabot F. and Chave J., *in press*. Analyzing tropical forest tree species abundance distributions using a nonneutral model and through Approximate Bayesian inference. *American Naturalist*.

Wright S., 1931. Evolution in Mendelian populations. *Genetics*, **16**: 97–159.