Help: gapw() function

The gap-weighting (Thiele, 1993):

The treatment of continuous characters in cladistics is an important issue. One of the most used methods is the gap-weighting (Thiele, 1993). It tends to keep the differences between taxa within the coding. Thiele's wish was to increase the number of states used to discretize the continuous characters in order to increase the resolution. Wiens (2001) proposed the step-matrix-gap-weighting, a derivative of Thiele's method which greatly increased the resolution (1000 states).
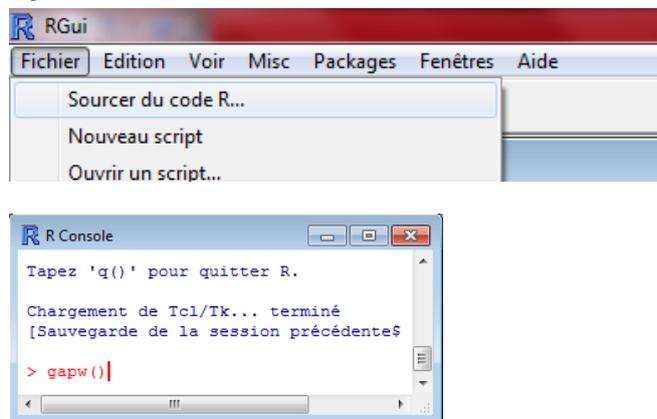
The function gapw():

The "gapw" function can transform data using the method of gap-weighting up to a maximum resolution of 65,000 states. The effects of resolution in phylogenetic analysis are discussed in the literature, this function can automatically transform the data set by several number of states, that is to say several resolutions. All matrices are generated in nexus or tnt format, and can be run with PAUP and TNT.

Two types of studies are facilitated by this function. First, users can treat imported data set with both discrete and continuous characters. Second, theoretical studies on gap-weighting can be performed thanks to the possibility to generate artificial data set. The user create a matrix and choose the number of continuous, discrete characters, taxons, the minimum and maximum values and the number of digits for the continuous characters, the number of different states and their occurrence probability for the discrete characters. All the generated characters are parsimony-informative.
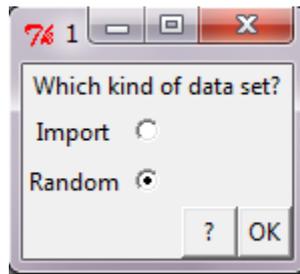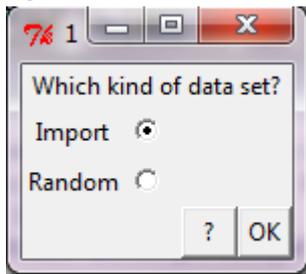
How to use de function:

1) Load the function:

You have to load the function.

Fichier->Sourcer du code R and select the function.

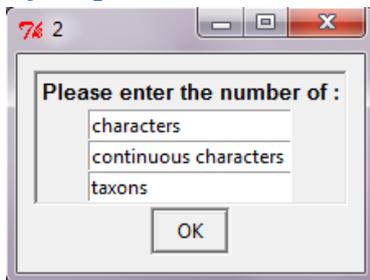Next, you have to write the function in the R console and [enter].

2) Kind of data set:

To import a data set, click on « Import » and OK . A file .csv is required. Taxons have to be on rows and characters on columns. For digits, the comma is required. Continuous characters have to be located at the beginning.

To create data sets with random data, click on « Random » and OK .
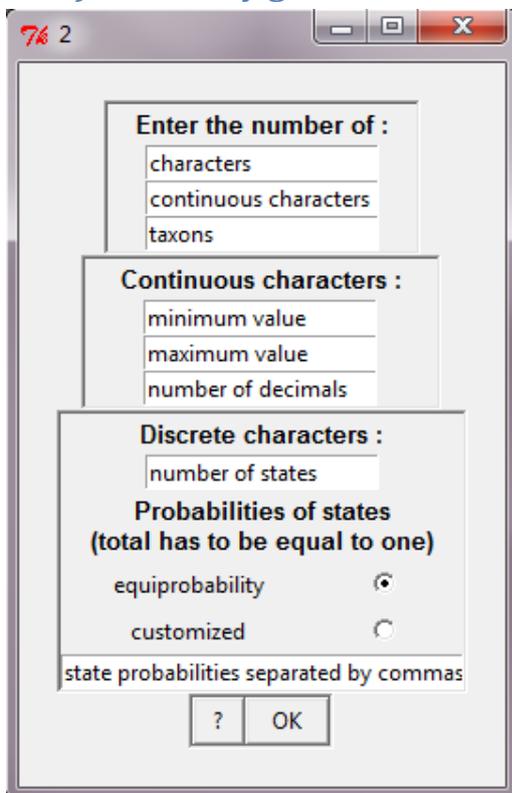
3) Imported data set:

Fill in the box:

- Number of characters in the matrix
- Number of continuous characters (at the beginning of the matrix)
- Number of taxons

3bis) Randomly generated data set:

Fill in the three first boxes to determine the size of the matrix :
- Number of characters in the matrix
- Number of continuous characters (at the beginning of the matrix)
- Number of taxons

In the second insert enter the parameters for the sampling of the continuous characters :
- Minimum value
- Maximum value
- Number of decimals

In the third insert enter the parameters for the sampling of the discrete characters :
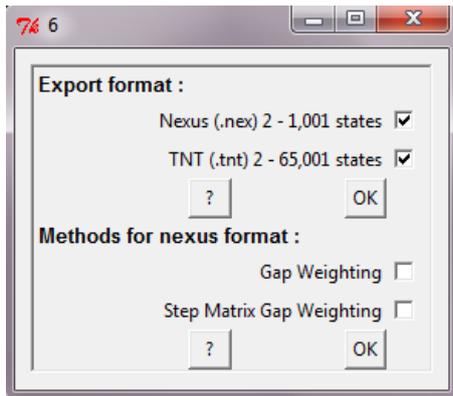- Number of different states
- Probability of each state:
  - o equiprobability
  - o customized : you have to enter in the box a list with the probability of each state separated by commas. WARNING : the decimal separator is the dot ".".

4) Export format and discretization:
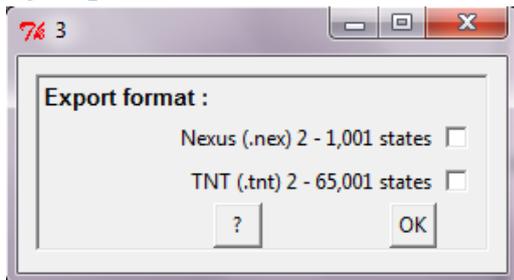
Choose the format:
- Nexus or tnt (.nex ou .tnt)
- The numbers of states next to the format are the maximum one for the current phylogenetic programms.
- For more informations on these methods, see Thiele (1993) for gap-weighting and Wiens (2001) for step-matrix-gap-weighting.

For each couple (method / format) selected, you have to enter the number of states for discretization. For each method, parameters can be enter in two different ways.

First, you can enter a list containing one or more values. Two values are separated by a comma. Example: 2,5,10.

Second, you can enter a sequence. For example all values between 10 (from) and 50 (to) every 20 unities (by). Example : From 10 to 50 by 20 corresponds to the following list : 10,30,50.
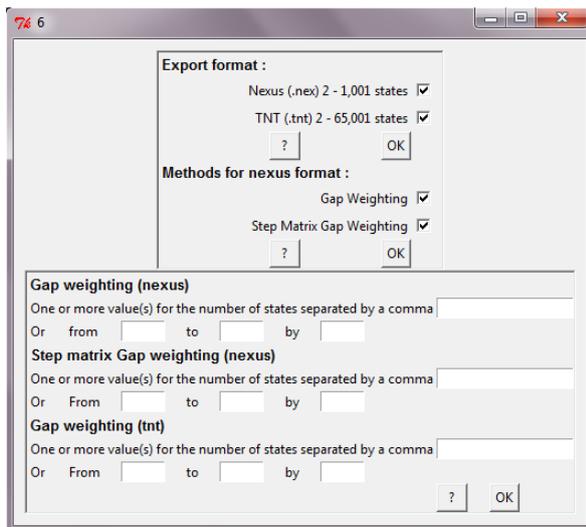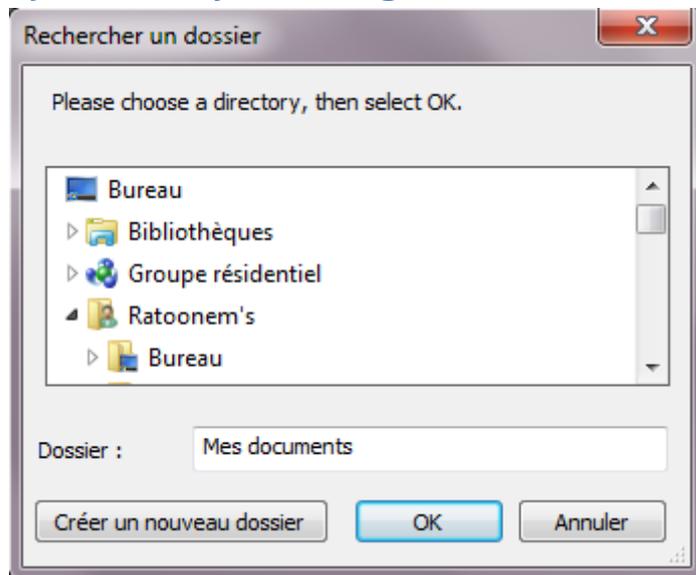
5) Matrix selection (for imported data set only):
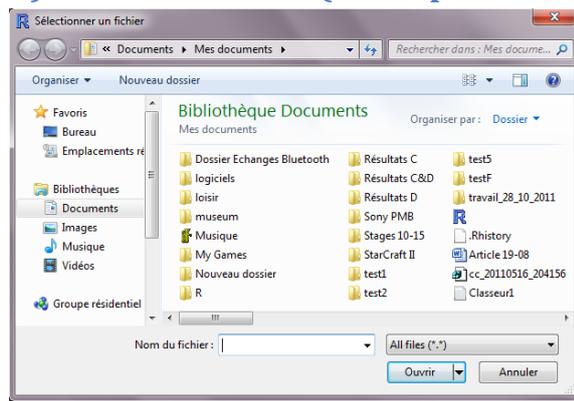
Choose your matrix on the window. A .csv file is required. Taxons have to be on rows and characters on columns. For digits, the comma is required. Continuous characters have to be located at the beginning. For decimals, the character is the comma ",".

6) A directory for saving :

Choose the directory for saving the files generated.

References :

Thiele, K. 1993. The holy grail of the perfect character: The cladistic treatment of morphometric data. Cladistics 9, 275-304.

Wiens, J. J. 2001. Character analysis in morphological phylogenetics: Problems and solutions. Syst. Biol. 50, 689-699.