How to analyze morphological data

Saturation

Cite (Dávalos *et al.* 2012) when using this protocol.

Morphological data sets can exhaust character states as they incorporate more species, so that further changes erode phylogenetic signal (Wagner 2000). Constraints on design arising from development or function can result in homoplasy and lead to long-branch attraction. We evaluated saturation—or exhaustion—in character states in the morphological data using the approach of (Wagner 2000). First, we converted all ordered characters to unordered, optimized these on the maximum parsimony phylogeny using acctran (accelerated transformation), and counted the number of states and steps for each branch, beginning with the oldest branches in the phylogeny.

1) Obtain a MP phylogeny. Apply character ordering as needed. Save.
2) Convert all characters to unordered. Optimize on the saved phylogeny. The pictured results are from PAUP (Swofford 2002). Use **deltran**. Save the optimization log.
3) Use Excel to process the resulting log so that one column has the step, and another the state, beginning with the oldest branches. Example:

```
Character                        CI   Steps            Changes

--------------------------------------------------------------------------------
-

1 (Pelage differentiated)       0.100    1        node_148 1 ==> 0 Ardops
nichollsi

                                         1        node_137 1 ==> 0 Sturnira lilium

                                         1        node_133 1 --> 0 node_102

                                         1         node_98 0 --> 1 node_95

                                         1         node_93 1 ==> 0 Lionycteris
spur

                                         1        node_100 0 --> 1 Diphylla
ecaudat

                                         1        node_109 1 ==> 0 node_108

                                         1        node_121 1 ==> 0 node_120

                                         1        node_114 1 ==> 0 Thyroptera
trico
```

You need to first organize the text so that it will go on excel. Use a text editor (e.g., Bbedit) or sed to edit the text to that the number of steps (this will be a

long series of "1"), the node, and the transformation (change from one state to another) align perfectly.
4) Fix by hand if necessary. Do not sort.
5) Insert a column to the right of the character number, which should be the first column on the left and should have numbers interspersed with empty cells, but nothing else.
6) In the second row of the newly inserted blank column type the following formula:

`=IF(A2>0,A2,B1)`

Where A is the column with character numbers and blank cells, and B is the column where you perform the operation.
7) Pull down the formula to the entire array so that the column becomes a character number column without empty cells. Fill the first row with the number of the character. The blank cells are now populated by the character number.
8) After checking visually that it all aligns and all values match the character number in the first column, copy the new column and paste special "values".
9) Select the columns that hold the starting node and end node or terminal. Search for "node_" and replace with nothing. This will leave the start column as a series of numbers and the end column as a combination of numbers and tip names. At this point you might want to label the headers to guide subsequent steps.
10) Sort the data ascending first by **character**, then by **end state**, then by **start node**. This will rearrange the data so that the oldest transformations within a character are above and the most recent below (but check how nodes are numbered and sort start node ascending or descending depending on how the nodes are numbered).
11) All these sorts will fail if you have not copy-pasted as values the character number column.
12) After the last sort you can tally which steps result in state changes. Insert a new column to the right of existing columns. Enter the following formula in the second cell (the first cell of the column should be a header indicating "end state". Pull down to copy to the whole column.

`=IF(A2=A1,IF(G2=G1,0,1),1)`

Where A is the character column and G is the end state column.
13) The formula will produce a 1 every time there is a new value, and a 0 otherwise. Copy the new column and paste special "values". **This is a critical step**.
14) Sort the data ascending by age, with oldest nodes first by sorting first by **start node** then by **end node**. This may be ascending or descending on the start and end nodes depending on how nodes are numbered on the tree, what matters is for backbone nodes to go first in the start node and for tips to be last on the end node. This will rearrange the data from oldest to most

recent changes, independent of what character the states changed in. **This is a critical step**.

15) Create a column with the number of steps, this is easily achieved by using the following formula on a new column:

```
=SUM(B$2:B2)
```

Where B is the number of steps (from PAUP log "steps") column.

16) Pull down the formula, copy the new column and paste special "values".

17) Create a column with the number of states, this is easily achieved by using the following formula on a new column:

```
=SUM(H$2:H2)
```

Where H is the column with states changing or not that you created earlier.

18) Pull down the formula, copy the new column and paste special "values".

19) You now have the three main elements of the state to step curve, the column with the number of steps, the column with the number of states, and the column with whether steps have been seen before or not. Copy/paste them to a new spreadsheet with three columns and save as .csv file. The headers in this file should short and crisp.

20) Save the operations spreadsheet for your records.

Dávalos, L. M., A. L. Cirranello, J. H. Geisler and N. B. Simmons, 2012 Understanding phylogenetic incongruence: lessons from Phyllostomid bats. Biological Reviews 87**:** 991-1023.

Swofford, D. L., 2002 PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). pp. Sinauer Associates, Sunderland, Massachusetts.

Wagner, P. J., 2000 Exhaustion of morphologic character states among fossil taxa. Evolution 54**:** 365-386.