

## DATA FOR A STRATAPHY STRATOCLADISTIC SEARCH

Stratocladistic analysis requires additional data to supplement what is traditionally collected for cladistic analysis. Specifically, stratigraphic data and autapomorphies are necessary components of a stratocladistic data set.

As previously mentioned, one of the chief differences between stratocladistics and cladistics is the use of stratigraphic data in the process of phylogenetic inference in stratocladistics. These data are in the form of the temporal or stratigraphic intervals of first and last appearance of taxa. They are coded as discrete states in the character matrix, rather than as continuous values (e.g., absolute dates or meters of section). Taxa that span multiple intervals are coded as being polymorphic, and possess all states corresponding to intervals in which they are sampled. Exactly how stratigraphic intervals should be defined for stratocladistic analysis and how finely they should be divided remains an unresolved matter, and one for further research (see Smith 2000; Fisher et al. 2002).

Autapomorphies are often explicitly excluded from cladistic analysis, as they offer no information regarding sister-clade relationships among taxa. However, they are required for stratocladistic analysis because they potentially provide information about the optimality of hypotheses of ancestor-descendant relationships among taxa. Operationally, they affect the optimality of assignments of taxa as ancestors, in which consideration of a single taxon as an ancestor weighs the possible stratigraphic debt savings against the possible morphologic debt incurred. This morphologic debt will equal the number of autapomorphies, if character changes have weight of one, so the exclusion of autapomorphies will bias the results toward finding more optimally assigned ancestors than would be the case if autapomorphies were included. Aside from this obvious issue of erroneously assigning ancestors, failure to include autapomorphies can be a computational challenge, as it can result in a large number of optimal trees when multiple taxa can be either ancestors or terminal taxa on a tree without a change in total debt. In such a case, for a given topology, numerous permutations of ancestral assignments of such taxa will be equally optimal. Searching vast numbers of equally optimal trees dramatically increases search times. Therefore, it is strongly suggested that users resist the temptation to append a single stratigraphic character to existing cladistic matrices for analysis in StrataPhy without first adding autapomorphies of included taxa.

## FORMATTING NEXUS FILES FOR STRATAPHY ANALYSIS

StrataPhy reads files in the standard NEXUS format (D. R. Maddison et al. 1997) common to many modern computer programs for phylogenetic analysis. See Maddison et al. (1997) for a more detailed description of the NEXUS format, and for the syntax of commands common to most NEXUS files. Here, we restrict our description of NEXUS syntax to only that necessary to perform stratocladistic analyses in StrataPhy. An example NEXUS file with a StrataPhy block is included in the Appendix. In general, NEXUS files are ASCII text files and may be created in any text editor. Alternatively, they can be created and edited in programs explicitly designed to produce NEXUS files, such as MacClade or Mesquite (W. P. Maddison and Maddison 2004). NEXUS files for StrataPhy include labels of taxa, character data, as well as a stratigraphic character. Note that at the present time MacClade is the only program that fully supports stratigraphic data and the corresponding character type, as well as the assignment of taxa as ancestors. Ideally, StrataPhy and MacClade should be used in concert for the straightforward creation of data sets. In MacClade, stratigraphic characters can be specified, and will automatically be incorporated into the NEXUS file. Other programs such as Mesquite do not support stratigraphic data explicitly, and assumptions about stratigraphic characters must be added manually in an additional step.

StrataPhy currently supports both ordered and unordered categorical character types, as well as the stratigraphic character type in MacClade (D. R. Maddison and Maddison 2005). To make StrataPhy compatible with MacClade for NEXUS file creation and tree viewing and analysis, StrataPhy shares some critical restrictions on character states with MacClade. StrataPhy support a maximum of 34 character states: the numbers "0" through "9" and letters "A" through "Z", regardless of case, are all valid character states, except for "I" and "O", which are invalid. Therefore due to restrictions imposed by MacClade, the stratigraphic character can include no more than 34 intervals.

StrataPhy allows differential user-defined weighting of characters, including the stratigraphic character. Character types and weights are specified using the TYPESET and WTSET options, respectively, in the ASSUMPTIONS block of the NEXUS data file, as is standard NEXUS format. Characters present in a data matrix contained in the NEXUS file can be excluded from analysis by creating an exclusion set (EXSET) in the ASSUMPTIONS block. Note that MacClade refers to exclusion sets as "Inclusion Sets." This option records the character numbers of characters to be excluded prior to StrataPhy analysis. StrataPhy only performs stratocladistic analysis using the "active" TYPESET, WTSET or EXSET set, which are each designated with an asterisk (\*) in the NEXUS file (see Appendix). Typically, these sets are those active (i.e., selected) at the time the

file was saved to disk. It is possible that a user might explicitly store a single TYPESET, for example, but if the user were to change the types of any character, then save the entire file to disk, the untitled TYPESET in effect at the time of the save would be that used by StrataPhy. Care must be taken to designate the TYPESET, WTSET and EXSET immediately before saving NEXUS files in MacClade, and we encourage users to review their selections prior to analysis by opening NEXUS files in any standard text editor. Individual taxa in a matrix cannot be excluded from analysis within StrataPhy within the ASSUMPTIONS block, so taxa to be excluded must be deleted from the NEXUS file prior to analysis with StrataPhy.

The current version of StrataPhy utilizes a command line for inputting the file name of the NEXUS file to be analyzed (instead of a graphic user interface), so all parameter settings for the analysis must be specified within the NEXUS file prior to analysis. These parameters are set by placing commands in a new block labeled "STRATAPHY" in the NEXUS file. Syntax of the StrataPhy block is given below. Terms in italics would be replaced by user-defined values, and those in braces represent the range of possible parameters with the default underlined.

```
BEGIN STRATAPHY;  
  MAXTREES = maximum-number-of-trees;  
  SEARCHREPS = number-of-branch-swapping-replicates;  
  ANCREPS = number-of-ancestral-search-replicates;  
  ANCTYPE = { heuristic | exhaustive };  
  OUTGROUP outgroup1 outgroup2 outgroup3;  
END;
```

**MAXTREES** – This command sets the maximum number of optimal trees to be saved in a heuristic search. It is also the maximum number of trees saved during single reps of a heuristic search, although if these are not globally optimal over all previous reps, they will not be saved. StrataPhy currently cannot adjust MAXTREES dynamically during a search, so this must be specified prior to the analysis. The default setting of MAXTREES is set to 1000.

**SEARCHREPS** – This command sets the number of heuristic search replicates to be performed. Each of these replicates begins with a tree built by random taxon addition. The default number of search replicates is 10.

**ANCREPS** – This command sets the number of replicate searches for optimal assignments of ancestors reps per tree produced by branch swapping. By default, StrataPhy performs only a single ancestral search replicate.

**ANCTYPE** – This command designates the type of search for optimal assignments of ancestors, either heuristic or exhaustive. Heuristic search (described above) is the default.

**OUTGROUP** – This command designates taxa that will be used as the outgroup. Taxon labels must exactly match the spelling of those in the translation table and matrix in the NEXUS file, and multiple outgroup taxa are separated by single spaces. If no outgroup is specified, StrataPhy assumes all taxa are in the ingroup. In this case, if a stratigraphic character is present, StrataPhy will search over all possible rooted trees, so it is recommended that at least one outgroup taxon is specified.