Command Reference for MrBayes ver. 3.1

(c) John P. Huelsenbeck and Fredrik Ronquist

1. Command summary

Commands that are available from the command line or from a MrBayes block include:

```
About           -- Describes the program
Acknowledgments -- Shows program acknowledgments
Charset         -- Assigns a group of sites to a set
Charstat        -- Shows status of characters
Citations       -- Appropriate citation of program
Comparetree     -- Compares the trees from two tree files
Constraint      -- Defines a constraint on tree topology
Ctype           -- Assigns ordering for the characters
Databreaks      -- Defines nucleotide pairs (doublets) for stem models
Delete          -- Deletes taxa from the analysis
Deroot          -- Deroots user tree
Disclaimer      -- Describes program disclaimer
Exclude         -- Excludes sites from the analysis
Execute         -- Executes a file
Help            -- Provides detailed description of commands
Include         -- Includes sites
Link            -- Links parameters across character partitions
Log             -- Logs screen output to a file
Lset            -- Sets the parameters of the likelihood model
Manual          -- Prints a command reference to a text file
Mcmc            -- Starts Markov chain Monte Carlo analysis
Mcmcp           -- Sets the parameters of a chain (without starting analysis)
Outgroup        -- Changes outgroup taxon
Pairs           -- Defines nucleotide pairs (doublets) for stem models
Partition       -- Assigns a character partition
Plot            -- Plots parameters from MCMC analysis
Prset           -- Sets the priors for the parameters
Props           -- Set proposal probabilities
Quit            -- Quits the program
Report          -- Controls how model parameters are reported
Restore         -- Restores taxa
Root            -- Roots user tree
Set             -- Sets run conditions and defines active data partition
Showmatrix      -- Shows current character matrix
Showmodel       -- Shows model settings
Showtree        -- Shows user tree
Sump            -- Summarizes parameters from MCMC analysis
```

```
Sumt            -- Summarizes trees from MCMC analysis
Taxastat        -- Shows status of taxa
Taxset          -- Assigns a group of taxa to a set
Unlink          -- Unlinks parameters across character partitions
Usertree        -- Defines a single user tree
Version         -- Shows program version

Commands that should be in a NEXUS file (data
block or trees block) include:

Begin           -- Denotes beginning of block in file
Dimensions      -- Defines size of character matrix
End             -- Denotes end of a block in file
Endblock        -- Alternative way of denoting end of a block
Format          -- Defines character format in data block
Matrix          -- Defines matrix of characters in data block
Translate       -- Defines alternative names for taxa
Tree            -- Defines a tree from MCMC analysis

Note that this program supports the use of the shortest unambiguous
spelling of the above commands (e.g., "exe" instead of "execute").
------------------------------------------------------------------------

************************************************************************
*                                                                      *
*  2. MrBayes commands                                                 *
*                                                                      *
************************************************************************

------------------------------------------------------------------------
About

This command provides some general information about the program.
------------------------------------------------------------------------
------------------------------------------------------------------------
Acknowledgments

This command shows the authors' acknowledgments.
------------------------------------------------------------------------
------------------------------------------------------------------------
Charset

This command defines a character set. The format for the charset command
is

   charset <name> = <character numbers>

For example, "charset first_pos = 1-720\3" defines a character set
called "first_pos" that includes every third site from 1 to 720.
The character set name cannot have any spaces in it. The slash (\)
is a nifty way of telling the program to assign every third (or
second, or fifth, or whatever) character to the character set.
This option is best used not from the command line, but rather as a
line in the mrbayes block of a file. Note that you can use "." to
stand in for the last character (e.g., charset 1-.\3).
------------------------------------------------------------------------
------------------------------------------------------------------------
```

Charstat

This command shows the status of all the characters. The correct usage is

   charstat

After typing "charstat", the character number, whether it is excluded or included, and the partition identity are shown. The output is paused every 100 characters. This pause can be turned off by setting autoclose to "yes" (set autoclose=yes).

Citations

This command shows a thorough list of citations you may consider using when publishing the results of a MrBayes analysis.

Comparetree

This command compares the trees in two files, called "filename1" and "filename2". It will output a bivariate plot of the split frequencies as well as plots of the tree distance as a function of the generation. The plots can be used to get a quick indication of whether two runs have converged onto the same set of trees. The "Comparetree" command will also produce a ".parts" file and a ".dists" file (these file endings are added to the end of the "Outputname"). The ".parts" file contains the paired split frequencies from the two tree samples; the ".dists" file contains the tree distance values. Note that the "Sumt" command provides a different set of convergence diagnostics tools that you may also want to explore. Unlike "Comparetree", "Sumt" can compare more than two tree samples and will calculate consensus trees and split frequencies from the pooled samples.

Parameter	Options	Current Setting
Filename1	<name>	temp.t
Filename2	<name>	temp.t
Outputname	<name>	temp.comp
Burnin	<number>	0

Constraint

This command defines a tree constraint. The format for the constraint command is

   constraint <constraint name> <probability> = <list of taxa>

A list of taxa can be specified using a taxset, taxon names, or taxon numbers. A probability must also be specified. For now, MrBayes ignores this probability value and treats the constraint as an absolute requirement of trees. That is, trees that are not compatible with the constraint have zero prior (and hence zero posterior) probability.

Future releases of MrBayes will use the probability value to determine how

much more probable a tree is that contains the constraint than a tree without the constraint. For example, the following command

```
   constraint example 100 = taxon_2 taxon_3
```

defines a constraint called "example" that includes two taxa. In future releases of MrBayes, trees that contain a clade with those two taxa together will have a prior probability that is 100 times that of trees without the constraint. In the current version, the probability value will be ignored and trees without the two taxa together will not be sampled.

If you are interested in inferring ancestral states for a particular node, you need to constrain that node first using the 'constraint' command. For more information on how to infer ancestral states, see the help for the 'report' command.

It is important to note that simply defining a constraint using this command is not sufficient for the program to actually implement the constraint in an analysis. You must also specify the constraints using 'prset topologypr = constraints (<name of constraint>)'. For more information, see the help on the 'prset' command.

Ctype

This command sets the character ordering for standard-type data. The correct usage is:

```
   ctype <ordering>:<characters>
```

The available options for the <ordering> specifier are:

```
  unordered    -- Movement directly from one state to another is
                  allowed in an instant of time.
  ordered      -- Movement is only allowed between adjacent characters.
                  For example, perhaps only between 0 <-> 1 and 1 <-> 2
                  for a three state character ordered as 0 - 1 - 2.
  irreversible -- Rates of change for losses are 0.
```

The characters to which the ordering is applied is specified in manner that is identical to commands such as "include" or "exclude". For example,

```
   ctype ordered: 10 23 45
```

defines charactes 10, 23, and 45 to be of type ordered. Similarly,

```
   ctype irreversible: 54 - 67  71-92
```

defines characters 54 to 67 and characters 71 to 92 to be of type irreversible. You can use the "." to denote the last character, and "all" to denote all of the characters. Finally, you can use the specifier "\" to apply the ordering to every n-th character or you can use predefined charsets to specify the character.

Only one ordering can be used on any specific application of ctype.

If you want to apply different orderings to different characters, then you need to use ctype multiple times. For example,

```
ctype ordered: 1-50
ctype irreversible: 51-100
```

sets characters 1 to 50 to be ordered and characters 51 to 100 to be irreversible.

The ctype command is only sensible with morphological (here called "standard") characters. The program ignores attempts to apply character orderings to other types of characters, such as DNA characters.

Databreaks

This command is used to specify breaks in your input data matrix. Your data may be a mixture of genes or a mixture of different types of data. Some of the models implemented by MrBayes account for nonindependence at adjacent characters. The autocorrelated gamma model, for example, allows rates at adjacent sites to be correlated. However, there is no way for such a model to tell whether two sites, adjacent in the matrix, are actually separated by many kilobases or megabases in the genome. The databreaks command allows you to specify such breaks. The correct usage is:

```
databreaks <break 1> <break 2> <break 3> ...
```

For example, say you have a data matrix of 3204 characters that include nucleotide data from three genes. The first gene covers characters 1 to 970, the second gene covers characters 971 to 2567, and the third gene covers characters 2568 to 3204. Also, let's assume that the genes are not directly adjacent to one another in the genome, as might be likely if you have mitochondrial sequences. In this case, you can specify breaks between the genes using:

```
databreaks 970 2567;
```

The first break, between genes one and two, is after character 970 and the second break, between genes two and three, is after character 2567.

Delete

This command deletes taxa from the analysis. The correct usage is:

```
delete <name and/or number and/or taxset> ...
```

A list of the taxon names or taxon numbers (labelled 1 to ntax in the order in the matrix) or taxset(s) can be used. For example, the following:

```
delete 1 2 Homo_sapiens
```

deletes taxa 1, 2, and the taxon labelled Homo_sapiens from the analysis. You can also use "all" to delete all of the taxa. For example,

```
delete all
```

deletes all of the taxa from the analysis. Of course, a phylogenetic analysis that does not include any taxa is fairly uninteresting.

Deroot

This command deroots the user tree. If the tree is already unrooted, a warning is issued. The correct usage is "deroot".

Disclaimer

This command shows the disclaimer for the program. In short, the disclaimer states that the authors (John Huelsenbeck and Fredrik Ronquist) are not responsible for any silly things you may do to your computer or any unforseen but possibly nasty things the computer program may inadvertently do to you.

Exclude

This command excludes characters from the analysis. The correct usage is

```
exclude <number> <number> <number>
```

or

```
exclude <number> - <number>
```

or

```
exclude <charset>
```

or some combination thereof. Moreover, you can use the specifier "\" to exclude every nth character. For example, the following

```
exclude 1-100\3
```

would exclude every third character. As a specific example,

```
exclude 2 3 10-14 22
```

excludes sites 2, 3, 10, 11, 12, 13, 14, and 22 from the analysis. Also,

```
exclude all
```

excludes all of the characters from the analysis. Excluding all characters does not leave you much information for inferring phylogeny.

Execute

This command executes a file called <file name>. The correct usage is:

```
execute <file name>
```

For example,

```
execute replicase.nex
```

would execute the file named "replicase.nex". This file must be in the same directory as the executable.

Help

This command provides useful information on the use of this program. The correct usage is

```
help
```

which gives a list of all available commands with a brief description of each or

```
help <command>
```

which gives detailed information on the use of <command>.

Include

This command includes characters that were previously excluded from the analysis. The correct usage is

```
include <number> <number> <number>
```

or

```
include <number> - <number>
```

or

```
include <charset>
```

or some combination thereof. Moreover, you can use the specifier "\" to include every nth character. For example, the following

```
include 1-100\3
```

would include every third character. As a specific example,

```
include 2 3 10-14 22
```

includes sites 2, 3, 10, 11, 12, 13, 14, and 22 from the analysis. Also,

```
include all
```

includes all of the characters in the analysis. Including all of the characters (even if many of them are bad) is a very total-evidence-like thing to do. Doing this will make a certain group of people very happy. On the other hand, simply using this program would make those same people unhappy.

```
------------------------------------------------------------------------
Link

This command links model parameters across partitions of the data. The
correct usage is:

   link <parameter name> = (<all> or <partition list>)

The list of parameters that can be linked includes:

   Tratio          -- Transition/transversion rate ratio
   Revmat          -- Substitution rates of GTR model
   Omega           -- Nonsynonymous/synonymous rate ratio
   Statefreq       -- Character state frequencies
   Shape           -- Gamma shape parameter
   Pinvar          -- Proportion of invariable sites
   Correlation     -- Correlation parameter of autodiscrete gamma
   Switchrates     -- Switching rates for covarion model
   Brlens          -- Branch lengths of tree
   Topology        -- Topology of tree
   Speciationrates -- Speciation rates for birth-death process
   Extinctionrates -- Extinction rates for birth-death process
   Theta           -- Parameter for coalescence process
   Growthrate      -- Growth rate of coalescence process

For example,

   link shape=(all)

links the gamma shape parameter across all partitions of the data.
You can use "showmodel" to see the current linking status of the
characters. For more information on this command, see the help menu
for link's converse, unlink ("help unlink");
------------------------------------------------------------------------
------------------------------------------------------------------------
Log

This command allows output to the screen to also be output to a file.
The useage is:

   log start/stop filename=<name> append/replace

The options are:

Start/Stop     -- Starts or stops logging of output to file.
Append/Replace -- Either append to or replace existing file.
Filename       -- Name of log file (currently, the name of the log
                  file is "log.out").
------------------------------------------------------------------------
------------------------------------------------------------------------
Lset

This command sets the parameters of the likelihood model. The likelihood
function is the probability of observing the data conditional on the phylo-
genetic model. In order to calculate the likelihood, you must assume a
model of character change. This command lets you tailor the biological
assumptions made in the phylogenetic model. The correct usage is
```

```
   lset <parameter>=<option> ... <parameter>=<option>
```

For example, "lset nst=6 rates=gamma" would set the model to a general model of DNA substition (the GTR) with gamma-distributed rate variation across sites.

Options:

Applyto -- This option allows you to apply the lset commands to specific partitions. This command should be the first in the list of commands specified in lset. Moreover, it only makes sense to be using this command if the data have been partitioned. A default partition is set on execution of a matrix. If the data are homogeneous (i.e., all of the same data type), then this partition will not subdivide the characters. Up to 30 other partitions can be defined, and you can switch among them using "set partition=<partition name>". Now, you may want to specify different models to different partitions of the data. Applyto allows you to do this. For example, say you have partitioned the data by codon position, and you want to apply a nst=2 model to the first two partitions and nst=6 to the last. This could be implemented in two uses of lset:

```
   lset applyto=(1,2) nst=2

   lset applyto=(3) nst=6
```

The first applies the parameters after "applyto" to the first and second partitions. The second lset applies nst=6 to the third partition. You can also use applyto=(all), which attempts to apply the parameter settings to all of the data partitions. Importantly, if the option is not consistent with the data in the partition, the program will not apply the lset option to that partition.

Nucmodel -- This specifies the general form of the nucleotide substitution model. The options are "4by4" [the standard model of DNA substitution in which there are only four states (A,C,G,T/U)], "doublet" (a model appropriate for modelling the stem regions of ribosomal genes where the state space is the 16 doublets of nucleotides), and "codon" (the substitution model is expanded around triplets of nucleotides--a codon).

Nst -- Sets the number of substitution types: "1" constrains all of the rates to be the same (e.g., a JC69 or F81 model); "2" allows transitions and transversions to have potentially different rates (e.g., a K80 or HKY85 model); "6" allows all rates to be different, subject to the constraint of time-reversibility (e.g., a GTR model).

Code -- Enforces the use of a particular genetic code. The default is the universal code. Other options include "vertmt" for vertebrate mitocondrial DNA, "mycoplasma", "yeast", "ciliates", and "metmt" (for metazoan mitochondrial DNA except vertebrates).

Ploidy -- Specifies the ploidy of the organism. Options are "Haploid" or "Diploid". This option is used when a coalescence prior is used on trees.

Rates -- Sets the model for among-site rate variation. In general, the

rate at a site is considered to be an unknown random variable. The valid options are:

* equal -- No rate variation across sites.
* gamma -- Gamma-distributed rates across sites. The rate at a site is drawn from a gamma distribution. The gamma distribution has a single parameter that describes how much rates vary.
* adgamma -- Autocorrelated rates across sites. The marginal rate distribution is gamma, but adjacent sites have correlated rates.
* propinv -- A proportion of the sites are invariable.
* invgamma -- A proportion of the sites are invariable while the rate for the remaining sites are drawn from a gamma distribution.

Note that MrBayes versions 2.0 and earlier supported options that allowed site specific rates (e.g., ssgamma). In versions 3.0 and later, site specific rates are allowed, but set using the 'prset ratepr' command for each partition.

Ngammacat -- Sets the number of rate categories for the gamma distribution. The gamma distribution is continuous. However, it is virtually impossible to calculate likelihoods under the continuous gamma distribution. Hence, an approximation to the continuous gamma is used; the gamma distribution is broken into ncat categories of equal weight (1/ncat). The mean rate for each category represents the rate for the entire cateogry. This option allows you to specify how many rate categories to use when approximating the gamma. The approximation is better as ncat is increased. In practice, "ncat=4" does a reasonable job of approximating the continuous gamma.

Nbetacat -- Sets the number of rate categories for the beta distribution. A symmetric beta distribution is used to model the stationary frequencies when morphological data are used. This option specifies how well the beta distribution will be approximated.

Omegavar -- Allows the nonsynonymous/synonymous rate ratio (omega) to vary across codons. Ny98 assumes that there are three classes, with potentially different omega values (omega1, omega2, omega3): omega2 = 1; 0 < omega1 <1; and omega3 > 1. Like the Ny98 model, the M3 model has three omega classes. However, their values are less constrained, with omega1 < omega2 < omega3. The default (omegavar = equal) has no variation on omega across sites.

Covarion -- This forces the use of a covarion-like model of substitution for nucleotide or amino acid data. The valid options are "yes" and "no". The covarion model allows the rate at a site to change over its evolutionary history. Specifically, the site is either on or off. When it is off, no substitutions are possible. When the process is on, substitutions occur according to a specified substitution model (specified using the other lset options).

Coding -- This specifies how characters were sampled. If all site patterns had the possibility of being sampled, then "all" should be specified (the default). Otherwise "variable" (only variable characters had the possibility of being sampled), "noabsence" (characters for which all taxa were coded as absent were not sampled), and "nopresence" (characters for which all taxa were coded as present were not sampled. "All" works for all data types. However, the others only work for

morphological (all/variable) or restriction site (all/variable/
noabsence/nopresence) data.

Parsmodel -- This forces calculation under the so-called parsimony model described by Tuffley and Steel (1998). The options are "yes" or "no". Note that the biological assumptions of this model are anything but parsimonious. In fact, this model assumes many more parameters than the next most complicated model implemented in this program. If you really believe that the parsimony model makes the biological assumptions described by Tuffley and Steel, then the parsimony method is miss-named.

Default model settings:

Parameter	Options	Current Setting
Nucmodel	4by4/Doublet/Codon	4by4
Nst	1/2/6	1
Code	Universal/Vertmt/Mycoplasma/ Yeast/Ciliates/Metmt	Universal
Ploidy	Haploid/Diploid	Diploid
Rates	Equal/Gamma/Propinv/Invgamma/Adgamma	Equal
Ngammacat	<number>	4
Nbetacat	<number>	5
Omegavar	Equal/Ny98/M3	Equal
Covarion	No/Yes	No
Coding	All/Variable/Noabsencesites/ Nopresencesites	All
Parsmodel	No/Yes	No

Manual

This command allows you to generate a text file containing help information on all the available commands. This text file can be used as an up-to-date command reference. You can set the name of the text file using the "filename" option; the default is "commref_mb<version>.txt".

Parameter	Options	Current Setting
Filename	<name>	commref_mb3.1.txt

Mcmc

This command starts the Markov chain Monte Carlo (MCMC) analysis. The posterior probability of phylogenetic trees (and other parameters of the substitution model) cannot be determined analytically. Instead, MCMC is used to approximate the posterior probabilities of trees by drawing (dependent) samples from the posterior distribution. This program can implement a variant of MCMC called "Metropolis-coupled Markov chain Monte Carlo", or MCMCMC for short. Basically, "Nchains" are run, with Nchains - 1 of them heated. The chains are labelled 1, 2, ..., Nchains. The heat that is applied to the i-th chain is B = 1 / (1 + temp X i). B is the power to which the posterior probability is raised. When B = 0, all trees have equal probability and the chain freely visits trees. B = 1 is

the "cold" chain (or the distribution of interest). MCMCMC can mix better than ordinary MCMC; after all of the chains have gone through one cycle, two chains are chosen at random and an attempt is made to swap the states (with the probability of a swap being determined by the Metropolis et al. equation). This allows the chain to potentially jump a valley in a single bound. The correct usage is

```
mcmc <parameter> = <value> ... <parameter> = <value>
```

For example,

```
mcmc ngen=100000 nchains=4 temp=0.5
```

performs a MCMCMC analysis with four chains with the temperature set to 0.5. The chains would be run for 100,000 cycles.

Options:

Seed -- Sets the seed number for the random number generator. The random number seed is initialized haphazardly at the beginning of each MrBayes session. This option allows you to set the seed to some specific value, thereby allowing you to exactly repeat an analysis. If the analysis uses swapping between cold and heated chains, you must also set the swap seed (see below) to exactly repeat the analysis.

Swapseed -- Sets the seed used for generating the swapping sequence when Metropolis-coupled heated chains are used. This seed is initialized haphazardly at the beginning of each MrBayes session. This option allows you to set the seed to some specific value, thereby allowing you to exactly repeat a swap sequence. See also the 'Seed' option.

Ngen -- This option sets the number of cycles for the MCMC algorithm. This should be a big number as you want the chain to first reach stationarity, and then remain there for enough time to take lots of samples.

Nruns -- How many independent analyses are started simultaneously.

Nchains -- How many chains are run for each analysis for the MCMCMC variant. The default is 4: 1 cold chain and 3 heated chains. If Nchains is set to 1, MrBayes will use regular MCMC sampling, without heating.

Temp -- The temperature parameter for heating the chains. The higher the temperature, the more likely the heated chains are to move between isolated peaks in the posterior distribution. However, excessive heating may lead to very low acceptance rates for swaps between different chains. Before changing the default setting, however, note that the acceptance rates of swaps tend to fluctuate during the burn-in phase of the run.

Reweight -- Here, you specify three numbers, that respectively represent the percentage of characters to decrease in weight, the percentage of characters to increase in weight, and the increment. An increase/decrease in weight is acheived by replicating/removing a character in the matrix. This is only done to non-cold chains. The format for this parameter is "reweight=(<number>,<number>)" or "reweight=(<number>,"<number>,<number>)".

Swapfreq -- This specifies how often swaps of states between chains are attempted. You must be running at least two chains for this

option to be relevant. The default is Swapfreq=1, resulting in Nswaps (see below) swaps being tried each generation of the run. If Swapfreq is set to 10, then Nswaps swaps will be tried every tenth generation of the run.

Nswaps -- The number of swaps tried for each swapping generation of the chain (see also Swapfreq).

Samplefreq -- This specifies how often the Markov chain is sampled. You can sample the chain every cycle, but this results in very large output files. Thinning the chain is a way of making these files smaller and making the samples more independent.

Printfreq -- This specifies how often information about the chain is printed to the screen.

Printall -- If set to NO, only cold chains in a MCMC analysis are printed to screen. If set to YES, both cold and heated chains will be output. This setting only affects the printing to screen, it does not change the way values are written to file.

Printmax -- The maximum number of chains to print to screen.

Mcmcdiagn -- Determines whether acceptance ratios of moves and swaps will be printed to file. The file will be named similarly to the '.p' and '.t' files, but will have the ending '.mcmc'. If more than one independent analysis is run simultaneously (see Nruns below), convergence diagnostics for tree topology will also be printed to this file. The convergence diagnostic used is the average standard deviation in partition frequency values across independent analyses. The Burnin setting (see below) determines how many samples will be discarded as burnin before calculating the partition frequencies. The Minpartfreq setting (see below) determines the minimum partition frequency required for a partition to be included in the calculation. As the independent analyses approach stationarity (converge), the value of the diagnostic is expected to approach zero.

Diagnfreq -- The number of generations between the calculation of MCMC diagnostics (see Mcmcdiagn above).

Minpartfreq -- The minimum frequency required for a partition to be included in the calculation of the topology convergence diagnostic. The partition is included if the minimum frequency is reached in at least one of the independent tree samples that are compared.

Allchains -- If this option is set to YES, acceptance ratios for moves are recorded for all chains, cold or heated. By default, only the acceptance ratios for the cold chain are recorded.

Allcomps -- If this option is set to YES, topological convergence diagnostics are calculated over all pairwise comparisons of runs. If it is set to NO, only the overall value is reported.

Relburnin -- If this option is set to YES, then a proportion of the sampled values will be discarded as burnin when calculating the convergence diagnostic. The proportion to be discarded is set with Burninfrac (see below). By default, the Relburnin option is set to NO, resulting in a specific number of samples being discarded instead. This number is set by Burnin (see below).

Burnin -- Determines the number of samples (not generations) that will

be discarded when convergence diagnostics are calculated. The value of this option is only relevant when Relburnin is set to NO.

BurninFrac -- Determines the fraction of samples that will be discarded when convergence diagnostics are calculated. The value of this option is only relevant when Relburnin is set to YES. Example: A value for this option of 0.25 means that 25
4525774f
the samples will be discarded.

Stoprule -- If this option is set to NO, then the chain is run the number of generations determined by Ngen. If it is set to YES, and topological convergence diagnostics are calculated (Mcmcdiagn is set to YES), then the chain will be stopped before the
pre-
determined number of generations if the convergence
diagnostic
falls below the stop value.

Stopval -- The critical value for the topological convergence
diagnostic.
Only used when Stoprule and Mcmcdiagn are set to yes, and more than one analysis is run simultaneously (Nruns > 1).

Filename -- The name of the files that will be generated. Two files are generated: "<Filename>.t" and "<Filename>.p". The .t file contains the trees whereas the .p file con-
tains the sampled values of the parameters.

Startingtree -- The starting tree for the chain can either be randomly selected or user-defined. It might be a good idea to start from randomly chosen trees; convergence seems likely if independently run chains, each of which started from different random trees, converge to the same answer.

Nperts -- This is the number of random perturbations to apply to the user starting tree. This allows you to have something between completely random and user-defined trees start the chain.

Savebrlens -- This specifies whether branch length information is saved on the trees.

Ordertaxa -- Determines whether taxa should be ordered before trees are printed to file. If set to 'Yes', terminals in the sampled trees will be reordered to match the order of the taxa in the data matrix as closely as possible. By default, trees will be printed without reordering of taxa.

Parameter	Options	Current Setting
Seed	<number>	1116355510
Swapseed	<number>	1116355510
Ngen	<number>	1000000
Nruns	<number>	2
Nchains	<number>	4
Temp	<number>	0.200000
Reweight	<number>,<number>	0.00 v 0.00 ^
Swapfreq	<number>	1
Nswaps	<number>	1
Samplefreq	<number>	100
Printfreq	<number>	100
Printall	Yes/No	Yes

Printmax	<number>	8
Mcmcdiagn	Yes/No	Yes
Diagnfreq	<number>	1000
Minpartfreq	<number>	0.10
Allchains	Yes/No	No
Allcomps	Yes/No	No
Relburnin	Yes/No	Yes
Burnin	<number>	0
Burninfrac	<number>	0.25
Stoprule	Yes/No	No
Stopval	<number>	0.01
Filename	<name>	temp.out.<p/t>
Startingtree	Random/User	Random
Nperts	<number>	0
Savebrlens	Yes/No	Yes
Ordertaxa	Yes/No	No

Mcmcp

This command sets the parameters of the Markov chain Monte Carlo (MCMC) analysis without actually starting the chain. This command is identical in all respects to Mcmc, except that the analysis will not start after this command is issued. For more details on the options, check the help menu for Mcmc.

Parameter	Options	Current Setting
Seed	<number>	1116355510
Swapseed	<number>	1116355510
Ngen	<number>	1000000
Nruns	<number>	2
Nchains	<number>	4
Temp	<number>	0.200000
Reweight	<number>,<number>	0.00 v 0.00 ^
Swapfreq	<number>	1
Nswaps	<number>	1
Samplefreq	<number>	100
Printfreq	<number>	100
Printall	Yes/No	Yes
Printmax	<number>	8
Mcmcdiagn	Yes/No	Yes
Diagnfreq	<number>	1000
Minpartfreq	<number>	0.10
Allchains	Yes/No	No
Allcomps	Yes/No	No
Relburnin	Yes/No	Yes
Burnin	<number>	0
Burninfrac	<number>	0.25
Stoprule	Yes/No	No
Stopval	<number>	0.01
Filename	<name>	temp.out.<p/t>
Startingtree	Random/User	Random
Nperts	<number>	0
Savebrlens	Yes/No	Yes
Ordertaxa	Yes/No	No

Outgroup

This command assigns a taxon to the outgroup. The correct usage is:

```
outgroup <number>/<taxon name>
```

For example, "outgroup 3" assigns the third taxon in the matrix to be the outgroup. Similarly, "outgroup Homo_sapiens" assings the taxon "Homo_sapiens" to be the outgroup (assuming that there is a taxon named "Homo_sapiens" in the matrix). Only a single taxon can be assigned to be the outgroup.

Pairs

This command is used to specify pairs of nucleotides. For example, your data may be RNA sequences with a known secondary structure of stems and loops. Substitutions in nucleotides involved in a Watson-Crick pairing in stems are not strictly independent; a change in one changes the probability of a change in the partner. A solution to this problem is to expand the model around the pair of nucleotides in the stem. This command allows you to do this. The correct usage is:

```
pairs <NUC1>:<NUC2>, <NUC1>:<NUC2>,..., <NUC1>:<NUC2>;
```

For example,

```
pairs 30:56, 31:55, 32:54, 33:53, 34:52, 35:51, 36:50;
```

specifies pairings between nucleotides 30 and 56, 31 and 55, etc. Only nucleotide data (DNA or RNA) may be paired using this command. Note that in order for the program to actually implement a "doublet" model involving a 16 X 16 rate matrix, you must specify that the structure of the model is 16 X 16 using "lset nucmodel=doublet".

Partition

This command allows you to specify a character partition. The format for this command is

```
partition <name> = <num parts>:<chars in first>, ...,<chars in last>
```

For example, "partition by_codon = 3:1st_pos,2nd_pos,3rd_pos" specifies a partition called "by_codon" which consists of three parts (first, second, and third codon positions). Here, we are assuming that the sites in each partition were defined using the charset command. You can specify a partition without using charset as follows:

```
partition by_codon = 3:1 4 6 9 12,2 5 7 10 13,3 6 8 11 14
```

However, we recommend that you use the charsets to define a set of characters and then use these predefined sets when defining the partition. Also, it makes more sense to define a partition as a line in the mrbayes

block than to issue the command from the command line (then again, you may be a masochist, and want to do extra work).

Plot

This command plots specified parameters in the .p file created by the program. The program prints two files during a MCMC analysis: a tree file and a parameter file. The parameter file has the extension ".p". This command, plot, makes an x-y graph of the parameter over the course of the chain. The command can be useful for visually diagnosing convergence for many of the parameters of the phylogenetic model. The parameter to be plotted is specified by the "parameter" option. Several parameters can be plotted at once by using the "match" option, which has a default value of "perfect". For example, if you were to set "parameter = pi" and "match = consistentwith", then all of the state frequency parameters would be plotted. You can also set "match=all", in which case all of the parameters are plotted.

Parameter	Options	Current Setting
Filename	<name>	temp.p
Burnin	<number>	0
Parameter	<name>	lnL
Match	Perfect/Consistentwith/All	Perfect

Prset

This command sets the priors for the phylogenetic model. Remember that in a Bayesian analysis, you must specify a prior probability distribution for the parameters of the likelihood model. The prior distribution represents your prior beliefs about the parameter before observation of the data. This command allows you to tailor your prior assumptions to a large extent.

Options:

Applyto -- This option allows you to apply the prset commands to specific partitions. This command should be the first in the list of commands specified in prset. Moreover, it only makes sense to be using this command if the data have been partitioned. A default partition is set on execution of a matrix. If the data are homogeneous (i.e., all of the same data type), then this partition will not subdivide the characters. Up to 30 other partitions can be defined, and you can switch among them using "set partition=<partition name>". Now, you may want to specify different priors to different partitions of the data. Applyto allows you to do this. For example, say you have partitioned the data by codon position, and you want to fix the statefreqs to equal for the first two partitions but apply a flat Dirichlet prior to the statefreqs of the last. This could be implemented in two uses of prset:

```
                    prset applyto=(1,2) statefreqs=fixed(equal)

                    prset applyto=(3) statefreqs=dirichlet(1,1,1,1)

                  The first applies the parameters after "applyto"
                  to the first and second partitions. The second prset
                  applies a flat Dirichlet to the third partition. You can
                  also use applyto=(all), which attempts to apply the para-
                  meter settings to all of the data partitions. Importantly,
                  if the option is not consistent with the data in the part-
                  ition, the program will not apply the prset option to
                  that partition.
   Tratiopr    -- This parameter sets the prior for the transition/trans-
                  version rate ratio (tratio). The options are:

                    prset tratiopr = beta(<number>, <number>)
                    prset tratiopr = fixed(<number>)

                  The program assumes that the transition and transversion
                  rates are independent gamma-distributed random variables
                  with the same scale parameter when beta is selected. If you
                  want a diffuse prior that puts equal emphasis on transition/
                  transversion rate ratios above 1.0 and below 1.0, then use a
                  flat Beta, beta(1,1), which is the default. If you wish to
                  concentrate this distribution more in the equal-rates
region,
                  then use a prior of the type beta(x,x), where the magnitude
                  of x determines how much the prior is concentrated in the
                  equal rates region. For instance, a beta(20,20) puts more
                  probability on rate ratios close to 1.0 than a beta(1,1). If
                  you think it is likely that the transition/transversion rate
                  ratio is 2.0, you can use a prior of the type beta(2x,x),
                  where x determines how strongly the prior is concentrated on
                  tratio values near 2.0. For instance, a beta(2,1) is much
                  more diffuse than a beta(80,40) but both have the expected
                  tratio 2.0 in the absence of data. The parameters of the
                  Beta can be interpreted as counts: if you have observed x
                  transitions and y transversions, then a beta(x+1,y+1) is a
                  good representation of this information. The fixed option
                  allows you to fix the tratio to a particular value.
   Revmatpr    -- This parameter sets the prior for the substitution rates
                  of the GTR model for nucleotide data. The options are:

                    prset revmatpr =
dirichlet(<number>,<number>,...,<number>)
                    prset revmatpr = fixed(<number>,<number>,...,<number>)

                  The program assumes that the six substitution rates
                  are independent gamma-distributed random variables with the
                  same scale parameter when dirichlet is selected. The six
                  numbers in brackets each corresponds to a particular substi-
                  tution type. Together, they determine the shape of the
prior.
                  The six rates are in the order A<->C, A<->G, A<->T, C<->G,
                  C<->T, and G<->T. If you want an uninformative prior you can
                  use dirichlet(1,1,1,1,1,1), also referred to as a 'flat'
                  Dirichlet. This is the default setting. If you wish a prior
```

where the C<->T rate is 5 times and the A<->G rate 2 times higher, on average, than the transversion rates, which are all the same, then you should use a prior of the form dirichlet(x,2x,x,x,5x,x), where x determines how much the prior is focused on these particular rates. For more info, see tratiopr. The fixed option allows you to fix the substitution rates to particular values.

Aamodelpr -- This parameter sets the rate matrix for amino acid data. You can either fix the model by specifying aamodelpr= fixed(<model name>), where <model name> is 'poisson' (a glorified Jukes-Cantor model), 'jones', 'dayhoff', 'mtrev', 'mtmam', 'wag', 'rtrev', 'cprev', 'vt', 'blosum', 'equalin' (a glorified Felsenstein 1981 model), or 'gtr'. You can also average over the first ten models by specifying aamodelpr= mixed. If you do so, the Markov chain will sample each model according to its probability. The sampled model is reported as an index: poisson(0), jones(1), dayhoff(2), mtrev(3), mtmam(4), wag(5), rtrev(6), cprev(7), vt(8), or blosum(9). The 'Sump' command summarizes the MCMC samples and calculates the posterior probability estimate for each of these models.

Aarevmatpr -- This parameter sets the prior for the substitution rates of the GTR model for amino acid data. The options are:

```
prset revmatpr =
dirichlet(<number>,<number>,...,<number>)
prset revmatpr = fixed(<number>,<number>,...,<number>)
```

The options are the same as those for 'Revmatpr' except that they are defined over the 190 rates of the time-reversible GTR model for amino acids instead of over the 6 rates of the GTR model for nucleotides. The rates are in the order A<->R, A<->N, etc to Y<->V. In other words, amino acids are listed in alphabetic order based on their full name. The first amino acid (Alanine) is then combined in turn with all amino acids following it in the list, starting with amino acid 2 (Arginine) and finishing with amino acid 20 (Valine). The second amino acid (Arginine) is then combined in turn with all amino acids following it, starting with amino acid 3 (Asparagine) and finishing with amino acid 20 (Valine), and so on.

Omegapr -- This parameter specifies the prior on the nonsynonymous/synonymous rate ratio. The options are:

```
prset omegapr = uniform(<number>,<number>)
prset omegapr = exponential(<number>)
prset omegapr = fixed(<number>)
```

This parameter is only in effect if the nucleotide substitution model is set to codon using the lset command (lset nucmodel=codon). Moreover, it only applies to the case when there is no variation in omega across sites (i.e., "lset omegavar=equal").

Ny98omega1pr -- This parameter specifies the prior on the nonsynonymous/synonymous rate ratio for sites under purifying selection. The options are:

```
                    prset Ny98omega1pr = beta(<number>,<number>)
                    prset Ny98omega1pr = fixed(<number>)

                  This parameter is only in effect if the nucleotide sub-
                  stitution model is set to codon using the lset command
                  (lset nucmodel=codon). Moreover, it only applies to the
                  case where omega varies across sites using the model of
                  Nielsen and Yang (1998) (i.e., "lset omegavar=ny98"). If
                  fixing the parameter, you must specify a number between
                  0 and 1.
Ny98omega3pr  -- This parameter specifies the prior on the nonsynonymous/
                  synonymous rate ratio for positively selected sites. The
                  options are:

                    prset Ny98omega3pr = uniform(<number>,<number>)
                    prset Ny98omega3pr = exponential(<number>)
                    prset Ny98omega3pr = fixed(<number>)

                  This parameter is only in effect if the nucleotide sub-
                  stitution model is set to codon using the lset command
                  (lset nucmodel=codon). Moreover, it only applies to the
                  case where omega varies across sites according to the
                  NY98 model. Note that if the NY98 model is specified
                  that this parameter must be greater than 1, so you should
                  not specify a uniform(0,10) prior, for example.
M3omegapr     -- This parameter specifies the prior on the nonsynonymous/
                  synonymous rate ratios for all three classes of sites for
                  the M3 model. The options are:

                    prset M3omegapr = exponential
                    prset M3omegapr = fixed(<number>,<number>,<number>)

                  This parameter is only in effect if the nucleotide sub-
                  stitution model is set to codon using the lset command
                  (lset nucmodel=codon). Moreover, it only applies to the
                  case where omega varies across sites using the M3 model of
                  Yang et al. (2000) (i.e., "lset omegavar=M3"). Under the
                  exponential prior, the four rates (dN1, dN2, dN3, and dS)
                  are all considered to be independent draws from the same
                  exponential distribution (the parameter of the exponential
                  does not matter, and so you don't need to specify it). The
                  rates dN1, dN2, and dN3 are taken to be the order statistics
                  with dN1 < dN2 < dN3. These three rates are all scaled to
                  the same synonymous rate, dS. The other option is to simply
                  fix the three rate ratios to some values.
Codoncatfreqs -- This parameter specifies the prior on frequencies of sites
                  under purifying, neutral, and positive selection. The
                  options are:

                    prset codoncatfreqs = dirichlet(<num>,<num>,<num>)
                    prset codoncatfreqs = fixed(<number>,<number>,<number>)

                  This parameter is only in effect if the nucleotide sub-
                  stitution model is set to codon using the lset command
                  (lset nucmodel=codon). Moreover, it only applies to the
                  case where omega varies across sites using the models of
```

```
                  Nielsen and Yang (1998) (i.e., "lset omegavar=ny98")
                  or Yang et al. (2000) (i.e., "lset omegavar=M3")
                  Note that the sum of the three frequencies must be 1.
Statefreqpr    -- This parameter specifies the prior on the state freq-
                  uencies. The options are:

                     prset statefreqpr = dirichlet(<number>)
                     prset statefreqpr = dirichlet(<number>,...,<number>)
                     prset statefreqpr = fixed(equal)
                     prset statefreqpr = fixed(empirical)
                     prset statefreqpr = fixed(<number>,...,<number>)

                  For the dirichlet, you can specify either a single number
                  or as many numbers as there are states. If you specify a
                  single number, then the prior has all states equally
                  probable with a variance related to the single parameter
                  passed in.
Shapepr        -- This parameter specifies the prior for the gamma shape
                  parameter for among-site rate variation. The options are:

                     prset shapepr = uniform(<number>,<number>)
                     prset shapepr = exponential(<number>)
                     prset shapepr = fixed(<number>)

Pinvarpr       -- This parameter specifies the prior for the proportion of
                  invariable sites. The options are:

                     prset pinvarpr = uniform(<number>,<number>)
                     prset pinvarpr = fixed(<number>)

                  Note that the valid range for the parameter is between 0
                  and 1. Hence, "prset pinvarpr=uniform(0,0.8)" is valid
                  while "prset pinvarpr=uniform(0,10)" is not. The def-
                  ault setting is "prset pinvarpr=uniform(0,1)".
Ratecorrpr     -- This parameter specifies the prior for the autocorrelation
                  parameter of the autocorrelated gamma distribution for
                  among-site rate variation. The options are:

                     prset ratecorrpr = uniform(<number>,<number>)
                     prset ratecorrpr = fixed(<number>)

                  Note that the valid range for the parameter is between -1
                  and 1. Hence, "prset ratecorrpr=uniform(-1,1)" is valid
                  while "prset ratecorrpr=uniform(-11,10)" is not. The
                  default setting is "prset ratecorrpr=uniform(-1,1)".
Covswitchpr    -- This option sets the prior for the covarion switching
                  rates. The options are:

                     prset covswitchpr = uniform(<number>,<number>)
                     prset covswitchpr = exponential(<number>)
                     prset covswitchpr = fixed(<number>,<number>)

                  The covarion model has two rates: a rate from on to off
                  and a rate from off to on. The rates are assumed to have
                  independent priors that individually are either uniformly
                  or exponentially distributed. The other option is to
                  fix the switching rates, in which case you must specify
```

both rates. (The first number is off->on and the second is on->off).

Symdirihyperpr - This option sets the prior for the stationary frequencies of the states for morphological (standard) data. There can be as many as 10 states for standard data. However, the labelling of the states is somewhat arbitrary. For example, the state "1" for different characters does not have the same meaning. This is not true for DNA characters, for example, where a "G" has the same meaning across characters. The fact that the labelling of morphological characters is arbitrary makes it difficult to allow unequal character-state frequencies. MrBayes gets around this problem by assuming that the states have a dirichlet prior, with all states having equal frequency. The variation in the dirichlet can be controlled by this parameter--symdirihyperpr. Symdirihyperpr specifies the distribution on the variance parameter of the dirichlet. The valid options are:

```
prset Symdirihyperpr = uniform(<number>,<number>)
prset Symdirihyperpr = exponential(<number>)
prset Symdirihyperpr = fixed(<number>)
prset Symdirihyperpr = fixed(infinity)
```

If "fixed(infinity)" is chosen, the dirichlet prior is fixed such that all character states have equal frequency.

Topologypr -- This parameter specifies the prior probabilities of phylogenies. The options are:

```
prset topologypr = uniform
prset topologypr = constraints(<list>)
```

If the prior is selected to be "uniform", the default, then all possible trees are considered a priori equally probable. The constraints option allows you to specify complicated prior probabilities on trees (constraints are discussed more fully in "help constraint"). Note that you must specify a list of constraints that you wish to be obeyed. The list can be either the constraints' name or number. Also, note that the constraints simply tell you how much more (or less) probable individual trees are that possess the constraint than trees not possessing the constraint.

Brlenspr -- This parameter specifies the prior probability distribution on branch lengths. The options are:

```
prset brlenspr = unconstrained:uniform(<num>,<num>)
prset brlenspr = unconstrained:exponential(<number>)
prset brlenspr = clock:uniform
prset brlenspr = clock:birthdeath
prset brlenspr = clock:coalescence
```

Trees with unconstrained branch lengths are unrooted whereas clock-constrained trees are rooted. The option after the colon specifies the details of the probability density of branch lengths. If you choose a birth-death or coalescence prior, you may want to modify the details of the parameters of those processes.

Treeheightpr -- This parameter specifies the prior probability distribution on the tree height, when a clock model is specified. The options are:

```
prset treeheightpr = Gamma(<num>,<num>)
prset treeheightpr = Exponential(<number>)
```

(And, yes, we know the exponential is a special case of the gamma distribution.) The tree height is the expected number of substitutions on a single branch that extends from the root of the tree to the tips. This parameter does not come into play for the coalescence prior. It insures that the prior probability distribution for unconstrained and birth-death models is proper.

Ratepr -- This parameter allows you to specify the site specific rates model. First, you must have defined a partition of the characters. For example, you may define a partition that divides the characters by codon position, if you have DNA data. Second, you must make that partition the active one using the set command. For example, if your partition is called "by_codon", then you make that the active partition using "set partition=by_codon". Now that you have defined and activated a partition, you can specify the rate multipliers for the various partitions. The options are:

```
prset ratepr = fixed
prset ratepr = variable
prset ratepr = dirichlet(<number>,<number>,...,<number>)
```

If you specify "fixed", then the rate multiplier for that partition is set to 1 (i.e., the rate is fixed to the average rate across partitions). On the other hand, if you specify "variable", then the rate is allowed to vary across partitions subject to the constraint that the average rate of substitution across the partitions is 1. You must specify a variable rate prior for at least two partitions, otherwise the option is not activated when calculating likelihoods. The variable option automatically associates the partition rates with a dirichlet(1,...,1) prior. The dirichlet option is an alternative way of setting a partition rate to be variable, and also gives accurate control of the shape of the prior. The parameters of the Dirichlet are listed in the order of the partitions that the ratepr is applied to. For instance, "prset applyto=(1,3,4) ratepr = dirichlet(10,40,15)" would set the Dirichlet parameter 10 to partition 1, 40 to partition 3, and 15 to partition 4.

Speciationpr -- This parameter sets the prior on the speciation rate. The options are:

```
prset speciationpr = uniform(<number>,<number>)
prset speciationpr = exponential(<number>)
prset speciationpr = fixed(<number>)
```

This parameter is only relevant if the birth-death process is selected as the prior on branch lengths.

```
   Extinctionpr  -- This parameter sets the prior on the extinction rate. The
                    options are:

                       prset extinctionpr = uniform(<number>,<number>)
                       prset extinctionpr = exponential(<number>)
                       prset extinctionpr = fixed(<number>)

                    This parameter is only relevant if the birth-death
                    process is selected as the prior on branch lengths.
   Sampleprob    -- This parameter sets the fraction of species that are
                    sampled in the analysis. This is used with the birth-
                    death prior on trees (see Yang and Rannala, 1997).
   Thetapr       -- This parameter sets the prior on the coalescence para-
                    meter. The options are:

                       prset thetapr = uniform(<number>,<number>)
                       prset thetapr = exponential(<number>)
                       prset thetapr = fixed(<number>)

                    This parameter is only relevant if the coalescence
                    process is selected as the prior on branch lengths.

   Default model settings:

   Parameter        Options                     Current Setting
   ------------------------------------------------------------------
   Tratiopr         Beta/Fixed                  Beta(1.0,1.0)
   Revmatpr         Dirichlet/Fixed
Dirichlet(1.0,1.0,1.0,1.0,1.0,1.0)
   Aamodelpr        Fixed/Mixed                 Fixed(Poisson)
   Aarevmatpr       Dirichlet/Fixed             Dirichlet(1.0,1.0,...)
   Omegapr          Dirichlet/Fixed             Dirichlet(1.0,1.0)
   Ny98omega1pr     Beta/Fixed                  Beta(1.0,1.0)
   Ny98omega3pr     Uniform/Exponential/Fixed   Exponential(1.0)
   M3omegapr        Exponential/Fixed           Exponential
   Codoncatfreqs    Dirichlet/Fixed             Dirichlet(1.0,1.0,1.0)
   Statefreqpr      Dirichlet/Fixed             Dirichlet
   Treeheightpr     Exponential/Gamma           Exponential(1.0)
   Ratepr           Fixed/Variable=Dirichlet    Fixed
   Shapepr          Uniform/Exponential/Fixed   Uniform(0.0,50.0)
   Ratecorrpr       Uniform/Fixed               Uniform(-1.0,1.0)
   Pinvarpr         Uniform/Fixed               Uniform(0.0,1.0)
   Covswitchpr      Uniform/Exponential/Fixed   Uniform(0.0,100.0)
   Symdirihyperpr   Uniform/Exponential/Fixed   Fixed(Infinity)
   Topologypr       Uniform/Constraints         Uniform
   Brlenspr         Unconstrained/Clock         Unconstrained:Exp(10.0)
   Speciationpr     Uniform/Exponential/Fixed   Uniform(0.0,10.0)
   Extinctionpr     Uniform/Exponential/Fixed   Uniform(0.0,10.0)
   Sampleprob       <number>                    1.00
   Thetapr          Uniform/Exponential/Fixed   Uniform(0.0,10.0)
   ------------------------------------------------------------------

   ------------------------------------------------------------------------
   Props

   This command allows the user to change the details of the MCMC mechanism
   that updates the state of the chain. The useage is:
```

```
   props
```

On typing "props", you will get a list of parameters to change. The program works as follows: On typing "mcmc", MrBayes figures out which model parameters need to be updated. For example, if you include a transition/transversion rate parameter, then the program needs to update this parameter along with others, such as the tree and branch lengths. Once MrBayes figures out which moves are needed, it figures out the probability of making each move on every cycle of the chain. MrBayes updates parameters in blocks; it decides which parameter to update, changes the parameter, and then accepts or rejects the move according to the Metropolis-Hastings equation. The probability of making a move is calculated as the proposal rate for the move divided by the sum of the proposal rates for all of the other parameters that need to be updated. This command also allows you to change the details of each proposal mechanism. Many of the moves change parameters using sliding windows centered on the current value of the parameter. If you increase or decrease the window size, you will respectively decrease or increase the acceptance rate of the move. Some of the other moves update using a dirichlet or beta distribution, centered on the current values. You can change the variance parameter of the dirichlet or beta distribution. Finally, a few of the topology moves have a tuning parameter which influences the degree to which branch lengths are modified. If you increase this tuning parameter, you will make more radical changes to the branch lengths.

One word of warning: You should be extremely careful when modifying any of the chain parameters using "props". It is quite possible to completely wreck any hope of achieving convergence by inappropriately setting the chain parameters. Please exercise this command with caution.

Quit

This command quits the program. The correct usage is:

```
   quit
```

It is a very easy command to use properly.

Report

This command allows you to control how the posterior distribution is reported. For rate parameters, it allows you to choose among several popular parameterizations. The report command also allows you to request printing of some model aspects that are usually not reported. For instance, if a node is constrained in the analysis, MrBayes can print the probabilities of the ancestral states at that node. Similarly, if there is rate variation in the model, MrBayes can print the inferred site rates, and if there is omega variation, MrBayes can print the inferred omega (positive selection) values for each codon. In a complex model with several partitions, each partition is controlled separately using the same 'Applyto' mechanism as in the 'Lset' and 'Prset' commands.

Options:

Applyto -- This option allows you to apply the report commands to specific partitions. This command should be the first in the list of commands specified in 'report'.
For example,

```
report applyto=(1,2) tratio=ratio

report applyto=(3) tratio=dirichlet
```

would result in the transition and transversion rates of the first and second partitions in the model being reported as a ratio and the transition and transversion rates of the third partition being reported as proportions of the rate sum (the Dirichlet parameterization).

Tratio -- This specifies the report format for the transition and transversion rates of a nucleotide substituion model with nst=2. If 'ratio' is selected, the rates will be reported as a ratio (transition rate/transversion rate). If 'dirichlet' is selected, the transition and transversion rates will instead be reported as proportions of the rate sum. For example, if the transition rate is three times the transversion rate and 'ratio' is selected, this will reported as a single value, '3.0'. If 'dirichlet' is selected instead, the same rates will be reported using two values, '0.75 0.25'. The sum of the Dirichlet values is always 1.

Although the Dirichlet format may be unfamiliar to some users, it is more convenient for specifying priors than the ratio format.

Revmat -- This specifies the report format for the substitution rates of a GTR substitution model for nucleotide or amino acid data. If 'ratio' is selected, the rates will be reported scaled to the G-T rate (for nucleotides) or the Y-V rate (for amino acids). If 'dirichlet' is specified instead, the rates are reported as proportions of the rate sum. For instance, assume that the C-T rate is twice the A-G rate and four times the transversion rates, which are equal. If the report format is set to 'ratio', this would be reported as '1.0 2.0 1.0 1.0 4.0 1.0' since the rates are reported in the order rAC, rAG, rAT, rCG, rCT, rGT and scaled relative to the last rate, the G-T rate. If 'dirichlet' is selected instead, the same rates would have been reported as '0.1 0.2 0.1 0.1 0.4 0.1' since the rates are now scaled so that they sum to 1.0. The Dirichlet format is the parameterization used for formulating priors on the rates.

Ratemult -- This specifies the report format used for the rate multiplier of different model partitions. Three formats are available. If 'scaled' is selected, then rates are scaled such that the mean rate per site across partitions is 1.0. If 'ratio' is chosen, the rates are scaled relative to the rate of the first partition. Finally, if 'dirichlet' is chosen, the rates are given as proportions of the rate sum. The latter is the format used when formulating priors on the rate multiplier.

Ancstates -- If this option is set to 'yes', MrBayes will print the probability of the ancestral states at all constrained nodes. Typically, you are interested in the ancestral states of only a few

```
                characters and only at one node in the tree. To perform such
                an analysis, first define and enforce a topology constraint
                using 'constraint' and 'prset topologypr = constraints (...)'.
                Then put the character(s) of interest in a separate partition
and
                set MrBayes to report the ancestral states for that partition.
                For instance, if the characters of interest are in partition 2,
                use 'report applyto=(2) ancstates=yes' to force MrBayes to print
                the probability of the ancestral states of those characters at
                the constrained node to the '.p' file.
   Siterates -- If this option is set to 'yes' and the relevant model has rate
                variation across sites, the mean site rate in the posterior will
                be reported for each site to the '.p' file.
   Possel    -- If this option is set to 'yes' and the relevant model has omega
                variation across sites, the mean omega value for each model site
                (codon in this case) will be written to the '.p' file.

   Current settings:

   Parameter       Options                 Current Setting
   -------------------------------------------------------
   Tratio          Ratio/Dir               Ratio
   Revmat          Ratio/Dir               Dirichlet
   Ratemult        Scaled/Ratio/Dir        Scaled
   Ancstates       Yes/No                  No
   Siterates       Yes/No                  No
   Possel          Yes/No                  No

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

   ------------------------------------------------------------------------
   Restore

   This command restores taxa to the analysis. The correct usage is:

      restore <name and/or number and/or taxset> ...

   A list of the taxon names or taxon numbers (labelled 1 to ntax in the order
   in the matrix) or taxset(s) can be used.  For example, the following:

      restore 1 2 Homo_sapiens

   restores taxa 1, 2, and the taxon labelled Homo_sapiens to the analysis.
   You can also use "all" to restore all of the taxa. For example,

      restore all

   restores all of the taxa to the analysis.
   ------------------------------------------------------------------------
   ------------------------------------------------------------------------
   Root

   This command roots the tree. If the tree is already rooted, a warning
   is issued. The tree is rooted at the outgroup species. the correct
   usage is "root".
   ------------------------------------------------------------------------
   ------------------------------------------------------------------------
```

Set

This command is used to set some general features of the model or program behavior. The correct usage is

```
set <parameter>=<value> ... <parameter>=<value>
```

Only four parameters can be changed using "set". First, you can set the autoclose feature:

```
set autoclose=<yes/no>
```

If autoclose is set to yes, then the program will not prompt you during the course of executing a file. Second, you can set the partition that is in effect:

```
set partition=<partition id>
```

A valid partition ID is either a number or a partition name. This command enforces use of a specific partitioning of the data. When the program executes, a default partition (that may not divide the data at all) is created called "Default". You can always go back to the original or default partition by typing

```
set partition=default
```

or

```
set partition=1
```

Third, you can set the nowarnings feature:

```
set nowarnings=<yes/no>
```

If nowarnings is set to yes, then the program will not prompt you when overwriting an output file that is already present. If nowarnings=no (the default setting) then the program prompts the user before over-writing output files.

Fourth, you can set the quitonerror feature:

```
set quitonerror=<yes/no>
```

If quitonerror is set to yes, then the program will quit when an error is encountered, after printing an error message. If quitonerror=no (the default setting) then the program will wait for additional commands from the command line after printing the error message.

Showmatrix

This command shows the character matrix currently in memory.

Showmodel

This command shows the current model settings. The correct usage is

showmodel

After typing "showmodel", the modelling assumptions are shown on a partition-by-partition basis.

Showtree

This command shows the current user tree. The correct usage is "showtree".

Sump

During a MCMC analysis, MrBayes prints the sampled parameter values to a tab-delimited text file. This file has the extension ".p". The command 'Sump' summarizes the information in the parameter file. By default, the name of the parameter file is assumed to be the name of the last matrix-containing nexus file, but with a '.p' extension. You can set 'Sump' to summarize the information in any other parameter file by setting the 'filename' option to the appropriate file name. The 'Sump' command does not require a matrix to be read in first. When you invoke the 'Sump' command, three items are output: (1) a generation plot of the likelihood values; (2) estimates of the marginal likelihood of the model; and (3) a table with the mean, variance, and 95 percent credible interval for the sampled parameters. Each of these items can be switched on or off using the options 'Plot', 'Marglike', and 'Table'. By default, all three items are output but only to the screen. If output to a file is also desired, set 'Printtofile' to 'Yes'. The name of the output file is specified by setting the 'Outputname' option. When a new matrix is read in or when the 'Mcmc' output filename or 'Sump' input filename is changed, the 'Sump' outputname is changed as well. If you want to output to another file than the default, make sure you specify the outputname every time you invoke 'Sump'. If the specified outputfile already exists, you will be prompted about whether you like to overwrite it or append to it. This behavior can be altered using 'Set nowarn=yes'; see the help for the 'Set' command. When running 'Sump' you typically want to discard a specified number of samples from the beginning of the chain as the burn in. Note that the 'Burnin' value of the 'Sump' command is set separately from the 'Burnin' values of the 'Sumt' and 'Mcmc' commands. That is, if you issue

```
sump burnin = 4000
sumt burnin = 2000
sump
```

the burnin of the last 'Sump' command is 4000 and not 2000. The burnin values are reset to 0 every time a new matrix is read in. Similarly, 'Plot', 'Marglike' and 'Table' are all set to 'Yes' and 'Printtofile' to 'No' (the default values) when a new matrix is processed. If you have run several independent MCMC analyses, you may want to summarize and compare the samples from each of these runs. To do this, set 'Nruns' to the number of runs you want to compare and make sure that the '.p' files are named using the MrBayes convention (<filename>.run1.p, <filename>.run2.p, etc). When you run several independent analyses simultaneously in MrBayes, the 'Nruns' and 'Filename' options are automatically set such that 'Sump' will summarize all the resulting output files.

```
   Options:

   Burnin       -- Determines the number of samples that will be discarded from
                   the input file before calculating summary statistics. If
there
                   are several input files, the same number of samples will be
                   discarded from each. Note that the burnin is set separately
                   for the 'sump', 'sumt', and 'mcmc' commands.
   Nruns        -- Determines how many '.p' files from independent analyses that
                   will be summarized. If Nruns > 1 then the names of the files
                   are derived from 'Filename' by adding '.run1.p', '.run2.p',
                   etc. If Nruns=1, then the single file name is obtained by
                   adding '.p' to 'Filename'.
   Filename     -- The name of the file to be summarized. This is the base of
the
                   file name to which endings are added according to the current
                   setting of the 'Nruns' parameter. If 'Nruns' is 1, then only
                   '.p' is added to the file name. Otherwise, the endings will
                   be '.run1.p', '.run2.p', etc.
   Printtofile  -- Determines whether results will be printed to file.
   Outputname   -- Name of the file to which 'sump' results will be printed if
                   'Printtofile' is set to YES.
   Plot         -- Determines whether a likelihood plot should be output.
   Marglike     -- Determines whether estimates of marginal model likelihoods
                   should be calculated. The marginal model likelihoods are use-
                   ful in Bayesian model testing.
   Table        -- Determines whether the table summarizing the parameter value
                   samples should be output.

   Current settings:

   Parameter       Options                 Current Setting
   -------------------------------------------------------
   Burnin          <number>                0
   Nruns           <number>                1
   Filename        <name>                  temp.p<.p>
   Printtofile     Yes/No                  No
   Outputname      <name>                  temp.p.stat
   Plot            Yes/No                  Yes
   Marglike        Yes/No                  Yes
   Table           Yes/No                  Yes

   ------------------------------------------------------------------------
   ------------------------------------------------------------------------
   Sumt

   This command is used to produce summary statistics for trees sampled during
   a Bayesian MCMC analysis. You can either summarize trees from one individual
   analysis, or trees coming from several independent analyses. In either case,
   all the sampled trees are read in and the proportion of the time any single
   taxon bipartition is found is counted. The proportion of the time that the
   bipartition is found is an approximation of the posterior probability of the
   bipartition. (Remember that a taxon bipartition is defined by removing a
   branch on the tree, dividing the tree into those taxa to the left and right
   of the removed branch. This set is called a taxon bipartition.) The branch
   length of the bipartition is also recorded, if branch lengths have been saved
   to file. The result is a list of the taxon bipartitions found, the frequency
```

with which they were found, the posterior probability of the bipartition and, if the branch lengths were recorded, the mean and variance of the length of the branch.

The partition information is output to a file with the suffix ".parts". A consensus tree is also printed to a file with the suffix ".con" and printed to the screen as a cladogram, and as a phylogram if branch lengths have been saved. The consensus tree is either a 50 percent majority rule tree or a majority rule tree showing all compatible partitions. If branch lengths have been recorded during the run, the ".con" file will contain a consensus tree with branch lengths and interior nodes labelled with support values. This tree can be viewed in a program such as TreeView.

Finally, MrBayes produces a file with the ending ".trprobs" that contains a list of all the trees that were found during the MCMC analysis, sorted by their probabilities. This list of trees can be used to construct a credible set of trees. For example, if you want to construct a 95 percent credible set of trees, you include all of those trees whose cumulated probability is less than or equal to 0.95. You have the option of displaying the trees to the screen using the "Showtreeprobs" option. The default is to not display the trees to the screen; the number of different trees sampled by the chain can be quite large. If you are analyzing a large set of taxa, you may actually want to skip the calculation of tree probabilities entirely by setting "Calctreeprobs" to NO.

When calculating summary statistics you probably want to skip those trees that were sampled in the initial part of the run, the so-called burn-in period. The number of skipped samples is controlled by the "burnin" setting. The default is 0 but you typically want to override this setting.

If you are summarizing the trees sampled in several independent analyses, such as those resulting from setting the "Nruns" option of the "Mcmc" command to a value larger than 1, MrBayes will also calculate convergence diagnostics for the sampled topologies and branch lengths. These values can help you determine whether it is likely that your chains have converged.

The "Sumt" command expands the "Filename" according to the current values of the "Nruns" and "Ntrees" options. For instance, if both "Nruns" and "Ntrees" are set to 1, "Sumt" will try to open a file named "<Filename>.t". If "Nruns" is set to 2 and "Ntrees" to 1, then "Sumt" will open two files, "<Filename>.run1.t" and "<Filename>.run2.t", etc. By default, the "Filename" option will be set such that "Sumt" automatically summarizes all the results from your immediately preceding "Mcmc" command. You can also use the "Sumt" command to summarize tree samples in older analyses. If you want to do that, remember to first read in a matrix so that MrBayes knows what taxon names to expect in the trees. Then set the "Nruns", "Ntrees" and "Filename" options appropriately.

Options:

Burnin -- Determines the number of samples that will be discarded from the input file before calculating summary statistics. If there are several input files, the same number of samples will be discarded from each. Note that the burnin is set separately for the 'sumt', 'sump', and 'mcmc' commands.

```
   Nruns         -- Determines how many '.t' files from independent analyses
that
                    will be summarized. If Nruns > 1 then the names of the files
                    are derived from 'Filename' by adding '.run1.t', '.run2.t',
                    etc. If Nruns=1 and Ntrees=1 (see below), then only '.t' is
                    added to 'Filename'.
   Ntrees        -- Determines how many trees there are in the sampled model. If
                    'Ntrees' > 1 then the names of the files are derived from
                    'Filename' by adding '.tree1.t', '.tree2.t', etc. If there
                    are both multiple trees and multiple runs, the filenames
will
                    be '<Filename>.tree1.run1.t', '<Filename>.tree1.run2.t',
etc.
   Filename      -- The name of the file(s) to be summarized. This is the base
of
                    the file name, to which endings are added according to the
                    current settings of the 'Nruns' and 'Ntrees' options.
   Displaygeq    -- The minimum probability of partitions to display.
   Contype       -- Type of consensus tree. 'Halfcompat' results in a 50 major-
                    ity rule tree, 'Allcompat' adds all compatible groups to
such
                    a tree.
   Calctreeprobs -- Determines whether tree probabilities should be calculated.
   Showtreeprobs -- Determines whether tree probabilities should be displayed on
                    screen.

   Current settings:

   Parameter       Options                 Current Setting
   -------------------------------------------------------
   Burnin          <number>                0
   Nruns           <number>                1
   Ntrees          <number>                1
   Filename        <name>                  temp.t<.t>
   Displaygeq      <number>                0.05
   Contype         Halfcompat/Allcompat    Halfcompat
   Calctreeprobs   Yes/No                  Yes
   Showtreeprobs   Yes/No                  No

   ------------------------------------------------------------------------
   ------------------------------------------------------------------------
   Taxastat

   This command shows the status of all the taxa. The correct usage is

      taxastat

   After typing "taxastat", the taxon number, name, and whether it is
   excluded or included are shown.
   ------------------------------------------------------------------------
   ------------------------------------------------------------------------
   Taxset

   This command defines a taxon set. The format for the taxset command
   is

      taxset <name> = <taxon names or numbers>
```

For example, "taxset apes = Homo Pan Gorilla Orang gibbon" defines a taxon set called "apes" that includes five taxa (namely, apes). You can assign up to 30 taxon sets. This option is best used not from the command line but rather as a line in the mrbayes block of a file.

Unlink

This command unlinks model parameters across partitions of the data. The correct usage is:

```
unlink <parameter name> = (<all> or <partition list>)
```

A little background is necessary to understand this command. Upon execution of a file, a default partition is set up. This partition referenced either by its name ("default") or number (0). If your data are all of one type, then this default partition does not actually divide up your characters. However, if your datatype is mixed, then the default partition contains as many divisions as there are datatypes in your character matrix. Of course, you can also define other partitions, and switch among them using the set command ("set partition=<name/number>"). Importantly, you can also assign model parameters to individual partitions or to groups of them using the "applyto" option in lset and prset. When the program attempts to perform an analysis, the model is set for individual partitions. If the same parameter applies to different partitions and if that parameter has the same prior, then the program will link the parameters: that is, it will use a single value for the parameter. The program's default, then, is to strive for parsimony. However, there are lots of cases where you may want unlink a parameter across partitions. For example, you may want a different transition/transversion rate ratio to apply to different partitions. This command allows you to unlink the parameters, or to make them different across partitions. The converse of this command is "link", which links together parameters that were previously told to be different. The list of parameters that can be unlinked includes:

```
Tratio          -- Transition/transversion rate ratio
Revmat          -- Substitution rates of GTR model
Omega           -- Nonsynonymous/synonymous rate ratio
Statefreq       -- Character state frequencies
Shape           -- Gamma shape parameter
Pinvar          -- Proportion of invariable sites
Correlation     -- Correlation parameter of autodiscrete gamma
Switchrates     -- Switching rates for covarion model
Brlens          -- Branch lengths of tree
Topology        -- Topology of tree
Speciationrates -- Speciation rates for birth-death process
Extinctionrates -- Extinction rates for birth-death process
Theta           -- Parameter for coalescence process
Growthrate      -- Growth rate of coalescence process
```

For example,

```
unlink shape=(all)
```

unlinks the gamma shape parameter across all partitions of the data.
You can use "showmodel" to see the current linking status of the
characters.

Usertree

This command allows you to specify a user tree. The user tree can then be
used as a starting tree for a MCMC analysis. The format for the command is

```
   usertree = <tree in Newick format>
```

For example,

```
   usertree = (A,B,(C,D))
```

specifies an unrooted tree of four species. Note that the program re-
quires that trees are binary (i.e., strictly bifurcating). Hence, there
can be only one three-way split, as shown in the example. If the tree
is not binary, the program will return an error.

Version

This command shows the release version of the program.

```
**************************************************************************
*                                                                        *
*  3. 'Data' or 'tree' block commands (in #NEXUS file)                   *
*                                                                        *
**************************************************************************
```

Begin

This command is used to format data or commands in the program. The correct
usage is

```
   begin <data or mrbayes>;
```

The two valid uses of the "begin" command, then, are

```
   begin data;
   begin mrbayes;
```

The "data" specifier is used to specify the beginning of a data block; your
character data should follow. For example, the following is an example of
a data block for four taxa and ten DNA sites:

```
   begin data;
      dimensions ntax=4 nchar=10;
      format datatype=dna;
      matrix
      taxon_1  AACGATTCGT
      taxon_2  AAGGATTCCA
      taxon_3  AACGACTCCT
```

```
      taxon_4  AAGGATTCCT
      ;
   end;
```

The other commands -- dimensions, format, and matrix -- are discussed in the appropriate help menu. The only thing to note here is that the block begins with a "begin data" command. The "mrbayes" command is used to enter commands specific to the MrBayes program into the file. This allows you to automatically process commands on execution of the program. The following is a simple mrbayes block:

```
   begin mrbayes;
      charset first  = 1-10\3;
      charset second = 2-10\3;
      charset third  = 3-10\3;
   end;
```

This mrbayes block sets off the three "charset" commands, used to predefine some blocks of characters. The mrbayes block can be very useful. For example, in this case, it would save you the time of typing the character sets each time you executed the file. Also, note that every "begin <data or mrbayes>" command ends with an "end". Finally, you can have so-called foreign blocks in the file. An example of a foreign block would be "begin paup". The program will simply skip this block. This is useful because it means that you can use the same file for MrBayes, PAUP* or MacClade (although it isn't clear why you would want to use those other programs).

Dimensions

This command is used in a data block to define the number of taxa and characters. The correct usage is

```
   dimensions ntax=<number> nchar=<number>
```

The dimensions must be the first command in a data block. The following provides an example of the proper use of this command:

```
   begin data;
      dimensions ntax=4 nchar=10;
      format datatype=dna;
      matrix
      taxon_1  AACGATTCGT
      taxon_2  AAGGATTCCA
      taxon_3  AACGACTCCT
      taxon_4  AAGGATTCCT
      ;
   end;
```

Here, the dimensions command tells MrBayes to expect a matrix with four taxa and 10 characters.

End

This command is used to terminate a data or mrbayes block. The correct

usage is

```
   end;
```

For more information on this, check the help for the "begin" command.

Endblock

This is an older, deprecated version of "End", see that command.

Format

This command is used in a data block to define the format of the character matrix. The correct usage is

```
   format datatype=<name> ... <parameter>=<option>
```

The format command must be the second command in a data block. The following provides an example of the proper use of this command:

```
   begin data;
      dimensions ntax=4 nchar=10;
      format datatype=dna gap=-;
      matrix
      taxon_1  AACGATTCGT
      taxon_2  AAGGAT--CA
      taxon_3  AACGACTCCT
      taxon_4  AAGGATTCCT
      ;
   end;
```

Here, the format command tells MrBayes to expect a matrix with DNA characters and with gaps coded as "-".

The following are valid options for format:

Datatype -- This parameter MUST BE INCLUDED in the format command. Moreover, it must be the first parameter in the line. The datatype command specifies what type of characters are in the matrix. The following are valid options:

- Datatype = Dna: DNA states (A,C,G,T,R,Y,M,K,S,W,H,B, V,D,N)
- Datatype = Rna: DNA states (A,C,G,U,R,Y,M,K,S,W,H,B, V,D,N)
- Datatype = Protein: Amino acid states (A,R,N,D,C,Q,E, G,H,I,L,K,M,F,P,S,T,W,Y,V)
- Datatype = Restriction: Restriction site (0,1) states
- Datatype = Standard: Morphological (0,1) states
- Datatype = Continuous: Real number valued states
- Datatype = Mixed(<type>:<range>,...,<type>:<range>): A mixture of the above datatypes. For example, "datatype=mixed(dna:1-100,protein:101-200)" would specify a mixture of DNA and amino acid characters with the DNA characters occupying the first 100 sites and the amino acid char-

acters occupying the last 100 sites.

Interleave -- This parameter specifies whether the data matrix is in interleave format. The valid options are "Yes" or "No", with "No" as the default. An interleaved matrix looks like

```
format datatype=dna gap=- interleave=yes;
matrix
taxon_1  AACGATTCGT
taxon_2  AAGGAT--CA
taxon_3  AACGACTCCT
taxon_4  AAGGATTCCT

taxon_1  CCTGGTAC
taxon_2  CCTGGTAC
taxon_3  ---GGTAG
taxon_4  ---GGTAG
;
```

Gap -- This parameter specifies the format for gaps. Note that gap character can only be a single character and that it cannot correspond to a standard state (e.g., A,C,G,T,R,Y, M,K,S,W,H,B,V,D,N for nucleotide data).

Missing -- This parameter specifies the format for missing data. Note that the missing character can only be a single character and cannot correspond to a standard state (e.g., A,C,G,T,R,Y, M,K,S,W,H,B,V,D,N for nucleotide data). This is often an unnecessary parameter to set because many data types, such as nucleotide or amino acid, already have a missing character specified. However, for morphological or restriction site data, "missing=?" is often used to specify ambiguity or unobserved data.

Matchchar -- This parameter specifies the matching character for the matrix. For example,

```
format datatype=dna gap=- matchchar=.;
matrix
taxon_1  AACGATTCGT
taxon_2  ..G...--CA
taxon_3  .....C..C.
taxon_4  ..G.....C.
;
```

is equivalent to

```
format datatype=dna gap=-;
matrix
taxon_1  AACGATTCGT
taxon_2  AAGGAT--CA
taxon_3  AACGACTCCT
taxon_4  AAGGATTCCT
;
```

The only non-standard NEXUS format option is the use of the "mixed", "restriction", "standard" and "continuous" datatypes. Hence, if

you use any of these datatype specifiers, a program like PAUP* or MacClade will report an error (as they should because MrBayes is not strictly NEXUS compliant).

Matrix

This command specifies the actual data for the phylogenetic analysis. The character matrix should follow the dimensions and format commands in a data block. The matrix can have all of the characters for a taxon on a single line:

```
   begin data;
      dimensions ntax=4 nchar=10;
      format datatype=dna gap=-;
      matrix
      taxon_1  AACGATTCGT
      taxon_2  AAGGAT--CA
      taxon_3  AACGACTCCT
      taxon_4  AAGGATTCCT
      ;
   end;
```

or be in "interleaved" format:

```
   begin data;
      dimensions ntax=4 nchar=20;
      format datatype=dna gap=- interleave=yes;
      matrix
      taxon_1  AACGATTCGT
      taxon_2  AAGGAT--CA
      taxon_3  AACGACTCCT
      taxon_4  AAGGATTCCT

      taxon_1  TTTTCGAAGC
      taxon_2  TTTTCGGAGC
      taxon_3  TTTTTGATGC
      taxon_4  TTTTCGGAGC
      ;
   end;
```

Note that the taxon names must not have spaces. If you really want to indicate a space in a taxon name (perhaps between a genus and species name), then you might use an underline ("_"). There should be at least a single space after the taxon name, separating the name from the actual data on that line. There can be spaces between the characters.

If you have mixed data, then you specify all of the data in the same matrix. Here is an example that includes two different data types:

```
   begin data;
      dimensions ntax=4 nchar=20;
      format datatype=mixed(dna:1-10,standard:21-30) interleave=yes;
      matrix
      taxon_1  AACGATTCGT
      taxon_2  AAGGAT--CA
```

```
      taxon_3  AACGACTCCT
      taxon_4  AAGGATTCCT

      taxon_1  0001111111
      taxon_2  0111110000
      taxon_3  1110000000
      taxon_4  1000001111
      ;
   end;
```

The matrix command is terminated by a semicolon.

Finally, just a note on data presentation. It is much easier for others to (1) understand your data and (2) repeat your analyses if you make your data clean, comment it liberally (using the square brackets), and embed the commands you used in a publication in the mrbayes block. Remember that the data took a long time for you to collect. You might as well spend a little time making the data file look nice and clear to any that may later request the data for further analysis.

Translate

This command is used by MrBayes to specify the mapping between taxon names and taxon numbers in a Nexus tree file. For instance,

```
   translate
      1 Homo,
      2 Pan,
      3 Gorilla,
      4 Hylobates;
```

establishes that the taxon labeled 1 in the trees that follow is Homo, the taxon labeled 2 is Pan, etc.

Tree

This command is used by MrBayes to write trees to a nexus tree file. Trees are written in the Newick format. For instance,

```
   tree ((1,2),3,4);
```

describes an unrooted tree with taxa 1 and 2 being more closely related to each other than to taxa 3 and 4. If branch lengths are saved to file, they are given after a colon sign immediately following the terminal taxon or the interior node they refer to. An example of an unrooted tree with branch lengths is:

```
   tree ((1:0.064573,2:0.029042):0.041239,3:0.203988,4:0.187654);
```

Trees that are rooted (clock trees) are written with a basal dichotomy instead of a basal trichotomy. If the tree described above had been rooted on the branch leading to taxon 4, it would have been represented as:

```
tree (((1,2),3),4);
```
