Command Reference for MrBayes ver. 3.0B4, Serial Version

(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
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 certain 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

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
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.
You can assign up to 30 character sets. This option is best used
not from the command line but rather as a line in the mrbayes block
of a file.
------------------------------------------------------------------------
------------------------------------------------------------------------
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".

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

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 unordered: 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

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).

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 station-
             ary frequencies when morphological data are used. This option
             specifies how well the beta distribution will be approx-
             imated.
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 poss-
             ible. 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 pat-
             terns had the possibility of being sampled, then "all" should
             be specified (the default). Otherwise "variable" (only var-
             iable 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 imple-
             mented in this program. If you really believe that the pars-
             imony 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
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_mb3.0B4.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 a analysis.

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.

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.

Swapfreq -- This specifies how often a swap of the states of the chain is attempted. You must be running at least two chains for this option to be relevant. The default is Swapfreq=1.

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

Nchains -- How many chains are run for the MCMCMC variant.

Temp -- The temperature parameter for heating the chains.

```
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>)".
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.
Burnin       -- Which portion of the chain is discarded. It may take a
                while for the chain to reach stationarity. Samples taken
                when the chain is not at stationarity (the early phase of
                the chain) should be discarded. This parameter is one way
                of simply discarding early samples.
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.

Parameter       Options               Current Setting
-------------------------------------------------
Seed            <number>              1047501578
Ngen            <number>              1000000
Samplefreq      <number>              100
Swapfreq        <number>              1
Printfreq       <number>              100
Nchains         <number>              4
Temp            <number>              0.200000
Reweight        <number>,<number>     0.00 v 0.00 ^
Filename        <name>                temp.out.<p/t>
Burnin          <number>              0
Startingtree    Random/User           Random
Nperts          <number>              0
Savebrlens      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>              1047501578
Ngen            <number>              1000000
Samplefreq      <number>              100
```

```
Swapfreq        <number>              1
Printfreq       <number>              100
Nchains         <number>              4
Temp            <number>              0.200000
Reweight        <number>,<number>     0.00 v 0.00 ^
Filename        <name>                temp.out.<p/t>
Burnin          <number>              0
Startingtree    Random/User           Random
Nperts          <number>              0
Savebrlens      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 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 prset option to that partition.

Tratiopr -- This parameter sets the prior for the transition/transversion 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. 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 substitution 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 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 focussed
                    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), "equalin" (a glorified
                    Felsenstein 1981 model), " jones", "dayhoff", "mtrev",
                    "mtmam", "wag", "rtrev", "cprev", "vt", or
                    "blossum". You can also average over models by specifying
                    aamodelpr=mixed.
   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 sub-
                    stitution 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 ex-
                 ample, 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 diri-
                 chlet 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 constraints"). 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 dist-
                  ribution 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.
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 para-
                  meter 10 to partition 1, 40 to partition 3, and 15 to parti-
                  tion 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)
   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
   Ratepr           Fixed/Variable=Dirichlet    Fixed
   Shapepr          Uniform/Exponential/Fixed   Uniform(0.1,50.0)
   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)
   Symmetricbetapr  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 is reported. Although the program conistently uses the Dirichlet for specifying priors on rate sets, it reports the posterior in more familiar formats. Thus, it reports the transition/transversion ratio instead of the proportion of the rate sum occupied by the transition and transversion rates, which is the Dirichlet format. Similarly, it reports the subtitution rates of the GTR model scaled to the G <-> T rate, and the partition rates of a rate multiplier scaled such that the average rate is 1.0 over partitions. The "Report" command allows you to change this behavior. For instance, "report revmat=dir" will cause the program to report each substitution rate of the GTR model as the proportion of the rate sum that it represents. This can be useful if you want to use the posterior from one analysis as the prior in another analysis.

Parameter	Options	Current Setting
Tratio	Ratio/Dir	Ratio
Revmat	Ratio/Dir	Ratio
Ratemult	Scaled/Ratio/Dir	Scaled

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 three 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 the 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.

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

This command summarizes the information in a file named "filename".
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, sump, summarizes the information in the file <filename.p>.
The output is to the screen, and provides the mean, variance, and
95 percent credible interval for the parameter. You may want to
discard a specified number of observations from the chain as the
burn-in.

Parameter	Options	Current Setting
Filename	<name>	temp.p
Burnin	<number>	0

Sumt

This command summarizes the trees in a file named "<filename>". All of
the trees are read from the file and the proportion of the time any sin-
gle taxon bipartition is found is counted. The proportion of the time
that the bipartition is found is an approximation of the posterior prob-
ability of the bipartition. (Remember that a taxon bipartition is def-
ined 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 record-
ed. 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
the length of the branch. The partition information is output to a file
called "<filename>.parts". A consensus tree is also printed to a file
called "<filename>.con" and printed to the screen. The consensus tree
is either a 50 percent majority rule tree or a majority rule tree show-
ing all compatible partitions. You can also display the majority rule con-
sensus tree using a program such as PAUP*. The program also produces a

file called "<filename>.trprobs" that contains a sorted 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. Finally, the number of trees in the tree file that are skipped is controlled by the "burnin". The default is 0, but you may want to discard those trees that were sampled while the chain was not at stationarity. You can also display the majority rule consensus tree using a program such as PAUP*.

Parameter	Options	Current Setting
Filename	<name>	temp.t
Burnin	<number>	0
Displaygeq	<number>	0.05
Contype	Halfcompat/Allcompat	Halfcompat
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
```

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 requires 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.

```
**************************************************************************
*                                                                        *
*  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 char-
acter 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.

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. More-
              over, 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 char-
              acter 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

To be completed.

Tree

To be completed.
