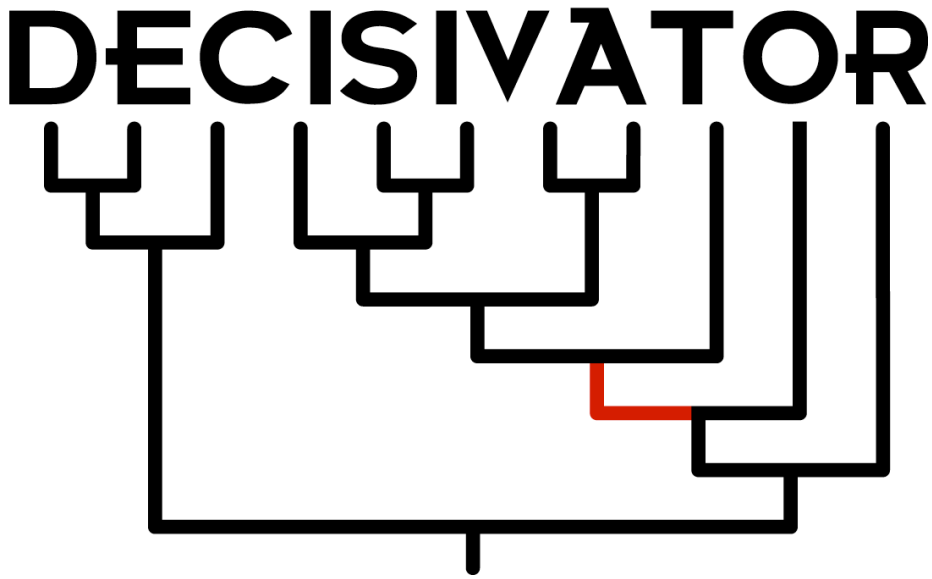
# DECISIVATOR



User Manual

A PuRGe Product
University of Idaho
Department of Biological Sciences
Complaints: josephwb@uidaho.edu

November, 2011

## Introduction and Overview

The impetus for Decisivator was a series of excellent papers by Michael Sanderson, Mike Steel, and Michelle McMahon (Sanderson, et al. 2010; Sanderson, et al. 2011; Steel and Sanderson 2010). These papers characterize phylogenetic 'decisiveness', which speaks to whether parts of a tree topology *can* be recovered given a particular taxon-character coverage pattern. It turns out (not surprisingly, but surprisingly unaddressed rigorously for so long) that certain edges in a phylogenetic tree topology *cannot* be reconstructed because of a lack of taxon information overlap. In order for a matrix to be decisive (that is, be able to reconstruct all possible trees relating the taxa), it must satisfy the 'four-way partition property'. I'll briefly outline this below, and then summarize in non-mathspeak.

Definition (Steel, Sanderson, and McMahon nomenclature, with extra things defined)

$X$ = {all taxa}
$N$ = number of taxa
$E$ = {all subsets of $X$ represented in the empirical sample}
= all unique column patterns in taxon-gene matrix; genes are exchangeable (for now)
$Q_E$ = {all unique taxon quartets present in $E$}
$Q_X$ = {all taxon quartets present in $X$}
= all possible quartets
= $N$ choose 4
$W$ = {all possible ways to partition $X$ into 4 non-empty sets}
= Stirling number of the second kind $S(n,k)$:

$$S(n,k) = \frac{1}{k!}\sum_{i=0}^{k}(-1)^i\binom{k}{i}(k-i)^n$$

So for our problem, there exists:

$$S(N,4) = \frac{1}{24}\sum_{i=0}^{4}(-1)^i\binom{4}{i}(4-i)^N$$

possible partitions. This increases exponentially with the number of taxa:

| N | $S(N,4)$ |
|---|---|
| 5 | 10 |
| 10 | 34105 |
| 15 | 42355950 |
| 25 | $4.677 \times 10^{13}$ |
| 50 | $5.282 \times 10^{28}$ |
| 100 | $6.696 \times 10^{58}$ |
| 250 | $1.364 \times 10^{149}$ |
| 500 | $4.465 \times 10^{299}$ |

This is actually the number of total internal edges across all possible

$$T(N) = \frac{(2N-2)!}{2^{N-3}(N-3)!}$$

unique trees with $N$ taxa (Cavalli-Sforza and Edwards 1967). Note that these numbers are not reported in any of the original decisiveness papers.

What the four-way partition property says is that for a matrix to be decisive, for every taxon partition $W_i$ (that is, all $S(N,4)$ possible way of splitting up the taxa into 4 non-empty sets), there exists an empirical quartet $Q_{Ei}$ such that the union of the elements of $W_i$ and $Q_{Ei}$ is non-empty. In normal-speak, no matter how you group the taxa into 4 partitions, there exists a gene such that you have *at least* one sequenced taxon for each partition. Put another, more visual way, for a given internal edge on a given tree to be recoverable (pendant edges are, of course, always recoverable), character information (say, a sequenced gene) must be available from at least one member in each of the four clades branching off of that internal edge. This is demonstrated in the figure below:

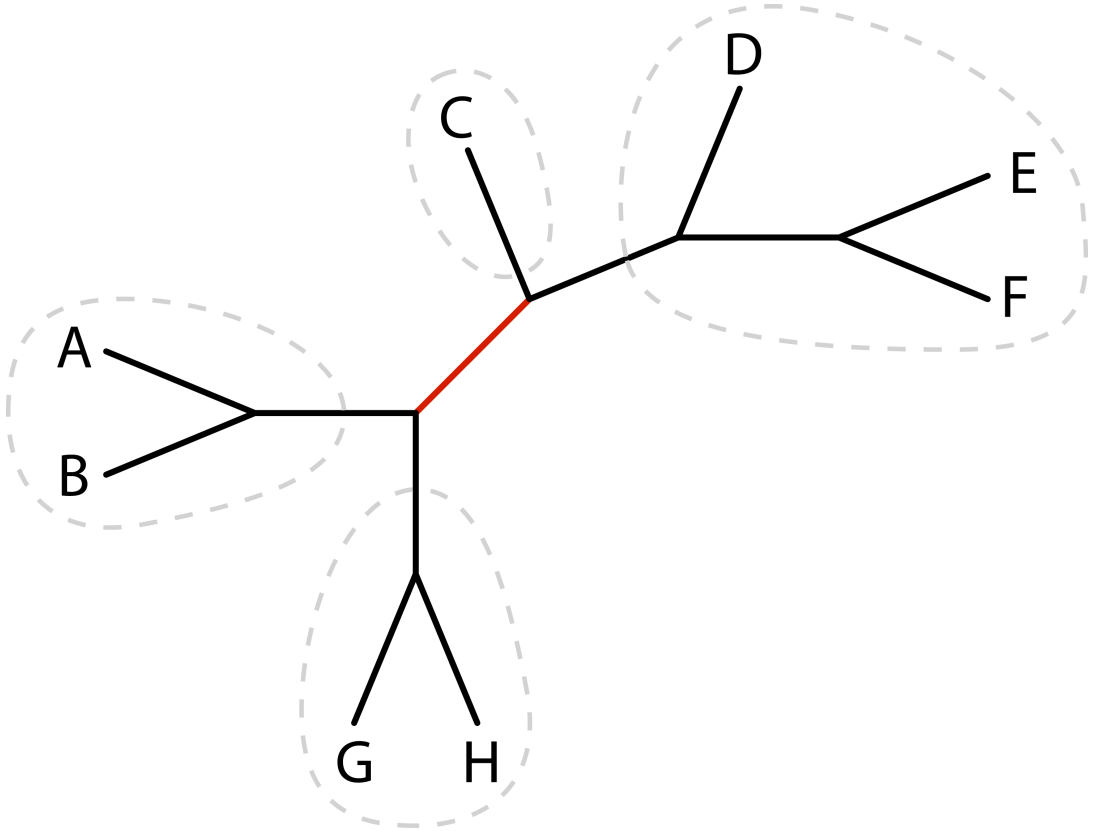



Figure 1. Demonstrating decisiveness for a single $W_i$ (or single internal edge), assuming a taxon-gene matrix of sequenced genes. In order to be able to a particular reconstruct edge $x$, there must exist at least one gene for which there exists at least one sequenced taxon from each of the clades $w_1$, $w_2$, $w_3$, and $w_4$. In the above tree, there must at least one gene that is sequences for one taxon in each of {A,B}, {G,H}, {C}, and {D,E,F}. If this is satisfied, the edge *can* be reconstructed with the given matrix. For the matrix to be decisive (that is, for all possible trees), all possible edges like the one above must be recoverable.

Determining Decisiveness

Steel, Sanderson, and McMahon give some useful conditions to evaluate for determining decisiveness.

1. *Non-decisiveness*: it may be of interest to determine if your matrix is *not* decisive, that is, whether your matrix *cannot* possibly reconstruct all possible tree topologies. This is easily ascertained. A necessary condition for decisiveness is that all taxon triplets are present in the matrix (that is, in $E$ above). So, checking how many of the $N$ choose 3 taxon triplets are present in $E$ can reject (but *not* confirm) decisiveness. In fact, if one just wants to know *if* their matrix is indecisive, they need not consider all $N$ choose 3 possible taxon triplets, but instead stop at the first absent triplet.

2. *Decisiveness*: As described above, decisiveness is ensured if each $W_i$ is satisfied. Steel and Sanderson give a sufficient condition for decisiveness. This relies upon the presence of a "reference-taxon", that is, a taxon that is sequenced for all genes. [The original calculations actually prune the matrix until a reference-taxon is 'created'. Decisivator does not require this]. So, given a reference-taxon, $r$, a sufficient condition is that all taxon quartets containing $r$ are present in the matrix (that is, in $Q_E$). An alternate sufficient condition that does not require a reference-taxon (probably what we will be working with) is that all possible taxon quartets (that is, $Q_X$) are present in $Q_E$. [A shortcut to determine this, without searching for each $Q_{Xi}$, would be to just see if the length of $Q_E = N$ choose 4].

However, while this is sufficient, it is not required. In other words, we can save ourselves some time if we were able somehow to cut down on the number of checks we need to make, and Decisivator tries to do just this. For example, each partitioning strategy $W_i$ can be satisfied by multiple (potentially *very* many) taxon quartets $Q_{Ei}$. In fact, for a given $W_i$ with partitions containing $w_1$, $w_2$, $w_3$, and $w_4$ taxa, there are:

$$\prod w_i$$

possible quartets that will satisfy $W_i$. This is maximized when partitions contain the same number of taxa (or closest thereto). This can be a very large number! Fortunately, **u**nique taxon quartets $Q_{Ei}$ can also satisfy multiple $W_i$. So, with some intelligent programming, we are able to avoid needlessly crunching through the huge number of possibilities (potentially ending with the answer: no, the matrix is *not* decisive for all trees)/

3. *Partial decisiveness:* As alluded to above, knowing whether a matrix is decisive or not may be of limited interest. For example, a matrix that can reconstruct 99.99999% of all possible trees will be deemed indecisive. A more useful metric is one of partial decisiveness, which computes the proportion of trees or edges a given matrix can potentially reconstruct. Branch-wise, the four-way partition property above can be applied, calculating the proportion of edges that could be recovered. Tree-wise, we take the approach of simulating a large number of random trees, and calculating the proportion of trees for which all edges can be reconstructed. Because this method is fast, we also use it for branch-wise partial decisiveness.

Dummy Example

Below is a dummy matrix to illustrate the points above. 'X' means that a particular taxon-gene has been sequenced.

| Taxon | Gene1 | Gene2 | Gene3 | Gene4 | Gene5 | Gene6 |
|---|---|---|---|---|---|---|
| A | | X | X | X | X | X |
| B | X | X | X | | X | X |
| C | X | | X | X | X | |
| D | X | X | X | X | | X |
| E | X | X | | X | | X |

Here we have 5 empirical taxon sets:
$E$ = {{B,C,D,E},{A,B,D,E},{A,B,C,D},{A,C,D,E},{A,B,C}}

Gene6 is redundant to Gene2 (i.e. they contain the same information) so it is not considered below. Gene5 is not useful to us (yet! It could be useful, in a non-trivial matrix, to add sequences to this gene) as it has less than 4 taxa sampled. The empirical quartets are thus:

$Q_E$ = {{B,C,D,E},{A,B,D,E},{A,B,C,D},{A,C,D,E}}

Notice that this is short of the (5 choose 4) = 5 total possible quartets ($Q_X$; missing quartet {A,B,C,E}). Below are $S(5,4) = 10$ the possible partitions ($W_i$):

| Partition 1 | Partition 2 | Partition 3 | Partition 4 | Genes that satisfy |
|---|---|---|---|---|
| {A} | {B} | {C} | {D,E} | 3 |
| {A} | {B} | {C,D} | {E} | 2 |
| {A} | {B,C} | {D} | {E} | 2,4 |
| {A,B} | {C} | {D} | {E} | 1,4 |
| {A} | {B} | {D} | {C,E} | 2,3 |
| {A} | {C} | {D} | {B,E} | 3,4 |
| {B} | {C} | {D} | {A,E} | 1,3 |
| {A,C} | {B} | {D} | {E} | 1,2 |
| {A,D} | {B} | {C} | {E} | 1 |
| {A} | {B,D} | {C} | {E} | 4 |

So, this matrix is decisive despite lacking a particular quartet. Obviously this is a trivial example, as each gene (1-4) is missing exactly 1 cell and so all are contributing equally to satisfying $W_i$.

Decisiveness of an individual tree

For a given tree we no longer need to consider all $S(N,4)$ possible partitions. Rather, we need only consider the $N$-3 internal edges of the tree and which taxon sets could resolve those edges. This drastically cuts down on the number of conditions we need to satisfy.

## Compiling

To compile Decisivator, uncompress the .tgz file using:

```
tar -xvzf Decisivator.tgz
```

Now, navigate to the `src` directory and type:

```
make
```

This will create the executable 'Decisivator'. That's it! Put this wherever in your system you prefer.

If you have troubles compiling (say, when moving from one computer to another), type:

```
make clean
```

and then continue with the `make` command. (If you have trouble getting this to work on your system, please complain to be at: josephwb@uidaho.edu).

## Input Files

Multiple user files may be passed in to Decisivator, but this must minimally include a file containing information regarding which taxa possess data for each partition. Normally a partition will refer to a sequenced genetic fragment (e.g. DNA/RNA or amino acids), but this could refer to any kind of character that might be used for phylogenetic reconstruction (e.g. morphology). Each independent set of characters should be designated as a separate partition: for genetics this will likely involve contiguous genes; for morphology, each character should be designated as a separate partition (unless characters can be identified as *non*-independent, in which case characters should be partitioned together). This taxon-partition information will most often take the form of a Nexus-formatted data file. Regardless of data type, partition information is gleaned from `CHARSET` declarations. `CHARSET`s can either be contiguous or interval in nature:

```
CHARSET ziggy = 1-2043;
CHARSET iggy = 2044-2099\3; [e.g. for codon data]
CHARSET joey = 2100-2100; [e.g. morph. character]
```

At the moment, the formatting of these `CHARSET` declarations is rather Draconian. For example, there should be no spaces around the '`-`' and '`\`'. More annoyingly, Decisivator does not yet support referential `CHARSET` declarations. For example:

```
CHARSET nineteenseventytwo = iggy david;
```

Will cause Decisivator to die a horrible death. If you wish to use referential `CHARSET` declarations, first rearrange your data into contiguous format (sorry).

A second option exists for passing in taxon-partition information. This is a 'legacy' format for which support will likely not continue, although in some instances (e.g. large/sparse matrices) it

may be easier to construct than a Nexus-formatted file. This second option involves a tab-delimited file. The first row should give partition names, and the first column should give taxon names. Existence of information for a taxon-partition is indicated by a '1', non-existence by a '0'. Below is a dummy example:

```
Gene_1      Gene_2      Gene_3      Gene_4
Taxon_1     1     0     1     1
Taxon_2     1     1     0     1
Taxon_3     0     1     1     1
Taxon_4     0     0     0     1
Taxon_5     1     1     1     0
Taxon_6     0     0     1     0
Taxon_7     1     0     1     0
Taxon_8     0     1     0     0
Taxon_9     0     0     1     1
```

In addition to a required data file, several optional files are possible. Perhaps most importantly is a file containing user-tree(s). These trees must be Nexus-formatted, and may (but not necessarily) utilize a translation table. The tree file may contain multiple trees; this may involve possible trees that are of interest as hypotheses, or it may involve a distribution of trees (i.e. posterior or bootstrap distribution). In the case of a distribution of trees, conventional options (e.g. burnin and thinning) are available (see below).

The last type of file that can be passed in indicates 'weights', one file for taxa and one for partitions. These weights are used to help determine optimal targeted data acquisition (e.g. targeted sequencing). By default, all taxa and all partitions have a weight of 1.0. Weight files, if present, need only identify those taxa or partitions which receive a weight *not* equal to 1.0. Below is an example weight file for taxa:

```
Taxon       Weight
Taxon_3     5
Taxon_7     3
Taxon_8     0.5
Taxon_2     0.1
```

Note that a header line is expected.

## Usage Options

While analysis conditions are accessed via a interactive menu prompt, several variables can be passed in to Decisivator via the command-line. To see these options, type:

```
./Decisivator -h
```

which will print out a slew of information to the screen:

```
*********************************************
            Decisivator version 0.47
                A PuRGe Product
              University of Idaho
       Department of Biological Sciences
        Complaints: josephwb@uidaho.edu
                 November, 2011
*********************************************


Program description: Calculates phylogenetic 'decisiveness' sensu
Sanderson and Steel.

To compile, type the following in a unix prompt:

make

To run, type:

./Decisivator [-d data_file] [-m taxon-gene_matrix] [-t tree_file] [-b
burnin] [-n thinning]
   [-w taxon_weights] [-l locus_weights]

where:

'data_file' is a simple 'vanilla' Nexus file containing sequences and
defined CHARSETs.
  - PLEASE NOTE! Only simple CHARSETs are currently supported.
    - e.g. contiguous (X-Y) or interval (e.g. codon: X-Y\3) data are
fine.
    - CHARSET referencing is NOT allowed at present (but will be!)

'taxon-gene_matrix' is a table listing taxa (rows) and genes (columns).
This is a legacy format.
  '1' indicates cell has been sequenced, while '0' indicates it has
not.
  First row should give locus names. First column should give taxon
names.

'tree_file' contains user tree(s) in Nexus format to evaluate
decisiveness upon. Tree(s) must be fully bifurcating.

'burnin' is the number of trees to ignore. Only makes sense with a
distribution of trees.

'thinning' is the interval between sampling trees (i.e. where every
nth tree sample will be retained).
  Only makes sense with a distribution of trees.

'taxon_weights' is a two-column (taxon, weight; with headers) file
listing weights for taxa
  based on some arbitrary accessibility criterion.
```

```
'locus_weights' is the analogous two-column (locus, weight; with
headers) file for locus weights.

NOTE: The taxon and locus weight files need not be complete. All
weights are 1.0 by default.
  Enter only weights which should be changed from the default.
```

For a description of the various input files, see the 'Input Files' section above.

If a distribution of trees is being passed in, it might not be of interest to analyze *every* tree, but instead a sampling of this distribution. Decisivator allows both '`burnin`' (where the first N trees are discarded) and '`thinning`' (where only every N tree is retained).

## Program Menu Options

Once in the program itself, several options are available:

```
Programs options:
 [A]dd virtual taxon-character(s)
 [M]erge taxa (i.e create a chimeric taxon)
 [E]xclude taxa
 [D]elete partition(s)
 [P]rint current matrix to the screen
 [C]alculate partial decisiveness using random trees
 [T]est of complete decisiveness
 [I]nvestiagte decisiveness on a provided user-tree
 [S]ummarize current status
 [W]rite current matrix
 [R]evert to original matrix
 [Q]uit
```

The `Add virtual taxon-character(s)` option allows a user to investigate how hypothetical data collection influences matrix decisiveness.

The `Merge taxa` option allows one to create a chimeric taxon. This can be useful for closely related taxa which, alone, have limited data, but combined have more partition-specific data. This will be useful for deep relationships (where the variance in tip values is much less than the variance between clades), but should not be done for 'shallow' trees.

The `Exclude taxa` option includes many sub-options for removing taxa according to some criterion dealing with missing data. The sub-options are:

```
Exclude:
 Taxa by [I]ndex
 Taxa by [N]ame
 Taxa missing [E]xactly N partitions
 Taxa missing N or [M]ore partitions
 Taxa possessing [O]nly N partitions
 Taxa possessing N or [F]ewer partitions
 Taxa possessing data for only a [S]pecific partition
```

```
Taxa exhibiting minimal [P]artition overlapping
[B]ack to main menu
```

Most of these options are self-explanatory. '`Index`' refers to the position of the taxon in the alignment (e.g. the 13th taxon). '`Partition overlap`' is an index that quantifies the number of taxa with which a particular taxon shares partition information. Rationale: taxa with high 'overlap' will satisfy more taxon quartets, and so contribute more to decisiveness. On the other hand, taxa with low 'overlap' feature in few quartets, and so are less pivotal to decisiveness. In the best scenario, taxon X shares (across all partitions) a partition with all N-1 taxa. This *could* be completely decisive, but not necessarily. Take the following dummy matrix (partition names excluded for clarity):

```
A     1     1     0     0     1
B     0     1     0     0     0
C     0     0     0     0     1
D     1     0     0     0     0
E     0     0     0     0     1
F     0     1     1     0     0
G     0     0     1     0     0
H     0     0     1     0     0
I     0     1     1     0     0
J     0     0     1     0     0
K     1     0     0     0     0
L     0     0     0     0     1
M     0     0     0     1     0
N     0     1     0     0     1
O     1     0     0     0     0
P     1     0     0     0     0
Q     1     0     0     0     0
R     0     0     1     1     0
S     0     0     0     1     0
T     0     0     0     1     0
U     0     0     1     1     0
V     1     0     0     1     0
W     0     0     0     1     0
X     0     0     0     1     0
Y     0     1     0     0     0
Z     1     0     0     0     0
```

Given that we are dealing with 26 taxa, a particular taxon can potentially share data with at most 25 other taxa. If we walk through the matrix, we come up with the following overlap counts:

```
Overlap   Shares a gene with:
A    15   B, C, D, E, F, I, K, L, N, O, P, Q, V, Y, Z
B    5    A, F, I, N, Y
C    4    A, E, L, N
D    7    A, K, O, P, Q, V, Z
E    4    A, C, L, N
F    10   A, B, G, H, I, J, N, R, U, Y
G    6    F, H, I, J, R, U
```

```
H    6    F, G, I, J, R, U
I    8    F, G, H, J, N, R, U, Y
J    6    F, G, H, I, R, U
K    7    A, D, O, P, Q, V, Z
L    4    A, C, E, N
M    8    R, S, T, U, V, W, X, Y
N    8    A, B, C, E, F, I, L, Y
O    7    A, D, K, P, Q, V, Z
P    7    A, D, K, O, Q, V, Z
Q    7    A, D, K, O, P, V, Z
R    12   F, G, H, I, J, M, S, T, U, V, W, X
S    7    M, R, T, U, V, W, X
T    7    M, R, S, U, V, W, X
U    12   F, G, H, I, J, M, R, S, T, V, W, X
V    10   A, D, K, R, S, T, U, W, X, Z
W    7    M, R, S, T, U, V, X
X    7    M, R, S, T, U, V, W
Y    5    A, B, F, I, N
Z    7    A, D, K, O, P, Q, V
```

As an example, let us focus on taxon A. Taxon A only has data for partitions 1, 2, and 5, so only consider these partitions for this taxon. Taxon A only shares partition information with taxa B, C, D, E, F, I, K, L, N, O, P, Q, V, Y, Z, but no others. Taxon A thus has an 'overlap' of 15 (i.e. do not count sharing with taxon N twice). This is the highest overlap, so we definitely want to keep it. Taxa C, E, and L have an overlap of only 4. Throwing out these taxa will leave a more overly-decisive matrix. We can therefore use minimal overlap as a criterion for throwing out taxa. Decisivator supports two exclusion criteria based on 'overlap':

`Exclude [M]inimally overlapping taxa (i.e. just the worst) or implement [T]hreshold`

That it, exclude the taxon (or taxa, in the case of a tie) that exhibits the worst overlap, or exclude *all* taxa below a particular threshold (chosen by the user). Use of the threshold criterion tends to yield matrices with 1) more taxa, 2) lower coverage, but 3) higher decisiveness than naïve exclusion practices (such as excluding all taxa with information for only a single partition).

The `Delete partition(s)` option probably is not useful, as deleting a partition cannot possibly increase decisiveness, although it can be informative as to the relative influence on decisiveness of different partitions. The option is mainly used for the situation where more `CHARSET`s exist in a file than are to be analyzed (say, overlapping partitions or `CHARSET`s grouped according to some criterion).

The `Print current matrix to the screen` option is meant only to double-check that matrix changes are recorded correctly.

The `Calculate partial decisiveness using random trees` option is the meat-and-potatoes function of Decisivator. Ideally it would be preferable to know whether a particular

matrix is decisive *for all possible trees*. However, in practice this might not be possible if taxon sampling is extensive enough, as there are N choose 4 quartet checks that must be made (see Introduction and Overview above). The way to get around this is to instead analyze a *large number of random trees*. In many ways, this test is much more useful than testing for complete decisiveness (see below), as this particular metric tells one *how decisive* (or, if you are a pessimist, *how indecisive*) a particular matrix is.

Decisivator offers two flavours of partial decisiveness:

```
Calculate partial decisiveness:
 [T]ree-wise
 [B]ranch-wise
```

These are very different takes on phylogenetic decisiveness. The first option gives the estimated proportion of trees for which the given matrix is decisive. This analysis proceeds very quickly, as only a single satisfaction of a given edge need be found; subsequent satisfactions of the edge are not considered. Furthermore, as it only take a single *unsatisfied* edge to make a given tree indecisive; Decisivator will bail on the current tree and move on to the next one. The second option gives the estimates proportion of decisive *branches* across random tree for which the given matrix is decisive. This analysis is considerably slower, as for each edge all possible relevant taxon quartets are investigated for the purpose of determining, edge-by-edge, the proportion of taxon quartets that satisfy the edge. These decisiveness scores can inform targeted sampling, and can be correlated with nodal support values, etc. These two decisiveness numbers can vary, as it only takes a single indecisive branch to make an entire tree indecisive.

The `Test of complete decisiveness` option checks to see if the current matrix is decisive for *all possible trees*. As mentioned above, this can take a large (potentially prohibitive) amount of time. The outcome of this test is rather limited: either a matrix is decisive or it is not (and a matrix that is 99.9999999% partially decisive is simply *not* completely decisive). This option will thus be less useful than the partial decisiveness analogue above, although it may be of use to users with nearly completely decisive matrices hoping to make it wholly decisive.

Beyond the test for partial decisiveness, the `Investigate decisiveness on a provided user-tree` option may be the most useful. Here, a user may pass in a tree or distribution of trees for decisiveness analysis. Tree distributions may undergo burnin and thinning via command-line options (see above). Note that all such tree distributions will necessarily contain decisive trees (or they wouldn't have been inferred in the first place!). Much more useful is comparing the branch-wise decisiveness of tree representing distinct phylogenetic hypotheses. Decisivator calculates these numbers, and can attach them to the tree(s) for viewing in a tree visualization program (e.g. FigTree (Rambaut 2006)). The results are also printed out to a bipartition table similar to those from PAUP*(Swofford 2003) or MrBayes(Ronquist and Huelsenbeck 2003):

```
Bipartition Decisiveness Scores:

12345678  Freq Poss %
------------------------
..**....  4    8    50
.....*.*  6    8    75
..**.*.*  12   12   100
..****.*  8    8    100
..******  5    5    100
```

These numbers may be useful in determining, say, whether a particular branch is really more highly supported or simply that more data are present with respect to that branch. Trees from these analyses can also be printed out to file with decisiveness annotations.

`Summarize current status` simply prints summary information to the screen (the number of 'reference taxa', the number of trees in memory, matrix coverage, and results of the various decisiveness tests).

The `Write current matrix` option allows the user to export a modified matrix in either Nexus or Phylip format. Obviously 'virtual taxon-characters' cannot be included; this function is mainly use for exporting matrices where taxa (or, possibly, partitions) have been excluded, so that matrix phylogenetic-sensitivity can be explored.

Finally, the `Revert to original matrix` option is self-explanatory: go back to the original taxon-partition matrix and original weights, and reset all decisiveness scores.

## References


Cavalli-Sforza LL, Edwards AWF. 1967. Phylogenetic analysis: models and estimation procedures. Am J Hum Genet, 19:233-257.

Rambaut A. 2006. FigTree. Available from the author (http://evolve.zoo.ox.ac.uk/software.html?id=figtree).

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics, 19:1572-1574.

Sanderson M, McMahon M, Steel M. 2010. Phylogenomics with incomplete taxon coverage: the limits to inference. BMC Evol Biol, 10:155.

Sanderson MJ, McMahon MM, Steel M. 2011. Terraces in phylogenetic tree space. Science, 333:448-450.

Steel M, Sanderson MJ. 2010. Characterizing phylogenetically decisive taxon coverage. Applied Mathematics Letters, 23:82-86.

Swofford DL. 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Sinauer Associates, Sunderland, Massachusetts.