#NEXUS

[
This is a fully commented setup file that can be used to implement the
likelihood ratchet using the PAUPRat program of Derek Sikes and Paul Lewis:
http://www.ucalgary.ca/~dsikes/software2.htm
Sikes, D.S. & Lewis, P.O. 2001. Beta software, version 1.
PAUPRat: PAUP* implementation of the parsimony ratchet.
Distributed by the authors. Department of Ecology and Evolutionary
Biology, University of Connecticut, Storrs, USA. June 2001.

You need to obtain a copy of the PAUPRat program, and to read its
Instruction manual. Basically, you first run this setup file through
PAUPRat, and then you run the PAUPRat output file through PAUP*.

The input files for PAUP* are your data file and the control file created
by PAUPRat from this setup template. The output files from PAUP* are:
lratchet.log - a text file with the results
lratchet.tre - a treefile with the optimal trees from each iteration
model.out   - a text file with the model parameter values used
lratchet.tmp - a temporary file that you can discard.
]
[
The original idea for the parsimony ratchet was by Kevin Nixon:
Nixon, K.C. 1999. The parsimony ratchet: a new method for rapid parsimony
analysis. Cladistics 15: 407-414.
]
[
The original version of the likelihood ratchet was by Rutger Vos:
Vos, R.A. 2003. Accelerated likelihood surface exploration: the likelihood
ratchet. Systematic Biology 52: 368-373.
The original setup file (June 2002) was downloaded from:
http://www.sfu.ca/~rvosa/likelihoodratchet
]
[
Modifications (November 2006) were by David Morrison, to implement the
'ratchet' part of the procedure, as this was missing from the Vos version
(which generates a new starting tree for each iteration). Also, the
strategy now provides a series of initial "successive approximations" to
estimate both the starting tree and the substitution-model parameter values.
Finally, the tree-search strategy has been optimized for maximum-likelihood
analyses of up to 150 sequences.
]
[
The successive approximations were based on the ideas of:
Sullivan, J., Abdo, Z., Joyce, P. & Swofford, D.L. 2005. Evaluating
the performance of a successive-approximations approach to parameter
optimization in maximum-likelihood phylogeny estimation. Molecular
Biology & Evolution 22: 1386-1392.
The specific implementation was inspired by:
Sullivan, J. 2005. Maximum likelihood methods for phylogeny estimation.
Methods in Enzymology 395: 757-779.
and by Peter Foster:
http://bioinf.ncl.ac.uk/molsys/data/like.pdf

http://www.ch.embnet.org/CoursEMBnet/PHYL03/Slides/unix_like_pfoster.pdf
Foster, P.G. 2001. Likelihood in Molecular Phylogenetics. Unpublished
notes used for Molecular Systematics course. Natural History Museum,
London, UK. July, 2001; September 2003.]
]
[
There are seven settings in this setup file that you might need to change:

(1) You must modify the 'nchar' command to match your data set.
(2) The default number of re-weighting iterations is 10, and the
percentage of characters to re-weight is 25. This produces 11 trees (the
initial tree plus 10 attempts to change island). You can change these
values using the 'nreps' and 'pct' commands (e.g. nreps=20 pct=15).
(3) The default re-weighting scheme treats all of the characters as
equal. You can change this using the 'wtmode' command.
(4) The default substitution model is GTR+G+I (general time reversible, with
gamma-distributed site-to-site variation and a proportion of invariable
sites). If you want to use a different model, then you need to change all of
the 'LScores' and 'LSet' commands. Note that the complexity of the model
does not affect the speed of the tree searches (since the model is fixed for
all searches), but does affect the speed of model estimation during the
initial successive approximations.
(5) The default tree-search strategy is SPR (subtree-prune-regraft) (the
PAUP* default is TBR, intended for parsimony searches). If you want to use
a different strategy, then you need to change all of the 'Swap=spr'
commands. If you want separate strategies for the re-weighted and
unweighted searches, then you need to change the commands labelled
'rewtdcmd' and 'normcmd', respectively.
(6) During the tree search the log-likelihood scores are not fully optimized
unless they are within 2% of the current optimum value (the PAUP* default is
5%, intended for <50 sequences). If you want to use a different strategy,
then you can change this using the 'ApproxLim' command (e.g. ApproxLim=1 for
data sets with larger negative log-likelihoods). Note that this value can
make a big difference to how long the ratchet takes to run; even a change in
value of 0.01% can be important for large data sets (multiple genes for >100
sequences). For a discussion, see:
Rogers J.S. & Swofford, D.L. 1998. A fast method for approximating maximum
likelihoods of phylogenetic trees from nucleotide sequences. Systematic
Biology 47: 77-89.
(7) Only one tree is saved during the re-weighted tree search, on the
principle that the optimal tree does not necessarily have to be found for
this search (only for the unweighted search). If you do want to find the
optimal tree, then you need to change the 'MulTrees=no' command. Also, you
might like to consider using the 'RearrLimit' or 'TimeLimit' commands if you
wish to prevent unduly long re-weighted searches.
]

[Start of instructions. Don't change.]

begin pauprat;

[Enter the number of characters after nchar= on the following line.]

  dimensions nchar=10922;

[Enter the number of iterations after 'nreps=' and the fraction of
characters drawn after 'pct=' on the following line. The default
values seem to work, but you can always use more replicates and a
greater percentage (probably up to 35%, as for the parsimony
ratchet) if you expect a very complex landscape, or if you have a small
data set and/or a very fast computer. 'Seed=0' sets a randomly chosen
random-number seed, but you can pre-specify a particular seed if you
want exact repetition of the characters chosen for re-weighting.]

    set seed=0 nreps=10 pct=25;

[Choose the weighting mode. The choices are: additive, multiplicative,
uniform. Typically, the default works fine unless you are using a
weighting scheme (i.e. a 'WtSet' command) based on codon positions, in
which case you might want to try 'mult'.]

    set wtmode=uniform;

[Don't change this unless you want a lot of output.]

    set terse;

[Opening message.]

    startcmd "[!* * * * * * * * * * * * * * * * * *]";
    startcmd "[!* ----- Likelihood Ratchet v2 ----- *]";
    startcmd "[!*   David A. Morrison   *]";
    startcmd "[!*   Sveriges Lantbruksuniversitet   *]";
    startcmd "[!*   November, 2006   *]";
    startcmd "[!* * * * * * * * * * * * * * * * * *]";

[Record the current time.]

    startcmd "Time";

[The *.log file stores PAUP*'s display buffer.]

    startcmd "Log File=lratchet.log";

[Automatically increase the 'maxtrees' setting. Don't change.]

    startcmd "Set Increase=auto";

[Get the starting tree. No need to change unless you want to specify
a user starting tree, in which case use the 'GetTrees' command.]

    startcmd "DSet Dist=logdet Objective=ME Rates=equal PInv=0 Subst=all
NegBrLen=setzero";
    startcmd "NJ BioNJ=yes ShowTree=no BrLens=no BreakTies=systematic";

[Set the optimality criterion to ML. Don't change.]

    startcmd "Set Criterion=likelihood";

[Optimize the substitution-model parameters.]

   startcmd "LScores 1 / NST=6 BaseFreq=estimate RMatrix=estimate Rates=*gamma*
Shape=estimate PInvar=estimate";

[The *.tmp file contains the current working tree. It can be used
to re-start a ratchet run that has been interrupted. Don't change.]

   startcmd "SaveTrees File=lratchet.tmp Replace";
   startcmd "Time";

[Do an NNI search based on these parameter estimates, and then
optimize the substitution-model parameters again.]

   startcmd "LSet BaseFreq=previous NST=6 RMatrix=previous Rates=gamma
Shape=previous PInvar=previous ApproxLim=2 AdjustAppLim=no";
   startcmd "HSearch Status=no Start=current Swap=nni MulTrees=yes";
   startcmd "SaveTrees File=lratchet.tmp Replace";
   startcmd "LScores 1 / NST=6 BaseFreq=estimate RMatrix=estimate Rates=*gamma*
Shape=estimate PInvar=estimate";
   startcmd "Time";

[Do an SPR search based on these parameter estimates, and then
optimize the substitution-model parameters again. Save the model
parameter values to the model.out file. The 'LongFmt' option is used only
to deal with a long-standing bug in PAUP* version 4b10.]

   startcmd "LSet BaseFreq=previous NST=6 RMatrix=previous Rates=gamma
Shape=previous PInvar=previous ApproxLim=2 AdjustAppLim=no";
   startcmd "HSearch Status=no Start=current Swap=spr MulTrees=yes";
   startcmd "SaveTrees File=lratchet.tmp Replace";
   startcmd "Default LScores LongFmt=yes";
   startcmd "LScores 1 / NST=6 BaseFreq=estimate RMatrix=estimate Rates=*gamma*
Shape=estimate PInvar=estimate ScoreFile=model.out Replace";
   startcmd "Default LScores LongFmt=no";
   startcmd "Time";

[The *.tre file contains the set of solutions for the initial tree
plus all subsequent iterations. There will thus be at least nreps+1
trees in this file at the end. Don't change.]

   startcmd "SaveTrees File=lratchet.tre Replace";

[Set the substitution-model parameters for the likelihood model used
in all subsequent iterations.]

   startcmd "LSet BaseFreq=previous NST=6 RMatrix=previous Rates=gamma
Shape=previous PInvar=previous ApproxLim=2 AdjustAppLim=no";

[Commands for the branch-swapping cycles under the re-weighted scheme.
This is the tree search that tries to get to another island of trees.]

   rewtdcmd "HSearch Status=no Start=1 Swap=spr MulTrees=no";

[Updates the *.tmp file to contain the current tree. Don't change.]

```
  rewtdcmd "SaveTrees File=lratchet.tmp Replace";
  rewtdcmd "Time";
```

[Commands for the branch-swapping cycles under the original weighting scheme. This is the tree search that tries to find the peak of the island.]

```
  normcmd "HSearch Status=no Start=1 Swap=spr MulTrees=yes";
```

[Update the *.tmp file to contain the current starting tree. Don't change.]

```
  normcmd "SaveTrees File=lratchet.tmp Replace";
```

[Update the set of optimal trees over all iterations. Note that both the 'GetTrees' and 'SaveTrees' commands are used in order to get all of the trees into a single Trees block in the treefile (the default in PAUP* is to create a separate block for each ratchet iteration). Don't change.]

```
  normcmd "GetTrees Rooted=no Unrooted=yes File=lratchet.tre Mode=7";
  normcmd "SaveTrees File=lratchet.tre Replace";
  normcmd "GetTrees Rooted=no Unrooted=yes File=lratchet.tmp Mode=3
Warntree=no";
  normcmd "Time";
```

[Retrieve the final set of optimal trees at the end of the ratchet search. Don't change.]

```
  stopcmd "GetTrees File=lratchet.tre Mode=3";
```

[Print the negative log-likelihoods and the between-tree distances for the set of optimal solutions. Note that the trees are numbered in reverse order (i.e. the final-iteration tree is #1). There will be more than nreps+1 trees if some of the iterations found several equally optimal trees. Don't change.]

```
  stopcmd "LScores All / SortTrees=yes";
  stopcmd "TreeDist Metric=symdiff";
  stopcmd "Time";
```

[Stop the logging of the display buffer.]

```
  stopcmd "Log Stop";
```

[Final message.]

```
  stopcmd "[!* * * * * * * * * * * * * * * *]";
  stopcmd "[!* -- THIS SEARCH IS COMPLETE -- *]";
  stopcmd "[!*  A LOG FILE HAS BEEN WRITTEN  *]";
  stopcmd "[!* AND ALL TREES HAVE BEEN SAVED *]";
  stopcmd "[!*    IT IS OKAY TO QUIT PAUP    *]";
```

```
    stopcmd "[!* * * * * * * * * * * * * * * *]";
    stopcmd "Quit";

[Define the name of the ratchet script file.]

    write file=lratchet.nex;

end;
```