

HIVtree Manual

Anna Nagel and Bruce Rannala

May 20, 2022

HIVtree is a Bayesian phylogenetic inference program that estimates HIV latent integration times and node ages on a fixed phylogenetic tree. This program was originally modified from PAML version 4.9. The program takes as input (1) a tree in newick format without branch lengths, (2) an alignment with tip dates, (3) a file with the list of latent sequences, and (4) a control file. The control file contains model specifications and specifications of the MCMC.

The control file is very similar to the control file in `mcmctree`, a program in PAML. Here, only new or changed options to the control file will be detailed. We refer readers to the `mcmctree` manual for a full description.

1 Control File

clock: 1 must be used for the clock model, which specifies a strict clock. Other clock models are available in PAML but not in HIVtree.

latentFile: This is the name of text file that provides the name of all of the latent sequences. The sequence names should match the names in the alignment and tree file with one name per line.

latentBound: This provides a hard upper bound on all of the latent integration times in the analysis. This is specified in backward time in the time units specified by the `TipDate` option. For example, consider the options "`latentBound = 3`", "`tipDate = 1 1000`", and time specified in days in the sequence names. This means no latent ages can be more than 3000 days older than the time of the last sample.

RootAge: This specifies the prior on the root age. There are two options, either a shifted gamma prior, $G(\alpha, \beta)$, or a uniform prior, $U(a, b)$. The gamma distribution is shifted by adding the first sample time to the distribution. This ensures there is no density after sequences are sampled. The uniform prior has hard bound, so there is no density outside of the range between a and b . The parameters for both the uniform and gamma distributions must also be chosen with the time unit transformation going backward in time. For example, with option "`tipDate = 1 1000`" and the dates for the sequences specified in days, $U(3,4)$ would be a uniform root age prior between 3000 and 4000 days prior to the *last* sample time. $G(1, 1)$ would be a gamma prior with mean 1000 days prior to the *first* sample time with variance 1000 days. Note that the user specified prior will not match the induced prior when running without data (option `usedata = 0`) because of the constraints imposed by the tip ages and rank order of the node. The user should run without data to see what the induced prior will be.

2 Running the Program

The program is run in the same way as `mcmctree`. All path names must be relative to the current directory.

```
./mcmctree control.ctl
```