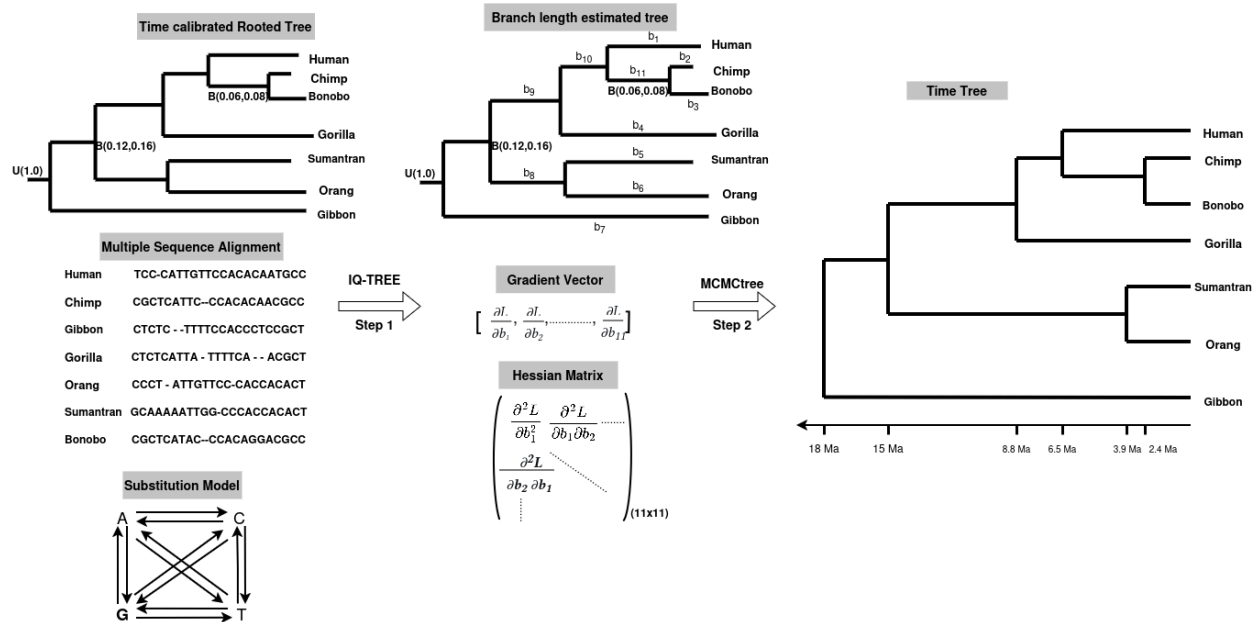


Recipe for Phylogenetic dating with IQ2MC workflow

Here, we demonstrate a worked example of the IQ2MC workflow for phylogenetic dating using MCMCtree's approximate likelihood method. For demonstration purposes, we have used the example available in [MCMCtree tutorials](#). The DNA data of 7 primate species and the tree file used for the example are available [here](#).



In this example, a rooted tree with time calibrations, a multiple sequence alignment and substitution model needed to be provide to IQ-TREE.

- The MSA could be in any format IQ-TREE supports for single or partition models.
- The substitution model could be any single, mixture or partition model available in IQ-TREE such as GTR+G4 for DNA or LG+G4+C60 for protein data.
- The rooted tree should be annotated with time calibrations and should looks like the following.

```
(((((human,(chimpanzee,bonobo)))'B(.06,.08)',gorilla),(orangutan,sumatran))'B(.12,.16)',gibbon)'U(1.0)';
```

MCMCtree support multiple time calibrations types depending on the nature of fossil/sampling times available. Here, we have applied a uniformly distributed time calibartion for (human,(chimpanzee,bonobo)) clade using the node label 'B(.06,.08)' which setup the lower bound to 6 million years and upper bound to 8 million years as MCMCtree by default use 100 million years time scale. The same strategy can be use to calibrate other clades. Since MCMCTree required a root calibrations, here we have applied 'U(1.0)' tip label at the root as the root calibration. This setup 100 million years as upper bound the root. Other calibrations methods such as Gamma and Skew normal distributions are available in MCMCtree and more details can be found at [PAML documentation](#).

After calibrating tree, following command can be used to generate the necessary files to perform phylogenetic dating with MCMCtree using approximate likelihood.

```
./iqtree3 -s mtCDNApri123.phy -m GTR+G4 -te mtCDNApri.tree --dating mcmctree -prefix mtCDNApri123
```

Here, IQ-TREE will generate following files to perform phylogenetic dating with MCMCtree.

- [mtCDNApri123.mcmctree.hessian](#): the hessian file which contains the gradients vector and the Hessian for approximate likelihood dating.
- [mtCDNApri123.rooted.nwk](#): the rooted tree file which is compatible with the Hessian file. It is necessary to use this tree file with MCMCtree for dating as the Hessian is calculated with respect to the ordering of taxa of this tree file.
- [mtCDNApri123.mcmctree.ctl](#): the control file that can be used directly to run MCMCtree from IQ-TREE output of step 2.
- [mtCDNApri123.dummy.phy](#): It is not necessary to use the alignment with MCMCtree under approximate likelihood dating. However, in the current format MCMCtree requires an alignment, and you can simply use this dummy alignment file as the input to MCMCtree to save compute.

In the above command, we have used GTR+G4 as the substitution model. User can specify any substitution model supported by IQ-TREE. Also, the parameters in the [mtCDNApri123.mcmctree.ctl](#) file can be manually changed or can be specify using the commands [available here](#) to achieve the convergence of MCMC estimations. The control file has the following format.

```
seed = -1
seqfile = mtCDNApri123dummy.phy * A dummy alignment only allow to run MCMCtree
treefile = mtCDNApri123.rooted.nwk * Rooted newick tree with annotated fossil/tip dates
mcmcfile = mtCDNApri123.mcmctree.log * MCMC log of parameters that can be examined in Tracer
outfile = mtCDNApri123.mcmctree.out * Output of the summerized results of MCMCtree
ckpfile = mtCDNApri123.mcmctree.ckp * Checkpoint file of MCMCtree
hessianfile = mtCDNApri123.mcmctree.hessian * File with gradient and hessian matrix

checkpoint = 1 * 0: nothing; 1 : save; 2: resume
ndata = 1 * number of partitions
seqtype = 0 * 0 : nucleotides; 1: codons; 2: AAs (not required if the approximate likelihood method is used)
usedata = 2 * 0: sampling from priors with no data; 1: exact slow likelihood; 2: approximate likelihood
clock = 3 * 1: global clock with equal rates; 2: independent rates; 3: correlated rates
RootAge = <1.0 * safe constraint on root age, used if no fossil for root.
```

```

BDparas = 1 1 0.5 * birth rate, death rate, sampling priors for sampling times
finetune = 1: 0.1 0.1 0.1 0.01 .5 * auto (0 or 1) : times, musigma2, rates, mixing, paras, FossilErr
print = 1 * 1: normal output; 2: verbose output

*** These parameters are used for multi-loci partitioned data (ndata > 1), see dos Reis et al .(2013)

rgene_gamma = 2 2 * alpha and beta parameter of Dirichlet-gamma prior for mean rate across loci
for clock=2 or 3
sigma2_gamma = 1 10 * alpha and beta parameter of Dirichlet-gamma prior for rate variance across
loci for clock=2 or 3

*** These parameters control the MCMC run

burnin = 20000
sampfreq = 100
nsample = 20000

*** Note: Total number of MCMC iterations will be burnin + (sampfreq * nsample)

*** Following parameters only needed to run MCMCtree with exact likelihood (usedata = 1), no need
to change anything for approximate likelihood (usedata = 2)

model = 0 * 0:JC69, 1:K80, 2:F81, 3:F84, 4:HKY85
alpha = 0 * 0: No rate heterogeneity across sites; otherwise: initial alpha parameter of the Gamma
distribution
ncatG = 0 * Number of rate categories for the discrete Gamma distribution

cleandata = 0 * remove sites with ambiguity data (1:yes, 0:no)?

kappa_gamma = 6 2 * gamma prior for kappa of the HKY model
alpha_gamma = 1 1 * alpha and beta parameter of Gamma distribution for heterogeneous rates
across sites

```

After, obtaining the above files, following command can be used to run MCMCtree to obtain time estimates.

```
mcmctree mtCDNApri123.mcmctree.ctl
```

MCMCtree generates following files which provides and summarizes the MCMC estimates for each parameter of the analysis.

- [mtCDNApri123.mcmctree.log](#): All MCMC samples obtained during the analysis. [Tracer](#) can be used to check the convergence of the MCMC run.
- [mtCDNApri123.phy.mcmctree.out](#): Summary of the MCMC estimates. Here, the means and confidence intervals of time and rate estimates can be found.
- [FigTree.tre](#): The time tree which can be viewed using [Figtree](#).

The FigTree.tre file obtained after the MCMCtree run should look like the following.

