

Example of using CreDating

Xavier Didelot

2017-11-22

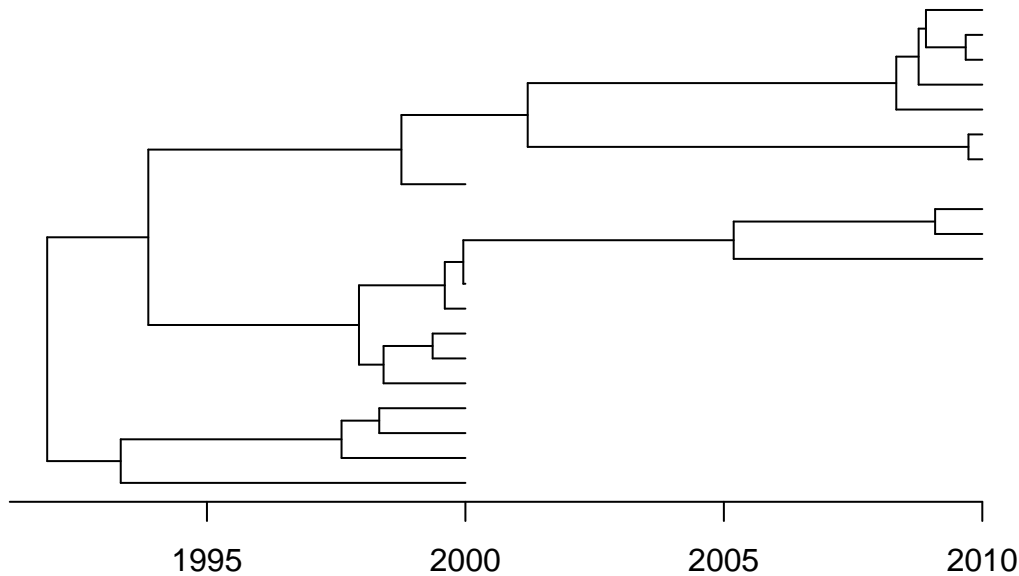
Initialisation

```
library(CreDating)
library(ape)
set.seed(0)
```

Data

We start by generating a coalescent tree with 10 leaves sampled in 2000, 10 leaves sampled in 2010, and a coalescent time unit of 10 years:

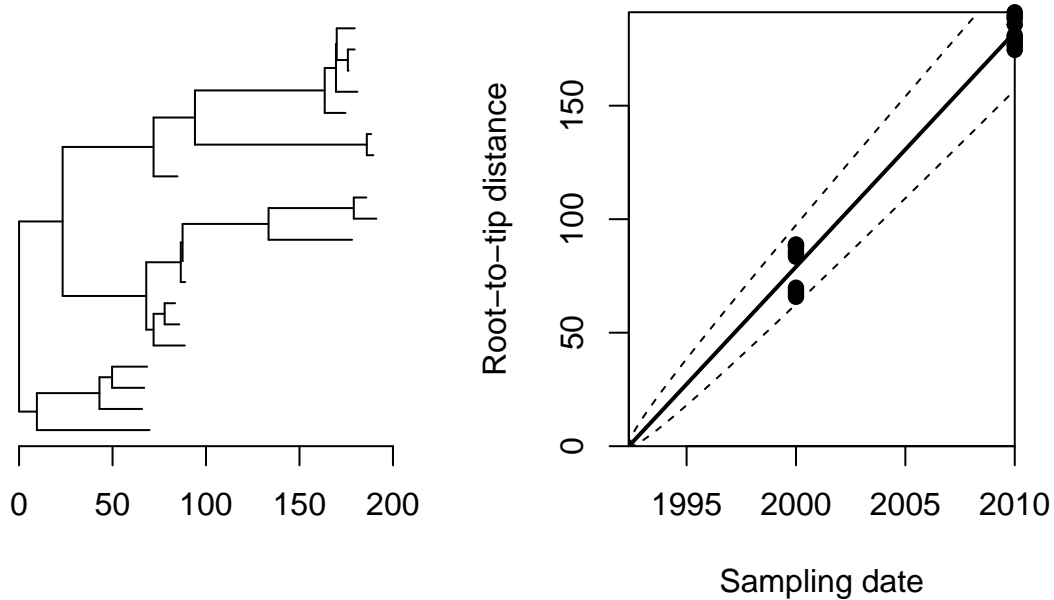
```
dates=c(rep(2000,10),rep(2010,10))
phy=simcoaltree(dates,neg=10)
plot(phy,show.tip.label = F)
axisPhylo(backward = F)
```



On each branch we observe a number of substitutions which is distributed $\text{Gamma}(rl, 1)$ where l is the branch length and $r = 10$ per year is the substitution rate. We can simulate an observed phylogenetic tree and perform a root-to-tip analysis as follows:

```
obsphy=simobsphy(phy,rate=10)
res=roottotip(obsphy,dates)
```

Rate=1.03e+01,MRCA=1992.36,R2=0.98,p<1.00e-04



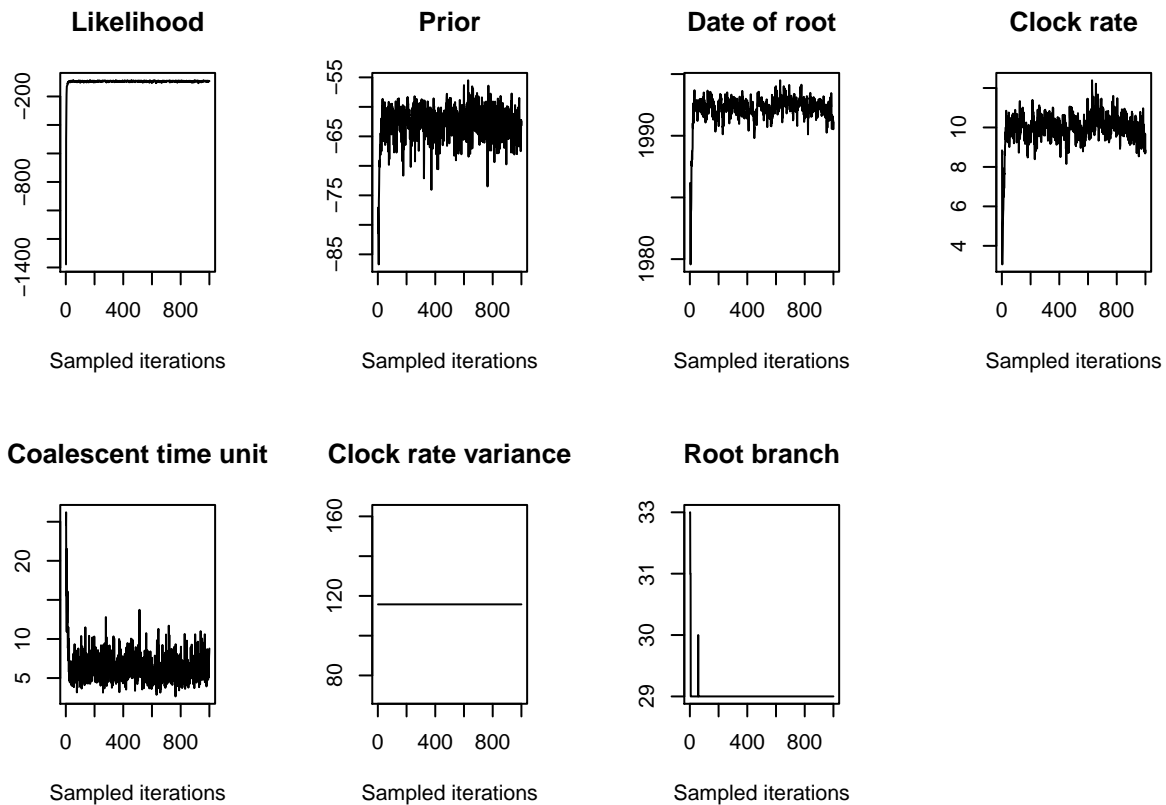
Let's consider that we do not know where is the root of the observed phylogeny, and that the data of the last sample has been lost:

```
obsphy=unroot(obsphy)
dates[length(dates)]=NA
```

First analysis

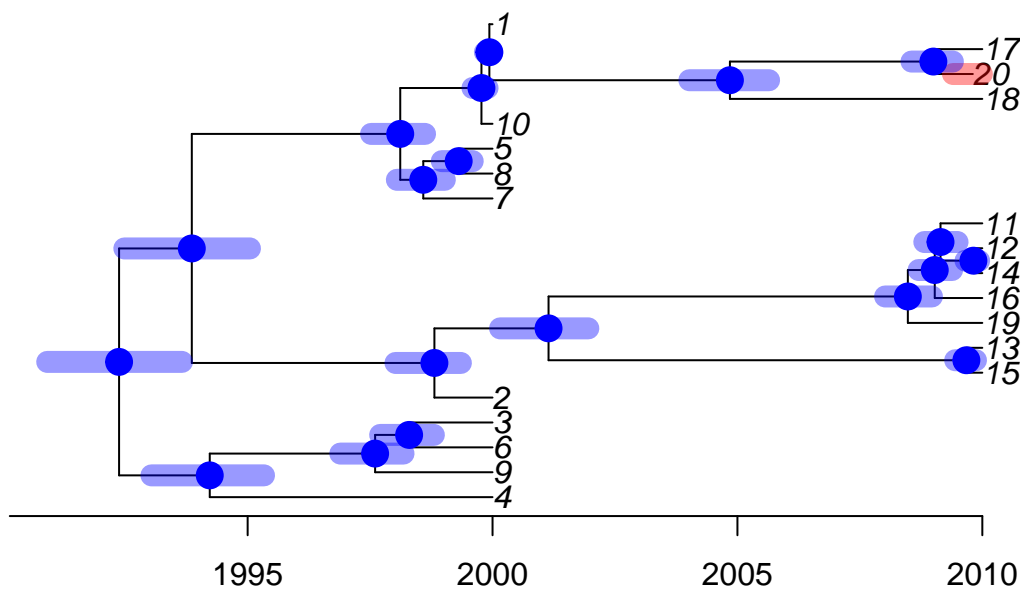
We run the dating analysis as follows:

```
res=credat(obsphy,dates)
plot(res,'trace')
```



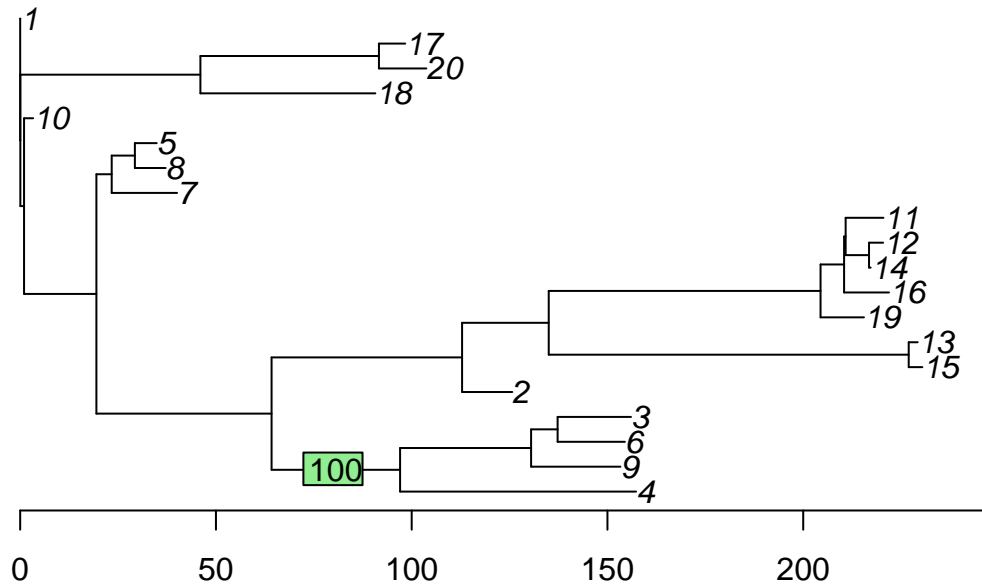
Let's see what the result looks like:

```
plot(res, 'treeCI')
```



Let's see where the root was inferred:

```
plot(res, 'treeRoot')
```



Testing significance of clock signal

We run the algorithm again, with all sampling dates forced equal, and compare the fit of the two runs in order to assess the statistical significance of the temporal signal:

```
res2=ccreate(obsphy,rep(2015,length(dates)))
```

```
## Warning in roottotip(tree, date, showFig = F): Warning: All dates are identical.
```

```
modelcompare(res,res2)
```

```
## The first model has DIC=230.07 and the second model has DIC=1068.75.
```

```
## Model 1 is definitely better.
```

Second analysis using a relaxed clock model

Let's do another analysis, using a relaxed clock model, and compare to the first analysis:

```
res3=ccreate(obsphy,dates,model='relaxedgamma')
```

```
modelcompare(res,res3)
```

```
## The first model has DIC=230.07 and the second model has DIC=327.12.
```

```
## Model 1 is definitely better.
```

Analysis of data generated with relaxed clock model

Let's start everything again with a new dataset using the relaxed model:

```
rm(list=ls())
```

```
set.seed(0)
```

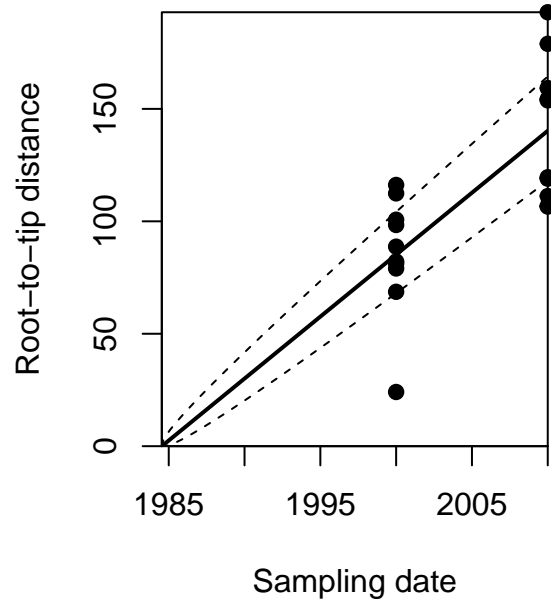
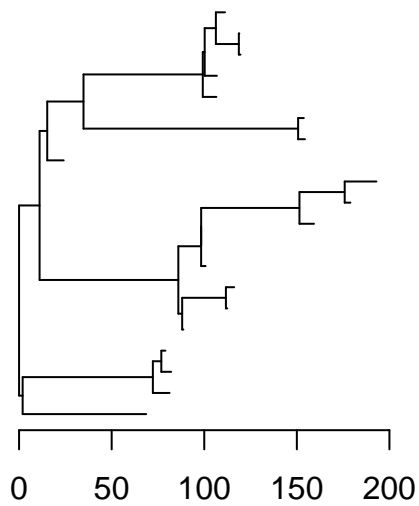
```
dates=c(rep(2000,10),rep(2010,10))
```

```
phy <- simcoaltree(dates,neg=10)
```

```
obsphy=simobsphy(phy,rate=10,model='relaxedgamma',ratevar = 100)
```

```
res=roottotip(obsphy,dates)
```

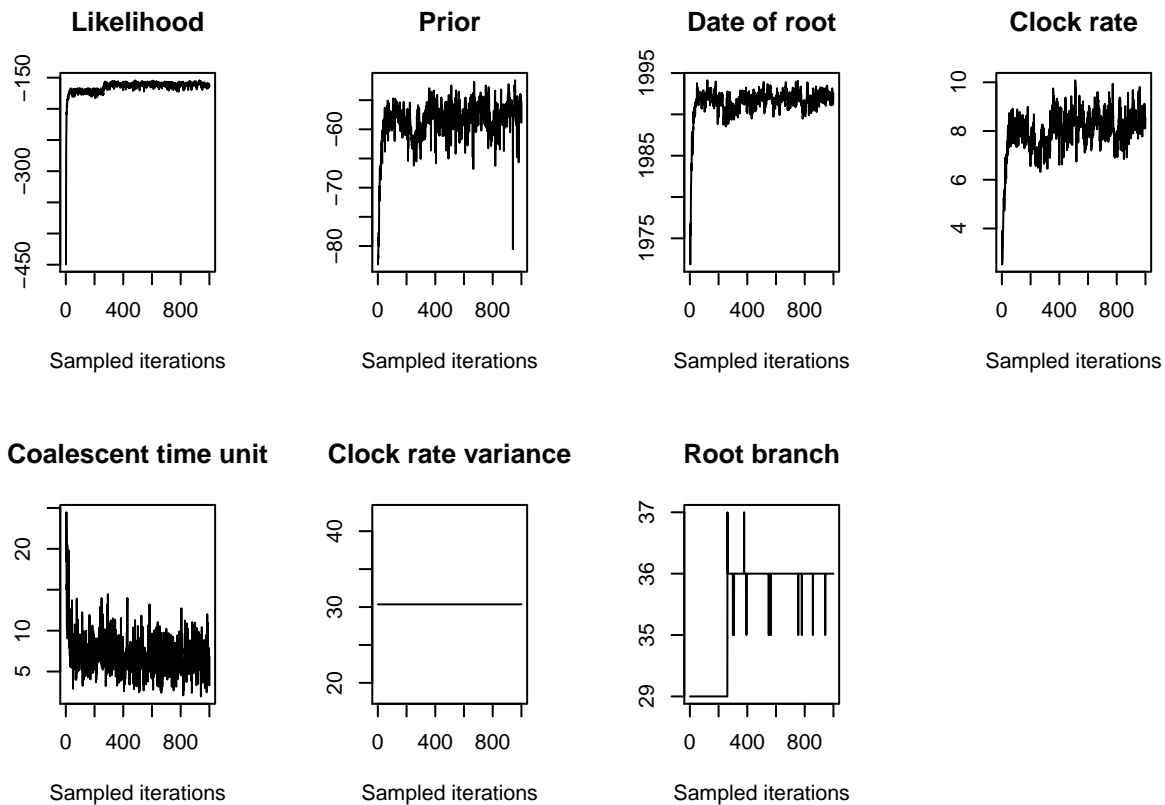
Rate=5.51e+00,MRCA=1984.54,R2=0.50,p<1.00e-04



analyse using a strict clock model:

```
res=cdate(obsphy,dates)
plot(res,'trace')
```

First we



Let's see if the temporal signal is significant:

```
res2=ccreate(obsphy,rep(2015,length(dates)))
```

```
## Warning in roottotip(tree, date, showFig = F): Warning: All dates are identical.
```

```
modelcompare(res,res2)
```

```
## The first model has DIC=352.64 and the second model has DIC=633.33.
```

```
## Model 1 is definitely better.
```

Finally we analyse with the relaxed clock model:

```
res3=ccreate(obsphy,dates,model='relaxedgamma')
```

```
plot(res3,'trace')
```

```
modelcompare(res,res3)
```

```
## The first model has DIC=352.64 and the second model has DIC=274.29.
```

```
## Model 2 is definitely better.
```

