

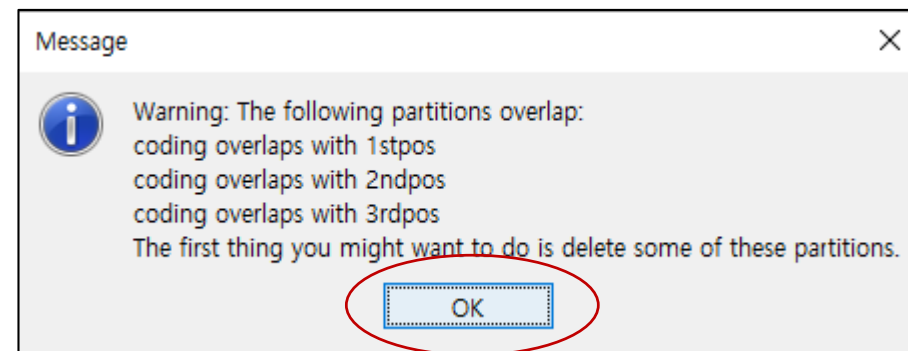
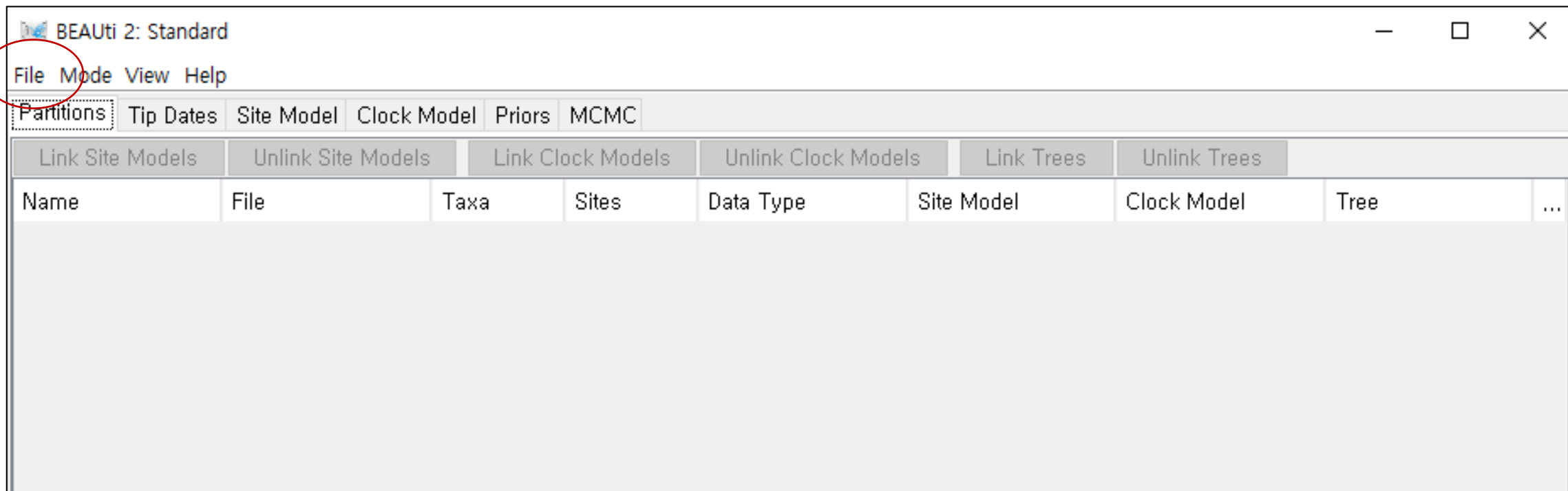
BEAST 프로그램 사용법 초간단 설명

(한국진화학회 2024 겨울학교 자료)

극지연구소 서태건

- 베이esian 분기연대 추정 초간단 실습
- <https://taming-the-beast.org/tutorials/Introduction-to-BEAST2/> 의 내용 일부 + α (본 자료에서는 prior가 약간 다르기 때문에 결과가 정확히 일치하지는 않지만 큰 차이는 없음)
- ① BEAUti 프로그램으로 설정파일 만들고, 이를 ②Beast 프로그램으로 읽어 들여 MCMC를 실행함. 결과 파일을 ③Tracer, TreeAnnotator, Densitree, FigTree 프로그램을 이용하여 확인함.

(1) BEAUi를 실행시키고 File/import alignment / 클릭, primate-mtDNA.nex (examples\nexus 폴더에 있음) 읽어 들인다. 파티션 중복 정의 (뒤에서 설명)에 따른 경고 메시지는 무시하고 OK 누름



(primate-mtDNA.nex 파일 내부. {begin taxa; end;}, {begin characters; end;} 블록의 정의 방식을 눈여겨 보자. Nexus format에 대한 자세한 사항은 PAUP 프로그램 설명서 참조.

<http://phylosolutions.com/paup-documentation/paupmanual.pdf>)

```
#NEXUS
[!Data from:
      Hayasaka, K., T. Gojobori, and S.
Horai. 1988. Molecular phylogeny
      and evolution of primate
mitochondrial DNA. Mol. Biol. Evol.
      5:626-644.
]
begin taxa;
      dimensions ntax=12;
      taxlabels
Lemur_catta
Homo_sapiens
Pan
Gorilla
Pongo
Hylobates
Macaca_fuscata
M._mulatta
M._fascicularis
M._sylvanus
Saimiri_sciureus
Tarsius_syrichta
;
end;
```

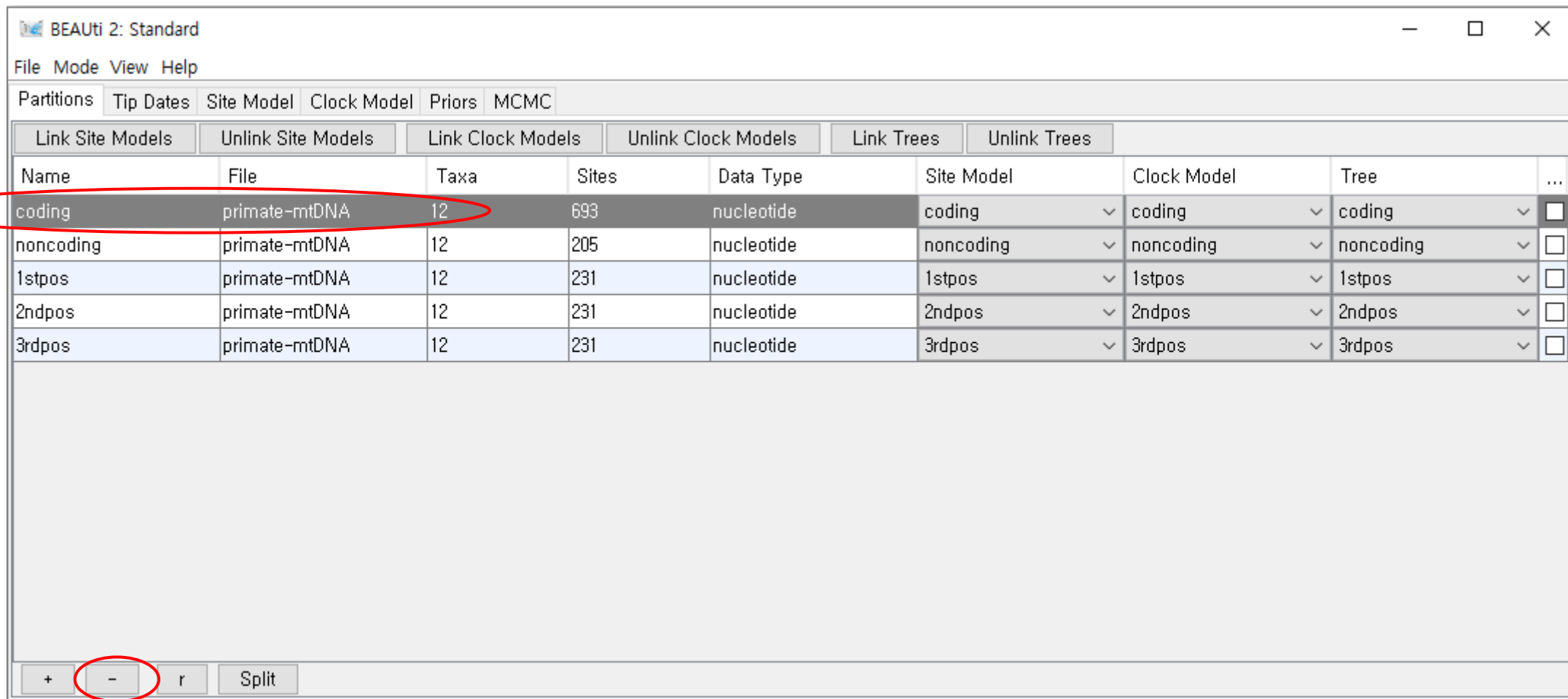
```
begin characters;
      dimensions nchar=898;
      format missing=? gap=- matchchar=. datatype=dna;
      options gapmode=missing;
      matrix

[
[
[
      1      2      3      ... ]
1234567890123456789012345678901234567... ]

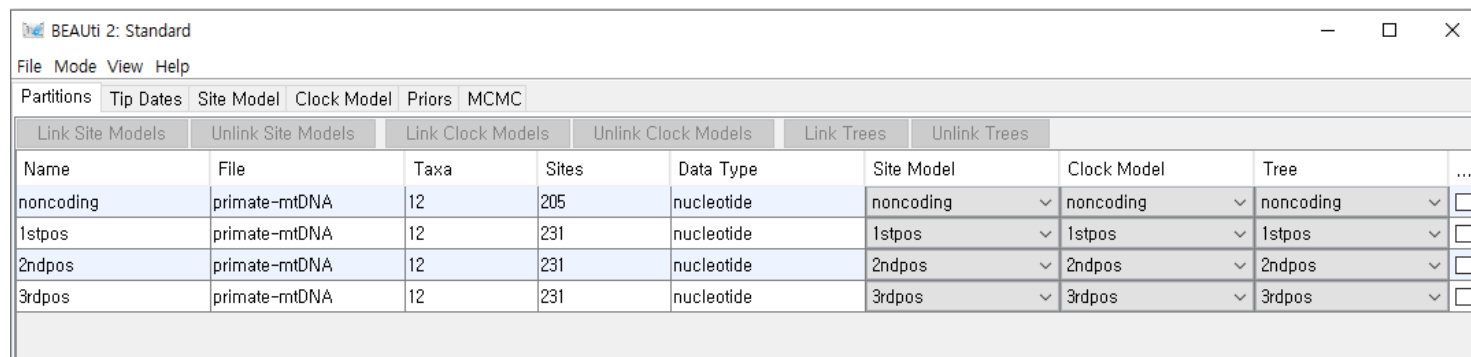
Lemur_catta      AAGCTTCATAGGAGCAACCATTCTAATAATCGCA...
Homo_sapiens      .....CC..C...GT.....C.....C...
Pan      .....CC..C....TT..C..C.....C...
Gorilla      .....CC..C...GTTG....T....T..C...
Pongo      .....CC..C.....CC..C..G..T..C...
Hylobates      .....T.C...T.....G.C..C.....C...
Macaca_fuscata      .....TTCC..C.....C..T..G.....T...
M._mulatta      .....TTCT..C.....C..C..G..T..T...
M._fascicularis      .....TCC..C.....CC..T.....C...
M._sylvanus      .....TCC..T....T..C..T..G..T..C...
Saimiri_sciureus      .....CC..C....TG..C.....T...
Tarsius_syrichta      ...T.....T....C....C...T.....T..C...
      ;
end;
```

(primate-mtDNA.nex 파일 내부. {begin assumptions; end;} 블록에서 데이터 파티션을 정의함. coding 파티션과 1stpos, 2ndpos, 3rdpos파티션이 중복되어 있기때문에 파일 읽어 들일때 경고 발생(뒤에서 해결할 예정)

```
begin assumptions;  
    charset coding = 2-457 660-896;  
    charset noncoding = 1 458-659 897-898;  
    charset 1stpos = 2-457\3 660-896\3;  
    charset 2ndpos = 3-457\3 661-896\3;  
    charset 3rdpos = 4-457\3 662-.\3;  
  
    exset coding = noncoding;  
    exset noncoding = coding;  
  
(생략)  
  
end;  
#      taxset hominoids = Homo_sapiens Pan Gorilla Pongo Hylobates;
```



(2) 첫째 행을 클릭하고 하단의 '-' 버튼을 눌러 삭제한다. 이로써 파티션 중복 문제 해결.



(3) 상단의 Link Clock Models, Link Trees 탭을 클릭하여 동일하게 세팅. 클릭하면 자동으로 세팅되어야 하나 안되는 경우도 많아(프로그램 버그인듯) 자동으로 설정되지 않으면 수동으로 설정.

BEAUi 2: Standard D:\Seo\class\beast_practice\run_20240124\beauti_setting_20240124.xml

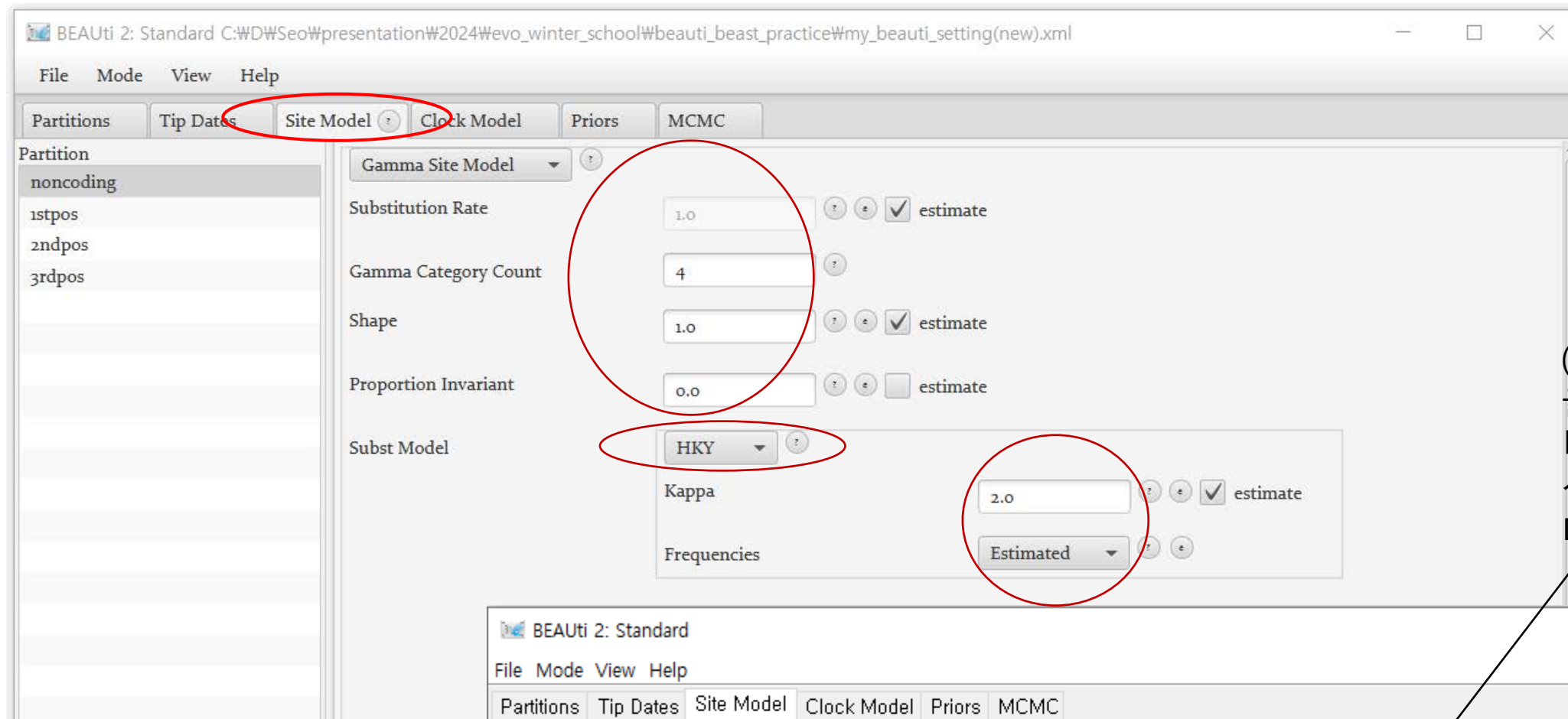
File Mode View Help

Partitions Tip Dates Site Model Clock Model Priors MCMC

Link Site Models Unlink Site Models **Link Clock Models** Unlink Clock Models **Link Trees** Unlink Trees

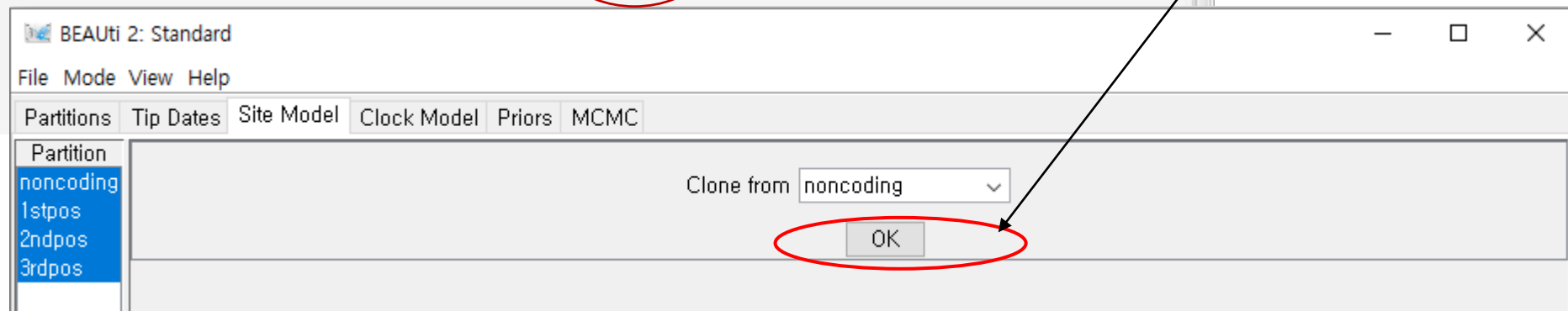
Name	File	Taxa	Sites	Data Type	Site Model	Clock Model	Tree	...
noncoding	primate-mtDNA	12	205	nucleotide	noncoding ▾	noncoding ▾	noncoding ▾	<input checked="" type="checkbox"/>
1stpos	primate-mtDNA	12	231	nucleotide	1stpos ▾	noncoding ▾	noncoding ▾	<input type="checkbox"/>
2ndpos	primate-mtDNA	12	231	nucleotide	2ndpos ▾	noncoding ▾	noncoding ▾	<input type="checkbox"/>
3rdpos	primate-mtDNA	12	231	nucleotide	3rdpos ▾	noncoding ▾	noncoding ▾	<input type="checkbox"/>

(4) 각 파티션 별로 Site Model 설정. DNA 치환 모형에 관해서는 서태건(2022) 참조.

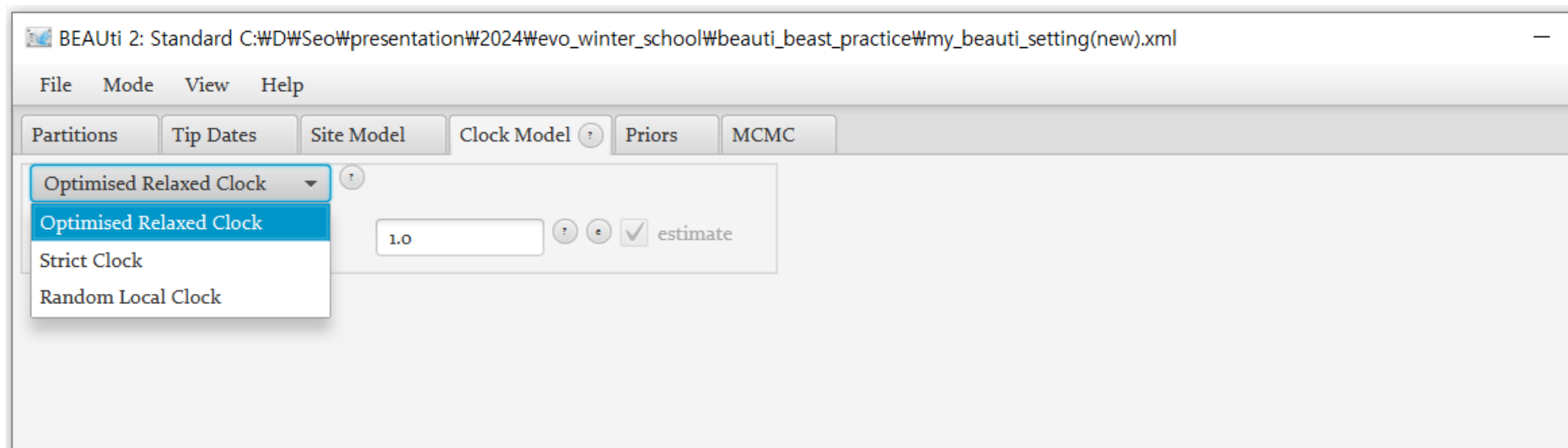


(Site Model 설정을 noncoding 파티션과 동일한 방식으로 한다는 의미)

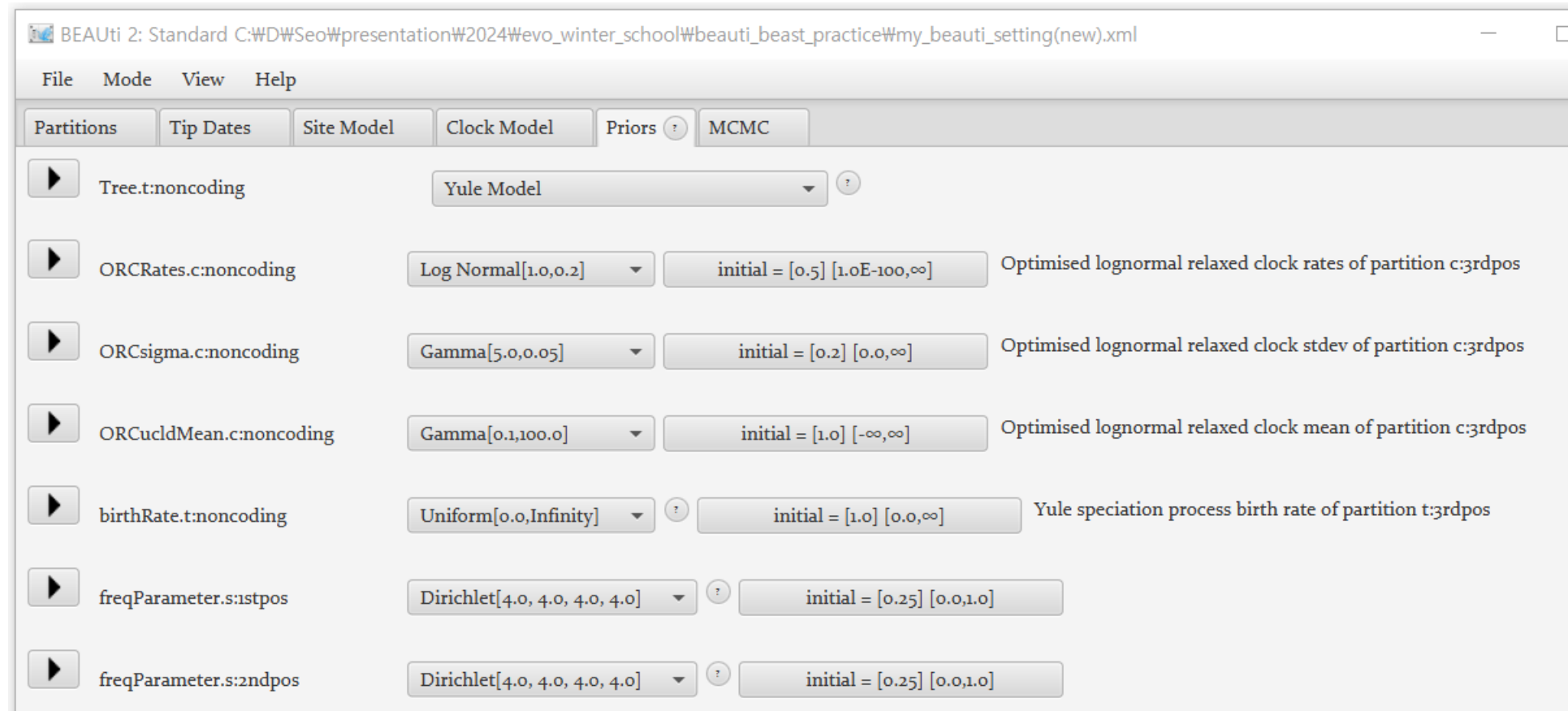
(4-1) Ctrl 누른상태로 파티션을 하나씩 클릭하여 모두 선택한후 OK 누름



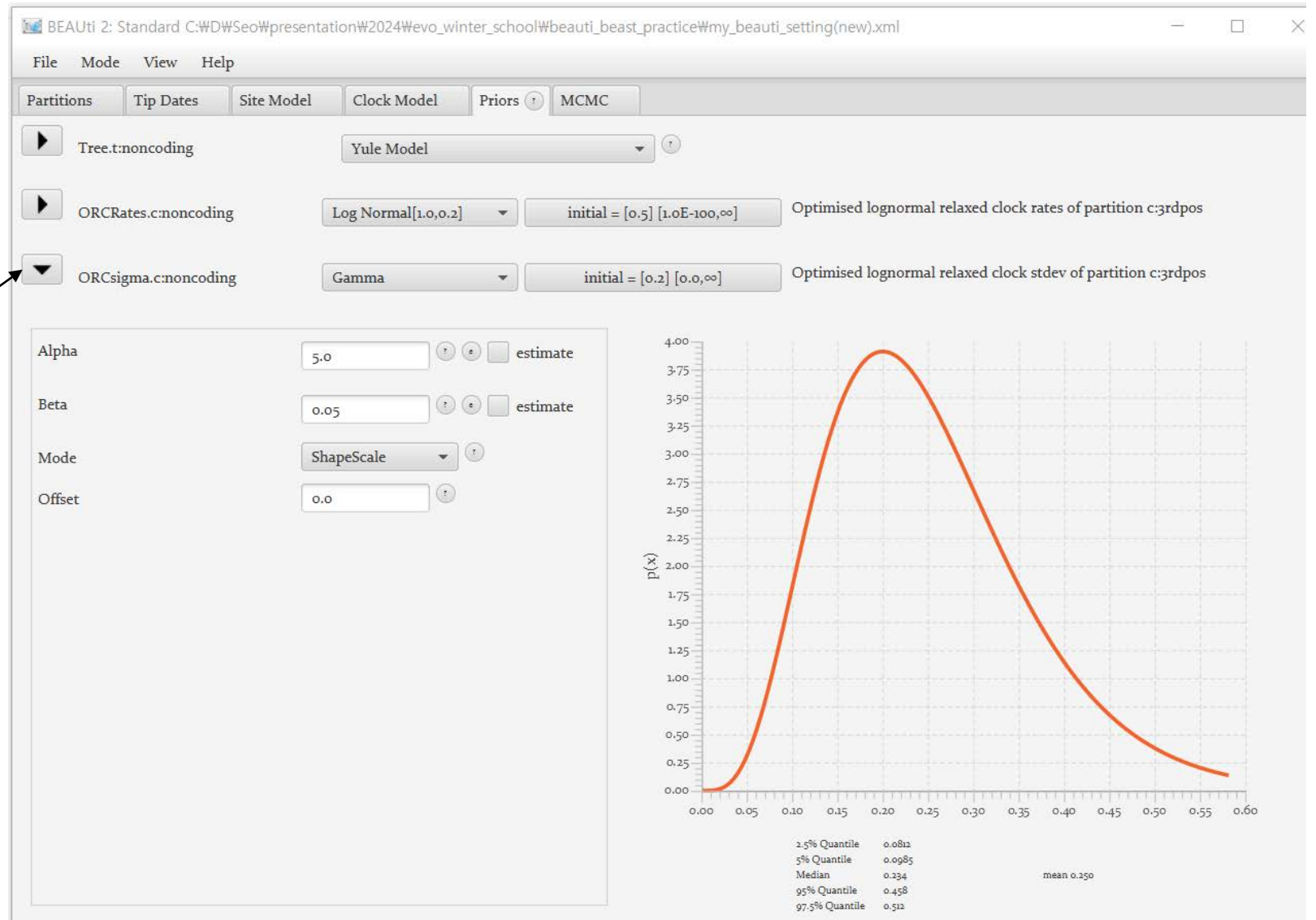
(5) Clock Model 탭을 클릭하여 진화속도 변화 모형을 설정한다. Optimised Relaxed Clock 으로 설정한다. 나머지 항목은 그대로.



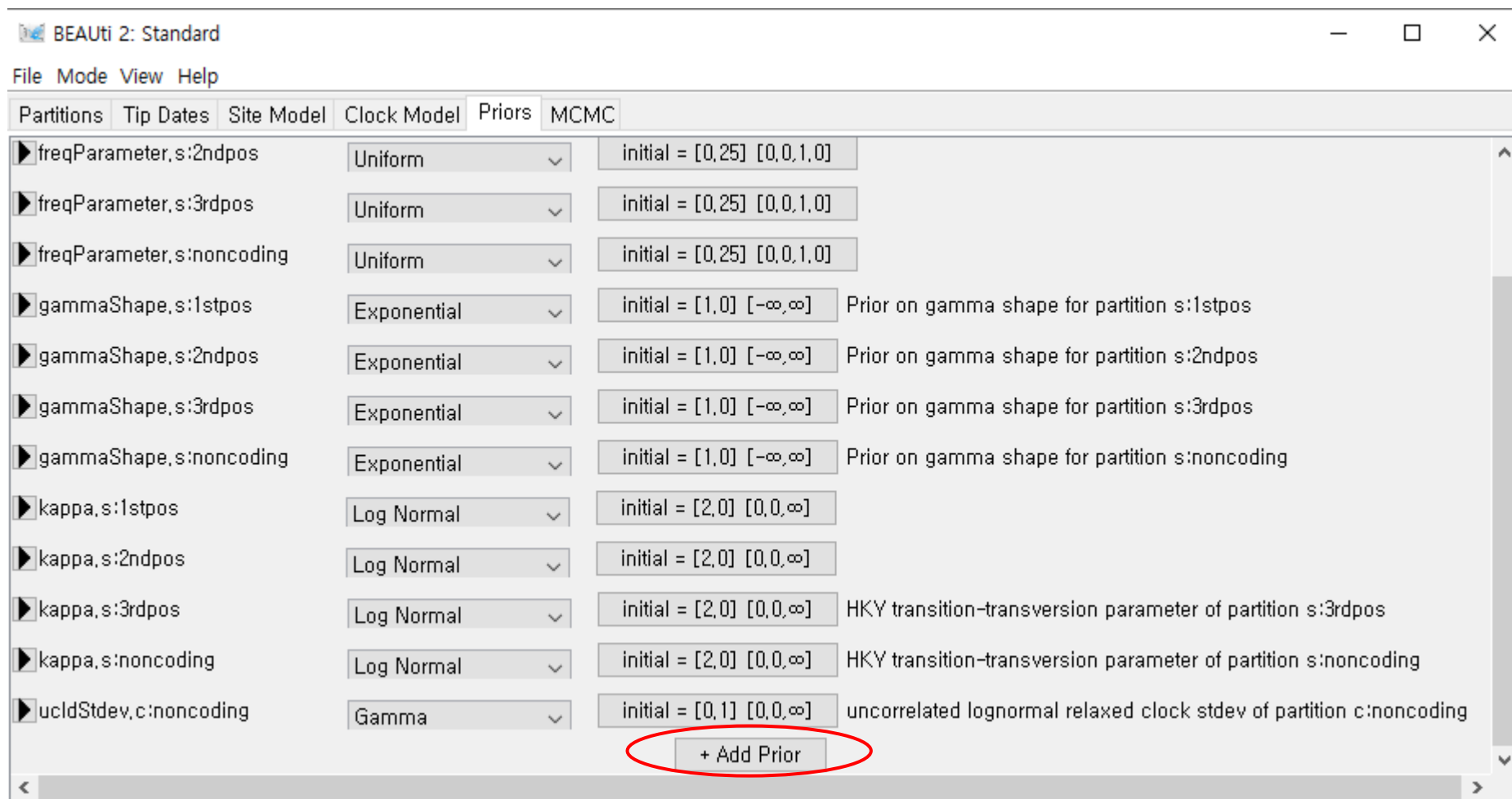
(6) Priors 탭을 눌러 각종 사전분포를 설정한다 (잘 모르면 디폴트 설정에 맡긴다)



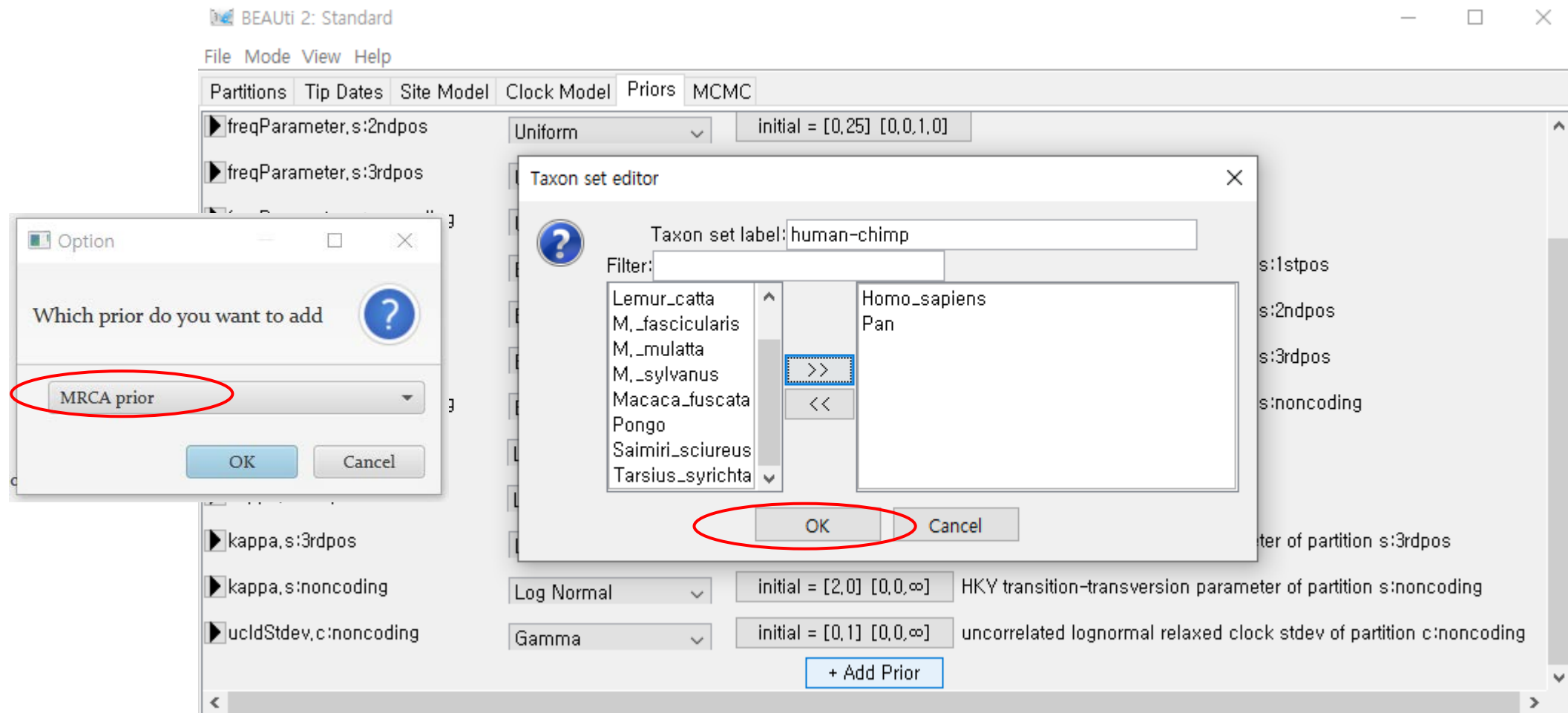
(화살표를 누르면 사전분포에 대한 자세한 설명 및 그래프 형태까지 볼 수 있다)



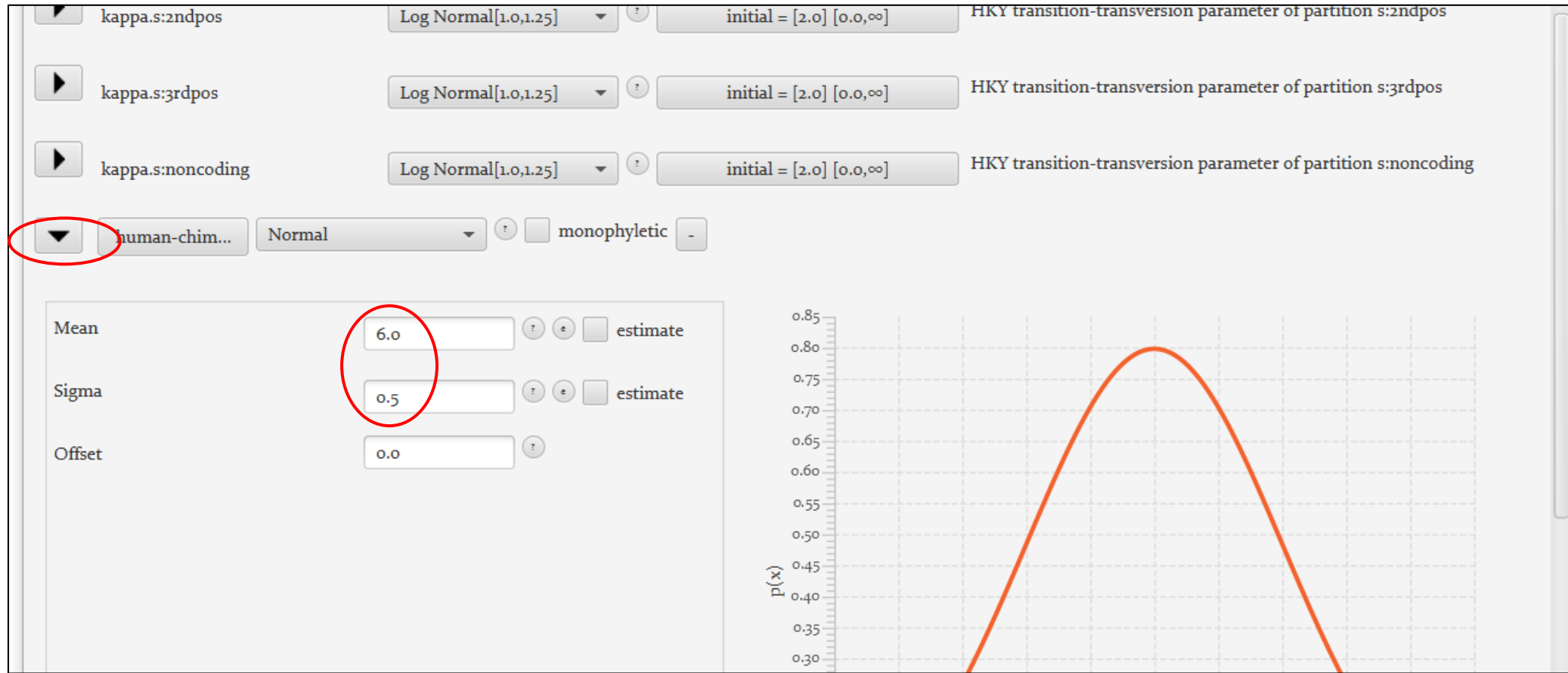
(7) 하단의 『+Add Prior』를 눌러 Human-Chimp 사이의 Time calibration 정보를 입력한다.



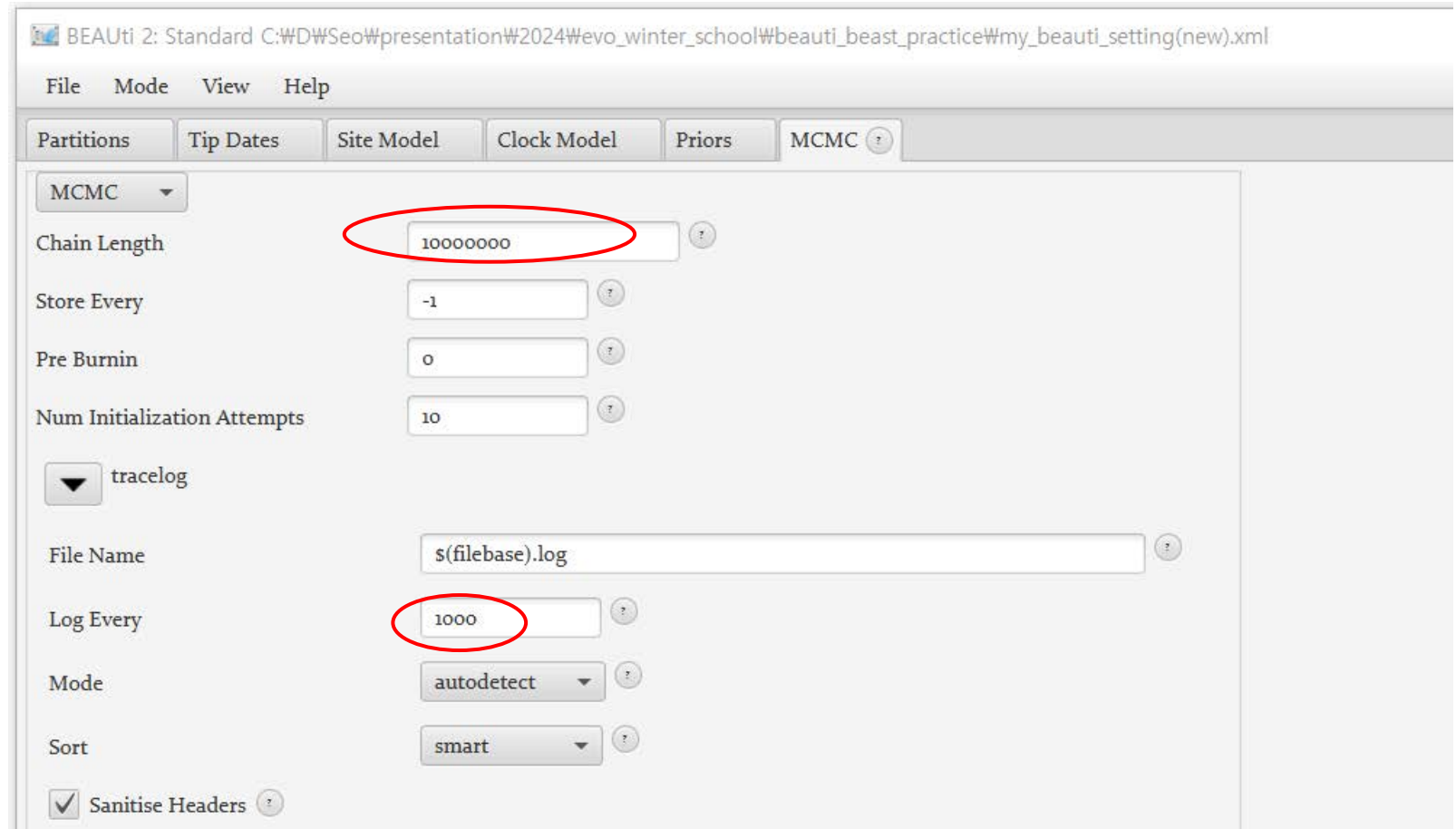
(7 계속) Human, Chimp 공동조상 노드 이름을 적당히 지정하고 해당 taxa들을 선택한다. OK 누른다.



(7 계속) Human-Chimp 노드의 시간에 대한 사전 분포를 입력한다. 일단은 Normal 분포로, 시간은 6.0 ± 0.5 MYA(Million Years Ago)로 설정한다. 우측의 그래프를 보면 이렇게 설정해도 분포의 꼬리가 현재시점 (time 0)에 닿지 않음을 알 수 있다.

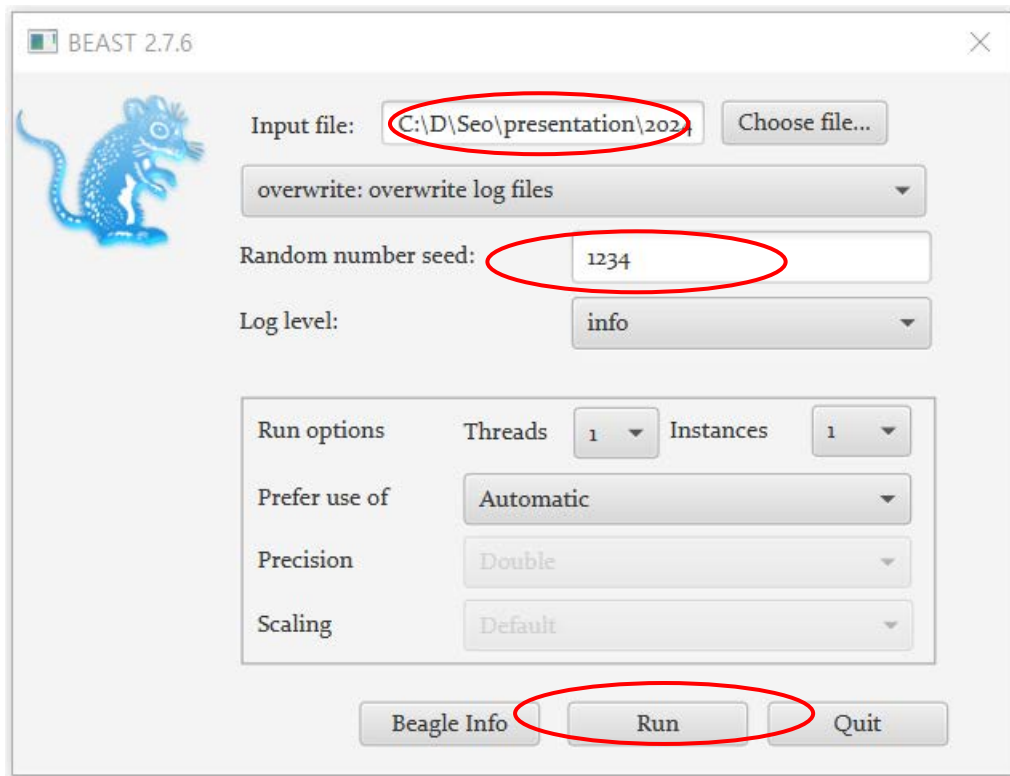


(8) MCMC 탭을 클릭하고
tracelog 앞에 있는 화살표를 누
른다. Total generation은 1000
만, sampling interval은 1000
generation 으로 설정되어 있음
을 알 수 있다. 일단은 이대로 실행한다. File/Save를 클릭하고 파
일 이름을 적당히
{my_beauti_setting(new).xml} 지
정하고 저장한다.



주의: (9)를 실행하기 전에 반드시 BEAUti 프로그램을 종료(파일을 저장하느냐 다시 물어 옴)하여 *.xml 파일을 저장한다 (단순히 save 버튼만으로는 저장이 잘 안되는 경우가 있음)

(9) Beast.exe를 실행하고 생성된 my_beauti_setting(new).xml 파일을 선택한 후 좌측과 같이 설정하고 Run 누른다. MCMC 실행이 종료되면 아래와 같은 내용이 화면에 표시된다.

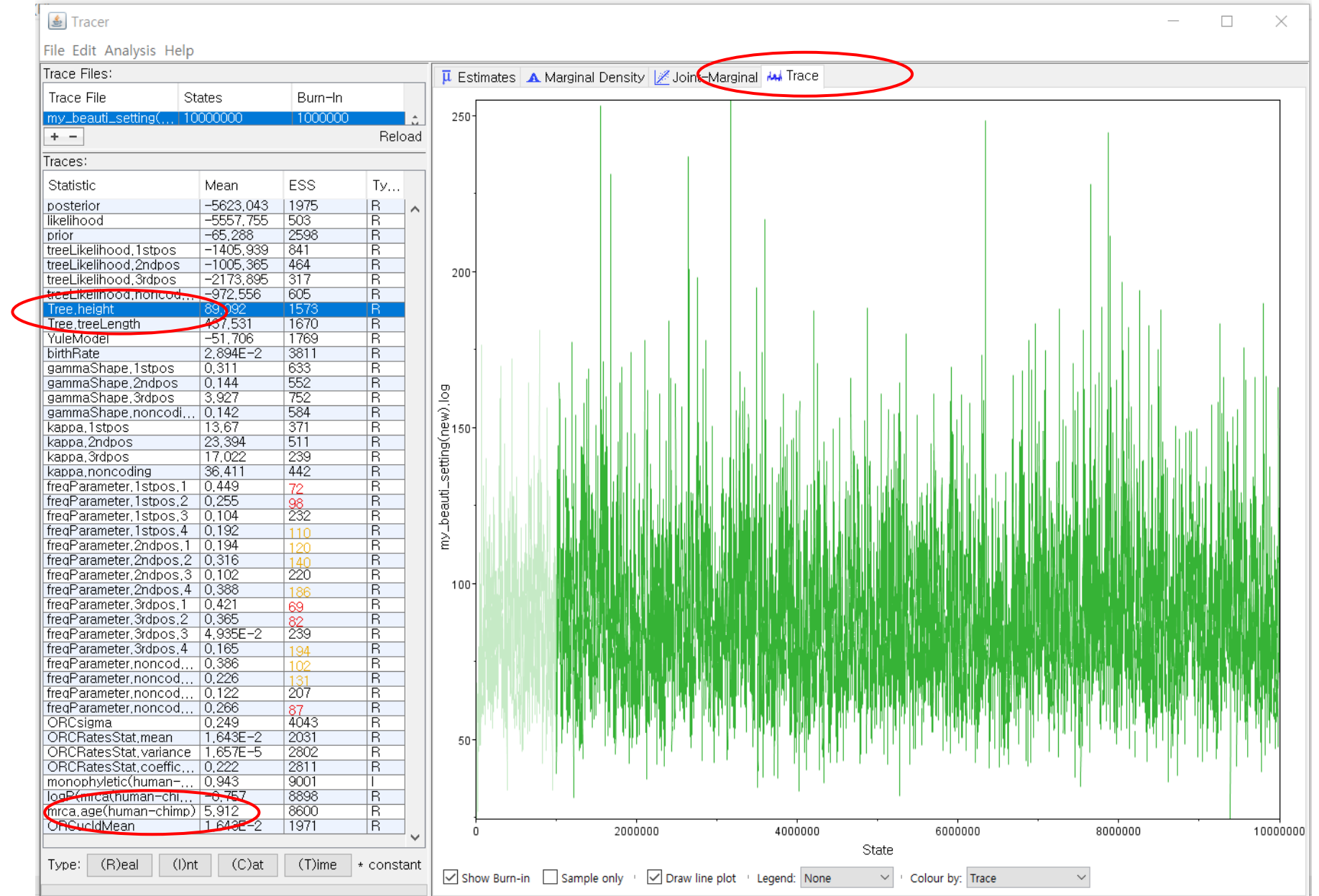


leOperatorSampler(gammaShapeScaler.s:3rdpos)	-	1105	3135	0.00042	0.26061
leOperatorSampler(KappaScaler.s:3rdpos)	-	1250	2913	0.00042	0.30026
leOperatorSampler(FrequenciesExchanger.s:3rdpos)	-	1058	3171	0.00042	0.25018
leOperatorSampler(ORCAdaptableOperatorSampler_sigma.c:noncoding)	-	78558	174380	0.02530	0.3105
leOperatorSampler(ORCAdaptableOperatorSampler_rates_root.c:noncoding)	-	61494	22670	0.00843	0.730
leOperatorSampler(ORCAdaptableOperatorSampler_rates_internal.c:noncoding)	-	582421	1103737	0.16863	0.34
leOperatorSampler(ORCAdaptableOperatorSampler_NER.c:noncoding)	-	10497	831606	0.08432	0.0124
erator(ORCuclMeanScaler.c:noncoding)		0.65794	54460	198676	0.02530 0.21514
se.inference.operator.kernel.BactrianUpDownOperator(ORC.UpDown.c:noncoding)		0.19547	70726	182868	0.02530

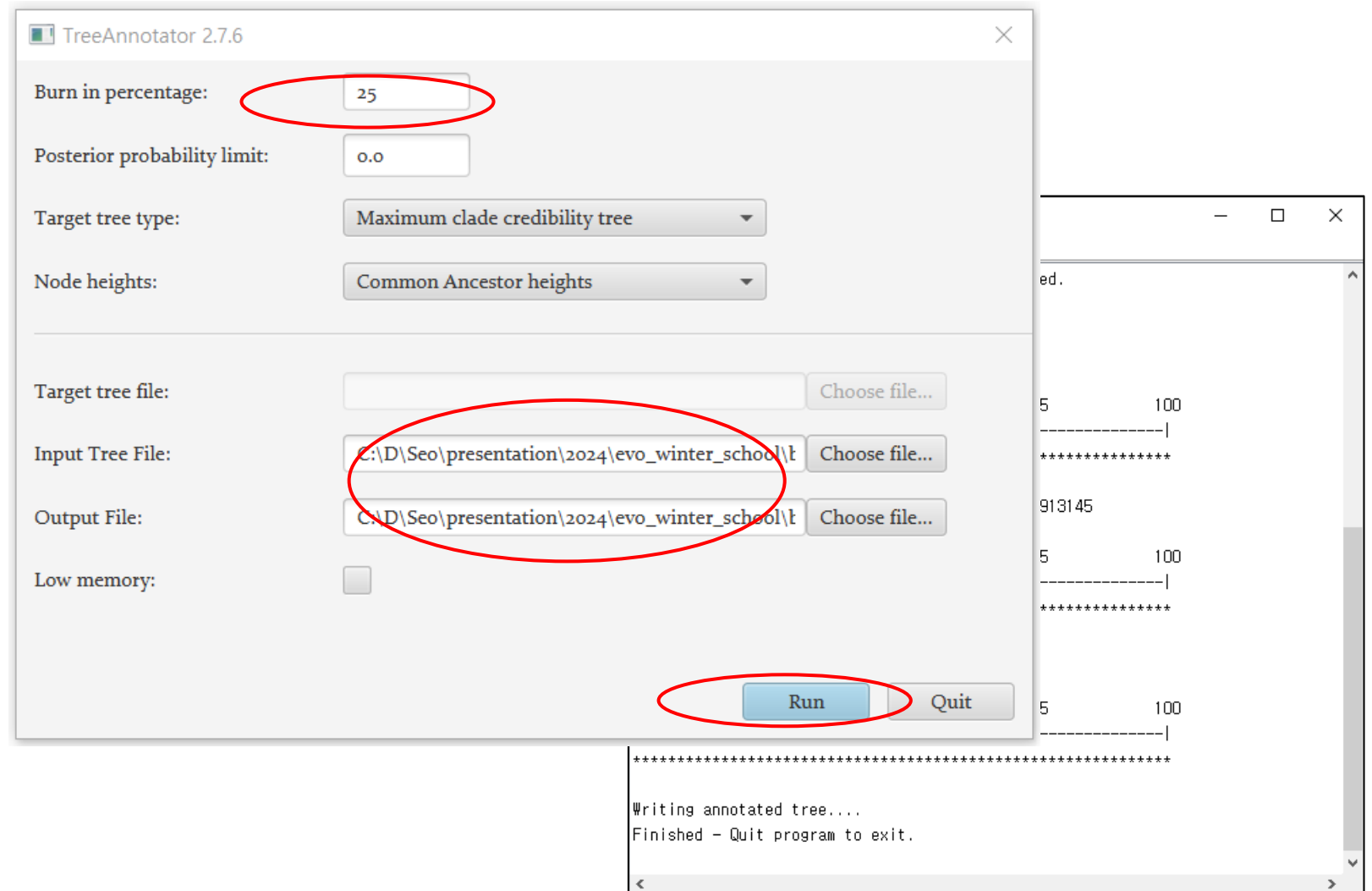
g: The value of the operator's tuning parameter, or '-' if the operator can't be optimized.
t: The total number of times a proposal by this operator has been accepted.
r: The total number of times a proposal by this operator has been rejected.
p: The probability this operator is chosen in a step of the MCMC (i.e. the normalized weight).
n: The acceptance probability (#accept as a fraction of the total proposals for this operator).

Total calculation time: 895.079 seconds
End likelihood: -5626.327742430513

(10) Tracer 프로그램 실행하여 my_beauti_setting(new)-noncoding.log 파일 읽어들인다 (File/Import Trace File에서 선택). TreeHeight는 공동조상의 분기연대를 의미한다 (MYA). 우측의 Trace 탭을 눌러 궤적을 확인한다. 상당수의 모수에서 ESS가 충분하지 않음을 알 수 있다. 좌측 하단에 Human-chimp 분기연대의 사후평균이 표시되어 있다.



(11) TreeAnnotator (BEAST 프로그램 폴더에 있음)를 실행하고
인풋으로
my_beauti_setting(new)-
noncoding.trees 파일을, 아웃풋
으로 적당한 이름
(tree.annotation.out(new).txt)을
지정하고 Run 누른다.



(12) TreeAnnotator 실행
결과로 생성된 파일
(tree.annotation.out(new)
.txt)을 FigTree로 읽어들
인다(File/Open 으로 선택
택).

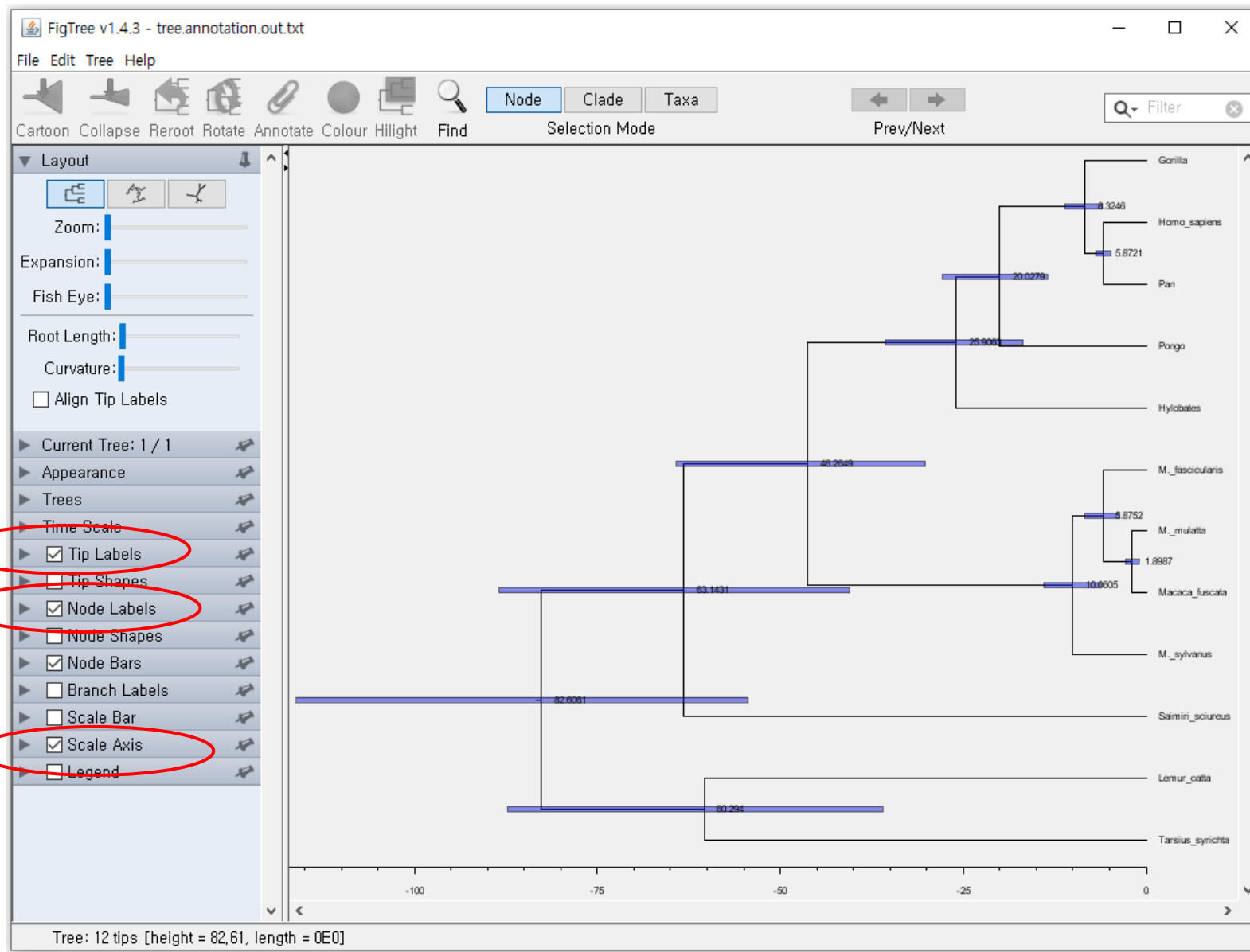
*** 분기연대 사후분포의 95%
CI (credibility interval) 를 보는
방법 ***

(a) Tip Labels 체크 (폰트사이
즈 조절 가능)

(b) Node Labels 체크/Display :
height 선택

(c) Node Bars 체크/Display:
Height_95%_HPD 선택

(d) Scale Axis 체크/Reverse
axis : 체크



(13) DensiTree (BEAST 프로그램
램 폴더에 있음)를 실행하고
my_beauti_setting(new)-
noncoding.trees 파일 읽어 들
인다(File/Load에서 선택).

우하 'Burn In' 클릭 25% 설정;
'Grid' 클릭 Full grid 선택;

