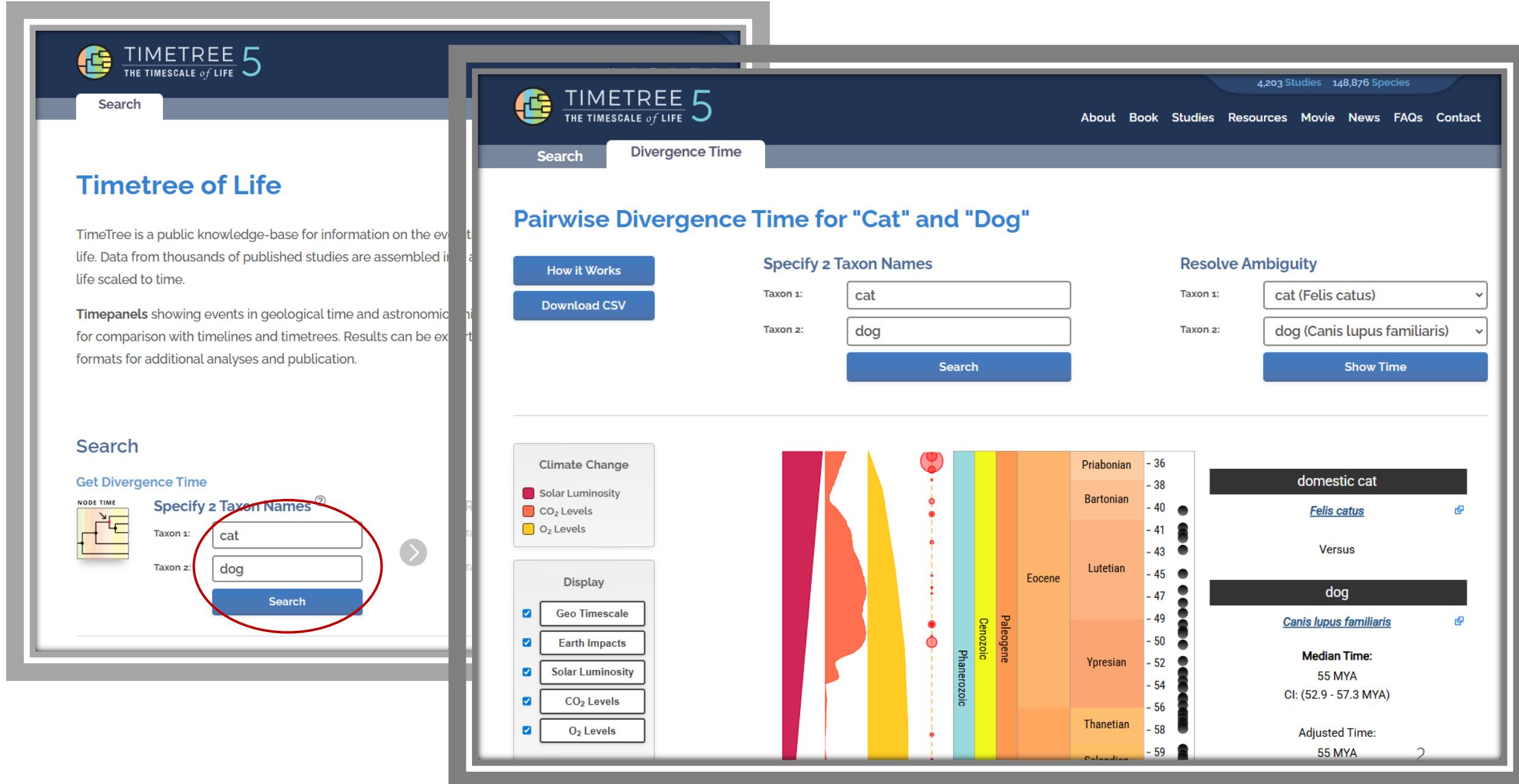


분기연대 베이즈 추정

(분기연대 베이즈 추정(서태건 2025)논문의 요약 및 추가 설명입니다 ; ver.2025.01.24)

< 분기연대 검색 사이트 > <https://timetree.org/>



The screenshot shows the homepage of the OneZoom tree of life explorer. At the top left is the logo 'ONE ZOOM'. To the right are four navigation links: 'For Everyone', 'For Education', 'For Science', and 'About us'. Below the navigation is a large title: 'Welcome to the OneZoom tree of life explorer...'. A descriptive paragraph explains the purpose of the tree: 'The Tree of Life shows how all life on Earth is related. In our interactive tree of life you can explore the relationships between 2,228,001 species and wonder at 105,508 images on a single zoomable page. Each leaf represents a different species, and the branches illustrate how these many species evolved from common ancestors over billions of years. Explore the relationships between species, learn about their common ancestors, or embark on a guided tour around the tree. Zoom around the tree to be amazed by the diversity of life on Earth.' The main feature is a zoomable tree of life diagram. A specific branch of the tree is highlighted, showing the evolutionary timeline and relationships for primates. Key nodes on this branch include: Cercopithecinae (8.52 Ma), Cercopithecidae (13.6 Ma), Colobinae (8.97 Ma), Catarrhini, Hominoidea (18 Ma), Hylobatidae (5.4 Ma), and Callitrichidae (18.8 Ma, 51 species). Small images of primates are shown at each node. Below the tree, a section titled 'Popular places to start exploring...' lists several categories with corresponding circular images: Birds, Butterflies, Fish, Insects, Spiders, Monkeys, Frogs, and Owls.

For Everyone For Education For Science About us

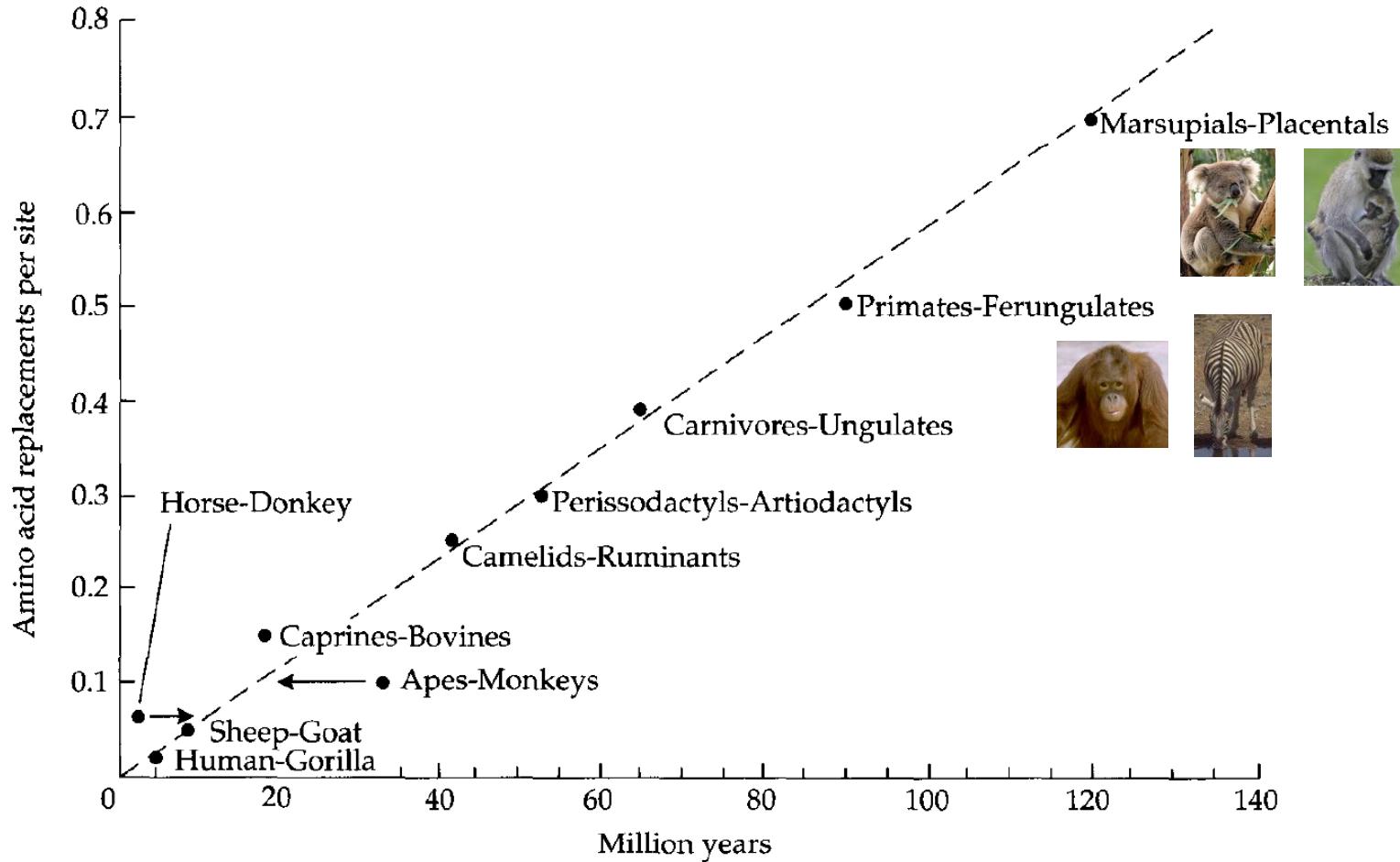
Welcome to the OneZoom tree of life explorer...

The Tree of Life shows how all life on Earth is related. In our interactive tree of life you can explore the relationships between 2,228,001 species and wonder at 105,508 images on a single zoomable page. Each leaf represents a different species, and the branches illustrate how these many species evolved from common ancestors over billions of years. Explore the relationships between species, learn about their common ancestors, or embark on a guided tour around the tree. Zoom around the tree to be amazed by the diversity of life on Earth.

Popular places to start exploring...

- Birds
- Butterflies
- Fish
- Insects
- Spiders
- Monkeys
- Frogs
- Owls

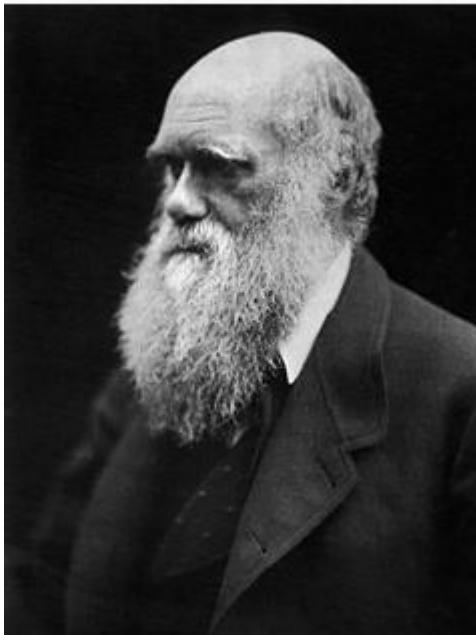
분자진화의 중립설 (neutral theory of molecular evolution) 과 분자시계 (molecular clock)



헤모글로빈 α , β , 시토크롬c, 피브리노펩타이드A 아미노산 치환 속도
(Langley and Fitch 1974; Graur & Li 2000)



라마르크(용불용설;
Theory of Use and
Disuse)
(1744~1829)



다윈 (종의기원)
(1809~1882)



기무라 (분자진화의
중립설)
(1924~1994)

돌연변이의 종류 및 고정(fixation)

(Ohta 2008)

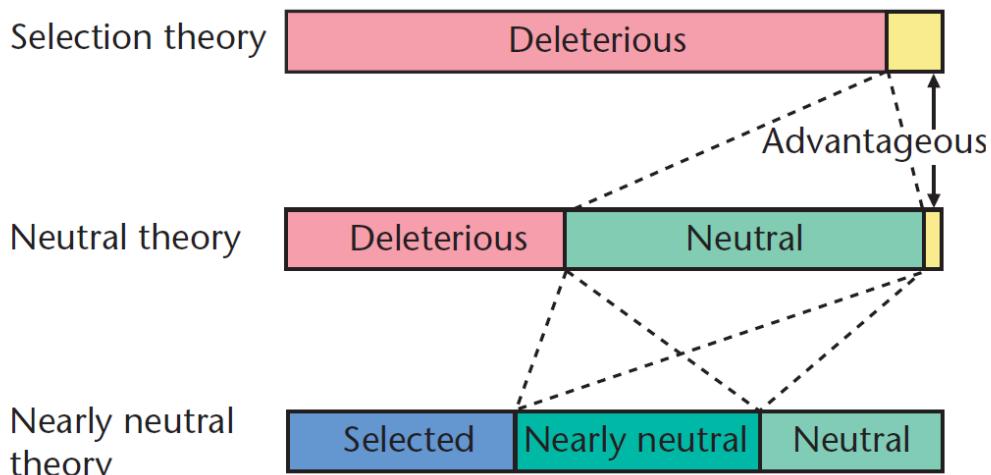


Figure 1 The classification of new mutants under the selection, neutral and nearly neutral theories. Note that while most selected mutants are deleterious, the group also includes advantageous mutants.

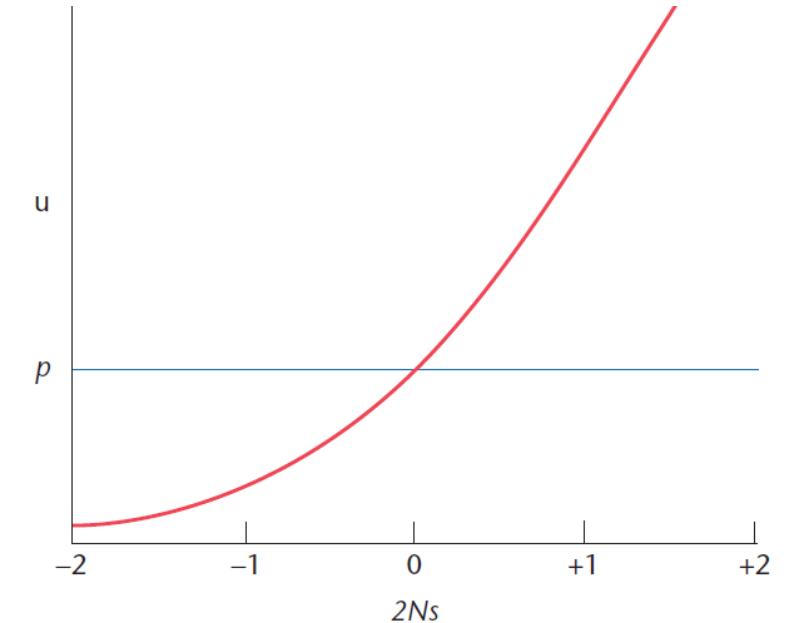


Figure 2 Fixation probability of a mutant in a finite population as a function of $2Ns$. p is the initial frequency of the mutant. The region of $2Ns < 0$ is that of slightly deleterious mutations.

Ohta's Nearly Neutral Theory

- N 이 크면 Ns 는 큰 음수 . 음의 선택이 강하게 작용 . 진화속도(단위 세대 기준) 느림
 - N 이 작으면 Ns 는 작은 음수. Genetic drift 가 강하게 작용 . 진화속도가 빠름
 - N 이 크면 일반적으로 세대 길이가 짧음
 - N 이 작으면 일반적으로 세대 길이가 깊
- } 음의 상관관계
- } 음의 상관관계
- } 음의 소멸
-

그러므로 chronological time
단위(년, 백만년)로 molecular
clock이 성립

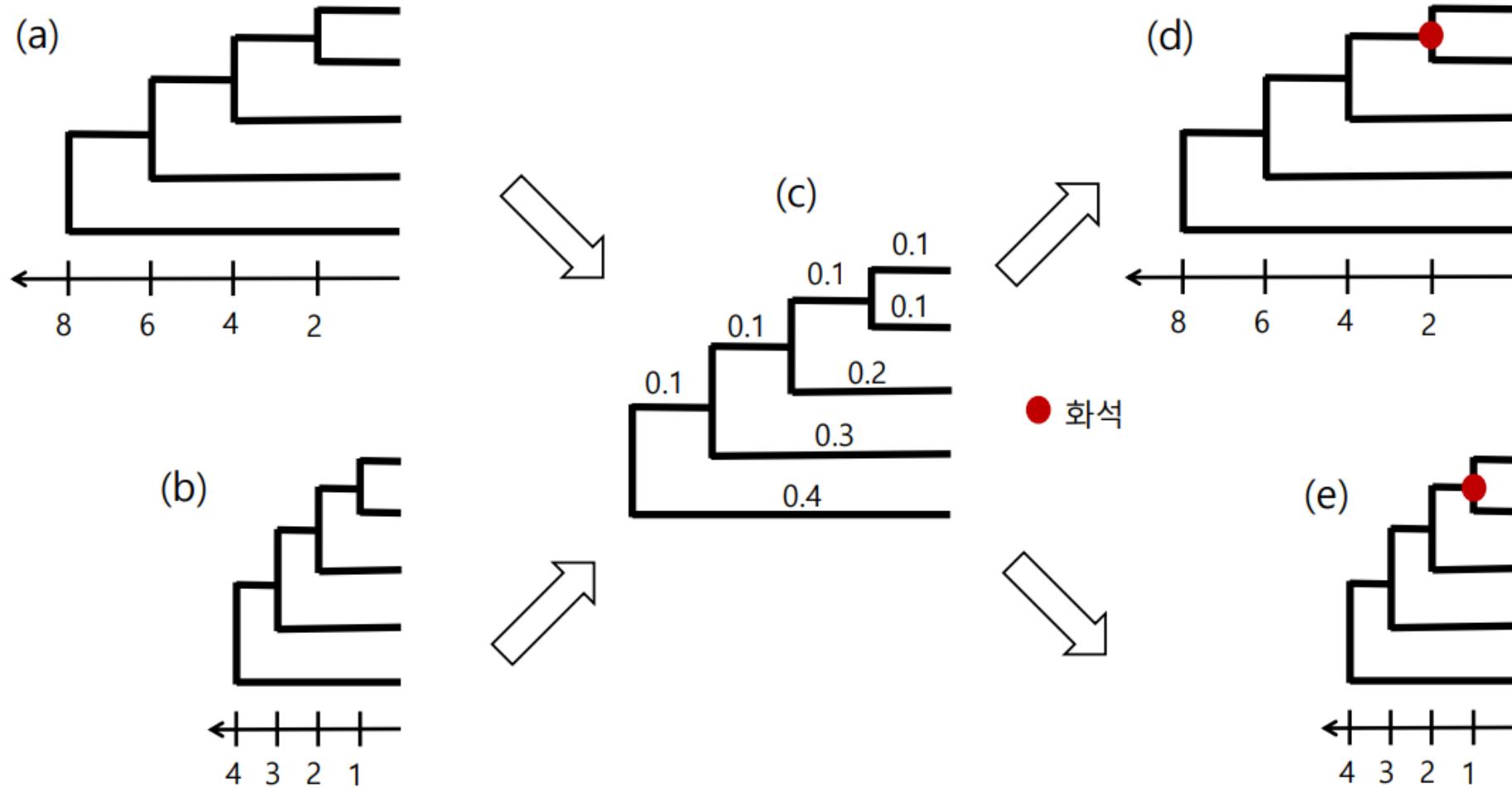
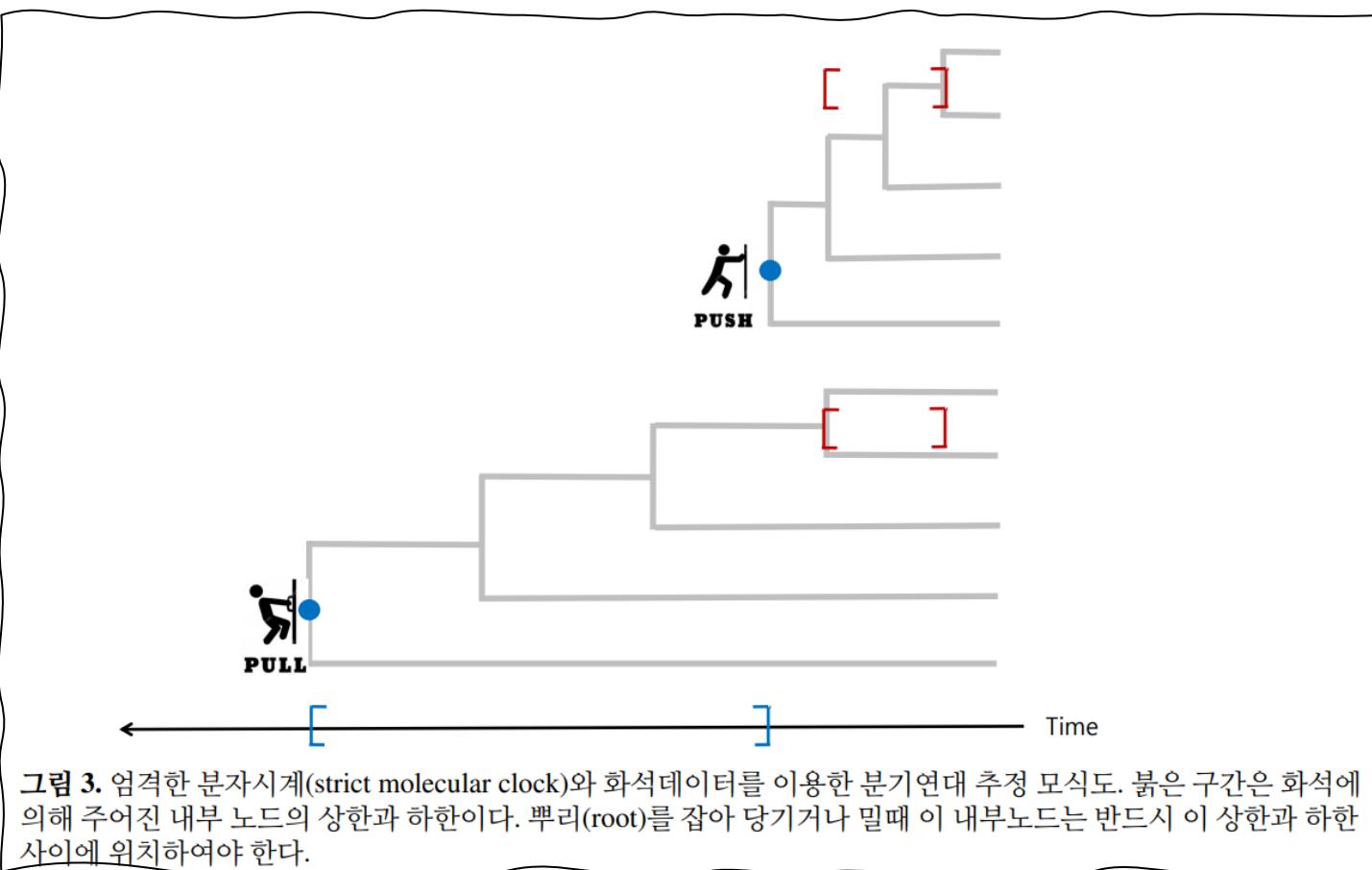
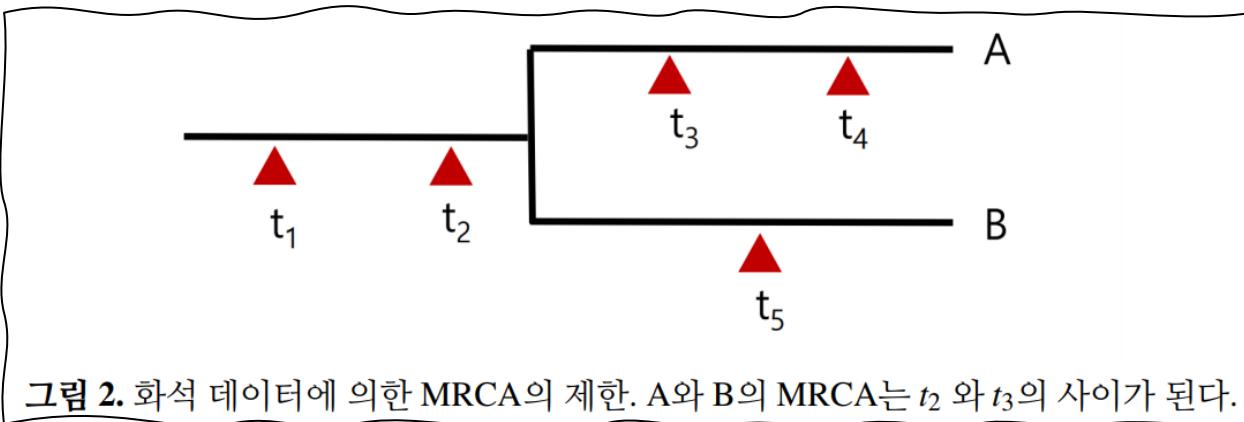


그림 1. Molecular clock과 화석데이터를 이용한 분기연대 추정 예시



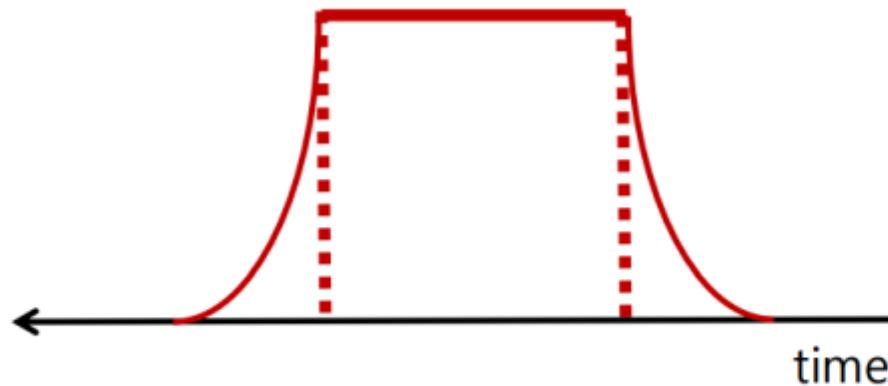
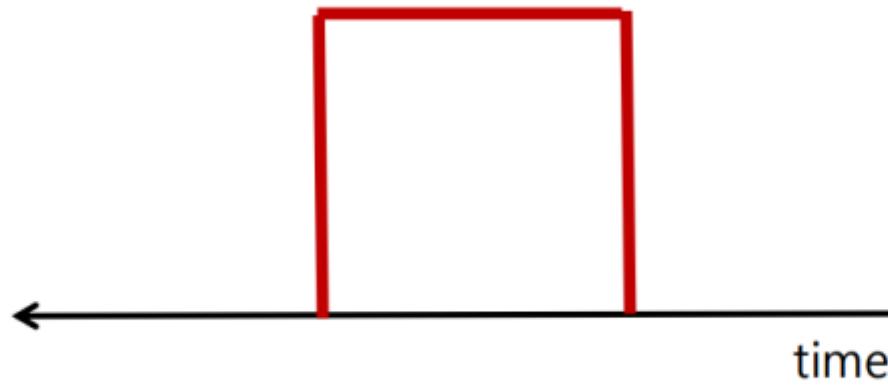


그림 4. Hard bound (위)와 soft bound (아래)에 의한 분기연대 분포 모식도

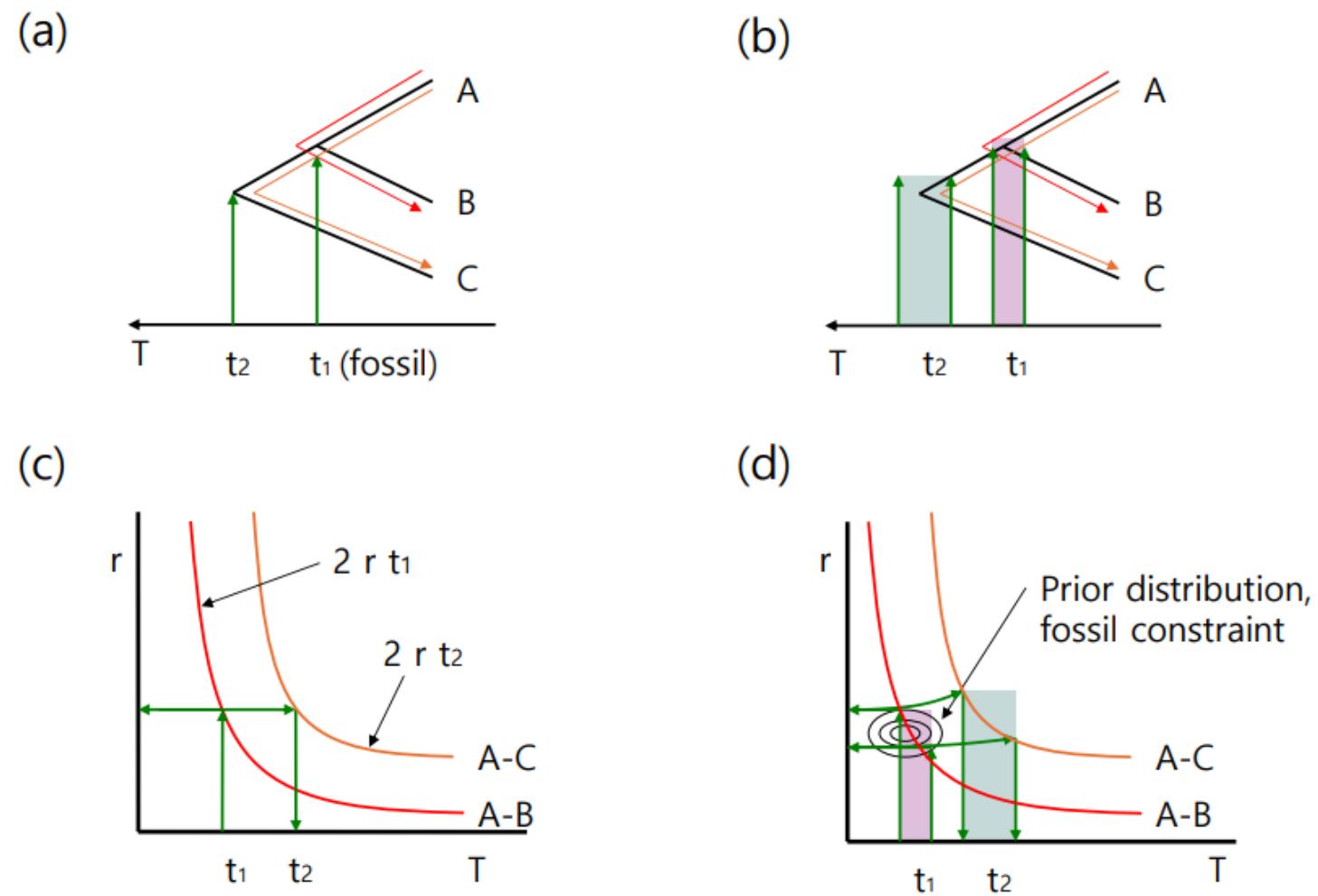
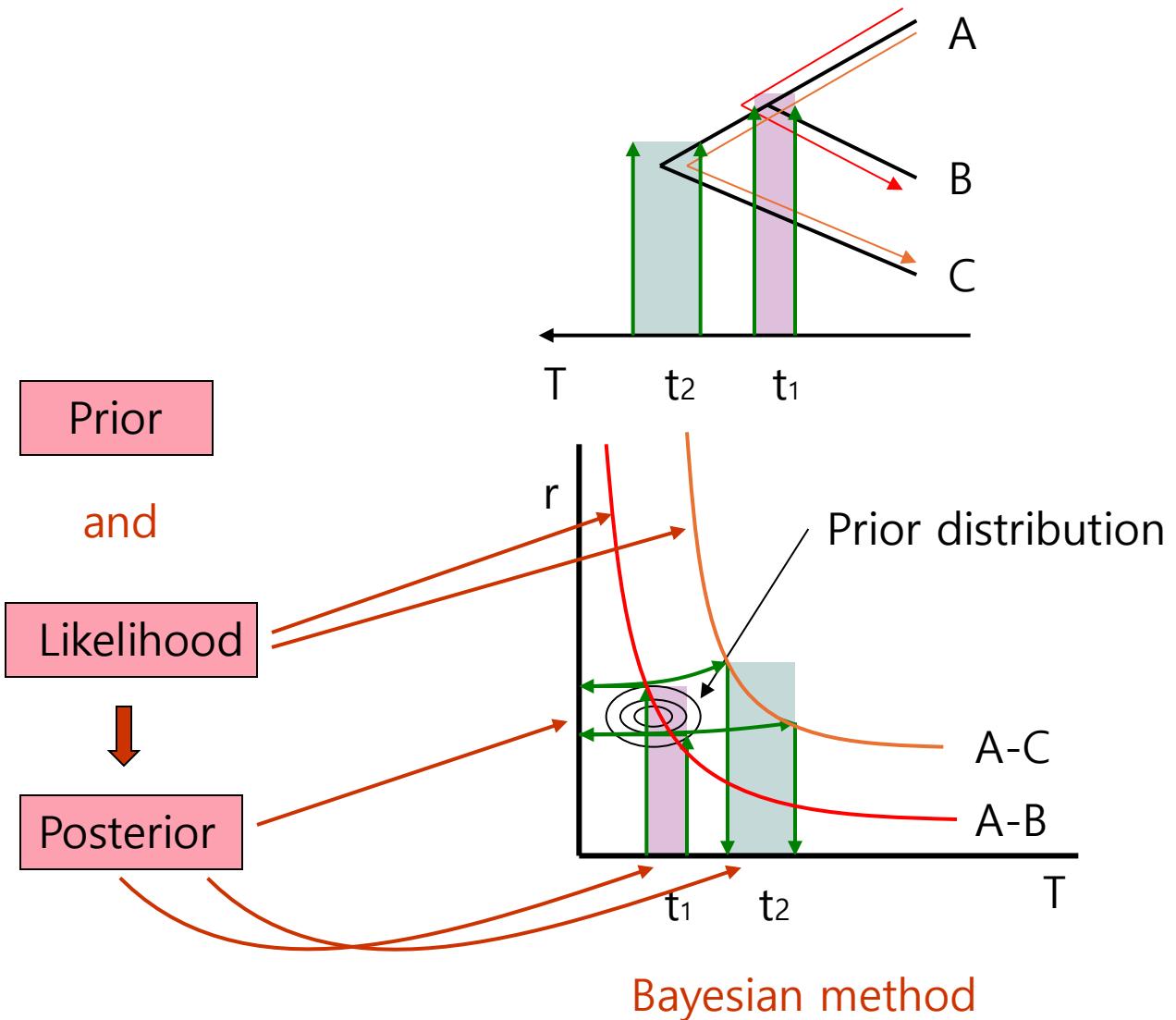


그림 5. 엄격한 분자시계와 베이즈 방법에 의한 분기연대 추정 모식도



화석데이터 \mathbf{C} 와 염기서열 데이터 \mathbf{X} 가 주어졌을 때, 분기연대 \mathbf{T} , 진화속도 \mathbf{r} 그리고 각종 모형의 모수¹⁴ $\boldsymbol{\theta}$ 의 사후 분포를 베이즈 추정법으로 구하는 것이 분석의 목적이다. 베이즈 정리를 이용하면 $(\mathbf{T}, \mathbf{r}, \boldsymbol{\theta})$ 의 사후분포는 다음과 같이 구할 수 있다.

$$\begin{aligned}
 p(\mathbf{T}, \mathbf{r}, \boldsymbol{\theta} | \mathbf{X}, \mathbf{C}) &= \frac{p(\mathbf{T}, \mathbf{r}, \boldsymbol{\theta}, \mathbf{X}, \mathbf{C})}{p(\mathbf{X}, \mathbf{C})} \\
 &= \frac{p(\mathbf{T}, \mathbf{r}, \boldsymbol{\theta}, \mathbf{X} | \mathbf{C}) p(\mathbf{C})}{p(\mathbf{X} | \mathbf{C}) p(\mathbf{C})} \\
 &= \frac{p(\mathbf{X} | \mathbf{T}, \mathbf{r}, \boldsymbol{\theta}, \mathbf{C}) p(\boldsymbol{\theta}) p(\mathbf{r} | \mathbf{T}, \mathbf{C}) p(\mathbf{T} | \mathbf{C})}{p(\mathbf{X} | \mathbf{C})} \\
 &\propto p(\mathbf{X} | \mathbf{T}, \mathbf{r}, \boldsymbol{\theta}) p(\boldsymbol{\theta}) p(\mathbf{r} | \mathbf{T}, \mathbf{C}) p(\mathbf{T} | \mathbf{C})
 \end{aligned} \tag{1}$$

여기에서, $p(\mathbf{X} | \mathbf{T}, \mathbf{r}, \boldsymbol{\theta})$ 는 염기서열 정보만을 이용하는 가능도 함수, $p(\boldsymbol{\theta})$ 는 각종 모수의 사전확률 밀도함수, $p(\mathbf{r} | \mathbf{T}, \mathbf{C})$ 는 분기연대와 화석데이터가 주어졌을 때 진화속도의 변화양상을 설명하는 확률밀도 함수, $p(\mathbf{T} | \mathbf{C})$ 화석데이터에 의해 제한이 가해지는 분기연대의 사전확률밀도 함수를 나타낸다. 식 (1)의 여러 항 중에 $p(\mathbf{r} | \mathbf{T}, \mathbf{C})$ 에 특히 주목할 필요가 있다.

(1) Independent rates 모형

Independent rates 모형은 계통수의 각 가지가 갖는 진화 속도는 어떤 확률분포로부터 독립적으로 얻어진 랜덤 샘플이라는 가정을 한다(그림 6). 즉, 어떤 진화속도를 갖고 진화하던 종이 종분화 했을 때, 두 자손 종의 진화속도는 완전히 독립적인 진화속도를 가지며 확률분포에 의해 결정된다. 이후 각각 새로운 종분화가 발생하기 전까지는 주어진 일정한 진화속도로 진화한다. 진화속도를 결정하는 분포로서 흔히 로그정규분포(lognormal distribution) 혹은 감마분포를 가정한다.¹⁵

로그정규분포를 이용한 진화속도 변화 모형에서는 진화 속도에 로그를 취한 값이 아래와 같이 평균 μ , 분산 σ^2 인 정규분포를 따른다고 가정한다.

$$\log r \sim N(\mu, \sigma^2) \quad (2)$$

이 분포의 확률밀도함수는

$$f(r|\mu, \sigma^2) = \frac{1}{r\sigma\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left(\frac{\log r - \mu}{\sigma} \right)^2 \right\}$$

이며 평균과 분산은

$$E[R] = \exp \{ \mu + \sigma^2/2 \} \quad (3)$$

$$\text{Var}[R] = \{ \exp(\sigma^2) - 1 \} E[R]^2 \quad (4)$$

가 됨이 알려져 있다(Johnson et al. 1994).

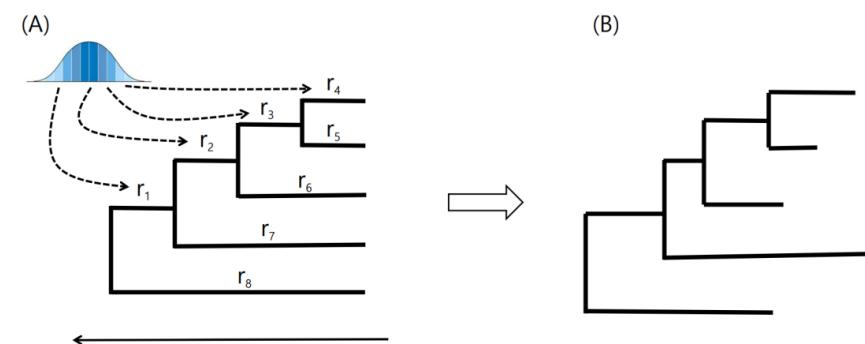


그림 6. Independent rates 모형

감마분포(김우철 2021)를 independent rates 모형화에 사용하기도 한다. 감마분포의 확률밀도함수는 모수 α, β 가 주어졌을 때 다음과 같고

$$f(r|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} r^{\alpha-1} e^{-\beta r}$$

평균과 분산은 다음과 같다.

$$\text{E}[R] = \frac{\alpha}{\beta} \tag{5}$$

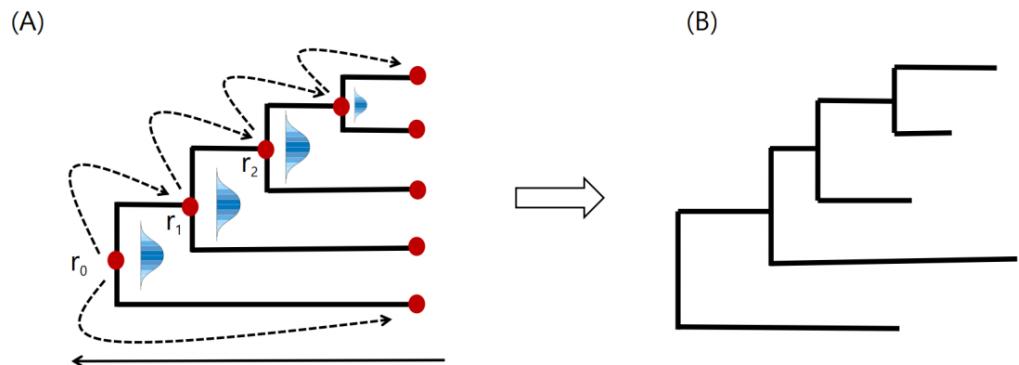
$$\text{Var}[R] = \frac{\alpha}{\beta^2} \tag{6}$$

(2) Correlated rates 모형

Correlated rates 모형¹⁶은 진화속도가 연속적으로 변하는 상황을 상정한 모형이다. 계통수의 내부노드에 진화속도가 할당되며 이들끼리 상관관계를 유지하며 변화한다고 가정하는 모형이다(그림 7). 분산이 시간 간격에 의존하는 상관관계를 표현하기 위해 식 (2)를 변형시켜 아래와 같이 모형화 한다.

$$\log r_{i+1} \sim N(\log r_i, t_i \sigma^2) \quad (7)$$

즉, $\{i+1\}$ 번째 노드의 로그-진화속도의 평균은 직전 조상인 i 번째 노드의 로그-진화속도와 같고 분산은 시간 간격 t_i 에 비례한다. 즉, 시간 간격 t_i 가 매우 작을 경우 분산이 매우 작아져 $\log r_{i+1}$ 은 $\log r_i$ 과 매우 유사한 값을 갖게 되며 시간 간격이 매우 클 경우 크게 다른 값을 가질 수 있다. 이는 진화속도의 독립성을 가정하는 식 (2)과 대조되는 큰 차이이다. Root에서의 로그-진화 속도는 $\log r_0 = \mu$ 로 가정한다.



실습

대용량의 염기서열 데이터를 MCMCTree 프로그램으로 분석할 때 보통 두 단계로 진행한다. 첫번째 단계는 가능도 함수를 다변량 정규분포 확률밀도 함수로 근사하는 과정이다. 염기서열 데이터의 규모가 클 경우 가능도 함수의 계산에 많은 시간이 소요되는데 가능도 함수를 다변량 정규분포로 근사하면 (Thorne et al. 1998; Seo et al. 2004) 계산 시간을 단축 시킬 수 있다. 두번째 단계는 본격적으로 분기연대와 각종 모수의 사후분포를 베이즈 방법으로 구하는 과정이다. 이 과정에서 MCMC 알고리즘이 사용된다.

```

seed = 1
seqfile = mtCDNApri123.txt
treefile = mtCDNApri.trees
outfile = out

ndata = 3
seqtype = 0 * 0: nucleotides; 1:codons; 2:AAs
usedata = 3 * 0: no data; 1:seq like; 2:use in.BV; 3: out.BV
clock = 3 * 1: global clock; 2: independent rates; 3: correlated rates
RootAge = <1.0 * safe constraint on root age, used if no fossil for root.

model = 4 * 0:JC69, 1:K80, 2:F81, 3:F84, 4:HKY85
alpha = 0.5 * alpha for gamma rates at sites
ncatG = 5 * No. categories in discrete gamma

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

BDparas = 1 1 0 * birth, death, sampling
kappa_gamma = 6 2 * gamma prior for kappa
alpha_gamma = 1 1 * gamma prior for alpha

```

그림 10. mcmctree.1.ctl.txt은 분석의 첫번째 단계를 실행하기 위한 컨트롤 파일이다. 붉은색 상자가 수정부분이고 수정이 필요없는 하단 부분은 표시를 생략하였다. 실행의 결과 out.BV 파일이 생성된다. 이를 복사하여 in.BV 파일을 만들고 두번째 단계의 분석에 사용한다.

C:\temp>mcmctree mcmctree.1.ctl.txt



그림 11. MCMCTree 프로그램 실행의 첫번째 단계. out.BV 파일이 생성된다.

```

7 3331
human AGACTGTTAGCGGCCGGCAACTCCCTACTCAAACCAACCTCAAGCATTACGTTAATAGAAA
TAACCACGCCAAACAATCTCGGACGAATCGCCACCGCAGCCTGAACATAATGGTTGGTTCTACGCACCCGGCTCCAGAAC
AAAACCTTTAAAAGAGAACGAGCTTGTGGACGACCATGCCCATCACGCATAATCATCTCAGACCCACTATTAAAGGATGGTAC
AGTACGTGTTACCTATTGCCAGCCCCGGGGATGCTGCCCTGAGGCACGTGCCCTAAAGGGGTTCACTCCCGAAC
chimpanzee AGACTGTTGGCAACGTGGCAACTCCCCGCCAAACCAACCTCACGCATTACGTTAATAAAAA
TTAACCCACGTCAAATAATCTCGGACGGATGCCACCGCAGCCTGAACATAATGGTTGGTTCTACGCACCCGGCTCCAGAAC
AAAACCTTTAAAAGAGAACGAGCTTGTGGACGACCATGCCCATCATGCATAATCATTTCAGACCCACTATTAAAGGATGGCA
GAGTGCACGGTCCCTATTGCCAAACCGTCCGGGGATGCCGCCCTGAGGCACGTGCCCTAAAGGGGATCCACTCCGGAAC
..... (이하 생략)

7 3331
human AAATCCTTCCCTTGTCCCTTTTCCCTTCATTAAGTCCAAGTTAGGCCTTTCTTTCCAT
TCTTTGTTCACTCTAAGGATCGAGCAAGAACCTAAGTGAGTTTCCATTTCGTTGCTAACGTCCTCATGGAGCCCGTCCT
CCAAACCTTTCTCTATGTCATCAGCCTCGTTTCAATCCCTATGTTCTTCTTCGGGGGTAAC
GTTTCTCGTTACACGACAATCCCTGTTCCGACCATGTACGTCCTAGCCCTCCTAACGCTTCGTTGACTCAAGCTT
chimpanzee AAATTCCCTCCCTTGTCCCTTTTCCATTAAAGTCCAAGTTAGGCCTTTCTTTCCC
ACCCCTCTGTTCACTCTAAGGATTGAGCAAGAACCTAAGTGAGTTTCCATTTCGTTGCTAACGTCCTCATGGAGCCCGTC
CACCAAACCTTTCTCTATGTCATCAGCCTCGTTTCAATCCCTATATTCTTCTTCGGGGGTA
ATGTTCTCGTTACACGACAATTCCCTGTTCCGACCATGTACGTCCTAGCCCTCCTAACGCTTCGTTGACTCAAAC
..... (이하 생략)

7 3331
human CATGCTACTCCACACACCAAGCTATCTAGCCTCCCAATCCAAAACAAACATTAAACACTTCA
AGCCAACGACCAACACAAGCTAAAACACACCCAAATCAACTACGACATTCAATTCAACTCCTCAAGCGCAACCCGA
TCTACAAAACAGTAAAAAAACCTCCACAGTCACCTGACCGTTCAAAACCCACACACTCAACACTCCTCACACCAACTATAACAC
TCCCCAAATCACCCAAACAAAACCAAAACAAACATCCCACCAACCCACACACCAACTATAACCCACCCCCCAATCAAC
chimpanzee CATACTACTCCACACACCAAAACTACCTAGCCTCCCAATCCAAAATAACATCAAACACTTCA
CCAAACAAACAATCAGCATAAAACTAAACACACCCAAACCAATCATGATTTCAATTCTAATTATCCCCACTCGAAACGCAATT
CTTCTACAAAATAAGAAAAAAACCTCTATAATCACCCAAACACCAAAACTATACACCAAGCACTCCCTACGCTAACCAAC
TTTCACAAATCACCCAAATAAAACTAAACCCAAACAAATAACTCCACCAACTCTTACACCCACCAACACCCCCCAAGTCA
..... (이하 생략)

```

그림 12. 데이터파일 mtCDNApri123.txt; 세개의 파티션 데이터가 차례대로 나열되어 있다. 각 파티션은 phylip 형식이다. 본 데이터는 코돈의 1,2,3번째 사이트이다. 일반적인 파티션도 이와 동일한 형식으로 데이터를 구성하면 된다.

< 계통수 파일 mtCDNApri.trees >

```
7 1
```

```
((((human, (chimpanzee, bonobo)) '>.06<.08', gorilla), (orangutan, sumatran)) '>.12<.16', gibbon);
```

```
//end of file
```

```
((((human, (chimpanzee, bonobo)) '>.06<.08', gorilla), (orangutan, sumatran)) '>.12<.16', gibbon);
```

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

화석에 의한 내부노드의 시간 제한; 1단위 시간은 1억년(100백만년)이다.

```

seed = 1
seqfile = mtCDNApri123.txt
treefile = mtCDNApri.trees
outfile = out.txt

ndata = 3
seqtype = 0 * 0: nucleotides; 1:codons; 2:AAs
usedata = 2 * 0: no data; 1:seq like; 2:use in.BV; 3: out.BV
clock = 3 * 1: global clock; 2: independent rates; 3: correlated rates
RootAge = <1.0 * safe constraint on root age, used if no fossil for root.

model = 4 * 0:JC69, 1:K80, 2:F81, 3:F84, 4:HKY85
alpha = 0.5 * alpha for gamma rates at sites
ncatG = 5 * No. categories in discrete gamma

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

```

```

print = 1
burnin = 2000
sampfreq = 2
nsample = 20000

```

(계산 시간을 줄이기 위해
위의 설정을 그대로 사용;
일반적으로 큰 값으로 설
정하고 여러 번 반복하여
결과의 수렴여부 확인해
야 함)

그림 13. mcmctree.2.ctl.txt은 분석의 두번째 단계를 실행하기 위한 컨트롤 파일이다. 붉은색 상자가 수정부분이고
수정이 필요없는 하단 부분은 표시를 생략하였다.

C:\temp>mcmctree mcmctree.2.ctl.txt



그림 14. MCMCTree 프로그램 실행의 두번째 단계. 본격적으로 MCMC 알고리즘을 실행한다.

5%	0.24	0.26	0.38	0.36	0.31	0.177	0.157	0.089	0.062	0.023	0.038	-0.292	0.843	-17.7	0:01
10%	0.24	0.26	0.38	0.33	0.31	0.176	0.156	0.088	0.062	0.023	0.038	-0.312	0.842	-17.2	0:01
15%	0.23	0.27	0.38	0.33	0.31	0.178	0.156	0.088	0.062	0.023	0.038	-0.331	0.826	-17.0	0:01
20%	0.23	0.27	0.38	0.32	0.31	0.177	0.156	0.088	0.062	0.023	0.037	-0.332	0.821	-16.9	0:01
25%	0.23	0.27	0.38	0.33	0.31	0.177	0.156	0.088	0.062	0.023	0.037	-0.335	0.831	-16.8	0:01
30%	0.23	0.27	0.38	0.33	0.31	0.177	0.156	0.088	0.062	0.023	0.037	-0.339	0.828	-16.8	0:02
35%	0.23	0.27	0.38	0.34	0.31	0.177	0.156	0.088	0.062	0.023	0.037	-0.350	0.836	-16.7	0:02
40%	0.23	0.26	0.38	0.34	0.31	0.178	0.156	0.088	0.062	0.023	0.038	-0.347	0.841	-16.8	0:02
45%	0.23	0.27	0.37	0.34	0.31	0.178	0.156	0.088	0.062	0.023	0.037	-0.353	0.835	-16.8	0:02
50%	0.23	0.26	0.37	0.34	0.31	0.178	0.156	0.089	0.062	0.023	0.038	-0.348	0.835	-16.8	0:02
55%	0.23	0.26	0.38	0.34	0.31	0.179	0.156	0.088	0.062	0.023	0.037	-0.351	0.832	-16.8	0:03
60%	0.23	0.26	0.37	0.34	0.31	0.179	0.156	0.088	0.062	0.023	0.037	-0.349	0.827	-16.8	0:03
65%	0.23	0.26	0.37	0.34	0.31	0.178	0.156	0.088	0.062	0.023	0.038	-0.347	0.832	-16.8	0:03
70%	0.23	0.26	0.37	0.34	0.31	0.179	0.156	0.088	0.062	0.023	0.038	-0.346	0.831	-16.9	0:03
75%	0.23	0.26	0.37	0.34	0.31	0.179	0.156	0.088	0.062	0.023	0.038	-0.349	0.825	-16.9	0:03
80%	0.23	0.26	0.37	0.34	0.31	0.179	0.156	0.088	0.062	0.023	0.038	-0.346	0.825	-16.9	0:03
85%	0.23	0.26	0.37	0.34	0.31	0.179	0.156	0.088	0.062	0.023	0.037	-0.348	0.826	-16.9	0:04
90%	0.23	0.26	0.37	0.34	0.31	0.180	0.156	0.088	0.062	0.023	0.038	-0.346	0.827	-16.9	0:04
95%	0.23	0.26	0.37	0.34	0.31	0.180	0.156	0.088	0.062	0.023	0.037	-0.349	0.829	-16.9	0:04
100%	0.23	0.26	0.37	0.34	0.31	0.180	0.156	0.088	0.062	0.023	0.037	-0.351	0.827	-16.9	0:04

그림 15. MCMC 알고리즘 진행 화면

Posterior mean (95% Equal-tail CI) (95% HPD CI) HPD-CI-width

t_n8	0.1796	(0.1566 , 0.2140)	(0.1552 , 0.2108)	0.0556	(Jnode 12)
t_n9	0.1562	(0.1446 , 0.1624)	(0.1464 , 0.1635)	0.0170	(Jnode 11)
t_n10	0.0885	(0.0807 , 0.0976)	(0.0803 , 0.0970)	0.0167	(Jnode 10)
t_n11	0.0620	(0.0588 , 0.0679)	(0.0584 , 0.0671)	0.0087	(Jnode 9)
t_n12	0.0232	(0.0186 , 0.0284)	(0.0184 , 0.0280)	0.0096	(Jnode 8)
t_n13	0.0374	(0.0292 , 0.0460)	(0.0294 , 0.0461)	0.0167	(Jnode 7)
mu1	0.5149	(0.3979 , 0.6417)	(0.4033 , 0.6465)	0.2432	
mu2	0.1761	(0.1295 , 0.2257)	(0.1301 , 0.2259)	0.0958	
mu3	5.6300	(3.9849 , 7.5471)	(3.9429 , 7.4926)	3.5497	
sigma2_1	0.2410	(0.0068 , 0.7363)	(0.0017 , 0.6158)	0.6141	
sigma2_2	0.3507	(0.0300 , 0.9282)	(0.0039 , 0.7930)	0.7891	
sigma2_3	0.8274	(0.2998 , 1.6802)	(0.2646 , 1.5698)	1.3052	
lnL	-16.8512	(-25.1920 , -10.3060)	(-24.4020 , -9.7220)	14.6800	

그림 16. 두번째 단계 실행 이후 out.txt 파일 하단에는 내부 노드의 분기연대 사후분포의 평균 및 95 % 신뢰구간이 표시된다. 내부 노드의 위치에 관한 정보는 out.txt 파일내에서 찾을 수 있다.

Gen	t_n8	t_n9	t_n10	t_n11	t_n12	t_n13	mul
1	0.1828834		0.1615101		0.0876872		0.0614
2	0.1828834		0.1526452		0.0854514		0.0614
4	0.1828834		0.1526452		0.0854514		0.0608
6	0.1828834		0.1582246		0.0879473		0.0608
8	0.1847655		0.1518980		0.0924822		0.0637
10	0.1847655		0.1607646		0.0910202		0.0608
12	0.1667355		0.1555069		0.0868780		0.0608
14	0.1840996		0.1555069		0.0936563		0.0608
16	0.1840996		0.1484759		0.0872715		0.0608
18	0.1840996		0.1559565		0.0930122		0.0643
20	0.1702507		0.1504006		0.0930122		0.0660
22	0.1702507		0.1504006		0.0888878		0.0660
24	0.1618147		0.1528821		0.0888878		0.0627
26	0.1618147		0.1528821		0.0888878		0.0654
28	0.1618147		0.1528821		0.0956387		0.0646
30	0.1614548		0.1528821		0.0956387		0.0622
32	n 173901n		n 1604972		n 0956387		n 0622

그림 17. mcmc.txt 파일에는 사후분포로 부터 얻는 모수들의 샘플이 저장되어 있다.

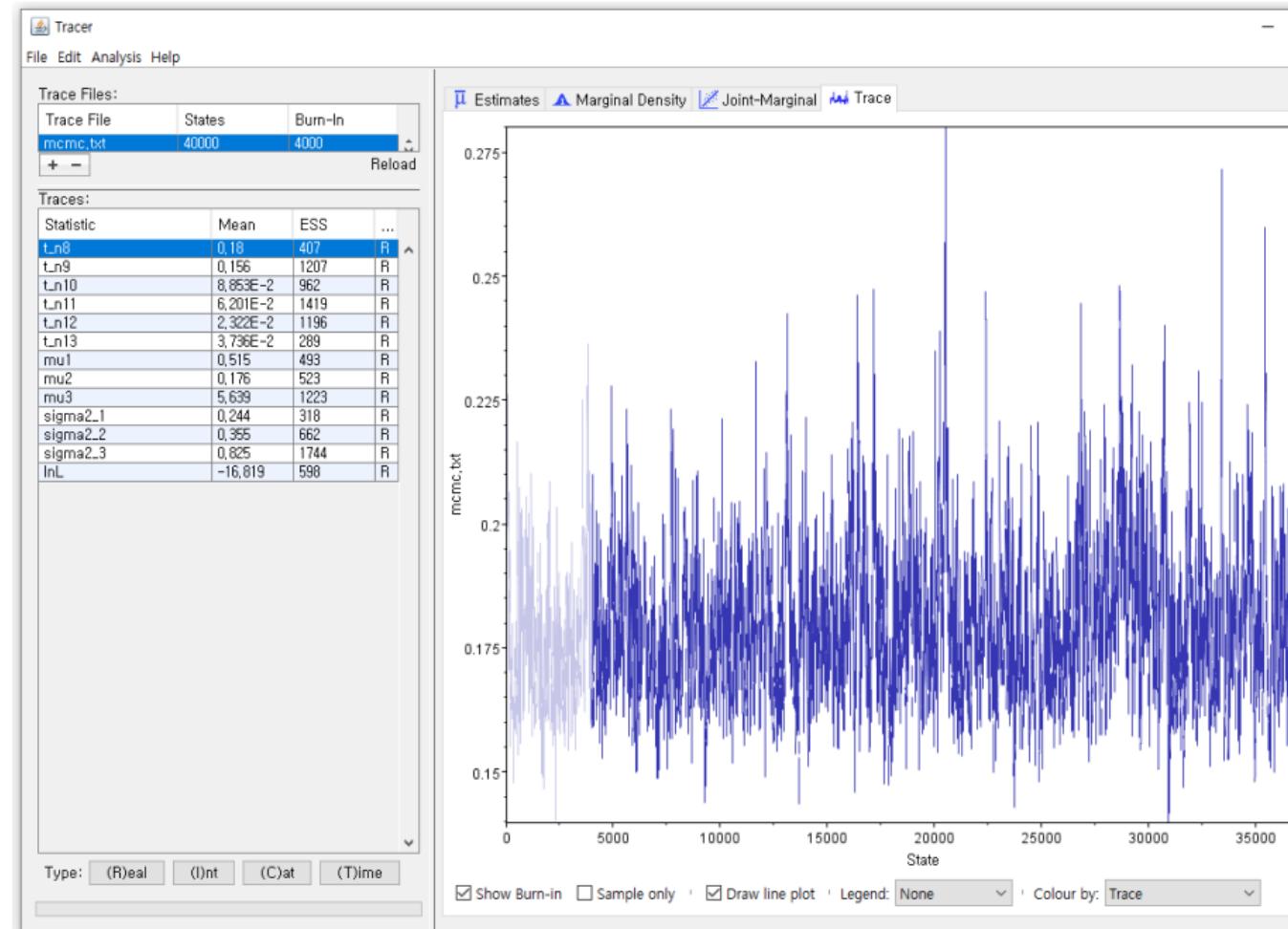


그림 18. mcmc.txt 파일을 Tracer 프로그램으로 읽어들인 화면.

* ESS(Effective Sample size)가 작을 때는 붉은(주황)색으로 표시됨 → Posterior sample 수를 늘릴 필요가 있음.

```
print = 1
burnin = 2000
sampfreq = 2
nsample = 20000
```

디폴트로 지정된 figtree.tre 파일에는 분기연대의 사후분포에 대한 정보가 저장된다. 이를 FigTree프로그램²²으로 열어보자(그림 19). 좌측 패널에서 Node Bar 항목에 체크하고 하부의 95%HPD를 선택한다. 또한, Scale Axis에 체크하고 하부의 Reverse axis 항목을 체크한다. 그외 취향에 따라 폰트의 크기나 계통수 가지의 굵기등을 조절하면 보기 좋은 그림을 얻을 수 있다.

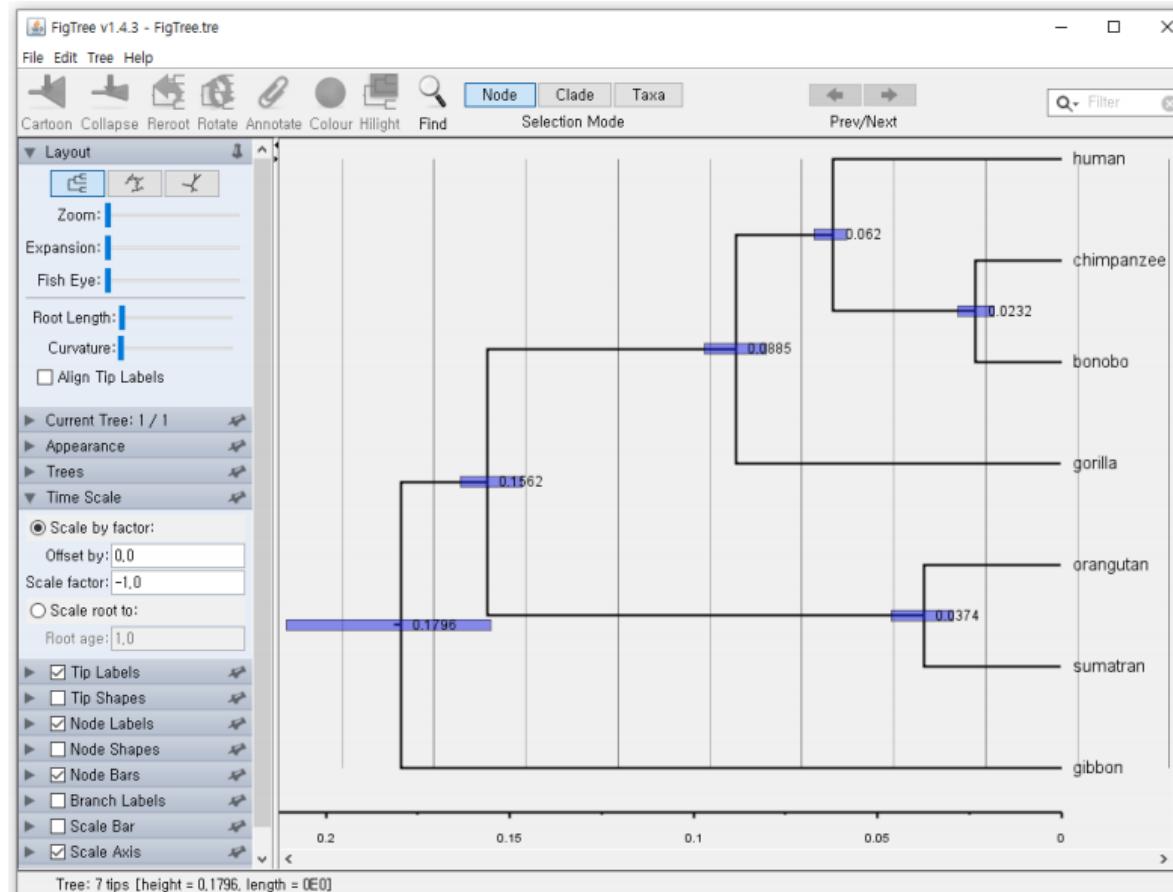
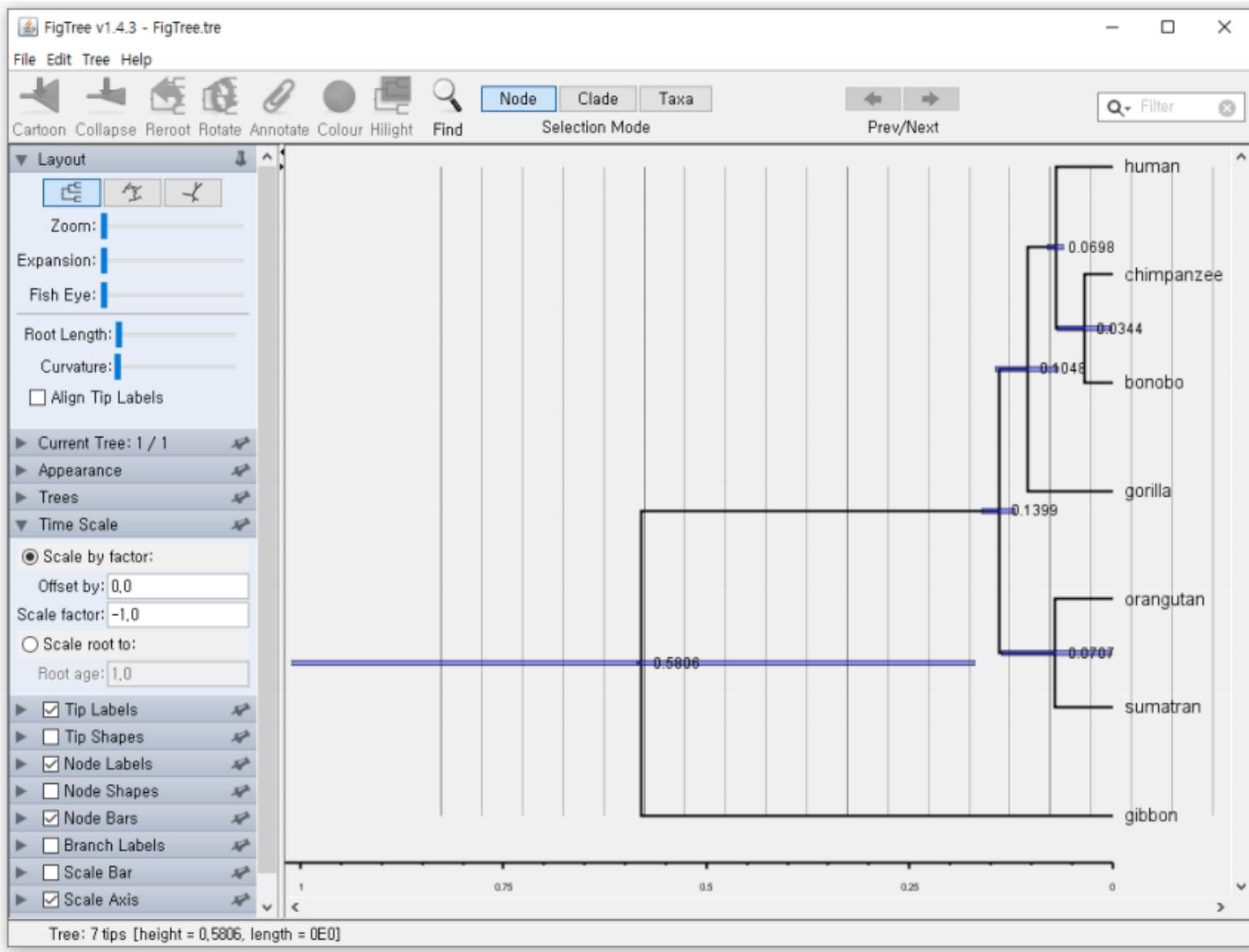


그림 19. 두번째 단계 실행후 생성된 figtree.tre 파일을 FigTree프로그램으로 읽어 들인 화면 모습. 각 내부노드에 분기연대의 사후분포 평균과 95 % 신뢰구간이 표시되어 있다.



두번째 단계에서 다른 설정은 모두 동일하게 하고 usedata 설정만 'usedata=0'을 지정하고 실행. 이는 사전분포와 화석데이터만 정보로서 입력하고 실행한 것.

그림 20. 화석정보는 포함하나 염기서열데이터는 포함하지 않은 분기연대 추정값.

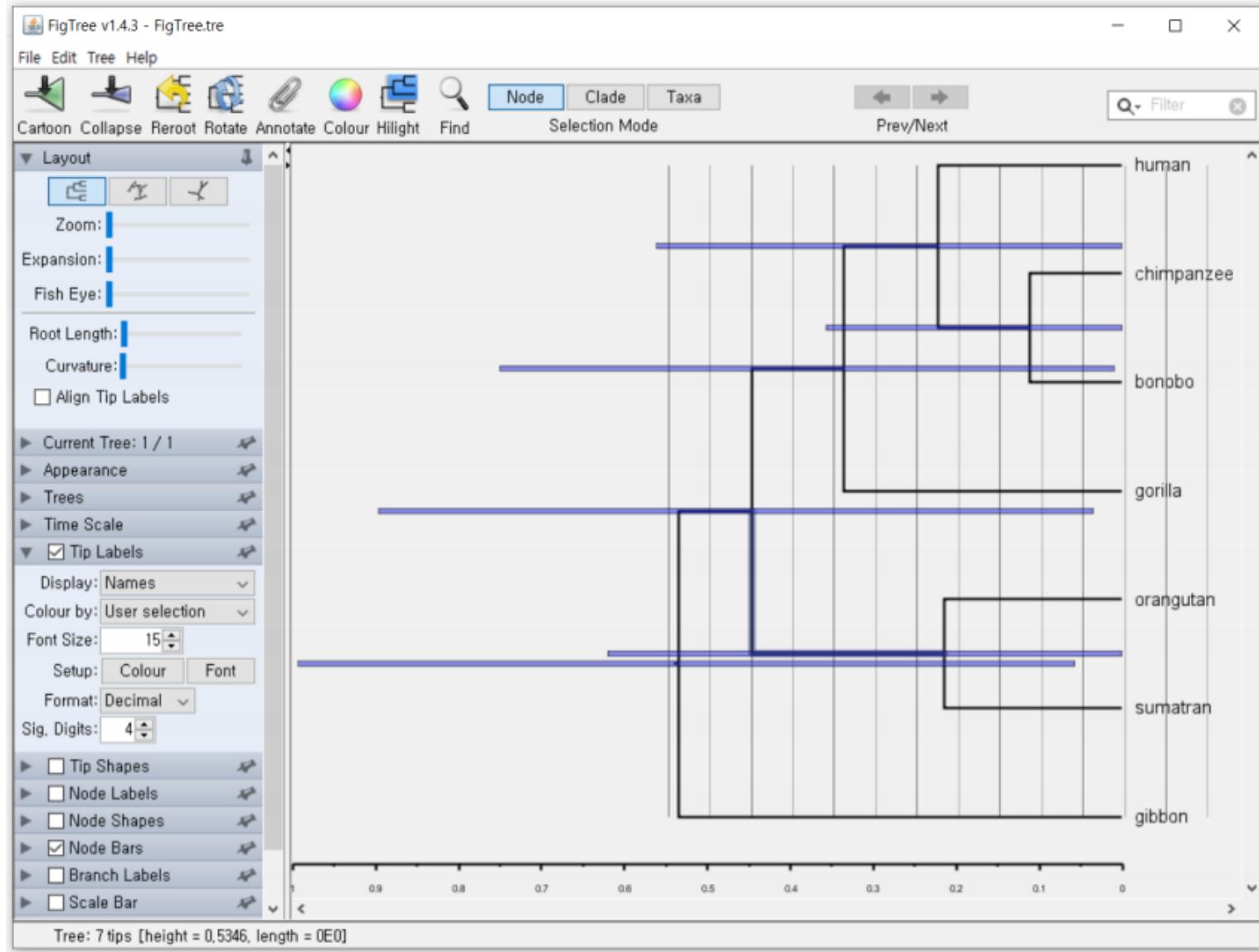


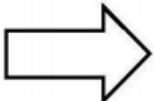
그림 21. 화석정보와 염기서열 정보 모두 포함하지 않았을때의 분기연대 사전분포.

화석 정보를 포함하고 있는 mtCDNApri.trees 파일에서 화석 정보를 삭제한 후 'usedata=0'을 지정하고 두번째 단계의 분석을 실행

<화석 정보 삭제 방법>
계통수 내의 ' $>0.06 <0.08$ ', ' $>0.12 <0.16$ ' 의 화석정보 문자열을 삭제한다. 화석정보가 전무할 경우 컨트롤 파일에서 RootAge는 양방향 제한이 필요하므로 하한값을 임으로 0.01로 지정하여 'RootAge = $<1.0 >0.01$ '로 설정한다.

<단위 시간 변경 방법>

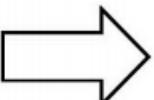
```
clock = 2  
RootAge = '<1.0'  
BDparas = 1 1 0  
kappa_gamma = 6 2  
alpha_gamma = 1 1  
rgene_gamma = 2 20 1  
sigma2_gamma = 1 10 1
```



```
clock = 2  
RootAge = '<100.0'  
BDparas = 0.01 0.01 0  
kappa_gamma = 6 2  
alpha_gamma = 1 1  
rgene_gamma = 2 2000 1  
sigma2_gamma = 1 10 1
```

그림 22. independent rate 모형에서 단위시간이 1/100로 변경된 경우.

```
clock = 3  
RootAge = '<1.0'  
BDparas = 1 1 0  
kappa_gamma = 6 2  
alpha_gamma = 1 1  
rgene_gamma = 2 20 1  
sigma2_gamma = 1 10 1
```



```
clock = 3  
RootAge = '<100.0'  
BDparas = 0.01 0.01 0  
kappa_gamma = 6 2  
alpha_gamma = 1 1  
rgene_gamma = 2 2000 1  
sigma2_gamma = 1 1000 1
```

그림 23. Correlated rates 모형에서 단위시간이 1/100로 변경된 경우.