MrBayes를 이용한 데이타 분석



MrBayes: Bayesian Inference

of Phylo



MrBayes is a pro across a wide ra MrBayes uses Ma estimate the pos

Program feature:

Manua
Bug Rep

A common
Author



Vetenskapsrådet





- A common Windows, a
- Extensive h
- Analysis of morphologic
- Mixing of data
 characters,
- Easy linking partitions;
- An abundar doublet, an the standar
- Estimation
 Bayesian fr



Home
» Download «
Manual
Bug Report
Authors
Links

MrBayes: Bayesian Inference of Phylogeny

MrBayes may be downloaded as a pre-compiled executable or in source form (recommended).

Current release

The most recent release version of MrBayes is 3.2.7a, released March 6, 2019.

The 3.2.7a source code is available for compilation on Unix machines

Pre-compiled (provisional) executables are available for Windows (MrBayes-3.2.7-WIN.zip). These are, however, serial versions compiled without the Beagle library. The serial version works well for smaller analyses but if you plan to run large analyses using many parallel chains, you should use the MPI version instead. Refer to the User Manual, and the INSTALL document on GitHub for help with installation of the program.

MrBayes may also be installed through the Homebrew package manager on macOS, Linux, and Windows Subsystem for Linux (WSL). Please see the <u>INSTALL</u> document for instructions.

Older releases

You can get access to older releases (from release 3.2.0 onwards), by browsing

(1) 명령프롬프트을 열어 MrBayes프로그램을 실행한다

C:\temp>mb.3.2.7-win64.exe

MrBayes 3.2.7a ×86_64

(Bayesian Analysis of Phylogeny)

Distributed under the GNU General Public License

Type "help" or "help <command>" for information on the commands that are available.

Type "about" for authorship and general information about the program.

(본 자료는 MrBayes 매뉴얼 Chapter 3 Tutorial: A partitioned analysis 의 일부 내용 + α 입니다. 보다 자세한 내용은 매뉴얼을 참조해주세요)

MrBayes > exec cynmix.nex_

MrBayes > exec cynmix.nex

" bayes - exec cyrminx mex

DOS line termination

Longest line length = 217

Executing file "cvnmix.nex"

Parsing file

Expecting NEXUS formatted file

Reading data block

Allocated taxon set

Allocated matrix

Defining new matrix with 32 taxa and 3246 characters

Data is Mixed

Data for partition 1 is Standard

Data for partition 2 is Dna

There are a total of 2 default data divisions
Data matrix is interleaved

Gaps coded as -

명령어를 사용하여 읽어들인다.

(2) cynmix.nex 데이터를 exec

(3) cynmix.nex 파일을 메모장위에 drag/drop 하여 내용을 살펴보자.

각진 괄호('[' 와 ']')는 주석부분이라 프로그램 실행시 무시된다.

Dimensions 명령어를 이용하여 taxa의 수 character 수를 지정한다.

Format 명령어를 이용하여 데이터의 종류를 설정한다. Standard는 Morphological character를 지정하는 명령어이다. Interleave, gap, missing의 설정방식을 눈여겨 보자.

```
#NEXUS
[ Data from: Nylander JAA, Ronquist F, Huelsenbeck JP, Nieves-Aldrey JL. 2004. Bayesian phylogenet
Begin data;
 Dimensions ntax=32 nchar=3246;
 Format datatype=mixed(Standard:1-166,DNA:167-3246) interleave=yes gap=- missing=?:
 Matrix
Ibalia
    Synergus
Periclistus
    Ceroptres
    Synophromorpha
    Xestophanes
    Diastrophus
    Gonaspis
    1-010120000?01001010000110110001000001020000101? ??00101111111002100000000221001110
Antistrophus
Rhodus
    Hedickiana
    Neaylax
Isocolus
    Aulacidea
    Panteliella
    00000030001002007?170007?7?07?0107007??7?7?72007??17101010111102070?700000121001010
Barbotinia
```

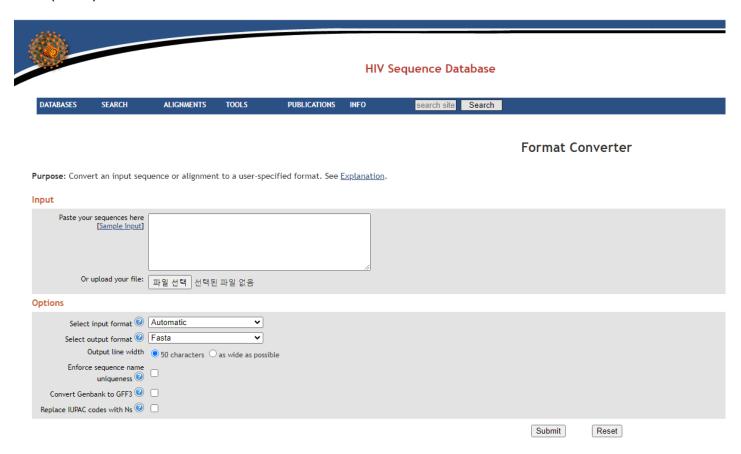
(4) cynmix.nex 파일의 하단 부분을 살펴보자.

{ begin data; end; } 로 지정한 데이터 블록이 표시된 이후에 { begin mrbayes; end; } 블록에 MrBayes 실행을 위한 설정이 나열된다.

각진 괄호는 주석 부분이라 프로그램 실행 시 무시된다(주석부분을 살린 후, 즉, 각진 괄호를 삭제한 후, cynmix.nex 파일을 읽어 들이면 설정이 자동으로 실행된다).

```
begin mrbayes;
    [This block defines several different character sets that could be used in partitioning the
    and then defines and enforces a partition called favored.]
    charset morphology = 1-166;
    charset COI = 167-1244;
    charset EF1a = 1245-1611;
    charset LWRh = 1612-2092;
    charset 28S = 2093-3246;
    partition favored = 5: morphology, COI, EF1a, LWRh, 28S;
    [The following lines set up a particular model (the one discussed in the MrBayes manual). Up
    if you want to set up the model when executing this file instead of specifying it yourself.
    set partition=favored;
    lset app=(1) rates=gamma;
    lset app=(2,3,4,5) rates=invgamma nst=6;
    unlink revmat=(all) pinvar=(all) shape=(all) statefreq=(all);
    prset ratepr=variable;
end:
```

(참고) 서열데이터의 포맷 변환 사이트



https://www.hiv.lanl.gov/content/sequence/FORMAT_CONVERSION/form.html

(5) 결과의 재현가능성을 위해 seed=1 swapseed=1을 설정한다(다른 양의 정수도 무방).

서열데이터를 읽어 들인다.

partition=favored 설정한다.

```
MrBayes > set seed=1 swapseed =1
   Setting seed to 1
   Setting swapseed to 1
MrBayes > exec cynmix.nex
          ( . . . . . )
          (화면 출력 부분 생략)
          (\ldots )
MrBayes > set partition = favored
   Setting favored as the partition, dividing characters into 5
parts.
   Setting model defaults
   Seed (for generating default start values) = 1911300560
   Expecting command
MrBayes >
```

(6) showmodel 명령어로 설정을 확인해보자. partition 1은 morphological character, partition2-5는 DNA 로 동일

```
MrBayes > showmodel
  Model settings:
     Settings for partition 1 --
       Datatype = Standard
       Coding = Variable
       # States = Variable, up to 10
                  State frequencies are fixed to be equal
       Rates = Equal
        ( . . . . . )
        (화면 출력 부분 생략)
        (\ldots )
  Active parameters:
                     Partition(s)
                    1 2 3 4 5
     Parameters
     Statefreq 1 2 2 2 2
     Ratemultiplier 3 3 3 3 3
                   4 4 4 4 4
     Topology
     Brlens
        ( . . . . . )
        (화면 출력 부분 생략)
        (....)
```

(7) Iset 명령어로 파티션 1-5의 모형을 설정한다 (DNA 모형에 관한 사항은 서태건2022 논문 참조).

```
MrBayes > lset applyto=(1) rates=gamma;
   Setting Rates to Gamma for partition 1
   Successfully set likelihood model parameters to
  partition 1 (if applicable)
   Expecting command
MrBayes > lset applyto=(2,3,4,5) rates=invgamma nst=6;
   Setting Rates to Invgamma for partition 2
   Setting Rates to Invgamma for partition 3
   Setting Rates to Invgamma for partition 4
   Setting Rates to Invgamma for partition 5
   Setting Nst to 6 for partition 2
   Setting Nst to 6 for partition 3
   Setting Nst to 6 for partition 4
   Setting Nst to 6 for partition 5
   Successfully set likelihood model parameters to
  partitions 2, 3, 4, and 5 (if applicable)
   Expecting command
MrBayes >
```

(8) showmodel 명령어로 설정을 확인해보자. partition 1은 morphological character, partition2-5는 DNA 로 동일

```
MrBayes > showmodel;
  Model settings:
     Settings for partition 1 --
        Datatype = Standard
        Coding = Variable
        # States = Variable, up to 10
                    State frequencies are fixed to be equal
         ( . . . . . )
         (화면 출력 부분 생략)
         (....)
 Active parameters:
                        Partition(s)
     Parameters
     Revmat
     Statefreq
     Shape
     Pinvar
     Ratemultiplier 7 7 7 7 7
     Topology
     Brlens
         (....)
         (화면 출력 부분 생략)
         ( . . . . . )
```

(9) unlink 명령어로 파티션별로 다른 모수를 할당하도록 설정한 후 showmodel 명령어로 설정을 확인해보자.

앞페이지의 showmodel 결과와 비교해보자.

```
MrBayes > unlink revmat=(all) pinvar=(all) shape=(all)
statefreq=(all);
  Unlinking
  Expecting command
MrBayes > showmodel;
         ( . . . . . )
         (화면 출력 부분 생략)
         ( . . . . . )
Active parameters:
                        Partition(s)
     Parameters
     Revmat
     Statefreq 5 6 7 8 9
                    10 11 12 13 14
     Shape
     Pinvar
                      . 15 16 17 18
     Ratemultiplier 19 19 19 19 19
     Topology 20 20 20 20 20
     Brlens
                      21 21 21 21 21
         ( . . . . . )
         (화면 출력 부분 생략)
         (\ldots )
```

(10) prset 명령어로 ratepr의 설정을 바꾼다.

showmodel 명령어로 설정을 확인한다. 그림의 붉은색 부분이 prset설정 이전과 어떻게 다른지 살펴보자.

```
19 --
           Parameter = Ratemultiplier{all}
           Type
                     = Partition-specific rate multiplier
           Prior
                     = Fixed(1.0)
           Partitions = All
         (화면 출력 생략)
MrBayes > prset ratepr=variable;
         (화면 출력 생략)
MrBayes > showmodel
         (화면 출력 부분 생략)
19 -- Parameter = Ratemultiplier{all}
                     = Partition-specific rate multiplier
           Type
                      = Dirichlet(1.00,1.00,1.00,1.00,1.00)
           Prior
           Partitions = All
```

(11) 『help Iset 』명령으로 likelihood setting 확인. 그림은 파티션 5의 예시이다. 파티션 1과는 어떻게 다른지 살펴보자.

```
MrBayes > help lset;
           (화면 출력 생략)
  Model settings for partition 5:
   Parameter Options
                                                   Current Setting
  Nucmodel 4by4/Doublet/Codon/Protein
                                                     4bv4
             1/2/6/Mixed
   Nst
               Universal/Vertmt/Invermt/Yeast/Mycoplasma/
   Code
               Ciliate/Echinoderm/Euplotid/Metmt
                                                     Universal
   Ploidy
               Haploid/Diploid/Zlinked
                                                     Diploid
   Rates
               Equal/Gamma/LNorm/Propinv/
               Invgamma/Adgamma/Kmixture
                                                     Invqamma
  Ngammacat <number>
  Nlnormcat <number>
  Nmixtcat
             <number>
  Nbetacat <number>
  Omegavar Equal/Ny98/M3
                                                     Equal
   Covarion
              No/Yes
                                                     No
   Coding
               All/Variable/Informative/Nosingletons
               Noabsencesites/Nopresencesites/
               Nosingletonabsence/Nosingletonpresence All
   Parsmodel
               No/Yes
                                                     Nο
MrBayes >
```

(12) 『help prset 』명령으로 prior setting 확인. 그림은 파티션 5의 예시이다.

```
MrBayes > help prset;
          (화면 출력 생략)
  Model settings for partition 5:
  Parameter Options
                             Current Setting
  Tratiopr Beta/Fixed
                                         Beta(1.0,1.0)
  Revmatpr Dirichlet/Fixed Dirichlet(1.0,1.0,1.0,1.0,1.0,1.0)
  Aamodelpr Fixed/Mixed
Aarevmatpr Dirichlet/Fixed
                                         Fixed(Poisson)
                                    Dirichlet(1.0,1.0,...)
          Dirichlet/Fixed Dirichlet(1.0,1.0)
  Omegapr
          (화면 출력 생략)
  TK02varpr Fixed/Exponential/Uniform
                                        Exponential(1.00)
              Fixed/Exponential/Uniform
Fixed/Variable=Dirichlet
                                          Exponential(10.00)
  Igrvarpr
                                          Dirichlet(...,1.0,...)
  Ratepr
  Generatepr Fixed/Variable=Dirichlet
                                        Fixed
MrBayes >
```

(13) 『help mcmcp』 명령으로 MCMC parameter setting 확인. 디폴트 세팅이다.

```
MrBayes > help mcmcp;
   mcmcp
  Mcmcp
  This command sets the parameters of the Markov chain Monte Carlo (MCMC)
  analysis without actually starting the chain. This command is identical
  in all respects to Mcmc, except that the analysis will not start after
  this command is issued. For more details on the options, check the help
  menu for Mcmc.
  Parameter Options Current Setting
  Ngen <number>
                         1000000
            <number>
  Nruns
  Nchains
  Temp
           <number>
                          0.100000
  Reweight <number>,<number> 0.00 v 0.00 ^
Swapfreq <number> 1
  Nswaps
           <number>
                                    1
  Samplefreq <number>
Printfreq <number>
                                   500
                                   1000
           Yes/No
  Printall
                                    Yes
  (화면 출력 생략)
MrBayes >
```

(14) MCMC 설정을 바꾸려면 아래와 같이 mcmcp 명령으로 바꿀 수 있다. 『help mcmcp』명령으로 MCMC parameter setting 확인하고 어느 부분이 바뀌었는지 살펴보자.

```
MrBayes > mcmcp ngen=1000000 samplefreq=500 printfreq=100 diagnfreq=1000;
          (화면 출력 생략)
MrBayes > help mcmcp;
   mcmcp
  Mcmcp
  This command sets the parameters of the Markov chain Monte Carlo (MCMC)
  analysis without actually starting the chain. This command is identical
  in all respects to Mcmc, except that the analysis will not start after
  this command is issued. For more details on the options, check the help
  menu for Mcmc.
  Parameter Options
                         Current Setting
              <number>
                                   1000000
  Ngen
           <number>
  Nruns
           <number>
  Nchains
              <number>
                                  0.100000
  Temp
           <number>,<number> 0.00 v 0.00 ^
  Reweight
            <number>
  Swapfreq
              <number>
  Nswaps
  Samplefreq <number>
                                  500
  Printfreq <number>
                                  100
  Printall
             Yes/No
                                   Yes
  (화면 출력 생략)
MrBayes >
```

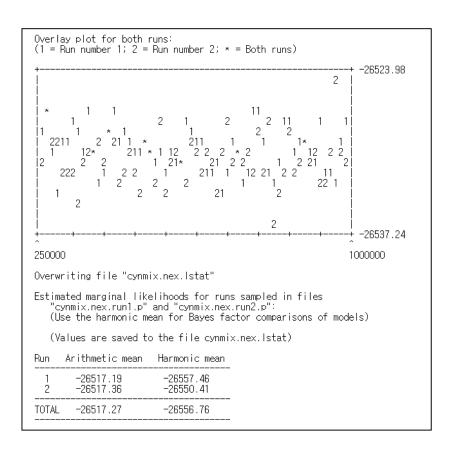
(15) 『mcmc』를 입력하여 MCMC를 실행한다. 설정한대로 2개의 MCMC run (가운데 '*' 로구분) 각 run당 chain 4개 (cold 1, hot 3) 가 실행됨을 볼 수 있다. 설정한 100만세대에 도달하면 계속 진행할지 종료할지 물어온다. No를 입력.

```
MrBayes > mcmc
```

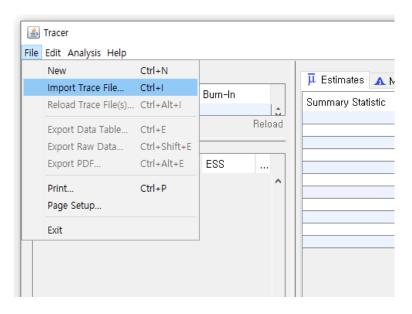
```
Chain results (1000000 generations requested):
                                                                                  (-36535.731)
(-33392.604)
                                                    (-36315.410) * [-36398.170]
                        (-36612.898)
                                                                                                 (-36347.760)
    0 -- [-36420.775]
                                      (-36521.562)
                                                                                                               (-36231.370)
  100
           -33056.218]
                         -33453.289)
                                      (-34107.536)
                                                    (-34076.557) *
                                                                    (-33764.603)
                                                                                                [-32423.715]
                                                                                                                -32517.073)
                                                    (-32043.488) *
          -31774.7651
                        -32229.865)
                                      (-32461.079)
                                                                    (-31670.658)
                                                                                   (-31390.711)
                                                                                                 (-31251.369)
  200 ---
                                                                                                                -31072.3831
                                                                                                                             -- 2:46:38
          -30321.925]
                        -31283.716)
                                      (-31156.318)
                                                    (-31389.299) *
                                                                    (-30856.835)
                                                                                   (-30666.852)
                                                                                                 (-30596.379)
                                                                                                                -29999.069]
  300 ---
                                                                                                                             -- 2:46:37
                                                    (-30533.192) *
                                                                                                 (-30199.993)
           -29977.0231
                        -30688.639)
                                      (-30607.615)
                                                                     (-29903.261)
                                                                                    -30231.651)
                                                                                                                -29381.049]
  400 ---
                                                                                                                             -- 2:46:36
                                                    (-29817.365) *
                                                                                                 (-30000.738)
                                                                                                                [-28721.935]
          -29035.889]
                        (-29767.380)
                                      (-29943.057)
                                                                     (-29423.371)
                                                                                   (-29873.605)
  500 ---
                                                                                                                             -- 2:46:35
                                      (-29331.651)
(-29086.234)
                                                    (-29442,260) *
                                                                                                 (-29742.622)
          -28667.397]
                        (-29342.569)
                                                                    (-28927.953)
                                                                                    -29207.586)
  600 ---
                                                                                                                -28528.715]
                                                   (-28989.418) *
           -28331.464<u>1</u>
                        (-28718.483)
                                                                    (-28727.895)
                                                                                    -28890.244)
                                                                                                 (-29191.674)
  700 ---
                                                                                                               [-28460.043]
                                                                                                                             -- 3:10:20
                                                                                                 (-28850.931)
  800 ---
                                      (-28999.631) (-28779.440) * (-28395.527)
                                                                                   -28482.847)
                                                                                                               [-28227.453]
          [-28196.837]
                        (-28321.949)
                                                                                                                             -- 3:07:21
900 -- [-27
1000 -- [-27
                Average standard deviation of split frequencies: 0.005460
                                        [-26526.585]
                                                        -26533.179) (-26548.298) *
                                                                                    [-26535.568]
                                                                                                   (-26538, 489) (-26532, 042)
                                                                                                                               (-26544,724) -- 0:00:19
                998100 -- (-26550.291)
Average stand
                                                                     (-26543.154)
                                                                                                   (-26536.281)
                998200 -- (-26544.423)
                                          -26532.126]
                                                        -26535.483)
                                                                                     [-26529.331]
                                                                                                                 (-26526.384)
                                                                                                                                -26545.223) -- 0:00:18
                                          -26532.941]
                           (-26546.771)
                                                        -26536.762)
                                                                     (-26554.364)
                                                                                     (-26532.636)
                                                                                                   [-26534.869]
                                                                                                                 (-26528.616)
                                                                                                                               (-26545,230) -- 0:00:17
                998300 ---
1100 -- [-27
                                          -26533.143]
                                                                                                   (-26550.687)
                                                                                                                  -26530.388)
                           (-26553.216)
                                                        -26546.007)
                                                                     (-26553.919) *
                                                                                    [-26528.170]
                                                                                                                               (-26540.282)
                998400 ---
1200 -- [-27
                                                                    (-26559.953) * (-26538.783)
(-26555.832) * (-26543.522)
(-26552.799) * (-26537.804)
(-26551.396) * (-26534.674)
(-26551.624) * [-26534.6071
                                                        -26544.599)
-26541.835)
                            (-26546.576)
                                          -26533.058]
                                                                                                   (-26546.102)
                                                                                                                  -26528.781)
                                                                                                                               [-26540.176]
                998500 ---
                                                                                                                                            -- 0:00:15
1300 -- [-27
                                                                                                   (-26551.674)
                           (-26539.610)
                                          -26531.967]
                                                                                                                  -26526.8601
                                                                                                                                (-26547,567) -- 0:00:14
                998600 ---
1400 ---
                                          -26523.269]
                                                                                                  (-26554.226)
[-26553.409]
                                                        -26544.875)
                998700 ---
                            (-26533.468)
                                                                                                                  -26531.164]
                                                                                                                                -26547.811) -- 0:00:13
1500 -- (-27
                                          -26521.952]
                998800 -- (-26532,194)
                                                        -26550.870)
                                                                                                                 (-26541.061)
                                                                                                                               (-26544.463) -- 0:00:12
1600 -- (-27
                                          -26522.219]
                                                                                                                (-26540.621)
                998900 -- (-26532.885)
                                                        -26548.156)
                                                                                                  (-26550.197)
                                                                                                                               (-26547,552) -- 0:00:11
1700 -- (-27
                999000 -- (-26539.707) [-26522.163]
                                                       (-26547.147) (-26548.942) * (-26534.987) (-26548.281) (-26546.160) [-26550.281] -- 0:00:10
1800 ---
         (-27)
1900 -- [-27
2000 -- (-27
                Average standard deviation of split frequencies: 0.005402
                                                                                                  (-26546.231)
(-26546.769)
                999100 -- (-26543.177) [-26517.502]
                                                       (-26547.697) (-26546.545) *
                                                                                     [-26522.247]
                                                                                                                (-26548.351)
                                                                                                                               (-26550.283) -- 0:00:09
Average stand
                                                                    (-26537.982)
                                                                                     [-26523.955]
                                                                                                                 (-26563.648)
                999200 -- (-26534.871)
                                         [-26515.092]
                                                                                                                                -26543.853) -- 0:00:08
                                                        -26555.376)
                                                        -26552.697)
                                          [-26512.130]
                                                                     (-26532.226)
                                                                                     (-26523.747)
                                                                                                   [-26542.773]
                                                                                                                 (-26562.142)
                                                                                                                                -26549.338) -- 0:00:07
                999300 ---
                            (-26541.478)
                                                                                  *
                                                                                     [-26522.047]
                999400 -- (-26538.384)
                                         (-26518.928)
                                                        -26555.241)
                                                                     [-26535.862]
                                                                                                    (-26546.667)
                                                                                                                  (-26567.006)
                                                                                                                                -26543.911) -- 0:00:06
                                                        -26552.608)
                                         [-26519.030]
                                                                                      -26522.902]
                                                                                                                 (-26562.522)
                999500 -- (-26535.859)
                                                                     (-26540.379)
                                                                                                    (-26549.997)
                                                                                                                                -26546.062) -- 0:00:05
                                                                                                                                (-26544.703) -- 0:00:04
                999600 -- (-26532.493)
                                         (-26524.546)
                                                        -26545.202)
                                                                     [-26537.017]
                                                                                     [-26526.572]
                                                                                                   (-26552.777)
                                                                                                                 (-26562.637)
                                         (-26528.291)
                                                                     (-26530.230)
                                                                                                                 (-26563.025)
                           [-26527.338]
                                                        -26539.300)
                                                                                      -26526.252]
                                                                                                    -26554.610)
                                                                                                                                -26546,465) -- 0:00:03
                999700 ---
                                          (-26531.298)
                                                                                                                (-26561.622)
                999800 -- [-26532.021]
                                                        -26542.708)
                                                                     (-26534.073)
                                                                                     [-26523.469]
                                                                                                    -26553.909)
                                                                                                                                -26544.778) -- 0:00:02
                Average standard deviation of split frequencies: 0.005449
                Continue with analysis (yes/no): no
```

- (16) Burn-in step의 비율을 적절히 설정하고 sump 명령을 실행한다. 텍스트 모드로 보여진 로그가능도 plot과 산술/조화 평균을 보면 1번,2번 run이 비슷한 듯 보인다. 보다 자세히는 Tracer 프로그램을 실행시켜 『*.p』파일을 열어보면 알 수 있다(18번 참조).
- (17) Burn-in step의 비율을 적절히 설정하고 sumt 명령을 실행한다. 이는 tree를 summary 하는 명령어이다.

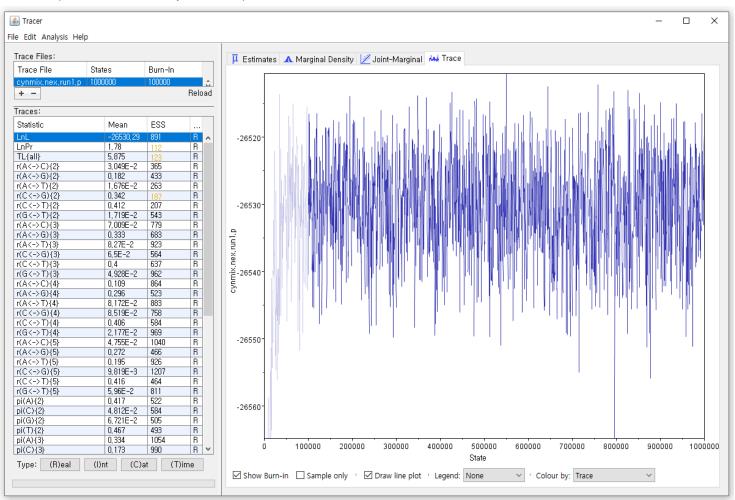
```
MrBayes > sump relburnin=yes burninfrac=0.25 ;
MrBayes > sumt relburnin=yes burninfrac=0.25 conformat=simple;
```



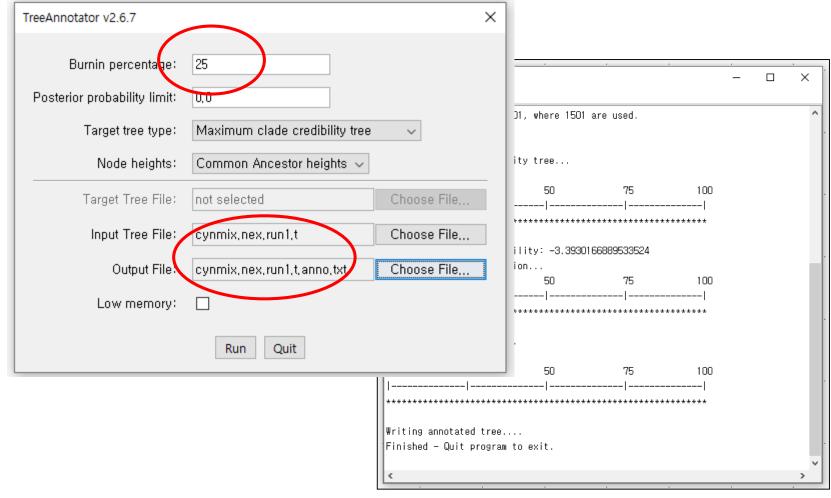
(18) Tracer 프로그램을 실행시켜 cynmix.nex.run1.p 파일을 연다



(18 계속) 왼쪽 패널에서 항목을 선택해서 오른쪽 패널에서 Trace 탭을 누른다. 비교적 랜덤한 패턴이 보이면 총세대수와 sampling interval이 적절하다고 볼 수 있다. ESS(Effective Sample Size) 값은 일부 항목에서 다소 부족한 듯 보인다.



(19) sumt 명령어로 생성된 『*.t 』파일에는 사후분포로부터 샘플링된 개별적인 계통수가 저장되어 있다. 이를 이용하여 각 노드별 사후확률을 계산하기 위해서는 TreeAnnotator(BEAST 프로그램 패키지의 일부)의 실행이 필요하다. Burn-in 비율과 입출력 파일명을 적절히 선택하여 실행한다.



(참고) TreeAnnotator와 FigTree를 이용한 노드별 사후확률 확인

- (1)"figtree phylogeny"로 검색
- (2) 최신버전을 다운로드 및 FigTree***.exe 실행
- (2-1) 자바가 설치되어 있지 않거나 오래된 버전이면 안내에 따라 최신 자바를 설치
- (3) File -> open선택하여 TreeAnnotator 로 생성한 파일 (cynmix.nex.run1.t.anno.txt)을 읽어 들임
- (4) 좌측패널에서 Trees/Order nodes(increasing)에 체크, Tree/Transform branches(cladogram)에 체크
- (5) 좌측패널에서 node label에 체크 display에 posterior 선택
- (6) 기타 폰트 사이즈를 적절히 조정. Root 노드의 위치를 적절히 조정.

