

MrBayes の使い方

目 次

| | |
|----------------------------|-----------------|
| 1 使用法概説 | |
| 1.1 MrBayes のダウンロードとインストール | 1 |
| 1.2 データファイルの作成 | 2 |
| 1.3 データファイルの読み込みと事前設定 | 3 |
| 1.4 解析の実行 | 5 |
| 2 塩基の置換モデルと置換率の決定 | … MrAIC の使用法 |
| 2.1 準備 | 6 |
| 2.2 データファイルの用意 | 6 |
| 2.3 MrAIC の実行 | 7 |
| 2.4 実行結果 | 7 |
| 3 GAP コーディング | … SeqState の使用法 |
| 3.1 ダウンロードとインストール | 8 |
| 3.2 実 行 | 8 |
| 3.3 Nexus 形式のデータファイルの修正 | 9 |

本冊子の著作権はすべて、当サイト euphrasia.sanya-net.ddo.jp の作者に属します。勝手にコピー・改変等を行ったり、第3者に配布することはお断りいたします。また、本冊子の内容をもとに MrBayes による解析を行って、仮に不具合が生じても、責任は著作者にはないことをご了承願います。

1 使用法概説

1.1 MrBayes のダウンロードとインストール

MrBayes の HP である MrBayes: Bayesian Inference of Phylogeny (図 1)、アドレスは

<http://mrbayes.sourceforge.net/download.php> からダウンロードする。WindowsPC を使用している場合は図の矢印の箇所をクリックすると新たにウィンドウが立ち上がり、ダウンロードサイトが開く。

ただし、ブラウザソフトがインターネットエクスプローラの場合、エクスプローラの設定によってはダウンロードサイトが開かない場合がある。その時は、エクスプローラの「インターネットオプション」→「詳細設定」と進んで、「TLS1.0を使用する」、「TLS1.1の使用」、「TLS1.2の使用」それぞれに図2のようにチェックが入っているか確認する。

ダウンロードに際しては、「ファイルを開くか、または保存しますか?」というメッセージが表示されるが、「保存」を選択する。

ファイルは MrBayes-3.2.6_WIN32_X64.ZIP という名称で、ZIP 形式で圧縮されているので、ファイルを圧縮・解凍ができるアーカイブソフトを予めインストールしておくが良い¹。

解凍された結果生じた MrBsyes フォルダはデスクトップに置いておく。

フォルダの内容は図3のようになっている。documentation フォルダには、PDF 形式のマニュアルファイルが入っている²。MrBayes の実行は、Windows が 32 ビットの場合は「Mrbayes_X86.exe」、64 ビットの場合は「Mrbayes_X64.exe」をダブルクリックすることで開始される。



MrBayes: Bayesian Inference of Phylogeny

Current release

Current version of MrBayes is 3.2.6, released November 25, 2015. All downloads of MrBayes are managed through the MrBayes download page on SourceForge. You can download executable versions for [Windows \(32bit and 64bit\)](#), and [Macintosh \(64bit\)](#) computers, as well as the [source code](#) for compilation on Unix machines. The Mac installer includes both the serial and parallel (MPI) version of the program, while the Windows installer only includes the serial version; both include the BEAGLE library (v2.1). 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](#) for help with installation of the program.

図 1 MrBayes の HP

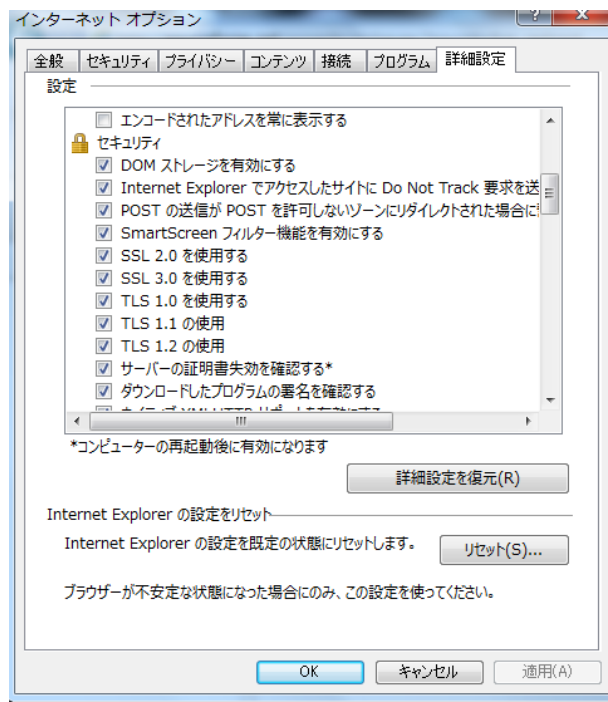


図 2

¹ アーカイブソフトは「窓の杜」の Lhaplus などを使い易い。
² その中の「2.2 Thorough Version」をひととおり読んでおくが良い。

1.2 データファイルの作成

MrBayes の実行にはデータファイルを nexus 形式で作成する必要があるが、Paup などの他の nexus 形式のデータファイルを用いるソフトとは異なる MrBayes 独自の仕様がある。

例えば、MEGA でアライメントを実行して、その結果を Paup 形式(nexus 形式) で出力することができるが、MrBayes では taxlabel(MEGA では "Sequence Name"に相当する)に"."や"(", ")", スペースなどは用いることができない。"."も GAP と混乱して扱われるので避けた方が良い。また datatype 項目に設定可能なのは「DNA」, 「RNA」及び「standard」の3種のみで、"nucleotide"等の用語は設定できない。(図4) MrBayes に限らず、nexus 形式においては[]で囲まれた部分は注釈やコメント行とみなされる³。

ClustalX2 でもアライメントした後、データを nexus 形式で出力することができる。

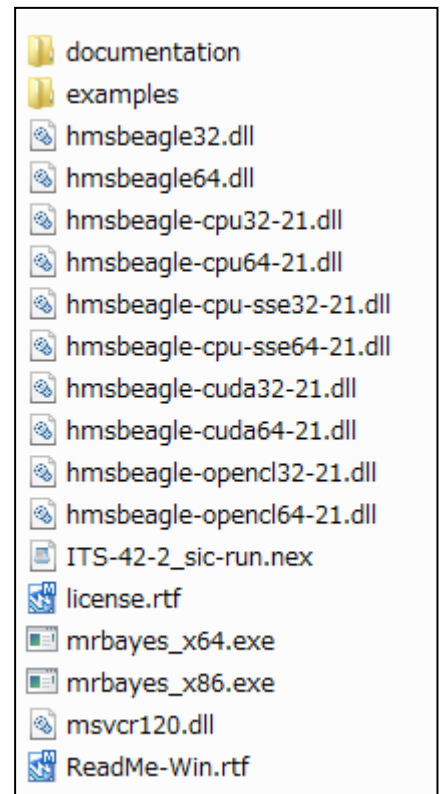


図 3

```
#NEXUS
[ Title Chrysosplenium matK gene ]
begin taxa;
  dimensions ntax=11;
  taxlabels
    C_nagasei
    C_echinus
    C_macrostemon
    C_album
    C_flagelliferum
    C_grayanum
    C_japonicum
    C_pilosum_var_sphaerospermum
    C_fauriei
    T_polyphylla
    R_podophylla
;
end;
begin characters;
  dimensions nchar=1290;
  format missing=? gap=- matchchar=. datatype=dna;
  matrix

C_nagasei
CGAATGCTGTCAACTAAATCATTTGGTATATTTCTGCTAATGATTACAAACGAAAAATACCTT
```

図4 MEGA でアライメントし、PAUP 形式でエクスポートしたデータファイルの先頭部分。このままでは MrBayes では読み込めないので修正する必要がある。修正項目等については9頁を参照。

³ 9頁を参照

1.3 データファイルの読み込みと事前設定

データファイルの読み込みは、MrBayes のコマンドプロンプト ("MrBayes >")から、
 > execute 「データファイル名 (パス名を含む。拡張子は .nex)」

と入力して、enter キーを打つ。nexus 形式のデータファイルが問題なく読み込まれると、図 5 のように "Reached end of file" というメッセージが表示され、コマンド待ちの状態となる。

続いて、解析にあたって必要なパラメータを設定する。おおよその項目については、デフォルトのままで良い。以下、いくつかのパラメータについて概説するが詳しい内容については、マニュアルを参照。

```

Longest line length = 82
Parsing file
Expecting NEXUS formatted file
Reading data block
  Allocated taxon set
  Allocated matrix
  Defining new matrix with 11 taxa and 1290 characters
  Missing data coded as ?
  Data is Dna
  Gaps coded as -
  Taxon 1 -> C_nagasei
  Taxon 2 -> C_echinus
  Taxon 3 -> C_macrostemon
  Taxon 4 -> C_album_
  Taxon 5 -> C_flagelliferum
  Taxon 6 -> C_gravanum
  Taxon 7 -> C_japonicum
  Taxon 8 -> C_pilosum_var_sphaerospermum
  Taxon 9 -> C_fauriei
  Taxon 10 -> T_polyphylla
  Taxon 11 -> R_podophylla
Successfully read matrix
Setting default partition (does not divide up characters)
Setting model defaults
Seed (for generating default start values) = 1449019110
Setting output file names to "sample-2.nex.run<i>.<pt>"
Exiting data block
Reached end of file
MrBayes >
  
```

図 5

1.3.1 塩基の置換モデルと置換率

デフォルトでは F81 モデルになっている。GTR モデルに変更したい場合は、MrBayes のコマンドプロンプトより、

```
Lset nst=6
```

と入力する。また置換率はデフォルトでは同じ ("equal") になっているが、ガンマ設定したい場合は、

```
Lset rates=gamma
```

とする。

モデルの設定状態を確認するには

```
help lset
```

と入力すると、図 6 のように設定状態が表示される。

Nucmodel は DNA の塩基配列による解析ならば、このままの "4by4" で良い。

デフォルトの NST=1 は F81 モデルだが、NST=2 ならば HKY、NST=6 ならば GTR モデルとなる。置換率の Rates 項目については、図のように 6 種類の設定が可能なので、必要に応じて lset rates コマンドで設定すれば良い。他の項目については、一般的にはデフォルトのままで良い⁴。

| Parameter | Options | Current Setting |
|-----------|--|-----------------|
| Nucmodel | 4by4/Doublet/Codon/Protein | 4by4 |
| Nst | 1/2/6/Mixed | 1 |
| Code | Universal/Vertmt/Mycoplasma/Yeast/Ciliates/Metmt | Universal |
| Ploidy | Haploid/Diploid/Zlinked | Diploid |
| Rates | Equal/Gamma/LNorm/Propinv/Invgamma/Adgamma | Equal |
| Ngammacat | <number> | 4 |
| Nbetacat | <number> | 5 |
| Omegavar | Equal/Ny98/M3 | Equal |
| Covarion | No/Yes | No |
| Coding | All/Variable/Noabsencesites/Nopresencesites | All |
| Parsmodel | No/Yes | No |

図 6

1.3.2 その他のモデルパラメータ

Help prset コマンドでは他の様々なパラメータが表示されるが、ほとんどのケースでデフォルトで良い。詳しいことは、マニュアルを参照。

⁴ 置換モデルや置換率の決定については、6～7頁を参照。

1.3.3 解析にあたっての事前設定

コマンドプロンプトから

`Help mcmc`

と入力すると、図の用に **MrBayes** の実行パラメータのデフォルトが表示される。

・ 解析世代数 `Ngen`

デフォルトでは 1,000,000 回となっている。

必要に応じて設定するが、データ量が多ければそれだけ多くの世代数を設定した方が良い。ただし、設定した世代数の解析が完了した時点で、追加設定ができる。また、20,000 世代に設定変更したい場合は

```
mcmc ngen=20000
```

と入力する。

・ 異なる Tree sample 間の比較頻度 `Diagnfreq`

MrBayes では同時に 4 つの解析(run)を実行している。そのうちの 1 つ **Cold Chain**、他の 3 つは **Heat Chain** となっている。実行中にそれぞれの **Chain** でできる系統樹の比較を行う頻度について、デフォルトの値は 5000 世代ごとになっているが、マニュアルでは、50~60 種で、サイト数 2000 程度のデータであれば、1000 世代ごとに変更することを薦めている。この場合以下のように入力する。

```
mcmc diagnfreq=1000
```

・ サンプルング頻度 `samplefreq`

上述のようなデータサイズで、`diagnfreq` がデフォルトよりも小さい設定であれば、サンプルング頻度もデフォルトの 500 より小さい設定が望ましい。例えば以下のように設定する。

```
mcmc samplefreq=100
```

・ 進行状況を何世代ごとに表示するか `printfreq`

サンプルングの頻度に応じて設定する。上述のように `samplefreq` を設定したのであれば、

```
mcmc printfreq=100
```

と入力して、デフォルトよりも下げておく。

解析を実行する前に注意する項目はおよそ上記のとおり。

なお、デフォルトでは **MrBayes** は解析をスタートしてから **Cold Chain** に生成した系統樹の最初の 25%を破棄し、残りの 75%から合意樹をつくるように設定されている。

| | | |
|--------------------------|------------------------------|--------------------|
| <code>Ngen</code> | <number> | 1000000 |
| <code>Nruns</code> | <number> | 2 |
| <code>Nchains</code> | <number> | 4 |
| <code>Temp</code> | <number> | 0.100000 |
| <code>Reweight</code> | <number>, <number> | 0.00 v 0.00 ^ |
| <code>Swapfreq</code> | <number> | 1 |
| <code>Nswaps</code> | <number> | 1 |
| <code>Samplefreq</code> | <number> | 500 |
| <code>Printfreq</code> | <number> | 1000 |
| <code>Printall</code> | Yes/No | Yes |
| <code>Printmax</code> | <number> | 8 |
| <code>Mcmcdiag</code> | Yes/No | Yes |
| <code>Diagnfreq</code> | <number> | 5000 |
| <code>Diagnstat</code> | Avgstddev/Maxstddev | Avgstddev |
| <code>Minpartfreq</code> | <number> | 0.10 |
| <code>Allchains</code> | Yes/No | No |
| <code>Allcomps</code> | Yes/No | No |
| <code>Relburnin</code> | Yes/No | Yes |
| <code>Burnin</code> | <number> | 0 |
| <code>Burnfrac</code> | <number> | 0.25 |
| <code>Stoprule</code> | Yes/No | No |
| <code>Stopval</code> | <number> | 0.05 |
| <code>Savetrees</code> | Yes/No | No |
| <code>Checkpoint</code> | Yes/No | Yes |
| <code>Checkfreq</code> | <number> | 2000 |
| <code>Filename</code> | <name> | sample-2.nex.<p/t> |
| <code>Startparams</code> | Current/Reset | Current |
| <code>Starttree</code> | Current/Random/ Parsimony | Current |
| <code>Nperts</code> | <number> | 0 |
| <code>Data</code> | Yes/No | Yes |
| <code>Ordertaxa</code> | Yes/No | No |
| <code>Append</code> | Yes/No | No |
| <code>Autotune</code> | Yes/No | Yes |
| <code>Tunefreq</code> | <number> | 100 |

1.4 解析の実行

Bayes 解析による系統樹の作成は、コマンドプロンプトから

```
mcmc
```

と入力することで実行される。

完了までにかかる時間は PC のスペックやデータの大きさによるが、64 ビットの Windows PC で、CPU が 3 G/Hz、データファイルは種数が 50 でサイト数が 600~700 の場合、2 時間程度が見込まれる。

実行中は図のような表示が MrBayes のウィンドウに表示される。

図は `printfreq` コマンドで 100 世代ごとの表示を設定してあるので、100 世代ごとに左端の数字に実行結果が表示されている。そして「Average stANDARD

```
Average standard deviation of split frequencies: 0.004824
15100 -- (-3417.359) (-3431.630) (-3413.415) [-3416.994] * (-3416.316) [-3413.564] (-3419.842) (-3416.235) -- 0:13:02
15200 -- (-3417.388) (-3420.072) (-3415.927) [-3420.017] * (-3412.829) (-3415.263) [-3420.131] (-3421.303) -- 0:12:57
15300 -- [-3415.636] (-3417.495) (-3417.443) (-3420.963) * (-3413.592) (-3412.704) (-3416.430) [-3414.142] -- 0:12:52
15400 -- (-3411.662) (-3420.219) (-3416.052) [-3415.197] * (-3411.961) (-3416.533) [-3412.886] (-3418.386) -- 0:12:47
15500 -- (-3411.518) [-3424.470] (-3414.651) (-3416.770) * (-3409.807) (-3418.225) (-3411.896) [-3411.404] -- 0:12:42
15600 -- [-3415.188] (-3426.097) (-3412.856) (-3415.403) * (-3411.956) (-3419.677) [-3411.005] (-3413.712) -- 0:12:37
15700 -- (-3416.699) [-3421.621] (-3414.342) (-3413.875) * (-3413.721) (-3420.574) [-3411.512] (-3417.144) -- 0:12:32
15800 -- [-3416.769] (-3426.859) (-3420.522) (-3416.628) * (-3419.897) [-3416.402] (-3414.778) (-3420.039) -- 0:12:27
15900 -- (-3415.278) (-3419.372) [-3412.810] (-3422.676) * (-3416.886) (-3412.423) [-3417.912] (-3420.437) -- 0:12:22
16000 -- (-3419.256) (-3422.055) (-3414.321) [-3417.648] * (-3415.652) (-3413.107) (-3430.797) [-3416.873] -- 0:12:18

Average standard deviation of split frequencies: 0.003247
16100 -- (-3418.898) (-3420.640) (-3414.762) [-3424.161] * (-3416.478) [-3414.796] (-3421.185) (-3422.397) -- 0:12:13
```

図 7 MrBayes 実行中の表示画面

deviation split frequencies」の値が解析の指標となる。世代が進むとともに、一般にその値は小さくなっていくが、この値が 0.01 より小さければ、非常に良好にマルコフ連鎖が定常状態に達したと判断され、0.01 から 0.05 の間であれば、解析の目的に応じて解析を終了するかどうか判断して良いとされている。

事前に `Ngen` コマンドで設定した世代数に世代数に達すると、

```
Continue with analysis? (Yes/no):
```

というメッセージが表示されるので、この「Average stANDARD deviation split frequencies」値が十分であれば `no` を入力して解析を終了する。

1.4.1 終了操作

コマンドプロンプトから

```
sump
```

と入力すると、データに対して設定したパラメータや解析に用いたパラメータ等の記録ファイル(拡張子.p) が作成される。

続いて同じくコマンドプロンプトから

```
sumt relburnin=yes burninfrac=0.25
```

と入力すると、系統樹等が作成される。作成された 50%合意樹は拡張子.con ファイルに保存される。また、「データファイル名」.nex.run1.t 及び「データファイル名」.nex.run2.t という 2 つの系統樹ファイルも作成される。前者は Cold Chain、後者は heat chain で作成された系統樹が、設定した解析世代数+1 の数保存されている。

系統樹は、「FigTree」「TreeGraph2」等のソフト⁵で各枝の事後確率等のデータもあわせて表示させることができる。

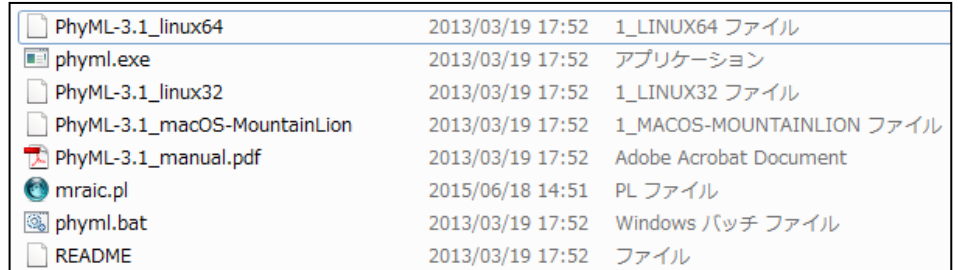
⁵ いずれも Java で書かれているので、事前に Java をインストールしておく。Java の HP からダウンロード及びインストールできる。

2 塩基の置換モデルと置換率の決定

MrBayes を実行するにあたって必要なモデルと置換率の決定は、PhyML⁶と組み合わせて用いる MrAIC を使うと便利である。

2.1 準備

PhyML をダウンロードしてする。ダウンロードファイルは Zip 形式で圧縮されているので、MrBayes の時と同様に、アーカイブソフトで解凍する。解凍すると 2015 年 12 月の時点では



| | | |
|------------------------------|------------------|---------------------------|
| PhyML-3.1_linux64 | 2013/03/19 17:52 | 1_LINUX64 ファイル |
| phyml.exe | 2013/03/19 17:52 | アプリケーション |
| PhyML-3.1_linux32 | 2013/03/19 17:52 | 1_LINUX32 ファイル |
| PhyML-3.1_macOS-MountainLion | 2013/03/19 17:52 | 1_MACOS-MOUNTAINLION ファイル |
| PhyML-3.1_manual.pdf | 2013/03/19 17:52 | Adobe Acrobat Document |
| mraic.pl | 2015/06/18 14:51 | PL ファイル |
| phyml.bat | 2013/03/19 17:52 | Windows バッチ ファイル |
| README | 2013/03/19 17:52 | ファイル |

PhyML-3.1 というフォルダが生じ、その中にマニュアルをはじめ数種のファイルが入っている。

図 8

続いて MrAIC をダウンロードして保存する

MrAIC の使用法は <https://www.abc.se/~nylander/mraic/mraic.html> に詳しく書いてあるが、MrAIC は Perl 言語で書かれているので、実行にあたっては使用する PC に ActivePerl がインストールされている必要がある。Perl のインストールに関しては <http://www.perlplus.jp/perlinstall/install/index1.html> に説明があるので参照しながらインストールする。

MrAIC 本体は、上記の使用法を説明したサイトから圧縮形式のもの（拡張子.gz）がダウンロードできる。これも保存したのち、解凍すると MrAIC143 というフォルダにフォルダ名と同じ MrAIC143 というファイルができていますので、このファイル名を mraic.pl に変更（するとファイルのアイコンも変わる）。続いて、このファイルを先の PhyML フォルダに移す。以上で準備完了だが、作業の結果として PhyML フォルダ内には、図 8 のようなファイルがあるはずである。

2.2 データファイルの用意

MrAIC を用いて、データファイルの塩基配列の置換モデルの決定を行う場合、データファイルの形式はフィリップ形式（拡張子.ph）である必要がある。また、データが複数のパーティションで構成されている場合⁷は、パーティションごとに ph 形式のファイルを作成する必要がある。

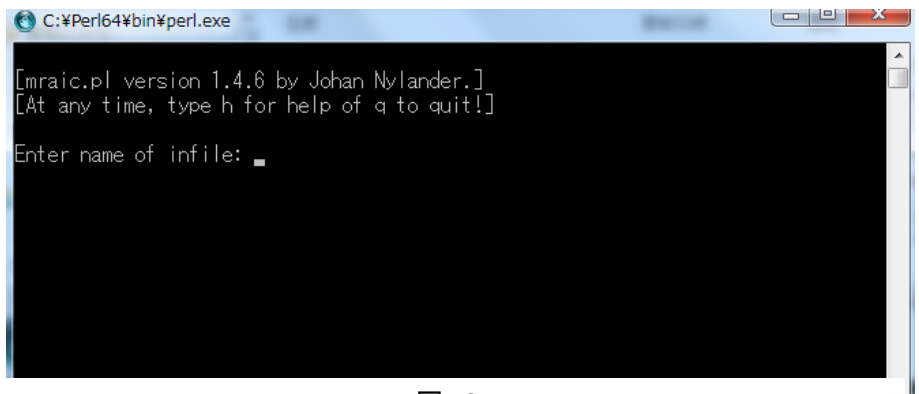


図 9

Fasta 形式や nexus 形式のデータファイルを ph 形式に変換する方法は複数あるが、Fasta 形式のファイルを ClustalX2 で読み込み、ph 形式で出力することが簡便であるように思う。

⁶ IFB (French National institute of Bioinformatics)が提供する ATGC のホームページ (<http://www.atgc-montpellier.fr/phyml/binaries.php>) からダウンロードできる。

⁷ MrBayes では、複数のパーティションでの解析が可能。「1.6 GAP コーディング」参照。

2.3 MrAICの実行

完成した ph 形式のファイルを MrAIC で解析実行するには、図 8 の mraic.pl をダブルクリックする。すると、図 9 のようなウィンドウが立ちあがるので、「Enter name of infile:」のあとにパス名付きで作成した ph 形式のデータファイル名を入力する。

ただ、WindowsPC を使用し、デスクトップに PhyML-3.1 フォルダを置いた場合は、データファイルのパス名がややこしくならないよう、PhyML-3.1 フォルダ内にデータファイルを置いた方が良い様に思う。

実行中は、様々なモデルで ML 法で系統樹成を試みている様子が表示されている。実行時間は PC のスペックやデータの大きさによるが、50~60 種で 800 サイト程度の大きさのデータで 2 時間程度かかる。

2.4 実行結果

図 10 は rps2.ph というデータファイルで MrAIC を実行した結果である。PhyML-3.1 のフォルダ内の先の図

| | | | |
|----------------------|------------------|-------------|-------|
| rps2.ph | 2015/09/11 10:59 | PH ファイル | 13 KB |
| rps2.ph.AICc-GTR.tre | 2015/09/11 11:05 | TRE ファイル | 1 KB |
| rps2.ph.AIC-GTR.tre | 2015/09/11 11:05 | TRE ファイル | 1 KB |
| rps2.ph.BIC-K2P.tre | 2015/09/11 11:05 | TRE ファイル | 1 KB |
| rps2.ph.MrAIC.txt | 2015/09/11 11:05 | テキスト ドキュ... | 5 KB |

図 10

9 に示したファイルの他に 3 種のモデル (AICc,AIC,BIC)で作成した TRE ファイルと 1 つのテキストファイルが生成している。実行結果はこのテキストファイルにまとめられている。

右の図 11 は、このテキストファイルの最後の部分を表示したものだが、ここに Mr Bayes を実行するとき、モデル等の事前設定をどうするか、まとめられている。例えば、AIC モデルで得られた値で Bayes 解析の設定をするならば、図の□で囲んだ部分を MrBayes のコマンドプロンプトから入力する。この場合は、2 回 (Lset と Prset コマンド) 入力する。applyto=(1)というのは、ここで解析したデータがパーティション 1 の場合を前提として、Lset あるいは Prset コマンドをこのパーティションに適用する、という意味で、実際には MrBayes で解析するデータファイルで、この rps2 遺伝子データがパーティション 2 にある、ということもある。そのような場合は、Lset(あるいは Prset) に続いて applyto=(2)としなければならない。

```
rps2.ph.MrAIC.txt - メモ帳
ファイル(F) 編集(E) 書式(O) 表示(V) ヘルプ(H)
To specify default values for best AIC, AICc, or BIC models in MrBayes:
[Mrbayes block for the best AIC model (GTR)]
BEGIN MRBAYES;
  Charset rps2.ph = 1 - 487;
  Partition Dummy = 1:rps2.ph;
  Set partition = Dummy;
  Lset applyto=(1) nst=6 rates=equal;
  Prset applyto=(1) revmatpr=Dirichlet(1.0,1.0,1.0,1.0,1.0,1.0)
  statefreqpr=Dirichlet(1.0,1.0,1.0,1.0);
END;

[Mrbayes block for the best AICc model (GTR)]
BEGIN MRBAYES;
  Charset rps2.ph = 1 - 487;
  Partition Dummy = 1:rps2.ph;
  Set partition = Dummy;
  Lset applyto=(1) nst=6 rates=equal;
  Prset applyto=(1) revmatpr=Dirichlet(1.0,1.0,1.0,1.0,1.0,1.0)
  statefreqpr=Dirichlet(1.0,1.0,1.0,1.0);
END;

[Mrbayes block for the best BIC model (K2P)]
BEGIN MRBAYES;
  Charset rps2.ph = 1 - 487;
  Partition Dummy = 1:rps2.ph;
  Set partition = Dummy;
  Lset applyto=(1) nst=2 rates=equal;
  Prset applyto=(1) statefreqpr=Fixed(Equal) tratioopr=Beta(1.0,1.0);
END;
```

図 11

他の AICc や BIC モデルの場合についても同じである。

また、ここでの設定内容をいちいちコマンドプロンプトから手入力するのは面倒であり、誤りもありうるので、MrBayes で解析するデータファイル内に Begin MRBayes ではじまる Bayes ブロックを作成しておき、そこにコピー&ペーストしておいても良い。

3 GAP コーディング

3.1 ダウンロードとインストール

塩基配列に欠失や挿入 (Indels)があると、アライメントの後に GAP が生じるが、SeqState⁸を使えば、GAP をコード化して系統樹に反映させることができる。

SeqState の HP からダウンロードしてインストールする (図 12)。

ダウンロードしたファイルは SeqState.zip という名称で、zip 形式で圧縮されているので、解凍して用いる (図 13)。

なお、SeqState

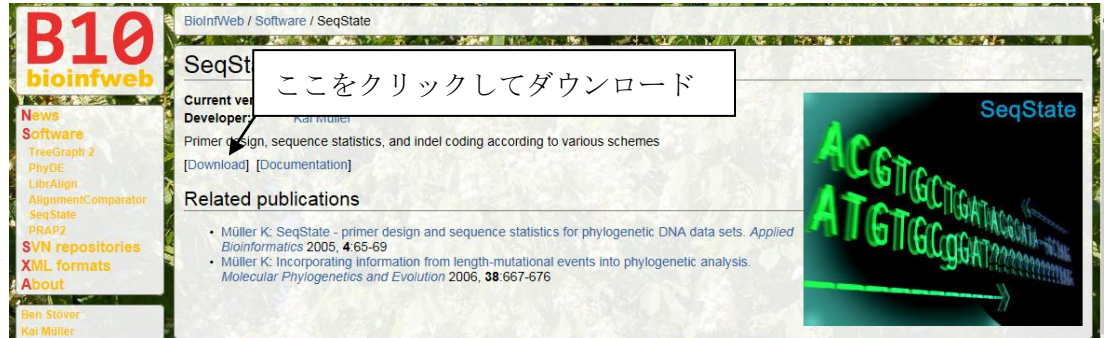


図 12

| | | | |
|-----------------------|------------------|--------------------|----------|
| SeqState.dsc | 2008/03/06 19:54 | DSC ファイル | 1 KB |
| SeqState.jar | 2009/10/03 0:17 | Executable Jar ... | 1,139 KB |
| ValidNexusExample.nex | 2012/02/08 11:26 | NEX ファイル | 1 KB |

図 13

は java で書かれたプログラムなので、その実行には予め PC に java がインストールされている必要がある。

3.2 実行

解凍して生じたフォルダ内の SeqState.jar をダブルクリックすると、図 14 のようなウィンドウが立ち上がる。

ここで、MrBayes で解析するデータファイルを読み込むが、この時点では FASTA 形式のものを用意しておき、メニューの [File] から [LOAD FASTA FILE] を選択して、そのデータファイルを読み込む。

続いて、メニューの [IndelCoder] をクリックし、IndelCoding のメニューから必要なものを選択してクリックすれば、その形式でアライメントファイル中の GAP が 1、0、? の 3 つのパラメータでコード化され、nexus 形式の [ファイル名]_sic.nex という名称で保存される。しかし、出来上がったファイルはそのままでは MrBayes ではエラーが出て実行できないので、Notepad や Wordpad など

のテキストエディタで修正を加える必要がある。



図 14

⁸ アドレスは <http://bioinfweb.info/Software/SeqState>

なお、複数の遺伝子データを Bayes 解析する場合は、その遺伝子ごとに FASTA ファイルを用意して GAP コーディングし、1つの遺伝子について塩基配列のパーティションと GAP コード化したパーティションの2パーティションとする。例えば遺伝子 A と B の2つで系統樹を作成したい、という場合は、データファイルは以下のように4つのパーティションを持つことになる。

- パーティション 1 遺伝子 A の塩基配列
- パーティション 2 遺伝子 A の GAP コード
- パーティション 3 遺伝子 B の塩基配列
- パーティション 4 遺伝子 B の GAP コード

3.3 Nexus 形式のデータファイルの修正

下の図 15 は、SeqState で simple coding し、出力されたファイルの先頭部分である。

```
#NEXUS
[matrix with indels coded according to Simmons & Ochoterena (2000) - simple coding]
[compiled on Fri Jun 26 13:28:32 JST 2015 with SeqState version 1.4.1 build 02.10.2009]

BEGIN CHARACTERS;
DIMENSIONS newtaxa ntax=42 NCHAR=675;
FORMAT
  DATATYPE = standard
  GAP = -
  MISSING = ?
  symbols="0123456789A#C$EFG.IJ&L%>OPQ/TU:*X<Z"
  equate="R={AG} Y={CT} M={AC} K={GT} S={CG} W={AT} H={ACT} B={CGT} V={ACG} D={AGT} N={ACGT} r={AG} y={CT} m={AC} k={GT} s={CG} w={AT} h={ACT} b={CGT} v={ACG} d={AGT} n={ACGT}"
INTERLEAVE;
CHARSTATELABELS
646 indel1_10_11
647 indel2_11_11
648 indel3_40_40
649 indel4_53_54
650 indel5_62_62
651 indel6_73_73
652 indel7_93_95
653 indel8_104_114
```

図 15

この部分に関しては以下のように修正する。(太字部分)

```
BEGIN DATA;
DIMENSIONS ntax=42 NCHAR=675;
FORMAT
  DATATYPE = mixed(dna:1-645,standard:646-675)
  GAP = -
  MISSING = ?
  [symbols="0123456789A#C$EFG.IJ&L%>OPQ/TU:*X<Z" ]
  [equate="R={AG} Y={CT} M={AC} K={GT} S={CG} W={AT}
H={ACT} B={CGT} V={ACG} D={AGT} N={ACGT} r={AG} y={CT} m={AC} k={GT} s={CG} w={AT} h={ACT}
b={CGT} v={ACG} d={AGT} n={ACGT}"]
INTERLEAVE = yes;
[CHARSTATELABELS
  646 indel1_10_11
  647 indel2_11_11
```

"newtaxa"を除く

塩基配列部分は dna、Gap コード部分は standard

symbols 行と equate 行を [] でコメントアウト

CHARSTATELABELS ブロックを [] でコメントアウト

```
648 indel3_40_40
649 indel4_53_54
650 indel5_62_62
651 indel6_73_
```

<以下略>

また、GAP コードよりも後ろの部分に、
BEGIN PAUP;
END;
と入っている。

ここを、以下の例のように PAUP を MRBAYES に変え、**BEGIN MRBAYES;**から **END;**までの MRBAYES ブロックとして、この間にパーティションごとの置換モデルの設定コマンドや解析実行にあたっての事前設定を入れることができる。

```
begin mrbayes;
  lset applyto=(1) nst=6 rates=gamma;
  lset applyto=(3) nst=6 rates=equal;
  Prset applyto=(1)
    revmatpr=Dirichlet(1.0,1.0,1.0,1.0,1.0,1.0)
    statefreqpr=Dirichlet(1.0,1.0,1.0,1.0) shapepr=Uniform(0.1,50.0);
  Prset applyto=(3)
    revmatpr=Dirichlet(1.0,1.0,1.0,1.0,1.0,1.0)
    statefreqpr=Dirichlet(1.0,1.0,1.0,1.0);
  mcmc ngen=2000000;
  mcmc diagnfreq=1000 samplefreq=100 printfreq=100;
  mcmc relburnin=yes burninfrac=0.25;
end;
```

パーティション 1 について MrAIC の結果からコピー

パーティション 3 について MrAIC の結果からコピー

パーティション 1 について MrAIC の結果からコピー

パーティション 3 について MrAIC の結果からコピー

MrBayes の実行パラメータ

なお、GAP コードのパーティションについては、デフォルト設定のままで良い。

また、当然であるが、このように MRBAYES ブロックとせず、**BEGIN PAUP;**から **END ;** までを削除してしまい、必要なパラメータは MrBayes を起動し、ファイルを execute コマンドで読み込んだ時点で、コマンドプロンプトから入力することもできる。