

- 1 例題 6.2
- 2 6.4 節と 6.5 節で紹介したハチドリのミトコンドリア全タンパク質コード遺伝子の解析を
- 3 実践せよ。

4



5

6 **ホバリング飛行により花の蜜を吸うハチドリの 1 種(*Eulampis jugularis*)**

- 7 ホバリング飛行 (高速ではばたくことで空中で静止する飛行) を行うハチドリは動物の中で
- 8 最も代謝率が高い分類群のひとつです。好気呼吸によりエネルギーを生産するミトコンド
- 9 リアはハチドリが活発な代謝を維持するために特殊な進化を遂げてきた可能性があります。
- 10 ここではハチドリのミトコンドリアゲノムの進化のテンポとモードを調べるために系統樹
- 11 推定、選択圧の推定、分岐年代推定を行いたいと思います。

12 *Eulampis jugularis* の写真は Wikipedia(<https://en.wikipedia.org/wiki/Hummingbird>)より

13

1 **解説編**

2 ここでは6章4節「選択圧の推定」と5節「分岐年代推定」で紹介したハチドリのミト
3 コンドリア全タンパク質コード遺伝子の解析を実際に行ってみたいと思います。なおこの
4 例題はおもにWindowsユーザーを想定していますが、ここで紹介するプログラムはMacや
5 LINUXでも使用可能です。

6 なお本解説で推定される尤度やパラメータは、皆さんが推定した値と小数点以下数桁の
7 ところで若干異なるかもしれませんが問題ありません。

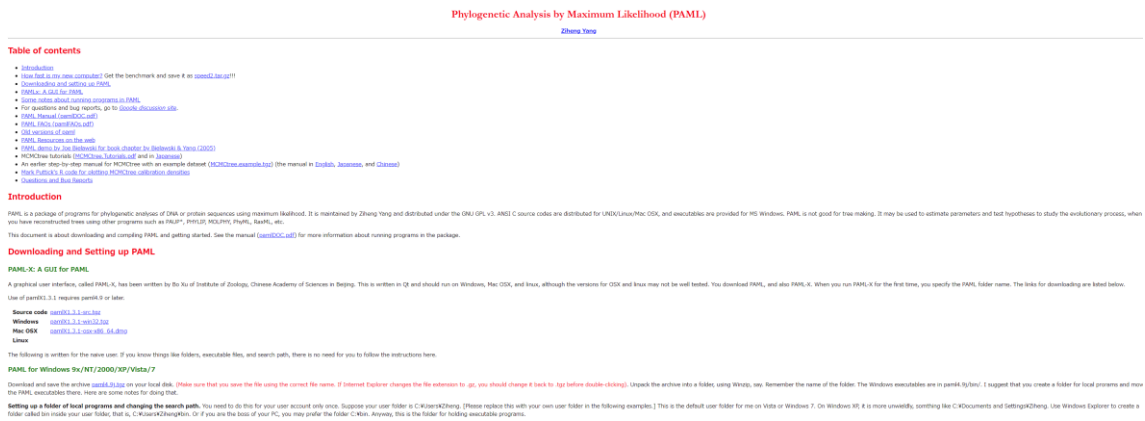
8

9 **序章：準備**

10 **1節. PAML**

11 解析は主にPAMLパッケージを用います。ロンドン大学のZiheng Yang教授のウェブサイ
12 トからPAMLをダウンロードしましょう。

13 <http://abacus.gene.ucl.ac.uk/software/paml.html>



14 ※ここではPAML-X: A GUI for PAMLではなく、コマンドラインバージョンを用います。
15 PAML for Windows 9x/NT/2000/XP/Vista/7 から最新版(paml4.9j.tgz: 2022年2月7日現在)
16 をダウンロードし、Lhaplus等を用いて解凍して用います。

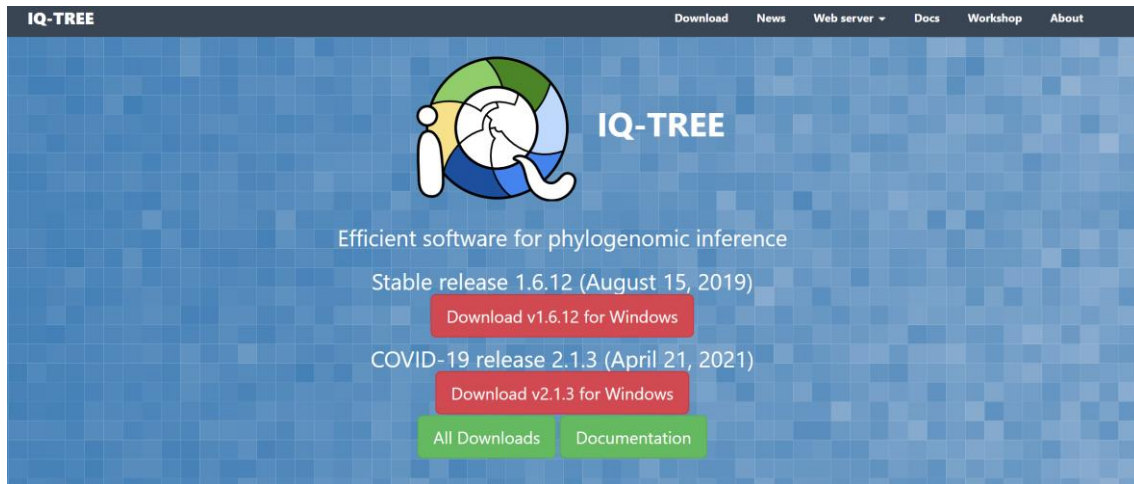
17

18

19 **2節. IQ-TREE2**

20 PAMLでの解析に先立ちIQ-TREE2プログラムを用いた最尤法による系統樹推定を行います。
21 オーストラリア国立大学のBui Quang Minh博士らによるウェブサイトからIQ-TREE2
22 をダウンロードしましょう。

23 <http://www.iqtree.org/>



1

2 2022年2月7日現在、最新版は version 2.1.3(2021年4月21日リリース)です。

3 Download v2.1.3 for Windows と書かれた赤いボタンをクリックすると、qtree-2.1.3-
4 Windows.zip というファイルがダウンロードされます。これを解凍して用います。

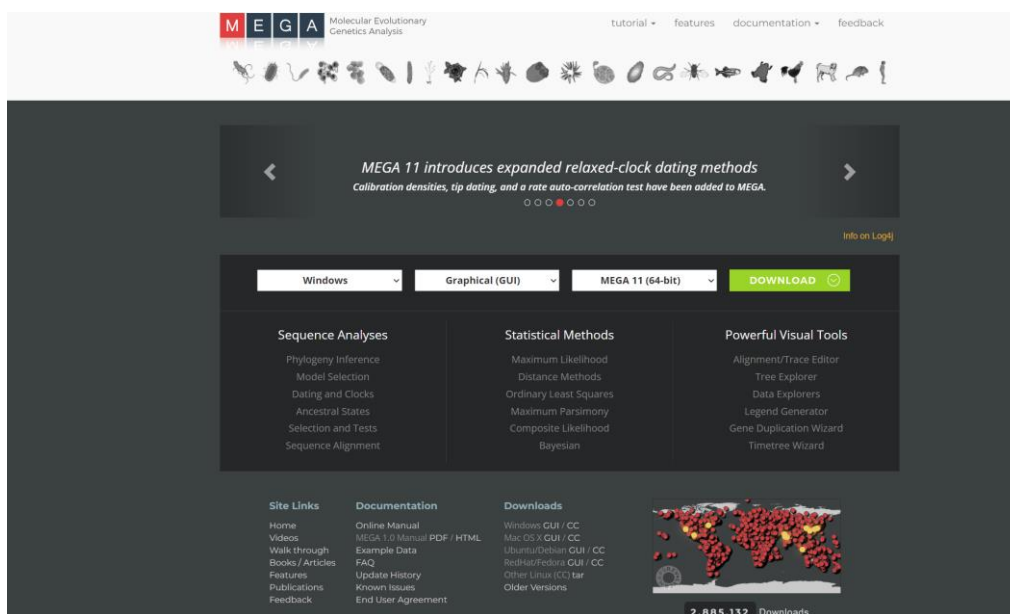
5

6 3節. MEGA

7 系統樹を描画する際に MEGA を用いると非常に便利です。東京都立大学の田村浩一郎教授
8 らによる MEGA の公式からダウンロードしましょう。

9 <https://www.megasoftware.net/>

10 MEGA は分子進化学及び集団遺伝学に関する様々な解析を実行でき、非常にユーザーフレ
11 ンドリーで定評のあるプログラムです。この例題では扱いませんが、Tajima's D (6.4.1「集
12 団遺伝学的アプローチによる選択圧の推定」参照)や RELTIME 法(6.5.2「最尤法による局
13 所的分子時計と緩和型分子時計」参照)などは、この MEGA プログラムにより実行できま
14 す。最新版は MEGA 11 です (2022年2月7日現在)。



1 2 第一章：系統樹推定

3 ハチドリのミトコンドリア全タンパク質コード遺伝子から最尤系統樹を推定してみよう

4 5 1 節：準備

6 IQ-TREE2 プログラムを使い最尤系統樹を推定します。

7 ここで必要なものは①IQ-TREE2 プログラムの実行ファイルと②アラインメントファイ
8 ル、そして必要に応じて③パーティションファイルです。

9 10 ① 実行ファイル

11 IQ-TREE のサイトからダウンロードした iqtrees-2.1.3-Windows.zip をデスクトップ上に解
12 凍すると、bin というフォルダと 4 つの例題ファイルが入っているのが分かります。

名前	更新日時	種類	サイズ
bin	2022/02/07 14:12	ファイルフォルダー	
example.cf	2022/02/07 14:12	CF ファイル	2,185 KB
example.nex	2022/02/07 14:12	NEX ファイル	1 KB
example.phy	2022/02/07 14:12	PHY ファイル	34 KB
models.nex	2022/02/07 14:12	NEX ファイル	119 KB

13 14 この bin というフォルダの中には実行ファイルが入っています。

名前	更新日時	種類	サイズ
iqtrees2.exe	2022/02/07 14:12	アプリケーション	9,761 KB
iqtrees2-click.exe	2022/02/07 14:12	アプリケーション	9,761 KB
libiomp5md.dll	2022/02/07 14:12	アプリケーション拡張	1,089 KB

15 16 17 ② アラインメントファイル

- 1 上記のフォルダの中にハチドリのみトコンドリア遺伝子のアラインメント
- 2 (*hummingbird_12mtCDS.fas*)を入れます。このファイルは、ミトコンドリアゲノムの H
- 3 鎖にコードされている 12 のタンパク質コード遺伝子を整列したうえで連結したもので
- 4 す。開始コドンや停止コドン、遺伝子間のオーバーラップ領域を除外しており、10,758 塩
- 5 基対 (3,586 コドン) の配列長です。このアラインメントファイルは FASTA 形式と呼ばれ
- 6 るフォーマットで作成しています。それぞれの配列名は学名と NCBI のアクセッション番
- 7 号を表しています。

```

hummingbird_12mtCDS.fas - メモ帳
ファイル(F) 編集(E) 書式(O) 表示(V) ヘルプ(H)
>Aesothelus cristatus_EU844979
TTC TTT GAC CAA TTT GCA AGC CCA AGC CTA ATG GGA ATC CCC TTG ATC CTA CTC TCC TTA TTA TTT CCT GCA CTA CTC CTC CCC ACG CCG GAC AAC CGA TGA CTC ACC AAG
AAC CCA CCA TCC AAC ATA AAC TTA ACA --- AGC ACT AAA GGC ACC TOC TGA GTC TGA ACC TTC ATT AAC CGA TGA TTA TTT TCA ACT AAC CAC AAA GAC ATT GGC ACC CT/
TTC GGA CAC CCA GAA GTC TAT ATC CTC ATC CTA CCA GGC TTT GGA ATC ATC TCA CAT GTA GTC GGC TAC TAC GCA GGA AAA AAA GAG CCA TTT GGC TAC ATA GGA ATA GT/
TGA ATC CAC GGC TGC CCG CCC CCA TAC CAC ACC TTT GAA GAA COG GOC TTC GTC CAA ATT CAA GAA GGC AAC CAC TCA CAA CTA GAA GAC GGC TOC TCC CCC ATC
ATA GTA GAC CCC AGC CCA TGA CCA ATC TTC GGA GCA ATC GCT GCT CTC CTT ACC ACT TCA GGC TTA GGC ATC TGA TTC CAC CAT AAC TCC ACC CAA CTT CTT AGC CTA G/
CGA AAG TCC CAC CCC CTA AAG ATA ATC AAC AGC TOC CTA ATC GAC CTT CCC ACT CCA TCA AAC ATC TCT GGC TGA TGA AAC TTC GGA TCT CTC CTA GGC ATC TG/
CCA CTA GTT ACT CCG CCA CAT ATC AAA CCA GAA TAC TTC TTA TTT GCA TAT GGC ATC CCA CGT TCA ATC CCT AAC AAA CTC GGA GGA GTG CTA GGC CTT GGC GGC T/
ATT GGG GGC CCA CCC CTA AAG ATA ATC AAC AGC TOC TAT GAA GTC ACA CTA GAT ATC ATC CTT TTA TCG ATC ATC CTA CTA AGC GGA AAC TAC TCT CTA AGC ACA CTA G/
CTC CTC TTA TTC TOG GGC ATA ATC AAT GGC TGA TCA AGA CAA TGG GAC ATT ACC CAA CTA ACC CAC CCA ACA TCA TOC ATC CTG TTA ATC CTA GCA GCA ATT GCA CTA A/
ATT GCT ATC CTC AGC TCC CTA ATC ATC CTA CTA CCA ATC TOC TAT GAA GTC ACA CTA GAT ATC ATC CTT TTA TCG ATC ATC CTA CTA AGC GGA AAC TAC TCT CTA AGC ACA CTA A/
AOC CTA ATC CTA ATC ACA GGA TGA GGA AGC CAA TCA GAA CGC CTA AGC GGC GGC ATT TAC CTT CTA TTC TAC ACT CTT ATC AGC TOC CTC CCT CTA CTA ACT TCT ATC CT/
ACA CTA TTC AAC TGA TCC CCA ACC ATC ATC CTA ACC GGA ACC GTA ACC TTC CTA ACC GGC TOC TAC ACC TTA TTC ATA CTC TTG ACC ACA CAA GGA ACC CTA CC/
TTC TTC CCT ATC GGC CTA TTT GTA ACC TGA TCA ATC CTT CAA TTC GCA ACA TGG TAC ATG GGC TCA GAA CCA CAC ATT ACA AAA TTC TTC CTC TAC TTG CTA CTG TTC CT/
ATC ATC CAC AGC CTC AAT GGT GAG CAA GAC ATT CGA AAA ATA GGC GGC CTA CAA AAA ACA CTC CCA ACA ACC ACA TCA TGC CTA ACC ATT GGC AAT CTT GCA CTA G/
ACA CAA
>Amazilia brevirostris_KP722043
TTC TTT GAC CAA TTT ATA AGC CCC TAC CTA CTC GGA ATT CCC CTA ATC CTT ATC TCT TTA CTA TTC CCA ACC CTA CTA TTC CCC TOC CCC GGC AAC CGA TGA ATC ACA AAK
AAC GGC CCC TCA GAG AAA GGC CCA ACA --- AGC ACA AAA ACC CCA TOC TGA CCC TGA ACT TTT ATC AAT CGA TGA TTA TTT TCA ACC AAC CAC AAG GAC ATT GGC ACC CT/
TTT GGC CAC CCT GAA GTT TAC ATC CTC ATC CTA CCA GGA TTC GGA ATC ATC TCA CAT GTG GTA ACA TAC TAC ACC GGC AAA AAG GAA CCA TTC GGC TAC ATG GGC ATG GT/
TGA ATC CAC GGC TGC CCT CCC CCA TAC CAC ACC TTT GAG GAG CCT GOC TTT GTC CAA ATC CAA GAA GGC AAC CAC TOC CAA CTA GGA TTC CAA GAC GCT TCA TOC CCA AT/
ATA GTA GAC CCA AGC CCA TGA CCC CTA TTC GSA GCA GGC GGC GGC TTA CTC ACC ACC TGA GCA CTC ATC ATA TGG TTC CAC CAC AAC TOC GCA ACC CTA TTG TOC TTA G/
CGT AAA TOC CAC CCC CTG CTA AAG ATA ATC AAC GAC TCA CTA ATC GAC CTC CCA ACC CCA TCA ACC ATC TCA GOC TGA TGA AAC TTC GGA TCC CTC CTA GGC CTA TG/
CCC CTA GTA ACA CCT CCC CAC ATC AAA CCA GAA TGG TAC TTC CTA TTC GCA TAC GCT ATC CTC CCA TCA AIT CCA AAC AAA CTG GGA GGA GTC TTG GGC TTA GGC GGC TC/
ATT GGG GGC CTC CGA GCA GTA GGC CAA ACT ATT TOC TAT GAA GTG ACT CTC GCT ATC ATC CTT CTC TOG CTA AIT ATT TTA ACC GGA AAC TAC ACC TTG AAC ACC CTA G/
TTA CTA TTA TTC TCA AGC ATA GTC AAT GCT TGA TAC ACA GGA CAA TGA GAT ATT ACA CAA CTA AAC CAC CCC ACA TCA TOC CTA CTA CTA ACT GCA GCA ATC GCA ATA A/
ATC GGC ACA CTC TOC TOC ATA TCA ACC CTC CTA CTC CCA CTA TOC CCC ATA ATC CTT GGC TCC ATC CTC ACC CTA TCA CTC GCA CTC ACT GCT CTC CTA GTC ACA CTA AAK
ACC CTA ATC CTA ATC ACA GGA TGA GGT AAC CAA CCA GAA CGC CTA AGC GGC GGC ATC TAC CTA TTA TTC TAT ACC CTA ATC AGC TOC CTC CCC CTA CTA GTT GCA ATC TT/
GOC TTA TTC AAT TGA TCC CCT CTC ACC ATC CTC ATC ACC GGA ACC GCA ACA TTC CTA ACA GGC TCA TAC ACC CTA TTT ATA TTT ACC ACT CCA CAA GGA GGC CCA CTA CC/
TTC CTA CCC ATT GGC CTA TTT GTA ACA TGG TCT ATC CTT CAA TTC GCA TCA TGA TAC ATA TCA TCA GAC CCC CAT ATC ACA AAA TTC TTT TCA TAT CTC CTA ACA TTT TT/
ATT ATT CAT AGC CTA AAT GGT GAG CAA GAC ATC CGG AAA ATA GGA GGA TTG CAA AAA ACA CTC CCG ACA ACA GGC TOC TGC CTA ACC ATC GGC AAC CTT GGC CTA ATA G/
TCA ATA
>Amazilia millerii_KP722042
TTC TTT GAC CAA TTT ATA AGC CCC TAC CTA CTC GGA ATT CCC CTA ATC CTC ATC TCT TTG CTA TTC CCA ACT CTG CTA TTC CCC TOC CCA GGT AAC CGA TGA ATC ACA AAK
AAC ACC CCC TCA GAA AAA GGC CCA ATA --- AGC ACA AAA ACC CCA TOC TGA CCC TGA ACT TTT ATC AAT CGA TGA TTA TTT TCA ACC AAC CAC AAG GAC ATT GGC ACC CT/
TTT GGC CAT OCT GAA GTT TAC ATC CTC ATC CTA CCA GGA TTC GGA ATC ATC TCA CAT GTA GTC ACA TAT TAC ACC GGT AAA AAG GAA CCA TTC GGC TAC ATG GGC ATG GT/
TGA ATC CAC GGC TGC CCT CCC CCA TAC CAC ACC TTT GAG GAG CCT GOC TTT GTC CAA ATC CAA GAA GGC AAC CAC TOC CAA CTA GGA TTC CAA GAC GCT TCA TOC CCA AT/
ATA GTA GAC CCA AGC CCA TGA CCC CTA TTC GSA GCA GGC GGC GGC TTA CTC ACC ACC TGA GCA CTC ATC ATA TGG TTC CAC CAC AAC TOC GCA ACC CTA TTA TOC TTA G/
CGT AAA TOC CAC CCA CTG CTA AAG ATA ATC AAC GAC TCA CTA ATC GAC CTC CCA ACC CCA TCA ACC ATC CTC CCA TOC TCA ACC TGA AAC TTC GGA TCC CTC CTA GGC CTA TG/
CCC CTA GTA ACA CCT CCC CAC ATC AAA CCA GAA TGG TAC TTC CTA TTC GCA TAT GCT ATC CTC CCA TOC ATC CCA AAC AAA CTG GGA GGA GTC TTG GGC CTA GGC GGC TC/
ATT GGG GGC CTC CGA GCA GTA GGC CAA ACT ATT TOC TAT GAA GTG ACT CTC GCT ATC ATC CTT CTC TOG CTA AIT ATT TTA ACC GGA AAC TAC ACC TTG AAC ACC CTA G/
TTA CTA CTA TTC TCA AGC ATA GTC AAT GCT TGG TAC ACA GGA CAA TGG GAC ATT ACA CAA CTA AAC CAC CCC ACA TCA TGC CTA CTA CTA ACT GCA GCA ATC GCA ATA A/
ATC GGC ACA CTC TOC TOC ATA TCA ACC CTC CTA CTC CCA CTA TOC CCC ATA ATC CTT GGC TTC ATC CTC ACC CTA TCA CTC GCA CTC ACT GCT CTC CTA GTC ACA CTA AAK
ACC CTA ATC CTA ATC ACA GGA TGA GGT AAC CAA CCA GAA CGC CTA AGC GGC GGC ATC TAC CTA TTA TTC TAT ACT CTA ATC AGC TOC CTC CCC CTA CTA GTT GCA ATC TT/
GCA TTA TTC AAT TGA TCC CCT CTC ACC ATC CTC ATC ACC GGA GGC GCA ACA TTC CTA ACA GGC TCA TAC ACC CTA TTC ATA TTC ACC ACT ACC CAA GGA GGC CCA CTA CC/
TTC CTA CCC ATT GGC CTA TTT GTA ACA TGG TCT ATC CTT CAA TTC GCA TCA TGA TAC ATA TCA TCA GAC CCC CAC ATC ACA AAA TTC TTT TCA TAT CTC CTA ACA TTT TT/
<
Windows (CRLF) 1行、1列 100%

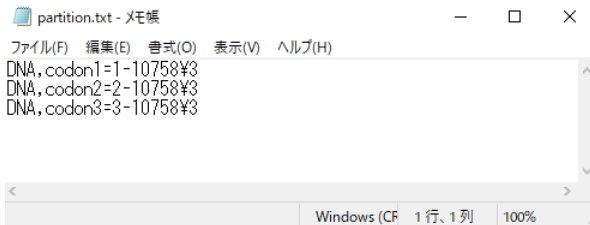
```

- 8
- 9 このアラインメントファイルはコドンとコドンの間を半角スペースで開けていますが、こ
- 10 の半角スペースの有無はこの後の解析には影響ありません。
- 11
- 12 ③ パーテーションファイル (オプション)
- 13 IQ-TREE2 プログラムは、系統樹推定を行う際に遺伝子領域ごとの進化速度や塩基組成の
- 14 違いなどを考慮したパーテーションモデルを用いることができます。パーテーションモデ
- 15 ルの詳細な説明については、橋本ほか(2008)の優れた研究詳解をご参照いただきたいと思
- 16 いますが、ここではコドンを構成するトリプレットのうち、1 番目の塩基、2 番目の塩
- 17 基、3 番目の塩基を区別して解析したいと思っています。
- 18

1 橋本哲男、有末伸子、坂口美亜子、稲垣 祐司 (2008) 複数遺伝子の結合データ
2 に基づく分子系統樹の推測 -真核生物の大系統の解析を例として-. 統計数理, 56:
3 145-164

4

5 パーテーションファイル(partition.txt)は以下のようになっています。



6

7 一行目の

8 **DNA,codon1=1-10758¥3**

9 は、DNA 配列である **codon1** という名前のパーテーションが、このアラインメントの 1 塩
10 基目からはじまり、10758 塩基目まで 3 塩基ごとにサンプリングされる塩基から構成され
11 ることを示しています。

12 同様にして

13 **DNA,codon2=2-10758¥3**

14 は 2 塩基目からはじまり、10758 塩基目まで 3 塩基ごとにサンプリングされる塩基から

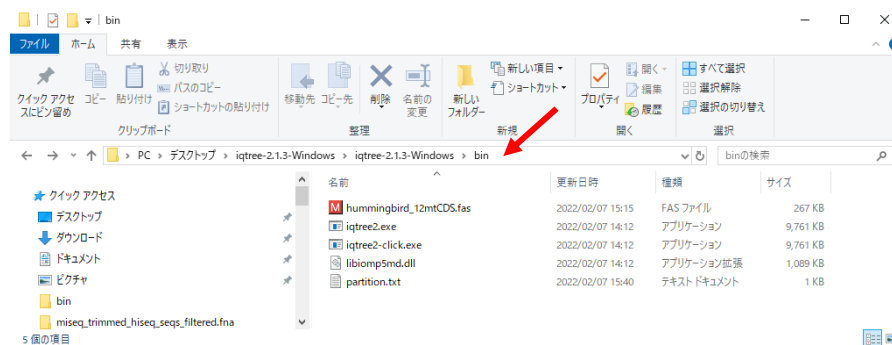
15 **DNA,codon3=3-10758¥3**

16 は 3 塩基目からはじまり、10758 塩基目まで 3 塩基ごとにサンプリングされる塩基から
17 それぞれ構成されていることを意味しています。

18

19 2 節：系統樹推定

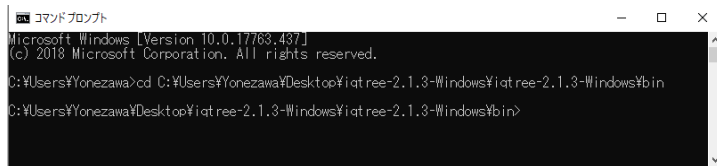
20 アラインメントファイルとパーテーションファイルを bin のフォルダにコピーしたら、コ
21 マンドプロンプトを起動し bin のフォルダへ移動します。



22

23 エクスプローラーにあるアドレスバー（上図の赤い矢印）をクリックするとフォルダーの
24 場所を表すアドレスのような文字列（パス）が表示されるので、それをコピーします。

1 コマンドプロンプトにカレントディレクトリの変更を意味する cd とタイプし、半角スペースをあけて上述のパスをペーストし、Enter キーを押すと bin のフォルダに移動できます。



```
コマンドプロンプト
Microsoft Windows [Version 10.0.17763.437]
(c) 2018 Microsoft Corporation. All rights reserved.
C:\Users\Yonezawa>cd C:\Users\Yonezawa\Desktop\iqtree-2.1.3-Windows\iqtree-2.1.3-Windows\bin
C:\Users\Yonezawa\Desktop\iqtree-2.1.3-Windows\iqtree-2.1.3-Windows\bin>
```

4
5

6 bin のフォルダへと移動したら以下のコマンドをタイプしましょう。

7

8 `iqtree2 -s hummingbird_12mtCDS.fas -m TEST -p partition.txt -B 1000 --prefix`
9 `hummingbird -T 4`

10

11 “iqtree2”は IQ-TREE2 プログラムを起動するコマンドです。

12

13 `-s` はアラインメントを指定するオプションで、半角スペースを空けてアラインメントファイル名（この場合 `hummingbird_12mtCDS.fas`）を入力します。

14

15 `-m` はモデルを指定するオプションです。

16 GTR+I+ Γ など前もって使いたいモデルが決まっている場合は

17

`-m GTR+F+I+G`

18 と入力しますが、ここにモデル名の代わりに **TEST** と入力すると IQ-TREE2 に搭載されている ModelFinder(Kalyaanamoorthy et al. 2017)というプログラムが起動し、ベイズ情報量規準(BIC)に基づきベストモデルを自動的に選択してくれます。パーテーションファイルがある場合は、パーテーションごとにベストモデルを選択します。

19

20 S. Kalyaanamoorthy, B.Q. Minh, T.K.F. Wong, A. von Haeseler, L.S. Jermini
21 (2017) ModelFinder: Fast model selection for accurate phylogenetic estimates.
22 Nat. Methods, 14:587-589. <https://doi.org/10.1038/nmeth.4285>

23

24 `-p` はパーテーションを指定するオプションで、半角スペースを空けてパーテーションファイル名（この場合 `partition.txt`）を入力します。パーテーションを使わずに解析する場合は、このオプションは不要です。橋本ほか(2008)に記載されているようにパーテーションモデルは様々な方法がありますが、`-p` というオプションを用いた場合、系統樹における各々の枝長は、パーテーション間で比例しているという仮定を置く「比例モデル」が用い

1 られます。一方、**-p** の代わりに**-Q** というオプションを用いると、個別のパーティション
2 それぞれについて独立に枝長の推定を行う「分離モデル」が適用されます。

3
4 **-B** というオプションは、ultrafast bootstrap 法によりノードの信頼性を評価するのに用い
5 ます。半角スペースをあけて試行回数(この場合 1000 回)を記します。Ultrafast bootstrap
6 法によるブートストラップ値はしばしば過大評価になることが知られています。この問題
7 を解決するために**-bnni** というオプションで ultrafast bootstrap 法で得られた系統樹を hill-
8 climbing nearest-neighbor interchange (NNI)により最適化する手法も考案されています
9 (Hoang et al. 2018)。また**-b** オプションにより通常のブートストラップ法を適用すること
10 もできます。

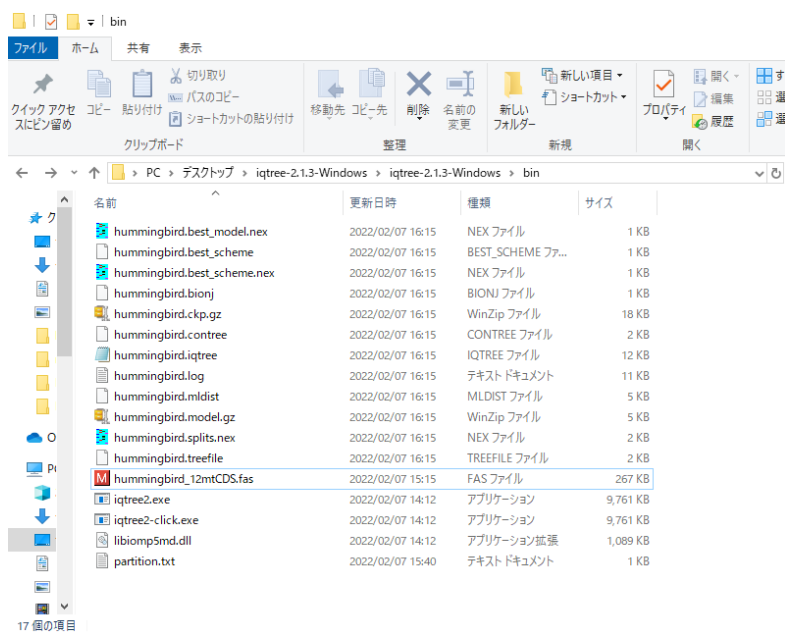
11 D.T. Hoang, O. Chernomor, A. von Haeseler, B.Q. Minh, L.S. Vinh (2018)
12 UFBoot2: Improving the ultrafast bootstrap approximation. Mol. Biol. Evol.,
13 35:518–522. <https://doi.org/10.1093/molbev/msx281>

14
15 **--prefix** というオプションはアウトプットファイルの名前を指定します。ここでは
16 **hummingbird** という名前を指定しています。このオプションを使わない場合、アラインメ
17 ントファイル名やパーティションファイル名がそのままアウトプットファイルの名前に用
18 いられます。

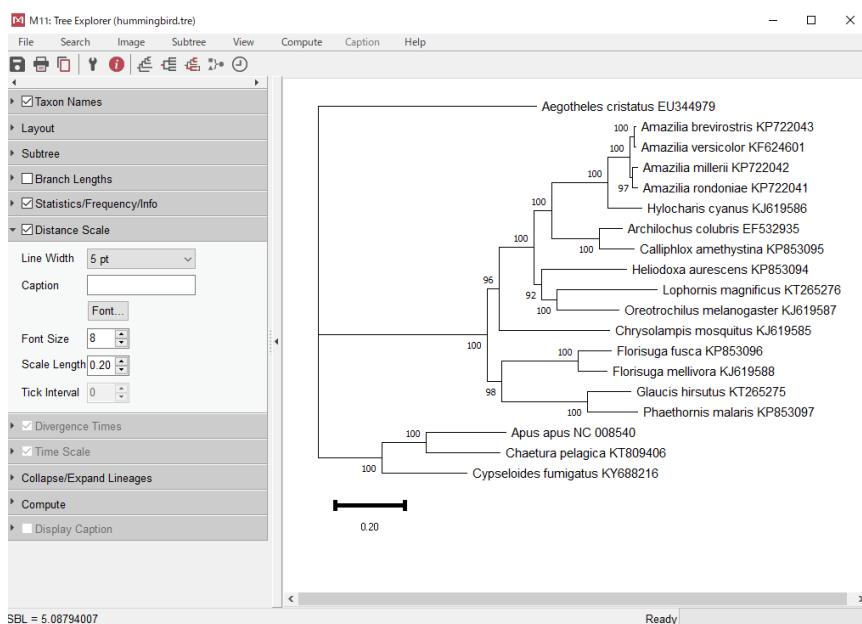
19
20 IQ-TREE2 プログラムはマルチコアに対応しており、これにより計算を高速化できます。
21 その場合は**-T** というオプションを使い、半角スペースをあけてコア数を記載します。

22 **3 節：解析結果の確認と系統樹の可視化**

23 計算が終了するとフォルダに 12 のファイルが新たに作られていることが分かります。
24 いずれも**--prefix** というオプションで指定したように **hummingbird** という名前がついてい
25 ます。
26



- 1
- 2 新たに作られたファイルのうち、特に重要なのが **iqtree** という拡張子のついたファイルと
- 3 **treefile** という拡張子のついたファイルです。
- 4
- 5 **iqtree** という拡張子のついたファイル（この場合、**hummingbird.iqtree**）には詳細な解析
- 6 結果（どのパーティションにどのモデルが用いられたのかや系統樹の尤度と標準誤差な
- 7 ど）が記載されています。**treefile** という拡張子のついたファイル（この場合、
- 8 **hummingbird.treefile**）にはベストの系統樹の樹形が、枝長とブートストラップ確率とともに
- 9 **newick** 形式のフォーマットで記録されています。
- 10
- 11 この **treefile** という拡張子を、**tre**（もしくは **nwk**）という拡張子に変更し、MEGA にドラ
- 12 ッグ&ドロップすれば系統樹を可視化することが出来ます。



1 SBL = 5.08794007 Ready

2 ノードの数値はブートストラップ確率です。なお、この系統樹は無根系統樹ですが、ズク

3 ヨタカ科である *Aegotheles cristatus* が外群になるため Subtree のタブから Root tree を選

4 択し、*Aegotheles cristatus* の枝をクリックすればこれを外群として表示することができま

5 す。

7 4 節：樹形ファイルの出力

8 この後の選択圧の推定や

9 分岐年代推定には、IQ-

10 TREE2 プログラムが推定

11 した系統樹の樹形のみを

12 用います。枝長やブート

13 ストラップ確率の情報は

14 不要なので、樹形の情報

15 のみを出力します。

16 MEGA で系統樹を描画す

17 る際、File のタブをクリ

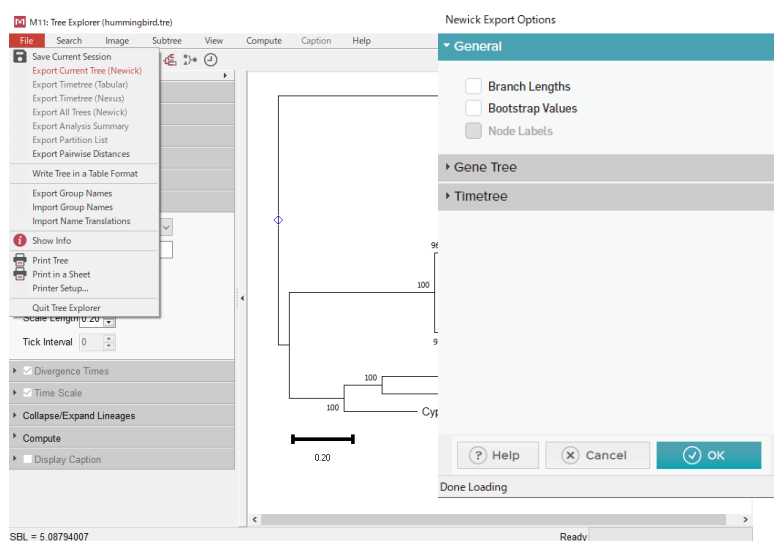
18 ックすると、オプション

19 が現れるので **Export**

20 **Current Tree (Newick)** を選択します。

21 すると **Newick Export Option** というウィンドウが出現するので、そのまま OK のボタンを

22 押し、枝長やブートストラップ確率の情報のない系統樹が **newick** 形式で出力されます。

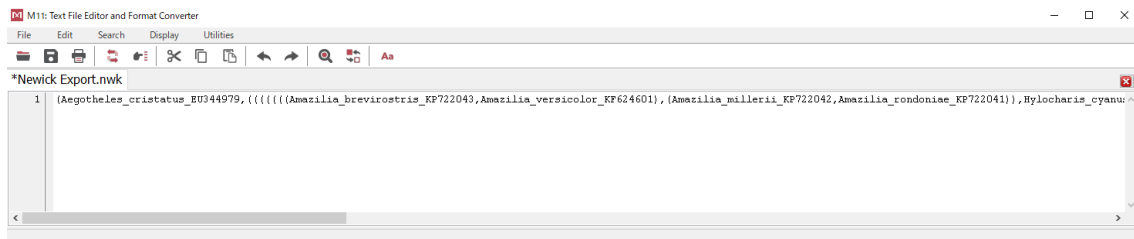


23

1 もし枝長やブートストラップ確率も一緒に出力したければ、**General** の項目の **Branch**
2 **Lengths** や **Bootstrap Values** にチェックを入れましょう。

3

4



5

6 デフォルトでは **Newick Export.nwk** という名前で保存されるので、**File** のタブで **Save as**
7 のオプションを使い、好きな場所に好きな名前で保存しましょう。

8

9 ※ここでは **ML.nwk** という名前を付けてこの後の解析に用います。

10

11

第二章：選択圧の推定

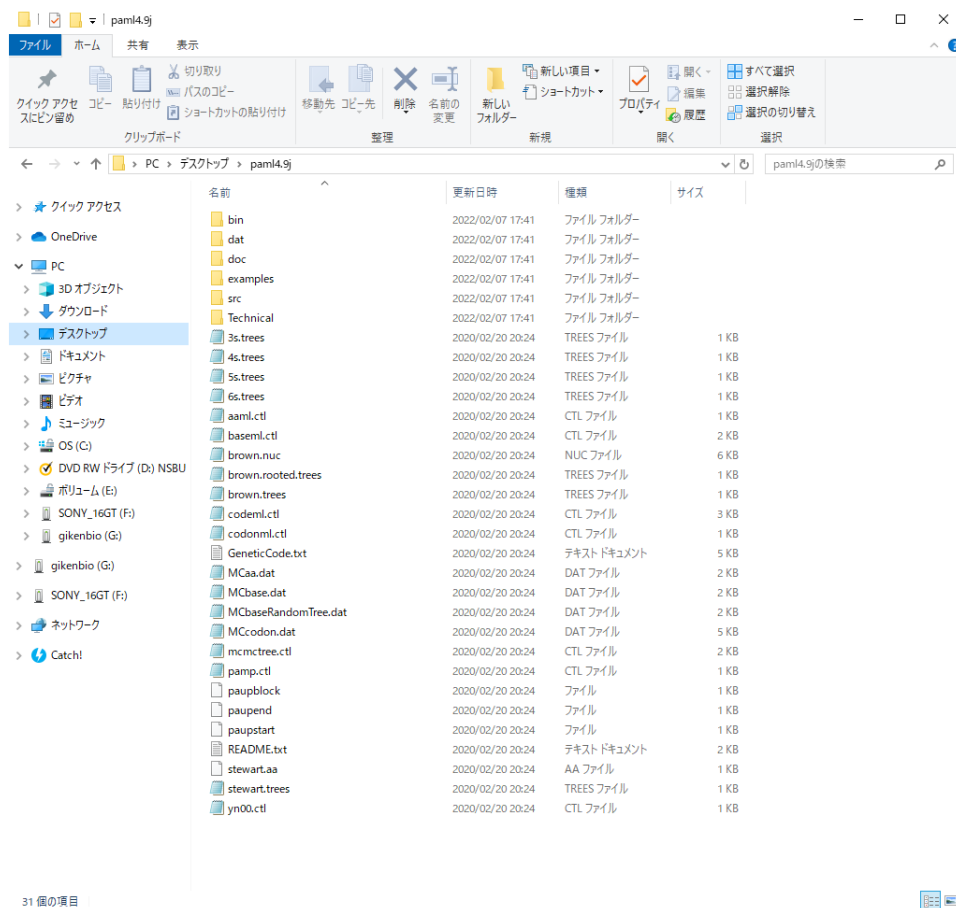
ハチドリのみトコンドリア全タンパク質コード遺伝子にかかる選択圧を推定してみよう

1 節：準備

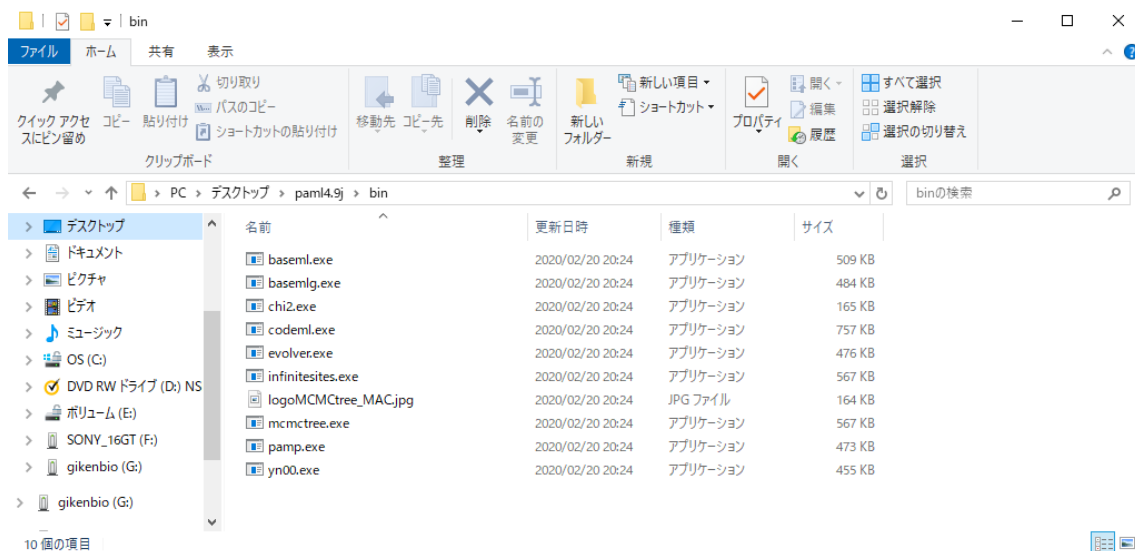
ここではコドン置換モデルによる選択圧の推定を行います。この解析に必要なのは、① PAML の CODEML プログラムの実行ファイル、② CODEML プログラムのコントロールファイル、③ アラインメントファイル、④ 樹形ファイルです。今回紹介する解析に限らず PAML は基本的に実行ファイル、コントロールファイル、アラインメントファイル、樹形ファイルの 4 点セットが必要になります。

① PAML の CODEML プログラムの実行ファイル

Ziheng Yang 教授のウェブサイトからダウンロードした **paml4.9j.tgz** を解凍すると、以下のようなフォルダやファイルが中に含まれていることが分かります。



PAML の様々なプログラムの実行ファイルは **bin** というフォルダの中に格納されています。



1 10 個の項目

2 バージョンによって若干の違いがありますが、PAML の bin には概ね 10 個程度の実行フ

3 ァイル(拡張子が exe になっているファイル)が入っており、特に **BASEML** プログラム、

4 **CODEML** プログラム、**MCMCTREE** プログラムが良く用いられます。

5

6 ここでは CODEML プログラムの実行ファイルである **codeml.exe** を、好きな場所に作っ

7 たフォルダにコピーします。

8 ここではデスクトップ上に **selection** というフォルダを作り、その中に **codeml.exe** を入れ

9 ます。

10

11 ② CODEML プログラムのコントロールファイル

12 コントロールファイルは、実行ファイルが読み込むアラインメントファイルや樹形ファイ

13 ルの情報や、実行ファイルが実際に行う解析内容、実行ファイルが出力するアウトファイ

14 ルの情報などが書き込まれたファイルです。paml4.9j の一番上の階層のフォルダには bin

15 のフォルダにほかに、多くのファイルが入っていました。このファイルのうち、**ctl** という

16 拡張子がついたものが、コントロールファイルになります。

17 ここでは特に CODEML プログラムのコントロールファイルである **codeml.ctl** を用いま

18 す。**codeml.ctl** をコピーし、前述のデスクトップ上に作られた **selection** というフォルダ

19 に入れます。

20

21 ③ アラインメントファイル

22 ここでは IQ-TREE2 プログラムで系統樹を推定する際に用いた

23 **hummingbird_12mtCDS.fas** をそのまま用いたいと思います。前述のデスクトップ上に作

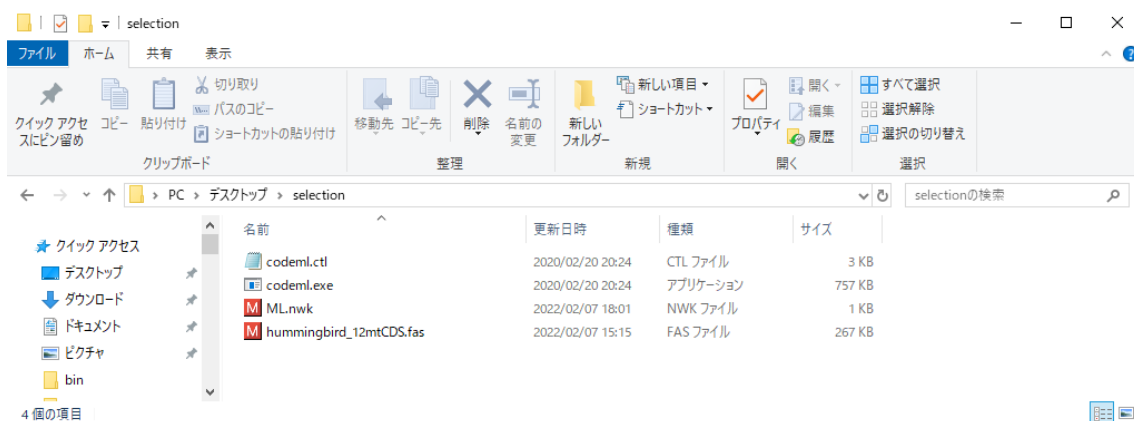
24 られた **selection** というフォルダに入れましょう。

25

1 ※PAML はもともと、PAML 形式と呼ばれるアラインメントファイルや、最尤法による系
2 統樹推定のパイオニアである PHYLIP プログラムに用いられる PHYLIP 形式のアライ
3 メントファイルが必要でしたが、PAML ver 4.3 以降は fasta 形式も利用可能になったの
4 で、今回はこのまま **hummingbird_12mtCDS.fas** を使います。ただし fasta 形式のアライ
5 メントファイルは、PAML でパーテーションモデルを用いる際に、制限が出てくる場合
6 があります。

7 8 ④ 樹形ファイル

9 ここでは IQ-TREE2 プログラムで推定し、MEGA で可視化したのちに、**ML.nwk** という
10 名前を付けて保存したファイルを前述のデスクトップ上に作られた **selection** というフォルダ
11 に入れて用います。



12
13
14

15 2 節：選択圧の推定

16 ここでは枝モデル、サイトモデル、枝サイトモデルについて紹介しますが、それぞれの
17 解析は独立した解析ですので、ここでは個別に紹介していきたいと思いますが、その前に
18 CODEML プログラムの**コントロールファイル**について、ここでもう少し詳しく紹介した
19 いと思います。

20 コントロールファイルを「Sublime」や「メモ帳」などのテキストエディタで開くと以
21 下のような情報が含まれています
22 なおアスタリスク(*)以下の情報は、コメントとして扱われていますので、オプションの内
23 容がよくわからない場合は、参照に用いてください。

```

seqfile = stewart.aa * sequence data filename
treefile = stewart.trees * tree structure file name
outfile = mlc * main result file name

noisy = 9 * 0,1,2,3,9: how much rubbish on the screen
verbose = 1 * 0: concise; 1: detailed, 2: too much
runmode = 0 * 0: user tree; 1: semi-automatic; 2: automatic
          * 3: StepwiseAddition; (4,5):PerturbationNNI; -2: pairwise

seqtype = 2 * 1:codons; 2:AAs; 3:codons-->AAs
CodonFreq = 2 * 0:1/61 each, 1:F1X4, 2:F3X4, 3:codon table

*
ndata = 10
clock = 0 * 0:no clock, 1:clock; 2:local clock; 3:CombinedAnalysis
aaDist = 0 * 0:equal, +:geometric; -:linear, 1-6:G1974,Miyata,c,p,v,a
aaRatefile = dat/jones.dat * only used for aa seqs with model=empirical(_F)
          * dayhoff.dat, jones.dat, wag.dat, mtmam.dat, or your own

model = 2
          * models for codons:
          * 0:one, 1:b, 2:2 or more dN/dS ratios for branches
          * models for AAs or codon-translated AAs:
          * 0:poisson, 1:proportional, 2:Empirical, 3:Empirical+F
          * 6:FromCodon, 7:AAClasses, 8:REVaa_0, 9:REVaa(nr=189)

NSsites = 0 * 0:one w;1:neutral;2:selection; 3:discrete;4:freqs;
          * 5:gamma;6:2gamma;7:beta;8:beta&w;9:beta&gamma;
          * 10:beta&gamma+1; 11:beta&normal>1; 12:0&2normal>1;
          * 13:3normal>0

icode = 0 * 0:universal code; 1:mammalian mt; 2-10:see below
Mgene = 0
          * codon: 0:rates, 1:separate; 2:diff pi, 3:diff kapa, 4:all diff
          * AA: 0:rates, 1:separate

fix_kappa = 0 * 1: kappa fixed, 0: kappa to be estimated
kappa = 2 * initial or fixed kappa
fix_omega = 0 * 1: omega or omega_1 fixed, 0: estimate
omega = .4 * initial or fixed omega, for codons or codon-based AAs

fix_alpha = 1 * 0: estimate gamma shape parameter; 1: fix it at alpha
alpha = 0. * initial or fixed alpha, 0:infinity (constant rate)
Malpha = 0 * different alphas for genes
ncatG = 8 * # of categories in dG of NSsites models

getSE = 0 * 0: don't want them, 1: want S.E.s of estimates
RateAncestor = 1 * (0,1,2): rates (alpha>0) or ancestral states (1 or 2)

Small_Diff = .5e-6
cleandata = 1 * remove sites with ambiguity data (1:yes, 0:no)?
* fix_blength = 1 * 0: ignore, -1: random, 1: initial, 2: fixed, 3: proportional
method = 0 * Optimization method 0: simultaneous; 1: one branch a time

* Genetic codes: 0:universal, 1:mammalian mt., 2:yeast mt., 3:mold mt.,
* 4: invertebrate mt., 5: ciliate nuclear, 6: echinoderm mt.,
* 7: euplotid mt., 8: alternative yeast nu. 9: ascidian mt.,
* 10: blepharisma nu.
* These codes correspond to transl_table 1 to 11 of GENEbank.

```

1

- 2 ここでは、枝モデル、サイトモデル、枝サイトモデルを用いて解析を実行する際に、共通
- 3 して変更しておくべきオプションについて概説します。

1

```
seqfile = stewart.aa * sequence data filename
treefile = stewart.trees * tree structure file name
outfile = mlc * main result file name
```

2

3 一行目の「seqfile」はアラインメントの情報です。今回は **hummingbird_12mtCDS.fas**
4 というファイルを用いるので、

5

6 **seqfile = hummingbird_12mtCDS.fas**

7

8 と書きます。

9 二行目の「treefile」は樹形ファイルの情報です。今回は **ML.nwk** というファイルを用いる
10 ので、

11

12 **treefile = ML.nwk**

13

14 と書きましょう。

15 三行目の「outfile」は尤度やパラメータなど、解析結果の基本情報が記録されるアウトフ
16 ァイルの情報です。ここは任意の名前を付けましょう。

17

```
seqtype = 2 * 1:codons; 2:AAs; 3:codons-->AAs
CodonFreq = 2 * 0:1/61 each, 1:F1X4, 2:F3X4, 3:codon table
```

18

19 CODEML プログラムは、コドン置換モデルもしくはアミノ酸置換モデルによる解析に特
20 化したプログラムです。従って現在扱っているアラインメントファイルがコドン配列なの
21 かアミノ酸配列なのかをプログラムに教えてあげる必要があります。

22 今回はコドン配列なので

23

24 **seqtype = 1**

25

26 を選択しましょう。

27 CodonFreq はコドンの使用頻度に関するオプションです。

28 0 を選ぶと、すべてのコドンが同じ頻度であることを仮定します。コドンの暗号表のう
29 ち標準遺伝コードは、停止コドンを除くと 61 種類のコドンがあるので、すべてのコドン
30 が 1/61 という頻度で存在することになります。なおこのオプションを選択すると、コド
31 ン使用頻度に関するパラメータ数は 0 個になります。

32 1 を選ぶと、アラインメント全体の塩基の頻度(π_A , π_T , π_G , π_C)をもとにコドン頻
33 度の期待値が計算されます。例えば ACT というコドンの場合、その頻度は $\pi_A \times \pi_C \times \pi_T$

1 になります。 $\pi_A + \pi_T + \pi_G + \pi_C = 1$ という制約があるので、塩基の頻度に関する自由パラメータ数は3になります。従ってこのオプションを選択すると、コドン使用頻度に関するパラメータ数は3個になります。

2 2を選ぶと、コドンの1番目、2番目、3番目ごと推定した塩基の頻度から、コドン頻度の期待値が計算されます。コドン内の3つの位置がそれぞれ塩基の頻度に関する自由パラメータを3つ持つため、このオプションを選択すると、コドン使用頻度に関するパラメータ数は9個になります。

3 3を選ぶと、コドンの使用頻度をアラインメントデータから直接カウントします。標準遺伝コードは61種類のコドンを持ちますが、コドン使用頻度に関するパラメータ数は(上述の塩基の頻度と同様に合計して1になるという制約があるため)60個になります。

4 どのオプションを選ぶかは、厳密にはAICなどを計算して決定すべきですが、脊椎動物のミトコンドリアゲノムの場合はコドンの1番目、2番目、3番目で塩基組成が大きく違うためオプション2もしくは3を選ぶと良いでしょう。

```
icode = 0 * 0:universal code; 1:mammalian mt; 2-10:see below
```

これはコドンのコード表を選ぶオプションになります。

詳しくはコントロールファイルの一番下の部分に

```
* Genetic codes: 0:universal, 1:mammalian mt., 2:yeast mt., 3:mold mt.,
* 4: invertebrate mt., 5: ciliate nuclear, 6: echinoderm mt.,
* 7: euplotid mt., 8: alternative yeast nu. 9: ascidian mt.,
* 10: blepharisma nu.
* These codes correspond to transl_table 1 to 11 of GENE BANK.
```

と書かれているので、自分のデータに合わせたコドンのコード表を使いましょう。

ここでは、鳥類のミトコンドリアゲノム(哺乳類のミトコンドリアゲノムのコード表と同一)を用いているので、

icode=1

を選択しましょう。

知っておくと便利なお役立ち情報

```
runmode = 0 * 0: user tree; 1: semi-automatic; 2: automatic
* 3: StepwiseAddition; (4,5):PerturbationNNI; -2: pairwise
```

今回の演習ではIQ-TREE2で推定した最尤系統樹の樹形を用いてPAMLによる解析を行います。その場合は、ユーザーが系統樹の樹形を与えるのでrunmode=0を選択します。1~5のオプションを選択するとPAMLを用いて系統樹の推定を行うことが出来ますが、やはりIQ-TREE2やRAxML-NGなどの系統樹推定に特化したプログラムのほうがより高

1 速かつ正確に樹形を推定できます。-2 のオプションを選択すると 2 配列間で尤度やパラメ
2 ータを総当たりで推定してくれます。

3

```
4 cleandata = 1 * remove sites with ambiguity data (1:yes, 0:no)?  
* fix_blength = 1 * 0: ignore, -1: random, 1: initial, 2: fixed, 3: proportional  
method = 1 * Optimization method 0: simultaneous; 1: one branch a time
```

5 曖昧塩基やギャップを含むサイトを解析から除外したい場合は cleandata=1 を選
6 択しましょう。曖昧塩基やギャップを含むサイトもすべて残したい場合は cleandata=0 に
7 します。

8 今回は、樹形ファイルを作るときに枝の長さの情報を除外して樹形のみにしまし
9 た。しかし IQ-TREE2 など推定した枝の長さを、そのまま使いたいときは（アスタリス
10 クを外して）fix_blength=2 を選択しましょう。また IQ-TREE2 など推定した枝の長さ
11 を初期値として使いたいときは fix_blength=1 を選択しましょう。今回は IQ-TREE2 で推
12 定した枝の長さは塩基サイトあたりの置換数になっており、PAML ではコドンサイトあた
13 りの置換数を推定することになるので、IQ-TREE2 など推定した枝の長さは使っていま
14 せん。

15 method は枝の長さを最適化する際のアルゴリズムです。0 はすべての枝を同時に
16 最適化し、1 は枝をひとつずつ最適化していきます。個人的な経験としては 1 のほうが計
17 算が早く終わりますが、尤度やパラメータの推定値は変わらないようです。

18

19

1 ■1項：枝モデルを用いた選択圧の推定

2 ホバリング飛行を行うグループと行わないグループではミトコンドリア全タンパク質コー
3 ド遺伝子にかかる選択圧に違いはあるだろうか？

4
5 1. コントロールファイルの編集

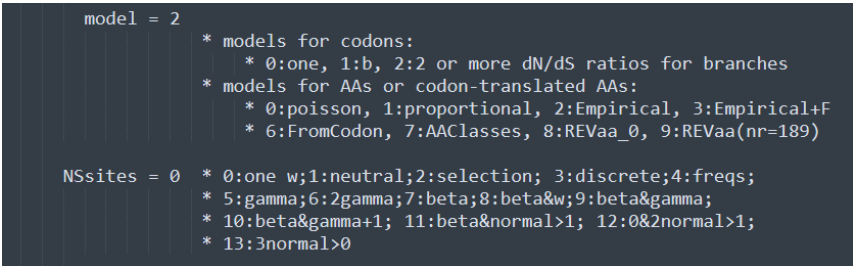
6 シーケンスファイル名や樹形ファイル名を指定しましょう。アウトファイル名は好きな名
7 前をつけましょう（ここでは **branch2w.out** としたいと思います）。

8
9 **seqfile = hummingbird_12mtCDS.fas**

10 **treefile = ML.nwk**

11 **outfile = branch2w.out**

12
13 枝モデル、サイトモデル、枝サイトモデルの切り替えは、コントロールファイルの
14 「**model**」および「**NSsites**」のオプションを変更します。

15
16 The screenshot shows a text-based control file with the following content:
model = 2
* models for codons:
* 0:one, 1:b, 2:2 or more dN/dS ratios for branches
* models for AAs or codon-translated AAs:
* 0:poisson, 1:proportional, 2:Empirical, 3:Empirical+F
* 6:FromCodon, 7:AAClasses, 8:REVaa_0, 9:REVaa(nr=189)
NSsites = 0 * 0:one w;1:neutral;2:selection; 3:discrete;4:freqs;
* 5:gamma;6:2gamma;7:beta;8:beta&w;9:betaγ
* 10:beta&gamma+1; 11:beta&normal>1; 12:0&2normal>1;
* 13:3normal>0

17 **model** では、どのようなコドン置換モデルを用いるかを選択できます。

18 枝モデルの場合、

19 **model = 2**

20 もしくは

21 **model = 1**

22
23 を選びましょう。

24 0を選択するとすべての枝が同じ ω を持つことを仮定します。

25 1を選択するとすべての枝が異なる ω を持つことを仮定します。このモデルは枝の数だけ
26 ω パラメータの数があるので非常にパラメータリッチになり、正しくパラメータ推定で
27 きない可能性もあります。

28 2を選択すると、系統樹の枝が2つ以上のグループに分かれ、グループごとに ω を推定し
29 ます。枝をどのようにグループ分けするかは樹形ファイルにより指定します。通常はこの
30 2を選択します。

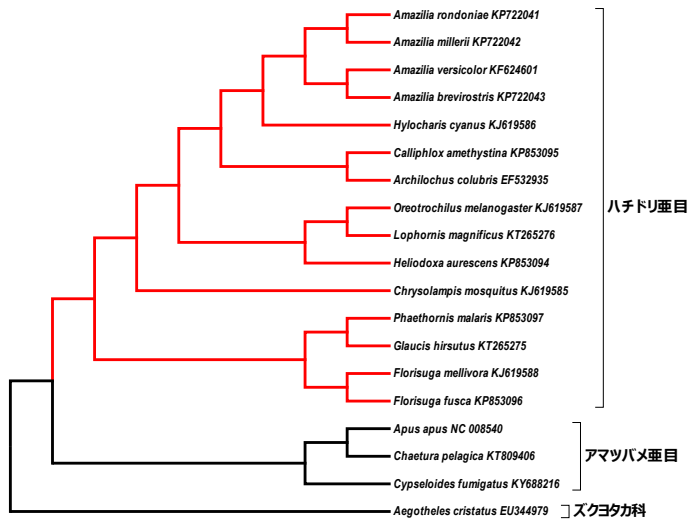
1 **NSsites** は、コドンサイトごとに異なる ω を持つことを許すモデルを選択するオプション
 2 ですが、枝モデルではすべてのコドンサイトを通して ω が一定であることを仮定している
 3 ため、ここでは

4 **NSsites=0**

5 を選択します。

7 2. 樹形ファイルの編集

8 樹形ファイルには、系統樹の枝をどのようにグループ分けするかという情報を与えます。
 9 今回はホバリング飛行をするグループ（ハチドリ亜目）とホバリング飛行をしないグルー
 10 プ（アマツバメ亜目・ズクヨタカ科）に分けますが、系統樹上では以下のようになりま
 11 す。赤い枝がホバリング飛行をするグループ、黒い枝がホバリング飛行をしないグループ
 12 になります。



13
 14
 15 第一章の系統樹推定で IQ-TREE2 プログラムにより推定した系統樹を MEGA プログラム
 16 により **ML.nwk** というファイルに保存しました。このファイルは下記のようになっていま
 17 す。

```

18 (Aegotheles_cristatus_EU344979,((((((Amazilia_brevirostris_KP722043
19 ,Amazilia_versicolor_KF624601),(Amazilia_millerii_KP722042,Amazilia_
  
```

1 枝モデルでは枝のグループは#により行います。ここではハチドリ亜目の枝に#1 をつけま
2 しょう。#1 がついていない枝は、「それ以外の枝」として一括して扱われます。

3

4 ハチドリ亜目の枝に#1 をつけると、以下のようになります。

```
(Aegotheles_cristatus_EU344979,(((((((Amazilia_brevirostris_KP722043  
#1,Amazilia_versicolor_KF624601 #1) #1,(Amazilia_millerii_KP72204  
2 #1,Amazilia_rondoniae_KP722041 #1) #1) #1,Hylocharis_cyanus_KJ6  
19586 #1) #1,(Archilochus_colubris_EF532935 #1,Calliphlox_amethyst  
ina_KP853095 #1) #1) #1,(Heliodoxa_aurescens_KP853094 #1,(Lophorn  
is_magnificus_KT265276 #1,Oreotrochilus_melanogaster_KJ619587 #1)  
#1) #1) #1,Chrysolampis_mosquitos_KJ619585 #1) #1,((Florisuga_fus  
ca_KP853096 #1,Florisuga_mellivora_KJ619588 #1) #1,(Glaucis_hirsut  
us_KT265275 #1,Phaethornis_malaris_KP853097 #1) #1) #1),((Apu  
s_apus_NC_008540,Chaetura_pelagica_KT809406),Cypseloides_fumigatus_K  
Y688216));
```

5

6 外部枝に#1 をつける際は OTU 名の直後に、内部枝に#1 をつける際はノードを定義す
7 る”の直後に#1 を挿入します。

8

9 #1 をつけ終わったら **ML.nwk** を保存します。これで準備完了です。

10

11 [知っておくと便利なお役立ち情報](#)

12 [上記の例ではハチドリ亜目内のすべての枝に#1 を付けましたが、今回のケースのようにあ](#)
13 [る特定の単系統群全体に#0 を付けたい場合、その共通祖先にあたる枝に\\$1 のシンボルを](#)
14 [つけると、その共通祖先の枝とその子孫にあたるすべての枝に#1 を付けるのと同じ意味に](#)
15 [なります。](#)

```
(Aegotheles_cristatus_EU344979,(((((((Amazilia_brevirostris_KP722043,  
Amazilia_versicolor_KF624601),(Amazilia_millerii_KP722042,Amazilia_ro  
ndoniae_KP722041)),Hylocharis_cyanus_KJ619586),(Archilochus_colubris_  
EF532935,Calliphlox_amethystina_KP853095)),(Heliodoxa_aurescens_KP853  
094,(Lophornis_magnificus_KT265276,Oreotrochilus_melanogaster_KJ61958  
7))),Chrysolampis_mosquitos_KJ619585,((Florisuga_fusca_KP853096,Flor  
isuga_mellivora_KJ619588),(Glaucis_hirsutus_KT265275,Phaethornis_mala  
ris_KP853097)))$1,((Apus_apus_NC_008540,Chaetura_pelagica_KT809406),C  
ypseloides_fumigatus_KY688216));
```

16

17

18

19

20 3. 解析の実行

21 解析に必要な必要なファイルが準備できたら、実行ファイル **codeml.exe** をダブルクリック
22 しましょう。ファイルに問題がなければ以下のようなターミナルが開いて計算が始まり
23 ます。

```

C:\Users\Yonezawa\Desktop\selection\codeml.exe
0.016550 0.008774 0.017470 0.055760 0.083286 0.023607 0.050919 0.015590 0.015979 0.080098
0.029246 0.027181 0.077124 0.064808 0.061225 0.074756 0.019246 0.029416 0.064753 0.059344
0.056230 0.079672 0.064383 0.065474 0.075092 0.424701 0.371625 0.389681

nrate & nrate & nrate: 35 3 38

Bounds (np=38):
0.000004 0.000004 0.000004 0.000004 0.000004 0.000004 0.000004 0.000004 0.000004 0.000004 0.000004
4 0.000004 0.000004 0.000004 0.000004 0.000004 0.000004 0.000004 0.000004 0.000004 0.000004 0.000004
04 0.000004 0.000004 0.000004 0.000004 0.000004 0.000004 0.000004 0.000004 0.000004 0.000004 0.000004
004 0.000004 0.000004 0.000100 0.000100 0.000100
0 50.000000 50.000000 50.000000 50.000000 50.000000 50.000000 50.000000 50.000000 50.000000 50.000000 50.000000
0 50.000000 50.000000 50.000000 50.000000 50.000000 50.000000 50.000000 50.000000 50.000000 50.000000 50.000000
10 50.000000 50.000000 50.000000 50.000000 50.000000 50.000000 50.000000 50.000000 50.000000 50.000000 50.000000
000 50.000000 50.000000 999.000000 999.000000 999.000000

np = 38
rfAL = -95530.625519

Iterating by ming2
Initial: fix= 95530.625519
xz= 0.03943 0.04651 0.10380 0.08284 0.02201 0.10564 0.03774 0.06735 0.01796 0.06582 0.01855 0.08977 0.01747
0.06576 0.08329 0.02361 0.05092 0.01559 0.01898 0.08010 0.02925 0.02718 0.07712 0.06481 0.06123 0.07476 0.
01925 0.02942 0.06475 0.05934 0.06623 0.07967 0.06438 0.06547 0.07509 0.42470 0.37162 0.38968

1 h-n-p 0.0000 0.0000 61140.3670 HCYCYCC 76938.554493 9 0.0000 58 | 0/38
2 h-n-p 0.0000 0.0000 16707.7542 CCC 76730.412275 4 0.0000 107 | 0/38
3 h-n-p 0.0000 0.0000 51986.3977 C

```

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9

しかし、コントロールファイル、アラインメントファイル、樹形ファイルのいずれかに問題があるとターミナルが一瞬だけ開き計算が途中で止まってしまいます。そのような場合は何が原因で計算が止まったのかわからないため、コマンドプロンプトから計算を実行すると良いです。

コマンドプロンプトを用いて計算を行いたいフォルダに移動し、**codeml** とタイプします。

```

コマンドプロンプト
Microsoft Windows [Version 10.0.17763.437]
(c) 2018 Microsoft Corporation. All rights reserved.

C:\Users\Yonezawa>cd C:\Users\Yonezawa\Desktop\selection
C:\Users\Yonezawa\Desktop\selection>codeml

```

- 10
- 11

ここでは例として、アラインメントファイルから”Aegotheles_cristatus_EU344979”をわざとのおいてみました。つまり樹形ファイルには Aegotheles_cristatus_EU344979 が存在するのに、アラインメントファイルにはこの配列が存在しない状態になっています。ここでコマンドプロンプトを用いて CODEML プログラムを実行すると

```

コマンドプロンプト
), 0:00
Counting codons..

1224 bytes for distance
3061440 bytes for conP
27866456493129829 bytes for fhK
5000000 bytes for space

Species Aegotheles_cristatus_EU344979?
C:\Users\Yonezawa\Desktop\selection>

```

- 16

1 このように

2 Species Aegotheles_cristatus_EU344979?

3

4 とアラートが出ています。このように解析がうまく進まない場合、コマンドプロンプトから CODEML プログラムを実行することでエラーを表示してくれるためトラブルシューティングを行うことができます。

7

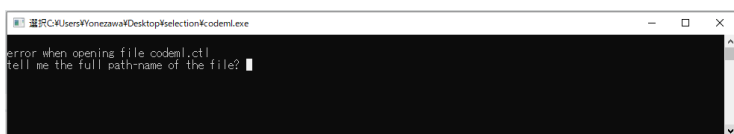
8 知っておくと便利なお役立ち情報

9 どのような設定で解析したのか記録を残しておくためにコントロールファイルを保存しておきたい場合もあると思います。今回の枝モデルのコントロールファイルも `branch2w.ctl` という名前で例題ファイルとして保存していますが (※)、コマンドプロンプトから CODEML プログラムを実行する場合

13 codeml branch2w.ctl

14 とタイプすればこの `branch2w.ctl` の内容でそのまま解析を行うことも出来ます。

15 ダブルクリックで CODEML を実行する場合も同じフォルダの中に `codeml.ctl` というコントロールファイルが存在しない場合、下記のようなエラーメッセージが出てコントロールファイル名を聞いてきますので、コントロールファイル名を入力してあげましょう。



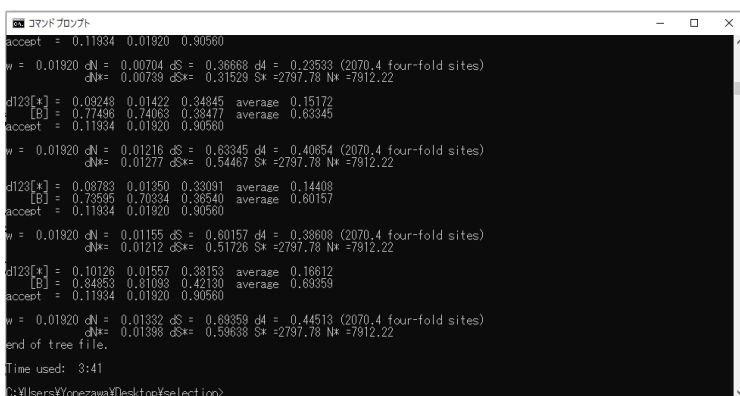
18

19

20 ※`help.zip` というフォルダに本解析で用いたコントロールファイルを保存しています。

21

22 実行ファイルをダブルクリックして計算を始めた場合、計算が終了すると自動的に window が閉じます。計算を行ったフォルダを確認するとアウトファイル (ここでは **branch2w.out**) などいくつかのファイルが新たに出来ていることが分かると思います。コマンドプロンプトから計算を始めた場合は、下のような画面になっていると思います。



26

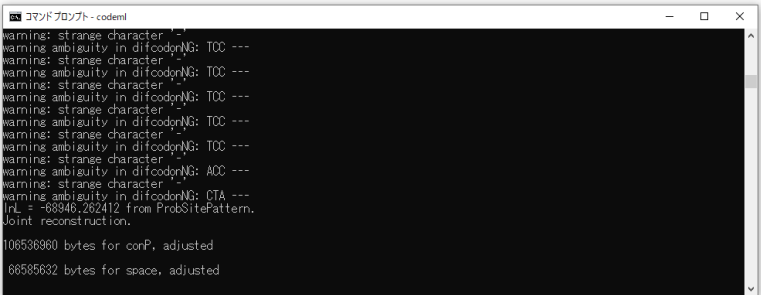
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27

知っておくと便利なお役立ち情報

PAML の便利な機能として、祖先ノードの配列を最尤推定してくれます。

```
RateAncestor = 1 * (0,1,2): rates (alpha>0) or ancestral states (1 or 2)
```

これは **RateAncestor=1** もしくは **RateAncestor=2** を選択すると良いのです。祖先配列の情報は **rst** というファイルに保存されます。このオプションを選ぶと計算が終了しても以下の画面のまま止まってしまう。しかしアウトファイルは正しく作られているようです。



4. 解析結果の確認

基本的な解析結果はアウトファイルに書き込まれているので **branch2w.out** をテキストエディターで開いてみましょう。

知っておくと便利なお役立ち情報

アウトファイルは、まずアラインメントそのものやサイトのパターンなどの情報が記録されています。こうした情報をわざわざアウトファイルに残しておきたくない場合は、コントロールファイルの

```
verbose = 1 * 0: concise; 1: detailed, 2: too much
```

verbose=0

を選択しておく、これらの情報は記録されません。

アラインメントの下には、コドンや塩基の使用頻度などが記録されています。

最も重要な情報である尤度やパラメータの推定値はアウトファイルの下のほうに記録されています。


```

TREE # 1: (1, ((((((2, 5), (3, 4)), 16), (7, 8)), (15, (17, 18))), 10), ((12, 13), (14, 19))), ((6, 9), 11)); MP score: 13598
lnL(ntime: 35 np: 38): -68637.606213 +0.000000
20..1 20..21 21..22 22..23 23..24 24..25 25..26 26..27 27..2 27..5 26..28 28..3 28..4 25..16 24..
1.299701 0.933482 0.069955 0.171350 0.095578 0.286230 0.140762 0.033073 0.017910 0.021282 0.023533 0.046428 0.045771 0.202760 0.252
Note: Branch length is defined as number of nucleotide substitutions per codon (not per nucleotide site).
tree length = 10.025336
(1: 1.299701, (((((((2: 0.017910, 5: 0.021282): 0.033073, (3: 0.046428, 4: 0.045771): 0.023533): 0.140762, 16: 0.202760): 0.286230,
(Aegotheles_cristatus_EU344979: 1.299701, (((((((Amazilia_brevirostris_KP722043: 0.017910, Amazilia_versicolor_KF624601: 0.021282):
Detailed output identifying parameters
kappa (ts/tv) = 9.38607
w (dN/dS) for branches: 0.01920 0.02248
dN & dS for each branch
branch      t      N      S      dN/dS      dN      dS      N*dN      S*dS
20..1      1.300  8304.3  2405.7  0.0192  0.0347  1.8089  288.4  4351.5
20..21     0.933  8304.3  2405.7  0.0225  0.0289  1.2855  240.0  3092.6
21..22     0.070  8304.3  2405.7  0.0225  0.0022  0.0963  18.0  231.8
22..23     0.171  8304.3  2405.7  0.0225  0.0053  0.2360  44.0  567.7
23..24     0.096  8304.3  2405.7  0.0225  0.0030  0.1316  24.6  316.6
24..25     0.286  8304.3  2405.7  0.0225  0.0089  0.3942  73.6  948.3
25..26     0.141  8304.3  2405.7  0.0225  0.0044  0.1938  36.2  466.3
26..27     0.033  8304.3  2405.7  0.0225  0.0010  0.0455  8.5  109.6
27..2      0.018  8304.3  2405.7  0.0225  0.0006  0.0247  4.6  59.3
27..5      0.021  8304.3  2405.7  0.0225  0.0007  0.0293  5.5  70.5
26..28     0.024  8304.3  2405.7  0.0225  0.0007  0.0324  6.0  78.0
28..3      0.046  8304.3  2405.7  0.0225  0.0014  0.0639  11.9  153.8
28..4      0.046  8304.3  2405.7  0.0225  0.0014  0.0630  11.8  151.6
25..16     0.203  8304.3  2405.7  0.0225  0.0063  0.2792  52.1  671.7
24..29     0.253  8304.3  2405.7  0.0225  0.0078  0.3482  65.0  837.7
29..7      0.163  8304.3  2405.7  0.0225  0.0051  0.2250  42.0  541.2

```

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15
- 16
- 17
- 18
- 19
- 20
- 21

TREE # 1: (1, (((((((2, 5), (3, 4)), 16), (7, 8)), (15, (17, 18))), 10), ((12, 13), (14, 19))), ((6, 9), 11)); MP score: 13598

は、系統樹の樹形を newick フォーマットで示しています。1~19 という数字はアラインメントファイルの配列名を降順で示したものです。MP score は最節約法の規準でのこの系統樹の置換数の総和になります。

lnL(ntime: 35 np: 38): -68637.606213 +0.000000

この行は、系統樹の対数尤度を示しています。
np: 38 は、この解析に用いられたパラメータの総数を表しています。

知っておくと便利なお役立ち情報

最尤法で系統樹を推定する際、パラメータの数は、系統樹の枝の総数と置換モデルの持つパラメータ数の和になります。系統樹の枝の総数ですが、この場合 19 の配列を用いているので、無根系統樹の場合、 $19 \times 2 - 3 = 35$ (本文参照) になります。置換モデルの持つパラメータですが、この場合、トランジション率/トランスバージョン率の比率(κ)と非同義置換率/同義置換率の比率(ω)になります。なお今回の解析では系統樹の枝を #0 と #1 に分けて、それぞれの ω を推定するため、 ω に関しては $\omega_{\#0}$ と $\omega_{\#1}$ の 2 つのパラメータを持つことになります。

1 置換モデルに関して、塩基やアミノ酸、コドンの使用頻度も通常はパラメータに含まれ
2 ますが、PAML の場合、これらの頻度は最尤推定しない限りはパラメータ数としてカウン
3 トしていないようです。従って置換モデル同士を AIC で比較したい場合は、これらの頻度
4 パラメータを加味して計算するようにしてください。

5
6 そのあとは、各枝の長さに関するパラメータの情報が続きます。枝の長さはコドン当たり
7 の置換数で表示しています。

```
Detailed output identifying parameters  
kappa (ts/tv) = 9.38612  
w (dN/dS) for branches: 0.01920 0.02248
```

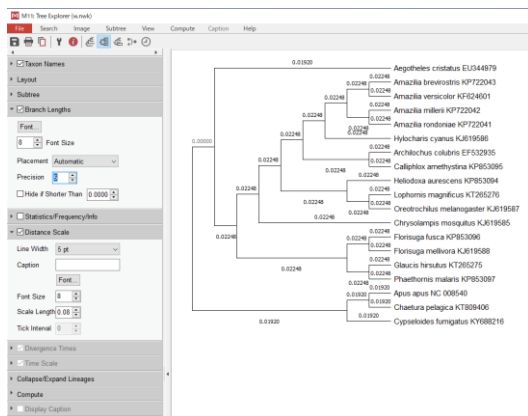
9
10 ここで、トランジション率/トランスバージョン率の比率(κ)や非同義置換率/同義置換率
11 の比率(ω)の情報が記録されています。

12 κ は 9.38612。 ω に関しては $\omega_{\#0}$ が **0.01920**、 $\omega_{\#1}$ が **0.02248** になります。
13 これはホバリング飛行をするグループでは非同義置換率が相対的に高くなっていることを
14 意味します。これは負の選択圧の緩和、もしくは正の選択圧が働いていると解釈します。

15
16 この下には、各枝の ω の情報がまとめられており、アウトファイルの一番下には、系統樹
17 の枝の長さを同義置換率のみで推定したものと、非同義置換率のみで推定したものを
18 newick 形式で示しており、最後に各枝の ω を newick 形式で示しています。

20 知っておくと便利なお役立ち情報

21 配列数が多い場合、newick 形式で表現した系統樹は()の数が多くなり、#1 などを書き込
22 んだり、アウトプットファイルを確認したりするのは大変です。newick 形式の系統樹が書
23 き込んであるファイルを nwk などの拡張子をつけて一度保存し、“#”を“:”に変換し MEGA
24 等で描画すると確認しやすいです。“:”のうしろの数値は枝の長さとして解釈されます。MEGA
25 の Tree Explorer は、View のタブで Show/Hide のオプションを選ぶことができます。そ
26 こで Toggle Display of Branch Lengths を選ぶと下図のように各枝の ω をチェックするこ
27 とができます。



1
2
3
4
5
6
7
8
9
10
11

5・統計的有意性の評価

アウトファイルを見ると $\omega_{\#0}$ が 0.01920、 $\omega_{\#1}$ が 0.02248 と推定されましたが、この値には有意な差はあるのでしょうか？

これを検証するためには、 $\omega_{\#0} = \omega_{\#1}$ を仮定した帰無仮説（1 ω モデルと呼びたいと思います）を比較する必要があります。

今回の場合は、 ω は二種類しかないので、コントロールファイルを以下のように変更してみましょう。

```

model = 0
* models for codons:
* 0:one, 1:b, 2:2 or more dN/dS ratios for branches
* models for AAs or codon-translated AAs:
* 0:poisson, 1:proportional, 2:Empirical, 3:Empirical+F
* 6:FromCodon, 7:AAClasses, 8:REVaa_0, 9:REVaa(nr=189)

NSsites = 0
* 0:one w;1:neutral;2:selection; 3:discrete;4:freqs;
* 5:gamma;6:2gamma;7:beta;8:beta&w;9:beta&gamma;
* 10:beta&gamma+1; 11:beta&normal>1; 12:0&2normal>1;
* 13:3normal>0

```

12
13
14
15
16
17
18
19
20

model=0

を選択すれば、系統樹上のすべての枝が同じ ω を持つことになるので、上記の帰無仮説の条件で解析を行うことが出来ます。前に計算したアウトファイルが上書きされないように、アウトファイル名も変更しておきましょう。

帰無仮説（1 ω モデル）の対数尤度やパラメータは以下のようにになっています。

```

TREE # 1: (1, (((((((2, 5), (3, 4)), 16), (7, 8)), (15, (17, 18)), 19), 20), 21), 22), 23), 24), 25), 26), 27), 28), 29), 30), 31), 32), 33), 34), 35);
lnL(nptime: 35 np: 37): -68640.899450 +0.000000
20..1 20..21 21..22 22..23 23..24 24..25 25..26
36..9 35..11
1.219520 0.956310 0.070409 0.172517 0.096024 0.287884 0.140993 0.021632
0.417906 0.479664 9.310184 0.021632

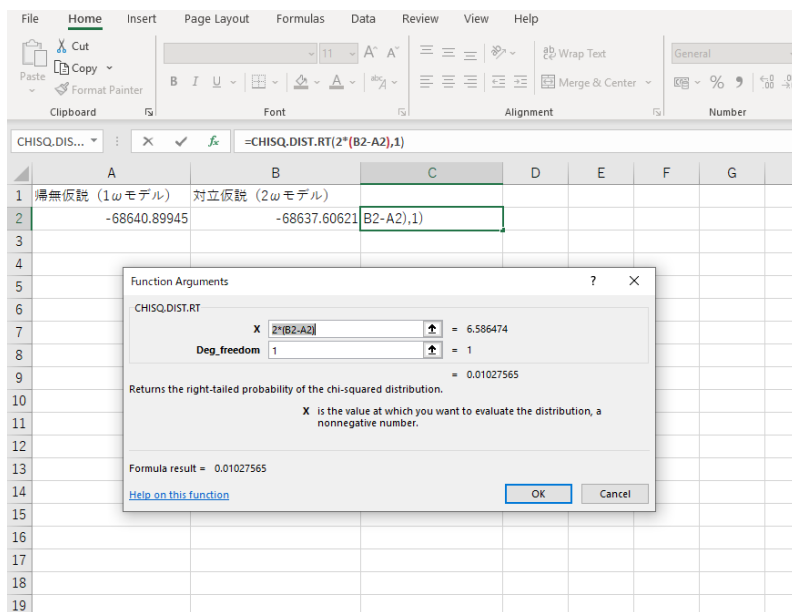
Note: Branch length is defined as number of nucleotide substitutions per site
tree length = 9.906797

(1: 1.219520, (((((((2: 0.017906, 5: 0.021285): 0.033084, (3: 0.0186213): 0.405762, (14: 0.245676, 19: 0.301732): 0.439758): 0.412625), 16): 0.172517, (7: 0.096024, (15, (17, 18)): 0.140993, 19): 0.301732), 20): 0.956310, (21, 22): 0.070409, (23, 24): 0.287884, (25, 26): 0.140993, 27): 0.417906, 28): 0.021632);
(Aegotheles_cristatus_EU344979: 1.219520, (((((((Amazilia_brevirostris_A00151: 0.163519, Calliphlox_amethystina_P00151: 0.593849): 0.070409, ((Florisuga_fusca_KP853096: 0.195105, Florisuga_fusca_KP853096: 0.412625), Cypseloides_fumigatus_KY688216: 0.479664): 0.412625);

Detailed output identifying parameters
kappa (ts/tv) = 9.31018
omega (dN/dS) = 0.02163

```

- 1
- 2 ω はひとつしか推定されておらず、0.02163 という値になっています。
- 3 また np:37 になっており、対立仮説 (2ω モデルと呼びたいと思います) と比較するとパラ
- 4 ラメータ数が1個分小さくなっています。
- 5 対数尤度も **-68640.899450** となっており、対立仮説 (**-68637.606213**) と比較するとわず
- 6 かにちいさくなっていることが分かります。
- 7
- 8 さて、ここで重要なことは帰無仮説 (1ω モデル) は、対立仮説 (2ω モデル) の特殊なケ
- 9 ースであり、 $\omega_{\#0}=\omega_{\#1}$ という条件下では 1ω モデルと 2ω モデルは同じモデルになるとい
- 10 うことです。この場合、ふたつのモデルは入れ子状の関係にあるといます。
- 11
- 12 ふたつのモデルが入れ子状になっている場合、尤度比検定を用いることで帰無仮説と対立
- 13 仮説を比較することができます。この場合、対立仮説のほうが帰無仮説よりも1個分パラ
- 14 メータが多いですが、それに見合うだけ尤度が上昇しているかどうかを調べます。
- 15
- 16 対立仮説と帰無仮説の対数尤度の差の2倍が、カイ二乗分布に従うため、エクセルを用い
- 17 て尤度比検定を行うことができます。
- 18
- 19 エクセルの CHISQ.DIST.RT 関数を用いて
- 20
$$X = 2 \times (\ln L_{2\omega \text{ model}} - \ln L_{1\omega \text{ model}})$$
- 21 を計算します。自由度はパラメータの数の差なので、この場合1になります。
- 22



1
2
3
4
5
6
7
8
9
10
11
12

この場合、p 値が 0.01027565 なので、5%の危険水準でバリング飛行を行うグループと行わないグループでは、 ω に有意差があることが分かります。

もしも ω が 3 種類以上あって、 $\omega_{\#0} = \omega_{\#1} \neq \omega_{\#2}$ などの条件で検証を行いたい場合は、コントロールファイルの model の設定ではなく、樹形ファイルの #1、#2 などに変更を加えるといいです。#2 を一括変換で #1 に変更して、CODEML プログラムで尤度を計算、有意差があるかを検証という流れになります。

1 ■2項：サイトモデルを用いた選択圧の推定
2 **ホバリング飛行を行うグループのミトコンドリア全タンパク質コード遺伝子には正の選択**
3 **圧が働いているのだろうか？**

4
5 ホバリング飛行を行うグループと行わないグループでは ω に有意な違いがあることがわか
6 りました。この場合、以下の2通りの解釈が可能です。

7
8 解釈1：ホバリング飛行を行うグループでは**負の選択圧が緩和している。**

9 解釈2：ホバリング飛行を行うグループでは**正の選択圧が働いている。**

10
11 枝モデルの解析結果のみでは、この二つの仮説を検証することが出来ません。
12 正の選択圧が働いている場合、 $\omega > 1$ になりますが、多くの場合正の選択圧はコドン配列
13 全体のうち、いくつかのコドンサイトにしか働かないため、正の選択圧が働いている場合
14 であってもコドン配列全体で推定した ω は1よりもずっと小さくなります。

15
16 このような場合、サイトモデルを用いて、正の選択圧が働いているコドンサイトが存在す
17 るか否かを検証することが出来ます。

18
19 **1. アラインメントファイルの編集**

20 ここではホバリング飛行を行うグループのデータのみを用いたいと思います。
21 ホバリング飛行を行わないアマツバメ亜目やズクヨタカ科の仲間を塩基配列ごとアライン
22 メントファイルから削除して、名前を付けて保存しましょう。

23
24 削除するのは
25 *Apus apus* NC 008540
26 *Chaetura pelagica* KT809406
27 *Cypseloides fumigatus* KY688216
28 *Aegotheles cristatus* EU344979
29 の4種です。

30
31 この例題では **hummingbird_12mtCDS15sp.fas** として保存することにします。

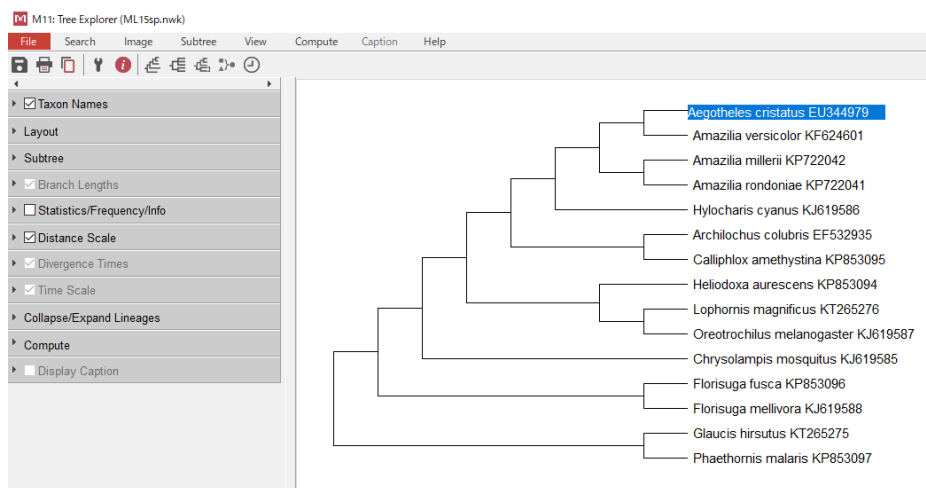
32
33 **2. 樹形ファイルの編集**

34 樹形ファイルからも上記の4種を削除しましょう。
35 赤く色をつけた部分を削除して名前を付けて保存すれば大丈夫です。

36

1 **(Aegotheles cristatus EU344979,(((((((Amazilia brevirostris_KP722043,Amazilia versicol**
 2 **or_KF624601),(Amazilia millerii_KP722042,Amazilia rondoniae_KP722041))),Hylocharis_**
 3 **cyanus_KJ619586),(Archilochus colubris_EF532935,Calliphlox amethystina_KP853095)),(**
 4 **Heliodoxa aurescens_KP853094,(Lophornis magnificus_KT265276,Oreotrochilus melano**
 5 **gaster_KJ619587))),Chrysolampis mosquitos_KJ619585),((Florisuga fusca_KP853096,Flor**
 6 **isuga mellivora_KJ619588),(Glaucis hirsutus_KT265275,Phaethornis malaris_KP853097)**
 7 **)),((Apus apus_NC_008540,Chaetura pelagica_KT809406),Cypseloides fumigatus_KY688**
 8 **216));**

9
 10 なれないうちは newick フォーマットをマニュアルで編集するのは難しいです。フォー
 11 ットが壊れていないか、ちゃんと希望する樹形になっているかは MEGA で可視化して確
 12 認しましょう。フォーマットが壊れていると MEGA がエラーメッセージを出してくれま
 13 す。



14 この例題では **ML15sp.nwk** として保存することにします。

17 3. コントロールファイルの編集

18 コントロールファイルは、自分が使うアラインメントファイル、樹形ファイル、そして好
 19 きなアウトファイルの名前を記入しておきます。

21 **seqfile = hummingbird_12mtCDS15sp.fas** * sequence data filename

22 **treefile = ML15sp.nwk** * tree structure file name

23 **outfile = site.out** * main result file name

24
 25 サイトモデルの設定で特に大事なのが **model** と **NSsites** のオプションです。

```

model = 0
* models for codons:
* 0:one, 1:b, 2:2 or more dN/dS ratios for branches
* models for AAs or codon-translated AAs:
* 0:poisson, 1:proportional, 2:Empirical, 3:Empirical+F
* 6:FromCodon, 7:AAClasses, 8:REVaa_0, 9:REVaa(nr=189)

NSsites = 0 1 2 * 0:one w;1:neutral;2:selection; 3:discrete;4:freqs;
* 5:gamma;6:2gamma;7:beta;8:beta&w;9:beta&gamma;
* 10:beta&gamma+1; 11:beta&normal>1; 12:0&2normal>1;
* 13:3normal>0

```

1

2 サイトモデルでは、系統樹の枝間では ω に違いがないことを仮定しているので、

3

model = 0

4 を選択します。

5 NSsites に関しては

6

NSsites = 0 1 2

7

8 としておくと3つのサイトモデル (NSsites=0 (一比率モデル), NSsites=1 (中立モデル) ,
9 NSsites=2 (選択モデル)) を順次計算してくれます。

10 一比率モデル(NSsites=0)は配列全体ですべてのコドンサイトが同じ ω を持つと仮定する
11 モデルです。従ってこのモデルでは ω に関連したパラメータは1つのみになります。

12 中立モデル(NSsites=1)は配列全体が、負の選択圧を受けているコドンサイトと中立に進化
13 するコドンサイトに分かれることを仮定しています。このモデルは、負の選択圧を受け
14 ているコドンサイトの ω (ω_0 ;ただし $\omega_0 < 1$)、中立に進化するコドンサイトの ω (ω_1) の
15 ほか、負の選択圧を受けているコドンサイトの割合 P_0 と中立に進化するコドンサイトの割
16 合 P_1 をパラメータとして持ちますが、 $\omega_1 = 1$ であり、また $P_0 + P_1 = 1$ という制約を持つ
17 ので、自由パラメータの数は2つになります。

18 選択モデル(NSsites=2)は中立モデルに加えて、配列中には正の選択圧を受けているコド
19 ンサイトが存在すること仮定しています。正の選択圧を受けているコドンサイトの ω は ω
20 ω_2 (ただし $\omega_2 > 1$)、その割合は P_2 (ただし $P_0 + P_1 + P_2 = 1$) になり、 ω に関連したパラメータ
21 を4つ持つこととなります。

22 ここで、中立モデルは選択モデルの特殊なケース、一比率モデルは中立モデルの特殊な
23 ケースであることに留意しましょう。例えば選択モデルにおいて $P_2 = 0$ もしくは $\omega_2 = 1$ と
24 いう状況下は中立モデルと等しくなります。このようにモデル同士が入れ子状になってい
25 る場合は、尤度比検定によりモデルを評価することが出来ます。

26

27 **知っておくと便利なお役立ち情報**

28 サイトモデルでは、一比率モデル、中立モデル、選択モデルがよく用いられますが、ほか
29 にも ω_1 や ω_2 のとりうる範囲に $=1$ や >1 などの制約を持たない離散モデル(NSsites=3)、負
30 の選択圧を受けているコドンサイトで ω が β 分布をしていることを仮定する β モデル

1 (NSsites=7)、そこにさらに正の選択圧を受けているコドンサイトが存在するとする β & ω
2 モデル(NSsites=8)など様々なモデルがあります。

3

4 3. 解析の実行

5 実行ファイル **codeml.exe** をダブルクリックするか、コマンドプロンプトで **codeml** とタイ
6 プしてエンターキーを押すと、CODEML プログラムによる解析が始まります。

7

8 5. 解析結果の確認と統計的有意性の評価

9 アウトファイルの下部のほうに、3つのモデル（一比率モデル、中立モデル、選択モデ
10 ル）による解析結果がまとめて表示されます。

```
461
462 Model 0: one-ratio
463
464
465 TREE # 1: ((((((1, 4), (2, 3)), 12), (5, 6)), (11, (13, 14))), 7), (8, 9), (10, 15)); MP score: -1
466 lnL(ntime: 27 np: 29): -52375.675174 +0.000000
467 16..17 17..18 18..19 19..20 20..21 21..22 22..1 22..4 21..23 23..2 23..3 20..12 1
468 0.149357 0.199762 0.106682 0.306663 0.144552 0.033247 0.017880 0.021339 0.023306 0.046943 0.046328 0.209087 0.
469
470 Note: Branch length is defined as number of nucleotide substitutions per codon (not per nucleotide site).
471
472 tree length = 6.044052
473
474 ((((((1: 0.017880, 4: 0.021339): 0.033247, (2: 0.046943, 3: 0.046328): 0.023306): 0.144552, 12: 0.209087): 0.
475
476 ((((((Amazilia_brevirostris_KP722043: 0.017880, Amazilia_versicolor_KF624601: 0.021339): 0.033247, (Amazilia_
477
478 Detailed output identifying parameters
479
480 kappa (ts/tv) = 11.16001
481
482 omega (dN/dS) = 0.02226
483
484 dN & dS for each branch
485
486 branch t N S dN/dS dN dS N*dN S*dS
487
488 16..17 0.149 8303.6 2454.4 0.0223 0.0045 0.2029 37.5 498.1
489 17..18 0.200 8303.6 2454.4 0.0223 0.0060 0.2714 50.2 666.2
490 18..19 0.107 8303.6 2454.4 0.0223 0.0032 0.1450 26.8 355.8
491 19..20 0.306 8303.6 2454.4 0.0223 0.0093 0.4159 76.9 1020.7
492 20..21 0.145 8303.6 2454.4 0.0223 0.0044 0.1964 36.3 482.1
493 21..22 0.033 8303.6 2454.4 0.0223 0.0010 0.0452 8.3 110.9
494 22..1 0.018 8303.6 2454.4 0.0223 0.0005 0.0243 4.5 59.6
495 22..4 0.021 8303.6 2454.4 0.0223 0.0006 0.0290 5.4 71.2
496 21..23 0.023 8303.6 2454.4 0.0223 0.0007 0.0317 5.9 77.7
497 23..2 0.047 8303.6 2454.4 0.0223 0.0014 0.0638 11.8 156.5
498 23..3 0.046 8303.6 2454.4 0.0223 0.0014 0.0629 11.6 154.5
499 20..12 0.209 8303.6 2454.4 0.0223 0.0063 0.2841 52.5 697.3
500 19..24 0.265 8303.6 2454.4 0.0223 0.0080 0.3602 66.6 884.2
501 24..5 0.167 8303.6 2454.4 0.0223 0.0051 0.2275 42.0 558.3
502 24..6 0.221 8303.6 2454.4 0.0223 0.0067 0.3006 55.6 737.9
503 18..25 0.051 8303.6 2454.4 0.0223 0.0015 0.0687 12.7 168.5
504 25..11 0.479 8303.6 2454.4 0.0223 0.0145 0.6510 120.3 1597.9
505 25..26 0.091 8303.6 2454.4 0.0223 0.0027 0.1230 22.7 301.0
506 26..13 0.598 8303.6 2454.4 0.0223 0.0181 0.8122 150.1 1993.3
507 26..14 0.348 8303.6 2454.4 0.0223 0.0105 0.4730 87.4 1160.8
508 17..7 0.627 8303.6 2454.4 0.0223 0.0190 0.8516 157.4 2090.1
509 16..27 0.443 8303.6 2454.4 0.0223 0.0134 0.6019 111.3 1477.4
510 27..8 0.197 8303.6 2454.4 0.0223 0.0060 0.2679 49.5 657.4
511 27..9 0.194 8303.6 2454.4 0.0223 0.0059 0.2648 48.8 647.9
512 16..28 0.483 8303.6 2454.4 0.0223 0.0146 0.6566 121.4 1611.6
513 28..10 0.255 8303.6 2454.4 0.0223 0.0077 0.3463 64.0 849.9
514 28..15 0.320 8303.6 2454.4 0.0223 0.0097 0.4353 80.5 1068.3
515
516 tree length for dN: 0.1828
517 tree length for dS: 8.2123
518
519
520 Time used: 1:39
521
522
523 Model 1: NearlyNeutral (2 categories)
524
525
526 TREE # 1: ((((((1, 4), (2, 3)), 12), (5, 6)), (11, (13, 14))), 7), (8, 9), (10, 15)); MP score: -1
527 lnL(ntime: 27 np: 30): -51949.796293 +0.000000
528 16..17 17..18 18..19 19..20 20..21 21..22 22..1 22..4 21..23 23..2 23..3 20..12 1
529 0.154342 0.202098 0.102482 0.312720 0.142951 0.033102 0.018004 0.021031 0.023230 0.046691 0.046002 0.206337 0.
530
531 Note: Branch length is defined as number of nucleotide substitutions per codon (not per nucleotide site).
532
533 tree length = 6.120714
534
535 ((((((1: 0.018004, 4: 0.021031): 0.033102, (2: 0.046691, 3: 0.046002): 0.023230): 0.142951, 12: 0.206337): 0.
536
537 ((((((Amazilia_brevirostris_KP722043: 0.018004, Amazilia_versicolor_KF624601: 0.021031): 0.033102, (Amazilia_
538
539 Detailed output identifying parameters
540
541 kappa (ts/tv) = 11.56937
542
543
544 MLEs of dN/dS (w) for site classes (K=2)
```

11

12

1 **Model 0: one-ratio** と書かれているところから**一比率モデル**の結果がまとめられています。

```
Model 0: one-ratio

TREE # 1: ((((((1, 4), (2, 3)), 12), (5, 6)), (11, (13, 14))), 7), (8, 9))
lnL(ntime: 27 np: 29): -52375.675174 +0.000000
 16..17 17..18 18..19 19..20 20..21 21..22 22..1 22..4
 0.149357 0.199762 0.106682 0.306063 0.144552 0.033247 0.017880 0.021339

Note: Branch length is defined as number of nucleotide substitutions per
tree length = 6.044052

((((((1: 0.017880, 4: 0.021339): 0.033247, (2: 0.046943, 3: 0.046328):
0.197135, 9: 0.194294): 0.443007, (10: 0.254863, 15: 0.320353): 0.483267

((((((Amazilia_brevirostris_KP722043: 0.017880, Amazilia_versicolor_KF6
Archilochus_colubris_EF532935: 0.167423, Calliphlox_amethystina_KP853095
Chrysolampis_mosquitus_KJ619585: 0.626749): 0.149357, (Florisuga_fusca_K

Detailed output identifying parameters

kappa (ts/tv) = 11.16001

omega (dN/dS) = 0.02226

dN & dS for each branch

branch      t      N      S      dN/dS      dN      dS      N*dN      S*dS
16..17      0.149  8303.6  2454.4  0.0223     0.0045  0.2029  37.5  498.1
17..18      0.200  8303.6  2454.4  0.0223     0.0060  0.2714  50.2  666.2
18..19      0.107  8303.6  2454.4  0.0223     0.0032  0.1450  26.8  355.8
```

2
3 このモデルでは、(コドン頻度に関するパラメータを除くと)パラメータ数は29、対数尤
4 度は**-52375.675174**であり、 $\omega = 0.02226$ になっています。

5
6 その下の**Model 1: NearlyNeutral (2 categories)**と書かれているところから**中立モデル**の
7 結果がまとめられています。

```
Model 1: NearlyNeutral (2 categories)

TREE # 1: ((((((1, 4), (2, 3)), 12), (5, 6)), (11, (13, 14))), 7), (8, 9))
lnL(ntime: 27 np: 30): -51949.796293 +0.000000
 16..17 17..18 18..19 19..20 20..21 21..22 22..1 22..4
 0.154342 0.202098 0.102482 0.312720 0.142951 0.033102 0.018004 0.0210

Note: Branch length is defined as number of nucleotide substitutions p
tree length = 6.120714

((((((1: 0.018004, 4: 0.021031): 0.033102, (2: 0.046691, 3: 0.046002):
0.185157, 9: 0.202747): 0.464650, (10: 0.244778, 15: 0.326188): 0.5172

((((((Amazilia_brevirostris_KP722043: 0.018004, Amazilia_versicolor_K
Archilochus_colubris_EF532935: 0.162912, Calliphlox_amethystina_KP8530
Chrysolampis_mosquitus_KJ619585: 0.637101): 0.154342, (Florisuga_fusca

Detailed output identifying parameters

kappa (ts/tv) = 11.56937

MLEs of dN/dS (w) for site classes (K=2)

p: 0.97090 0.02910
w: 0.01535 1.00000

dN & dS for each branch
```

8
9

- 1 このモデルでは、(コドン頻度に関するパラメータを除くと) パラメータ数は 30 になって
- 2 おりパラメータ数がひとつ増えていることが分かります。
- 3 ω に関するパラメータを詳細に見てみましょう。

```
MLEs of dN/dS ( $\omega$ ) for site classes (K=2)
```

```
p: 0.97090 0.02910  
w: 0.01535 1.00000
```

- 4
- 5 p: 0.97090 0.02910
- 6 は、負の選択圧を受けているコドンサイトの割合($P_0=0.97090$)と中立に進化するコドンサ
- 7 イトの割合($P_1=0.02910$)を表しています。(※ $P_0+P_1=1$ になることに注意)
- 8
- 9 w: 0.01535 1.00000
- 10 は、負の選択圧を受けているコドンサイトの ω ($\omega_0=0.01535$)と中立に進化するコドンサイ
- 11 トの ω ($\omega_1=1$)を表しています。中立に進化するコドンサイトが存在するため、 ω_0 は、一
- 12 比率モデルで推定した ω (=0.02226)よりもいくらか小さな値になっています。
- 13
- 14 中立に進化するコドンサイトの割合は全体の 2.91%と推定されており、ずいぶん少ないよ
- 15 うに感じますが、中立モデルのもとで対数尤度は-51949.796293 であり、一比率モデルの
- 16 それと比べると大きく改善しています (+ 425.878881)。尤度比検定により p 値を推定す
- 17 ると 3.016E-187 となり統計的に有意であることが分かります。このことは負の選択をう
- 18 けているサイトだけでなく、中立的に進化しているコドンサイトも存在すると考えたほう
- 19 が、今回のハチドリのみトコンドリアのデータへの当てはまりが良いことを示していま
- 20 す。
- 21
- 22 最後に、Model 2: PositiveSelection (3 categories) と書かれているところから選択モデル
- 23 の結果がまとめられています。
- 24
- 25

```

Model 2: PositiveSelection (3 categories)

TREE # 1: ((((((1, 4), (2, 3)), 12), (5, 6)), (11, (13, 14))), 7),
lnL(ntime: 27 np: 32): -51949.796293 +0.000000
 16..17 17..18 18..19 19..20 20..21 21..22 22..1 22..
0.154342 0.202099 0.102482 0.312721 0.142951 0.033102 0.018004 0.021
0.015354 1.000000

Note: Branch length is defined as number of nucleotide substitutions
tree length = 6.120731

(((((((1: 0.018004, 4: 0.021031): 0.033102, (2: 0.046691, 3: 0.046002
0.185157, 9: 0.202747): 0.464652, (10: 0.244778, 15: 0.326189): 0.517

(((((((Amazilia_brevirostris_KP722043: 0.018004, Amazilia_versicolor_
Archilochus_colubris_EF532935: 0.162912, Calliphlox_amethystina_KP853
Chrysolampis_mosquitus_KJ619585: 0.637104): 0.154342, (Florisuga_fusc

Detailed output identifying parameters

kappa (ts/tv) = 11.56942

MLEs of dN/dS (w) for site classes (K=3)

p: 0.97090 0.00000 0.02909
w: 0.01535 1.00000 1.00000

```

1
2 このモデルでは、(コドン頻度に関するパラメータを除くと) パラメータ数は 32 になって
3 います。

4 ω に関するパラメータに関しては

5
6 **p: 0.97090 0.00000 0.02909**

7 は、負の選択圧を受けているコドンサイトの割合($P_0=0.97090$)、中立に進化するコドンサ
8 イトの割合($P_1=0$)、正の選択圧を受けているコドンサイトの割合($P_2=0.02909$)を表してい
9 ます (※ $P_0+P_1+P_2=1$ になることに注意)

10
11 **w: 0.01535 1.00000 1.00000**

12 は、負の選択圧を受けているコドンサイトの ω ($\omega_0=0.01535$)、中立に進化するコドンサイ
13 トの ω ($\omega_1=1$)、正の選択圧を受けているコドンサイトの ω ($\omega_2=1$)を表しています。ここ
14 では正の選択圧を受けているコドンサイトの ω が 1 となっていることから、正の選択圧を
15 受けているコドンサイトは実質存在しないことが分かります。

16
17 実際、このモデルの対数尤度は -51949.796293 となっており、中立モデルで推定したそれ
18 と厳密に一致しています。尤度比検定で p 値をもとめると 1 になります。

19
20 このように尤度比検定を用いた選択モデルと中立モデルの比較から、ホバリング飛行を行
21 うハチドリの系統全体で、正の選択圧をうけているコドンサイトは存在しないことがわか
22 ります。

1 なお

2

```
Bayes Empirical Bayes (BEB) analysis (Yang, Wong & Nielsen 2005. Mol. Biol. Evol. 22:1107-1118)
Positively selected sites (*: P>95%; **: P>99%)
(amino acids refer to 1st sequence: Amazilia_brevirostris_KP722043)

      Pr(w>1)      post mean +- SE for w
1055 A      0.607      1.304 +- 0.244
2193 S      0.567      1.283 +- 0.248
2617 L      0.651      1.325 +- 0.238
3022 R      0.550      1.275 +- 0.249
```

3

4

5 計算終了時にフォルダ内に自動的に作成される rst ファイルには、すべてのコドンサイト
6 について、各々のサイトが負の選択圧を受けているコドンサイトに属する事後確率、中立
7 に進化するコドンサイトに属する事後確率、正の選択圧を受けているコドンサイトに属す
8 る事後確率が、各サイトの ω の事後平均値とともにまとめられています。

9

10 知っておくと便利なお役立ち情報

11 今回の例題では正の選択を受けているサイトの ω が 1 になっており、このデータからは正
12 の選択を受けているコドンサイトは存在しないということが示されましたが、同じデータ
13 を使って計算しても少し違った結果が出てくることがあります。

14 例えば

15 p: 0.97120 0.02880 0.00000

16 w: 0.01530 1.00000 43.70712

17 というように正の選択を受けているサイトの ω が 1 よりもずっと大きい値になりますが、
18 そのサイトの割合が 0 という場合です。

19 いずれの場合においても「正の選択を受けているコドンサイトは存在しない」という点で
20 は同じですが、このように一見異なる結果が出てくる場合もあります。

21

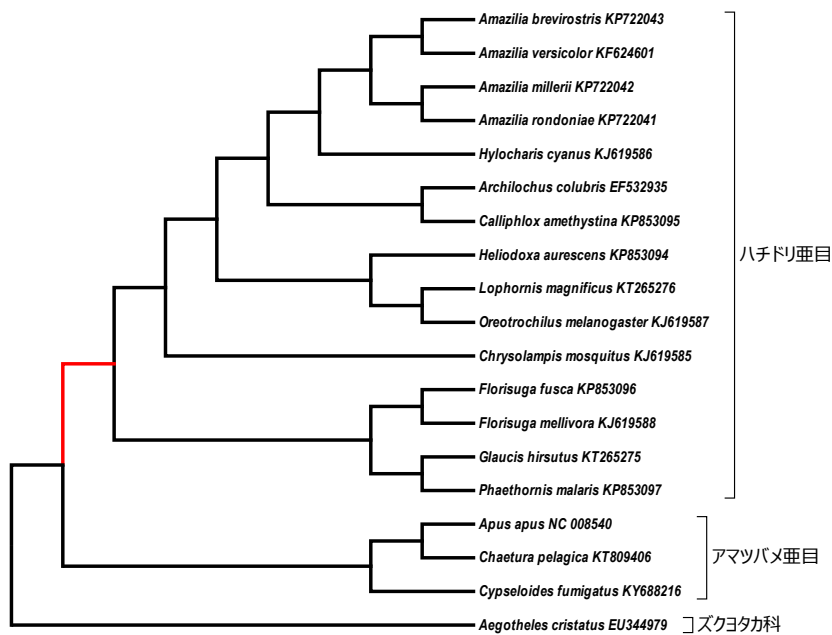
22

1 ■3項：枝サイトモデルを用いた選択圧の推定

2 非ホバリング飛行からホバリング飛行に移行する過程でミトコンドリア全タンパク質コード
3 遺伝子に正の選択圧が働いたのだろうか？

4

5 サイトモデルによる解析でホバリング飛行を行うハチドリの系統全体で、正の選択圧をう
6 けているコドンサイトは存在しないことが分かりましたが、非ホバリング飛行からホバリ
7 ング飛行に移行する過程（赤で示した枝）では何か特別なことが起きていた可能性がありま
8 す。



9

10 このように特定の枝で起きた正の選択の痕跡を検出したい場合は、枝サイトモデルを用い
11 ます。

12

13 1. コントロールファイルの編集

14 ここでは枝モデルを用いた解析で使ったアラインメントファイル

15 (hummingbird_12mtCDS.fas)をそのまま使うことにしましょう。コントロールファイル
16 と樹形ファイル(ML.nwk)は枝モデルの解析で使ったファイルに少しだけ手を加えます。

17 **Outfile** は好きな名前（ここでは **branchsite.out** としておきましょう）をつけます。

18

19 **seqfile = hummingbird_12mtCDS.fas** * sequence data filename

20 **treefile = ML.nwk** * tree structure file name

21 **outfile = branchsite.out** * main result file name

22

1 枝サイトモデルの設定でも、特に大事なのはやはり **model** と **NSsites** のオプションで
2 す。

3

```
model = 2
* models for codons:
* 0:one, 1:b, 2:2 or more dN/dS ratios for branches
* models for AAs or codon-translated AAs:
* 0:poisson, 1:proportional, 2:Empirical, 3:Empirical+F
* 6:FromCodon, 7:AAClasses, 8:REVaa_0, 9:REVaa(nr=189)

NSsites = 2
* 0:one w;1:neutral;2:selection; 3:discrete;4:freqs;
* 5:gamma;6:2gamma;7:beta;8:beta&w;9:beta&gamma;
* 10:beta&gamma+1; 11:beta&normal>1; 12:0&2normal>1;
* 13:3normal>0
```

4

5

6 枝サイトモデルでは、系統樹の枝を正の選択を受けている前景の枝と正の選択を受けてい
7 ない背景の枝という 2 群に分けます。従って枝モデルの時と同様に

8

9

model = 2

10

11 を選びましょう。(繰り返しになりますが 2 を選択すると、系統樹の枝が 2 つ以上のグル
12 ープに分かれ、グループごとに ω を推定します。枝をどのようにグループ分けするかは樹
13 形ファイルに # を書き込むことで指定します。)

14

15 また枝サイトモデルでは、コドンサイトを正の選択を受けているサイト、負の選択を受け
16 ているサイト、中立に進化するサイトに分けます。従ってサイトモデルの時にも選択した

17

18

NSsites = 2

19

20 としておくと NSsites=2 (選択モデル) で計算してくれます。サイトモデルの時と同様に
21 選択モデルは負の選択圧を受けているサイト、中立進化をするサイト、正の選択圧を受け
22 ているサイトが存在すること仮定しています。

23

24 これで枝サイトモデルの基本設定は完了です。

25

26 2. 樹形ファイルの編集

27 枝サイトモデルでは、前景となる枝をシャープ(#1)で指定します。ここではハチドリ亜目
28 全体の共通祖先の枝を前景にしたいので

29

1 (Aegotheles_cristatus_EU344979,(((((((Amazilia_brevirostris_KP722043,Amazilia_versicol
2 or_KF624601),(Amazilia_millerii_KP722042,Amazilia_rondoniae_KP722041)),Hylocharis_
3 cyanus_KJ619586),(Archilochus_colubris_EF532935,Calliphlox_amethystina_KP853095)),(
4 Heliodoxa_aurescens_KP853094,(Lophornis_magnificus_KT265276,Oreotrochilus_melano
5 gaster_KJ619587))),Chrysolampis_mosquitus_KJ619585),((Florisuga_fusca_KP853096,Flor
6 isuga_mellivora_KJ619588),(Glaucis_hirsutus_KT265275,Phaethornis_malaris_KP853097)
7))#1,((Apus_apus_NC_008540,Chaetura_pelagica_KT809406),Cypseloides_fumigatus_KY6
8 88216));

9

10 上記の赤い文字で#1と書いている箇所に、#1を書き込んで樹形ファイルを好きな名前で
11 保存しましょう（ここでは **ML.nwk** とします）。

12

13 3. 解析の実行

14 実行ファイル **codeml.exe** をダブルクリックするか、コマンドプロンプトで **codeml** とタイ
15 プしてエンターキーを押すと、CODEML プログラムによる解析が始まります。

16

17 4. 解析結果の確認

18 計算が終了するとアウトファイル（ここでは **branchsite.out**）に詳細な解析結果が記録さ
19 れます。アウトファイルの下部には樹形やパラメータの数、枝の長さなどの情報が記録さ
20 れています。

```
TREE # 1: (1, ((((((2, 5), (3, 4)), 16), (7, 8)), (15, (17, 18))), 10), ((12, 13), (14, 19)
lnl(ntime: 35 np: 40): -68067.657331 +0.000000
 20..1 20..21 21..22 22..23 23..24 24..25 25..26 26..27 27..2 27..5 26
 20..35 35..36 36..6 36..9 35..11
1.341661 1.040524 0.065828 0.172412 0.092155 0.292370 0.138915 0.032873 0.017989 0.020930 0.0
0.465409 0.270369 0.444724 0.421293 0.486134 9.802492 0.956590 0.030179 0.015539 4.367966

Note: Branch length is defined as number of nucleotide substitutions per codon (not per nucle
tree length = 10.191542

(1: 1.341661, ((((((2: 0.017989, 5: 0.020930): 0.032873, (3: 0.046143, 4: 0.045407): 0.023346
0.065828, ((12: 0.180872, 13: 0.195642): 0.419545, (14: 0.232202, 19: 0.307086): 0.453457): 0.
(Aegotheles_cristatus_EU344979: 1.341661, ((((((Amazilia_brevirostris_KP722043: 0.017989, Ama
0.200031): 0.292370, (Archilochus_colubris_EF532935: 0.159498, Calliphlox_amethystina_KP853095
0.042119): 0.172412, Chrysolampis_mosquitus_KJ619585: 0.590990): 0.065828, ((Florisuga_fusca_K
((Apus_apus_NC_008540: 0.444724, Chaetura_pelagica_KT809406: 0.421293): 0.270369, Cypseloides_

Detailed output identifying parameters

kappa (ts/tv) = 9.80249

MLEs of dN/dS (w) for site classes (K=4)
site class 0 1 2a 2b
proportion 0.95659 0.03018 0.01283 0.00040
background w 0.01554 1.00000 0.01554 1.00000
foreground w 0.01554 1.00000 4.36797 4.36797
```

21

22 ここで、**site class** と書かれている箇所に着目しましょう。

23 枝サイトモデルでは、コドンサイトが以下の4つの site class に分けられると仮定していま
24 す。

site class	0	1	2a	2b
割合	p_0	p_1	$(1-p_0-p_1)p_0/(p_0+p_1)$	$(1-p_0-p_1)p_1/(p_0+p_1)$
背景の ω	$0 < \omega_0 < 1$	$\omega_1 = 1$	$0 < \omega_0 < 1$	$\omega_1 = 1$
前景の ω	$0 < \omega_0 < 1$	$\omega_1 = 1$	$1 < \omega_2$	$1 < \omega_2$

1 Site class 0 は前景も背景も等しく負の選択を受けており、その ω (ω_0) は 0 以上 1 以下であ
2 ると仮定します。Site class 0 が全サイトに占める割合は p_0 です。Site class 1 は前景も背景
3 も中立進化をしていると仮定し、その ω (ω_1) は 1 です。Site class 1 の割合は p_1 です。Site
4 class 2a では背景は負の選択を受けており、その ω は ω_0 ですが、前景は正の選択を受けて
5 おりその ω は ω_2 ($1 < \omega_2$) です。Site class 2a の割合は $(1-p_0-p_1)p_0/(p_0+p_1)$ になります。Site
6 class 2b では背景は中立に進化している一方で ($\omega_1=1$)、前景は正の選択を受けています
7 ($1 < \omega_2$)。Site class 2a の割合は $(1-p_0-p_1)p_1/(p_0+p_1)$ になります。

8

9 以上のことから枝サイトモデルでは、枝の長さや κ (トランジション/トランスバージョン
10 比) などのほかに、これら 4 つの自由パラメータ (ω_0 、 ω_2 、 p_0 、 p_1) を持つことが分か
11 ります。

12

13 Site class 2a と site class 2b が正の選択を受けているサイトですが、今回のデータの解析結
14 果では、site class 2a と site class 2b の合計は、全体の約 1.3% を占めていることが分かりま
15 す。アラインメント全長が 3586 コドンサイトなので、約 47 コドンサイトが正の選択を受
16 けていると推定されました。これらの正の選択を受けているコドンサイトの ω (ω_2) は
17 4.36797 と推定されています。

18

19 最尤法では、全体のうちのどれだけの割合のサイトが正の選択を受けているか推定されま
20 すが、具体的にどのコドンサイトが正の選択を受けているかは経験ベイズにより推定さ
21 れ、アウトファイルの下部のほうに事後確率とともにまとめられています。事後確率が
22 0.95 以上のサイトにはアスタリスクがつけられます。

```

Bayes Empirical Bayes (BEB) analysis (Yang, Wong & Nielsen 2005. Mol. Biol. Evol. 22:1107-1118)
Positive sites for foreground lineages Prob(w>1):
49 L 0.782
78 M 0.790
261 N 0.802
1058 S 0.949
1899 L 0.502
2049 S 0.833
2677 L 0.606
2833 L 0.849
2846 S 0.847
3172 A 0.778
3393 L 0.802
3402 G 0.946
3504 S 0.959*

```

23

24

25 5. 統計的有意性の評価

1 正の選択を受けているコドンサイトの割合が0に近い場合や、これらのコドンサイトの
2 ω が1に近い場合は、実際には正の選択を受けていない可能性もあります。そこで統計的
3 有意性を評価する必要があります。ここでは正の選択を受けているサイト (site class2a と
4 site class2b) の前景の枝の ω が1と仮定し、これを帰無仮説として用います。

5
6 コントロールファイルの設定を以下のように変えましょう。ここではアウトファイルの名
7 前を **branchsite.null.out** とします。

```
9 seqfile = hummingbird_12mtCDS.fas * sequence data filename  
10 treefile = ML.nwk * tree structure file name  
11 outfile = branchsite.null.out * main result file name
```

12
13 **model** と **NSsites** のオプションは

14
15 **model = 2**

16
17 **NSsites = 2**

18
19 を選びますが、 ω の推定に関するオプションを以下のように

```
fix_kappa = 0 * 1: kappa fixed, 0: kappa to be estimated  
kappa = 2 * initial or fixed kappa  
fix_omega = 1 * 1: omega or omega_1 fixed, 0: estimate  
omega = 1 * initial or fixed omega, for codons or codon-based AAs
```

20
21 とします。ここで重要なことは

22 **fix_omega = 1**

23
24 **omega = 1**

25
26 とすることです。

27 **fix_omega = 1** にすることで、 ω は最尤推定されず固定された値になります。枝サイトモ
28 デルの設定では、この ω は正の前景における正の選択圧を受けているサイトの ω (ω_2)を意
29 味します。そして **omega = 1** にすることでこの ω_2 の初期値は1となり、これがそのまま
30 固定されます。こうすることで帰無仮説を設定することが出来ます。

31

- 1 ω_2 を自由パラメータとして尤度推定を行うと-68067.657331 という値になり、 ω_2 を1に
- 2 固定して尤度推定を行うと-68096.166835 という値になりました。これをもとに尤度比検
- 3 定を行うと p 値は 4.3162E-14 という非常に低い値になります。このことは ω_2 が1よりも
- 4 有意に大きいことを意味します。
- 5

第三章：分岐年代推定

ハチドリのミトコンドリア全タンパク質コード遺伝子から分岐年代を推定してみよう

ここでは分子時計を仮定した分岐年代推定と緩和型分子時計による分岐年代推定を行います。PAML の BASEML プログラムと MCMCTREE プログラムを用います。分子時計を仮定した分岐年代推定（最尤法）は BASEML プログラムを用いて、緩和型分子時計による分岐年代推定（ベイズ法）は BASEML プログラムと MCMCTREE プログラムの両方を用いて実行します。特に緩和型分子時計は①枝の長さとその分散・共分散の推定と②分岐年代の推定という 2 段階に分けて実行すること、ベイズ推定に必要な事前確率分布の設定を行う必要があることから少し複雑です。

※本文ではコドンの 1 番目、2 番目、3 番目を区別して分岐年代推定を行っていますが、例題としては複雑になりすぎるので、ここでは区別せずに解析しています。

1 節：準備

PAML を用います。PAML のダウンロードは序章 1 節を参照してください。PAML の基本的な構成と操作は第二章 1 節～2 節に書かれているように、PAML は基本的に実行ファイル、コントロールファイル、アラインメントファイル、樹形ファイルの 4 点セットを必要とします。

2 節：分子時計を仮定した分岐年代推定

1 項：準備

PAML の①実行ファイル（BASEML プログラム：**baseml.exe**）、②コントロールファイル（**baseml.cti**）、③アラインメントファイル（**hummingbird_12mtCDS.fas**）、④樹形ファイル（**ML.nwk**）を準備し、ひとつのフォルダ（ここではとりあえず **clock** という名前のフォルダにしましょう）にいます。**ML.nwk** については、**4 節：樹形ファイルの出力**で作成したものを uses。

（※**ML.nwk** については第二章の選択圧の推定で使った場合は#などは取り除いておいてください）

2 項：分子時計を仮定しないモデルによる尤度推定

まず分子時計を仮定せずに枝の長さなどのパラメータを最尤推定し、尤度を計算します。

1. コントロールファイルの編集

- 1 BASEML プログラムのコントロールファイルはデフォルトでは以下のようになっていま
2 す。

```
seqfile = brown.nuc
treefile = brown.trees

outfile = mlb      * main result file
noisy = 2          * 0,1,2,3: how much rubbish on the screen
verbose = 0       * 1: detailed output, 0: concise output
runmode = 0       * 0: user tree; 1: semi-automatic; 2: automatic
                  * 3: StepwiseAddition; (4,5):PerturbationNNI

model = 4         * 0:JC69, 1:K80, 2:F81, 3:F84, 4:HKY85
                  * 5:T92, 6:TN93, 7:REV, 8:UNREST, 9:REVu; 10:UNRESTu

Mgene = 0        * 0:rates, 1:separate; 2:diff pi, 3:diff kapa, 4:all diff

*
ndata = 100
clock = 0        * 0:no clock, 1:clock; 2:local clock; 3:CombinedAnalysis
fix_kappa = 0    * 0: estimate kappa; 1: fix kappa at value below; 2: kappa for branches
kappa = 5        * initial or fixed kappa

fix_alpha = 0    * 0: estimate alpha; 1: fix alpha at value below
alpha = 0.5     * initial or fixed alpha, 0:infinity (constant rate)
Malpha = 0      * 1: different alpha's for genes, 0: one alpha
ncatG = 5       * # of categories in the dG, AdG, or nparK models of rates
nparK = 0       * rate-class models. 1:rK, 2:rK&fK, 3:rK&MK(1/K), 4:rK&MK

nhomo = 0       * 0 & 1: homogeneous, 2: kappa for branches, 3: N1, 4: N2
getSE = 0       * 0: don't want them, 1: want S.E.s of estimates
RateAncestor = 0 * (0,1,2): rates (alpha>0) or ancestral states

Small_Diff = 7e-6
cleandata = 1   * remove sites with ambiguity data (1:yes, 0:no)?
* icode = 0     * (with RateAncestor=1. try "GC" in data,model=4,Mgene=4)
* fix_blength = 1 * 0: ignore, -1: random, 1: initial, 2: fixed, 3: proportional
method = 0     * Optimization method 0: simultaneous; 1: one branch a time
```

- 3
4 一番上のほうの

```
seqfile = brown.nuc
treefile = brown.trees

outfile = mlb      * main result file
```

- 5
6 という部分は BASEML プログラムが解析するアラインメントファイルと樹形ファイルの
7 情報です。ここでは
8
9 **seqfile = hummingbird_12mtCDS.fas**
10 **treefile = ML.nwk**
11 **outfile = noclock.out**
12 としておきます。outfile はアウトファイルの名前を指定する項目で、好きな名前をつけて
13 ください（ここでは **noclock.out** という名前にします）
14
15 以下は塩基置換モデルの選択に関するオプションです。

```
model = 4 * 0:JC69, 1:K80, 2:F81, 3:F84, 4:HKY85
          * 5:T92, 6:TN93, 7:REV, 8:UNREST, 9:REVu; 10:UNRESTu
```

1
2 デフォルトでは model=4 すなわち HKY85 モデルが選択されていますが、ここではよく用
3 いられる GTR モデルを用いたいと思います。PAML では GTR モデルは REV モデルと呼
4 ばれています。7 番目のモデルを選択しましょう。

5
6 **model = 7**

7
8 以下は分子時計モデルの選択に関するオプションです。

```
clock = 0 * 0:no clock, 1:clock; 2:local clock; 3:CombinedAnalysis
```

9
10 今回はここがキーになります。0 は分子時計を仮定しないモデル、1 は系統樹全体で分子
11 時計を仮定するモデルです。ここでは最初は分子時計を仮定せずに尤度計算をしたいので
12 デフォルト (clock=0) のままで行きましょう。

13
14 **clock = 0**

15
16 以下のオプションはサイト間の進化速度の不均一性を考慮して Γ モデルを適用するか否か
17 を選択するものです。 Γ モデルはサイトの進化速度と頻度は Γ 分布の関係にあることを仮
18 定しています。

```
fix_alpha = 0 * 0: estimate alpha; 1: fix alpha at value below
              alpha = 0.5 * initial or fixed alpha, 0:infinity (constant rate)
```

19
20
21 **fix_alpha = 0**

22
23 を選択すると Γ モデルが適用されます。

24
25 **alpha = 0.5**

26
27 は Γ 分布の形をきめる形状パラメータ (α) の初期値になります。今回のケースではデフォ
28 ルトオプションのままにしておきましょう。

29
30 もし (あまりそのようなケースはないと思いますが) Γ モデルを適用したくない場合、すな
31 わちすべてのサイトが同じ進化速度で置換することを仮定したい場合は

32
33 **fix_alpha = 1**

1 **alpha = 0.5**

2
3 とします。

4
5 今回のケースではほかのオプションはデフォルトのままです。これで分子時計を
6 仮定しないモデルによる尤度推定の準備完了です。

7 8 2. 解析の実行

9 実行ファイル **baseml.exe** をダブルクリックするか、コマンドプロンプトで **baseml** とタイ
10 プしてエンターキーを押すと、**BASEML** プログラムによる解析が始まります。

11 12 3. 解析結果の確認

13 計算が終了するとアウトファイル（ここでは **noclock.out**）に詳細な解析結果が記録され
14 ます。アウトファイルの下部には樹形やパラメータの数、枝の長さなどの情報が記録され
15 ています。

```
TREE # 1: (1, ((((((2, 5), (3, 4)), 16), (7, 8)), (15, (17, 18))), 10), ((12, 13), (14, 19))), ((6, 9), 11)); MP score: 14395.00
lnL(ntime: 35 np: 41): -74027.321471 +0.000000
 20..1 20..21 21..22 22..23 23..24 24..25 25..26 26..27 27..2 27..5 26..28 28..3 28..4 25..16 24..29 29
0.247636 0.143310 0.010976 0.028933 0.017689 0.064679 0.040437 0.010265 0.005910 0.007039 0.007652 0.014790 0.015399 0.058011 0.056837 0.0
0.173080 0.024132 0.264494

tree length = 2.18605

(Aegotheles_cristatus_EU344979, (((((((Amazilia_brevirostris_KP722043, Amazilia_versicolor_KF624601), (Amazilia_millerii_KP722042, Amazilia
Oreotrochilus_melanogaster_KJ619587))), Chrysolampis_mosquitus_KJ619585), (Florissuga_fusca_KP853096, Florissuga_mellivora_KJ619588), (Glaucis
(Aegotheles_cristatus_EU344979: 0.247636, (((((((Amazilia_brevirostris_KP722043: 0.005910, Amazilia_versicolor_KF624601: 0.007039): 0.010265
Calliphlox_amethystina_KP853095: 0.058049): 0.056837): 0.017689, (Heliodoxa_aurescens_KP853094: 0.119146, (Lophornis_magnificus_KT265276: 0
Florissuga_mellivora_KJ619588: 0.052170): 0.079455, (Glaucis_hirsutus_KT265275: 0.065383, Phaethornis_malaris_KP853097: 0.078268): 0.084175)

Detailed output identifying parameters

Parameters in the rate matrix (REV) (Yang 1994 J Mol Evol 39:105-111):

Rate parameters: 1.30001 0.13930 0.05813 0.17308 0.02413
Base frequencies: 0.24418 0.33824 0.29211 0.12548
Rate matrix Q, Average Ts/Tv = 4.8262
-1.402544 1.264549 0.117018 0.020977
0.912886 -1.066990 0.145396 0.008708
0.097817 0.168359 -0.627024 0.360848
0.040822 0.023473 0.840054 -0.904349

alpha (gamma, K=5) = 0.26449
rate: 0.00123 0.03282 0.21130 0.83533 3.91932
freq: 0.20000 0.20000 0.20000 0.20000 0.20000
```

16
17 ここで

```
18 lnL(ntime: 35 np: 41): -74027.321471 +0.000000
```

19 と書かれている箇所に着目しましょう。

20
21 np:41

22
23 という数値は、ここで使われたモデルのパラメータ数が 41 であることを意味していま
24 す。

1 知っておくと便利なお役立ち情報

2 ここでは分子時計を仮定していないためすべての枝の長さが自由パラメータとなります。枝の
3 数は $2N-3$ (※ N はシーケンス数) で計算でき、ここでは 35 になります。またここでは GTR+ Γ
4 モデルを用いていますが、そのパラメータ数が 6 なので合計 41 です (※本来、GTR モデルは
5 塩基組成を考慮するため、それらもパラメータ数に含めるべきあり、そうするとパラメータ数
6 は 9 になります。しかし PAML はアラインメントから直接カウントして塩基組成を計算する
7 「観察値」を用いる場合はパラメータ数には含みません)。

8
9 このモデルで推定された対数尤度が -74027.321471 という値になっています。

11 3項：分子時計を仮定したモデルによる尤度推定

12 続いて分子時計を仮定したモデルにより尤度とパラメータを推定してみましょう。

14 1. コントロールファイルの編集

15 コントロールファイルの上部のアラインメント名等をしてする個所を編集し

17 **seqfile = hummingbird_12mtCDS.fas**

18 **treefile = MLrt.nwk**

19 **outfile = clock.out**

21 とします。アラインメントファイルは前項 (分子時計を仮定しないモデルによる尤度推
22 定) で用いたものをそのまま使いますが、樹形ファイルは後述するように若干修正したも
23 のを使います (ここでは **MLrt.nwk** としています)。アウトファイル名は好きな名前を付け
24 ましょう (ここでは **clock.out** としています)。

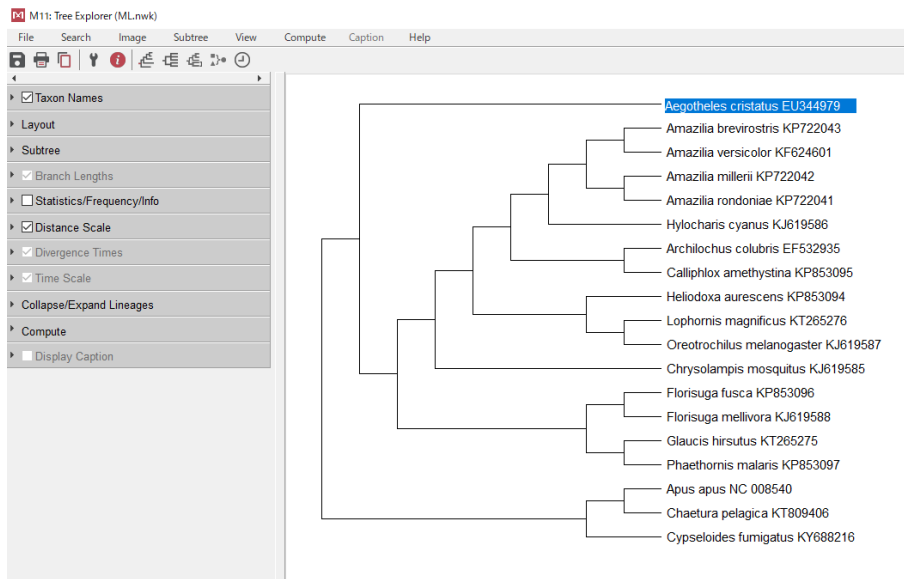
26 そして

28 **clock = 1**

30 とすることで、分子時計モデルが適用されます。ほかの設定は前項と同じで大丈夫です。

32 2. 樹形ファイルの編集

33 前項で用い **ML.nwk** に若干の修正を加えたものを使います。 **MLrt.nwk** を MEGA で開くと



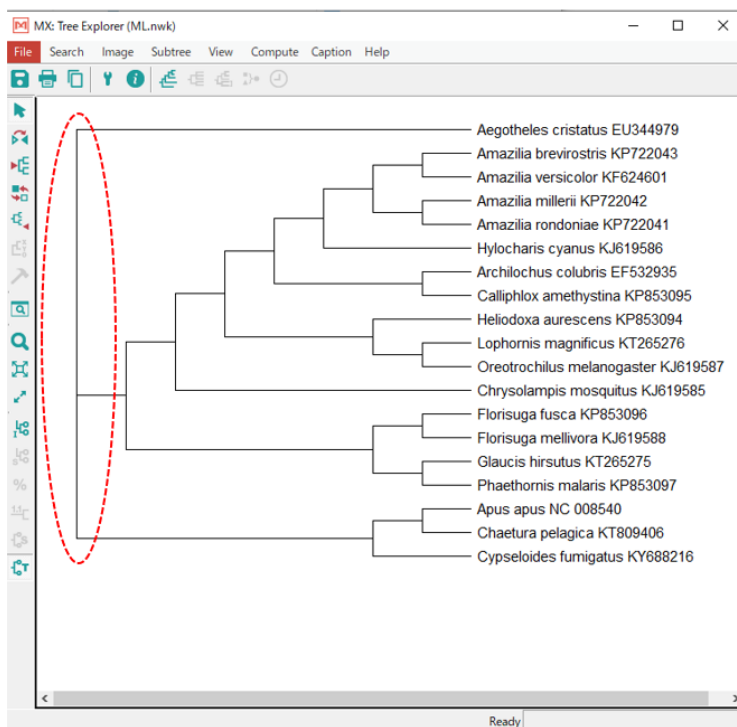
1

2 このように描画されます。

3

4 なお、ここではMEGA11を用いていますが、MEGA-X以前のバージョンでは下記に赤い

5 点線で囲ってあるように無根系統樹は根が三分岐になります。



6

7

1 分子時計を仮定したモデルでは、時間の方向性の概念があるために有根系統樹が必要になり
2 ます。そこで **ML.nwk** をメモ帳などのテキストエディターで開き、下記の赤いフォント
3 で()をつけている箇所に、()を加えましょう。

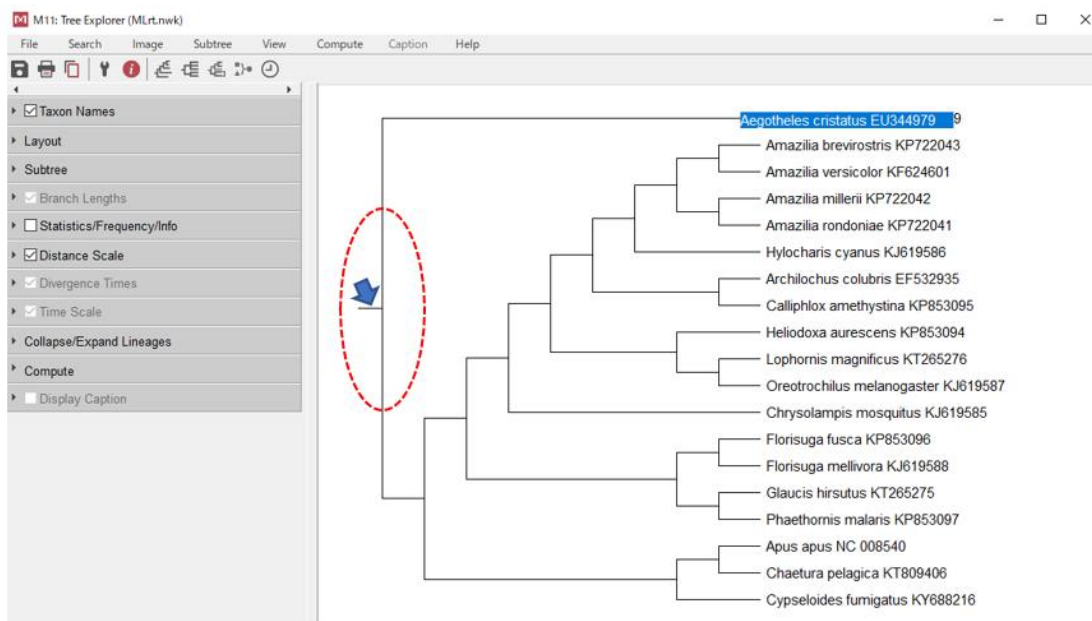
4

```
5 (Aegotheles_cristatus_EU344979,(((((((Amazilia_brevirostris_KP722043,Amazilia_versico  
6 lor_KF624601),(Amazilia_millerii_KP722042,Amazilia_rondoniae_KP722041)),Hylocharis  
7 _cyanus_KJ619586),(Archilochus_colubris_EF532935,Calliphlox_amethystina_KP853095)),  
8 (Heliodoxa_aurescens_KP853094,(Lophornis_magnificus_KT265276,Oreotrochilus_melan  
9 ogaster_KJ619587))),Chrysolampis_mosquitus_KJ619585),((Florisuga_fusca_KP853096,Fl  
10 orisuga_mellivora_KJ619588),(Glaucis_hirsutus_KT265275,Phaethornis_malaris_KP85309  
11 7))),((Apus_apus_NC_008540,Chaetura_pelagica_KT809406),Cypseloides_fumigatus_KY6  
12 88216)));
```

13

14 ここでは()を挿入することでハチドリ亜目とアマツバメ亜目が単系統になり、ズクヨタカ
15 科 (Aegotheles_cristatus_EU344979) が外群として位置づけられます。この段階でファイ
16 ルに好きな名前 (ここでは **MLrt.nwk** とします) を付けて再度 MEGA で開いてみると

17



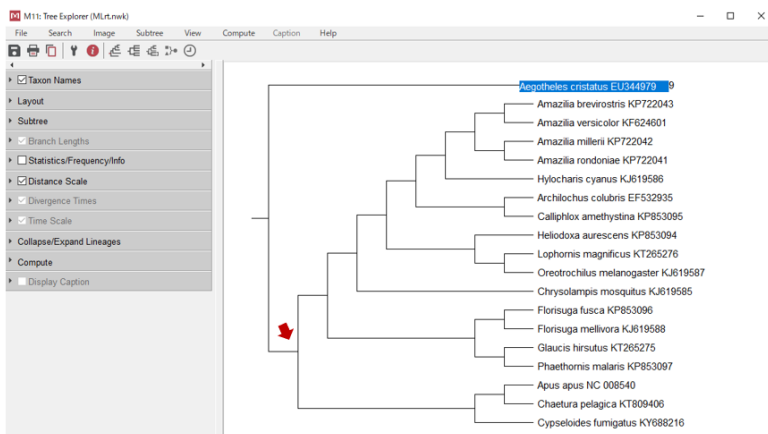
18

19

20 矢印で示しているように系統樹に「根」がついていること分かります。有根系統樹の完成
21 です。

22

- 1 次に化石記録を用いて分岐年代推定値の制約条件を加えましょう。
- 2
- 3 ここではハチドリ亜目とアマツバメ亜目の分岐年代に着目しましょう。これは下の系統樹
- 4 で赤い矢印がついているノードに相当します。



- 5
- 6
- 7 ハチドリ亜目の最古の化石記録の年代はリュプス期 (27.82~33.9Ma :Ma は Mega-annum
- 8 の略で、100 万年前を意味します) という地質年代、アマツバメ亜目の最古の化石記録はイ
- 9 ーベル期 (47.9~56Ma) という地質年代の地層から報告されていますので、ここではとり
- 10 あえずハチドリ亜目とアマツバメ亜目の分岐年代を 56Ma と仮定します。

- 11
- 12 **MLrt.nwk** をテキストエディターで開きハチドリ亜目とアマツバメ亜目の共通祖先に相当
- 13 する()のすぐ後ろに**@0.56** と記入しましょう。

- 14
- 15 (*Aegotheles cristatus*_EU344979,(((((((*Amazilia brevirostris*_KP722043,*Amazilia versico*
- 16 *lor*_KF624601),(*Amazilia millerii*_KP722042,*Amazilia rondoniae*_KP722041)),*Hylocharis*
- 17 *_cyanus*_KJ619586),(*Archilochus colubris*_EF532935,*Calliphlox amethystina*_KP853095)),
- 18 (*Heliodoxa aurescens*_KP853094,(*Lophornis magnificus*_KT265276,*Oreotrochilus melan*
- 19 *ogaster*_KJ619587))),*Chrysolampis mosquitus*_KJ619585),((*Florisuga fusca*_KP853096,*Fl*
- 20 *orisuga mellivora*_KJ619588),(*Glaucis hirsutus*_KT265275,*Phaethornis malaris*_KP85309
- 21 *7*))),((*Apus apus*_NC_008540,*Chaetura pelagica*_KT809406),*Cypseloides fumigatus*_KY6
- 22 *88216*)) **@0.56**);

- 23
- 24 PAML では分岐年代を推定する際に、大きな数字 (>10) を使うと計算がうまくいかない
- 25 ことがあるので@56 ではなく@0.56 にしておく方が良いです。つまりここでは 100Ma(1 億
- 26 年)が 1 単位時間になります。

27

1 3. 解析の実行

2 実行ファイル **baseml.exe** をダブルクリックするか、コマンドプロンプトで **baseml** とタイ
3 プしてエンターキーを押すと、BASEML プログラムによる解析が始まります。

4

5 4. 解析結果の確認

6 計算が終了するとアウトファイル（ここでは **clock.out**）に詳細な解析結果が記録されま
7 す。アウトファイルの下部には樹形やパラメータの数、枝の長さなどの情報が記録されて
8 います。

```
TREE # 1: (1, (((((((2, 5), (3, 4)), 16), (7, 8)), (15, (17, 18))), 10), ((12, 13), (14, 19))), ((6, 9), 11), 12), 13), 14), 15), 16), 17), 18), 19), 20), 21), 22), 23), 24), 25), 26), 27), 28), 29), 30), 31), 32), 33), 34), 35), 36), 37);
lnL(ntime: 18 np: 24): -74134.049288 +0.000000
0.560273 0.352414 0.334622 0.281599 0.248547 0.125136 0.043099 0.014846 0.030679 0.118910 0.265737 0.234445

Note: mutation rate is not applied to tree length. Tree has ages, for TreeView & FigTree

(Aegotheles_cristatus_EU344979, (((((((Amazilia_brevirostris_KP722043, Amazilia_versicolor_KF624601), (Amazilia_versicolor_KF624601, Amazilia_brevirostris_KP722043), 16), (7, 8)), (15, (17, 18))), 10), ((12, 13), (14, 19))), ((6, 9), 11), 12), 13), 14), 15), 16), 17), 18), 19), 20), 21), 22), 23), 24), 25), 26), 27), 28), 29), 30), 31), 32), 33), 34), 35), 36), 37);
(Aegotheles_cristatus_EU344979: 0.000000, (((((((Amazilia_brevirostris_KP722043: 0.000000, Amazilia_versicolor_KF624601: 0.000000), (Amazilia_versicolor_KF624601: 0.000000, Amazilia_brevirostris_KP722043: 0.000000), 16), (7, 8)), (15, (17, 18))), 10), ((12, 13), (14, 19))), ((6, 9), 11), 12), 13), 14), 15), 16), 17), 18), 19), 20), 21), 22), 23), 24), 25), 26), 27), 28), 29), 30), 31), 32), 33), 34), 35), 36), 37);
(1_Aegotheles_cristatus_EU344979, (((((((2_Amazilia_brevirostris_KP722043, 5_Amazilia_versicolor_KF624601), (2_Amazilia_brevirostris_KP722043, 5_Amazilia_versicolor_KF624601), 16), (7, 8)), (15, (17, 18))), 10), ((12, 13), (14, 19))), ((6, 9), 11), 12), 13), 14), 15), 16), 17), 18), 19), 20), 21), 22), 23), 24), 25), 26), 27), 28), 29), 30), 31), 32), 33), 34), 35), 36), 37);
(17_Lophornis_magnificus_KT265276, 18_Oreotrochilus_melanogaster_KJ619587) 32 ) 31 ) 24 , 10_Chrysolampis_mosquitus_KJ619585, 11_Cypseloides_fumigatus_KY688216) 36 ) 21 ) 20 ;

Detailed output identifying parameters

Substitution rate is per time unit
0.459203

Nodes and Times
(JeffNode is for Thorne's multidivtime. ML analysis uses ingroup data only.)

Node 20 (Jeffnode 36) Time 0.56027
Node 21 (Jeffnode 35) Time 0.56000
Node 22 (Jeffnode 34) Time 0.35241
Node 23 (Jeffnode 33) Time 0.33462
Node 24 (Jeffnode 32) Time 0.28160
Node 25 (Jeffnode 31) Time 0.24855
Node 26 (Jeffnode 30) Time 0.12514
Node 27 (Jeffnode 29) Time 0.04310
Node 28 (Jeffnode 28) Time 0.01485
Node 29 (Jeffnode 27) Time 0.03068
Node 30 (Jeffnode 26) Time 0.11891
Node 31 (Jeffnode 25) Time 0.26574
Node 32 (Jeffnode 24) Time 0.23444
Node 33 (Jeffnode 23) Time 0.31834
Node 34 (Jeffnode 22) Time 0.12316
Node 35 (Jeffnode 21) Time 0.15262
Node 36 (Jeffnode 20) Time 0.30088
Node 37 (Jeffnode 19) Time 0.21187

Parameters in the rate matrix (REV) (Yang 1994 J Mol Evol 39:105-111):

Rate parameters: 1.30987 0.13753 0.05727 0.17292 0.02519
Base frequencies: 0.24418 0.33824 0.29211 0.12548
Rate matrix Q, Average Ts/Tv = 4.8737
-1.404760 1.269099 0.115078 0.020583
0.916171 -1.069908 0.144685 0.009053
0.096196 0.167535 -0.623150 0.359419
0.040054 0.024405 0.836727 -0.901186

alpha (gamma, K=5) = 0.26226
rate: 0.00117 0.03190 0.20817 0.83056 3.92820
freq: 0.20000 0.20000 0.20000 0.20000 0.20000
```

9

1

2 いろいろな情報が書かれていますが、分岐年代の推定値は以下のようにノードごとにまと
3 められています。

4

```
Nodes and Times
(JeffNode is for Thorne's multidivtime. ML analysis uses ingroup data only.)

Node 20 (Jeffnode 36) Time 0.56027
Node 21 (Jeffnode 35) Time 0.56000
Node 22 (Jeffnode 34) Time 0.35241
Node 23 (Jeffnode 33) Time 0.33462
Node 24 (Jeffnode 32) Time 0.28160
Node 25 (Jeffnode 31) Time 0.24855
Node 26 (Jeffnode 30) Time 0.12514
Node 27 (Jeffnode 29) Time 0.04310
Node 28 (Jeffnode 28) Time 0.01485
Node 29 (Jeffnode 27) Time 0.03068
Node 30 (Jeffnode 26) Time 0.11891
Node 31 (Jeffnode 25) Time 0.26574
Node 32 (Jeffnode 24) Time 0.23444
Node 33 (Jeffnode 23) Time 0.31834
Node 34 (Jeffnode 22) Time 0.12316
Node 35 (Jeffnode 21) Time 0.15262
Node 36 (Jeffnode 20) Time 0.30088
Node 37 (Jeffnode 19) Time 0.21187
```

5

6

7 ここではノード 20~37 までの推定値が 100Ma を単位時間としてまとめられています。ど
8 のノードがどの番号に相当するかは、そのすぐ上の

9

```
(1_Aegotheles_cristatus_EU344979, (((((((2_Amazilia_brevirostris_KP722043, 5_Amazilia_versicolor_KF624601) 28, (3_Amazilia_millerii_KP722042,
4_Amazilia_rondoniae_KP722041) 29 ) 27, 16_Hylocharis_cyanus_KJ619586) 26, (7_Archilochus_colubris_EF532935, 8_Calliphlox_amethystina_KP853095) 30 ) 25,
(15_Helliodoxa_aurescens_KP853094, (17_Lophornis_magnificus_KT265276, 18_Oreotrochilus_melanogaster_KJ619587) 32 ) 31 ) 24,
10_Chrysolampis_mosquitus_KJ619585) 23, ((12_Florisuga_fusca_KP853096, 13_Florisuga_mellivora_KJ619588) 34, (14_Glaucis_hirsutus_KT265275,
19_Phaethornis_malaris_KP853097) 35 ) 33 ) 22, ((6_Apus_apus_NC_008540, 9_Chaetura_pelagica_KT809406) 37, 11_Cypseloides_fumigatus_KY688216) 36 ) 21 ) 20 ;
```

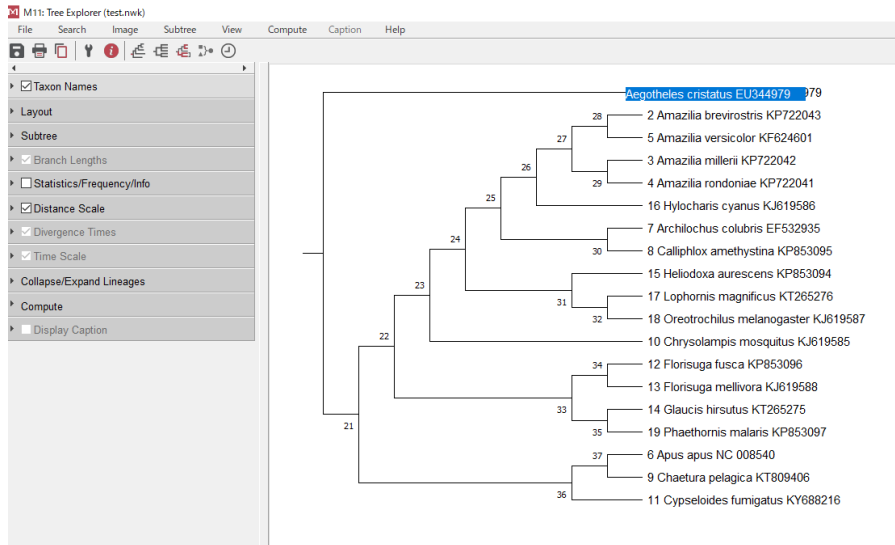
10

11

12 に newick 形式でまとめられています。

13 これでは見にくい、という人はこの個所をテキストエディターにコピー&ペーストして
14 nwk という拡張子をつけたうえで MEGA で描画すれば以下のように各ノード上にノード
15 番号が可視化されるのでわかりやすいと思います。

16



1
2
3
4
5
6
7
8
9

なおコントロールファイルで

```
getSE = 1 * 0: don't want them, 1: want S.E.s of estimates
```

getSE = 1

を選択すれば、年代推定値の標準誤差を推定できます。

```
Nodes and Times
(JeffNode is for Thorne's multidivtime. ML analysis uses ingroup data only.)

Node 20 (Jeffnode 36) Time 0.56027 +- 0.01342
Node 21 (Jeffnode 35) Time 0.56000
Node 22 (Jeffnode 34) Time 0.35241 +- 0.00712
Node 23 (Jeffnode 33) Time 0.33462 +- 0.00733
Node 24 (Jeffnode 32) Time 0.28160 +- 0.00669
Node 25 (Jeffnode 31) Time 0.24855 +- 0.00694
Node 26 (Jeffnode 30) Time 0.12514 +- 0.00478
Node 27 (Jeffnode 29) Time 0.04310 +- 0.00209
Node 28 (Jeffnode 28) Time 0.01485 +- 0.00129
Node 29 (Jeffnode 27) Time 0.03068 +- 0.00181
Node 30 (Jeffnode 26) Time 0.11891 +- 0.00502
Node 31 (Jeffnode 25) Time 0.26574 +- 0.00711
Node 32 (Jeffnode 24) Time 0.23445 +- 0.00757
Node 33 (Jeffnode 23) Time 0.31834 +- 0.00812
Node 34 (Jeffnode 22) Time 0.12316 +- 0.00532
Node 35 (Jeffnode 21) Time 0.15262 +- 0.00597
Node 36 (Jeffnode 20) Time 0.30088 +- 0.00942
Node 37 (Jeffnode 19) Time 0.21187 +- 0.00797
```

10
11
12
13
14

+のあとの数値が標準誤差です(ただしこの機能は PAML のバージョンによっては壊れおり、おかしい数値を返してくることがあります)。

1 また単位時間当たりの分子進化速度は

```
2 Substitution rate is per time unit  
0.459203
```

3 と推定されています。これは1塩基サイトあたり100億年間で0.459203回の置換が起き
4 ること(0.459203/塩基サイト/100億年)を示しています。

5

6 このようにして分子時計を仮定して分岐年代や分子進化速度を推定することができま
7 す。。。が、今回のケースではそもそも分子時計は成立しているのでしょうか？

8

9 パラメータ数と尤度を見てみると

```
10 lnL(ntime: 18 np: 24): -74134.049288 +0.000000
```

11 となっています。前項で分子時計を仮定せずに尤度を計算した際はパラメータ数が41で
12 したが、今回は24になっています。

13 分子時計を仮定する場合は、特定のノードに着目した場合そこから派生する系統の末端節
14 までの長さはすべて等しくなるため、枝の数ではなく、内部節の数が自由パラメータの数
15 になります。内部節の数は $N-1$ (※ N はシーケンス数)なので今回の場合は18になります。
16 これに塩基置換モデルのパラメータ数6を加えて合計24になります。

17

18 これらの情報をもとに尤度比検定を行うと

19 p値は $5.55785E-36$ となり、今回のケースでは分子時計は棄却されてしまいました。すな
20 わち分子時計を仮定して推定した分岐年代は、信頼性は低いことを意味しています。

21

22 種内やごく近縁種間のデータを扱う場合、分子時計が棄却できない例が多々あり、そのよ
23 うな場合では分子時計を仮定して分岐年代を推定することができます。しかし今回のハチ
24 ドリの例の様に科間などの系統的に離れたグループを扱う際は、分子時計が成立していな
25 いことが一般的です。このような場合は、系統間で進化速度が変動することを許す緩和型
26 分子時計が用いられます。

27

28 3節：緩和型分子時計による分岐年代推定

29

30 PAMLのMCMCTREEプログラムを用いて階層ベイズによる緩和型分子時計を用いて分
31 岐年代を推定します。ここでは特に正規近似法による年代推定について概説します。正規
32 近似法を用いる場合、以下の二段階の解析を行います。第一段階で最尤法により枝の長さ
33 とその分散共分散を推定します。第二段階では階層ベイズ法により分岐年代を推定しま
34 す。

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34

1項. 準備

1. 必要なファイル

MCMCTREE プログラムも他の PAML のプログラムと同様①実行ファイル、②コントロールファイル、③アラインメントファイル、④樹形ファイルを必要とします。

- ① 実行ファイルは **mcmctree.exe** と **baseml.exe** の2つの実行ファイルが必要です。
- ② コントロールファイルは **mcmctree.ctl** を用います。
- ③ アラインメントファイルは前節でも用いた **hummingbird_12mtCDS.fas** を用います。
- ④ 樹形ファイルは前節で作成した **MLrt.nwk** を若干編集して使います。

これらの5つのファイルを同じフォルダ（ここでは **Relaxedclock** という名前にしたいと思います）に入れて使います。

これらのファイルのほかに、系統樹の描画に Figtree プログラムが、ベイズの MCMC によるパラメータの収束の確認に Tracer プログラムがあると便利なので以下のサイトからダウンロードしておきましょう。

Figtree

<http://tree.bio.ed.ac.uk/software/figtree/>

(2022年2月7日現在最新版は version1.4.4 です)

Tracer

<https://github.com/beast-dev/tracer/releases>

(2022年2月7日現在最新版は version1.7.2 です)

Figtree プログラムと Tracer プログラムは JAVA 言語を用いているので、解析に用いるコンピューターに JAVA 言語がインストールされていない場合は、インストールしておきましょう。

<https://www.java.com/ja/>

2. コントロールファイルの編集

コントロールファイル **mcmctree.ctl** をテキストエディターで開くと以下のようになっています。


```
seed = -1
seqfile = examples/DatingSoftBound/mtCDNApri123.txt
treefile = examples/DatingSoftBound/mtCDNApri.trees
outfile = out

ndata = 3
seqtype = 0 * 0: nucleotides; 1:codons; 2:AAs
usedata = 1 * 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 = 0 * 0:JC69, 1:K80, 2:F81, 3:F84, 4:HKY85
alpha = 0 * 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

rgene_gamma = 2 2 * gamma prior for overall rates for genes
sigma2_gamma = 1 10 * gamma prior for sigma^2 (for clock=2 or 3)

finetune = 1: 0.1 0.1 0.1 0.01 .5 * auto (0 or 1) : times, musigma2, rates, mixing, paras, FossilErr

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

*** Note: Make your window wider (100 columns) before running the program.
```

1
2
3
4
5
6
7
8
9
10
11
12
13

シーケンスファイル名などは以下の個所を編集します。

```
seqfile = examples/DatingSoftBound/mtCDNApri123.txt
treefile = examples/DatingSoftBound/mtCDNApri.trees
outfile = out
```

ここでは

```
seqfile = hummingbird_12mtCDS.fas
treefile = MLrt.nwk
outfile = hummingbird.out
```

としておきたいと思います。

```
ndata = 3
seqtype = 0 * 0: nucleotides; 1:codons; 2:AAs
usedata = 1 * 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.
```

ndata は解析に用いるアラインメントの数を指定します。MCMCTREE プログラムは複数の遺伝子座位から分岐年代推定を行うことができます。今回のケースでは1座位（1アラインメント）のみのデータなので

18

1 **ndata=1**
2
3 にしましょう。
4
5 **seqtype** のオプションにより塩基配列(0)を扱うか、コドン配列を扱うか(1)、アミノ酸配列
6 (2)を扱うかを指定することができます。今回のケースでは塩基配列を用いるので

7
8 **seqtype=0**

9
10 を選びます。

11
12 **usedata** のオプションで、データを用いないで年代推定 (0)、アラインメントから直接尤
13 度関数を推定して年代推定(1)、枝の長さとその分散・共分散から近似的に尤度関数を推定
14 して年代推定(2)、枝の長さとその分散・共分散を推定(3)が選択できます。

15
16 データを用いないで年代推定 (0) は尤度関数を用いないので事前確率分布が推定される
17 こととなります。アラインメントから直接尤度関数を推定して年代推定(1)は正確ですが計
18 算が非常に遅く、アラインメントに含まれる配列数が 10 以上になると現実的ではないよ
19 うです。枝の長さとその分散・共分散から近似的に尤度関数を推定して年代推定(2)という
20 オプション枝の長さとその分散・共分散を推定(3)というオプションは正規近似法による分
21 岐推定に用います。ここはまた後程解説します。

22
23 **clock** のオプションでは分子進化速度の変動の様式を選択することができます。0 を選択す
24 ると系統樹全体を通して分子進化速度が一定 (分子時計) になり、1 を選択すると独立速
25 度モデルが、2 を選択すると自己相関モデルが選ばれます。独立速度モデルと自己相関モ
26 デルが緩和型分子時計です。これらのモデルの説明は本文をご参照ください。ここでは
27 (前節で分子時計モデルは棄却されているので) 緩和型分子時計を使います。とりあえず
28 自己相関モデルを選択してみましょう。

29
30 **clock=3**

31
32 **Rootage** のオプションは、系統樹の根の年代の最大値を指定するものです。この値を客観
33 的に決定することは難しいですがとりあえず 1 億年と仮定しておきましょう。

34
35 **RootAge = <1.0**

36

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

これらは塩基置換モデルの選択に関するオプションです。アラインメントから直接尤度関数を推定して年代推定(usedata=1)する場合は、**HKY+ Γ** が最も複雑なモデルになりますが、正規近似法を用いる場合は BASEML プログラムや CODEML プログラムで使える置換モデルは基本的にすべて使うことができます。ここでは **GTR+ Γ** モデルを用いたので

```
model = 7
alpha = 0.2
ncatG = 5
```

としておきましょう。alpha は Γ 分布の形状パラメータ(α)の初期値、ncatG は BASEML が Γ 分布を離散分布により近似的に推定しているため、そのカテゴリの数を指定するパラメータです。

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

BDparas は出生死滅過程に関するパラメータ、すなわち出生率(λ)、死滅率(ε)、標本抽出率(ρ)です(本文及びウェブ資料参照)。デフォルトの設定(1 1 0)では内部ノードの分布が均一分布になります。これは基本的にデフォルトの設定のままで大丈夫です。

kappa_gamma はトランジション/トランスバージョンの比率(κ)の、alpha_gamma はサイト間の進化速度の不均一性に関する Γ 分布の形状パラメータ(α)の事前確率分布です。正規近似法を用いる場合はこれらの二つのパラメータは使われないので変更する必要ありません。

```
rgene_gamma = 2 2 * gamma prior for overall rates for genes
sigma2_gamma = 1 10 * gamma prior for sigma^2 (for clock=2 or 3)
```

これらは進化速度の変動に関する事前分布を与えるパラメータです。自分自身のデータにあわせたパラメータを用いることが推奨されています。

rgene_gamma は枝の進化速度 r 、sigma2_gamma は幾何ブラウン運動の分散パラメータ σ^2 に関するものです(本文参照)。MCMCTREE はこれらのパラメータの事前確率分布を Γ 分布として与えます。 Γ 分布は形状パラメータ(α)と尺度パラメータ(β)により規定されます。いま進化速度 r の平均が m 、標準偏差を s とすると

1 $\alpha = (m/s)^2$

2 $\beta = m/s^2$

3

4 となるので進化速度 r の平均と標準偏差の大雑把な情報を知っておく必要があります。今
5 回のケースでは前節の3項（分子時計を仮定したモデルによる尤度推定）で推定した分子
6 進化速度を使うことにします。ここで推定された値は 0.459203 /塩基サイト/100 億年でし
7 た。これを平均値 m として用います。標準偏差 s は恣意的ではありますが $m/2$ と仮定しま
8 す。これで

9

10 $\alpha = (m/s)^2 = 4$

11 $\beta = m/s^2 \approx 8.7$

12

13 となるので

14

15 **rgene_gamma = 4 8.7**

16

17 になります。

18

19 幾何ブラウン運動の分散パラメータ σ^2 の平均値は、客観的に設定することが難しい値です
20 が、ここでは Thorne et al. (1998) により開発されたプログラム MULTIDIVTIME のマニ
21 ュアルで推奨されているように

22

23 σ^2 の平均値 \times rttm = 1

24

25 としたいと思います。rttm は根の年代です。今回のケースでは、樹形ファイルの編集のと
26 ころで紹介しますが 0.68(68Ma) にしておきたいと思います。従って σ^2 の平均値を $1/0.68$
27 ≈ 1.47 としましょう。 σ^2 の標準偏差は σ^2 の平均値と等しいと仮定します。

28

29 $\alpha = (m/s)^2 = 1$

30 $\beta = m/s^2 \approx 0.68$

31

32 となるので

33

34 **sigma2_gamma = 1 0.68**

35

36 になります。

1

2

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

3

4 これらは MCMC (マルコフ連鎖モンテカルロ) の設定条件です。

5 burnin は収束状態に達する前のサンプルを解析から除外するために初期のサンプルを切り
6 捨てる設定で、デフォルトでは最初の 2000 世代のサンプルは解析に用いられません。

7 sampfreq はサンプリングの頻度、nsample はサンプリングの回数です。従って MCMC の
8 長さの総和は

9

10 burnin+sampfreq×nsample 世代

11

12 になります。

13 ここでは

14

burnin = 50000

15

sampfreq = 50

16

nsample = 20000

17

18

19 という条件で MCMC を実行したいと思います。

20

21 3. 樹形ファイルの編集

22 ①MLrt.nwk をテキストエディターで開き、1 行目に解析に用いられる配列数(今回は 19)
23 と樹形の数 (今回は 1) を書きます。

24

25 **19 1**

```
(Aegotheles_cristatus_EU344979,(((((((Amazilia_brevirostris_KP722043,Amazilia_versico  
lor_KF624601),(Amazilia_millerii_KP722042,Amazilia_rondoniae_KP722041)),Hylocharis  
_cyanus_KJ619586),(Archilochus_colubris_EF532935,Calliphlox_amethystina_KP853095)),  
29 (Heliodoxa_aurescens_KP853094,(Lophornis_magnificus_KT265276,Oreotrochilus_melan  
ogaster_KJ619587))),Chrysolampis_mosquitus_KJ619585),((Florisuga_fusca_KP853096,Fl  
orisuga_mellivora_KJ619588),(Glaucis_hirsutus_KT265275,Phaethornis_malaris_KP85309  
7))),((Apus_apus_NC_008540,Chaetura_pelagica_KT809406),Cypseloides_fumigatus_KY6  
33 88216))));
```

34

35 ②年代制約の条件

1 前節で分子時計により最尤法で分岐年代を推定した際は、年代の制約条件を 56Ma という
2 ピンポイントの数値で与えましたが、ベイズ法では点ではなく分布の形で制約条件を与え
3 ることができます。

4
5 前項で書いたようにアマツバメ亜目の最古の化石記録はイーペル期 (47.9~56Ma) です。
6 これに従いアマツバメ亜目とハチドリ亜目の分岐を 47.9Ma 以前にしたいと思います。ま
7 たアマツバメ目全体の祖先にあたる最古の化石は北米の Green River 層の Fossil Butte 部
8 層から報告されています。ズクヨタカ亜目とアマツバメ亜目+ハチドリ亜目の分岐はこの
9 地層の年代幅の最小値 (51.85Ma) と現生鳥類の最古の化石記録の年代(68Ma)の間で起き
10 たと仮定したいと思います。

12 19 1

13 (Aegotheles_cristatus_EU344979,(((((((Amazilia_brevirostris_KP722043,Amazilia_versico
14 lor_KF624601),(Amazilia_millerii_KP722042,Amazilia_rondoniae_KP722041)),Hylocharis
15 _cyanus_KJ619586),(Archilochus_colubris_EF532935,Calliphlox_amethystina_KP853095)),
16 (Heliodoxa_aurescens_KP853094,(Lophornis_magnificus_KT265276,Oreotrochilus_melan
17 ogaster_KJ619587))),Chrysolampis_mosquitus_KJ619585),((Florisuga_fusca_KP853096,Fl
18 orisuga_mellivora_KJ619588),(Glaucis_hirsutus_KT265275,Phaethornis_malaris_KP85309
19 7))),((Apus_apus_NC_008540,Chaetura_pelagica_KT809406),Cypseloides_fumigatus_KY6
20 88216)) >0.479)>0.5158 <0.68;

21
22 これで樹形ファイルを保存したら解析準備は完了です。

25 2項. 最尤法による枝の長さや分散・共分散の推定

26
27 正規近似法では、第一段階として最尤法により枝の長さや分散・共分散を推定します。これ
28 により尤度関数を近似的に推定でき、枝の長さをどう変化させると尤度がどのように変化
29 するかを推定することが出来るようになるので次の段階 (階層ベイズ法による年代推定)
30 で MCMC を非常に高速に実行することができるようになります。本項では第一段階をどの
31 ように行うかを概説します。

33 1. コントロールファイルの編集

34 コントロールファイルの

```
35 usedata = 1 * 0: no data; 1:seq like; 2:use in.BV; 3: out.BV
```

36 に変更を加えます。

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36

usedata = 3

を選択してください。

2. 解析の実行

実行ファイル **mcmctree.exe** をダブルクリックするか、コマンドプロンプトで **mcmctree** とタイプしてエンターキーを押すと、MCMCTREE プログラムによる解析が始まります。

MCMCTREE プログラムは BASEML プログラムを起動し枝の長さや分散共分散を最尤推定します。ここで MCMCTREE プログラムは、自動的に tmp0001.ctl, tmp0001.trees, tmp001.txt というファイルを生じますが、これらは BASEML プログラムが解析できるようにコントロールファイル、樹形ファイル、シーケンスファイルの形を整えたものです。例えば樹形ファイルに関しては年代制約条件付きの有根系統樹を与えましたが、tmp0001.trees は無根系統樹に変換されたうえで、年代制約条件など BASEML プログラムが解析するうえで不要な情報は取り除かれます。解析に用いるシーケンスファイルが N 座位のアラインメントから構成される場合は、これらのコントロールファイル、樹形ファイル、シーケンスファイルは N セット分生成されます。

BASEML プログラムによる解析が正常に終了すれば、フォルダの中に out.BV というファイルが生成されます。この中に枝の長さや勾配ベクトル、分散共分散（ヘッセ行列）の情報が記録されます。

3項. 階層ベイズ法による分岐年代推定

1. out.BV ファイルの編集

out.BV ファイルの名前を変更し in.BV にします。

2. コントロールファイルの編集

コントロールファイルの

```
usedata = 1 * 0: no data; 1:seq like; 2:use in.BV; 3: out.BV
```

に変更を加えます。

usedata = 2

を選択してください。

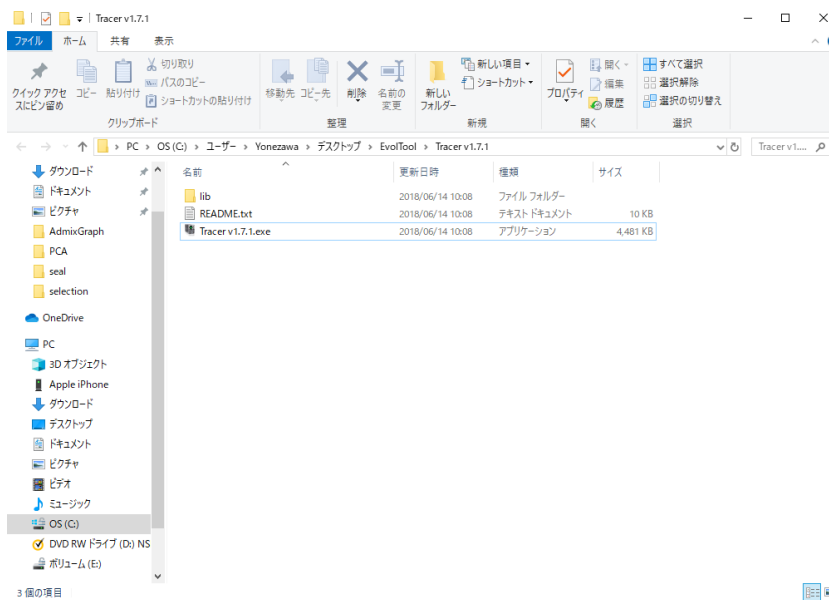
1 **3. 解析の実行**

2 実行ファイル mcmctree.exe をダブルクリックするか、コマンドプロンプトで **mcmctree** と
3 タイプしてエンターキーを押すと、MCMCTREE プログラムによる解析が始まります。

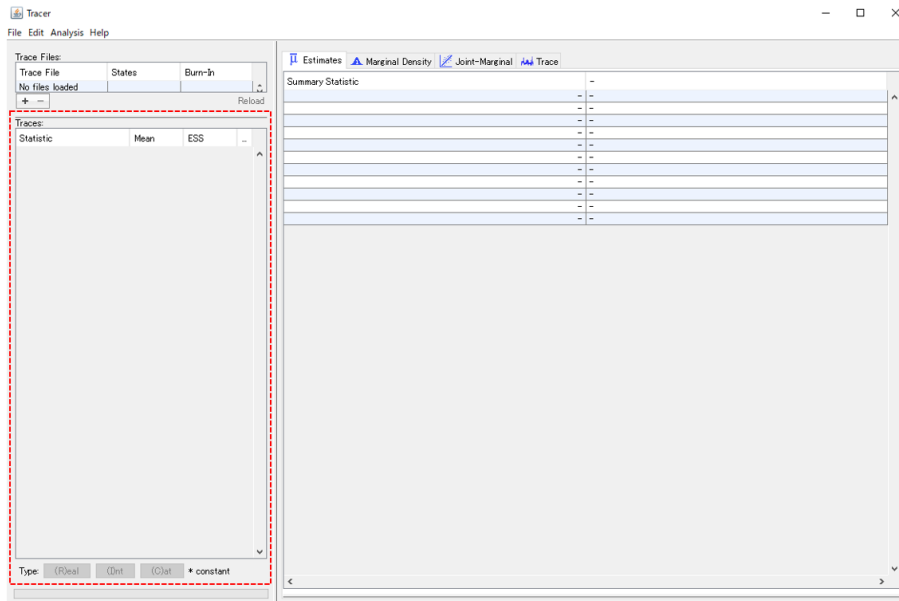
4
5 **4. パラメータの収束の確認**

6 解析が正常に終了すると mcmc.txt というファイルが生成され、その中に各サンプルのパ
7 ラメータが記録されています。TRACER を用いてパラメータが十分に収束しているか確認
8 しましょう。

9
10 TRACER をダウンロードしたら、その実行ファイル（ここでは Tracer v1.7.1.exe）をダブ
11 ルクリックで起動します。



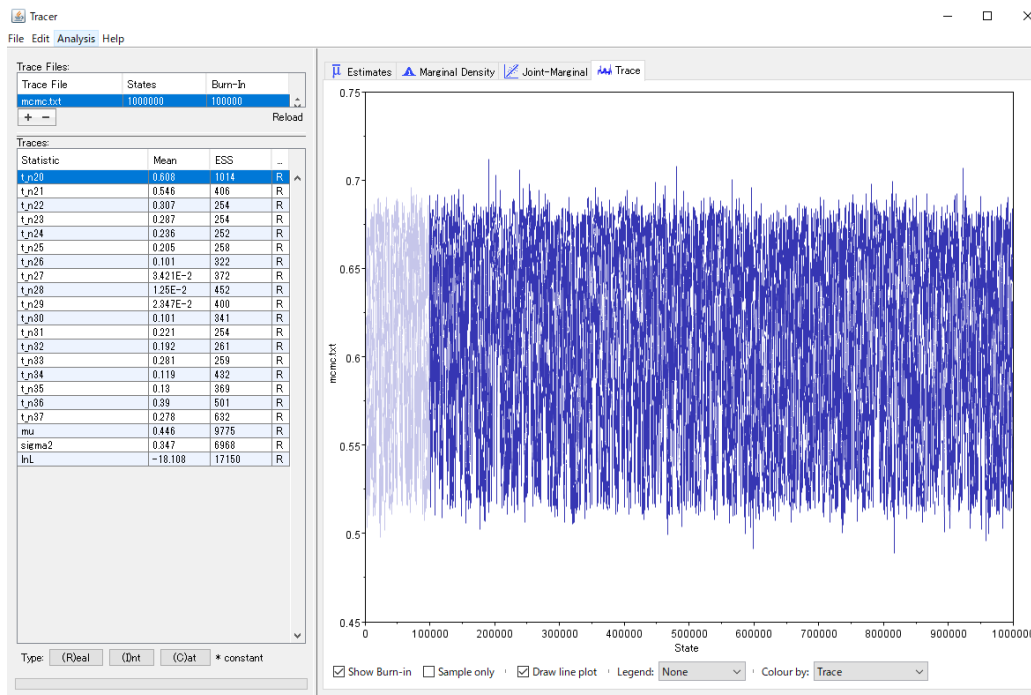
12
13
14 TRACER が起動すると以下のような画面が現れるので、赤い点線で囲っている箇所に
15 mcmc.txt をドラッグ&ドロップすると、TRACER は mcmc.txt を読み込みます。



1

2

3 こちらが TRACER が mcmc.txt を読み込んだ状態です。



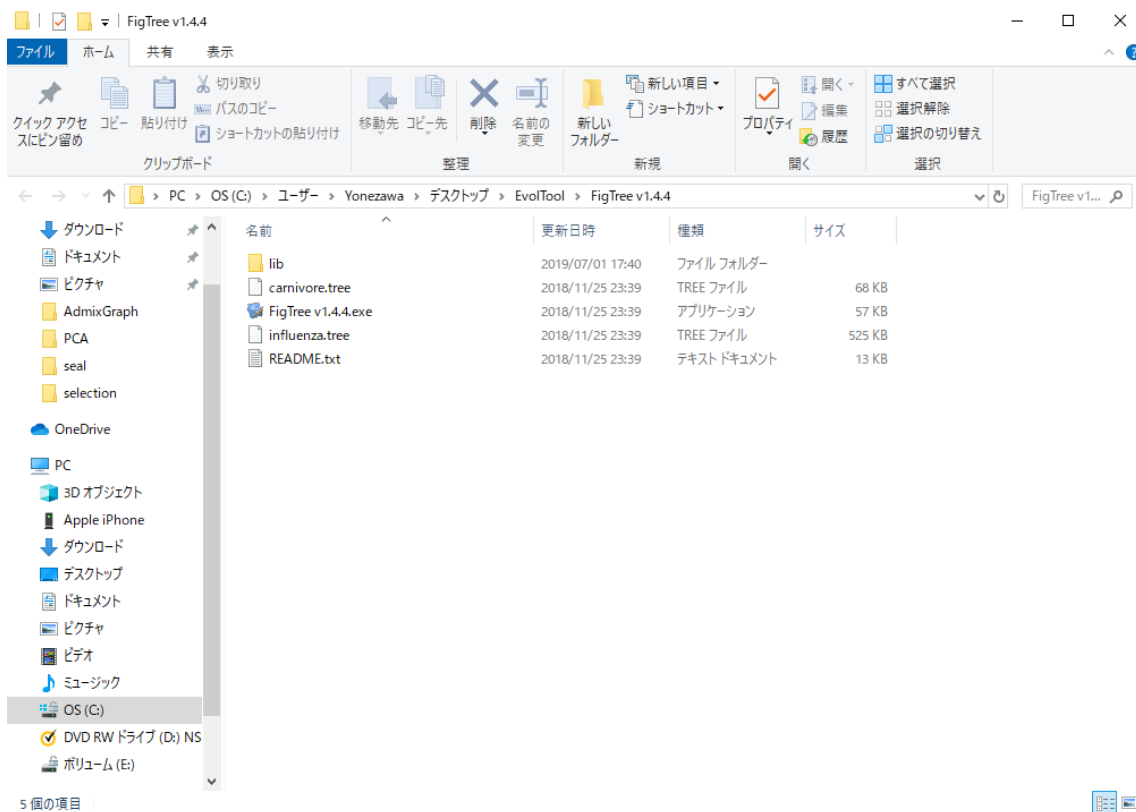
4

5 ESS は effective sample size の略ですが、ESS が 200 以上であればパラメータが十分に収束しているといえます。ESS が 200 未満の場合は MCMC の世代数を長めにとると良いでしょう。

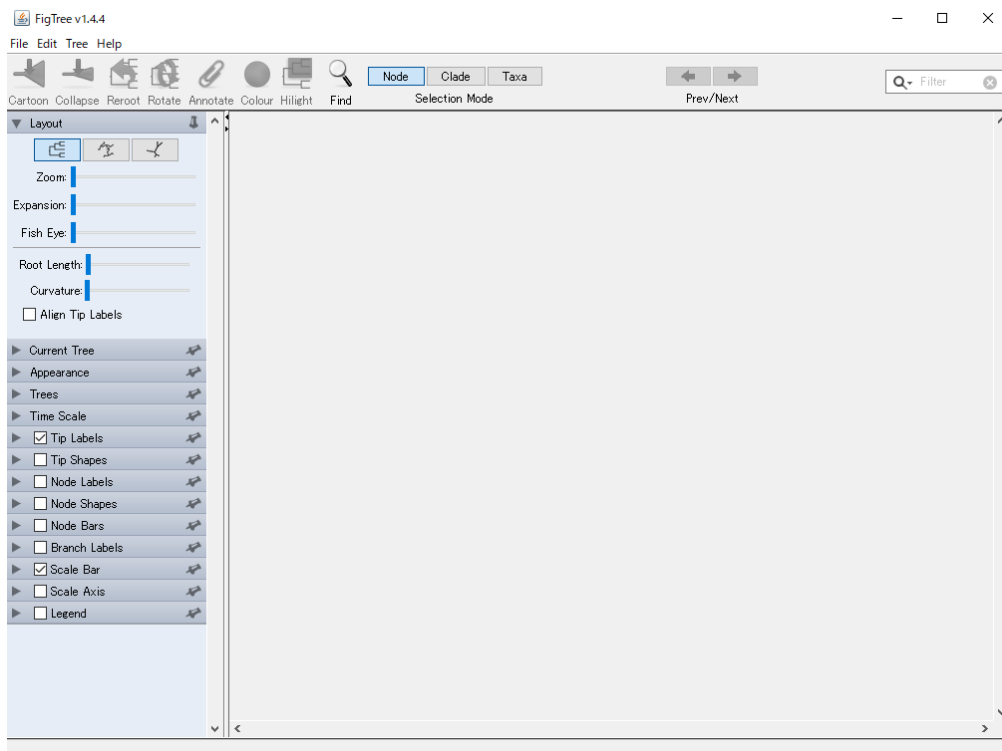
8

9 5. 結果の確認

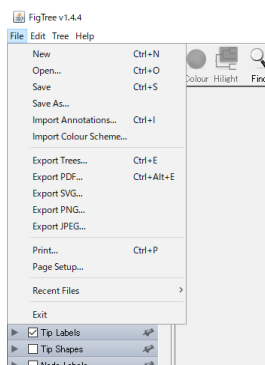
- 1 MCMCTREE の詳細な解析結果はアウトファイル(この場合 **hummingbird.out**)に書き込
- 2 まれますが、MCMCTREE は FIGTREE プログラムで可視化可能な **FigTree.tre** というフ
- 3 ァイルを自動的に生成してくれます。
- 4
- 5 そこでここでは FIGTREE プログラムによる結果の確認を紹介します。
- 6 FIGTREE をダウンロードしたら、その実行ファイル (ここでは FigTree v1.4.4.exe) をダ
- 7 ブルクリックで起動します。



- 8
- 9
- 10 そうすると以下のような画面が立ち上がります。

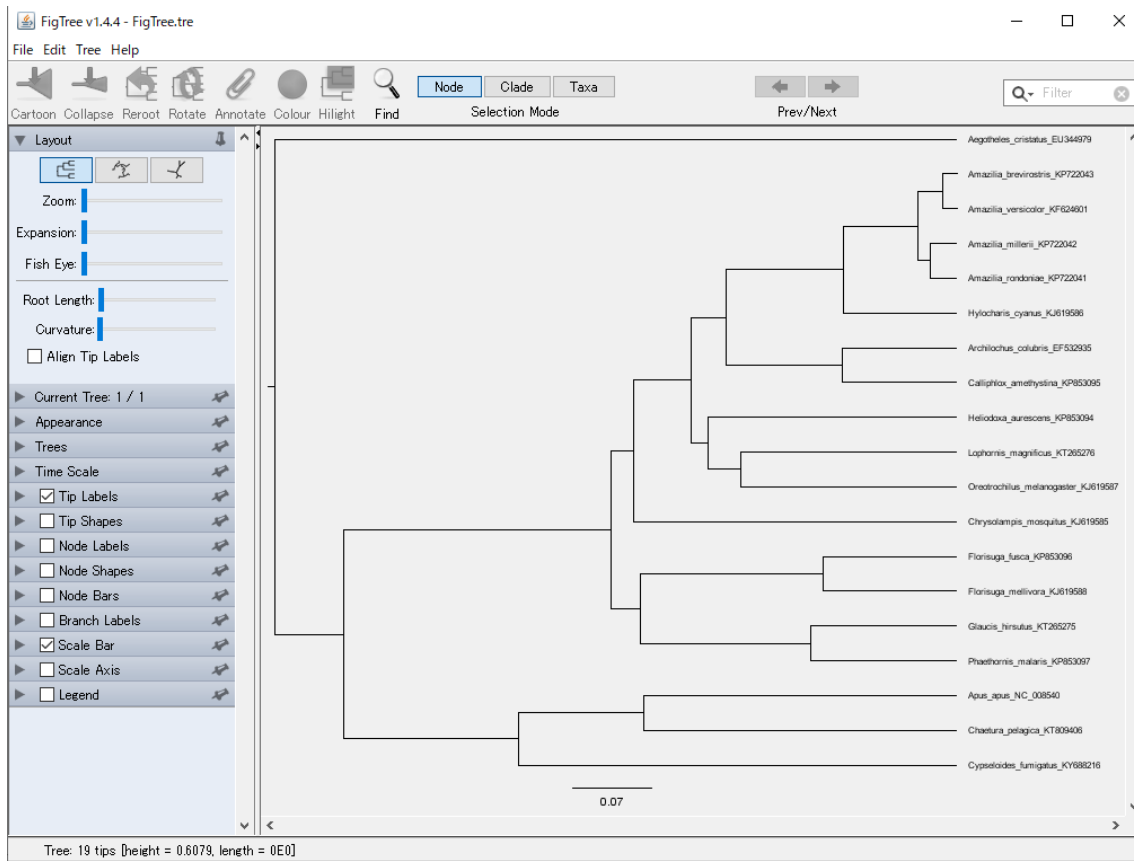


1
2



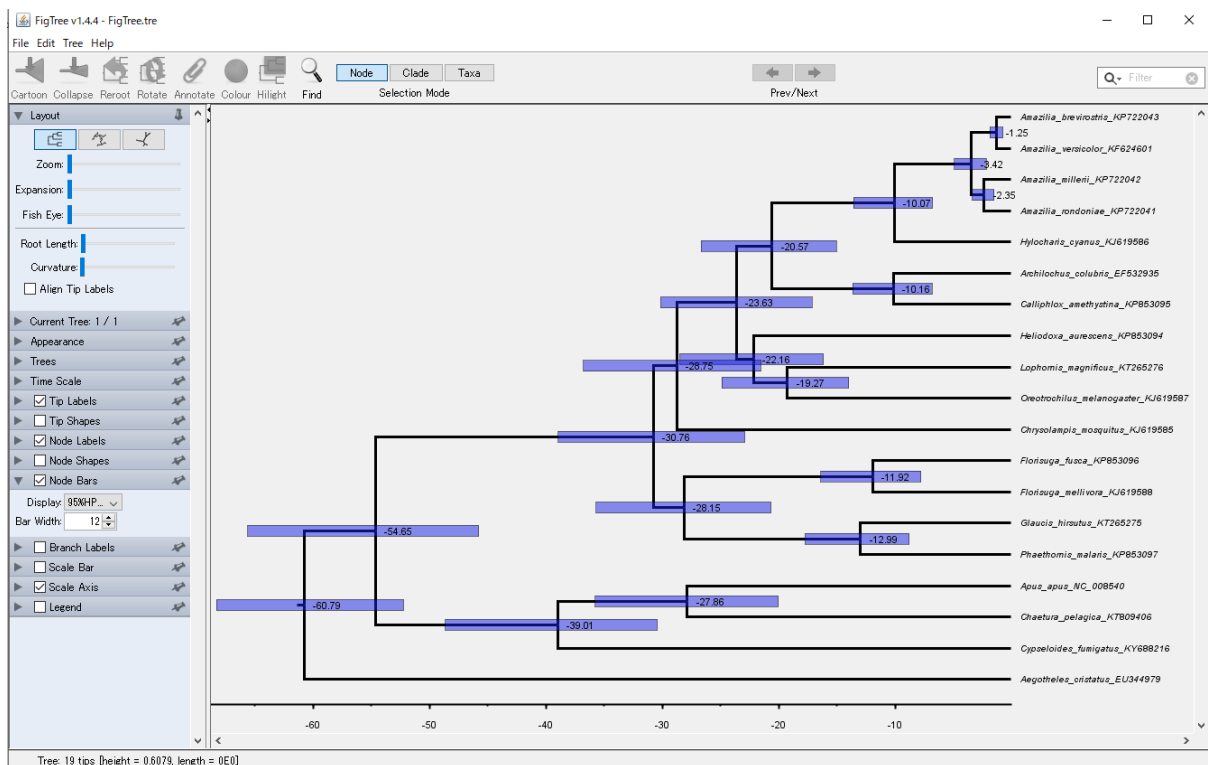
3
4
5
6

File → Open..を選択すると newick 形式や nexsus 形式の樹形ファイルを選択できるので、MCMCTREE によって生成された **FigTree.tre** を選択して開きましょう。



1
2

3 ノードを回転させたり、フォントのサイズを変えたり、時間スケールを変更したうえで時
4 間軸をつけたりと様々な描画のオプションがあります。また **FigTree.tre** にはノードの分
5 岐年代推定値の 95%最高事後密度信用区間の情報も書き込まれているので、以下の図のよ
6 うに Node Bars として年代推定値の信用区間を表示することも可能です。



- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8

謝辞

本例題の作成にあたり東京工業大学大学院生命理工学研究科・二階堂雅人先生と二階堂研究室の皆様より多くのご助言をいただきました。この場をお借りしてお礼申し上げます。

1 **help.zip** について
2 **help.zip** のフォルダの中に、本例題で用いたのとまったく同じ設定のファイルを入れてお
3 きました。PAML の解析がうまくいかない場合、これらのファイルを参考にしてご自身の
4 ファイルの設定を確認していただければ、解決策を見いだせると思います。

5
6 **help.zip** を解凍すると、6つのフォルダが入っています。これらは第二章～第三章の各々
7 の解析に対応していますので、

8
9 **第二章 2 節：選択圧の推定**

10 **1 項. 枝モデルを用いた選択圧の推定**

11 **Branch_model** というフォルダの中に

12 **branch2w.ctl** (コントロールファイル)

13 **branch1w.ctl** (コントロールファイル)

14 **2wML1.nwk** (樹形ファイル)

15 が入っています。

16 ここに PAML の codeml.exe プログラムと **hummingbird_12mtCDS.fas** を入れていただい
17 たうえで、コマンドプロンプトで以下のコマンドを使って実行しましょう

18

19 **codeml branch2w.ctl**

20

21 帰無仮説を試したい場合は

22 **branch1w.ctl** (コントロールファイル)

23 を使い以下のコマンドを使って実行しましょう

24

25 **codeml branch1w.ctl**

26

27

28

29 **2 項. サイトモデルを用いた選択圧の推定**

30 **Site_model** というフォルダの中に

31 **site.ctl** (コントロールファイル)

32 **ML15sp.nwk** (樹形ファイル)

33 **hummingbird_12mtCDS15sp.fas** (アラインメントファイル)

34 が入っています。

35 ここに PAML の codeml.exe プログラムを入れていただいたうえで、コマンドプロンプト
36 で以下のコマンドを使って実行しましょう

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36

codeml site.ctl

3 項. 枝サイトモデルを用いた選択圧の推定

Branchsite_model というフォルダの中に

branchsite.ctl (コントロールファイル)

branchsitenui.ctl (コントロールファイル)

2wML2.nwk (樹形ファイル)

が入っています。

ここに PAML の codeml.exe プログラムと **hummingbird_12mtCDS.fas** を入れていただい
たうえで、コマンドプロンプトで以下のコマンドを使って実行しましょう

codeml branchsite.ctl

帰無仮説を試したい場合は

branchsitenui.ctl (コントロールファイル)

を使い以下のコマンドを使って実行しましょう

codeml branchsitenui.ctl

第三章 2 節 : 分子時計を仮定した分岐年代推定

2 項 : 分子時計を仮定しないモデルによる尤度推定

Noclock_model というフォルダの中に

noclock.ctl (コントロールファイル)

ML.nwk (樹形ファイル)

が入っています。

ここに PAML の baseml.exe プログラムと **hummingbird_12mtCDS.fas** を入れていただい
たうえで、コマンドプロンプトで以下のコマンドを使って実行しましょう

baseml noclock.ctl

3 項 : 分子時計を仮定したモデルによる尤度推定

Clock_model というフォルダの中に

1
2 **clock.ctl** (コントロールファイル)
3 **MLrt.nwk** (樹形ファイル)
4 が入っています。
5 ここに PAML の baseml.exe プログラムと **hummingbird_12mtCDS.fas** を入れていただい
6 たうえで、コマンドプロンプトで以下のコマンドを使って実行しましょう

7
8 **baseml clock.ctl**

9
10
11 **3 節：緩和型分子時計による分岐年代推定**

12 **Relaxedclock_model** というフォルダの中に

13 **mcmctree.ctl** (コントロールファイル)

14 **MLrt2.nwk** (樹形ファイル)

15 が入っています。

16 ここに PAML の baseml.exe プログラムと mcmctree.exe プログラム、

17 **hummingbird_12mtCDS.fas** を入れていただいたうえで、コマンドプロンプトで以下のコ
18 マンドを使って実行しましょう

19
20 コマンドプロンプトで以下のコマンドを使って実行しましょう

21
22 **mcmctree**

23
24
25 これで枝の長さとその分散共分散行列が推定され、out.BV というファイルが作られます。

26 このファイルの名前を in.BV と変更したうえで、**mcmctree.ctl** の usedata オプションを

27
28 **usedata = 2**

29
30 に変更しましょう。そのうえで、コマンドプロンプトで以下のコマンドを使って実行しま
31 しょう

32
33 **mcmctree**

34
35