- 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賞コード遺伝子の解析を
- 3 実践せよ。
- 4



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

1 **解說編**

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

Phylogenetic Analysis by Maximum Likelihood (PAML)
2Deep Yoro
Table of contents
Installar
Introduction
APRE, Ex a balage of organs for single-process of the evolution of the evo
This document is about downloading and campiling Mark, and getting started. See the manual (see 50.02, ptf) for more information about numing programs in the package.
Downloading and Setting up PAML
PAML-X: A GUT for PAML
Server and a support of the server of the se
The following in writem for the naive user. If you know things like folders, executable files, and search path, there is no need for you to follow the matructions here.
PAML for Windows 5x/NT/2000/XP/Vista/7
Decisional or de norma en la entre associations de contra entre associations de contra entre
setting ap a failer of an anti-setting search and h. You need to do the for you can stand head h. You need to do the for you can stand head h. You need to do the for you can stand head h. You need to do the for you can stand head h. You need to do the for you can stand head h. You need to do the for you can stand head h. You need to do the for you can stand head h. You need to do the for you can stand head h. You need to do the for you can stand head h. You need to do the for you can stand head h. You need to do the for you can stand head h. You need to do the for you can stand head h. You need to do the for you can stand head h. You need to do the for you can stand head h. You need to do the for you can stand head h. You need to do the for you can stand head h. You need to do the for head head head head head head head head

- 15 ※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。
- 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)
 をダウンロードし、Lhaplus 等を用いて解凍して用います。
- 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
- 21 す。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/



- 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)。

E G A Cen	ecular Evolutionary etics Analysis	**●**	tutorial - fea	tures documentatio	n - feedback
<	MEGA 11 introc Calibration densities, tij	luces expanded rela dating, and a rate auto-cor 0000000	xed-clock datin relation test have beer	g methods n added to MEGA.	>
Windows	; v Gri	nphical (GUI) 🗸 🗸	MEGA 11 (64-bit)	V DOWNLO	AD 🚫
Sequence A	nalyses	Statistical Met	nods	Powerful Visu	al Tools
Site Links Home Videos Walk through Books / Articles Features Publications	Documentation Online Manual MEGA 10 Annual PDF / H Example Data FAQ Update History Known Issues	Downloads Windows GUI / CC Mindows GUI / CC Ubuntu/Debian GUI RetHat/Fedora GUI Chter Unius (CC) tar Older Versions	с. Эрания Эрани	Ţ.	

- 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パーテー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 iqtree-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	~ 3	更新日時	種類	サイズ
h bin	2	022/02/07 14:12	ファイル フォルダー	
example.cf	2	022/02/07 14:12	CF ファイル	2,185 KB
🔁 example.nex	2	022/02/07 14:12	NEX ファイル	1 KB
📴 example.phy	2	022/02/07 14:12	PHY ファイル	34 KB
🔁 models.nex	2	022/02/07 14:12	NEX ファイル	119 KB

13

14 この bin というフォルダに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
📧 iqtree2.exe	2022/02/07 14:12	アプリケーション	9,761 KB
📧 iqtree2-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

- 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ミ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 - 乄モ帳	- 🗆 X
ファイル(F) 編集(E) 書式(O) 表示(V) ヘルプ(H)	
*Aesotheles_cristatus_EU344979 TIC TIT GGA CAA TT GAA AGC CCA AGC CTA ATG GGA ATC CCC TTG ATC CTA CTC TCC JAC CCA CAC TC CAAC ATA AAC TTA ACA $$ ACC ACT AAA GCC ACC TCC TGA GTC TGA TTC GGA CAC CCA GAA GTC TAT ATC CTC ATC CTA CCA GGC TTT GGA ATC ATC TC CTA CAT TG GGA CAC CCA GAA GTC TAT ATC CTC ATC CTA CAC GGC TTT GGA ATC ATC TC CTA CAT TATA GTA GAC CCA GAC CAT TATA CCA TC ATC CTA CAC GGC TTT GTC CAA ATA GTA GAC CCA CAC ATC CTA CA ACA ATC ATC GTA GGC TTC GTC CAA ATTA GTA GAC CCC CCC CCA TATA CCAA ACC ATC GT GAC TTC GTT CCA CTT GGG CCC CCC CCA CTA CTA AAA CCA ACC ATC CTA GTT CAC TT CCA CT CCA CTA GTT ACT CCT CCA CAT ATC AAA CCA GAC TCC TT CTA TT GGA CTA GCT CTT CTT TTT TTC GG GGC ATA ATC AAA CCA ATC TTC CTA TGAA GTC ACC TA GTT ACT ATT GGG CCC CCA CCA GCA GTA GCA CAA ACC ATT CCA TC GAA GTC ACT CTA GTT ACT CTT CTT TTT TTC GG GGC ATA ATC AAA CCA ATC CTA CTA AGT CAC ATG GGC CTA ATT ACC ATT GGT ACT CCC CAA CAC ATC ATC CTA CTA ACA GGA CAA TGC ATT CTA CTT CTT TTA TTT CTG GGC ATA ATC AAT CCAA CCA ATC ATA CACA GGA CAA TGG GGC ATT ACC ATT GT ATC CTC AAT CC ACA GCA TGA GGA AGC CAA ACC ATTC CTA CACA ATA ATC CTT ACC CTA ATC TTA CTA CACA GGA TGA GGA AGC CAA TCA ACA GGA CAC TGG CCA CTA CT ACC CTA ATC CTA ATC ACA GGA TGA ACC ATC ATC CTA CACA ATA TC CTTC CT	: TTA TTA TTT CCT GCA CTA CTC CTC CCC ACG COG GAC AAC CGA TGA CTC ACC AAA 4 ACC TTC ATT AAC CGA TGA TTA TTT TCA ACT AAC CAC AAA GAC ATT GGE CAC CAC ACA CTT GGT GCC TAC TAC GCA GGA AAA AAA GAG CAT TTT GGE TAC ATA GGA ATA GTA ATT CAA GAA GAC ATT GGA CAC TAC GCA GGA AAA AAA GAG CAT TTT GGC TAC ATA GGA ATA GTA ATT CAA GAA GAC ATT GGC ACC ATT GGC CAC TTA GCC ATC CAC AAA CAA CTA GGA TTC CAA GAA GCC CAC TTA GCC ATC TAG CAC ATA GAA ATA GTA CAA GAC ATT CAA GAA CAC CAC GAC ATA GCA ATT GAA GAC ATC CTC AGC ATT GCC ATC CTA GCC ATC TCC TA GCC CAT TGC CTA GCC ATC ATC CTA ACC ATA ACC CAC AGA CTA CTC CTA GCC ATC ATC CAC ATA ACC CAC AGA CTA CTC CTA AGC ACA ACC ATC GAA GCA ACT ACC CAA CTA ACC CTA ACC ACA CTA GCC CAC ATT ACC CTA ACC ACA ACT ACC CAC ACT ACC ACA CTA GCC ATA ACA CTC ACC ATA ACA ACC CTA GCC ATC ATC CTC ATC ACC ATA CAAC ATA CAC ATC ACC CTA CAC ACA CTA ACC ACA ACT ACC ACA ACT ACC ACA ACA
*Amazilia_brevirostris_VP722043 *Amazilia_brevirostris_VP722043 TIC TIT GAC CAA ITT ATA AGE COC TAC CTA CTC GGA ATT CCC CTA ATC CTT ATC TCT JAC GCC COC TCA GAG ATT A AGE COC TAC CTA CTC GGA ATT CCC ATA CTC TCA CCC TGA TTT GGC CAC CTC GAA GTT TAC ATC CTC ATC CTA CCA GGA TTC GAA ATC ATC CTC CAC TAC AGE COC CTCA GAG GTT TAC ATC CTC ATC CTA CCA GGA TTC GAC ATC TAC GGC CAC CTC TCC CCC ATA CC CTA CTC GGA GCA GCC GC CTTA CTC CAC AT TA GTA GAC CCA AGC CTA TGA CCC TA TTC GGA GCA GCC GCC CTTA CTC CAC AC TA GGA CCC ACC CCC CTA GTA ATC AAC GAC CTA CTA ATC GAC CTC CCA ACT TAT GGA CCC CCC CCA GTA GAC CAA ACT ATT CAAC GAC TCA CTA ATC GAC CTC CCA ACT TAT GGA CCC CCC CCA GTA GAC CCA ACT ATT TTC CTA TGA GGA GCT CCC TTC CTC GTA CAA TTG GGC CCC CCC GCA GCA GTA ATC CAA CCA GTA TTC CTA TTC GCA TAC GAC TTA CTTA TTA TTC TCA AGC ATA GTC AAC CCT CTA CTC CCA ATG AGT ATT ACA ATC GCC CCA CTC CCC CAC GA AGT ATT CAAC GGA GCA CTA GTG ACT CAC GCA TAT GGG CCC CCC GA GCA GTA GCC CTA CTT CTTA CTA ATG AGC GCA ATG AGT ATT ACA ATC GCC CTC CAC AGC ATA TCA ACC CTA CTC CTA CTC CCA CTA GTG ACT CC CCC ATA CCC TTA TTA TTC TCA AGC ATA GTC AAT GCC TTA CTC CCA CAC ATG ACT ATT ACA ATC GCC CAC ACTC TCC TCC ACA CTA CTC CTA CTC CCA ACT AGC GCA ACA TTC CTA CCC TTA ATTC AAT GCA TTC GTA ACA GTG TCC TTA CTA ATTC CTA CCC ATA ATC CTA ATC AATA CCA GTA ACT CTTC CTA CTA ATC CTTA CTT ACC ATA GCC CTA CTC TCC CTA CAC CTC CTA ATC CTTA CTA ATT ATT CAT AGC CTA AAT GGT GAG CAA GAC ATC CGG AAA ATA GGA GGA TTG CAA AAA ATA GTA AGC	I TTA CTA TTC CCA ACC CTA CTA TTC CCC TCC CCC GGC AAC CGA TGA ATC ACA AA($_{\rm A}$ ACT TTT ATC AAT CGA TGA TTA TTT TCA AAC CAAC C
*Amazi I i a mi I eri i KP722042 TIC TIT GÃC CAA TIT ATA AGE CCC TAC CTA CTC GGA ATT CCC CTA ATC CTC ATC TC TTC GGC CAT CCT GAA AAA GCC CCA ATA AGE ACA AAA ACC CCA TCC TGA CCC TGA TTT GGC CAT CCT GAA GTT TAC ATC CTC ATC CTA CCA GGA TC GGA ATC ATC TCA CAT TT GGC CAT CCT GAA GTT TAC ATC CTC ATC CTA CCA GGA TC GGA ATC ATC TCA CAT TG ATC CAG GGC TGC CCT CCC CCA TAC CCA ACC TTT GAG GGC GGC CTT TT GTC CAA TA GTA GAC CCA AGC CCA TGA CCC CTA TTC GGA GCA GOC GCC CTA ATC CCA ACC CCC TA GTA ACA CCC CCC CCA CTA AAA ATA GTC AAC GAC TCA TTA ATC GGA CTC CCC AACC CCC TA GTA ACA CCC CCC CCA CATA ACA ATA GTC AAA GTG TAC TTC CTA TTC GCA TTA ATT GGG GCC CTC CGC CCA CATA GCC CAA ACT ATC TC TC TTT GCA GTG CCC CCC ATT ATT GGG GCC CTC CGC ACA GTA GCC CAA ACT ATC TCC TAT GAA GTG ACC CTC GCT ATC ATC GGC ACA CTC TCC TCC ATA TCA ACC CTA CTA CT	I TIG CTA TTE CCA ACT CTG CTA TTE CCC TCC CCA GGT AAC CGA TGA ATC ACA AA ACT TTT ATC AAT CGA TGA TTA TTT CCA ACC AAC CAC AAG GAC ATT GGE ACC CT/ I GTA GTA ACA TAT TAC ACC GGT AAA AAG GAA CCA TTC GGE TAC ATG GGE ATG GTI A TC CAA GAA GCC AAC CAC TCC GCA CTA GGA TTC CAA GAC GCT TCA TCC CCA AT TC CAA GAA GCC AAC CAC TCC CAA ACC TCC GCA ACC CCC ATT CTA GGA CTC ATC ATA TGG TTC CAC CAC AAC TCC GGA ACC CTA TTA TCC TTA GGE CCA TCA AAC ATC TCA ACC TGA TGA AAA CAC TAC GGA TCC CTC CTA GGC CTA TCC CAG CA ATC CATA TGG TTC CAA CAC CTC GGA ACC CTC CTA GGC CTA TCC CAG CCA ATC CCA AAC AAA CTA GGA AGC TCC TTG AGC CTA GGC CTA TCC CCG CGA TCC ATC CCA AAC AAA CTA GGA GGC TTT GCC CTA GGC CTA CCC TCC CGA TCC ATC ATT ATT TTA ACC GGA AAC TAC ACC TTG GAC ACC CTA GGC CAA CTA AAC CAC CCC CAA CAC TCA CTA GGC ACC ACC GCA ATC AGA CAC CTA AAC CAC CCC CAC CTA TCA TTA ATC GCA CCT GCC CTA CTC CTC CTC CTT CTT TCA TCA TCA TTA ACC GGC ACC CTA CTC CACA CTA AAC CTAC CTA ATTA TTC TA ACC GCA CTC ACT GCT CTC CTA GTC ACA CTA AAC CTAC CTA ATCA TCA TTA TTC TTA ATC CCC CTA CTA
	Windows (CRLF) 1 行、1 列 100%

- 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
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
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
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。

/// partiti	on.txt - 🗡	E帳			-		X	
ファイル(F)	編集(E)	書式(O)	表示(V)	ヘルプ(H)				
DNA,cod DNA,cod DNA,cod	on1=1-1 on2=2-1 on3=3-1	0758¥3 0758¥3 0758¥3						^
								\lor
<							>	
				Windows (CF	1行、1列	100%		

- 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。

📙 🛃 📮 🚽 bin				- 🗆	×
ファイル ホーム 共有 表示					^ 🕐
オートー オート オー オー	修動先 コビー先 移動先 コビー先	目 ▼ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □			
クリップボード	整理 新規	開く	選択		
← → ヾ ↑ PC > デスクトップ > iqtree-2.	1.3-Windows > iqtree-2.1.3-Windows > bin		✓ ひ binの検索		P
▲ カメッカ マクセラ	^ 名前 [^]	更新日時	重頼 サイ	ſズ	
=	M hummingbird_12mtCDS.fas	2022/02/07 15:15	AS ファイル	267 KB	
	iqtree2.exe	2022/02/07 14:12	アプリケーション	9,761 KB	
♣ 9000-F	iqtree2-click.exe	2022/02/07 14:12	アプリケーション	9,761 KB	
ドキュメント	ibiomp5md.dll	2022/02/07 14:12	アプリケーション拡張	1,089 KB	
■ ピクチャ	A partition.txt	2022/02/07 15:40	テキストドキュメント	1 KB	
📙 bin					
miseq_trimmed_hiseq_seqs_filtered.fna 5 個の項目	v				

- 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

- 2 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
- 3 す。

5	2 0
	International Content of Content
4	
5	
6 7	binのフォルダへと移動したら以下のコマントをダイブしましよう。
/ Q	istron? a humminghird 12mtCDS for m TEST in partition by R 1000 profix
0 Q	humminghird -T 4
10	
11	" igtree2 "は IO-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
14	イル名(この場合 hummingbird_12mtCDS.fas)を入力します。
15	
16	-m はモデルを指定するオプションです。
17	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
18	-m GTR+F+I+G
19	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
20	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
21	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
22	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。
23	
24	S. Kalyaanamoorthy, B.Q. Minh, T.K.F. Wong, A. von Haeseler, L.S. Jermiin
25	(2017) ModelFinder: Fast model selection for accurate phylogenetic estimates.
26	Nat. Methods, 14:587-589. https://doi.org/10.1038/nmeth.4285
27	
28	-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
29	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
30	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
31	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
32	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

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、個別のパーテーション 1 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 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 6 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., 35:518-522. https://doi.org/10.1093/molbev/msx281 13 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 ントファイル名やパーテー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 17 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。 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。 21 22 23 3節:解析結果の確認と系統樹の可視化 24 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。 25 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

26 ます。

ファイル ホ テ クイック アクセ スにピン留め	-ム 共有 表示 -ム 共有 表示 -ム 切り取り いmm パスのコピー コピー 貼り付け 戸 ショートカットの貼り付け クリップボード	移動先 1ビー先 整理 整理	新しい項目・ ゴショートカット・ オリングー 新規 新規	■ 日 間 < ~ 日 間 < ~ 日 間 < ~ 日 間 < ~ 日 道 す/ つ パティ ◎ 環座 開 < 8 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
← → °	↑ · · · · · · · · · · · · · · · · · · ·	1.3-windows > lqtree-2.1.3-window	s > bin	~ 0
1 n	名前	史新日時 裡規	717	
	🔁 hummingbird.best_model.nex	2022/02/07 16:15 NEX	ファイル 1 Ki	3
	hummingbird.best_scheme	2022/02/07 16:15 BEST	SCHEME 77 1 KI	3
*	🔰 hummingbird.best_scheme.nex	2022/02/07 16:15 NEX	ファイル 1 Ki	3
a	hummingbird.bionj	2022/02/07 16:15 BION	J ファイル 1 Ki	3
**	💐 hummingbird.ckp.gz	2022/02/07 16:15 WinZ	ip ファイル 18 Ki	3
	hummingbird.contree	2022/02/07 16:15 CON	TREE ファイル 2 KI	3
	hummingbird.iqtree	2022/02/07 16:15 IQTRI	E ファイル 12 Ki	3
	hummingbird.log	2022/02/07 16:15 テキス	トドキュメント 11 Ki	3
	hummingbird.mldist	2022/02/07 16:15 MLDI	ST ファイル 5 KI	3
	🔍 hummingbird.model.gz	2022/02/07 16:15 WinZ	ip ファイル 5 Ki	3
0	🔁 hummingbird.splits.nex	2022/02/07 16:15 NEX	ファイル 2 Ki	3
Di Di	hummingbird.treefile	2022/02/07 16:15 TREE	HLE ファイル 2 KI	3
	M hummingbird_12mtCDS.fas	2022/02/07 15:15 FAS 7	/アイル 267 Ki	3
	iqtree2.exe	2022/02/07 14:12 アプリ	ケーション 9,761 Ki	3
•	iqtree2-click.exe	2022/02/07 14:12 アプリ	ケーション 9,761 Ki	3
 1	libiomp5md.dll	2022/02/07 14:12 アプリ	ケーション拡張 1,089 KI	3
	partition.txt	2022/02/07 15:40 テキス	トドキュメント 1 KI	3
-				
M ~				
17個の項目				

- 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(どのパーテー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

- 9 に 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。



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	M 11: Tree Explorer (hummingbird.tre)	Newick Export Options
0		File Search Image Subtree View Compute Caption Help Save Current Session Save Current Session	▼ General
9	分岐年代推定には、IQ-	Export Timetree (Tabular) Export Timetree (Tabular) Export Timetree (Nexus)	Branch Lengths
10	TREE2 プログラムが推定	Export All Trees (Newick) Export Analysis Summary Export Partition List	Node Labels
11	した系統樹の樹形のみを	Export Pairwise Distances Write Tree in a Table Format	▶ Gene Tree
12	用います。枝長やブート	Export Group Names Import Group Names Import Name Translations	→ Timetree
10	フレラップ弦支の桂却は	1 Show Info	н Г
15	ストノッノ唯学の旧報は	Printer Setup	
14	不要なので、樹形の情報		
15	のみを出力します。	Divergence Times	-
16	MEGA で系統樹を描画す	Collapse/Expand Lineages	- X
17	る際、 File のタブをクリ	Compute Display Caption	Help Image: Cancel Image: OK
10	ックナフト ナプシーン		Done Loading
18	ツクすると、オノション	SBL = 5.08794007	Ready
19	が現れるので Export		
00		中レイナ	

- 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 押すと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。

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	
	M11: Text File Editor and Format Converter – 🗆 X
	1 (Aegotheles_cristatus_EU344979, (((((((Amazilia_brevirostris_KP722043, Amazilia_versicolor_KF624601), (Amazilia_millerii_KP722042, Amazilia_rondoniae_KP722041)), Hylocharis_cyanu o
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。

1 **第二章: 選択圧の推定**

2 ハチドリのミ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 3

4 1節:準備

5 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、②CODEML プログラムのコントロール
 7 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
 9 ファイルの4点セットが必要になります。

10

14 15 16

17 18

- 11 ① PAML の CODEML プログラムの実行ファイル
- 12 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
- 13 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。

★ 10 000 000 000 000 000 000 000 000 000	り取り スのコピー ョートカットの貼り付け	移動先 コピー先 削除	 三 名前の 変更 	「 通 新しい フォルダー	rしい項目 ▼ 'ヨートカット ▼	 ブロパティ ブロパティ 浸 編集 砂 履歴 	■ す	べて選択 訳解除 訳の切り替え	
クリップボード		整理		新規	ł	開く		選択	
← → ~ ↑ 📙 > PC > デス	クトップ > paml4.9j						ڻ ~	paml4.9jの検索	م
	名前	^	更	新日時	種類	サイズ			
🗼 🏄 クイック アクセス	bin		201	2/02/07 17:41	7741.7+	II.81_			
📥 OneDrive	dat		20/	2/02/07 17:41	ファイルフォ	ルダー			
	doc		202	2/02/07 17:41	ファイルフォ	ルダー			
PC	examples		200	2/02/07 17:41	ファイルフォ	ルダー			
> 🧊 3D オブジェクト	src		202	2/02/07 17:41	ファイル フォ	ルダー			
> 🕹 ダウンロード	- Technical		202	2/02/07 17:41	ファイル フォ	ルダー			
> 🛄 デスクトップ	3s.trees		202	20/02/20 20:24	TREES 771	CD4	1 KB		
> 🟥 ドキュメント	4s.trees		203	20/02/20 20:24	TREES 771	(JL	1 KB		
> 📰 ピクチャ	5s.trees		202	20/02/20 20:24	TREES 771	()L	1 KB		
> 📕 ビデオ	6s.trees		200	20/02/20 20:24	TREES 771	JU	1 KB		
> 🎝 ミュージック	aami.cti		202	20/02/20 20:24	CIL J711		1 KB		
> 🏭 OS (C:)	brown puc		20/	20/02/20 20:24	NUC 774	٠ ۱۱.	2 KB		
> 🧭 DVD RW ドライブ (D:) NSBU	brown rooter	trees	200	0/02/20 20:24	TREES 7p4	CIL.	1 KB		
> 🚅 ボリューム (E:)	brown.trees		200	20/02/20 20:24	TREES 771	(JL	1 KB		
> SONY_16GT (F:)	codeml.ctl		202	20/02/20 20:24	CTL ファイル		3 KB		
> 📋 gikenbio (G:)	codonml.ctl		202	20/02/20 20:24	CTL ファイル		1 KB		
	GeneticCode.	txt	202	20/02/20 20:24	テキストドキ	ュメント	5 KB		
gikenbio (G:)	/// MCaa.dat		202	20/02/20 20:24	DAT ファイル	ŀ	2 KB		
SONY_16GT (F:)	MCbase.dat		200	20/02/20 20:24	DAT ファイル	ŀ	2 KB		
- 📣 ネットローク	MCbaseRand	omTree.dat	200	20/02/20 20:24	DAT ファイル	ŀ	2 KB		
- 1717 V	MCcodon.dat		202	20/02/20 20:24	DATファイル	b in the second s	5 KB		
Gatch!	mcmctree.ctl		202	20/02/20 20:24	CIL J711		2 KB		
	pamp.cti		20/	20/02/20 20:24	ロビノアイル		1 KB		
	paupend		20	20/02/20 20:24	ファイル		1 KB		
	paupstart		202	20/02/20 20:24	ファイル		1 KB		
	README.txt		200	20/02/20 20:24	テキストドキ	ュメント	2 KB		
	stewart.aa		200	20/02/20 20:24	AA ファイル		1 KB		
	stewart.trees		202	20/02/20 20:24	TREES 771	(JL	1 KB		
	/// yn00.ctl		202	20/02/20 20:24	CTL ファイル		1 KB		
									·
311回以現日									8==
			_						. 16.
'AML の様々な	ふプログ	フムの実行	丁フ	アイル	はbir	nという	ファ	トルダの日	コに格為
-									

れていま

	ファイル ホーム 共有 表示 ^ メ ・ ・ ・ ・ ・ クリック アクセ コビー 貼り付け ショートカットの貼り付け ・ ・ 創除 名前の 変更 アメルダー ・ ・ ・ カイック アクセ コビー ・ ショートカットの貼り付け ・ ・
	クリップボード 整理 新規 開く 選択
	← → × ↑ → PC > テ X2 h y 7 > paml4.9j > bin
	> 筒 ドキュメント ■ basemLexe 2020/02/20 20:24 アプリケーション 509 KB
	> ■ ビブチャ ■ basemig.exe 2020/02/20 20:24 アブリケーション 484 KB > ■ ビデオ ■ chi2.exe 2020/02/20 20:24 アブリケーション 165 KB
	トレージック 「正 codeml.exe 2020/02/20 20:24 アブリケーション 757 KB E evolver.exe 2020/02/20 20:24 アブリケーション 476 KB
	>
	→ 量ポリューム(E) III mcmctree.exe 2020/02/20 20:24 アブリケーション 567 KB → III SONV_16GT (F) IIII mann gree 2020/02/20 20:24 アブリケーション 473 KB
	→ j gikenbio (G:) i yn00.exe 2020/02/22 20:24 アブリケーション 455 KB
	> 👔 gikenbio (G:)
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	
	13
	15

※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
 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 形式のアライン
 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
 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
 ンメントファイルは、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
 があります。

7

12 13 14

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
 11 ルダに入れて用います。

★ ↓ イックアクセ コピー ↓ スにピン留め	■ よ切 ■ パ ^{貼り付け} 記 シ	り取り スのコピー ヨートカットの貼り付け	移動先 パー先	★ 三 削除 名前 変更	し の 新しい こ フォルダー	「前新しい項目▼ 1 ショートカット▼	ער דין דער	開く → 🔡 編集 🔡 1 履歴 📲 1	ドベて選択 瞿択解除 瞿択の切り替え		
	クリップボード		整	理		新規	開く		選択		
← → * ↑ 🗌	→ PC → デス	マクトップ > selection						~ ē	selection)検索	ç
 オ クイック アクセス デスクトップ ダウンロード ドキュメント ビクチャ 	7 7 7 7	名前	_12mtCDS.fas		更新日時 2020/02/20 20 2020/02/20 20 2022/02/07 18 2022/02/07 15	裡類 :24 CTL ファイ :24 アプリケー :01 NWK ファ :15 FAS ファイ	サ ション パル イル	1 X 3 KB 757 KB 1 KB 267 KB			
」 bin 4 個の項目	~										

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
                                      * tree structure file name
       treefile = stewart.trees
                                 * main result file name
       outfile = mlc
          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:betaγ
                      * 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.<u>s of estimates</u>
  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.
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。
```

```
15
```

1	
2	<pre>seqfile = stewart.aa * sequence data filename treefile = stewart.trees * tree structure file name outfile = mlc * main result file name</pre>
2	
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	
18	<pre>seqtype = 2 * 1:codons; 2:AAs; 3:codons>AAs CodonFreq = 2 * 0:1/61 each, 1:F1X4, 2:F3X4, 3:codon table</pre>
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ブションです。
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	ち標準遺伝コートは、停止コトンを除くと61 種類のコトンがあるので、すべてのコトン
30 21	か1/61という頻度で仔仕することになります。なおこのオフションを選択すると、コド
ა1 აი	ノ(火用)須皮に送するハフノーダ (以口)値になります。 1 な深いと、マラインメントへ体の指すの短時()たりについい感
ა∠ ეე	I を迭かと、ノノインアント主体の塩基の頻度(π_A 、 π_T 、 π_G 、 π_C)をもとにコトン残
33	反の期付値が訂昇されます。例えるAUIというコトノの場合、ての頻度は $\pi_{A\times}\pi_{C\times}\pi_{T}$

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 1 2 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 3 ラメータ数は3個になります。 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 4 5 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 6 7 ータ数は9個になります。 8 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 9 10 (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。 11 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 12 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。 13 14 icode = 0 * 0:universal code; 1:mammalian mt; 2-10:see below 15 16 これはコドンのコード表を選ぶオプションになります。 17 18 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 GENEBANK. 19 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。 20 21 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 22 23 同一)を用いているので、 24 icode=1 25 を選択しましょう。 26 27 知っておくと便利なお役立ち情報 28 29 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。 30 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、 31 やはり IO-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 32

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
 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。

タ回は、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
た。しかし IQ-TREE2 などで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
クを外して)fix_blength=2 を選択しましょう。また IQ-TREE2 などで推定した枝の長さ
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
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
りの置換数を推定することになるので、IQ-TREE2 などで推定した枝の長さは使っていま
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	
	<pre>model = 2 * models for codons: * @:one, 1:b, 2:2 or more dN/dS ratios for branches * models for AAs or codon-translated AAs: * @:poisson, 1:proportional, 2:Empirical, 3:Empirical+F * 6:FromCodon, 7:AAClasses, 8:REVaa_0, 9:REVaa(nr=189) </pre>
16	<pre>NSsites = 0 * 0:one w;1:neutral;2:selection; 3:discrete;4:freqs; * 5:gamma;6:2gamma;7:beta;8:beta&w9:betaγ * 10:betaγ+1; 11:beta&normal>1; 12:0&2normal>1; * 13:3normal>0</pre>
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。 <u>通常はこの</u>
30	2を選択します。
31	

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。
6	
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。



- 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。

(Aegotheles_cristatus_EU344979,(((((((Amazilia_brevirostris_KP722043 ,Amazilia_versicolor_KF624601),(Amazilia_millerii_KP722042,Amazilia_ rondoniae_KP722041)),Hylocharis_cyanus_KJ619586),(Archilochus_colubr is_EF532935,Calliphlox_amethystina_KP853095)),(Heliodoxa_aurescens_K P853094,(Lophornis_magnificus_KT265276,Oreotrochilus_melanogaster_KJ 619587))),Chrysolampis_mosquitus_KJ619585),((Florisuga_fusca_KP85309 6,Florisuga_mellivora_KJ619588),(Glaucis_hirsutus_KT265275,Phaethorn is_malaris_KP853097))),((Apus_apus_NC_008540,Chaetura_pelagica_KT809 406),Cypseloides_fumigatus_KY688216));

18

- 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。



- 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。



- 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はじまり23 ます。

C:¥Users¥Yoneza 0.016550	wa¥Desktop¥s 0_089774	election¥coder	ml.exe						-	
0.016550	0.089774	0.017								
0.056230	0.027181 0.079672	0.0174 0.0771 0.0643	170 0.05 24 0.06 183 0.06	5760 54808 55474	0.083286 0.061225 0.075092	0.023607 0.074756 0.424701	0.050919 0.019246 0.371625	0.015590 0.029416 0.389681	0.018979 0.064753	0.08009 0.05934
itime & nrate	&np: 3	53								
ounds (np=38) 0.000004 0.000004 4 0.000004 50.000000 50.000000 50.000000 0 50.000000 0 50.000000	0.000004 0.000004 0.000004 0.000004 0.000000 0.000000 0.000000 50.000000 50.000000 50.000000	0.000004 0.000004 0.00000 4 0.0001 50.000000 50.000000 50.000000 0 999.0000	0.000004 4 0.0000 04 0.000 50.000000 50.00000 0 50.0000 0 50.0000	0.00 04 0.0 004 0. 000 0 00 50.0 00 50.0 000 50.0	00004 0.000 000004 0.00 0.000004 0.0 0.000100 00000 50.000 00000 50.00 000000 50.0 000000 50.0 000000 50.0	0004 0.00 00004 0.0 000004 0. 00000 50.00 00000 50.0 00000 50.0	0004 0.0000 00004 0.0000 000004 0.000 00000 50.0000 00000 50.0000 000000 50.000	4 0.000004 04 0.000004 004 0.00000 0 50.000000 00 50.000000 000 50.00000	0.000004 0.000004 4 0.000004 50.000000 50.000000 0 50.000000	0.0000 0.000 0.000 50.000 50.000 50.000
ip = 38 nL0 = -95530.1										
terating by m nitial: fx= 9 = 0.03943 0 0.05576 0.08 1925 0.02942	ing2 5530.62551 .04651 0. 329 0.023 0.06475	9 10380 0.0 61 0.0509 0.05934	18284 0.02 12 0.01558 0.05623 (2201 0. 0.018 0.07967	.10564 0.03 398 0.08010 0.06438 0	774 0.0673 0.02925 .06547 0.0	5 0.01796 0. 0.02718 0.077 7509 0.42470	06582 0.0165 12 0.06481 0.37162 0.3	5 0.08977 0.06123 0.0 8968	0.01747 07476 (
1 h-m-p 0.00 2 h-m-p 0.00 3 h-m-p 0.0	000 0.0000 000 0.0000 000 0.0000	61140.367 16707.754 51986.397	10 +YCYYCY 12 CCCCC 76 17 C	(CCC 769 3730.412	938.554493 s 2275 4 0.000	9 0.0000 00 107	58 0/38 0/38			

3 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
4 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
5 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
6 ると良いです。

7

8 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

9 す。



10 11

12 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

13 とのぞいてみました。つまり樹形ファイルには Aegotheles_cristatus_EU344979 が存在す

14 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

15 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



1	というように
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、 <u>コマンドプロンプトか</u>
5	<u>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</u>
6	<u>ィングを行うことが出来ます。</u>
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
10	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
11	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
12	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
16	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
17	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。
	■ #RC-VUsersWonezawaVDektopVselectionKodemLeve – □ X
	erfor when opening the coolent.ctil tell me the full path-name of the file?
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
23	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
24	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。
25	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。

לדעסד אלגעם אלגעסד 📾	-	×
accept = 0.11934 0.01920 0.90560		^
w = 0.01820 dN = 0.00704 dS = 0.36668 d4 = 0.23533 (2070.4 four-fold sites) dNx= 0.00739 dSx= 0.31529 Sx =2737.78 Nx =7912.22		
d123[1] = 0.09248 0.01422 0.24845 average 0.15172 B] = 0.77496 0.74058 0.38477 average 0.63345 accept = 0.11334 0.01820 0.38050		
w = 0.01920 dN = 0.01216 dS = 0.63345 d4 = 0.40854 (2070.4 four-fold sites) dNx= 0.01277 dSx= 0.54467 Sx =2797.78 Nx =7912.22		
d128[+1] = 0.08783 0.01250 0.33001 average 0.14408 B] = 0.78585 0.70834 0.38540 average 0.60157 accept = 0.11834 0.01520 0.38560		
w = 0.01920 dN = 0.01155 dS = 0.60157 d4 = 0.38808 (2070.4 four-fold sites) dN≭= 0.01212 dS≭= 0.51726 S≭ =2797.78 N# =7912.22		
d123[⊾] = 0.10126 0.01557 0.33153 average 0.16612 E] = 0.54853 0.81083 0.42130 average 0.69359 accept = 0.11334 0.01520 0.36560		
w = 0.01920 aN = 0.01332 aS = 0.89359 aH = 0.44513 (2070.4 four-fold sites) aN≈ 0.01388 aS≈ 0.59638 S× =2787.78 N× =7812.22 end of tree file.		
Time used: 3:41		
C:¥Users¥Yonezawa¥Desktop¥selection>		~

- 1 2 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。 3 RateAncestor = 1 * (0,1,2): rates (alpha>0) or ancestral states (1 or 2) 4 5 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 6 7 下の画面のまま止まってしまいます。しかしアウトファイルは正しく作られているようで す。 8 🚾 コマンド プロンプト - codeml nse character -uity in difcodonNG: TCC ---se character '-' uity in difcodonNG: TCC ---nse character '-' uity in difcodonNG: TCC ---uity in difcodonNG: TCC --ambiguity strange ambiguity er '-' codonŅG: TCC --odonNG: TCC ---Franse character "-" biguity in difcodonNG: CTA ----46.262412 from ProbSitePattern. Nstruction. 9 10 11 12 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 13 14 ディターで開いてみましょう。 15 知っておくと便利なお役立ち情報 16 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 17 18 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 19 トロールファイルの verbose = 1 * 0: concise; 1: detailed, 2: too much 20 21 verbose=0 22 を選択しておくと、これらの情報は記録されません。 23 24 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。 25 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 26
- 27 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 Int(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 0.331482 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 neucleotide 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 (uv/us) for branch
	branch t N S dN/dS dN dS N*dN S*dS
1	201 1.300 8304.3 2405.7 0.0192 0.0347 1.8089 288.4 4351.5 2021 0.933 8304.3 2405.7 0.0225 0.0289 1.2855 240.0 3092.6 2122 0.670 8304.3 2405.7 0.0225 0.0022 0.0963 18.0 231.8 2223 0.171 8304.3 2405.7 0.0225 0.0030 0.1316 24.6 516.6 2425 0.286 8304.3 2405.7 0.0225 0.0039 0.1316 24.6 516.6 2425 0.286 8304.3 2405.7 0.0225 0.0039 0.1316 24.6 516.6 2425 0.286 8304.3 2405.7 0.0225 0.0049 0.9342 73.6 948.3 2526 0.141 8304.3 2405.7 0.0225 0.0010 0.0455 8.5 109.6 272 0.018 8304.3 2405.7 0.0225 0.0007 0.0235 5.5 70.5 263 0.024 8304.3 2405.7 0.0225
2	
3	TREE # 1: (1, ((((((((2, 5), (3, 4)), 16), (7, 8)), (15, (17, 18))), 10), ((12, 13), (14,
4	19))), ((6, 9), 11)); MP score: 13598
5	
6	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
7	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
8	樹の置換数の総和になります。
9	
10	lnL(ntime: 35 np: 38): -68637.606213 +0.000000
11	、
10	
12	IP: 38 は、この脾材に用いられたハノタータの総数を衣しています。
13	
14	知っておくと便利なお役立ち情報
15	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
16	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
17	いるので、無根系統樹の場合、19×2-3=35(本文参照)になります。置換モデルの持つパラ
18	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
19	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
20	て、それぞれの ω を推定するため、ω に関しては ω#0と ω#1の2つのパラメータを持つ
21	ことになります。

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
 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
 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
 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。

8

9

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

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*ω_mが* 0.01920、*ω_m* 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。

19

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出来ます。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 とが出来ます。



3 **5**・統計的有意性の評価

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

6

7 これを検証するためには、 $\omega_{\#0} = \omega_{\#1}$ を仮定した帰無仮説(1ω モデルと呼びたいと思いま

- 8 す)を比較する必要があります。
- 9

10 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

11 みましょう。



	<pre>TREE # 1: (1, (((((((2, 5), (3, 4)), 16), (7, 8)), (15, (17, 18) InL(ntime: 35 np: 37): -68640.899450 +0.000000 201 2021 2122 2223 2324 2425 2526 369 3511 1.219520 0.956310 0.070409 0.172517 0.096024 0.287884 0.140993 0 0.417906 0.479664 9.310184 0.021632</pre>
	Note: Branch length is defined as number of nucleotide substitut:
	tree length = 9.906797
	(1: 1.219520, (((((((2: 0.017906, 5: 0.021285): 0.033084, (3: 0.0 0.186213): 0.405762, (14: 0.245676, 19: 0.301732): 0.439758): 0.0
	<pre>(Aegotheles_cristatus_EU344979: 1.219520, ((((((Amazilia_brevino Archilochus_colubris_EF532935: 0.163519, Calliphlox_amethystina_F 0.593849): 0.070409, ((Florisuga_fusca_KP853096: 0.195105, Floris Cypseloides_fumigatus_KY688216: 0.479664): 0.412625);</pre>
	Detailed output identifying parameters
	kappa (ts/tv) = 9.31018
1	omega (dN/dS) = 0.02163
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
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、ω#0=ω#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出来ます。
18	
19	エクセルの CHISQ.DIST.RT 関数を用いて
20	$X = 2 \times (\ln L_{2, \text{umodel}} - \ln L_{1, \text{umodel}})$
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	

		e Insert P	ige cuyour	Formulas	Data Revi	ew View	Help				
	Cut				~ A^ A*	= = =	≫~~ ab	Vrap Text		neral	
	Paste Sorma	at Painter B	t <u>u</u> ~ ⊞	- <u>A</u>	~ ^{abc} A ~		∈= →= [⊕]	vlerge & Cente	er ~ 🖸	~ % 9	€0 .00 .00 →0
	Clipboard	rs.		Font	F <u>N</u>		Alignment		F3	Number	6
	CHISQ.DIS *	: × 🗸	f _x =C	HISQ.DIST.RT(2*	(B2-A2),1)						
		A		B		С	D	E	F	G	H
	1 帰無仮説(2	(1ωモテル) -68640.89945	对立仮説(2ωモテル) -68637.606	21 B2-A2),1)						
	3						-				
	4	Function Arg	uments					?	×		
	6	CHISQ.DIST.F	π						_		
	7		x	2*(B2-A2)		★ = 6.586	474				
	8	_	Deg_treedom	1		T = 1 = 0.010	27565				
	10	Returns the ri	ght-tailed prot	bability of the chi-	quared distribut	tion.					
	11	_		nonne	gative number.	ou want to eva	uate the distri	Jution, a			
	12	Formula resul	t = 0.0102756	5							
	14	Help on this f	unction				OK	Cancel			
	15										
	17										
1	18										
1	18 19										
1 2	18 19										
L 2	18 19	م بر لار	አንበ በ	010275	65 to	わざ	5060	合除一	い准・	(
1 2 3	18 19 この場合	合 <u>、p</u> 値	[か 0.0	010275	65 ない	Dで、	5%の	危険	水準、	でバリ	
1 2 3 4	18 19 この場合 わない。	合 <u>、p</u> 値 グルーフ	[が 0.(プでは	010275 、ωに	<u>65</u> なの 有意差	Dで、 があ	5%の ること	危険	水準 [、] かり	でバリます。	
1 2 3 1	18 19 この場行 わない。	合 <u>、p</u> 値 グルーフ	[が 0.(?では	010275 、ωκ	<u>65</u> なの 有意差	Dで、 があ	5%の ること	危険が分け	水準 [・] かり	でバリ ま <u>す</u> 。	
1 2 3 4 5	18 19 この場行 わないか	合 <u>、p</u> 値 グルーフ	[が 0. (プでは	010275 、ωに	65 な ⁽ 有意差	ので、 があ	5%の ること	危険 が分:	水準 [・] かり	でバリます。	ン?
1 2 3 4 5	18 19 この場合 わないか	合 <u>、 p 値</u> グルーフ ωが 3 種	〔が 0. ([°] では 重類以	<u>010275</u> 、ωに 上あっ	<u>65 な</u> の 有意差 て、ω	<u>Dで、</u> があ #0=ω	<u>5%の</u> ること #1≠ u	危険 が分:	水準 [・] かり どの≶	<u>でバリ</u> ま <u>す</u> 。 条件で	· ン : ご検i
L 2 3 1 5 5 7	18 19 この場行 わない? もしもの	合 <u>、p値</u> グルーフ ωが3種	〔が 0.0 パでは 重類以	<u>010275</u> 、ωに 上あっ model	65 ない 有意差 て、 w	Dで、 があ #0= W	<u>5%の</u> ること #1≠ a	危険 が分 _{#2} な	水準 かり どの ア	でバリ ます。 条件で イル	ーン: ご検i
1 2 3 4 5 5 7	18 19 この場合 わない もしもの トローノ	合 <u>、p</u> 値 グルーフ ωが3種 ルファイ	iが 0.1 [°] では 重類以 [°] ルの	010275 、ωに 上あっ model	65 な ⁶ 有意差 て、 ⁶⁰	<i>Dで、</i> があ #0= ω Eでは	<u>5%の</u> ること #1≠ u なく、	危険 が分 ^{#2} な 樹形	水準 [・] かり どの ファ	でバリ ます。 条件で イルの	・ン ご検 の#1
1 22 33 44 55 56 77 83	18 19 この場合 わない もしも トローノ といい	合 <u>、p値</u> グルーフ ωが3種 ルファィ です。#	[が 0.1 パでは 重類以 2 を一	<u>010275</u> 、ωに 上あっ model 一括変換	65 なの 有意差 て、 w の 設定	Dで、 があ #0= w Eで変更	5% 0ること #1 $\neq u$ なく、 毛して	危険 が分: ^{#2} な 樹形 、CO	水準 かり どの ジ ファ DEN	でバリ ます。 条件て イル・ 4Lプ	レンジ ご検記 の#1 ロク
1 2 3 4 5 5 7 3 3	18 19 この場合 わない? もしもの トロー? 差があった	合、 p 値 グルーフ い が 3 種 ル フ す。 $ * *$	1が0.1 パでは 単加の 2 を レ	<u>010275</u> 、ωに 上あっ model 一括変換	65 なの 有意差 て、 w こで#1	<u>Dで、</u> があっ モロン は 更 +	5%の ること $\#1 \neq \alpha$ なして	危険 が分: ^{#2} 樹 、CO	水準 [→] かり どの ジファ DEN	<u>でバリ</u> ます。 条件で イル・ 4Lプ	レンジ ご検記 の#1 ロク
1 22 33 4 5 5 7 7 8 3 9	18 19 この場合 わないが もしもの といい 差がある	合 <u>、p値</u> グルーン が3種 ルフす。#	iが 0.1 プでは ガル を 2 証と	<u>010275</u> 、ωに 上あっ model 一括変換 いう流	65 なの 有意差 て、 の 定 #1 れにな	<u>Dで、</u> があ ========= こりま	5%0ること #1 $\neq a$ なく、 更してす。	<u>危険</u> が分; ^{#2} な 、 CO	<u>水準</u> かり どのジ ファ DEN	<u>でバリ</u> ます。 条件て イルの AL プ	- ンジ 検 ア #1
2 3 4 5 5 7 3 3 9 9	18 19 この場合 わない? もしもの といい 差があえ	合 <u>、p値</u> グルーフ が3種 ルファィ でかを様	が 0.0 では が ル の 2 証 と	<u>010275</u> 、ωに 上あっ model 一括変 流	65 な 有意差 て、設定 れにな	<u>Dで、</u> があ #10= いで変ま りま	5%0ること #1 $\neq a$ なく、 毛してす。	<u>危険</u> が分: #2な 、 てO	<u>水準</u> かり どのミ ファ DEN	<u>でバリ</u> ます。 そ件て イルの 4Lプ	- ンジ ジ検計 ロク
2 3 4 5 7 3 9	18 19 この場合 わない もしもの といい 差がある	合 <u>、p値</u> グルーフ が3種 ルファィ でかを様	が 0.0 パ で は が 0.0 パ で 類 ル を 一 2 証 と	<u>010275</u> 、ωに 上あっ model 一括変換 いう流	65 なの 有意差 て、 の 設定 れにな	<u>ので、</u> があーで変ま りま	5%の ること #1≠ a なく、 毛して す。	危険7 が分) #2なる 横形 、CO	<u>水準</u> かり ンの ジ ファ DEN	<u>でバリ</u> ます。 条件て イルの ALプ	- ンン **検 ア#1
2 3 5 5 7 3 9 9	18 19 この場合 もしもの といい 差がある	合 <u>、p値</u> グルーフ が3す ルフす。#	iが 0.0 パでは 加 ル を 2 証 と	<u>010275</u> 、ωに 上あっ model ・括変撩 いう流	65 なの 有意差 て、 w 記 で#1 れにな	<u>りで、</u> #0= 0 に りま	5%の ること #1≠ α 、 王 し す。	危険7 が分: #2なる 横形 、CO	<u>水準</u> かり どの∮ ファ DEN	<u>でバリ</u> ま <u>す</u> 。 条件て イルの AL プ	- ンジ ※ 検 部 の #1

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

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、ω>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	$cyanus_KJ619586), (Archilochus_colubris_EF532935, Calliphlox_amethystina_KP853095)), (Marchilochus_colubris_EF532935, Calliphlox_amethystina_KP853095)), (Marchilochus_KP853095)), (Marchilochus_KP853095)), (Marchilochus_KP853095)), (Marchilochus_KP853095)), (Marchilochus_KP853095)), (Marchilochus_KP853095)), (Marchilochus_KP853095))), (Marchilochus_KP853095)), (Marchilochus_KP853095))), (Marchilochus_KP853095))), (Marchilochus_KP853095))))))))))$
4	$Heliodoxa_aurescens_KP853094, (Lophornis_magnificus_KT265276, Oreotrochilus_melanometa) and the set of the s$
5	gaster_KJ619587))),Chrysolampis_mosquitus_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。フォーマ

- 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。

23

24

outfile = site.out



* main result file name

```
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。
25
```

	<pre>model = 0 * models for codons: * doine, 1:b, 2:2 or more dN/dS ratios for branches * models for AAs or codon-translated AAs: * doines, 1:proportional, 2:Empirical, 3:Empirical+F * 6:FromCodon, 7:AAClasses, 8:REVaa_0, 9:REVaa(nr=189) </pre>
1	<pre>NSSILES = 0 1 2 * 0:one w;1:neutral;2:Selection; 5:discrete;4:rreqs; * 5:gamma;6:2gamma;7:beta;8:beta&w9:betaγ * 10:betaγ+1; 11:beta&normal>1; 12:0&2normal>1; * 13:3normal>0</pre>
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 7	NSsites = 0 1 2
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
13	化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 $P_0 + P_1 = 1$ という制約を持つの
17	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	$_{2}$ (ただし ω_{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つ持つことになります。
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	いう状況ト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人れ子状になってい
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	知っておくと使利なお役立ち情報
28 20	サイトモブルでは、一几半モブル、中立モブル、選択モブルかよく用いられますか、はか によっ、あっ、のよりをス第回に、1 めい1 たいの知知たせたたいがサモニックバロ・ションを
29 30	にも ω_1 てい ω_2 のとりうる範囲に=1 ヤ>1 などの可約を持たない離散モデル(NSsites=3)、負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 $\beta \& \omega$ 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。

444 TREE # 1: (((((((1, 4), (2, 3)), 12), (5, 6)), (11, (13, 14))), 7), (8, 9), (10, 15)); MP score: -1 446 ▼ InL(ntime: 27 np: 29): -52375.673174 +0.000000 446 ▼ InL(ntime: 27 np: 29): -52375.673174 +0.000000 9.149357 0.199762 0.196682 0.306063 0.144552 0.33247 0.017880 0.021339 0.023306 0.046043 0.046338 0.2090 448 0.149357 0.199762 0.196682 0.306063 0.144552 0.033247 0.017880 0.021339 0.023306 0.046043 0.046338 0.2090 470 Note: Branch length is defined as number of nucleotide substitutions per codon (not per neucleotide site). 471 tree length * 6.044052 472 tree length * 6.044052 473 ((((((1: 0.017880, 4: 0.021339): 0.033247, (2: 0.046943, 3: 0.046328): 0.023306): 0.144552, 12: 0.209087) 474 (((((((1: 0.017880, 4: 0.021339): 0.033247, (2: 0.046943, 3: 0.046328): 0.023306): 0.144552, 12: 0.209087) 475 Ur((((((Amazilia_brevirostris_KP722043: 0.017880, Amazilia_versicolor_KF624601: 0.021339): 0.033247, (Amazi 476 Ur(((((Amazilia_brevirostris_KP722043: 0.017880, Amazilia_versicolor_KF624601: 0.021339): 0.033247, (Amazi 477 Ur.18 0.00226 488 db for each branch 484 db db for each branch 485 tree length * 0.149 8303.6 2454.4 0.0223 0.0005 0.0229 37.5 498.1 17.18 0.200 8303.6 2454.4 0.0223 0.0005 0.0229 37.5 498.1 18.19 0.167 8303.6 2454.4 0.0223 0.0005 0.2029 37.5 498.1 18.19 0.167 8303.6 2454.4 0.0223 0.0005 0.2029 37.5 498	2] 7 0. 0.3 ia_n								
Hatt # 1.1 (((((1, 4), (1, (2, 5)), 12), (2, 5)), (1, (1, (1, (1, (1, (1, (1, (1, (1, (1	0.3 .ia_n								
<pre>467 16.,17 17. 16 16.19 1920 2021 2122 221 224 2123 232 233 201 0.149357 0.199762 0.106682 0.306063 0.144552 0.033247 0.017880 0.021339 0.023306 0.046943 0.046328 0.20494 0.149357 0.199762 0.106682 0.306063 0.144552 0.03247 0.017880 0.021339 0.023306 0.046943 0.046328 0.2090 100 Note: Branch length is defined as number of nucleotide substitutions per codon (not per neucleotide site). 171 tree length = 6.044052 172 tree length = 6.044052 173 ((((((1: 0.017860, 4: 0.021339): 0.033247, (2: 0.046943, 3: 0.046328): 0.023306): 0.144552, 12: 0.209087) 174 ((((((Maszilia_brevirostris_KP722043: 0.017880, Amazilia_versicolor_KF624601: 0.021339): 0.033247, (Amazi 175 (((((Maszilia_brevirostris_KP722043: 0.017880, Amazilia_versicolor_KF624601: 0.021339): 0.033247, (Amazi 176 Detailed output identifying parameters 177 Detailed output identifying parameters 177 0.109 5383.6 2454.4 0.0223 0.0045 0.2029 37.5 498.1 177.18 0.200 5383.6 2454.4 0.0223 0.0045 0.2029 37.5 498.1 177.18 0.200 5383.6 2454.4 0.0223 0.0045 0.2029 37.5 498.1 177.18 0.200 5383.6 2454.4 0.0223 0.0045 0.0329 0.155 26.8 55.8 199.100 107 5383.6 2454.4 0.0223 0.0044 0.154 26.8 355.8 199.20 0.038 5383.6 2454.4 0.0223 0.0045 0.3029 37.5 498.1 199.20 0.366 5383.6 2454.4 0.0223 0.0046 0.3714 50.2 666.2 100.12 0.165 5383.6 2454.4 0.0223 0.0045 0.3029 37.5 498.1 199.20 0.38 5383.6 2454.4 0.0223 0.0045 0.3029 37.5 498.1 199.20 0.38 5383.6 2454.4 0.0223 0.0046 0.3024 3.5 50.6 120.21 0.165 5383.6 2454.4 0.0223 0.0045 0.3029 37.5 498.1 199.20 0.338 538.6 2454.4 0.0223 0.0046 0.3024 3.5 50.6 120.21 0.165 5383.6 2454.4 0.0223 0.0045 0.3029 5.6 67.3 109.22 0.003 5383.6 2454.4 0.0223 0.0046 0.3026 5.6 73.5 109.22 0.047 5383.6 2454.4 0.0223 0.0046 0.3026 5.6 73.5 109.22 0.047 5383.6 2454.4 0.0223 0.0046 0.3026 51.6 51.5 4.5 109.21 0.048 5383.6 2454.4 0.0223 0.0046 0.3026 51.6 51.5 4.5 109.21 0.047 5383.6 2454.4 0.0223 0.0046 0.3026 51.6 51.5 4.5 109.21 0.048 5383.6 2454.4 0.0223 0.0046 0.3026 51.6 51.5 4.5 109.21 0.048 5383.6 2454.4 0.0223 0.0046 0.3026</pre>	2 1 37 0. 0.3								
449 Note: Branch length is defined as number of nucleotide substitutions per codon (not per neucleotide site). 471 tree length = 6.044052 473 (((((((1: 0.017880, 4: 0.021339): 0.033247, (2: 0.046943, 3: 0.046328): 0.023306): 0.144552, 12: 0.209087) 474 (((((((1: 0.017880, 4: 0.021339): 0.033247, (2: 0.046943, 3: 0.046328): 0.023306): 0.144552, 12: 0.209087) 475 (((((((1: 0.017880, 4: 0.021339): 0.033247, (Amazilia_versicolor_KF624601: 0.02139, 0.03139): 0.033247, (Amazilia_versicolor_KF624601: 0.02139, 0.0140, 0.0535, 0.0321, 0.0214, 0.0214, 0.0214, 0.0214, 0.021, 0.0214, 0.0223, 0.0021, 0.0223, 0.02	0.3 lia_n								
<pre>tree length = 6.044052 tree length = 6.044052 ((((((1: 0.0178800, 4: 0.021339): 0.033247, (2: 0.046943, 3: 0.046328): 0.023306): 0.144552, 12: 0.209087) (((((((Amazilia_brevirostris_KP722043: 0.0178800, Amazilia_versicolor_KF624601: 0.021339): 0.033247, (Amazil Detailed output identifying parameters kappa (ts/tv) = 11.16001 mega (dW/dS) = 0.02226 dW dW & dK for each branch branch t N S dW/dS dM dS N*dN S*dS lenach t N S dW/dS 0.0022 0.0045 0.2029 37.5 498.1 1718 0.200 8303.6 2454.4 0.0223 0.0045 0.2029 37.5 498.1 1920 0.306 8303.6 2454.4 0.0223 0.0045 0.2029 37.5 498.1 1920 0.306 8303.6 2454.4 0.0223 0.0045 0.2029 37.5 498.1 1920 0.306 8303.6 2454.4 0.0223 0.0045 0.0229 37.5 498.1 1920 0.306 8303.6 2454.4 0.0223 0.0045 0.0229 37.5 498.1 1920 0.306 8303.6 2454.4 0.0223 0.0045 0.0229 37.5 498.1 1920 0.306 8303.6 2454.4 0.0223 0.0045 0.0229 37.5 498.1 1920 0.306 8303.6 2454.4 0.0223 0.0045 0.0229 37.5 498.1 2122 0.033 8303.6 2454.4 0.0223 0.0045 0.0229 37.5 498.1 2122 0.033 8303.6 2454.4 0.0223 0.0045 0.0229 37.5 498.1 2122 0.033 8303.6 2454.4 0.0223 0.0045 0.0229 5.4 71.2 224 0.011 8303.6 2454.4 0.0223 0.0045 0.0230 5.4 71.2 224 0.011 8303.6 2454.4 0.0223 0.0045 0.0230 5.4 71.2 224 0.021 8303.6 2454.4 0.0223 0.0045 0.0230 5.4 71.2 232 0.047 8303.6 2454.4 0.0223 0.0046 0.0230 5.4 71.2 243 0.023 8303.6 2454.4 0.0223 0.0041 0.0638 111.8 156.5 244 0.021 8303.6 2454.4 0.0223 0.0041 0.0653 11.6 154.5 252 0.047 8303.6 2454.4 0.0223 0.0041 0.0653 11.6 154.5 254 0.021 8303.6 2454.4 0.0223 0.0045 0.0657 0.73.9 2612 0.018 8303.6 2454.4 0.0223 0.0045 0.0657 0.73.9 272 0.041 8303.6 2454.4 0.0223 0.0045 0.0657 11.5 154.5 2526 0.051 8303.6 2454.4 0.0223 0.0454 0.0555 677.9 2526 0.051 8303.6 2454.4 0.0223 0.0457 0.0557 4.2 0.583.3 25. 0.051 8303.6 2454.4 0.0223 0.0457 0.0557 4.2 0.583.3 25. 0.051 8303.6 2454.4 0.0223 0.0457 0.0557 4.2 0.583.3 25. 0.051 8303.6 2454.4 0.0223 0.0457 0.0579 9.5 2526 0.051 8303.6 2454.4 0.0223 0.0459 0.0557 4.7 1.557.9 2526 0.051 8303.6 2454.4 0.</pre>	0.3								
<pre>474 ((((((1: 0.017860, 4: 0.021339): 0.033247, (2: 0.046943, 3: 0.046328): 0.023306): 0.144552, 12: 0.209087) 475 476 (((((((Amazilia_brevirostris_KP722043: 0.017860, Amazilia_versicolor_KF624601: 0.021339): 0.033247, (Amazi 477 478 Detailed output identifying parameters 479 480 kappa (ts/tv) = 11.16001 481 482 omega (dW/dS) = 0.02226 483 484 v dW & dS for each branch 485 484 484 484 484 484 485 branch t N S dW/dS dM dS N*dN S*dS 485 487 485 486 branch t N S dW/dS 0.0229 37.5 498.1 489 489 tal. 2.000 8303.6 2454.4 0.0223 0.0045 0.2029 37.5 498.1 489 1718 0.200 8303.6 2454.4 0.0223 0.0045 0.2029 37.5 498.1 489 1819 0.107 8303.6 2454.4 0.0223 0.0045 0.2029 37.5 498.1 489 1718 0.200 8303.6 2454.4 0.0223 0.0045 0.2029 37.5 498.1 489 1222 0.038 8303.6 2454.4 0.0223 0.0045 0.0042 0.1045 26.8 355.8 491 1920 0.306 8303.6 2454.4 0.0223 0.0045 0.0229 3.5 498.1 492 2021 0.145 8303.6 2454.4 0.0223 0.0045 0.0229 3.5 498.1 493 2122 0.033 8303.6 2454.4 0.0223 0.0045 0.0042 0.145 26.8 35.8 494 224 0.021 8303.6 2454.4 0.0223 0.0045 0.0243 4.5 59.6 495 224 0.021 8303.6 2454.4 0.0223 0.0045 0.0243 4.5 59.6 495 224 0.021 8303.6 2454.4 0.0223 0.0045 0.0230 5.4 71.2 496 2123 0.023 8303.6 2454.4 0.0223 0.0041 0.0633 11.5 19.7.7 497 232 0.047 8303.6 2454.4 0.0223 0.0041 0.0633 11.5 19.7.7 497 232 0.047 8303.6 2454.4 0.0223 0.0041 0.0653 11.5 155.6 499 2012 0.209 3303.6 2454.4 0.0223 0.0041 0.0653 11.5 155.5 499 2012 0.209 3303.6 2454.4 0.0223 0.0045 0.0259 5.4 71.2 406 212 0.033 8303.6 2454.4 0.0223 0.0041 0.0653 11.5 155.5 405 212 0.015 8303.6 2454.4 0.0223 0.0045 0.0250 5.4 71.2 409 2012 0.209 3303.6 2454.4 0.0223 0.0045 0.0583 11.5 155.5 405 212 0.015 8303.6 2454.4 0.0223 0.0045 0.0585 11.5 .5 95.6 405 212 0.047 8303.6 2454.4 0.0223 0.0045 0.0585 11.5 .5 95.6 405 212 0.058 3303.6 2454.4 0.0223 0.0045 0.0585 11.5 .5 95.6 405 11.2 0.159 3303.6 2454.4 0.0223 0.0045 0.0555 6.73.9 505 2526 0.091 8303.6 2454.4 0.0223 0.0045 0.0557 42.7 168.5 504 2510 0.79 8303.6 2454.4 0.0223 0.0045</pre>	0.: lia_n								
476 ((((((Amazilia_brevirostris_KF722043: 0.017880, Amazilia_versicolor_KF624601: 0.021339): 0.033247, (Amazi 477 Detailed output identifying parameters 478 Detailed output identifying parameters 479 magnitude 480 kappa (ts/tv) = 11.16001 481 omega (dW/dS) = 0.02226 484 W M & dS for each branch 485 branch t N S dN/dS dN dS N*dN S*dS 486 branch t N S dN/dS d. 0.0223 0.0045 0.0229 37.5 498.1 487 1118 0.200 8303.6 2454.4 0.0223 0.0045 0.01450 26.6 35.8 489 1718 0.200 8303.6 2454.4 0.0223 0.0042 0.1450 26.8 35.8 481 1920 0.386 8303.6 2454.4 0.0223 0.0042 0.1450 26.8 35.8 482 2021 0.145 8303.6 2454.4 0.0223 0.0042 0.01450 26.8 31.10.9 482 2122 0.033 8303.6 2454.4 0.0223 0.0040 0.0452 8.3 110.9 482 224 0.018 8303.6 2454.4 0.0223 0.0040 0.0231 5.9 7.7 483 2122 0.047 8303.6 2454.4 0.0223 0.0040 0.0231 5.9 7.7 484 221 0.018 8303.6 2454.4 0.0223 0.0041 0.0663 11.8 156.5 485 224 0.023 80.014 0.0663 11.8 156.5 485 232 0.047 8303.6 2454.4 0.0223 0.0041 0.06631 15.9 154.5 486 233 0.046 8303.6 2454.4 0.0223 0.0045	lia_n								
978 Detailed output identifying parameters 479 betailed output identifying parameters 480 kappa (ts/tv) = 11.16001 481 omega (dM/dS) = 0.02226 484 V dN & dS for each branch 485 branch t 486 branch t 487 OH & dS for each branch 488 V dN & dS for each branch 486 branch t 487 1617 0.149 S303.6 2454.4 0.0223 0.0065 0.2202 37.5 498.1 489 1718 0.200 S303.6 2454.4 0.0223 0.0063 0.4159 76.9 1020.7 489 1721 0.403 S303.6 2454.4 0.0223 0.0042 8.3 110.9 441 1922 0.030 S303.6 2454.4 0.0223 0.0042 8.3 110.9 442 2012 0.043 S303.6 2454.4 0.0223 0.0041 1.964 36.3 142.1 423 11.02 0.433 S303.6 2454.4									
480 kappa (ts/ty) = 11.16001 421 omega (dW/dS) = 0.02226 423 omega (dW/dS) = 0.02226 424 Wd & dS for each branch 425 branch t N S dW/dS dN dS N*dN S*dS 426 1617 0.149 & 5303.6 2454.4 0.0223 0.0045 0.2029 37.5 498.1 426 1718 0.2028 & 303.6 2454.4 0.0223 0.00450 0.2029 37.5 498.1 428 1718 0.2028 & 303.6 2454.4 0.0223 0.00450 0.2714 S0.2 666.2 439 1718 0.2028 & 303.6 2454.4 0.0223 0.00450 0.2714 S0.2 666.2 430 1819 0.107 S303.6 2454.4 0.0223 0.00450 0.2714 S0.2 666.2 442 2021 0.415 S303.6 2454.4 0.0223 0.00450 0.0221 1.16.19 452 211 0.418 S303.6 2454.4 0.0223 0.00460 0.0229 5.4 71.2 453 2123 0.047 S303.6 2454.4 0.0223 0.0041 0.0629 1.1.6 15.6 5 454 22.									
412 Omega (DM/G3) = 0.02220 423 434 40 Å Å Å 5 for each branch 424 40 Å Å Å 5 for each branch 425 branch t N S dM/d5 dM d5 N*dN S*d5 426 1617 0.149 8033.6 2454.4 0.0223 0.0045 0.2029 37.5 498.1 426 1718 0.200 8103.6 2454.4 0.0223 0.0045 0.2029 37.5 498.1 428 1718 0.200 8105.6 2454.4 0.0223 0.0042 0.1714 50.2 66.2 439 1718 0.107 8105.6 2454.4 0.0223 0.0042 0.1479 26.3 35.8 442 2122 0.633 8105.6 2454.4 0.0223 0.0043 0.157 42.9 7.1 452 2113 0.014 8105.6 2454.4 0.0223 0.0041 0.0453 8.3 110.9 454 2214 0.021 8105.6 2454.4 0.0223 0.0041 0.0453 11.8 156.5 2123 0.040 8105.6 <th></th>									
Has Has <th></th>									
4466 branch t N S dN/dS dN dS W*dN S*dS 4477 1617 0.149 8030.6 2454.4 0.0223 0.0045 0.229 37.5 498.1 4489 1718 0.200 8303.6 2454.4 0.0223 0.0042 0.1450 26.6 2 6666 0.2714 50.2 6662.2 469.18 13.19 0.106 2454.4 0.0223 0.0042 0.1450 26.8 355.8 451 1920 0.435 8306.2 2454.4 0.0223 0.0042 0.1450 26.8 355.8 452 2021 0.435 8306.2 2454.4 0.0223 0.0042 0.1450 26.8 355.8 452 224 0.0213 8306.2 2454.4 0.0223 0.0042 8.13 10.9 452 224 0.0213 8306.2 2454.4 0.0223 0.0413 15.5 96.7 12.2 452									
449 1718 0.200 8303.6 245.4 0.0223 0.0060 0.2714 50.2 666.2 491 1919 0.107 8303.6 245.4 0.0223 0.0030 0.126 26.8 355.8 491 1920 0.366 8303.6 245.4 0.0223 0.0043 0.1459 26.8 355.8 492 2021 0.145 8303.6 245.4 0.0223 0.0043 0.164 36.3 452.1 493 2122 0.033 8303.6 245.4 0.0223 0.0040 0.0452 8.3 110.9 442 221 0.013 8303.6 245.4 0.0223 0.0040 0.0217 5.9 7.7 457 232 0.047 8303.6 245.4 0.0223 0.0040 0.0211 5.9 5.6 459 2312 0.047 8303.6 245.4 0.0223 0.0041 0.052 11.6 15.4 5.5 450 <th></th>									
4300 1819 0.107 8303.6 2454.4 0.6223 0.0430 0.145 26.8 355.8 491 1920 0.368 8303.6 2454.4 0.0223 0.0493 0.155 76.9 1020.7 492 2021 0.145 8303.6 2454.4 0.0223 0.0493 0.4152 8.3 110.9 494 2222 0.033 8303.6 2454.4 0.0223 0.0405 0.425 8.3 110.9 494 2224 0.013 8303.6 2454.4 0.0223 0.0406 0.239 5.47 11.2 495 224 0.013 8303.6 2454.4 0.0223 0.0408 0.351 11.8 156.5 495 224 0.073 8303.6 2454.4 0.0223 0.0408 0.3620 16.5 5303.6 2454.4 0.0223 0.0463 0.2375 637.9 535.6 537.9 535.6 537.9 535.6 545.6 537.9									
20.21 0.145 0.145 0.021 0.023 0.5 0.021 0.023 0.5 0.021 0.023 0.5 0.021 0.023 0.5 0.021 0.5 0.021 0.5 0.021 0.5 0.021 0.011 0.021 0.011									
463 2122 0.033 838.6 244.4 0.0223 0.0045 0.3 310.0 444 221 0.018 838.6 244.4 0.0223 0.0045 0.0243 4.5 59.6 455 224 0.011 838.6 244.4 0.0223 0.0046 0.0243 4.5 59.6 452 224 0.021 838.6 244.4 0.0223 0.0046 0.0217 5.9 7.7 457 232 0.047 838.6 244.4 0.0223 0.0046 0.0529 11.6 15.4 5 462 232 0.047 838.6 244.4 0.0223 0.0048 0.0529 11.6 15.4 5 469 2312 0.046 838.6 244.4 0.0223 0.0040 0.0526 56.7 3 501 245 0.167 838.6 244.4 0.0223 0.0046 5.6 7.7 9 502 246									
494 22.1 0.018 6393.6 2454.4 0.0213 0.0069 0.0243 4.5 59.6 495 22.4 0.021 8303.6 2454.4 0.0223 0.0069 0.0203 5.4 59.6 495 22.4 0.021 8303.6 2454.4 0.0223 0.0069 0.0209 5.4 71.2 496 21.2 0.007 8303.6 2454.4 0.0223 0.0061 0.0628 11.6 154.5 492 232 0.046 8303.6 2454.4 0.0223 0.00614 0.629 11.6 154.5 492 232 0.046 8303.6 2454.4 0.0223 0.00614 0.620 11.6 154.5 501 12.4 -255 583.6 2454.4 0.0223 0.00614 0.2275 42.0 556.3 502 246 0.157 8303.6 2454.4 0.0223 0.0067 0.1206 55.6 73.9 503 1825<									
495 222.4 0.021 8303.6 2454.4 0.0223 0.0030 0.047 5.77.7 492 21.2.3 0.023 8303.6 2454.4 0.0223 0.0047 0.0331 5.9 77.7 497 232 0.047 8303.6 2454.4 0.0223 0.0040 0.0331 5.9 77.7 497 232 0.047 8303.6 2454.4 0.0223 0.0041 0.0623 11.6 154.5 492 2312 0.269 8303.6 2454.4 0.0223 0.0043 0.0223 11.6 154.5 499 2012 0.269 8303.6 2454.4 0.0223 0.0060 0.3262 66.6 844.2 501 245 0.167 8303.6 2454.4 0.0223 0.0051 0.3065 56.73.9 502 246 0.213 0.024 0.0627 0.12.7 168.5 56.4 56.25 56.26 56.3 56.56 56.73.9 56.25									
496 2123 0.043 8303.6 245.4 0.0223 0.0047 0.8317 5.9 77.7 497 232 0.047 8303.6 245.4 0.0223 0.0041 0.0633 11.8 156.5 498 233 0.046 8303.6 245.4 0.0223 0.0041 0.0623 11.6 156.5 498 2312 0.298 8303.6 245.4 0.0223 0.0063 0.281 152.5 607.3 500 1924 0.265 8303.6 245.4 0.0223 0.0064 0.362 66.6 884.2 502 245 0.167 8303.6 245.4 0.0223 0.0065 5.26 73.7 502 246 0.211 8303.6 245.4 0.0223 0.0045 0.6510 120.3 159.5 502 2526 0.091 8303.6 245.4 0.0223 0.0145 0.6510 120.3 1.99 0.357 9.05 130.5 <									
497 232 0.044 8303.6 243.4 0.0223 0.0213 11.8 156.5 482 233 0.046 8303.6 243.4 0.0223 0.0013 11.8 156.5 499 2012 0.209 8303.6 243.4 0.0223 0.0043 0.229 11.6 154.5 499 2012 0.209 8303.6 243.4 0.0223 0.0080 0.3202 66.6 844.2 501 1245 0.167 8303.6 243.4 0.0223 0.0061 0.2275 42.0 558.3 502 246 0.213 8303.6 243.4 0.0223 0.0067 0.306 55.6 73.9 503 1825 0.051 8303.6 243.4 0.0223 0.0467 0.3065 15.7 73.9 504 2511 0.479 8303.6 243.4 0.0223 0.0457 0.511<120.3 1597.9 505 2526 0.091 8303.6									
100 2012 0200 9183.6 2.444.4 0.022 0.0665 0.2841 52.5 607.3 501 122 0.205 5183.6 2444.4 0.022 0.0666 0.2841 52.5 607.3 501 245 0.167 5183.6 2444.4 0.022 0.0666 0.0206 66.6 884.2 501 245 0.167 5183.6 2444.4 0.022 0.0667 1.2.7 168.5 502 246 0.221 50.66 55.6 77.9 555 55.5 55.5 503 1825 0.051 5183.6 2444.4 0.022 0.0457 0.510 120.3 1597.9 505 2526 0.051 5183.6 2444.4 0.022 0.0150 121.3 120.3 1597.9 505 2526 0.051 5181.6 2444.4 0.022 0.0150 0.123 151.1 9.99.3 507 2614 0.348 5									
560 1924 0.265 8303.6 245.4 0.0223 0.0080 0.3622 66.6 884.2 512 245 0.167 8303.6 245.4 0.0223 0.0051 0.257 82.0 58.3 512 246 0.221 8303.6 245.4 0.0223 0.0067 0.3066 55.6 737.9 503 1825 0.491 8303.6 245.4 0.0223 0.0165 0.6510 120.7 168.5 542 5511 0.479 8303.6 245.4 0.0223 0.0127 0.1230 22.7 361.9 505 2526 0.091 8303.6 245.4 0.0223 0.0124 0.5101 120.3 1597.9 505 2526 0.091 8303.6 245.4 0.0223 0.0124 0.5101 130.3 1597.9 506 2613 0.5303 8303.6 245.4 0.0223 0.0126 0.4730 87.4 1160.3 507									
501 245 0.167 8303.6 244.44 0.0223 0.0627 0.2275 42.0 558.3 502 246 0.221 8303.6 2454.4 0.0223 0.0607 0.306 55.6 73.9 503 1825 0.051 8303.6 2454.4 0.0223 0.0605 0.6687 12.7 168.5 504 2511 0.479 8303.6 2454.4 0.0223 0.0457 0.510 120.3 1597.9 505 2526 0.091 8303.6 2454.4 0.0223 0.0457 0.510 120.3 1597.9 506 2513 0.508 8303.6 2454.4 0.0223 0.0130 0.8122 150.1 199.3 507 2516 0.493 8303.6 2454.4 0.0223 0.0160 0.8150 157.4 1668.8 507 1.6.77 0.6677 8303.6 2454.4 0.0223 0.0190 0.8151 157.4 1608.1 501									
502 2246 0.221 8303.6 245.4 0.0223 0.0055 0.066 55.6 737.9 51 1825 0.051 8303.6 245.4 0.0223 0.0015 0.0665 51.6 737.9 504 2511 0.479 8303.6 245.4 0.0223 0.015 0.0605 12.7 168.5 504 2512 0.479 8303.6 245.4 0.0223 0.015 0.0607 0.123 12.7 505 2526 0.091 8303.6 245.4 0.0223 0.0181 0.8122 12.7 10.9 506 2613 0.598 8303.6 245.4 0.0223 0.0181 0.8122 150.1 1993.3 507 2514 0.348 8303.6 245.4 0.0223 0.0180 0.815 157.4 4906.1 509 1627 0.433 8303.6 245.4 0.0223 0.0190 11.3 1477.4 510 279									
503 1825 0051 8303.6 245.4.4 0.0223 0.0457 0.2.7 168.5 504 2511 0.479 8303.6 245.4.4 0.0223 0.0455 0.510 120.3 1597.9 505 2526 0.091 8303.6 245.4.4 0.0223 0.0425 0.510 120.3 1597.9 505 2526 0.091 8303.6 245.4.4 0.0223 0.0130 0.2120 22.7 301.9 506 2613 0.598 8303.6 245.4.4 0.0223 0.0130 0.8120 150.7 507 7.6.27 803.6.2 245.4.4 0.0223 0.0190 0.8151 157.4 166.8 508 177 0.627 803.6.2 245.4.4 0.0223 0.0190 0.8151 157.4 160.8 509 1672 0.433 8303.6.2 245.4.4 0.0223 0.0190 0.813 11.3 1477.4 510 279 0.194 </th <th></th>									
564 2511 6.479 8363.6 245.44 0.6223 0.6119 160.3 1597.9 505 2525 0.6019 8303.6 245.44 0.6223 0.6027 0.123 22.7 301.9 506 2513 0.598 8303.6 245.44 0.6223 0.6019 0.1321 22.7 301.9 507 2613 0.598 8303.6 245.44 0.6023 0.6105 0.4730 87.4 1160.8 508 177 0.627 8303.6 245.44 0.6023 0.6199 11.3 1477.4 509 1627 0.443 8303.6 245.44 0.6223 0.6199 11.3 1477.4 509 1627 0.443 8303.6 245.44 0.6223 0.6060 0.2797 49.5 657.4 511 279 0.194 8303.6 245.44 0.6223 0.6059 0.2640 48.8 647.9 511 27.9 0.194 8303.6 <th></th>									
151 251.13 0.153 251.13 0.153 251.13 0.153 251.14 251.14									
507 2614 0.348 8303.6 245.4 0.0223 0.0196 0.4730 87.4 1166.8 508 177 0.627 8303.6 245.4 0.0223 0.0190 0.8516 157.4 2009.1 509 1627 0.433 8303.6 245.4 0.0223 0.0194 0.8516 157.4 2009.1 510 278 0.193 8303.6 245.4 0.0223 0.0194 0.8519 11.3 1477.4 510 278 0.193 9.356.7 4.9609 1 4.9679 4.95 567.4 511 279 0.194 8363.6 245.4 0.0223 0.0469 0.2409 48.8 647.9 512 169 9.044 8303.6 245.4 0.0223 0.0469 0.2409 48.8 647.9									
508 177 0.627 8383.6 2454.4 0.0223 0.0134 0.6619 111.3 1477.4 509 1627 0.443 8303.6 2454.4 0.0223 0.0134 0.6019 111.3 1477.4 510 278 0.197 8303.6 2454.4 0.0223 0.0066 0.2679 49.5 657.4 511 279 0.194 8303.6 2454.4 0.0223 0.0066 0.2679 49.5 657.4 511 279 0.194 8303.6 2454.4 0.0223 0.0066 0.2670 49.5 67.4 512 15. 28 0.493 8303.6 2454.4 0.0223 0.0066 0.2670 49.5 67.4 <th></th>									
509 16.27 0.443 8303.6 245.4 0.0223 0.0134 0.66019 111.3 1477.4 510 278 0.197 803.6 245.4 0.0223 0.0660 0.2579 49.5 657.4 511 279 0.194 8303.6 245.4 0.0223 0.0650 0.2640 48.8 647.9 512 16.7 0.494 4023 0.0464 0.6661 11.4 11.6									
27.1.8 0.197 8303.6 2454.4 0.0223 0.0056 0.2679 49.5 657.4 511 279 0.194 8303.6 2454.4 0.0223 0.0059 0.2640 48.8 647.9 512 16.8 0.433 8303 6.2454.4 0.0223 0.0059 0.2640 48.8 647.9									
513 28.10 0.255 8303.6 2454.4 0.0223 0.0077 0.3463 64.0 849.9									
514 2815 0.320 8303.6 2454.4 0.0223 0.0097 0.4353 80.5 1068.3									
516 tree length for dN: 0.1828									
517 tree length for dS: 8.2123									
519 520 Time used: 1:39									
523 Model 1: NearlyNeutral (2 categories) 524 555									
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 1617 1718 1819 1920 2021 2122 221 224 2123 232 233 201 529 0.154342 0.202098 0.102482 0.312720 0.142951 0.033102 0.018004 0.021031 0.023230 0.046691 0.046602 0.20633	1 70.								
530 531 Note: Branch length is defined as number of nucleotide substitutions per codon (not per neucleotide site).									
532 533 tree length = 6.120714									
534 535 (((((((1: 0.018004, 4: 0.021031): 0.033102, (2: 0.046691, 3: 0.046602): 0.023230): 0.142951, 12: 0.206337)	0.3								
536 537 (((((((Amazilia_brevirostris_KP722043: 0.018004, Amazilia_versicolor_KF624601: 0.021031): 0.033102, (Amazi	ia_n								
Detailed output identifying parameters									
kappa (ts/tv) = 11.56937									
543 544 MLEs of dN/dS (w) for site classes (K=2)									

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 InL(ntime: 27 np: 29): -52375.675174 +0.000000 1617 1718 1819 1920 2021 2122 221 224 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									
1617 0.149 8303.6 2454.4 0.0223 0.0045 0.2029 37.5 498.1									
1718 0.200 8303.6 2454.4 0.0223 0.0060 0.2714 50.2 666.2									
1819 0.10/ 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、*w*=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),
InL(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: 0.046602))
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
```

- 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	(w) for	site	classes	(K=2)
р: w:	0.97090 0.01535	0.02910 1.00000			

- 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(P₀=0.97090)と中立に進化するコドンサ
 7 イトの割合(P₁= 0.02910)を表しています。(※P₀+P₁=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 $\omega(\omega_0=0.01535)$ と中立に進化するコドンサイ
- 11 トの $\omega(\omega_1=1)$ 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 $\omega(=0.02226)$ 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 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 16 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 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 18 19 が、今回のハチドリのミ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)
     0.97090 0.00000 0.02909
      0.01535 1.00000 1.00000
w:
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

4 5

6

7 は、負の選択圧を受けているコドンサイトの割合(P₀=0.97090)、中立に進化するコドンサ
8 イトの割合(P₁=0)、正の選択圧を受けているコドンサイトの割合(P₂=0.02909)を表してい
9 ます(※P₀+P₁+P₂=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 $\omega(\omega_0=0.01535)$ 、中立に進化するコドンサイ 13 トの $\omega(\omega_1=1)$ 、正の選択圧を受けているコドンサイトの $\omega(\omega_2=1)$ を表しています。<u>ここ</u> 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 計算修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

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
 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
 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

1 ■3項:枝サイトモデルを用いた選択圧の推定

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。



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

9

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。

```
19seqfile = hummingbird_12mtCDS.fas * sequence data filename20treefile = 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

3	
4	<pre>model = 2 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&w9:betaγ * 10:betaγ+1; 11:beta&normal>1; 12:0&2normal>1; * 13:3normal>0</pre>
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。					
	<pre>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 201 2021 2122 2223 2324 2425 2526 2627 272 275 26 2035 3536 366 369 3511 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</pre>					
	Note: Branch length is defined as number of nucleotide substitutions per codon (not per neucle					
	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。

24 す。

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

site class	0	1	2a	2b
割合	\mathbf{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 $\omega(\omega_0)$ 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 $\omega(\omega_1)$ 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 $\omega_2(1 < \omega_2)$ 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。Site 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 class2a と site class2b 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 class2a と site class2b 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 $\omega(\omega_2)$ 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。
8	
9	<pre>seqfile = hummingbird_12mtCDS.fas * sequence data filename</pre>
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
20	<pre>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</pre>
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(ω2)を意
29	味します。そして omega = 1 にすることでこのω2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 <u>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。</u>

- 1
- 2 **第三章:**分岐年代推定

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分岐年代を推定してみよう

4

5 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

6 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
7 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
8 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

- 9 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10 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
- 11 行う必要があることから少し複雑です。
- 12

13 ※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、

- 14 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。
- 15

16 17 **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。

22

23 2節:分子時計を仮定した分岐年代推定

- 24 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.ctl)、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) にいれます。ML.nwk については、4節:樹形ファイルの出力で作
 成したものを用います。
- 30 (※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
- 31 ださい)
- 32

33 2項:分子時計を仮定しないモデルによる尤度推定

- 34 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。35
- 36 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。

5

6 7

8 9

10

11

12

13 14

```
seqfile = brown.nuc
      treefile = brown.trees
      outfile = mlb * main result file
noisy = 2 * 0,1,2,3: how much rubbish on the screen
                   * 1: detailed output, 0: concise output
       verbose = 0
                     * 0: user tree; 1: semi-automatic; 2: automatic
       runmode = 0
                     * 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
_kappa = 0 * 0: estimate kappa; 1: fix kappa at value below; 2: kappa for branches
     fix_kappa = 0
         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
                    * (0,1,2): rates (alpha>0) or ancestral states
 RateAncestor = 0
    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
一番上のほうの
       seqfile = brown.nuc
      treefile = brown.trees
       outfile = mlb * main result file
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
情報です。ここでは
       seqfile = hummingbird_12mtCDS.fas
       treefile = ML.nwk
       outfile = noclock.out
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
ください(ここでは noclock.out という名前にします)
以下は塩基置換モデルの選択に関するオプションです。
```

1	<pre>model = 4 * 0:JC69, 1:K80, 2:F81, 3:F84, 4:HKY85</pre>
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。
9	clock = 0 * 0:no clock, 1:clock; 2:local clock; 3:CombinedAnalysis
10	今回はここかキーになります。0は分子時計を仮定しないモテル、1は糸統樹全体で分子
11 12	時計を仮走するモテルです。ここでは最初は分子時計を仮走せずに尤度計算をしたいので
12	$\gamma \gamma \pi n r$ (clock=0) $\sigma x x c (1) e x c x \gamma_{o}$
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
19	alpha = 0.5 * initial or fixed alpha, 0:infinity (constant rate)
20	
21	fix_alpha = 0
22	
23	を選択すると「モテルが週用されます。
24 25	
20 26	aipha = 0.5
20	はΓ分布の形をきめる形状パラメータ(α)の初期値にたります。今回のケースではデフォ
28	μ h
29	
30	もし(あまりそのよう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
34	

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大丈夫です。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。
	<pre>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 InL(ntime: 35 np: 41): -74027.321471 +0.000000 201 2021 2122 2223 2324 2425 2526 2627 272 275 2628 283 284 2516 2429 2 0.247636 0.143316 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. 0.173080 0.024132 0.264494</pre>
	tree length = 2.18605
	(Aegotheles_cristatus_EU344979, ((((((Amazilia_brevirostris_KP722043, Amazilia_versicolor_KF624601), (Amazilia_millerii_KP722042, Amazili Oreotrochilus_melanogaster_KJ619587))), Chrysolampis_mosquitus_KJ619585), ((Florisuga_fusca_KP853096, Florisuga_mellivora_KJ619588), (Glau
	(Aegotheles_cristatus_EU344979: 0.247636, (((((((Amazilia_brevirostris_KP722043: 0.005910, Amazilia_versicolor_KF624601: 0.007039): 0.0102 Calliphlox_amethystina_KP853095: 0.058049): 0.056837): 0.017689, (Heliodoxa_aurescens_KP853094: 0.119146, (Lophornis_magnificus_KT265276: Florisuga_mellivora_KJ619588: 0.052170): 0.079455, (Glaucis_hirsutus_KT265275: 0.065383, Phaethornis_malaris_KP853097: 0.078268): 0.084175
	Detailed output identifying parameters

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	
201 2021 2122 2223 2324 2425 2526 2627 272 275 2628 283 284 2516 2429 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 0.056833 0.173080 0.024132 0.264494	7 0
tree length = 2.18605	
(Aegotheles_cristatus_EU344979, ((((((Amazilia_brevirostris_KP722043, Amazilia_versicolor_KF624601), (Amazilia_millerii_KP722042, Ama: Oreotrochilus_melanogaster_KJ619587))), Chrysolampis_mosquitus_KJ619585), ((Florisuga_fusca_KP853096, Florisuga_mellivora_KJ619588), (zil Gla
(Aegotheles_cristatus_EU344979: 0.247636, ((((((Amazilia_brevirostris_KP722043: 0.005910, Amazilia_versicolor_KF624601: 0.007039): 0.0 Calliphlox_amethystina_KP853095: 0.058049): 0.056837): 0.017689, (Heliodoxa_aurescens_KP853094: 0.119146, (Lophornis_magnificus_KT26527 Florisuga_mellivora_KJ619588: 0.052170): 0.079455, (Glaucis_hirsutus_KT265275: 0.065383, Phaethornis_malaris_KP853097: 0.078268): 0.08	010 76: 417
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	
ここで	
lnL(ntime: 35 np: 41): -74027.321471 +0.000000	
と書かれている箇所に着目しましょう。	
np:41	
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 す。	
47	

1 🛠	罰って	おく	5	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。
10	
11	3項: <u>分子時計を仮定したモデル</u> 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。
13	
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
16	
17	seqfile = hummingbird_12mtCDS.fas
18	treefile = MLrt.nwk
19	outfile = clock.out
20	
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)。
25 25	
26	そして
27	
28	Clock = 1
29 20	トナファルズ、ハフ味具ィズッズ演用されます。ほみの悲空は黄西ルロドズナままです。
3U 21	とすることで、万十時計モブルが適用されます。ほかの設定は削填と同しで入火天です。
31 22	9 母政マライルの毎年
ა∠ ვე	
33 24	前項で用す、WLIWKに石」の修正を加えたものを用います。WLITIWKをMEGAで用くと
54	



- 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。





- ogaster_KJ619587))), Chrysolampis_mosquitus_KJ619585), ((Florisuga_fusca_KP853096, Fl
 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。

- 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。



6

7 ハチドリ亜目の最古の化石記録の年代はリュプル期(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 ーペル期(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 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 12 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 _cyanus_KJ619586),(Archilochus_colubris_EF532935,Calliphlox_amethystina_KP853095)), 17 18 (Heliodoxa_aurescens_KP853094, (Lophornis_magnificus_KT265276, Oreotrochilus_melan ogaster_KJ619587))),Chrysolampis_mosquitus_KJ619585),((Florisuga_fusca_KP853096,Fl 19 20 orisuga_mellivora_KJ619588),(Glaucis_hirsutus_KT265275,Phaethornis_malaris_KP85309 21 7))),((Apus_apus_NC_008540,Chaetura_pelagica_KT809406),Cypseloides_fumigatus_KY6 88216)) @0.56); 22 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 にしておくと良いです。つまりここでは 100Ma(1 億

26 年)が1単位時間になります。

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 プレ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), InL(ntime: 18 np: 24): -74134.049288 +0.000000 20..1 20..21 21..22 22..23 23..24 24..25 25..26 26..27 27..28 28..2 28..5 27..29 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), (Amaz Oreotrochilus_melanogaster_KJ619587))), Chrysolampis_mosquitus_KJ619585), ((Florisuga_fusca_KP853096, Floris

(Aegotheles_cristatus_EU344979: 0.000000, ((((((((Amazilia_brevirostris_KP722043: 0.000000, Amazilia_versico Calliphlox_amethystina_KP853095: 0.000000): 0.000000): 0.000000, (Heliodoxa_aurescens_KP853094: 0.000000, (Lo Florisuga_mellivora_KJ619588: 0.000000): 0.000000, (Glaucis_hirsutus_KT265275: 0.000000, Phaethornis_malaris_

(1_Aegotheles_cristatus_EU344979, ((((((((2_Amazilia_brevirostris_KP722043, 5_Amazilia_versicolor_KF624601) (17_Lophornis_magnificus_KT265276, 18_Oreotrochilus_melanogaster_KJ619587) 32) 31) 24 , 10_Chrysolampis_mo 11_Cypseloides_fumigatus_KY688216) 36) 21) 20 ;

Detailed output identifying parameters

0.459203 Nodes and Times (JeffNode is for Thorne's multidivtime. ML analysis uses ingroup data only.) Node20(Jeffnode36)TimeNode21(Jeffnode35)TimeNode22(Jeffnode34)TimeNode23(Jeffnode33)TimeNode24(Jeffnode32)Time 0.56027 0.56000 0.35241 0.33462 0.28160 Node 25 (Jeffnode 31) Time Node 26 (Jeffnode 30) Time 0.12514 Node 27 (Jeffnode 29) Time Node 28 (Jeffnode 28) Time Node 29 (Jeffnode 27) Time Node 30 (Jeffnode 26) Time Node 31 (Jeffnode 25) Time 0.04310 0.01485 0.03068 0.11891 0.26574
 Node
 31
 (Jeffnode
 25)
 lime
 0.265/4

 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 = 1.269099 0.115078 -1.069908 0.144685 0.167535 -0.623150 1.404760 0.020583 0.916171 0.009053 0.096196 0.359419 -0.901186 0.040054 0.024405 0.836727 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

- 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

- 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_EU340979, (((((((2_Amazilia_brevirostris_KP722041, 5_Amazilia_versicolor_KF624601) 28, (3_Amazilia_millerii_KP722042, 4_Amazilia_rondoniae KP722041) 29, 7, 16 Hylocharis cynaus KJ619566) 26, (7 Archilochus_colubris_EF523925, 8_Calliphlox_amethystina_KP853095) 30) 25, (15 Heliodoxa_aurescens_KP853094, (17_Lophornis_magnificus_KT265276, 18_0reotorchilus_melanogaster_K5619587) 32) 31) 24, 10 Chrysolampis_mosquitus_KJ619585) 23 , (12_Florisuga_fusc_K7865966, 13_Florisuga_mellivora_KT619588) 34 , (14_Glaucis_hirsutus_KT265275, 10 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。



(Jeff	Node is for Th	orne's mult	tidivtime. 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

12 +-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

13 り、おかしな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

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す。
34 す。

3 4

5

6 7

8

13 14

26

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 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
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 version 1.4.4 です)
Tracer
https://github.com/beast-dev/tracer/releases
(2022 年 2 月 7 日現在最新版は version 1.7.2 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
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
しょう。
https://www.java.com/ja/
9 マントロールファイルの頃年
4. ユノトロールノナイルの棚朱

33 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
 34 います。

```
seed =
               seqfile = examples/DatingSoftBound/mtCDNApri123.txt
               treefile = examples/DatingSoftBound/mtCDNApri.trees
               outfile = out
                 ndata = 3
               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.</pre>
                 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
pa gamma = 6 2 * gamma prior for kappa
            kappa gamma = 62
            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
        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
 4
 5
        ここでは
 6
 7
 8
                                           seqfile = hummingbird 12mtCDS.fas
 9
                                                        treefile = MLrt.nwk
10
                                                  outfile = hummingbird.out
11
12
         としておきたいと思います。
13
                    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.
14
        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
15
        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
16
        インメント)のみのデータなので
17
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出来ます。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	ことになります。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	

	<pre>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</pre>
1	
2	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
3	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
4	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
5	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 モデルを用いたいので
6	
7	model = 7
8	alpha = 0.2
9	ncatG = 5
10	
11	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
12	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
13	ラメータです。
14	
	BDparas = 1 1 0 * birth, death, sampling kappa_gamma = 6 2 * gamma prior for kappa alpha gamma = 1 1 * gamma prior for alpha
15	
16	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)、標本推
17	出率(ρ)です(本文及びウェブ資料参照)。デフォルトの設定(110)では内部ノードの分布
18	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。
19	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
20	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
21	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
22	要ありません。
23	
24	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)
25	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。 <u>自分自身のデータ</u>
26	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。
27	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
28	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
29	$Γ$ 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
30	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
31	

 $\alpha = (m/s)^2$ $\beta = m/s^2$ 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 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 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 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 す。これで $\alpha = (m/s)^2 = 4$ $\beta = m/s^2 \approx 8.7$ となるので rgene_gamma = 4 8.7 になります。 幾何ブラウン運動の分散パラメータσ²の平均値は、客観的に設定することが難しい値です 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 ュアルで推奨されているように σ^2 の平均値×rttm=1 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 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 σ²の平均値を 1/0.68 ≈1.47 としましょう。 σ²の標準偏差は σ²の平均値と等しいと仮定します。 $\alpha = (m/s)^2 = 1$ $\beta = m/s^2 \approx 0.68$ となるので sigma2_gamma = 1 0.68 になります。

1	
2	
3	burnin = 2000 sampfreq = 2 nsample = 20000
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	
15	burnin = 50000
16	sampfreq = 50
17	nsample = 20000
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	① MLrt.nwk をアキストエティターで開き、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	10 1
25 26	
20 27	(Aegotheles_cristatus_EU344979,(((((((Amazilia_brevirostris_KP722043,Amazilia_versico
21 28	or_KF024001), (Amazina_millern_KF722042, Amazina_rondoniae_KF722041)), Hylocharis
20 20	
2)	(Therodoxa_autescens_K1 055094, Lophornis_magnificus_K1205270, Oreotrocinus_metan
30	orisuga mellivora KI619588) (Glaucis hirsutus KT265275 Phaethornis malaris KP85309
32	7))) ((Apus apus NC 008540 Chaetura pelagica KT809406) Cypseloides fumigatus KY6
33	88216))):
34	
35	②年代制約の条件

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
 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
 ることが出来ます。

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。

11

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。

23 24

- 25 2項. 最尤法による枝の長さと分散・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と分散・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 するの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。

32

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	usedata = 3
3	
4	を選択してください。
5	
6	2. 解析の実行
7	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
8	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。
9	
10	MCMCTREE プログラムは BASEML プログラムを起動し枝の長さと分散共分散を最尤推
11	定します。ここで MCMCTREE プログラムは、自動的に tmp0001.ctl, tmp0001.trees,
12	tmp001.txt というファイルを生成しますが、これらは BASEML プログラムが解析できる
13	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
14	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、
15	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
16	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
17	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
18	ル、シーケンスファイルは N セット分生成されます。
19	
20	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
21	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
22	報が記録されます。
23	
24	3頃、 階層ベイズ法による分岐年代推定
25	1. out.BV ファイルの編集
26	out.BV ファイルの名前を変更し in.BV にします。
27	
28	2. コントロールノアイルの編集
29	
30 21	usedata = 1 * 0; no data; 1:sed 11ke; 2:use 1n.bv; 5: out.bv
33	に変更を加えます。
32	usedata = 2
34	usevala - 2
35	を選択してください。
36	

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充分に収束しているか確認
8 しましょう。

9

12

- 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
 - 📙 | 🛃 📒 🖛 | Tracer v1.7.1 7ァイル ホーム 共有 表示 2 クリップボード 整理 新規 開く 選択 ✓ ♂ Tracer v1.... ,0 👃 ダウンロード * ^ 名前 更新日時 種類 サイズ F#1X2F * lib README.txt Tracer v1.7.1.exe 2018/06/14 10:08 ファイル フォルダー 2018/06/14 10:08 テキスト ドキュメント ー ビクチャ 2018/06/14 10:08 2018/06/14 10:08 10 KB AdmixGraph アプリケーション 4.481 KB PCA seal selection len OneDrive PC Apple iPhone ↓ ダウンロード
 ■ デスクトップ 🔮 ドキュメント 📰 ピクチャ 📕 ビデオ __ ♪ ミュージック 🏰 OS (C:) 🧭 DVD RW ドライブ (D:) NS 🚔 ボリューム (E:) ~ 3 個の項目
- 11 ルクリックで起動します。

- 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。

ce Files:				I Estimates A Marginal Density	t-Marginal dat Trace	
ace File	States	Burn-In				
files loaded			10	Summary Statistic	-	
			Reload			
ces:						
atistic	Mean	ESS	- 1			
			^			

2

3 こちらが TRACER が mcmc.txt を読み込んだ状態です。



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充分に収

- 6 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7 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 ア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。

アイル ホーム 共利 ★ □ □ □ □ イックアクセ コピー 貼りれ なにピン留め クリッ	■ 表示 】 よ切 [●] パ [●] ♪ パボード	た り取り スのコピー ョートカットの貼り付け	11、一先 前除 翌理	名前の 変更	■ 100 million mi	_い項目 ▼ Iートカット ▼	プロパティ プロパティ ☆ 履歴 開く	 すべて選択 選択解除 選択の切り 選択 	春え		^
÷ → ∽ ↑ 🔒 >	PC > OS	(C:) > ユーザー > Yonezaw	a » デスクトップ »	EvolToo	→ FigTree v1.4.	4			5 V	FigTree v1	م
🖊 ダウンロード	* ^	名前 ^		更亲	行時	種類	サイズ				
🔮 ドキュメント	*	, lib		201	9/07/01 17:40	ファイルフォ	<i>A</i> −				
■ ピクチャ	*	carnivore.tree		201	8/11/25 23:39	TREE 774	μ I	68 KB			
AdmixGraph		👹 FigTree v1.4.4.exe		201	8/11/25 23:39	アプリケーシ	ヨン	57 KB			
PCA		influenza.tree		201	8/11/25 23:39	TREE ファイ	ル 5	25 KB			
seal		README.txt		201	8/11/25 23:39	テキスト ドキ	コメント	13 KB			
selection											
loneDrive											
💻 PC											
🧊 3D オブジェクト											
Apple iPhone											
🖊 ダウンロード											
🛄 デスクトップ											
🗎 ドキュメント											
📰 ピクチャ											
📕 ビデオ											
 ♪ ミュージック											
🔛 OS (C:)											
🧭 DVD RW ドライブ (I) NS										
🚔 ボリューム (E:)											
個の項目	~										

10 そうすると以下のような画面が立ち上がります。

	E GE	9 🗶 💻	Node Clade	Taxa		Q- Filter	
artoon Collapse Rer	oot Kotate An	notate Colour Hilight	Find Selection Mode		Frev/hext		
 Layout 	4						
Ce 1	~						
Zoom:							
Evenneion							
Expansion							
Fish Eye:							
Root Length:							
-							
Curvature:							
Align Tip Labels							
Current Tree	P						
	~						
	4						
Trees	AF						
Time Scale	*						
Ip Labels	*						
Inp Shapes	*						
Node Labels	Se .						
Node Shapes	R						
Node Bars	R						
Branch Labels	R						
▶ 🗹 Scale Bar	R.						
Scale Axis	R						
Legend	R						
Legend	X						
Legend	×						
Legend	*						
Legend							
Legend	¥						
Egend	¥	v <					
Legend	<i></i>	v <					
Legend	*	v					
Legend	*	v c					
Logend	*	v					
▶ Legend	*	v c					
▶ Legend FigTree v1.44 iie Edit free Help Man		v «					
FigTre v1.44 ite Edit Tree Help New Open	Ctrl+N Ctrl+O	v (c					
☐ FigTree v1.44 ile Edit Tree Help New Open_ Save	Ctrl+N Ctrl+O Ctrl+S	v Kieht Fred					
▶ Legend ↓ Legend ↓ EgiTree v1.4.4 ↓ Legend ↓ EgiTree v1.4.4 ↓ Legend ↓ Lege	Ctrl+N Ctrl+O Ctrl+S	v Kient Find					
Ingree v1.4.4 FigTree v1.4.4 FigTree v1.4.4 New Open Swe A Swe A Swe A Swe A	Ctrl+N Ctrl+O Ctrl+S Ctrl+I	v Kiefer Fred					
FigTrev1.44 FigTr	Ctrl+N Ctrl+O Ctrl+S Ctrl+I	v I C					
FigTree v1.44 FigTree v1.44 FigTre	Ctri+N Ctri+O Ctri+C Ctri+E Ctri+E Ctri+E	v Kileh Fred					
▶ Legend Fighter v1.4.4 ite Edit Tree Help New Open Sive As Import Colour Scheme Export Trees Export PDF Export PDF Export PDF	Ctrl+N Ctrl+O Ctrl+O Ctrl+C Ctrl+I Ctrl+Alt+E	a Hildh Fire					
FigTree v1.4.4 FigTree v1.4.4 FigTree v1.4.4 New New New New Save A Import Annotations Import Annotations Export PRG Export PRG Export PRG	Ctrl+N Ctrl+O Ctrl+S Ctrl+L Ctrl+E Ctrl+Alt+E	r Hileht Find					
Image: IngTree v1.44 Image: Image	Ctrl+N Ctrl+O Ctrl+O Ctrl+C Ctrl+C Ctrl+E Ctrl+Alt+E	x Hileht Find					
▶ Leeend FigTree v1.4.4 Inter Edit Tree Help New Open Save As Import Colour Scheme Export PDF Export	Ctrl+N Ctrl+O Ctrl+S Ctrl+E Ctrl+Rt+E Ctrl+Rt+E	v Kiehr Fred					
Ideend FigTree v1.4.4	Ctrl+N Ctrl+O Ctrl+O Ctrl+C Ctrl+I Ctrl+I Ctrl+I Ctrl+I Ctrl+P	a Hildt Fro					
	Ctrl+N Ctrl+Q Ctrl+Q Ctrl+Q Ctrl+L Ctrl+Alt+E Ctrl+Alt+E Ctrl+Alt+E	V C					
	Ctrl+N Ctrl+O Ctrl+C Ctrl+E Ctrl+Alh+E Ctrl+P Ctrl+P	x Hileht Find					
	Ctrl+N Ctrl+O Ctrl+S Ctrl+I Ctrl+E Ctrl+Atr+E Ctrl+Atr+E	- Hight Free					

- 4 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、
- 5 MCMCTREE によって生成された **FigTree.tre** を選択して開きましょう。



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 3 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 4 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

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。 6



- 4 謝辞
- 5 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
- 6 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	holp in のフェルダの中に、大別販べ用いたのとす。たく同じ恐宅のファイルなりわてた
2	Theip.zipのノオルクの中に、本内越で用いたのとようたく同じ設たのノナイルを入れてお きました DAMLの敏振がらまくいかたい場合。これたのファイルを参考にしてご自身の
3 4	さました。FAIVILの脾机かりまくいかない場合、これらのノナイルを参考にしてこ日牙の ファイルの乳空な確認していただければ、個油等な目いだけても用います
4	ノナイルの設定を確認していたたければ、
5	holp -in を留声ナスレームへのフェルグが入っています。これとは第二音。第三音のタム
0	Neip.zip を解決すると、 0 つのフォルタか入っています。これらは第一早~第三早の谷々
/ 0	の時村に対応していますので、
0	第二音の第・潮田氏の推定
9	第二早 2 即・選び江の祖足 1 頂 は た デ ル な 田 い な 選切 耳 の 推 空
10	
11	Branch_model というノオルタの中に
12	branch2w.ctl $(\neg \gamma + \neg - \mu \gamma + 1 \mu)$
15	Dranch W.cti (コントロールファイル)
14	ZWMELT.NWK (個化ノアイル)
10	か入っています。 ここに DAML の and and プログラノト humminghind 12mtCDC for なしれていたざい
10	ここに PAML の codemi.exe ノロクノムと nummingDird_12mtcDS.tas を入れていたたい
17	にうえで、コマントノロンノトで以下のコマントを使って美行しましょう
10	
19 20	codemi brancnzw.cti
20 21	帰無仮説を封したい提合け
21	市
22	を伸い以下のコマンドを伸って実行しましょう
$\frac{23}{24}$	
25	codemi branch1w cti
25 26	
27	
28	
29	2 項 サイトチデルを用いた選択圧の推定
30	Site model k which k with k with k with k with k with k with k with k with k with k with k with k
31	site ctl $(\exists \forall h = h \forall z \neq \lambda h)$
32	MI 15sp pwk (樹形ファイル)
33	humminghird 12mtCDS15sp fas $(T \neg A \lor A$
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
00	

1	
2	codeml site.ctl
3	
4	
5	3項. 枝サイトモデルを用いた選択圧の推定
6	Branchsite_model というフォルダの中に
7	branchsite.ctl (コントロールファイル)
8	branchsitenull.ctl (コントロールファイル)
9	2wML2.nwk (樹形ファイル)
10	が入っています。
11	ここに PAML の codeml.exe プログラムと hummingbird_12mtCDS.fas を入れていただい
12	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
13	
14	codeml branchsite.ctl
15	
16	帰無仮説を試したい場合は
17	branchsitenull.ctl (コントロールファイル)
18	を使い以下のコマンドを使って実行しましょう
19	
20	codeml branchsitenull.ctl
21	
22	
23	
24	第三章2節:分子時計を仮定した分岐年代推定
25	2項:分子時計を仮定しないモデルによる尤度推定
26	Noclock_model というフォルダの中に
27	noclock.ctl (コントロールファイル)
28	ML.nwk (樹形ファイル)
29	が入っています。
30	ここに PAML の baseml.exe プログラムと hummingbird_12mtCDS.fas を入れていただい
31	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
32	
33	basemi noclock.cti
34	
35	3項:分子時計を仮定したモデルによる尤度推定
36	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	basemi clock.cti
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	このファイルの名削を in.BV と変更したうえで、 mcmctree.ctl の usedata オフションを
27	
28	usedata = 2
29 20	に亦再しましょう。そのうえで、コマンドプロンプトで川下のコマンドを使って実行しま
30 21	
32	
32	memetree
34	
35	
55	