

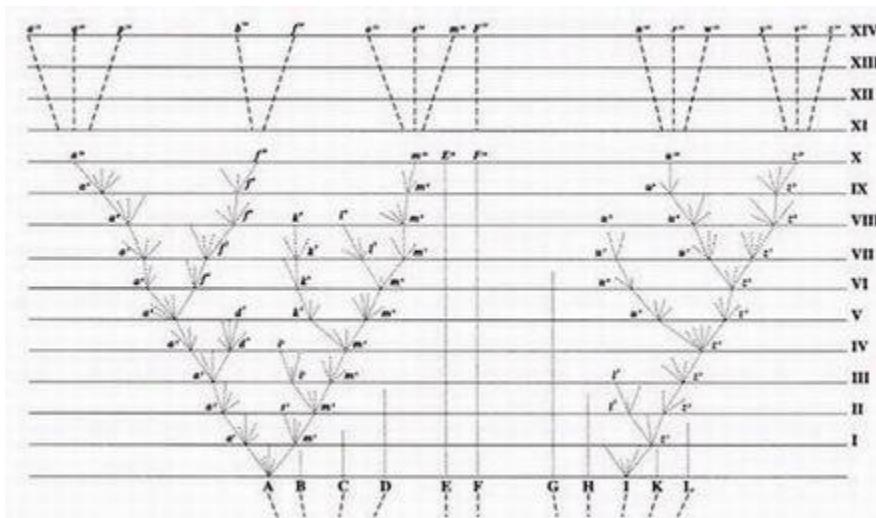
進化の解析

清水謙多郎

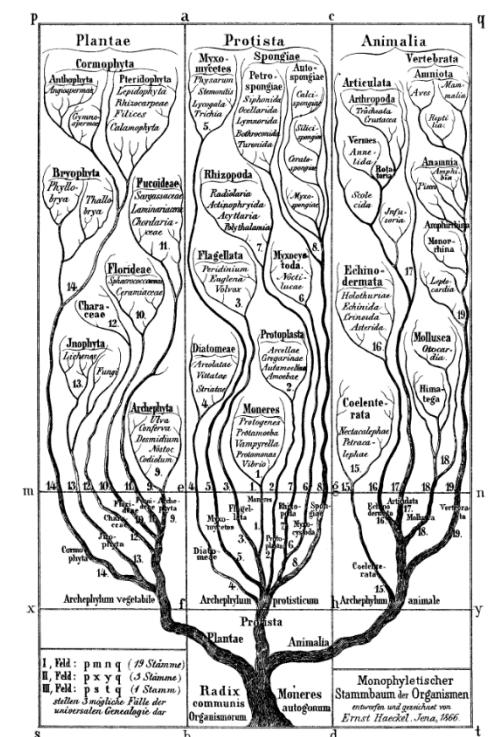
shimizuk@fc.jwu.ac.jp

進化系統樹

- 進化系統樹 (phylogenetic tree)
 - 生物の進化の系統関係を木構造で表したもの
 - 全生物は共通祖先から進化したという前提
- 形態的な特徴からの推定
 - 化石情報、形態学的形質
- 配列データからの推定 → **分子系統樹**
 - 塩基配列やアミノ酸配列などから推定



Charles Darwinの「種の起源」に描かれた系統樹

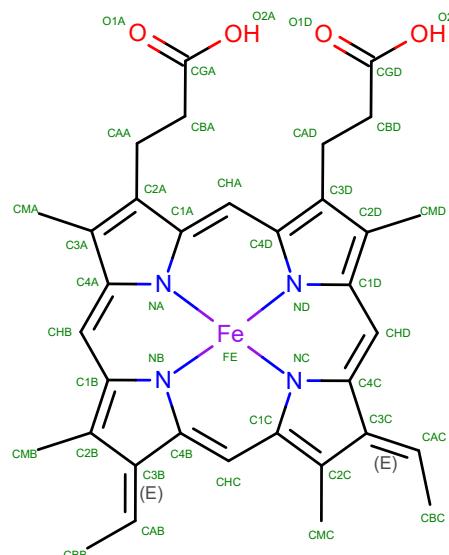
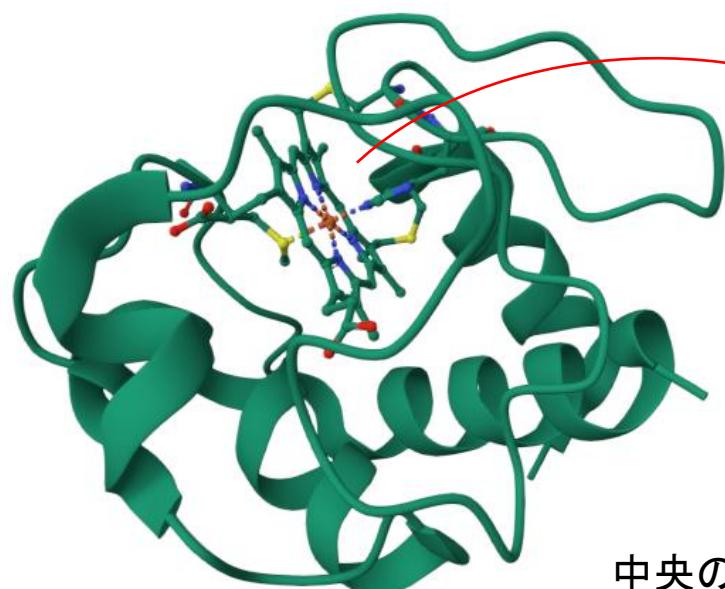


Ernst Haeckelによる系統樹

配列を比べてわかること

- 同じタンパク質の配列が生物種によってどのように違うかみてみよう
- シトクロムC: 細胞のミトコンドリアにあって、呼吸に重要な役割をもつタンパク質
- 多くの生物種に存在するので、並べて比較することができる

ヘムの構造



電子の受け渡しを行
うのに重要

中央の鉄は、 Fe^{2+} （還元状態）と Fe^{3+} （酸化状態）の間で可逆的に変化し、これにより、電子を受け取ったり渡したりすることができる

配列を並べてわかること

シトクロムCの配列

ヒト >cytochromeC:Homo sapiens

MGDVEKGKKIFIMKCSQCHTVEKGGKHTGPNLHGLFGRKTGQAPGYSYTAANKNKGIIWGEDTLMEYLENPKKYIPGTMIFVGICKKEERADLIAYLKKATNE

チンパンジー

>cytochromeC:Pan troglodytes

GDVEKGKKIFIMKCSQCHTVEKGGKHTGPNLHGLFGRKTGQAPGYSYTAANKNKGIIWGEDTLMEYLENPKKYIPGTMIFVGICKKEERADLIAYLKKATNE

マウス

>cytochromeC:Mus musculus

MGDVEKGKKIFVQKCAQCHTVEKGGKHTGPNLHGLFGRKTGQAAGFSYTDANKNGITWGEDTLMEYLENPKKYIPGTMIFAGIKKKGERADLIAYLKKATNE

ミツバチ

>cytochromeC:Apis

MGIPAGDPEKGKKIFVQKCAQCHTIESGGKHKGPNLYGVYGRKTGQAPGYSYTDANKKGITWNKETLFYELENPKKYIPGTMVFAGLKKPQERADLIAYIEQA

イネ

>cytochromeC:Oryza sativa

MASFSEAPPGNPKAGEKIFKTKCAQCHTVDKGAGHKQGPNLNGLFGRQSGTTPGYSYSTANKMAVIWEENTLYDYLLNPKKYIPGTMVFPLKKPQERADLISYI

ジャガイモ

>cytochromeC:S.Tuberosum

ASFGEAPPGNPKAGEKIFKTKCAQCHTVDKGAGHKQGPNLNGLFGRQSGTTAGSYSNSANKMAVTWGENTLYDYLLNPKKYIPGTMVFPLKKPQERADLIAYL

酵母

>cytochromeC:S.Cerevisiae

MTEFKAGSAKKGATLFKTRCLQCHTVEKGGPHKVGPNLHGIFGRHSGQAEGYSYTDANIKNVLDENNMSEYLTPKKYIPGTMMAFGGLKEKDRNDLITYLKKA

テトラヒメナ

>cytochromeC:tetrahymena

GDASAGRDIIFDSQCSACHAIEGDSTAAPVLGGVIGRKAGQEKFAYSKGKGSGITWNEDTLMEYLENPKKYIPGTMIFVGICKKEERADLIAYLK

発しんチフス

リケッチャ

>cytochromeC:RP253

REIAKKCLMCHSLDKDGPNKLGPHLWNIVGRP KASITDYKYSFAISKLGGVWDDENLFAFLHKPSSYAPGTMFSAGISKPQDIADVILFLKNYV

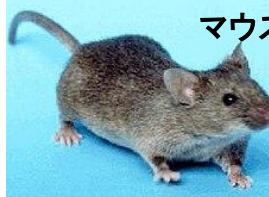
ヒト



チンパンジー



マウス



ミツバチ



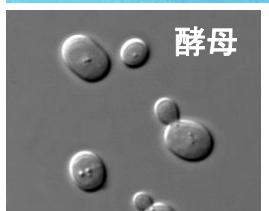
イネ



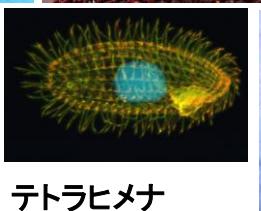
ジャガイモ



酵母



テトラヒメナ



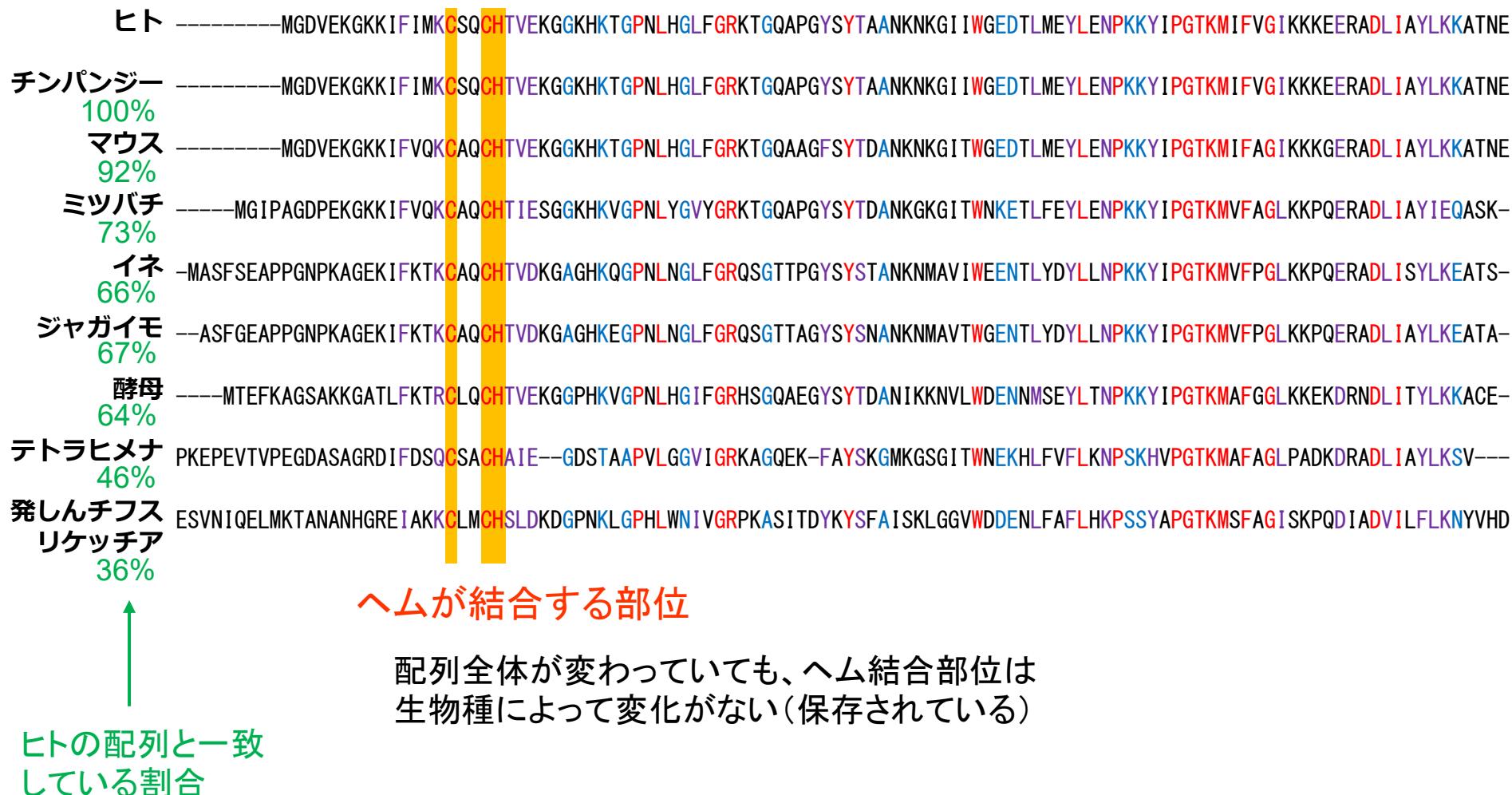
リケッチャ



すべてWikipediaより

配列を並べてわかること

シトクロムCの配列のアラインメント



配列データによる系統解析

- 分子進化
 - DNA、そこから変換されるRNAやタンパク質の配列が、長期間にわたる突然変異の蓄積によって変化していくこと
- 分子時計
 - 分子進化の速度が一定であること
 - 配列の変異（塩基、アミノ酸の置換）が、系統樹のどの場所においても同じ割合で起こる
 - 系統解析が容易
 - 分子時計が成り立たない例は多数見つかっている
 - 系統によって分子進化の速度が異なる
 - いつの時代でも進化の速度が一定とは限らない
 - 1962年、Linus Carl Pauling、Émile Zuckerkandlが提唱
- 中立進化説
 - DNAやタンパク質の配列が進化の過程で受ける変異の多くは、自然選択の上からは有利でも不利でもない「中立的」なもの
 - 1968年、木村資生きむら もとおが提唱、分子時計が成立することの論拠となる

分子時計の例

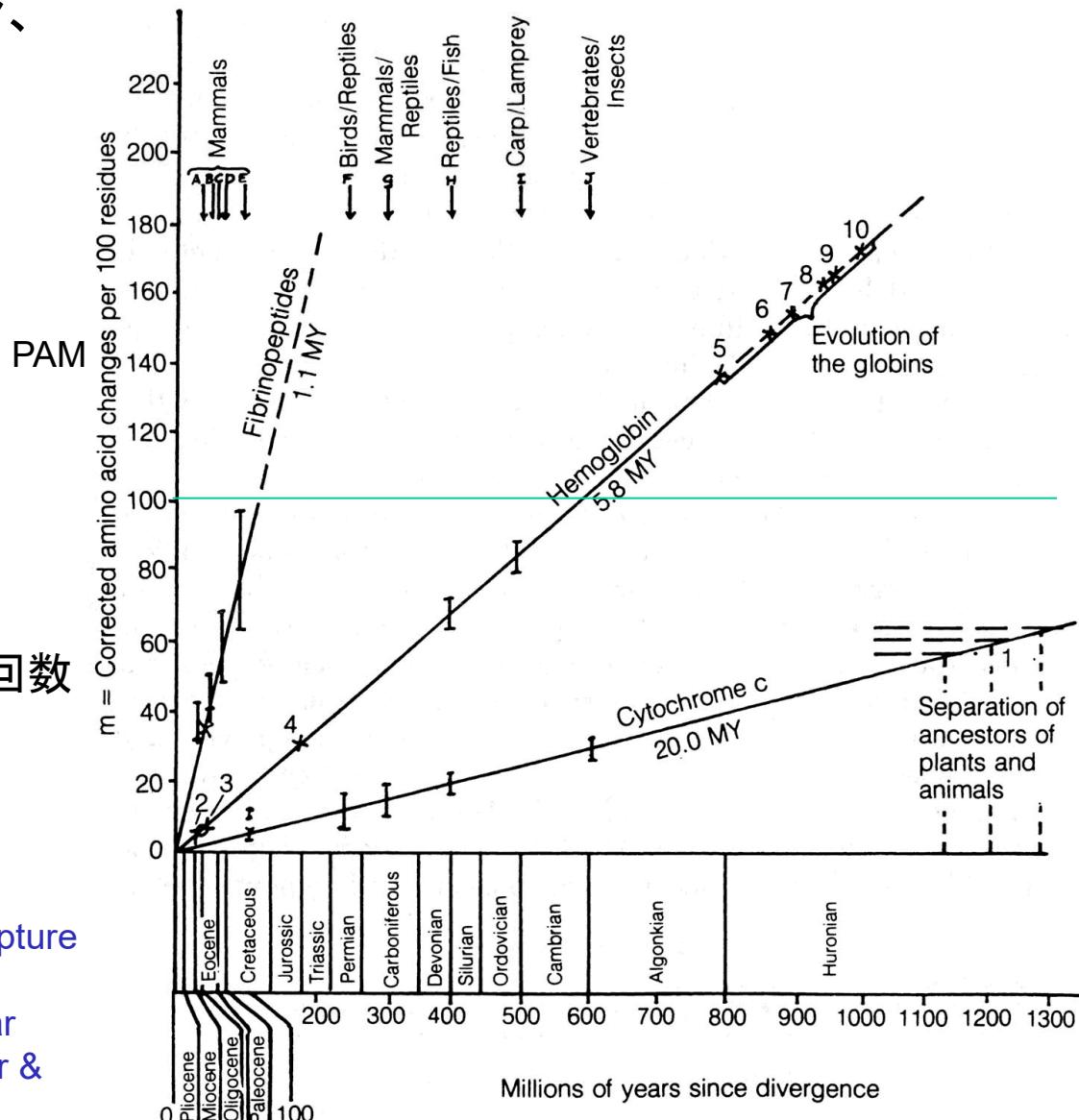
フィブリノペプチド、ヘモグロビン、シトクロームCのアミノ酸置換率

PAM: Point Accepted Mutation,
受容点突然変異

100アミノ酸あたり1個の突然変異が起きるという進化上の時間の単位

進化距離 = サイトあたりの置換の回数
進化速度 = 進化距離 / 時間

Pattle P. T. Pun. Evolution: nature and scripture in conflict?, 1982. Interdisciplinary Biblical Research Institute, 原図はNei M. Molecular population genetics and evolution, Elsevier & North Holland, 1975.



系統樹を描いてみよう

- いろいろな生物種のシトクロムCの配列
 - [cytoC.fasta](#)

ヒト	>cytoC Homo_sapiens MGDVERGKKIFIMKCSQCHTVEKGGKHKTGPNLHGLFGRKTGQAPGYSYTAANKNGIIWGEDTLMEYLENPKKYIPGTMIFVGIKKE
チンパンジー	>cytoC Pan_troglodytes GDVEKGKKIFIMKCSQCHTVEKGGKHKTGPNLHGLFGRKTGQAPGYSYTAANKNGIIWGEDTLMEYLENPKKYIPGTMIFVGIKKE
マウス	>cytoC Mus_musculus MGDVERGKKIFVQKCAQCHTVEKGGKHKTGPNLHGLFGRKTGQAAGFSYTDANKNGITWGEDTLMEYLENPKKYIPGTMIFAGIKE
ミツバチ	>cytoC Apis MGIPAGDPEKGKKIFVQKCAQCHTIESGGKHKGPNLYGVYGRKTGQAPGYSYTDANKKGITWNKETLFEYLENPKKYIPGTMVFAGIKE
イネ	>cytoC Oryza_sativa MASFSEAPPGNPKAGEKIFKTKCAQCHTVDKGAGHKQGPNLNGLFGRQSGTTPGYSYSTANKNMAVIWEENTLYDYLLNPKKYIPGTMVFAGIKE
ジャガイモ	>cytoC S.Tuberosum ASFGEAPPGNPKAGEKIFKTKCAQCHTVDKGAGHKEGPNIQNLFGRQSGTTAGYSYSNANKNMAVTWGENTLYDYLLNPKKYIPGTMVFAGIKE
酵母	>cytoC S.Cerevisiae MTEFKAGSAKKGATLFKTRCLQCHTVEKGGPHKVGPNLHGIFGRHSGQAEGYSYTDANIKNVLDENNMSYLTNPKKYIPGTMVFAGIKE
テトラヒメナ	>cytoC Tetrahymena GDASAGRDIIFDSQCSACHAIEGDSTAAPVLGGVIGRKAGQEKFAYSKGMKGSGITWNEDTLMEYLENPKKYIPGTMIFVGIKKE
発しんチフス リケッチャア	>cytoC RP253 REIAKKCLMCHSLDKDGPNLGPHLWNIVGRPKASITDYKYSFAISKLGGVWDDENLFAFLHKPSSYAPGTMFSAGISKPQDIADV

これらの配列を入力とする

系統樹を描いてみよう

- Clustal Wを使う
 - <https://www.genome.jp/tools-bin/clustalw>

Multiple Sequence Alignment by CLUSTALW

General Setting Parameters:

Output Format: CLUSTAL

Pairwise Alignment: SLOW/ACCURATE (highlighted with a red circle)

Enter your sequences (with labels) below (copy & paste): PROTEIN (highlighted with a red circle)

Support Formats: FASTA (Pearson), NBRF/PIR, EMBL/Swiss Prot, GDE, CLUSTAL, and GCG/MSF

```
>cytoC|Tetrahymena  
GDASAGRDIFDSCSACHAIEGDTSTAAPVLGGVIGRKAGQEKFAYSKGMKGSGITWNEDT  
LMEYLENPKKYIPGTKMIFVGIGKKEERADLIAYLK  
>cytoC|RP253  
REIAKKCLMCHSLDKDGPNKLGPHLWNIVGRPKASITDYKYSFAISKLGGVWDDENLFAF  
LHKPSSYAPGTKMSFAGISKPQDIADVILFLKNYY
```

Or give the file name containing your query
参照... ファイルが選択されていません。

Execute Multiple Alignment | Reset (highlighted with a red circle)

More Detail Parameters...

Pairwise Alignment Parameters:

For FAST/APPROXIMATE:

K-tuple(word) size: 1, Window size: 5, Gap Penalty: 3

Number of Top Diagonals: 5, Scoring Method: PERCENT

系統樹を描いてみよう

CLUSTALW Result

[clustalw.aln][clustalw.dnd][readme]

Select tree menu Exec

CLUSTAL 2.1 Multiple Sequence Alignments

```
Sequence type explicitly set to Protein
Sequence format is Pearson
Sequence 1: cytoC|Homo_sapiens      105 aa
Sequence 2: cytoC|Pan_troglodytes   104 aa
Sequence 3: cytoC|Mus_musculus     105 aa
Sequence 4: cytoC|Apis             108 aa
Sequence 5: cytoC|Oryza_sativa    112 aa
Sequence 6: cytoC|S.Tuberosum     111 aa
Sequence 7: cytoC|S.Cerevisiae   109 aa
Sequence 8: cytoC|Tetrahymena    96 aa
Sequence 9: cytoC|RP253          95 aa
Start of Pairwise alignments
Aligning...
```

入力した内容

```
Sequences (1:2) Aligned. Score: 100
Sequences (1:3) Aligned. Score: 91.4286
Sequences (1:4) Aligned. Score: 72.381
Sequences (1:5) Aligned. Score: 64.7619
Sequences (1:6) Aligned. Score: 64.7619
Sequences (1:7) Aligned. Score: 61.9048
Sequences (1:8) Aligned. Score: 59.375
Sequences (1:9) Aligned. Score: 36.8421
Sequences (2:3) Aligned. Score: 91.3462
Sequences (2:4) Aligned. Score: 73.0769
Sequences (2:5) Aligned. Score: 65.3846
Sequences (2:6) Aligned. Score: 65.3846
Sequences (2:7) Aligned. Score: 62.5
Sequences (2:8) Aligned. Score: 59.375
Sequences (2:9) Aligned. Score: 36.8421
Sequences (3:4) Aligned. Score: 76.1905
Sequences (3:5) Aligned. Score: 62.8571
Sequences (3:6) Aligned. Score: 66.6667
Sequences (3:7) Aligned. Score: 61.9048
Sequences (3:8) Aligned. Score: 58.3333
Sequences (3:9) Aligned. Score: 36.8421
Sequences (4:5) Aligned. Score: 60.1852
Sequences (4:6) Aligned. Score: 61.1111
Sequences (4:7) Aligned. Score: 56.4815
Sequences (4:8) Aligned. Score: 53.125
Sequences (4:9) Aligned. Score: 37.8947
```

マルチプルアラインメント(複数の配列
のアラインメント)構築の過程
→ 2本ずつのアラインメントのスコア

さらに下を見ると…

系統樹を描いてみよう

CLUSTAL 2.1 multiple sequence alignment

cytoC Oryza_sativa	MASFSEAPPGNPKAGEKIFKTCAQCHTVDKGAGHKQGPNLNGLFGRQSG
cytoC S.Tuberosum	-ASFGEAPPGNPKAGEKIFFKTCAQCHTVDKGAGHKQGPNLNGLFGRQSG
cytoC Homo_sapiens	-----MGDVEKGKKIFIMKCSQCHTVKEGGKHKTGPNLHGLFGRKKG
cytoC Pan_troglodytes	-----GDVEKGKKIFIMKCSQCHTVKEGGKHKTGPNLHGLFGRKKG
cytoC Mus_musculus	-----MGDVEKGKKIFVQKCAQCHTVKEGGKHKTGPNLHGLFGRKKG
cytoC Apis	----MGIIPAGDPEKGKKIFVQKCAQCHTVESGGKHVKVGPNLYGVYGRKG
cytoC Tetrahymena	----GDASAGRDFDSQCSACHAE--GDSTAAPVLLGGVIGRKAG
cytoC S.Cerevisiae	--MTEFKAGSAKKGATLFKTRCLQCHTVKEGGPHVKVGPNLHGIFGRHSG
cytoC RP253	-----REIAKKCLMCHSLDKDGPNKLGPHLWNIVGRPKA : : * ***: : . . . * .: ** .
cytoC Oryza_sativa	TPPGYSYSTANKMVAWIWEENTLYDYLLNPKKYIPGTTKMWFPGLKKPQER
cytoC S.Tuberosum	TTAGSYSNANKMVAWTWGENTLTYDYLNPKKYIPGTTKMWFPGLKKPQER
cytoC Homo_sapiens	QAPGYSYTAANKNGIWIWGEDTLMEYLENPKKYIPGTTKMFIVGIKKKER
cytoC Pan_troglodytes	QAPGYSYTAANKNGIWIWGEDTLMEYLENPKKYIPGTTKMFIVGIKKKER
cytoC Mus_musculus	QAAGFSYTDANKNGKGTIWGEDETLMEYLENPKKYIPGTTKMFIVGIKKKGER
cytoC Apis	QAPGYSYTDANKGKGTIWNEDTLMYELENPKKKYIPGTTKMFIVGIKKKER
cytoC Tetrahymena	QEK-FAYSKGMKGSGITWNEDTLMYELENPKKKYIPGTTKMFIVGIKKKER
cytoC S.Cerevisiae	QAEGFSYTDANIKNVLWDENNMSLEYLTNPKKYIPGTTKMAFFGLKKEKDR
cytoC RP253	SITDYKYSFAISKLGGWWDENLFAFLHKPSSYAPGTTKMSFAG1SKPQDI *: . * . .: * ; .. * **** * .: * :
cytoC Oryza_sativa	ADLISYLKEATS-
cytoC S.Tuberosum	ADLIAYLKEATA-
cytoC Homo_sapiens	ADLIAYLKKATNE
cytoC Pan_troglodytes	ADLIAYLKKATNE
cytoC Mus_musculus	ADLIAYLKKATNE
cytoC Apis	ADLIAYIEQASK-
cytoC Tetrahymena	ADLIAYLK----
cytoC S.Cerevisiae	NDLITYLKKACE-
cytoC RP253	ADVILFLKNVV-- *: .: :

clustalw.dnd

```
(  
(  
(  
(  
cytoC|Homo_sapiens:0.00153,  
cytoC|Pan_troglodytes:-0.00153)  
:0.04487,  
cytoC|Mus_musculus:0.04125)  
:0.06215,  
cytoC|Apis:0.15019)  
:0.00897,  
(  
(  
cytoC|Oryza_sativa:0.04091,  
cytoC|S.Tuberosum:0.03117)  
:0.16026,  
(  
cytoC|S.Cerevisiae:0.20907,  
cytoC|RP253:0.40146)  
:0.01998)  
:0.04706,  
cytoC|Tetrahymena:0.30344);
```

Select tree menu Exec

配列のマルチプルアラインメント

系統樹を描くための案内木(デンドログラム)「clustalw.dnd」

数字は推定される進化距離(進化的にどれだけ離れているか)を示す

系統樹を描いてみよう

結果を拡大したもの

clustalw.dnd

```
(  
(  
(  
(  
cyt BIONJ: 近隣結合法(NJ法)の改良  
cyt FastTree, PhyML, RAxML: 最尤法に  
:0 基づく手法  
:0.06215,  
cytC|Apis:0.15019)  
0.00000  
BIONJ
```

Select tree menu

FastTree

FastTree full

PhyML

PhyML bootstrap

RAxML

RAxML bootstrap

Select tree menu

0.04091,
.03117)

0.20907,
)

.30344);

Exec

クリックするとメニュー
が表示される

clustalw.dnd

```
(  
(  
(  
(  
cytC|Homo_sapiens:0.00153,  
cytC|Pan_troglodytes:-0.00153)  
:0.04487,  
cytC|Mus_musculus:0.04125)  
:0.06215,  
cytC|Apis:0.15019)  
0.00000  
BIONJ
```

Select tree menu

FastTree

FastTree full

PhyML

PhyML bootstrap

RAxML

RAxML bootstrap

BIONJ

分子系統樹の手法
を選択し、「Exec」
ボタンを押す

0.04091,
.03117)

0.20907,
)

.30344);

Exec

系統樹を描いてみよう



Phylogenetic analysis pipeline by ETE3

Workflow

none-none-none-bionj_default

Method

- Alignment and phylogenetic reconstructions were performed using the function "build" of ETE3 3.1.3 ([Huerta-Cepas et al., 2016](#)) as implemented on the GenomeNet (<https://www.genome.jp/tools/ete/>).
- User provided the multiple sequence alignment.
- A distance-based tree was inferred with the BioNJ algorithm ([Gascuel, 1997](#)) using PhyML v20160115 ([Guindon et al., 2010](#)) ran with model and parameters: --pinv e --alpha e --nclasses 4 -o lr -f m --bootstrap -2.
- Branch supports are the Chi2-based parametric values return by the approximate likelihood ratio test.

(These texts may be used for your publication.)

Result files

outTree_unrooted.nwk



outTree_midpointRooted.nwk



input.fa.final_tree.used_alg.fa



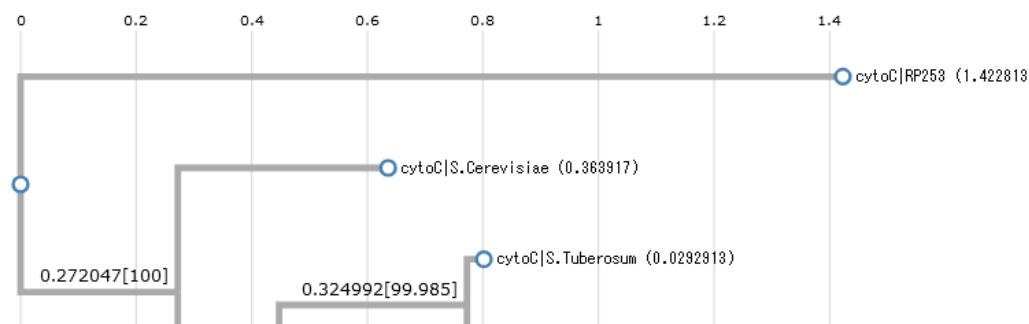
input.fa.final_tree.fa



Phylogram (midpoint rooted tree)

without branch length without branch length labels without leaf labels without ticks

JSON SVG PNG

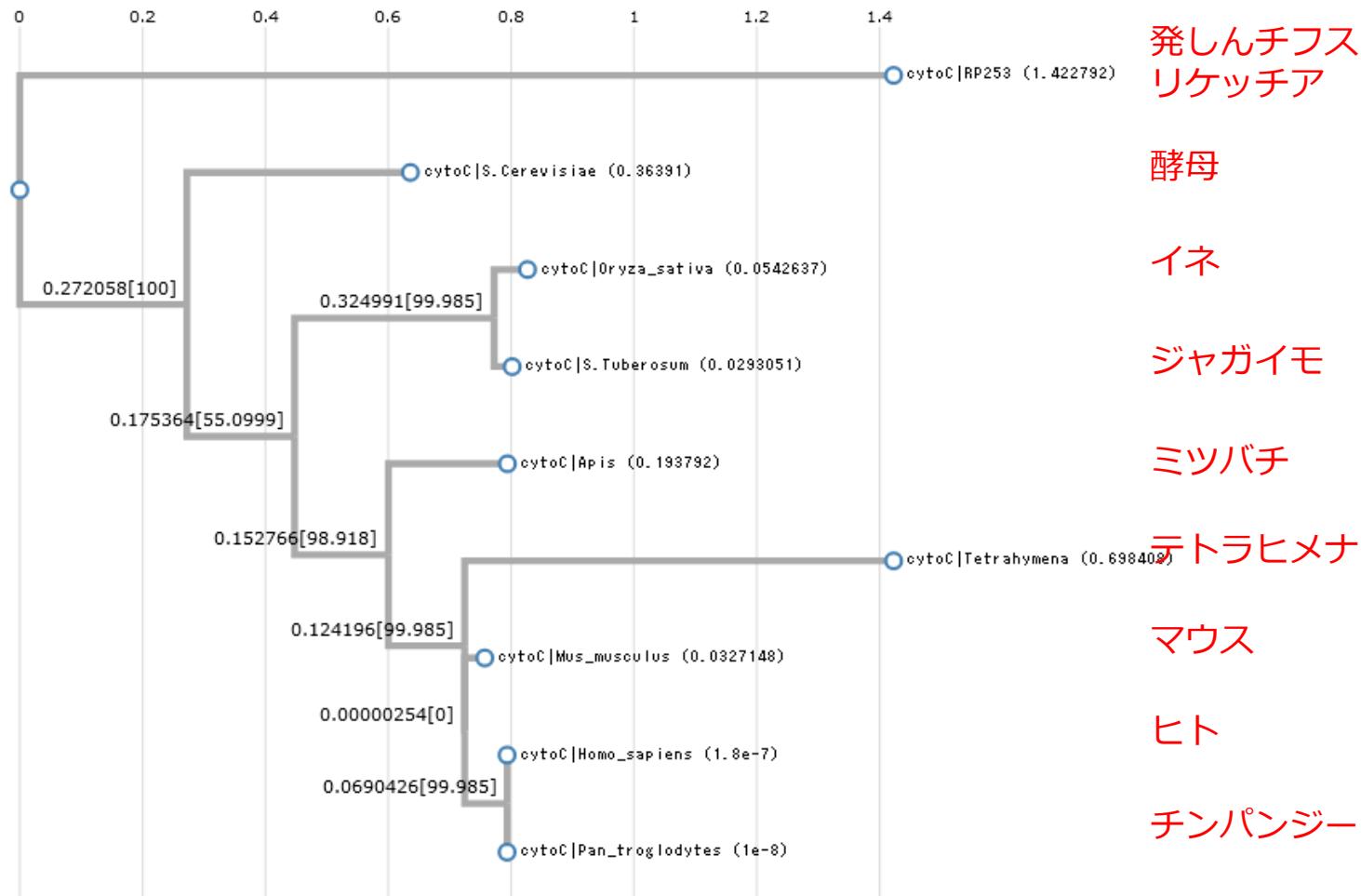


系統樹を描いてみよう

Phylogram (midpoint rooted tree)

without branch length without branch length labels without leaf labels without ticks

JSON  SVG  PNG 



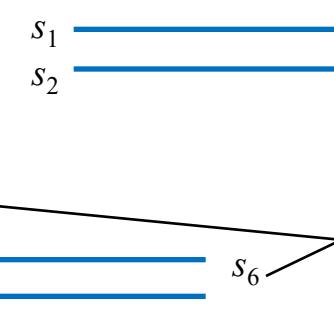
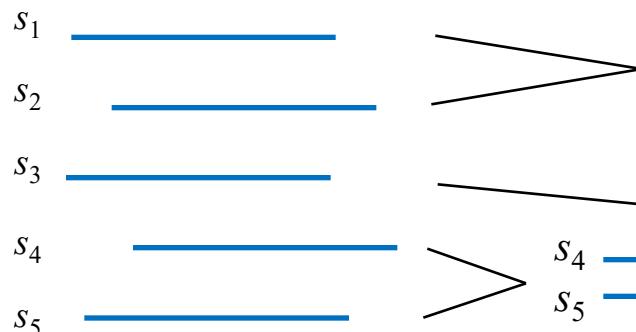
段階的なアラインメント

2本の配列のアラインメント（ペアワイズアラインメント）のスコア

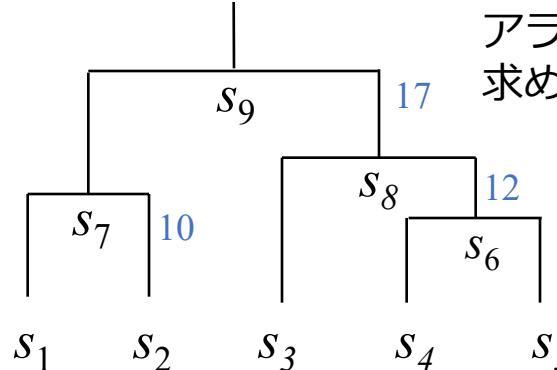
	s_1	s_2	s_3	s_4	s_5
s_1		10	18	24	20
s_2			12	18	14
s_3				14	10
s_4					8
s_5					

配列間の類似度スコアまたは
配列間の距離を計算

段階的アラインメント



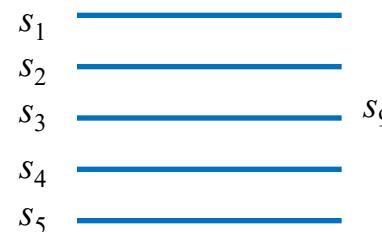
案内木（デンドログラム）の作成



アラインメントを段階的に
求める順序を決める



案内木に従ってアラインメントを構築



距離行列法

- 最初に配列間の距離を計算した後で、その距離情報のみから系統樹を推定する
- n 個の配列 s_1, s_2, \dots, s_n が与えられ、 $i < j$ なるすべての配列対 (s_i, s_j) について、その間の距離 d_{ij} を計算する
- $d_{ij} \geq 0$ 、 $d_{ij} = d_{ji}$ とする
- 主な手法
 - UPGMA法 (Unweighted Pair Group Method using Arithmetic averages)
 - 平均距離法のクラスタリングを適用
 - 1973年、Sneathらにより、形態的な類似度を距離として適用
 - 近隣結合法 (N-J法, Neighbor Joining method)

距離行列法

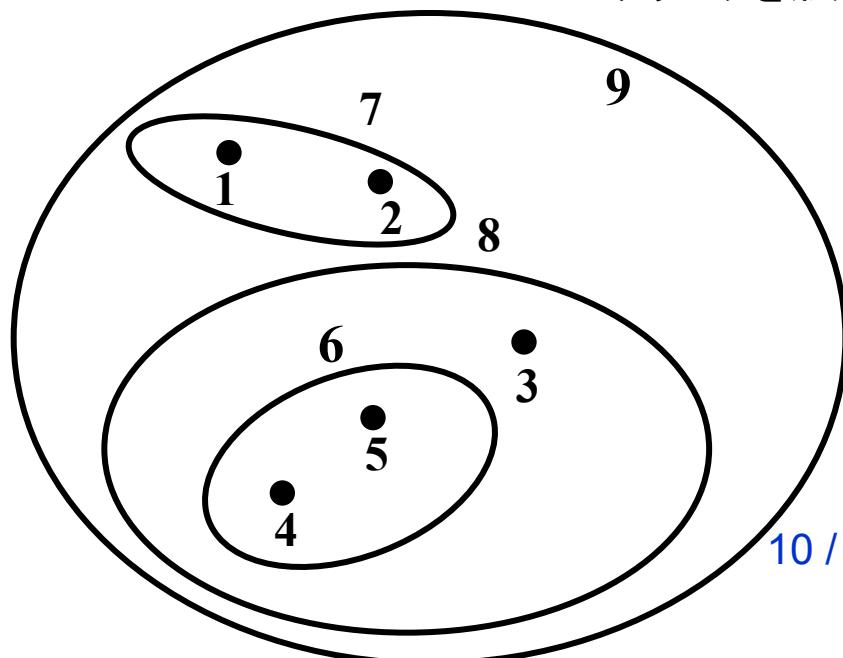
- 最初に配列間の距離を計算した後で、その距離情報のみから系統樹を推定
- 主な手法
 - UPGMA法** (Unweighted Pair Group Method using Arithmetic averages)
 - 近隣結合法** (N-J法, Neighbor Joining method)
- UPGMA法
 - 平均距離法のクラスタリングを適用
 - 分子時計が成り立つことを仮定
 - 1973年、Sneathらにより、形態的な類似度を距離として適用

案内木の作り方の例

配列	1	2	3	4	5
1		10	18	24	20
2			12	18	14
3				14	10
4					8
5					

距離行列

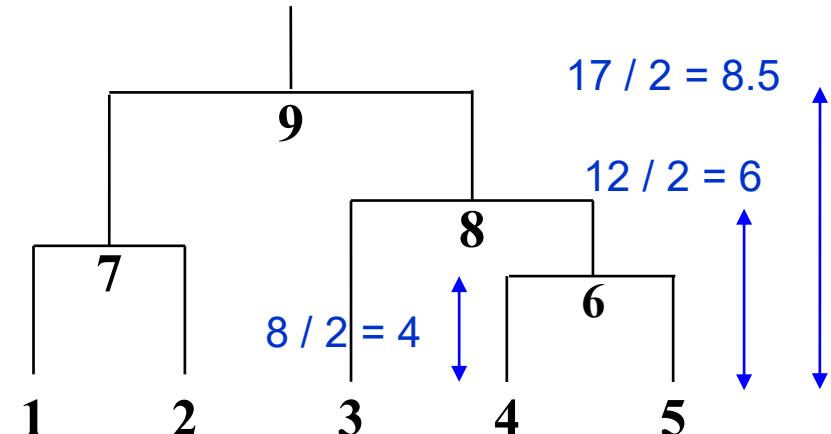
最小距離のものを選んでクラスタを形成



	1	2	3	6
1		10	18	22
2			12	16
3				12
6				

	7	3	6
7		15	19
3			12
6			

	7	8
7		17
8		



クラスタリングの考え方

- 基本的には、クラスタリング（クラスタ分析）の手法をとる
- 最初、すべての配列は、それぞれ1つのグループ（クラスタ）を形成しているとする
- すべてのグループ間の類似度を求め、最も似ているグループをまとめて、新しいグループを作る
- これを全体が1つのグループになるまで繰り返す
- グループをまとめていく過程を示したのが案内木

近隣結合法

- **近隣結合法 (N-J法, Neighbor Joining method)**
 - 近隣（1つのノードにつながる配列）の結合の各段階で、すべての枝の長さが最小になるように木を形成
 - 基本的には無根系統樹（根のない木）が形成される
 - 分子時計が成り立つことを仮定しない
 - 計算時間がそれほどかかるない
 - 1987年、斎藤成也と根井正利による
- アルゴリズムの主要な点
 - ノードの集合を L とし、以下を繰り返す
 - 距離 $D_{ij} = d_{ij} - (r_i + r_j)$ を最小化

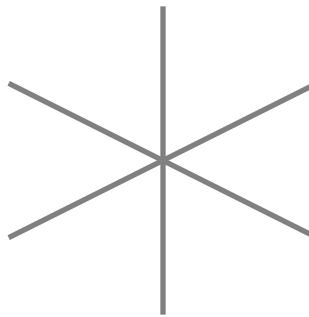
$$r_i = \frac{1}{|L|-2} \sum_{h \in L} d_{ih}, \quad r_j = \frac{1}{|L|-2} \sum_{h \in L} d_{hj}$$

- 各ステップでノード k を作成し、ノード i と j を結合する

$$d_{ik} = \frac{1}{2}(d_{ij} + r_i - r_j), \quad d_{jk} = \frac{1}{2}(d_{ij} - r_i + r_j)$$

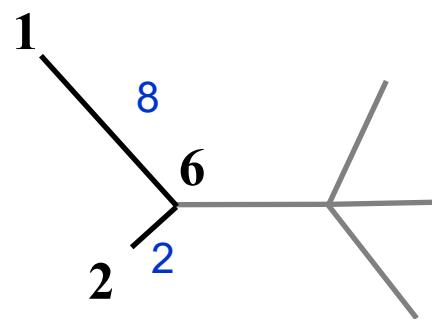
近隣結合法の実行例（1）

d_{ij}	1	2	3	4	5
1		10	18	24	20
2			12	18	14
3				14	10
4					8
5					
D_{ij}	1	2	3	4	5
1		-32	-24	-18	-22
2			-24	-18	-22
3				-25.3	-29.3
4					-30.7
5					
d_{ij}	6	3	4	5	
6		10	16	12	
3			14	10	
4				8	
5					



$$d_{16} = \frac{1}{2}(d_{12} + r_1 - r_2) = 8$$

$$d_{26} = \frac{1}{2}(d_{12} - r_1 + r_2) = 2$$



近隣結合法の実行例 (2)

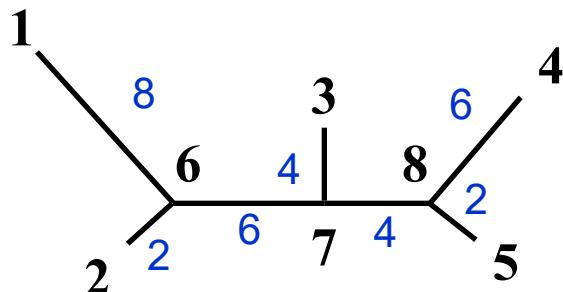
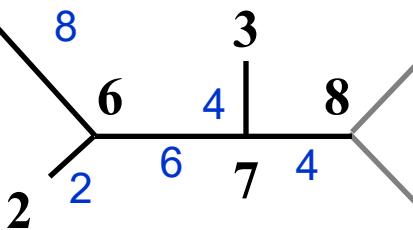
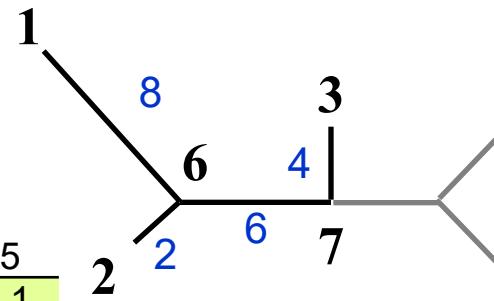
d_{ij}	6	3	4	5
6		10	16	12
3			14	10
4				8
5				

d_{ij}	7	4	5
7		10	6
4			8
5			

d_{ij}	8	4
8		6
4		

D_{ij}	6	3	4	5
6		-4	5	-1
3			1	0
4				-3
5				

D_{ij}	7	4	5
7		4	-4
4			-2
5			

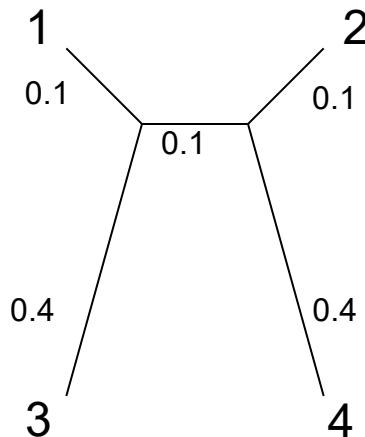


UPGMA法と近隣結合法の比較

分子時計(分子進化の速度が一定であること)に従わない系統樹

UPGMA法ではもとの系統樹を正しく推定できないが、
近隣結合法では正しく推定できる場合の例

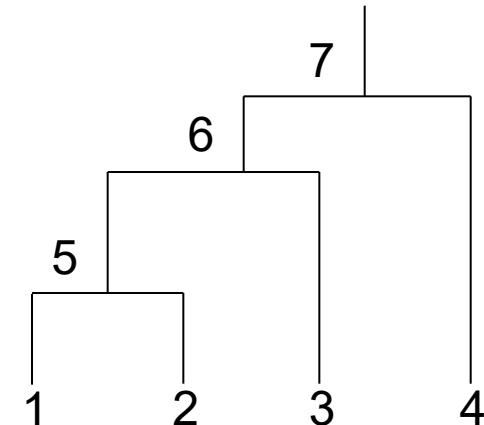
これが正しい系統樹と仮定



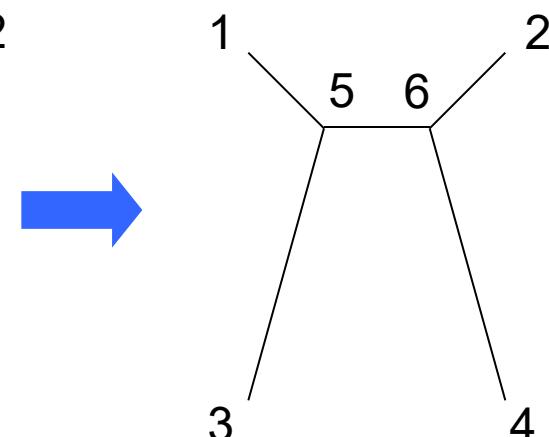
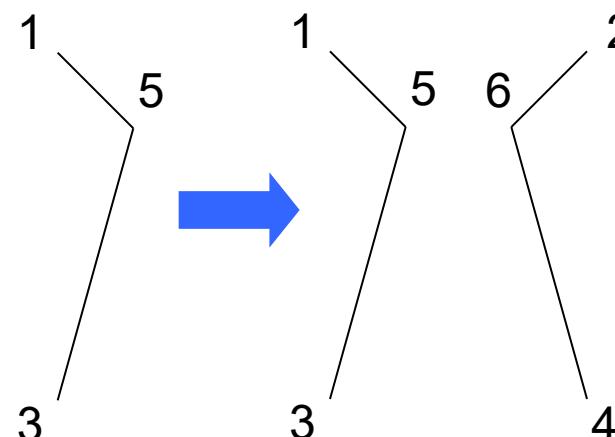
最短距離の葉のペア1,
2が系統樹のトポロジー
上で近隣でない

$$d_{12} = 0.3, d_{13} = 0.5, d_{14} = 0.6 \\ r_1 = 0.7, r_2 = 0.7, r_3 = 1.0, r_4 = 1.0 \\ D_{12} = -1.1, D_{13} = -1.2, D_{14} = -1.1$$

UPGMA法



近隣結合法



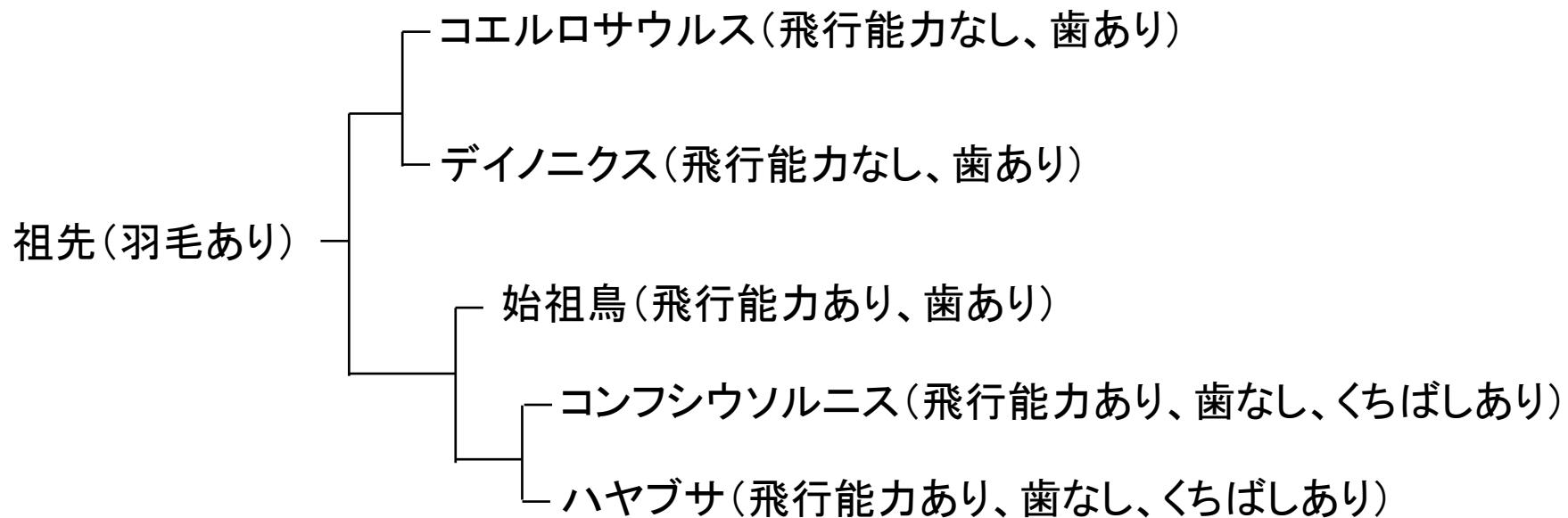
実際の進化の解析

- 系統樹を、特定の遺伝子、タンパク質の配列だけから求めるには限界がある
- 化石などから得られる形態的な特徴、埋もれていた地層などから推定される年代などは重要
- ただし、配列から系統樹を求める方法は、配列という共通の尺度が使えるので、動物、植物、微生物と、適用範囲が広い

形質からの進化系統樹

生物名	羽毛	飛行能力	歯の有無	尾羽の短縮	くちばしの発達
コエルロサウルス	1	0	1	0	0
ディノニクス	1	0	1	0	0
始祖鳥	1	1	1	0	0
コンフシウソルニス	1	1	0	1	1
ハヤブサ	1	1	0	1	1

進化による変化の回数が最小になるよう系統樹を描く



ヘモグロビンのホモログ

ヘモグロビンのホモログ

ヒトのαヘモグロビン	HBA_HUMAN
ハトのαヘモグロビン	HBA_COLLI
マグロのαヘモグロビン	HBA_THUTH
ヒトのβヘモグロビン	HBB_HUMAN
ヒトのミオグロビン	MYG_HUMAN
Lupinレグヘモグロビン	LGB2_LUPLU

ヒトのαヘモグロビン	HBA_HUMAN
ハトのαヘモグロビン	HBA_COLLI
マグロのαヘモグロビン	HBA_THUTH
ヒトのβヘモグロビン	HBB_HUMAN
ヒトのミオグロビン	MYG_HUMAN
Lupinレグヘモグロビン	LGB2_LUPLU

ヒトのαヘモグロビン	HBA_HUMAN
ハトのαヘモグロビン	HBA_COLLI
マグロのαヘモグロビン	HBA_THUTH
ヒトのβヘモグロビン	HBB_HUMAN
ヒトのミオグロビン	MYG_HUMAN
Lupinレグヘモグロビン	LGB2_LUPLU

ハト: *Columba livia* (rock dove, カワラバト)

マグロ: *Thunnus thynnus* (bluefin tuna, タイセイヨウクロマグロ)

-MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFSLFPTTKTYFPHF-DLSH----GS
-MVLSANDKSNVKAVFAKIQQAGDLGEALERLFITYPQTCKTYFPHF-DLSH----GS
-TTLSDKDKSTVKALWGKISKSSADAIGADALGRMLAVYPQTCKTYFSHPDMSP----GS
MVHLTPEEKSAVTALWGKVN--VDEVGGEALGRLLVVYPWTQRFESFGDLSTPDAMGN
-MGLSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKSEDEMKA
MGALESQAAVKSSWEEFNANIPKHTHRFFILVLEIAPAAKDLFSFLKGTS--VPQNN

ヘム結合

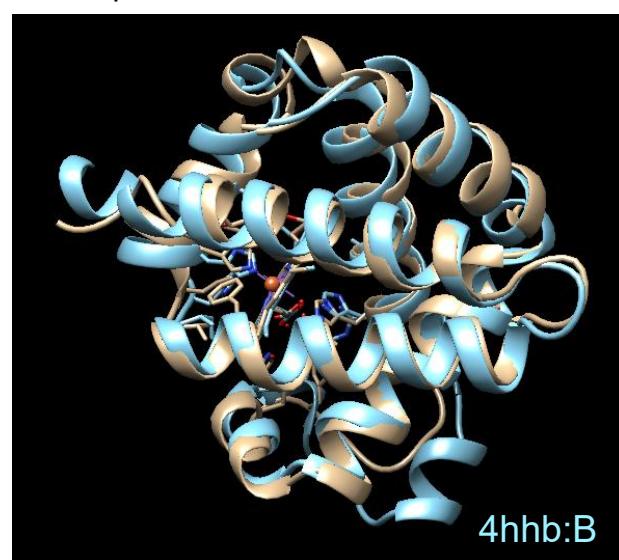
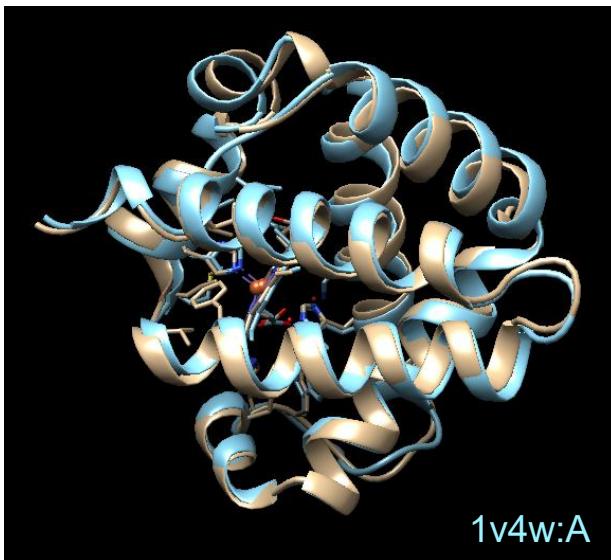
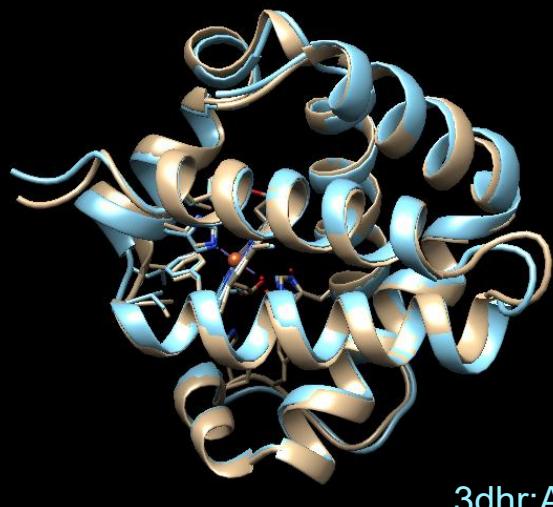
AQVKHGKKVADALTNAVAHV----DDMPNALSAISDIHAHKLRVDPVNFKLLSHCILV
AQIKHGKKVVAEALVVEAANHI----DDIAGAISKLSDLHAQKLRVDPVNFKLLGHCFLV
GPVKAHGKKVMGGVALAVTKI----DDLTTGLGDLSELHAFKMRVDPSNFKILSHCILV
PKVKAHGKKVLGAFSDGLAHL----DNLKGTATLSELHCDKLHVDPENFRLLGNVLVC
EDLKKHGATVLTAIIGGILKK----GHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQ
PELQAHAGKVFKLVYEAAIQLQVTGVVVTDATLKNLGSVHVSKGVA-DAHFPVVKEAILK

ヘム結合

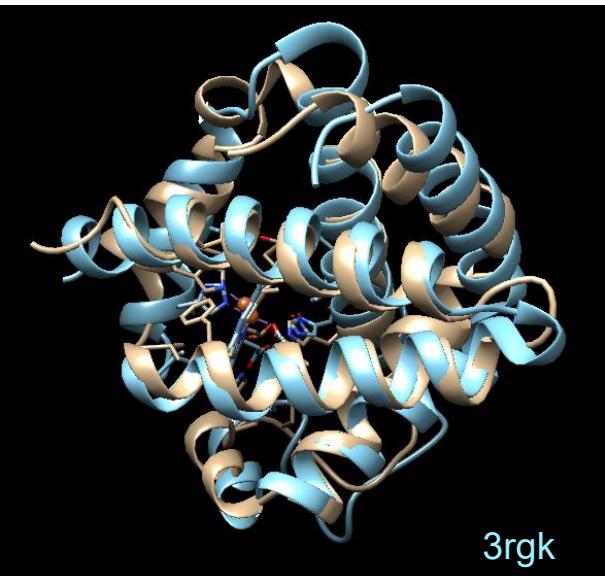
TLAAH----LPAEFTPAVHASLDKFLASVSTVLTSKYR-----	100 %
VVAVH----FPSLLTPEVHASLDKFVLAvgTVLTAKYR-----	69.0%
VVAKM----FPKEFTPDAHVSLDKFLASVALALAERYR-----	54.9%
VLAHH----FGKEFTPPVQAAVQKVAGVANALAHKYH-----	45.0%
VLQSK----HPGDFGADAQGMNKALELFRKDMASNYKELGFQG	27.5%
TIKEVVGAKWSEEINSAWTIAYDELAIVIKKEMNDAA-----	16.4%

ヘモグロビンのホモログの構造

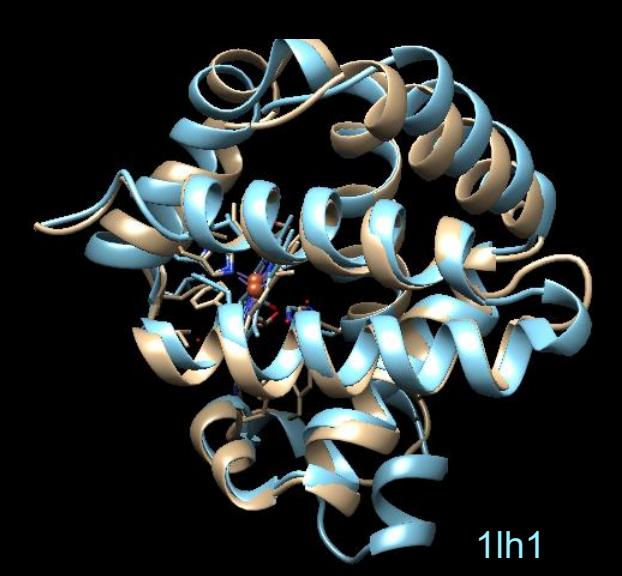
ハトのαヘモグロビン RMSD 0.722Å (141aa) マグロのαヘモグロビン RMSD 0.886Å (141aa) ヒトのβヘモグロビン RMSD 1.791Å (139aa)



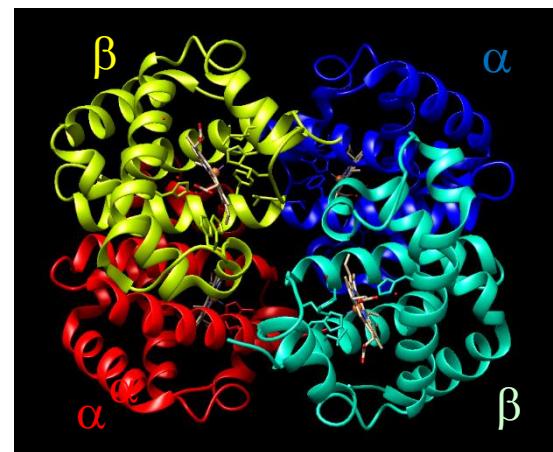
ヒトのミオグロビン RMSD 2.126Å (136aa)



Lupinレグヘモグロビン RMSD 3.641Å (141aa)



- ヒトのαヘモグロビン (4hhb:A)
- 比較対象の分子



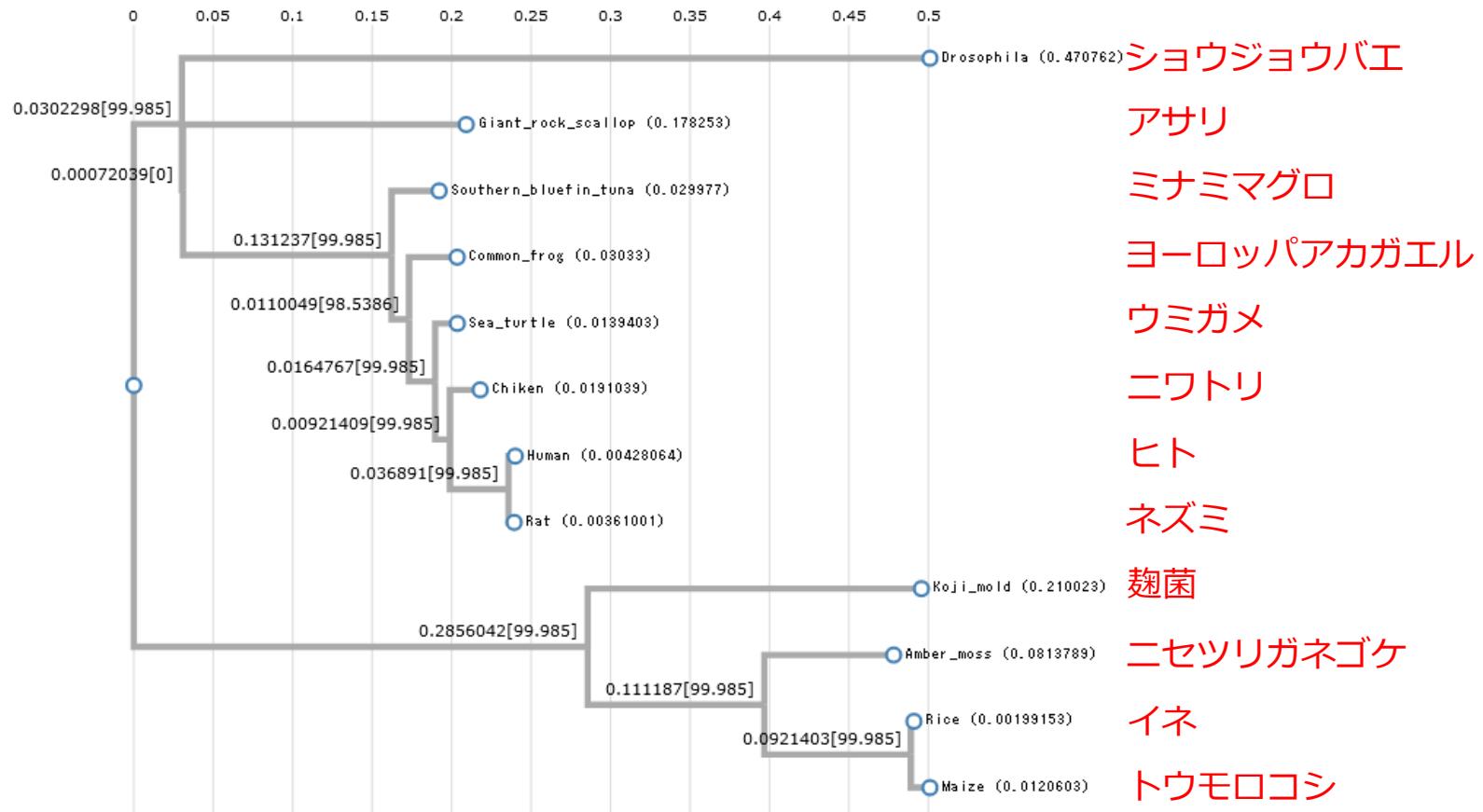
18S rRNAの配列による分子系統樹の作成

- さまざまな生物種の18S rRNA小サブユニットの配列 (18srRNA.fasta)

Phylogram (midpoint rooted tree)

without branch length without branch length labels without leaf labels without ticks

JSON  SVG  PNG 



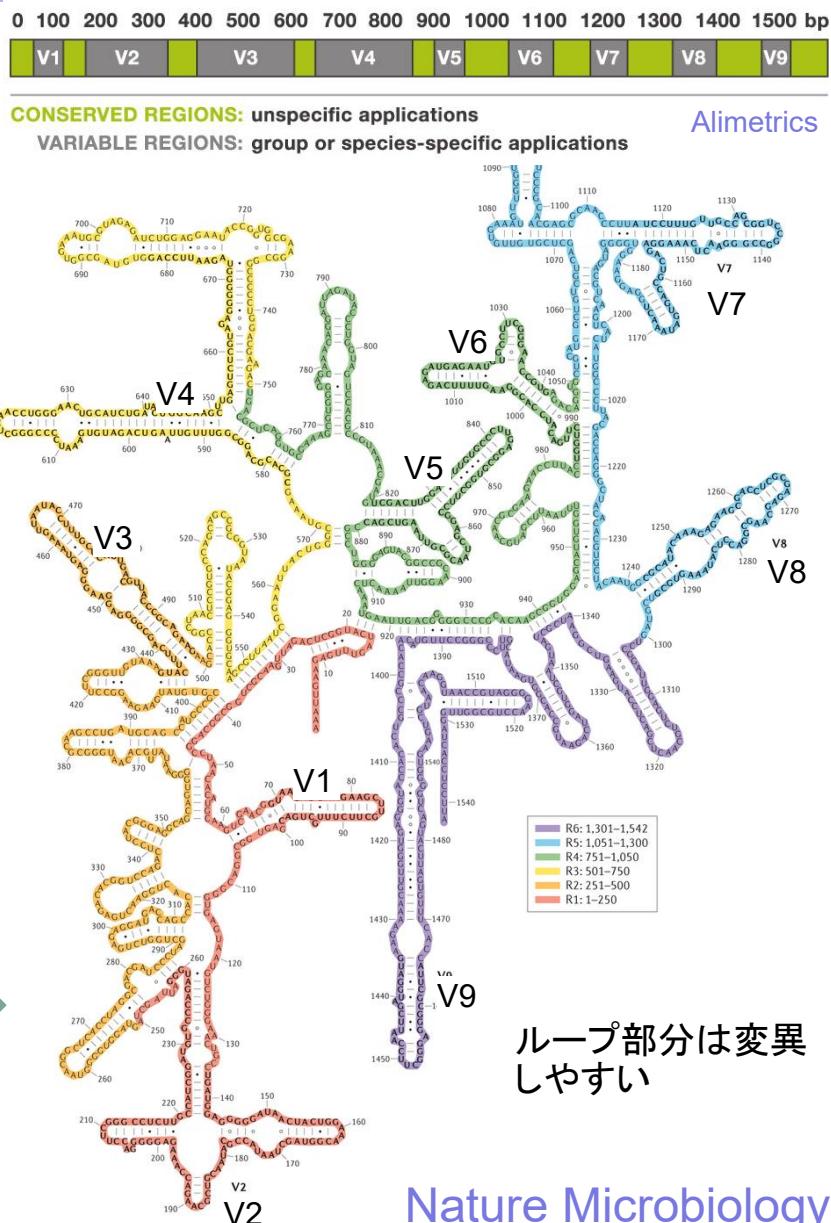
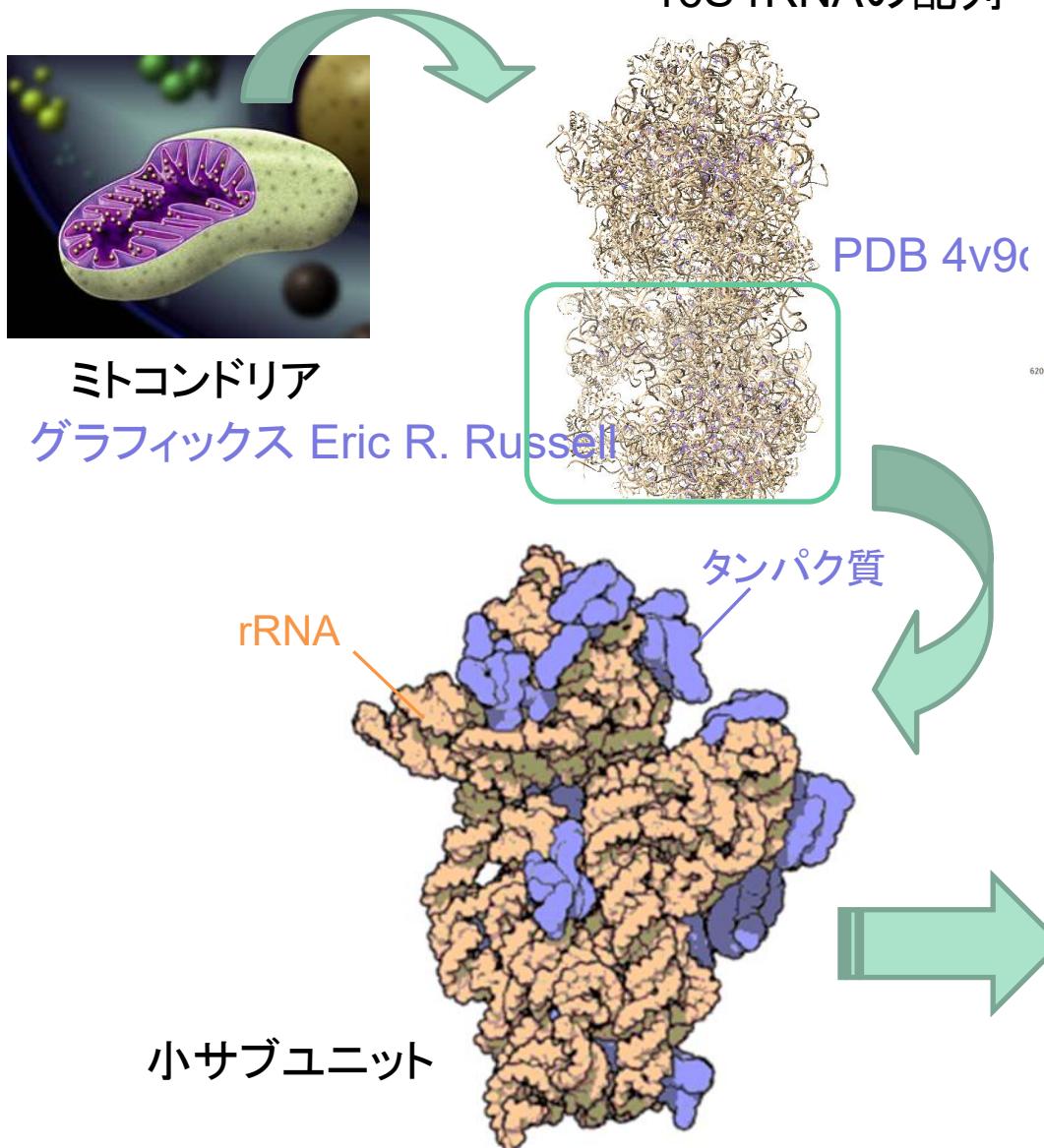
リボソームの小サブユニットの系統解析

- 16S rRNA系統解析
 - リボソームの小サブユニット16Sの遺伝子配列を用いた進化系統解析
 - 属レベルの推定、メタゲノム解析などにも利用される
 - 16S rRNAの特徴
 - さまざまな生物種に共通に存在
 - 1600塩基対程度の長さ
 - 保存性が高い部分と比較的変異しやすい部位が存在
 - 遠縁、近縁の両方の種の解析に適用できる
- 系統解析によく用いられるもの
 - 原核生物: ミトコンドリアゲノム、16S rRNA（小サブユニット）、23S rRNA（大サブユニット）
 - 真核生物: 18S rRNA（小サブユニット、16Sに対応）、28S rRNA（大サブユニット、23Sに対応）

リボソームの小サブユニットの系統解析

- 16S rRNA系統解析
 - リボソームの小サブユニット16Sの遺伝子配列を用いた進化系統解析
 - 属レベルの推定、メタゲノム解析などにも利用される
 - 16S rRNAの特徴
 - さまざまな生物種に共通に存在
 - 1600塩基対程度の長さ
 - 保存性が高い部分と比較的変異しやすい部位が存在
 - 遠縁、近縁の両方の種の解析に適用できる
- 系統解析によく用いられるもの
 - 原核生物: ミトコンドリアゲノム、16S rRNA（小サブユニット）、23S rRNA（大サブユニット）
 - 真核生物: 18S rRNA（小サブユニット、16Sに対応）、28S rRNA（大サブユニット、23Sに対応）

16S rRNAの遺伝子解析



進化系統樹の構築例

真核生物

- *Homo sapiens* (human) ヒト
- *Mus musculus* (house mouse) ハツカネズミ
- *Gallus gallus* (chicken) ニワトリ
- *Drosophila melanogaster* ショウジョウバエ
- *Oryza sativa* (rice) イネ
- *Zea mays* トウモロコシ
- *Lentinula edodes* シイタケ
- *Saccharomyces cerevisiae* 発芽酵母
- *Physarum polycephalum* モジホコリ
- *Euglena agilis* ミヅミドリムシ
- *Giardia intestinalis* ランブル鞭毛虫

真正細菌

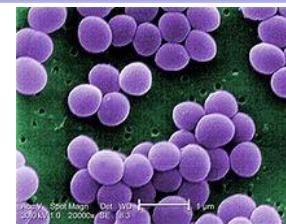
- *Staphylococcus aureus* 黄色ブドウ球菌
- *Bacillus subtilis* 枯草菌
- *Escherichia coli* 大腸菌
- *Chlamydia trachomatis* クラミジア・トラコマチス
- *Streptomyces griseus* ストレプトマイセス・グリザウス
- *Thermocrinis albus* サーモクリニス・アルバス
- *Thermotoga maritima* サーモトガ・マリチマ
- *Synechococcus sp.* シネココッカス

古細菌

- *Methanopyrus kandleri* メタノピュルス・カンドレリ
- *Thermococcus kodakarensis* テルモコックス・カダカレンシス
- *Sulfolobus solfataricus* スルホロブス・ソルファタリカス
- *Nitrososphaera viennensis* ニトロソスパエラ・ウィエンネンシス



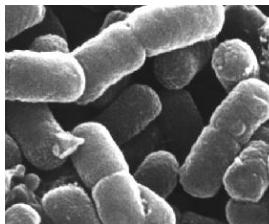
Giardia intestinalis



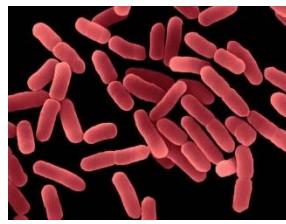
Staphylococcus aureus



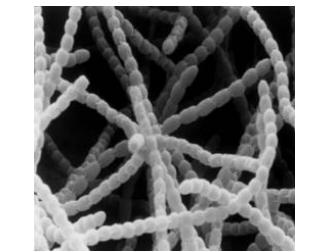
Synechococcus sp.



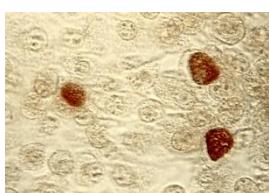
Escherichia coli



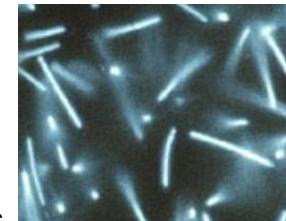
Bacillus subtilis



Streptomyces griseus



Chlamydia trachomatis



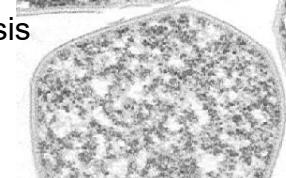
Methanopyrus kandleri



Saccharomyces cerevisiae



Thermotoga maritima



Nitrososphaera viennensis



Physarum polycephalum



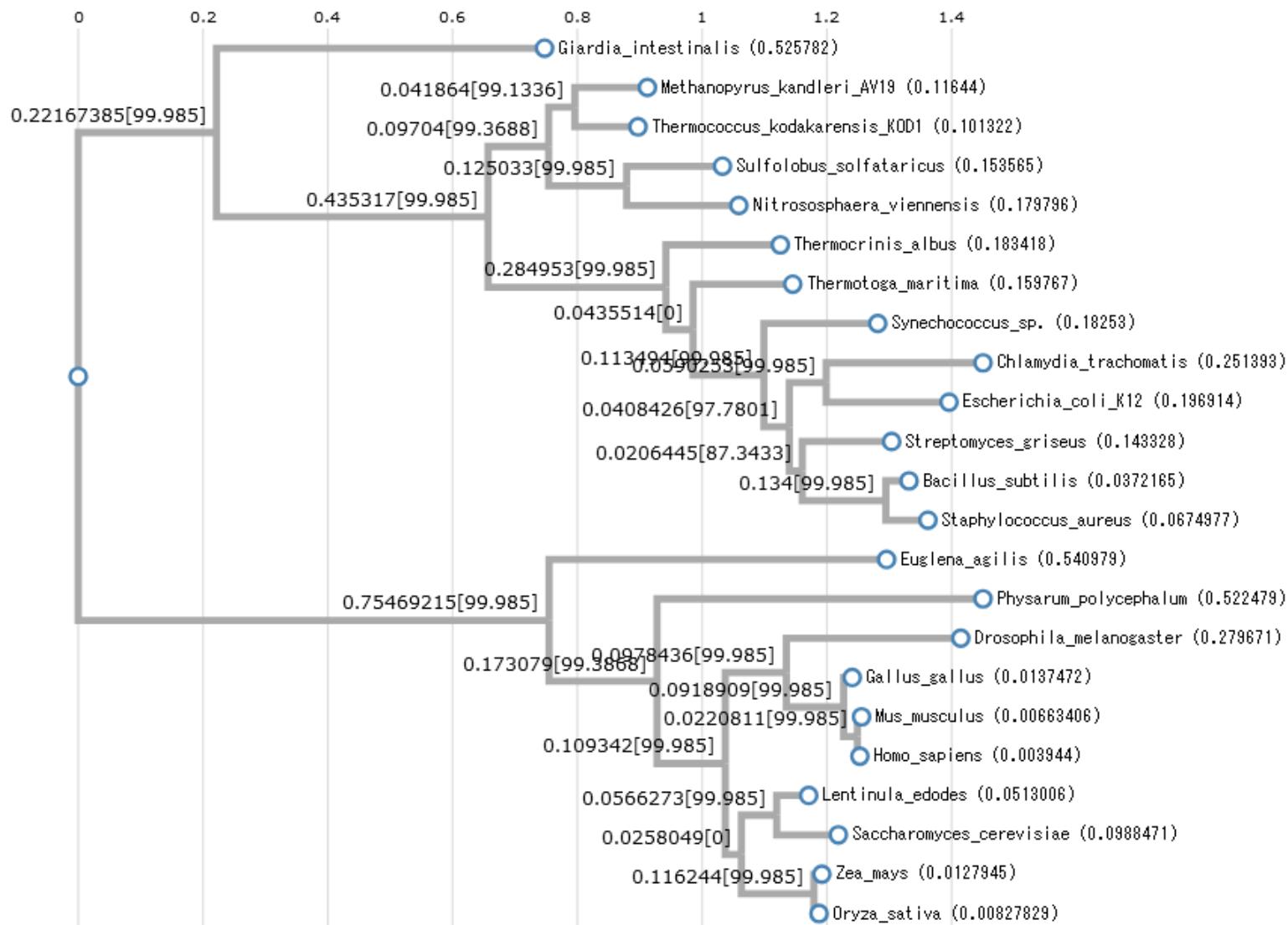
Sulfolobus solfataricus



Euglena agilis

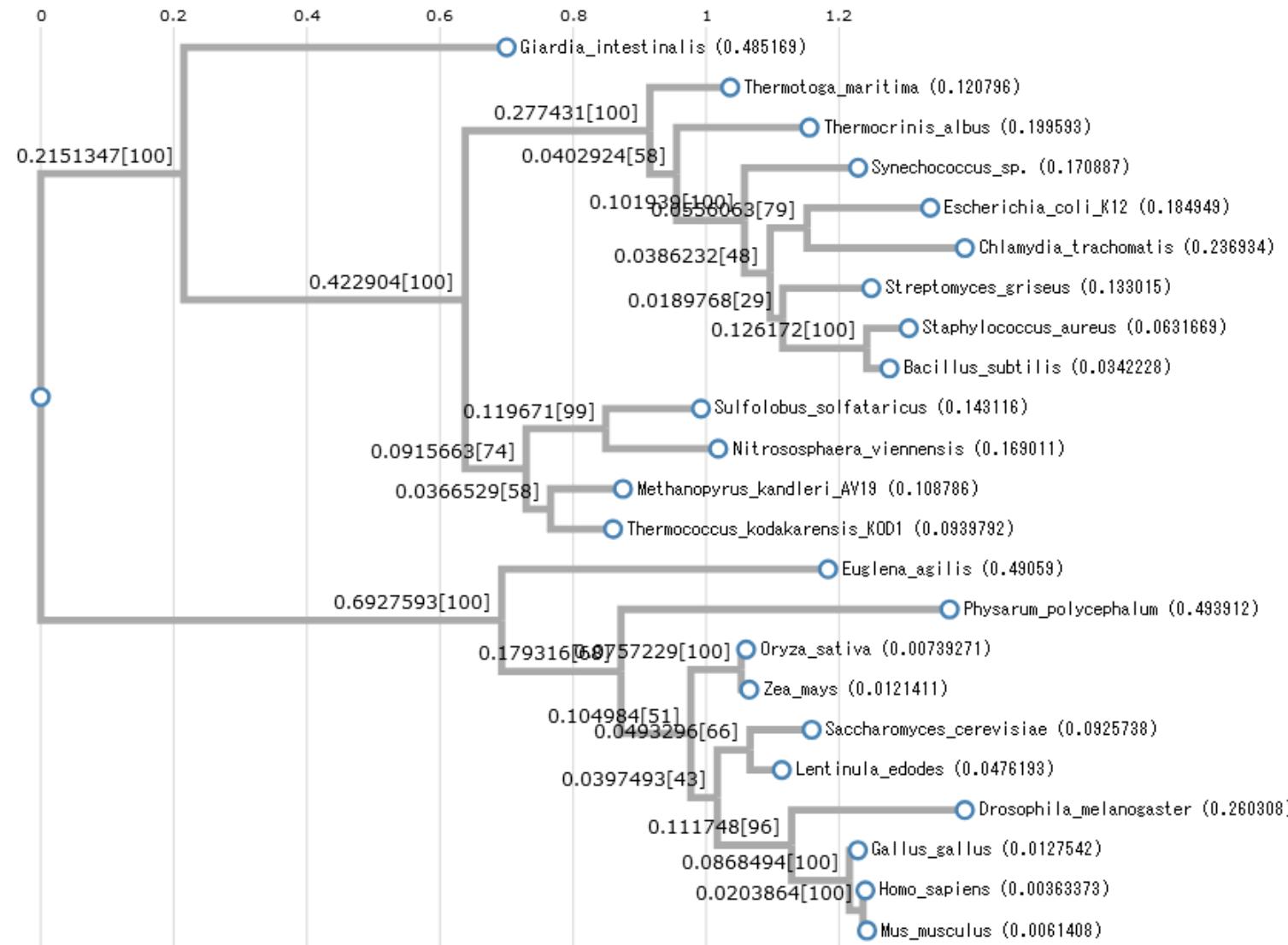
進化系統樹の構築例

平均距離法



進化系統樹の構築例

近隣結合法

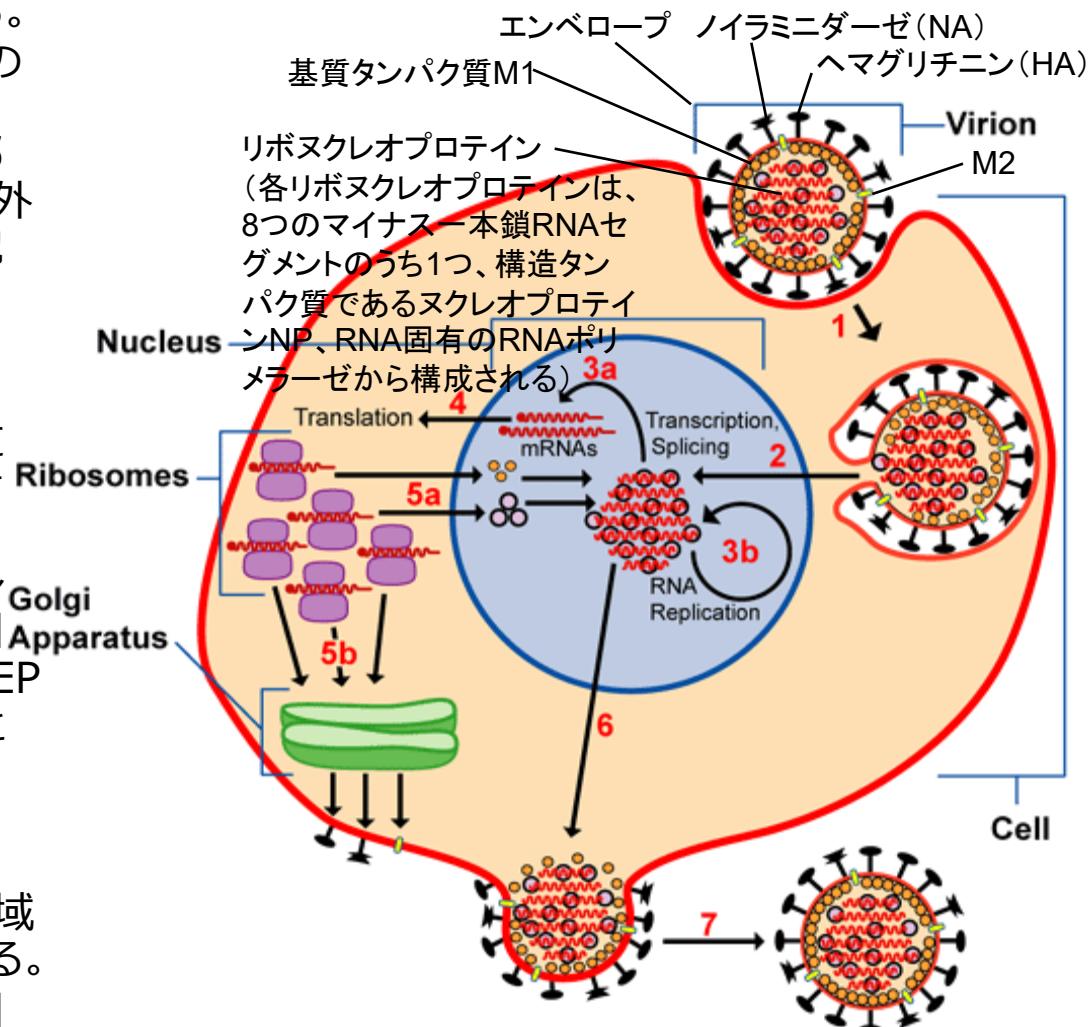


進化系統樹の構築例

- 前ページのさまざまな生物種のrRNA小サブユニット ([various-rRNA.fasta](#))
- インフルエンザウイルスのヘマグリチニンの配列 ([flu.fasta](#))
- インフルエンザウイルスのH1N1型のヘマグリチニンの配列 ([flu-h1.fasta](#))
- 新型コロナウイルスSARS-CoV-2のゲノム配列 ([SARS-CoV-2-genome.fasta](#))

インフルエンザウイルスの感染

- ウィルスは、HAにより細胞膜に付き、エンドサイトーシスにより細胞質に入り込み、エンドソームを形成する。
- HAがエンベロープとエンドソームの膜の融合を促進する。M2はイオンチャネルとして働き、ウイルス内部のphが下がり、エンベロープM1が外れ、リボヌクレオプロテインが細胞質に遊離し、核に入り込む。
- 核では、ウイルスピリメラーゼがRNAの転写と複製を行う。
- 新たに合成されたmRNAが細胞質に移動し、そこでタンパク質に翻訳される。
- HA, NA, M2 は、翻訳後プロセッシングにより、ゴルジ体を経由して細胞膜に運ばれる。NP, M1, NS1, NEP は、核に移動し、合成されたRNAと結合して、新たにリボヌクレオプロテインを形成する。
- リボヌクレオプロテインは細胞質に移動し、HA, NA, M2が刺さった領域の細胞膜とM1を介して相互作用する。
- 新たに合成されたウィルスが感染細胞から出芽する。



マルチプルアラインメントの高速化

- MAFFT

- 最初に k -tuple (k 個の文字) を単位に一致部分を高速検索する
- 複数の配列どうしのアラインメントに高速フーリエ変換 (FFT) を利用



Multiple Sequence Alignment by MAFFT

ETE3 **MAFFT** CLUSTALW PRRN

Multiple sequence alignment by MAFFT v6.864

Basic settings

Input:
Paste protein or DNA sequences in fasta format

SARS-CoV-2の系統樹



Multiple Sequence Alignment by MAFFT

ETE3

MAFFT

CLUSTALW

PRRN

Multiple sequence alignment by MAFFT v6.864

Basic settings

Input:

Paste protein or DNA sequences in fasta format

or upload a file: 参照... ファイルが選択されていません。

Submit **Reset**

Advanced settings

Output order:

- Same as input
 - Aligned

Strategy:

- Moderately accurate (FFT-NS-2, FFT-NS-i or L-INS-i; depends on data size)
 - FFT-NS-1 (Very fast; progressive method)
 - FFT-NS-2 (Fast; progressive method)
 - (Medium; iterative refinement method, two cycles only)
 - FFT-NS-i (Slow; iterative refinement method)
 - L-INS-i (Very slow; probably most accurate; recommended for <200 sequences)
 - G-INS-i (Very slow; suitable for sequences of similar lengths; recommended for <200 sequences)
 - E-INS-i (Very slow; suitable for sequences containing large unalignable regions; recommended

MAFFT Result

Select tree menu

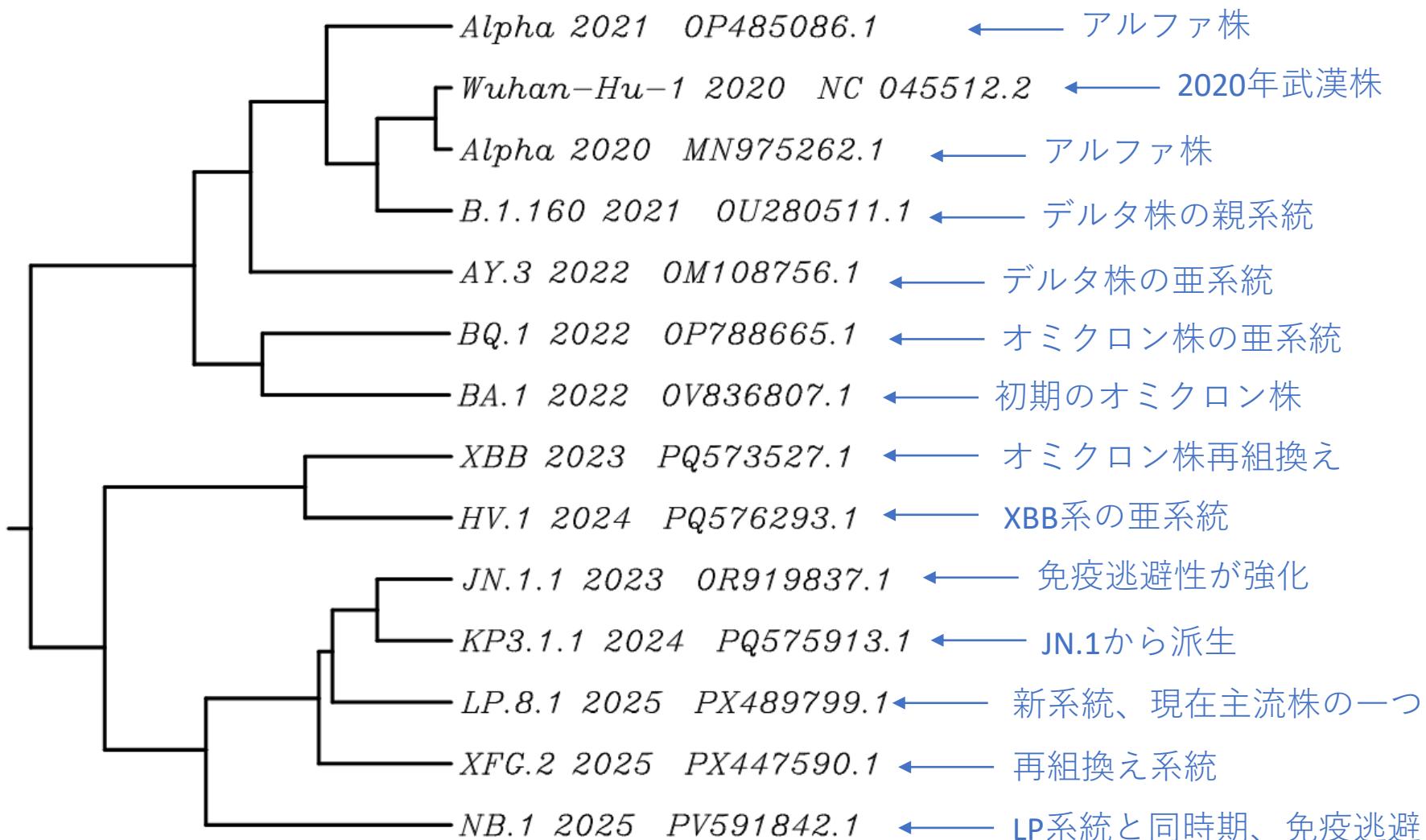
Exec

1. Alpha(2021) OP485086.1
 2. Wuhan-Hu-1(2020) NC_045512.2
 3. XBB(2023) PQ573527.1
 4. HV.1(2024) PQ576293.1
 5. B.1.160(2021) OU208511.1
 6. JN.1.1(2023) OR919837.1
 7. Alpha(2020) MN975262.1
 8. KP3.1.1(2024) PQ575913.1
 9. BQ.1(2022) OP788665.1
 10. BA.1(2022) OV836807.1
 11. AY.3(2022) OM108756.1
 12. LP.8.1(2025) PX489799.1
 13. NB.1(2025) PV591842.1
 14. XFG.2(2025) PX447590.1

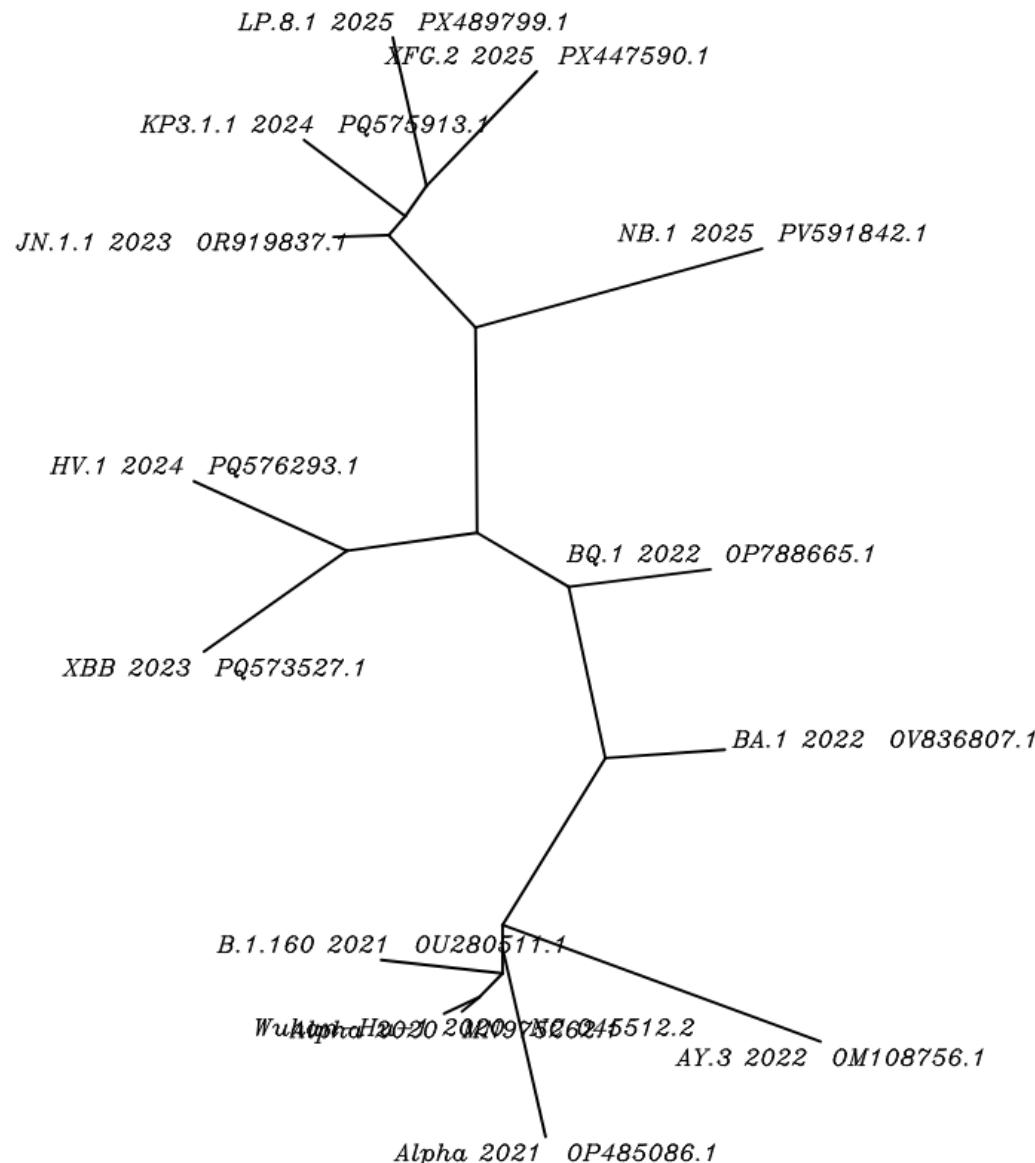
- | | |
|------------------|---|
| 1. Alpha(2021) | gttctctaaacgaactttaaaatctgtggctgtcactcggtgcacgtcttagtcact |
| 2. Wuhan-Hu-1(2 | gttctctaaacgaactttaaaatctgtggctgtcactcggtgcacgtcttagtcact |
| 3. XBB(2023) PQ | ----- |
| 4. HV.1(2024) P | -----aactttaaaatctgtggctgtcactcggtgcacgtcttagtcact |
| 5. B.1.160(2021) | gttctctaaacgaactttaaaatctgtggctgtcactcggtgcacgtcttagtcact |
| 6. JN.1.1(2023) | gttctctaaacgaactttaaaatctgtggctgtcactcggtgcacgtcttagtcact |
| 7. Alpha(2020) | gttctctaaacgaactttaaaatctgtggctgtcactcggtgcacgtcttagtcact |
| 8. KP3.1.1(2024) | gttctctaaacgaactttaaaatctgtggctgtcactcggtgcacgtcttagtcact |
| 9. BQ.1(2022) O | gttctctaaacgaactttaaaatctgtggctgtcactcggtgcacgtcttagtcact |
| 10. BA.1(2022) | gttctctaaacgaactttaaaatctgtggctgtcactcggtgcacgtcttagtcact |
| 11. AY.3(2022) | gttctctaaacgaactttaaaatctgtggctgtcactcggtgcacgtcttagtcact |
| 12. LP.8.1(2025) | gttctctaaacgaactttaaaatctgtggctgtcactcggtgcacgtcttagtcact |
| 13. NB.1(2025) | gttctctaaacgaactttaaaatctgtggctgtcactcggtgcacgtcttagtcact |
| 14. YEG_7(2025) | -----aactttaaaatctgtggctgtcactcggtgcacgtcttagtcact |

1. Alpha(2021) cacgcgatataattataactaattactgtcggtgacaggacacgagaactctgttata
2. Wuhan-Hu-1(2020) cacgcgatataattataactaattactgtcggtgacaggacacgagaactctgttata
3. XBB(2023) PQ -----ataactaattactgtcggtgacaggacacgagaactctgttata
4. HV.1(2024) P cacgcgatataattataactaattactgtcggtgacaggacacgagaactctgttata
5. B.1.160(2021) cacgcgatataattataactaattactgtcggtgacaggacacgagaactctgttata
6. JN.1.1(2023) cacgcgatataattataactaattactgtcggtgacaggacacgagaactctgttata
7. Alpha(2020) cacgcgatataattataactaattactgtcggtgacaggacacgagaactctgttata
8. KP3.1.1(2024) cacgcgatataattataactaattactgtcggtgacaggacacgagaactctgttata

SARS-CoV-2の系統樹



SARS-CoV-2の系統樹



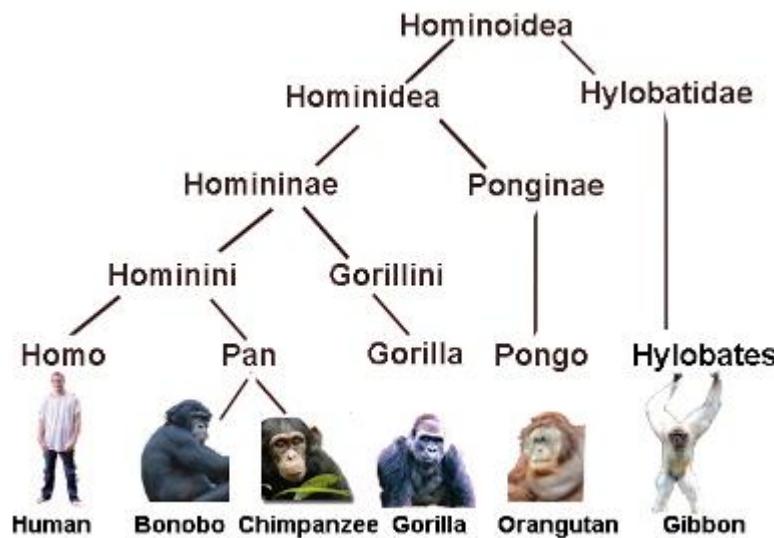
SARS-CoV-2のゲノムの変異

- 変異が大きい部分は、スパイク遺伝子（S）
 - オミクロンでは S に 30-50 超の置換/欠失が集中
 - 感染の鍵となる遺伝子（タンパク質）
 - ヒト細胞への結合と膜融合を担う
 - 中和抗体の標的 → ワクチン設計の基盤
 - 変異株の分化や感染性・免疫逃避の主要因

課題 5

霊長類のCOX遺伝子の配列 ([homoidea-cox1.fasta](#))について以下の問い合わせに答えよ。

1. Clustal Wのサイトを用いて、分子系統樹を作成せよ。
2. 実際のヒト上科の系統樹は以下の通りであることが知られている。2.で得られた結果と比較してみよう。



COX: ミトコンドリアのエネルギー生成に関わる酵素

ヒト、チンパンジー、ゴリラなどの分岐は短期間で起きているため、そうした微少な変異だけで分岐を識別することは難しい

<http://aquatic-human-ancestor.org/timeline.html>