

# 生体情報工学演習 第2-4回

北海道大学情報エレクトロニクス学科

生体情報コース3年

学生番号: 02180144 茂木貴紀

作成日: 2020/12/12

## はじめに

今回の演習では、これまで行ったアラインメントの学習の発展として、そこから分子系統樹を作成し、その評価について学習する。

まず、ケモカインレセプター のアミノ酸配列を相同性検索して得られたアミノ酸配列らを分子系統樹にし、考察した。同様のことを SARS-CoV-2 でも行った。さらに、血圧を制御する分子をコードする塩基配列について、霊長類のオルソログをまとめたものをアラインメントし、分子系統樹にすることで、種の系統関係を調べた。また、分子系統樹は近隣結合法と最尤法を用いて作成し、その違いについて分析した。

- 近隣結合法とは、全ての配列同士において、置換数の推定などから距離行列を作成し、その距離が近いもの同士を順々につないでいくアルゴリズムを利用した方法である。
- 最尤法とは、アミノ酸の単位時間内における置換の遷移確率行列を仮定し、尤度を求めていくことで一番高い系統樹を選択していく方法である。統計推定における最尤法を適用したものである。

## 1 アミノ酸配列における分子系統樹の作成

まず、前回の演習で作成した

ケモカインレセプター のアミノ酸配列を相同性検索して得られたアミノ酸配列ら

を分子系統樹にしてみる（その前に、何らかの条件でアラインメントする）。まず、それぞれの配列同士の進化距離を確認してみる。モデルはp距離を用いている。p距離とは配列の長さに対してどれだけ要素の違いがあるかの割合を示している。つまり、どれだけアミノ酸が異なるかを数値化したものであり、結果が近隣結合法で得られる系統樹に直結している。

進化距離および近隣結合法で得た系統樹を以下に示す。

2020/12/13

生体情報工学演習 第2-4回 - HackMD

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
_C-X-C_chemokine_receptor_type_1_Homo_sapiens																	
_C-X-C_chemokine_receptor_type_2_isoform_X1_Homo_sapiens	0.262																
_C-X-C_chemokine_receptor_type_2_Homo_sapiens	0.262	0															
_C-X-C_chemokine_receptor_type_2_Homo_sapiens	0.262	0	0														
_C-C_chemokine_receptor_type_6_Homo_sapiens	1.079	1.067	1.067	1.067													
_C-C_chemokine_receptor_type_6_Homo_sapiens	1.065	1.077	1.077	1.077	0.009												
_C-C_chemokine_receptor_type_4_Homo_sapiens	1.357	1.326	1.326	1.326	1.196	1.198											
_C-C_chemokine_receptor_type_9_isoform_B_Homo_sapiens	1.271	1.258	1.258	1.258	1.146	1.139	1.349										
_C-C_chemokine_receptor_type_9_isoform_B_Homo_sapiens	1.271	1.258	1.258	1.258	1.146	1.139	1.349	0									
_C-C_chemokine_receptor_type_9_isoform_X1_Homo_sapiens	1.271	1.258	1.258	1.258	1.146	1.139	1.349	0	0								
_C-C_chemokine_receptor_type_9_isoform_A_Homo_sapiens	1.271	1.258	1.258	1.258	1.146	1.139	1.349	0	0	0							
_C-X-C_chemokine_receptor_type_4_isoform_b_Homo_sapiens	1.365	1.375	1.375	1.375	1.375	1.373	1.357	1.325	1.325	1.325	1.325						
_C-X-C_chemokine_receptor_type_4_isoform_a_Homo_sapiens	1.394	1.381	1.381	1.381	1.384	1.382	1.373	1.347	1.347	1.347	1.347	0.016					
_C-X-C_chemokine_receptor_type_3_isoform_1_Homo_sapiens	1.252	1.174	1.174	1.174	1.296	1.315	1.459	1.509	1.509	1.509	1.509	1.365	1.361				
_C-X-C_chemokine_receptor_type_5_isoform_1_Homo_sapiens	1.436	1.273	1.273	1.273	1.444	1.428	1.523	1.472	1.472	1.472	1.472	1.381	1.382	1.237			
_C-X-C_chemokine_receptor_type_3_isoform_X1_Homo_sapiens	1.252	1.174	1.174	1.174	1.296	1.315	1.459	1.509	1.509	1.509	1.509	1.365	1.361	0	1.237		
_C-X-C_chemokine_receptor_type_3_isoform_2_Homo_sapiens	1.252	1.174	1.174	1.174	1.296	1.315	1.459	1.509	1.509	1.509	1.509	1.365	1.361	0	1.237	0	

表1 アミノ酸間の進化距離



図1 近隣結合法で作成した系統樹

次に、これを最尤法で作成したものを示す。



図2 最尤法で作成した系統樹

これらの系統樹の主な違いとして、進化距離があげられる。進化距離は系統樹における枝の長さに対応しており、その生物になった時期の推定にも役立つ。

近隣結合法では分子速度の一定性を仮定していないため、このような違いが生じると考えられる。

## 2 SARS-CoV-2の系統樹

次に、前章で行った分子系統樹の作成&考察を踏まえて、SARS-CoV-2 の分子系統樹を近隣結合法で作成する。以前、このウイルスに関して相同性検索を行い、アラインメントしたのち作成した系統樹は以下になった。時間がかかるため、bootstrap値は無しにした。

```
ABD72985.1:1-1255 spike glycoprotein Severe acute respiratory syndrome-related coronavirus
ABD73001.1:1-1255 spike glycoprotein Severe acute respiratory syndrome-related coronavirus
ABD72977.1:1-1255 spike glycoprotein Severe acute respiratory syndrome-related coronavirus
ABD72979.1:1-1255 spike glycoprotein Severe acute respiratory syndrome-related coronavirus
ABD72995.1:1-1255 spike glycoprotein Severe acute respiratory syndrome-related coronavirus
ABD72968.1:1-1255 spike glycoprotein Severe acute respiratory syndrome-related coronavirus
AEA10473.1:1-1255 spike glycoprotein precursor SARS coronavirus MA15 ExoN1
YP_009625051.1:1-1255 spike glycoprotein SARS coronavirus Tor2
AC_Z72093.1:1-1255 spike glycoprotein precursor SARS coronavirus wt/c-MB
AFR68672.1:1-1255 spike glycoprotein precursor SARS coronavirus Tor2
ADC35483.1:1-1255 spike glycoprotein precursor SARS coronavirus HKU-39849
ABD72970.1:1-1255 spike glycoprotein Severe acute respiratory syndrome-related coronavirus
AAU81608.1:1-1255 S protein SARS Coronavirus CDC*200301157
AC082725.1:1-1255 spike glycoprotein precursor SARS coronavirus P2
ABD72969.1:1-1255 spike glycoprotein Severe acute respiratory syndrome-related coronavirus
ABD72984.1:1-1255 spike glycoprotein Severe acute respiratory syndrome-related coronavirus
AFR68714.1:1-1255 spike glycoprotein precursor SARS coronavirus Tor2
AAR66775.1:1-1255 spike SARS coronavirus Shanghai/QX/C2
ABF68836.1:1-1255 S protein Severe acute respiratory syndrome-related coronavirus
AAP33697.1:1-1255 spike protein S SARS coronavirus Frankfurt 1
AAS75868.1:1-1255 SARS CoV S protein Expression vector pHCMV-SARS-S (FFM1)
AAR23250.1:1-1255 spike protein SARS coronavirus Sino1-11
AGT21078.1:1-1255 spike glycoprotein SARS coronavirus ExoN1
```

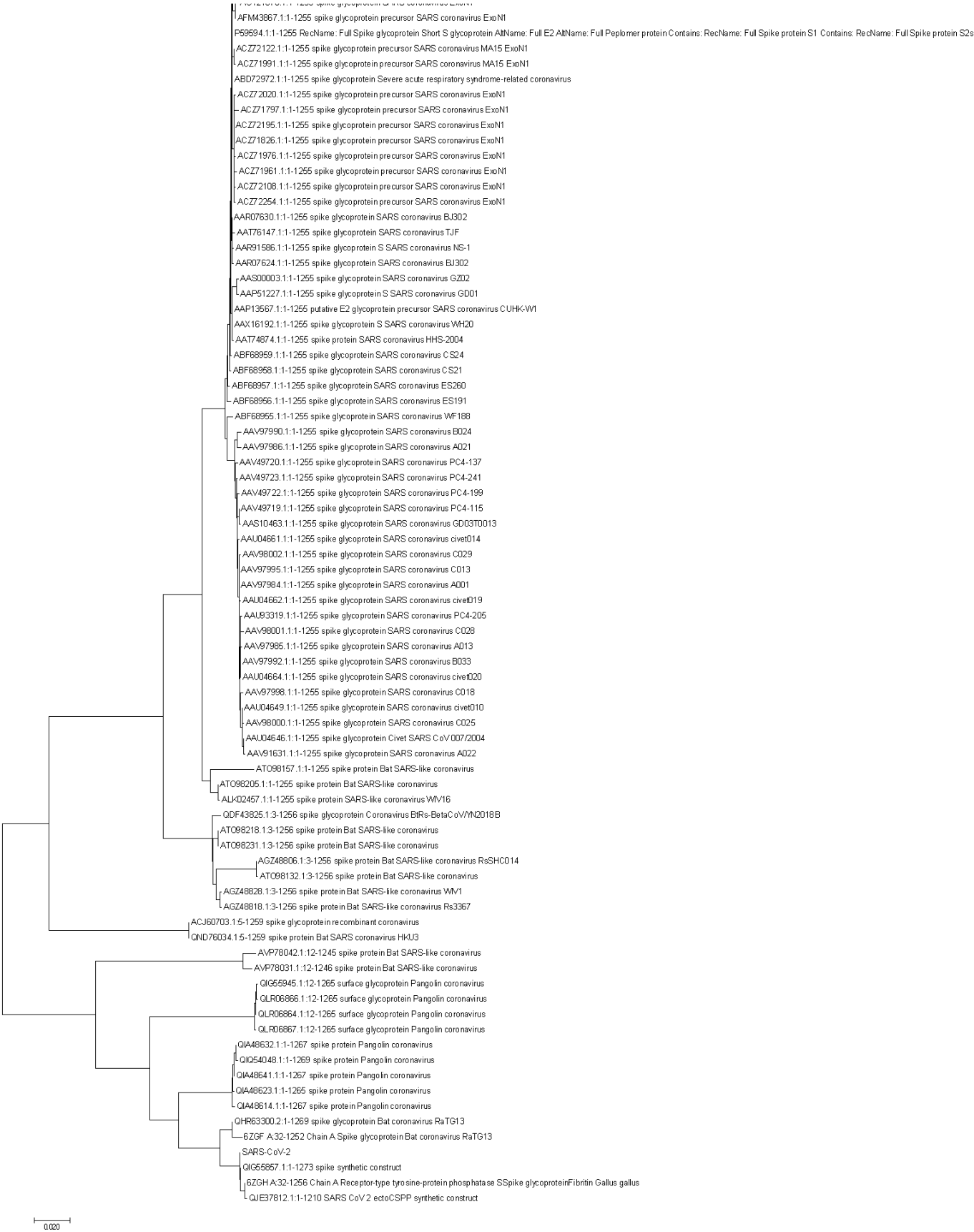


図3 SARS-CoV-2の系統樹

まず、下から4番目にquery arrayが存在している。横軸は時間であるから、query arrayが出現する前に何らかのスパイクタンパク質などが存在していたことが読み取れる。また、SARS-CoV-2が発見される前にBat（コウモリ）にウイルスが寄生していたと考えられる。

### 3 塩基配列における分子系統樹の作成

次に、これまでアミノ酸配列ベースで行ってきた系統樹作成を、塩基配列から作成する。塩基配列の時点でアラインメントするか、アミノ酸配列でアラインメントするかによって結果が変わるため、比較する必要がある。結果が変わるといのは、塩基配列をアラインメントするとコーディング領域にgapが生じ、発現するアミノ酸配列が変化する可能性があるということである。

さらに、これらに加え同義置換率及び非同義置換率を求め、アラインメントのタイミングによってこれらの値がどう変化するか分析した。

同義置換とは、発現するアミノ酸が変化しない置換のことをいい、非同義置換とは、発現するアミノ酸が変化してしまう置換のことをいう。

fx

	A	B	C	D	E	F	G	H	I	J
1	ENSP00000355627_Hsap/1-1455									
2	ENSOGAP00000010104_Ogar/1-1476	0.372								
3	ENSPTRP00000003543_Ptro/1-1455	0.038	0.368							
4	ENSNLEP00000002092_Nleu/1-1455	0.064	0.371	0.056						
5	ENSGGOP00000000194_Ggor/1-1473	0.029	0.374	0.012	0.047					
6	ENSMMP00000007736_Mmul/1-1455	0.107	0.389	0.101	0.119	0.095				
7	ENSCJAP00000022002_Cjac/1-1458	0.16	0.366	0.14	0.161	0.144	0.143			
8	ENSPANP00000007971_Panu/1-1470	0.101	0.373	0.098	0.117	0.092	0.029	0.141		
9	ENSPYP00000000122_Pabe/1-1455	0.055	0.395	0.046	0.063	0.037	0.111	0.153	0.107	
10	ENSCSAP00000006790_Csab/1-1455	0.094	0.367	0.084	0.108	0.083	0.042	0.148	0.035	0.098
11										

表2 同義置換/アミノ酸配列でアラインメント

	A	B	C	D	E	F	G	H	I	J
	ENSP00000355627_Hsap/1-1455									
	ENSOGAP00000010104_Ogar/1-1476	0.375								
	ENSPTRP00000003543_Ptro/1-1455	0.038	0.37							
	ENSNLEP00000002092_Nleu/1-1455	0.064	0.374	0.056						
	ENSGGOP00000000194_Ggor/1-1473	0.029	0.376	0.013	0.047					
	ENSMMP00000007736_Mmul/1-1455	0.107	0.392	0.102	0.119	0.096				
	ENSCJAP00000022002_Cjac/1-1458	0.161	0.368	0.141	0.162	0.145	0.145			
	ENSPANP00000007971_Panu/1-1470	0.102	0.376	0.099	0.118	0.092	0.029	0.142		
	ENSPYP00000000122_Pabe/1-1455	0.056	0.397	0.047	0.063	0.037	0.112	0.155	0.108	
	ENSCSAP00000006790_Csab/1-1455	0.095	0.369	0.085	0.109	0.083	0.043	0.149	0.035	0.099

表3 同義置換/塩基配列でアラインメント

	A	B	C	D	E	F	G	H	I	J
	ENSP00000355627_Hsap/1-1455									
	ENSOGAP00000010104_Ogar/1-1476	0.203								
	ENSPTRP00000003543_Ptro/1-1455	0.009	0.208							
	ENSNLEP00000002092_Nleu/1-1455	0.019	0.199	0.023						
	ENSGGOP00000000194_Ggor/1-1473	0.002	0.204	0.006	0.016					
	ENSMUP000000007736_Mmul/1-1455	0.043	0.207	0.045	0.035	0.04				
	ENSCJAP00000022002_Cjac/1-1458	0.077	0.211	0.079	0.068	0.075	0.063			
	ENSPANP00000007971_Panu/1-1470	0.043	0.203	0.045	0.031	0.04	0.01	0.057		
	ENSPYP00000000122_Pabe/1-1455	0.014	0.202	0.017	0.02	0.012	0.037	0.071	0.037	
	ENSCSAP00000006790_Csab/1-1455	0.04	0.203	0.042	0.031	0.038	0.007	0.057	0.004	0.035

表4 非同義置換/アミノ酸配列でアラインメント

	A	B	C	D	E	F	G	H	I	J
	ENSP00000355627_Hsap/1-1455									
	ENSOGAP00000010104_Ogar/1-1476	0.201								
	ENSPTRP00000003543_Ptro/1-1455	0.009	0.205							
	ENSNLEP00000002092_Nleu/1-1455	0.019	0.197	0.023						
	ENSGGOP00000000194_Ggor/1-1473	0.002	0.201	0.007	0.017					
	ENSMUP000000007736_Mmul/1-1455	0.043	0.203	0.045	0.035	0.041				
	ENSCJAP00000022002_Cjac/1-1458	0.077	0.206	0.08	0.068	0.075	0.063			
	ENSPANP00000007971_Panu/1-1470	0.043	0.198	0.045	0.031	0.04	0.01	0.057		
	ENSPYP00000000122_Pabe/1-1455	0.015	0.199	0.017	0.021	0.012	0.038	0.072	0.038	
	ENSCSAP00000006790_Csab/1-1455	0.039	0.2	0.041	0.03	0.037	0.009	0.058	0.005	0.034

表5 非同義置換/塩基配列でアラインメント

非同義置換は、どちらでアラインメントするかによって値が変化すると予想していたが、値を参照する限り、アミノ酸配列でアラインメントしても塩基配列でアラインメントしても割合はほとんど変化していなかった。

(ギリギリまで思考しましたが、これ以上わかりませんでした。すみません。)

次に、系統樹を近隣結合法と最尤法を用いて作成した。その結果を以下に示す。

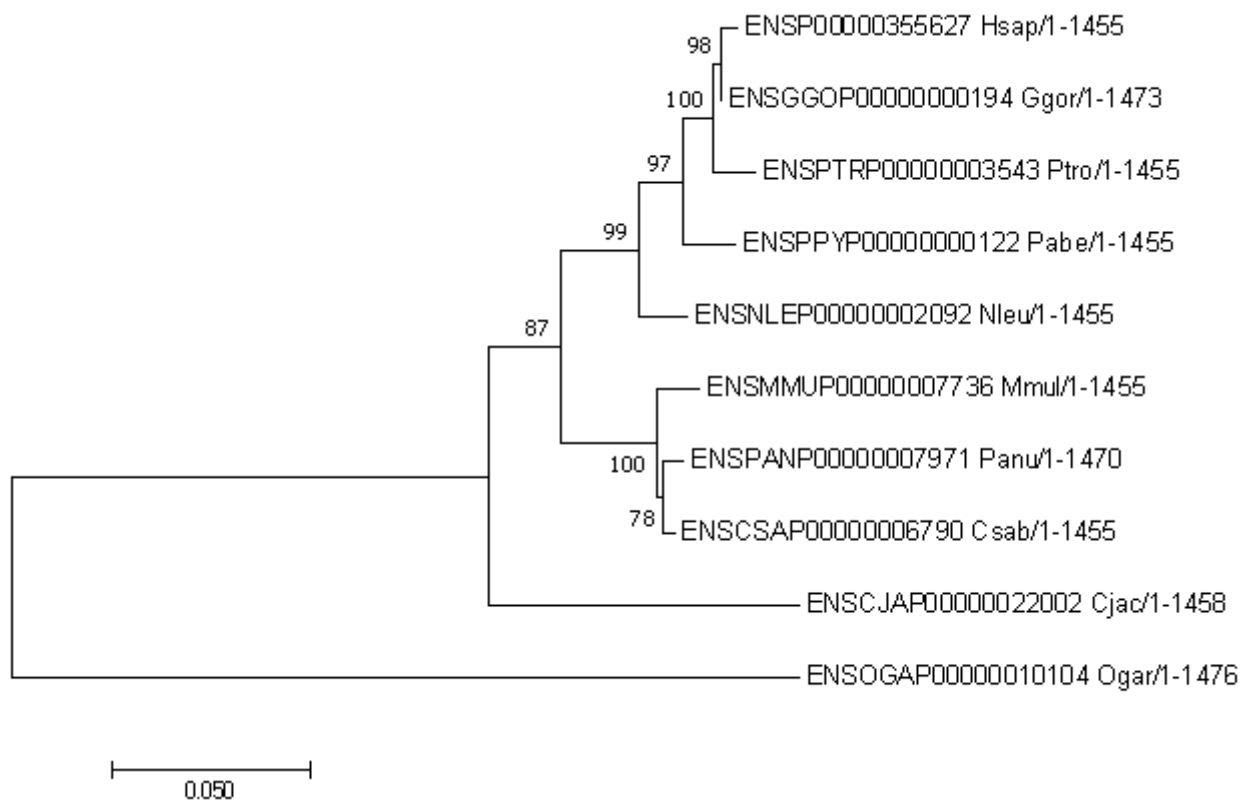


図4 近隣結合法を用いて作成した系統樹

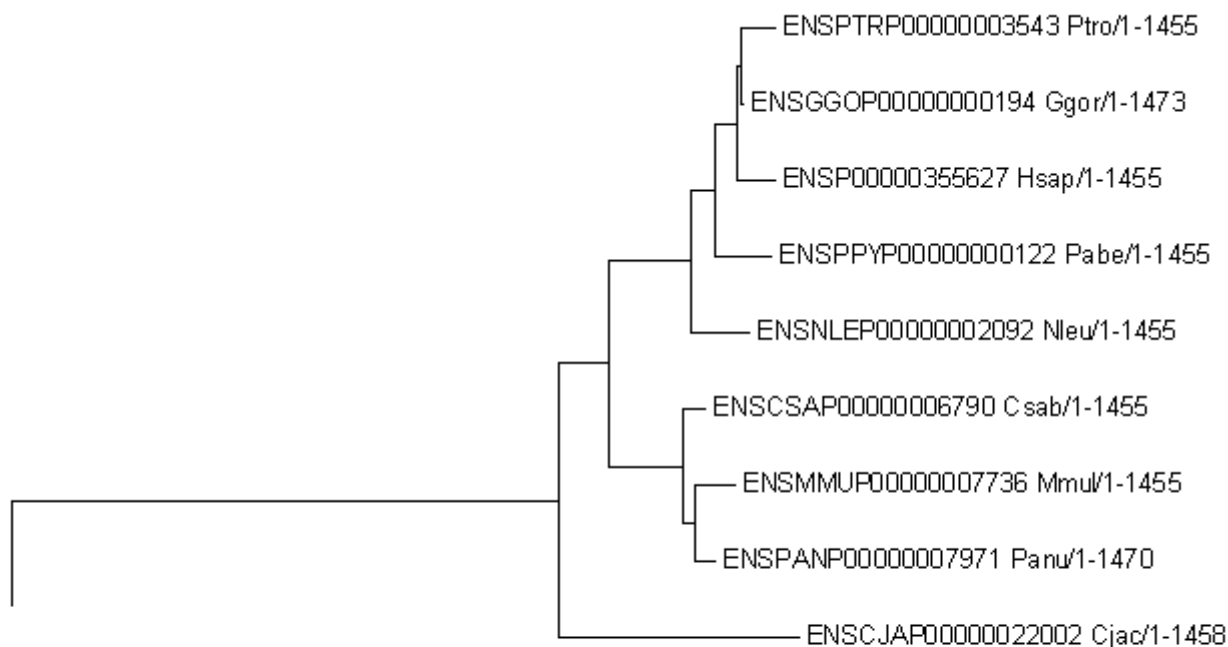


図5 最尤法を用いて作成した系統樹

系統樹を見ると、どちらも概ね一致していることがわかる。例えば、ヒトはゴリラ・チンパンジーと同じ山に位置しており、ヒトにより性質が離れていると考えられるテナガザルやガラゴ等は別の山に位置している。また、ヒトはガラゴに比べてゴリラと異常に進化距離が近いことも読み取れる。

**tags:** 生体情報工学演習