

2010/06/27

# 系統樹の工作

東京工業大学 下平英寿 <http://www.is.titech.ac.jp/~shimo/>



紙テープ(糸でもよい), はさみ, ものさし, のり, ペン, セロテープなど

# 置換数の長さに紙テープを切る

	 ヒト	 チンパンジー	 ゴリラ	 オラン
 チンパンジー	1185 119 mm			
 ゴリラ	1479 148 mm	1420 142 mm		
 オラン	2001 200 mm	2090 209 mm	2116 212 mm	
 マウス	4068 407 mm	4052 405 mm	4102 410 mm	4127 413 mm

DNAの長さ 10839

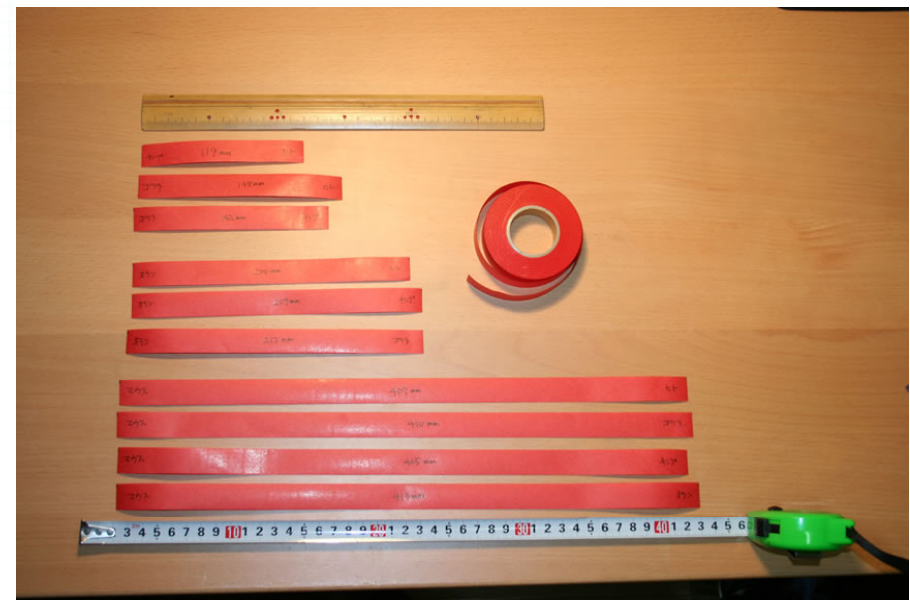
ミトコンドリアDNA 置換数(推定値)

ここでは置換数10個を1mmにした

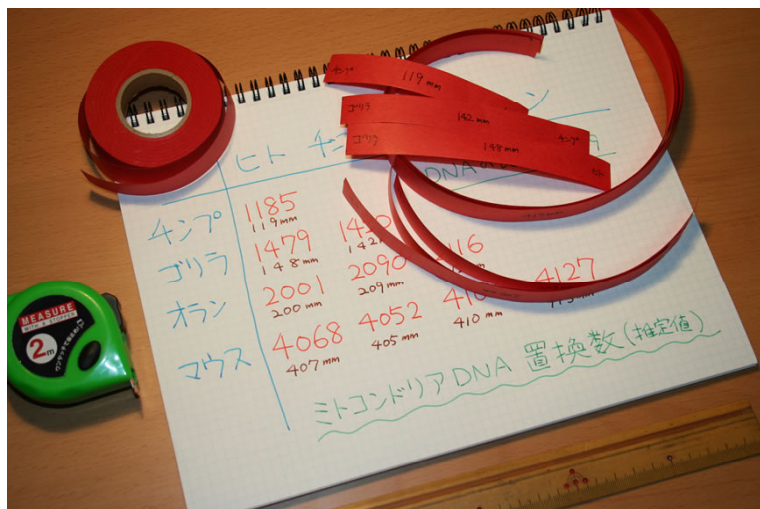
1センチがだいたい100万年です

## 参考：置換数は最尤法で推定した

- データは類人猿4種(ヒト, チンパンジー, ゴリラ, オランウータン)とマウスのミトコンドリアDNA配列をNCBIのウェブサイトから取得.
- 12個の遺伝子コーディングリージョンをclustalWでアライメントして位置合わせ. 長さ10839の配列を得る.
- R言語のapeパッケージにてDNA配列から最尤法で置換数を推定した. `dist.dna(dat, "TN93")`を実行後, 結果を10839倍して整数に丸めた. 単純な変異数(塩基が何個異なるか)は`dist.dna(dat, "raw")`で計算できる.
- 後で確かめたらJukes-Cantor (JC)モデルをつかってもほぼ同じ置換数が得られたので, TN93モデルより解説が容易なJCモデルを工作に使うべきだった.
- 置換数>変異数である. 今回のデータでは置換数があまり大きくないので, 置換数の代わりに変異数を用いても同じ系統関係が得られた.
- (工作しないで)系統樹まで求めるには, 置換数の行列をnj関数に与えれば, 近隣結合法の無根系統樹が得られる. `root()`で根を指定する.



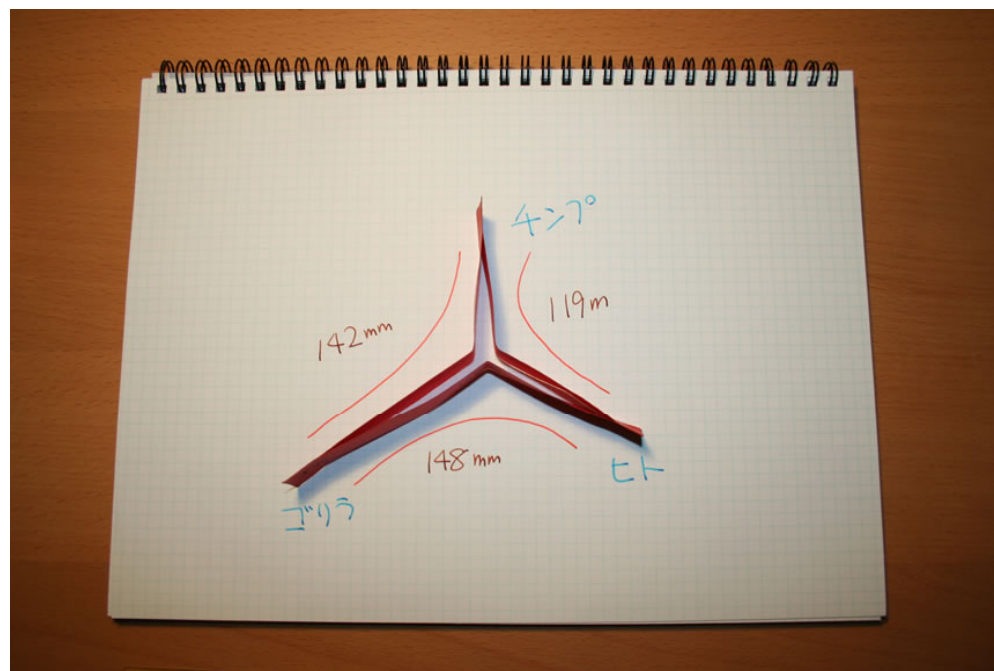
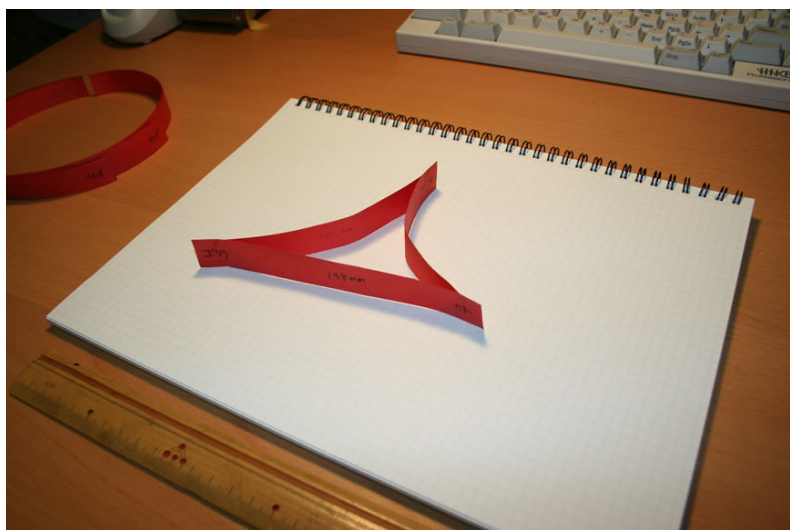
# 短い3本を貼り合わせる



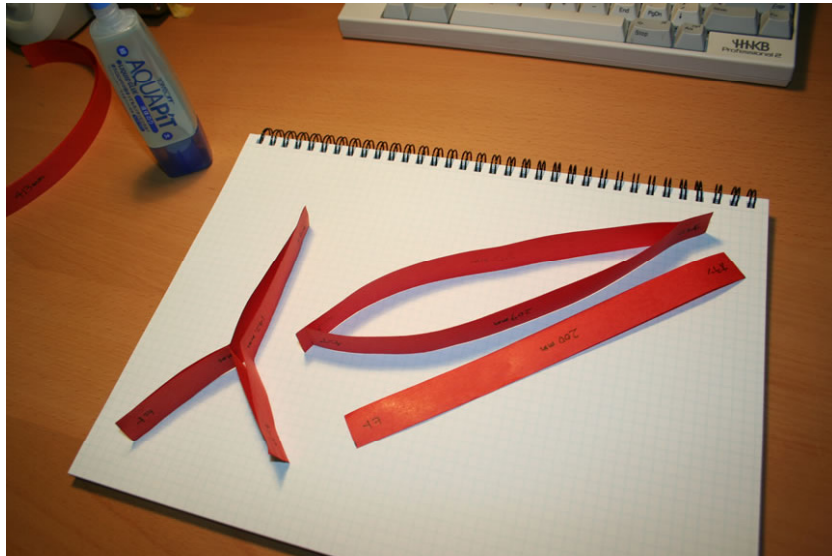
チンパンジー - ヒト

ゴリラ - ヒト

ゴリラ - チンパンジー



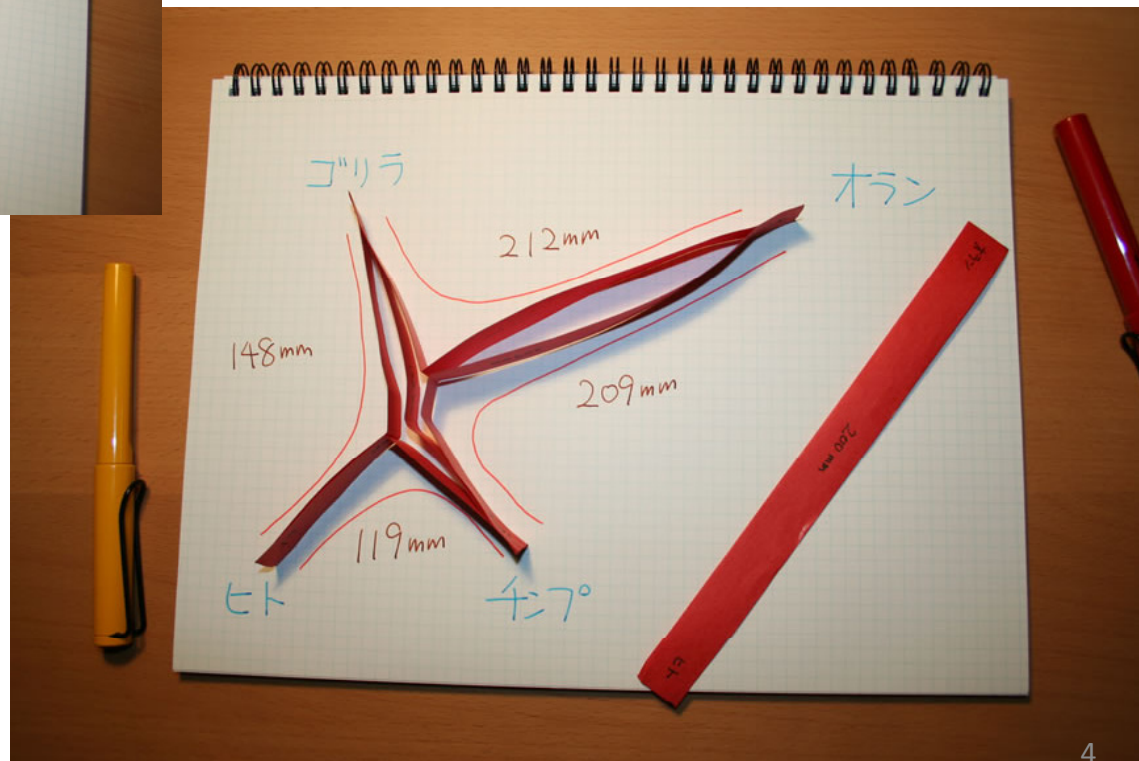
# 次に短い3本を追加する(1本余る)



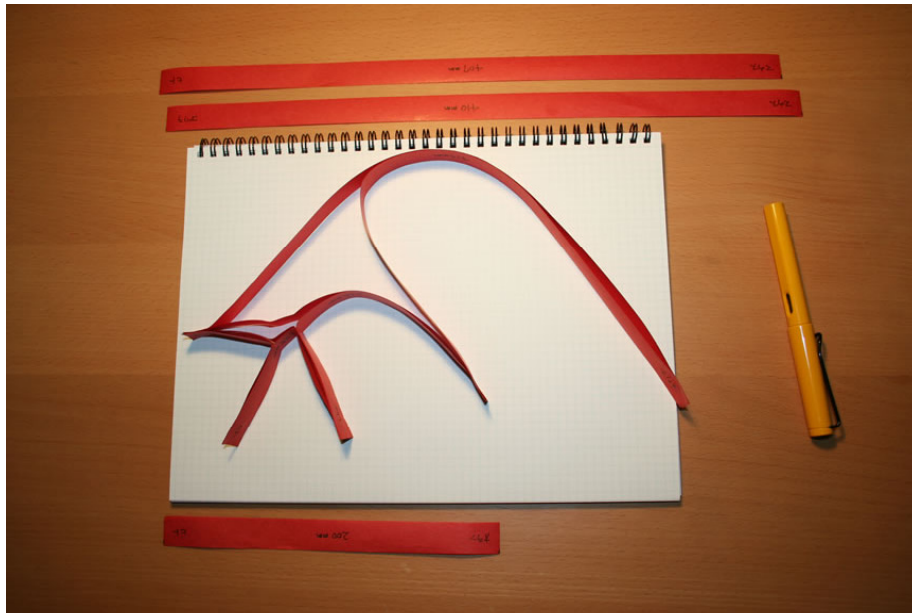
オラン - ヒト

オラン - チンプ

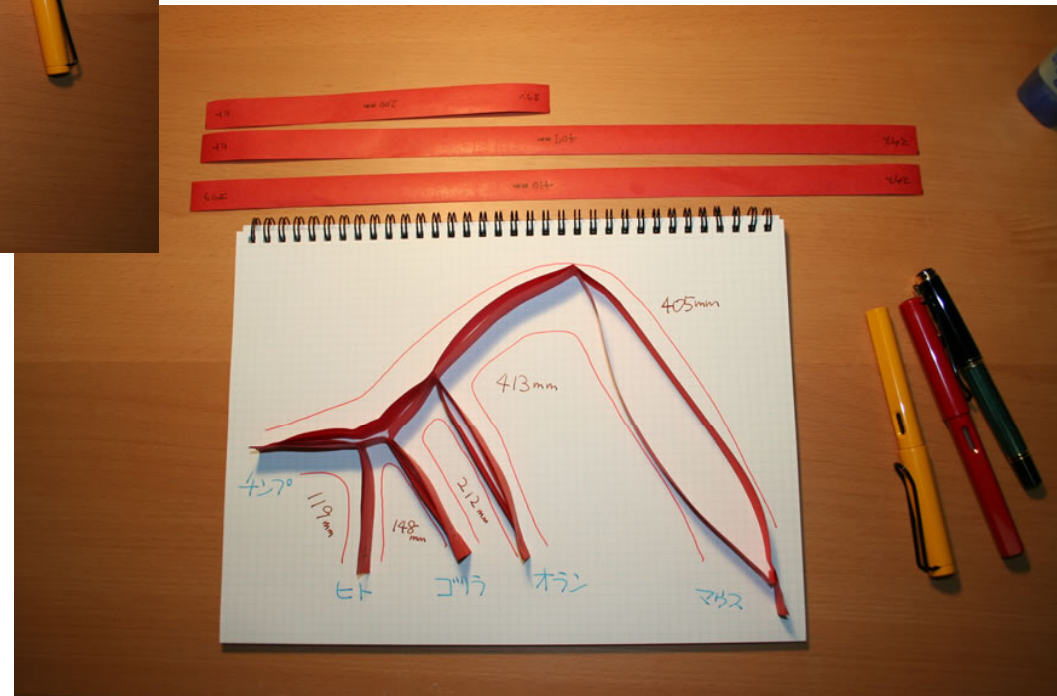
オラン - ゴリラ



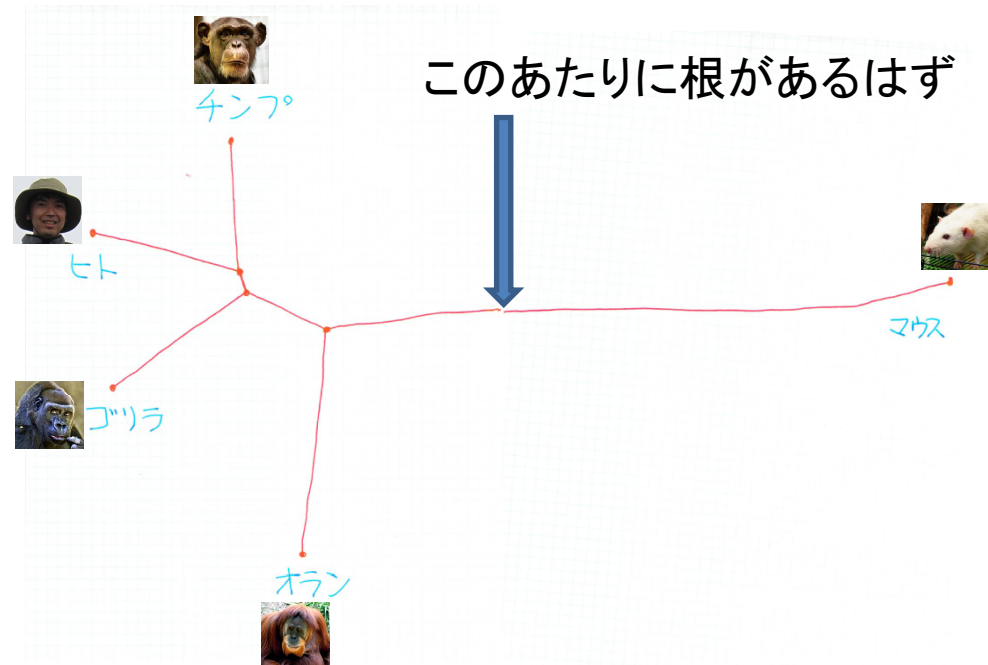
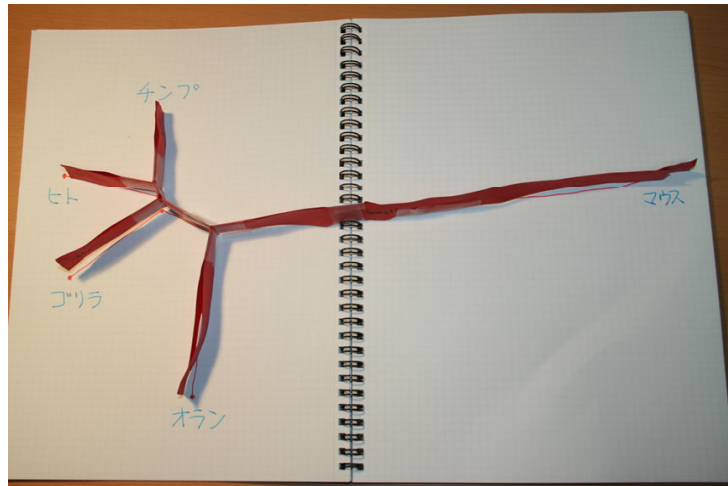
# 最後に長い4本を追加(2本余る)



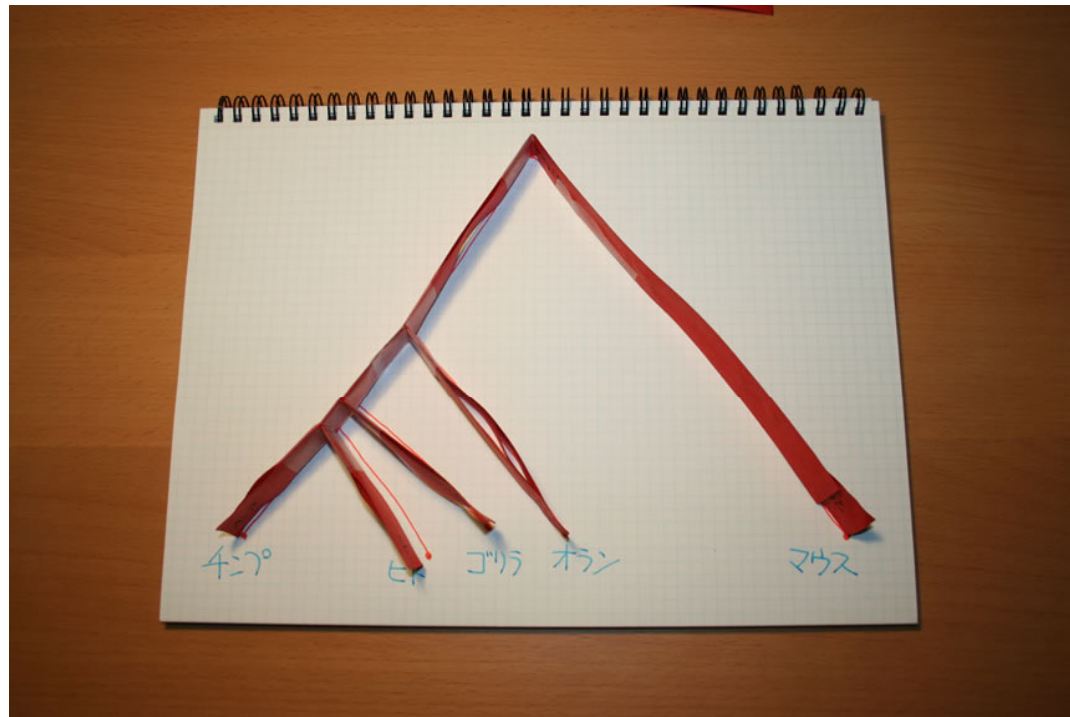
マウス - ヒト  
マウス - チンパンジー  
マウス - ゴリラ  
マウス - オラン



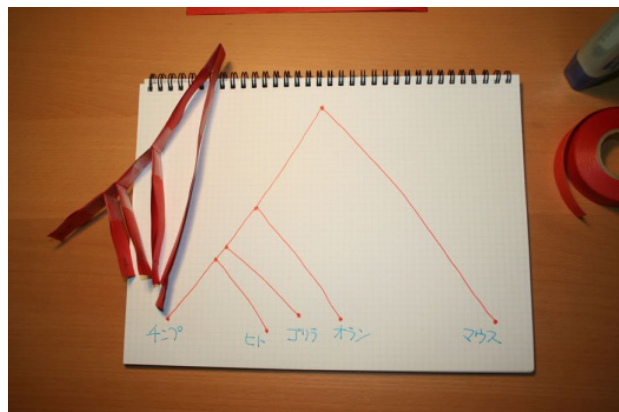
# 無根系統樹



# 系統樹の完成



# 紙に写しておく...



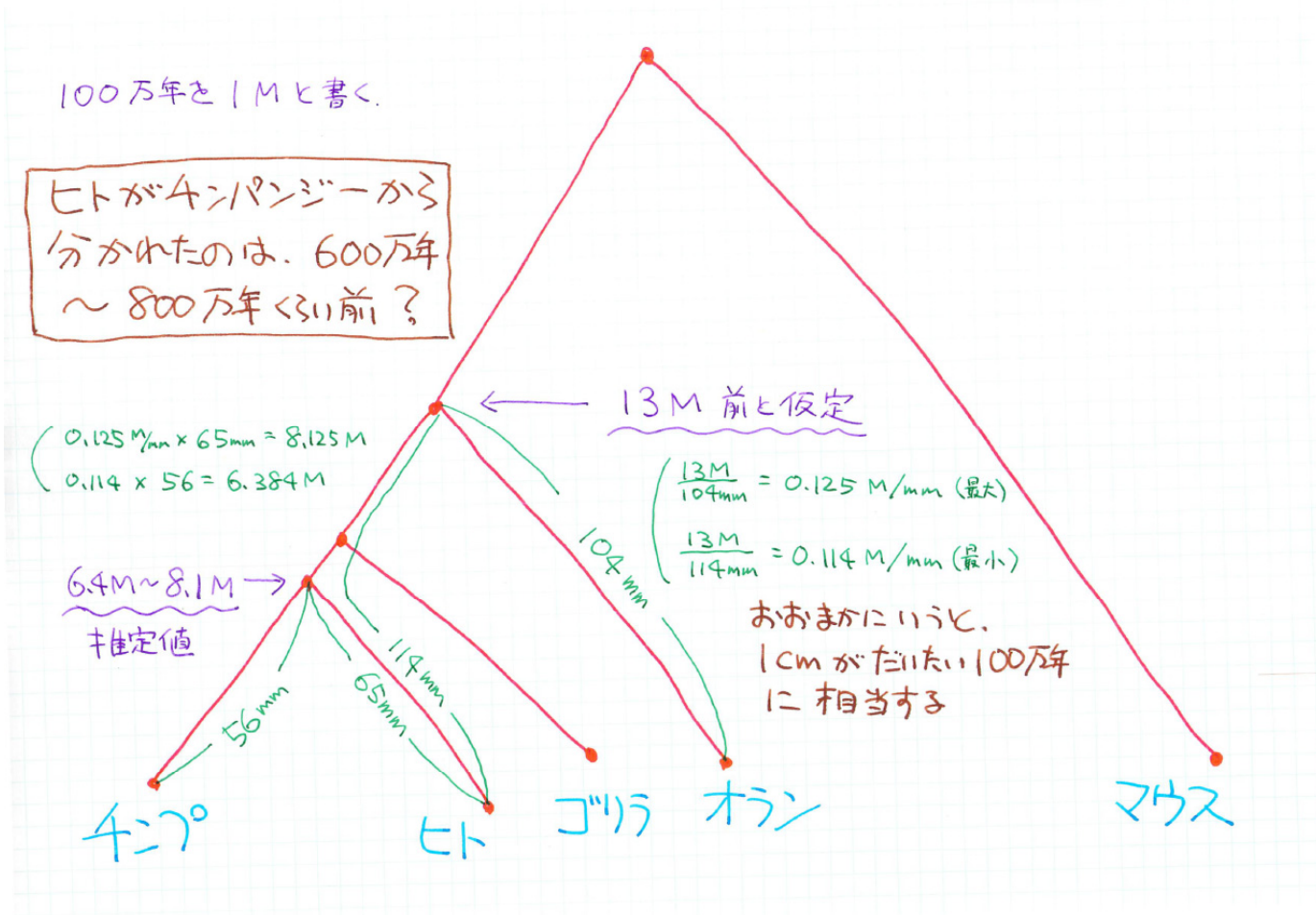
上図は近隣結合法で推定した系統樹

NC\_012920: Homo sapiens: human (the old refseq is NC\_001807)  
NC\_001643: Pan troglodytes: chimpanzee  
NC\_001645: Gorilla gorilla: Western Gorilla  
NC\_002083: Pongo abelii: Sumatran orangutan  
NC\_010339: Mus musculus musculus: eastern European house mouse





# ヒトがチンパンジーから分かれた時期

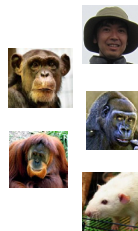


注意： 分岐年代推定の考え方を示すための図です。推定値はあまり正確でなく、改良の余地がいろいろあります。

# 参考：変異数と置換数



変異数 (単純に塩基が異なる個数を数えたもの)



	NC_001807	NC_012920	NC_001643	NC_001645	NC_002083	
NC_012920		18				
NC_001643		1063	1061			
NC_001645		1296	1294	1251		
NC_002083		1698	1703	1763	1780	
NC_010339		3148	3147	3134	3158	3180

Jukes-Cantorモデルで推定した置換数

	NC_001807	NC_012920	NC_001643	NC_001645	NC_002083	
NC_012920		18				
NC_001643		1139	1137			
NC_001645		1412	1409	1358		
NC_002083		1905	1911	1987	2009	
NC_010339		3982	3980	3959	3998	4034

TN93モデルで推定した置換数

	NC_001807	NC_012920	NC_001643	NC_001645	NC_002083	
NC_012920		18				
NC_001643		1188	1185			
NC_001645		1481	1479	1420		
NC_002083		1994	2001	2090	2116	
NC_010339		4069	4068	4052	4102	4127

DNA配列の長さ= $n = 10839$ , 変異数= $C$ ,  
Jukes-Cantorモデルの置換数を $T$ とすれば,

$$\longrightarrow T = -\frac{3n}{4} \log \left( 1 - \frac{4C}{3n} \right)$$