

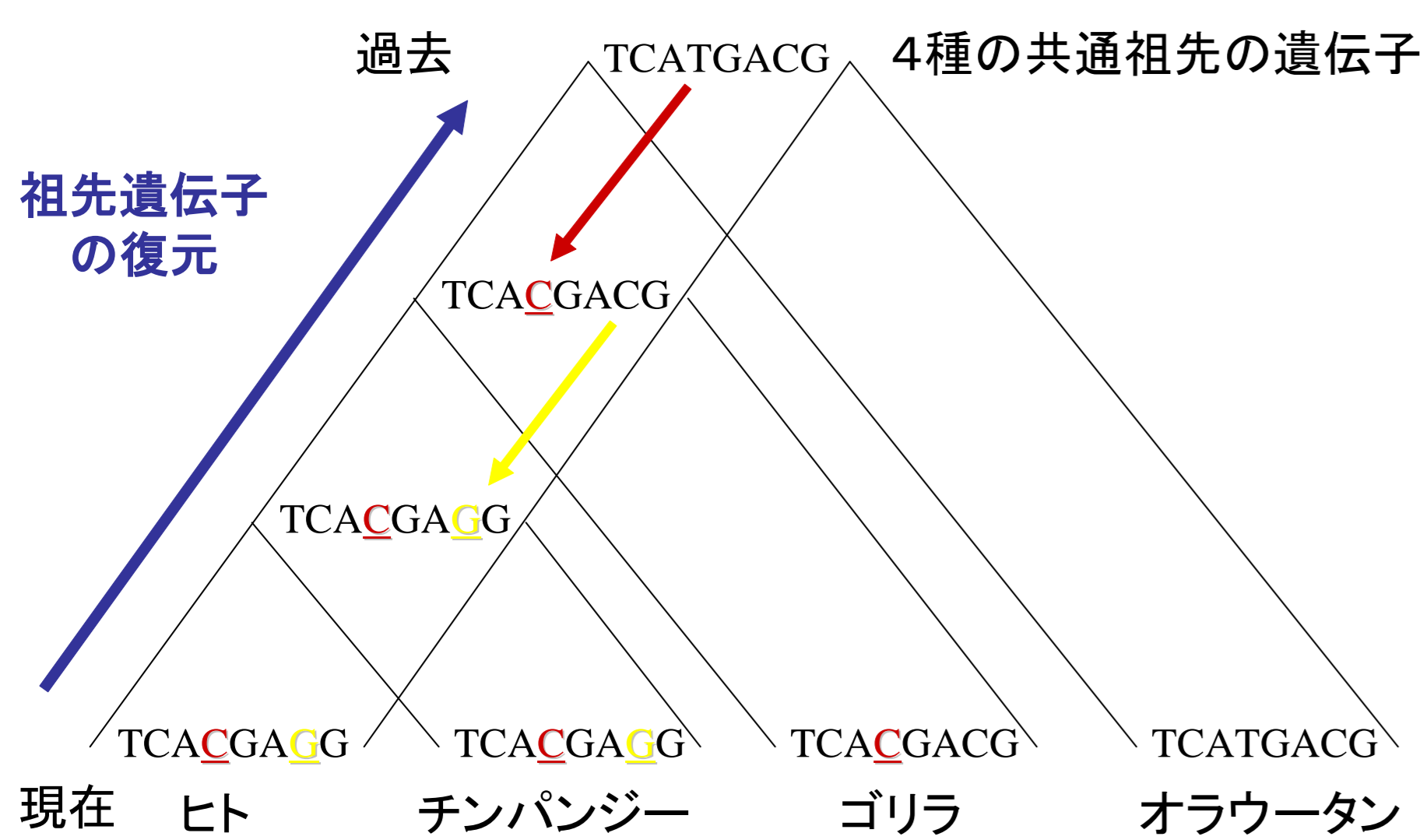
# 分子進化のモデリングと分子系統樹の推定

足立 淳 データ科学研究系・リスク科学NOE ゲノム解析 准教授

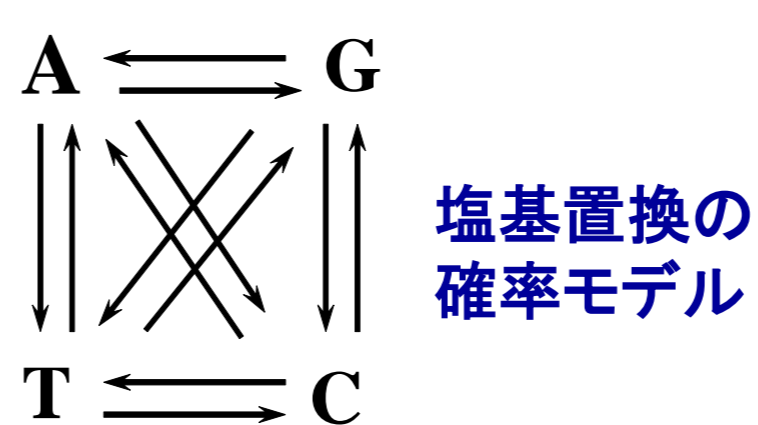
2012年6月15日 統計数理研究所 オープンハウス

## 分子系統学とは

遺伝子の変異差から種の系統を推定

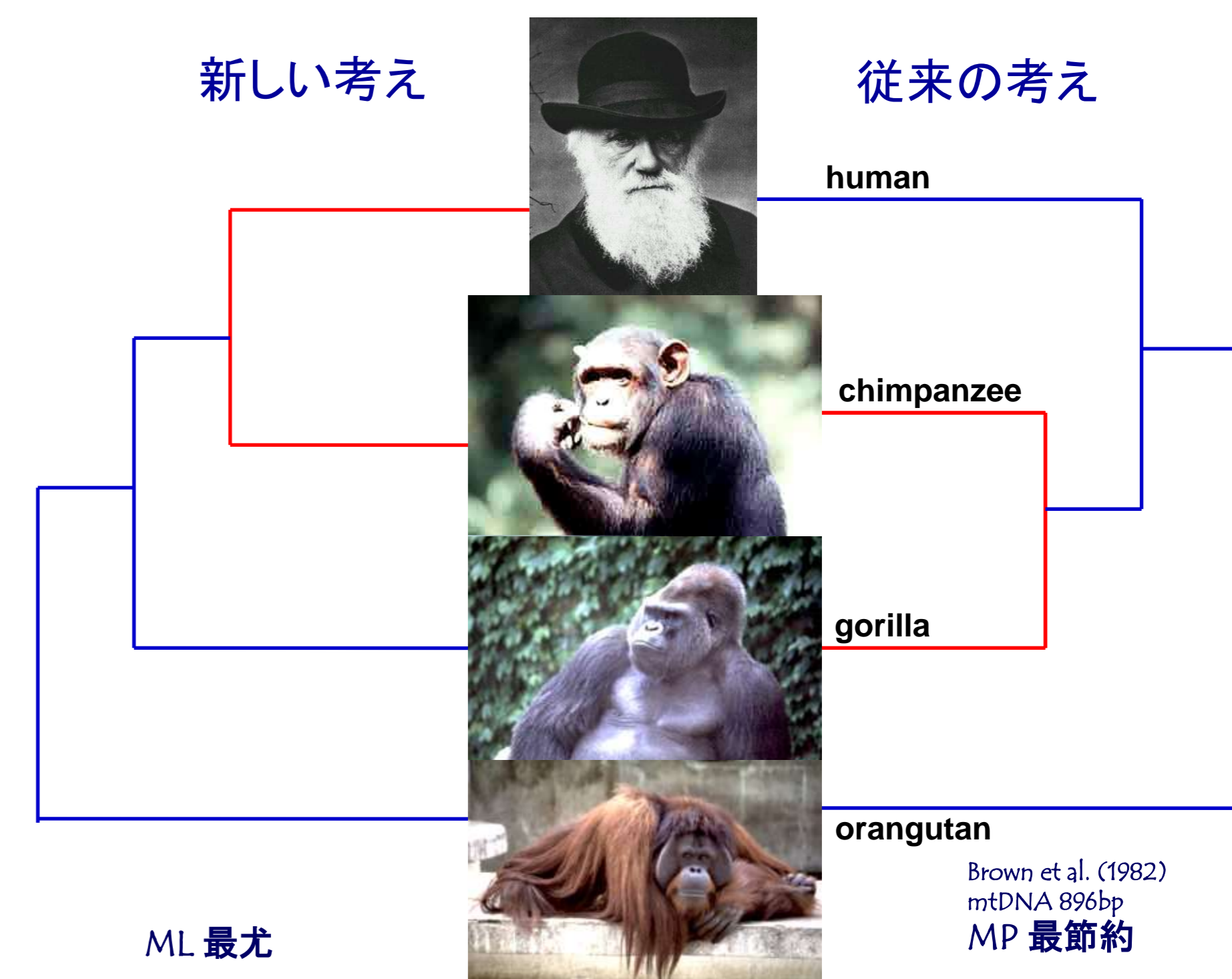


配列の違いは共通祖先から進化してきた結果生じたものであり、進化の歴史を反映している。  
→ 最尤法による系統樹推定

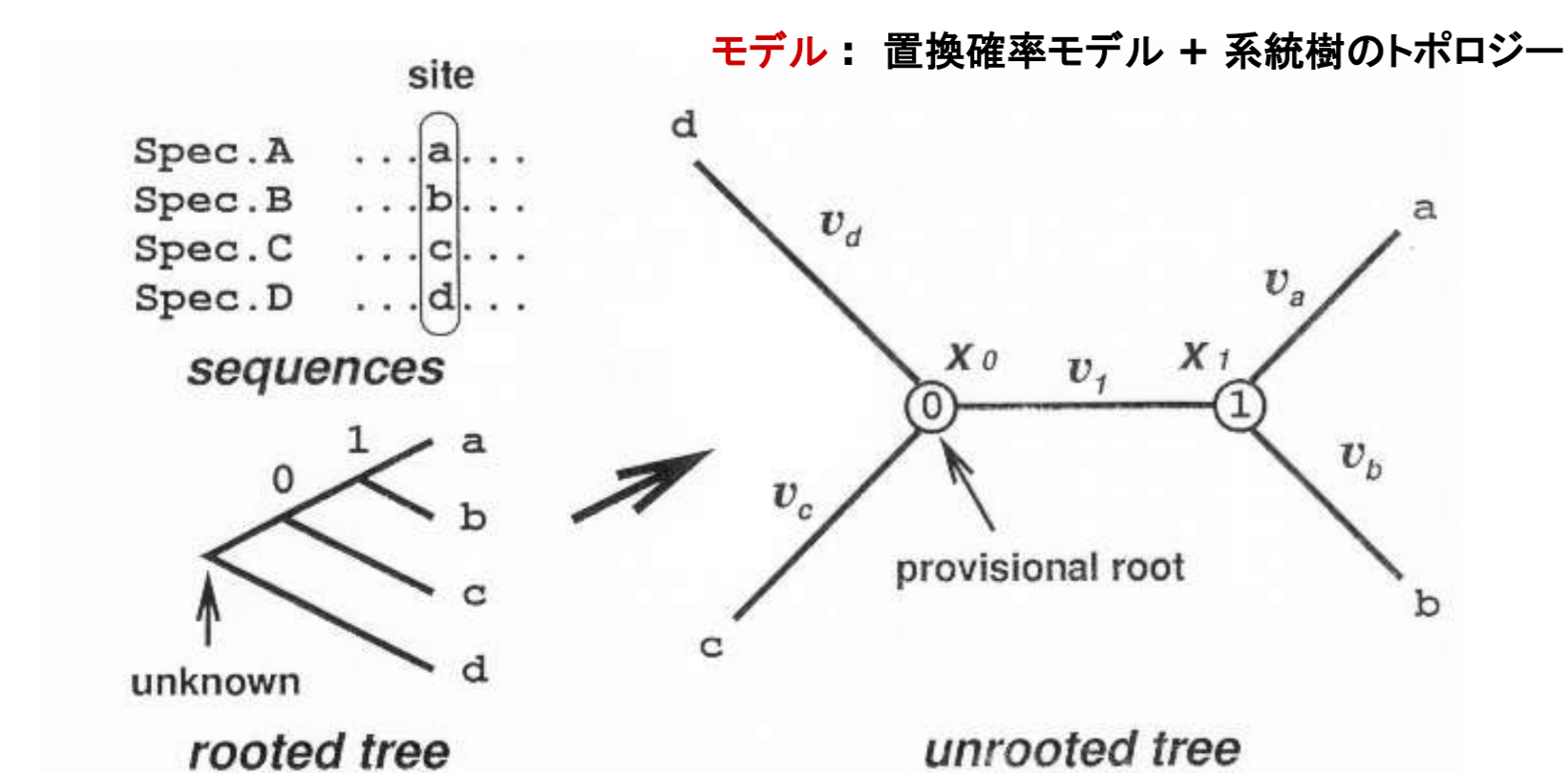


1. Human ヒト
2. Chimpanzee チンパンジー
3. Gorilla ゴリラ
4. Orangutan オランウータン

1 CTAGGCTATATACAACACTACGCAAAGGCCCAACGTTGTAGGCCCTAC  
2 CTAGGCTACATACAACACTACGCAAAGGTCCCAACATTGTAGGTCCTTAC  
3 TTAGGCTATATACAACACTACGTAAGGCCCAACGTCGTAGGCCCTAC  
4 CTAGGCTATACACAACACTACGCAAGGACCTAACATCGTAGGCCCTGC



## 最尤法と尤度(Likelihood) $L = P(\text{data}|\text{model})$



$$L = \sum_{x_0} \pi_{x_0} [P_{x_0c}(v_c) \times P_{x_0d}(v_d) \times \sum_{x_1} P_{x_0x_1}(v_1) P_{x_1a}(v_a) P_{x_1b}(v_b)]$$

## 置換モデル

塩基置換やアミノ酸置換のマルコフモデル  
Transition probability matrix  $P(t)$  during time  $t$   
 $P(t) = e^{Qt}$

- DNAモデル(塩基置換)
  - $A \leftrightarrow G, T \leftrightarrow C$  トランジション
  - $(A,G) \leftrightarrow (T,C)$  トランスバージョン
- タンパクモデル(アミノ酸置換)
  - 経験的に得られた  $20 \times 20$  の置換配列
- コドンモデル
  - アミノ酸置換 と 同義塩基置換 の組み合わせ
  - 深い分岐の系統と浅い分岐の系統を同時推定

## Hasegawa, Yano and Kishino (1984) model (HKY model)

$$Q = \begin{pmatrix} -\mu(\kappa\pi_G + \pi_Y) & \mu\pi_C & \mu\kappa\pi_G & \mu\pi_T \\ \mu\pi_A & -\mu(\kappa\pi_T + \pi_R) & \mu\pi_G & \mu\kappa\pi_T \\ \mu\kappa\pi_A & \mu\pi_C & -\mu(\kappa\pi_A + \pi_Y) & \mu\pi_T \\ \mu\pi_A & \mu\kappa\pi_C & \mu\pi_G & -\mu(\kappa\pi_C + \pi_R) \end{pmatrix}$$

where  $\alpha = \mu, \beta = \mu\kappa, \pi_R = \pi_A + \pi_G,$  and  $\pi_Y = \pi_C + \pi_T$ . This corresponds to

塩基組成の偏りを考慮:  
動物ミトコンドリアDNAのコードン3番目の塩基組成は非常に偏っている  
 $\pi_T=0.169, \pi_C=0.429, \pi_A=0.364, \pi_G=0.038$

## EF1 $\alpha$ のアミノ酸配列: 動物、菌類、植物、原生動物、細菌で共通の配列が見られ

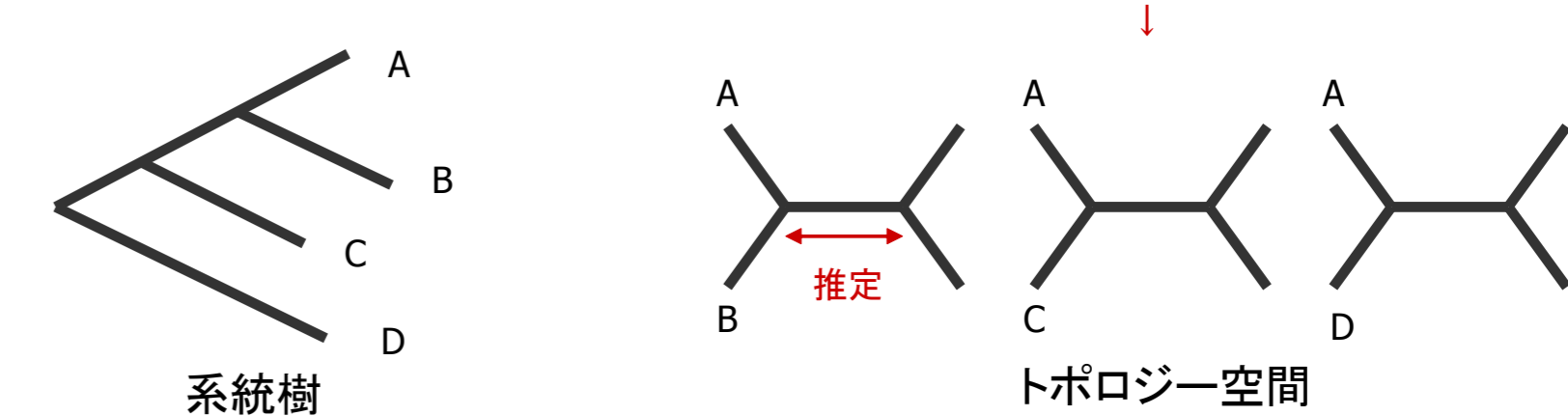
CONSENSUS	STTTTHLLYK	CGHIDKRTIE	KFEKKAEE	G	KSGSKYAWVL	DKLKAEREK	ITDIALWKF	ET	KY	VT	I	DAPGHDEIK
Homo sapiens	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Gallus gallus	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Menopus laevis	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Danio rerio	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Apis mellifera	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Bombus morio	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Onchocerca	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Saccharomyces	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Ashbya gossypii	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Candida albica	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Trichoderma re	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Podospora anse	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Puccinia grani	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Absidia glauca	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Arabidopsis th	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Glycine max	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Hordeum vulgare	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Trichium aesti	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Trichomonas te	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Giardia lamblia	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Hexamita	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Hexamita infie	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Glugea pleocoi	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Solifolobus aci	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Halobacterium	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Methanococcus	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....

## アミノ酸置換モデル mtREV

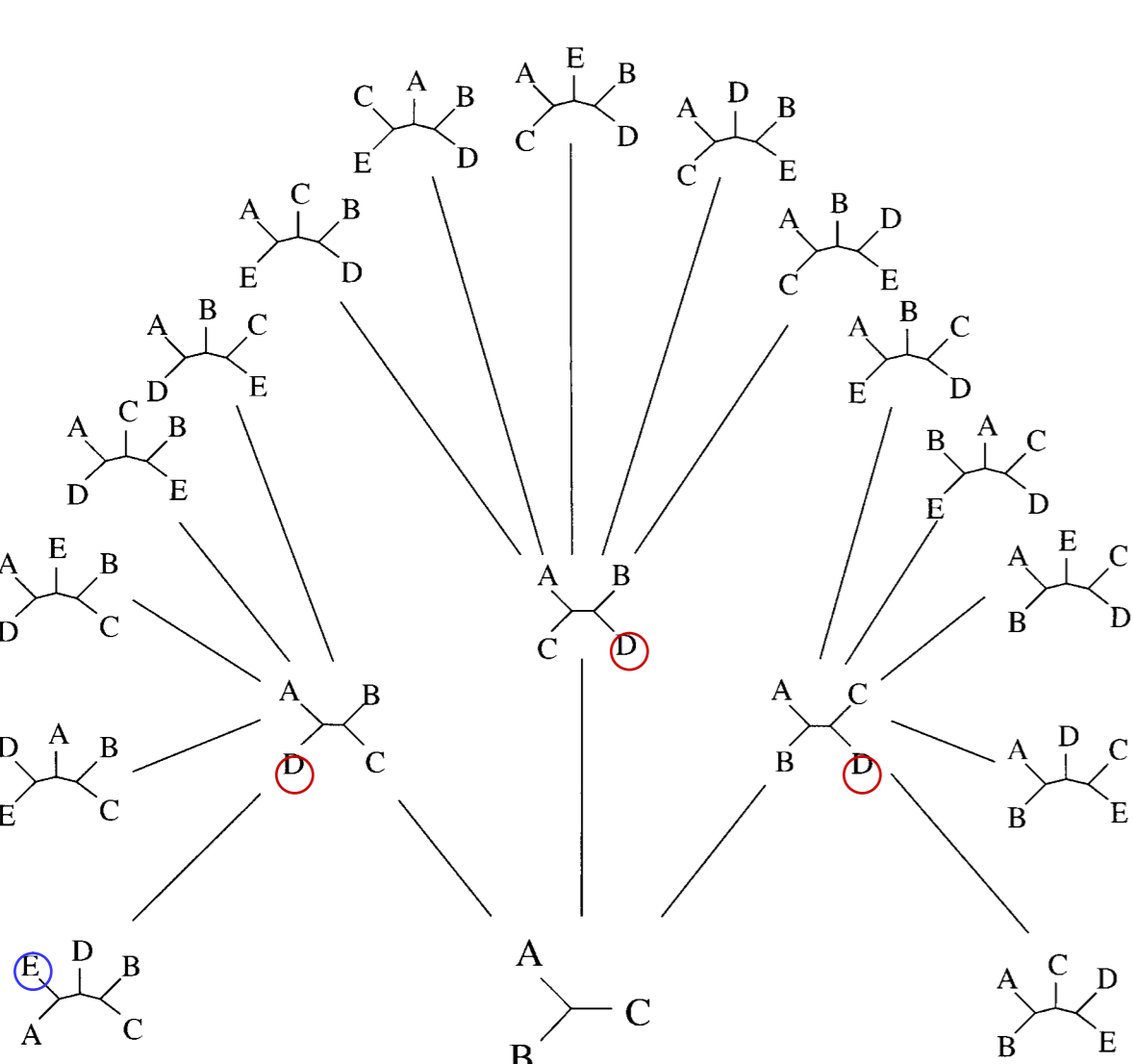
A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	31	135	69	328	41	63	727	84	595	165	1	772	35	321	2229	2078	1	41	818
R	31	135	1	1102	1	152	455	1	71	985	1	34	178	38	34	146	1	51	
N	135	135	4527	30	731	503	314	2656	132	151	2712	266	117	508	2607	1271	67	893	76
D	69	1	527	1	261	3495	334	828	88	4	175	1	35	1	319	98	63	74	1
C	328	1	30	1	242	1	214	723	256	247	1	547	101	1788	1046	222	897	1	
Q	41	1102	731	261	242	1798	83	3208	72	281	2631	349	160	801	447	524	1	260	78
E	63	1	503	3495	1	1798	191	306	1	5	1630	1	1	54	286	138	1	115	137
G	727	152	314	334	214	83	191	1	40	10	89	5	1	54	286	138	1	115	137
H	84	455	2656	828	723	3208	306	1	40	10	89	5	1	54	286	138	1	115	137
I	595	1	132	88	256	72	1	40	43	1875	65	2656	441	117	273	341	295	47	3171
L	165	71	151	4	247	281	5	10	62	1875	65	2656	441	117	273	341	295	47	3171
K	1	985	2712	175	1	2631	1630	89	470	65	37	1	440	52	293	765	808	183	233
M	772	1	266	1	1	349	1	5	13	2656	2938	440	1	546	89	714	2547	209	1745
F	35	34	117	35	547	160	1	1	177	441	1112	52	546	93	310	267	57	2534	34
P	321	178	508	1	101	801	54	1	273	117	184	293	89	93	720	732	33	102	48
S	2229	38	2607	319	1788	447	286	57	651	341	280	388	765	714	310	720	3426	148	218
T	2078	34	1271	98	1046	524	138	73	295	1805	717	808	2547	267	732	3426	125	158	1172
W	1	146	67	63	222	1	1	49	47	1	190	183	209	57	33	148	125	151	35
Y	41	1	893	74	897	260	115	1	3171	200	435	6	297	2534	102	218	158	151	35
V	818	51	76	1	1	78	137	20	1	6030	261	233	1745	34	48	1	1172	35	36

## 系統樹 = トポロジー + 距離

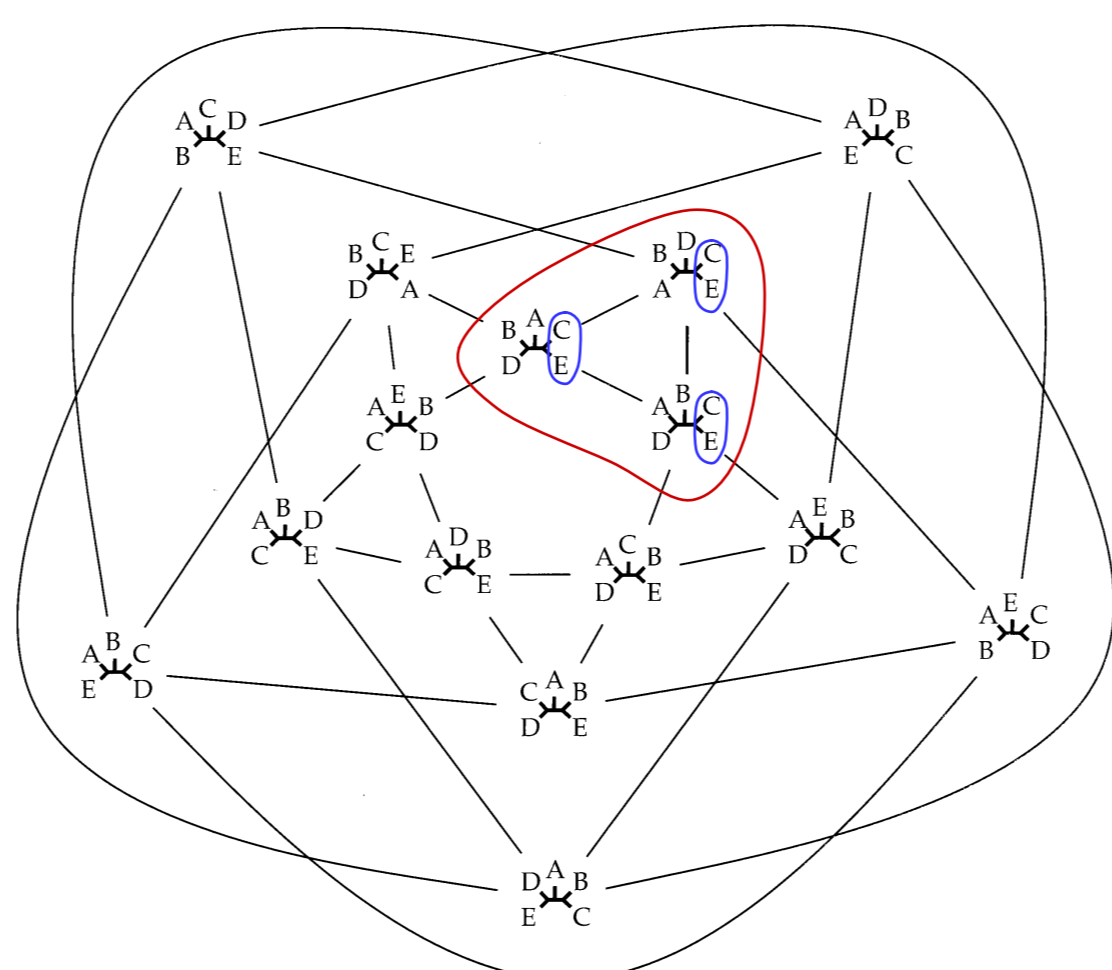
- トポロジー = 枝分かれの順番  
トポロジー空間から探索
- 距離 = 枝の長さ (推定置換数)  
置換率モデルから計算



## 5種によるトポロジー空間 (逐次付加アルゴリズム)

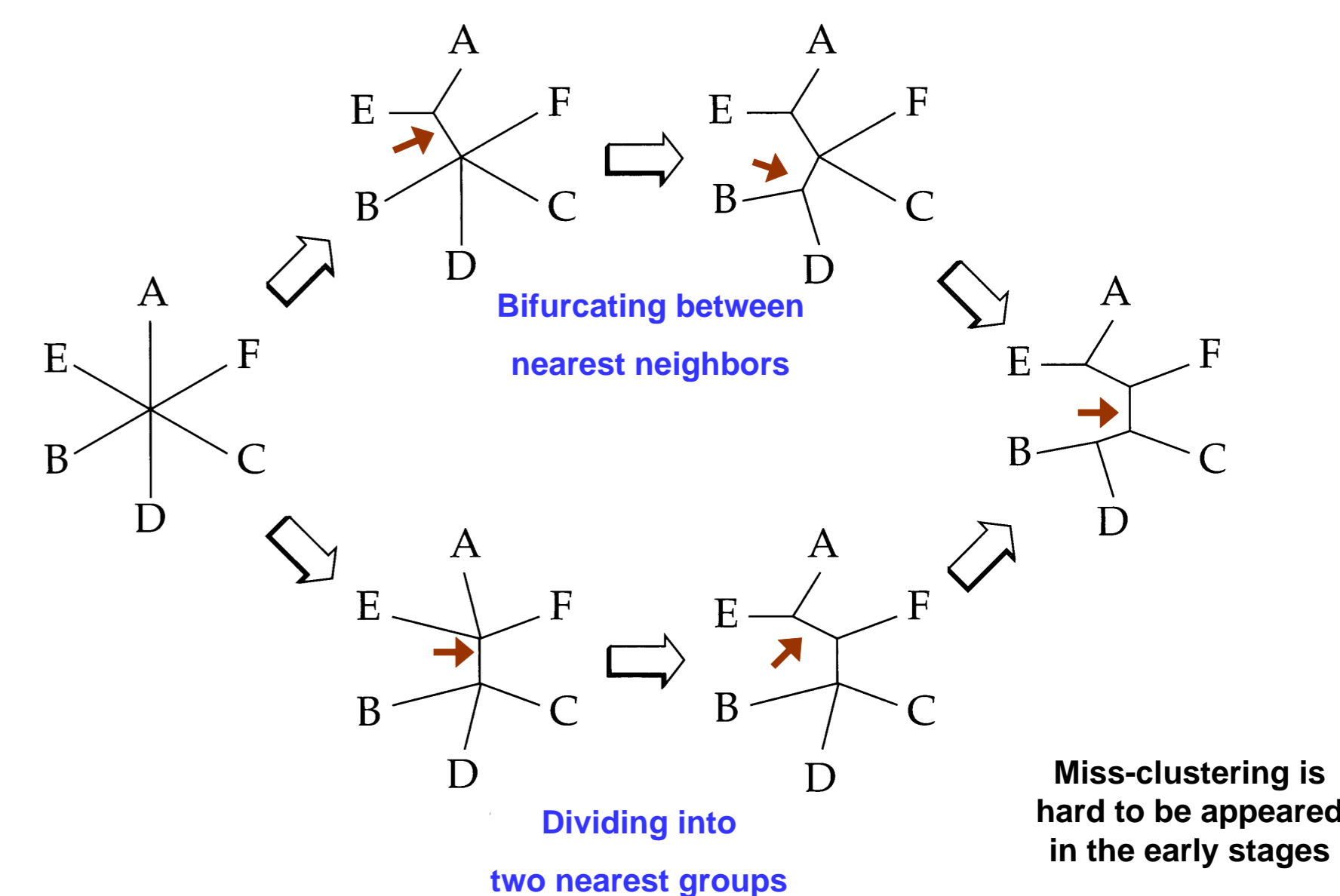


## トポロジー空間と探索法 最近隣木間の関係

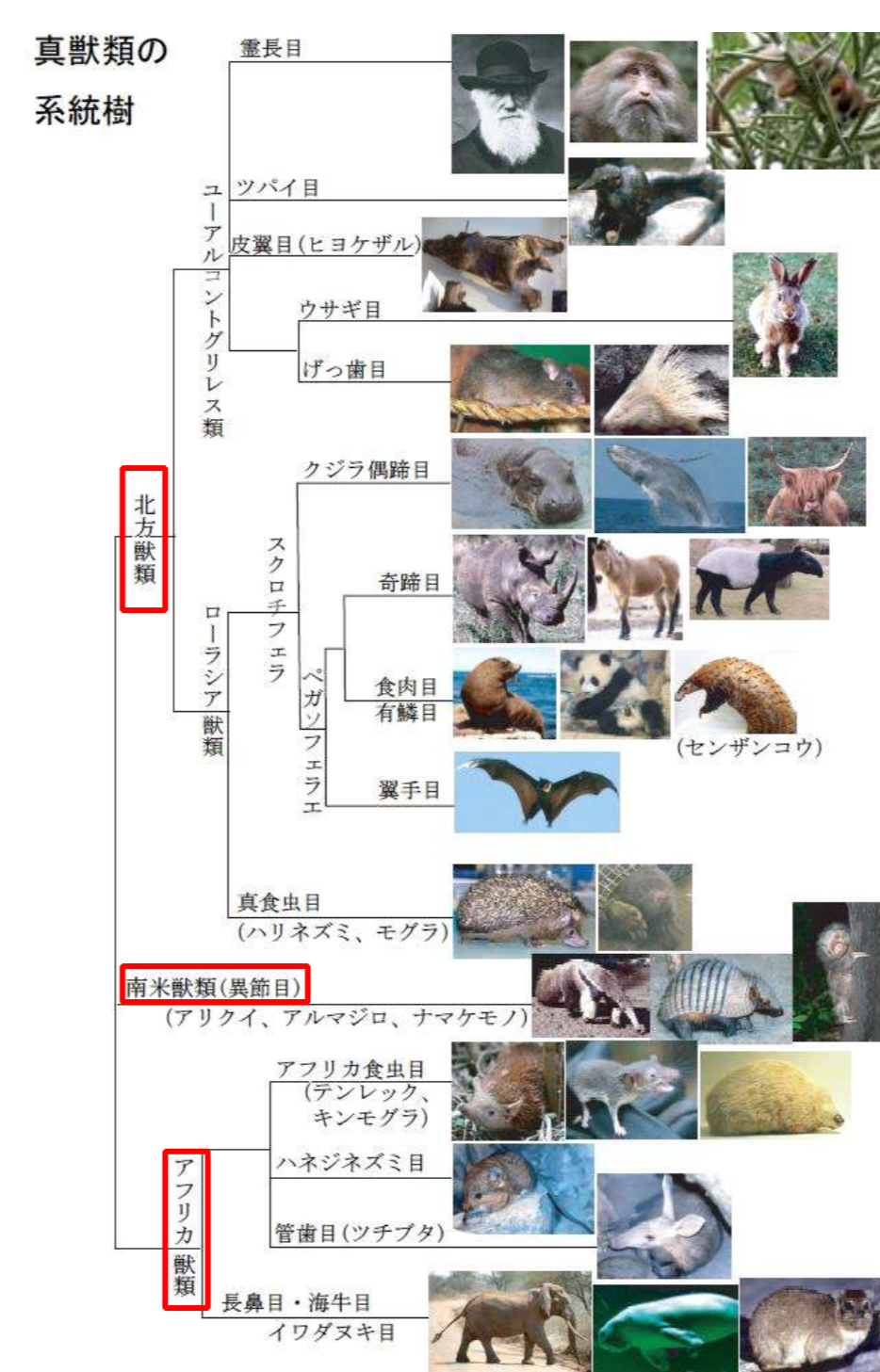


The space of all 15 possible unrooted trees with 5 tips

## New star decomposition



## 1億年前の大陸配置と真獣類の進化



## 遺伝子系統 × ゲノム構造比較 = 進化ダイナミクス

