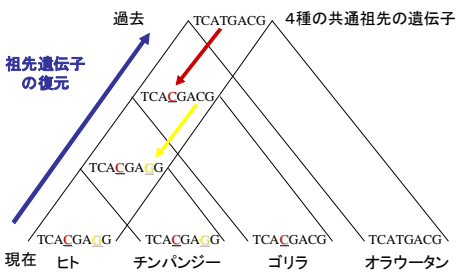


分子進化のモデリングと分子系統樹の推定

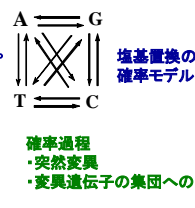
足立 淳 データ科学研究系・リスク科学NOE ゲノム解析 准教授

分子系統学とは

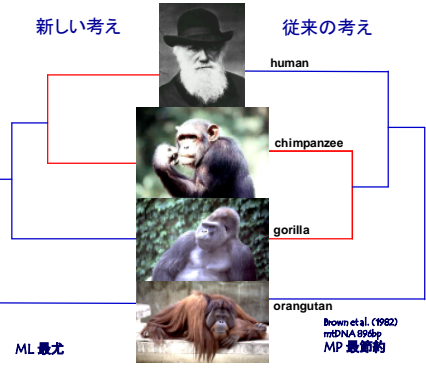
遺伝子の変異差から種の系統を推定



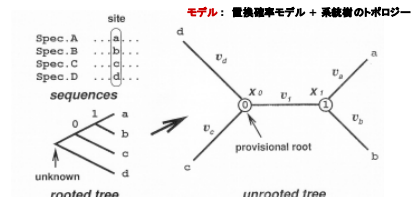
配列の違いは共通祖先から進化してきた結果生じたものであり、進化の歴史を反映している。
→ **最尤法による系統樹推定**



- 1 CTAGGCTATATACAACGCGCAAGGGCCCAACGTTGTAGGCCCTAC
- 2 CTAGGCTATATACAACGCGCAAGGGCCCAACATTGTAGGTCCTTAC
- 3 TTAGGCTATATACAACGCGTAAAGGCCCAACGTCGTAGGCCCTAC
- 4 CTAGGCTATACAGCAACTACGCAAGGGCCTAACATCGTAGGCCCTGCG



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

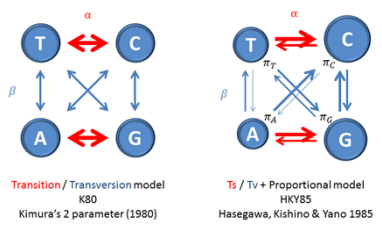


尤度は、与えられたモデルのもとで、データが実現する確率。最尤法ではこれを最大化するモデルを選択。モデルに書まれるパラメータも最尤推定。
$$L = \sum_{x_0} \pi_{x_0} [P_{x_0c}(v_c) \times P_{x_0d}(v_d)] \times \sum_{x_1} P_{x_1a}(v_1) P_{x_1a}(v_a) P_{x_1b}(v_b)$$

置換モデル 塩基置換やアミノ酸置換のマルコフモデル

- 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 20×20 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

塩基置換モデル



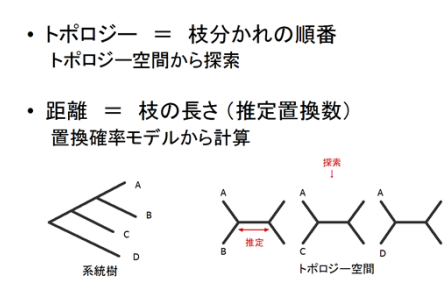
アミノ酸置換モデル mtREV

A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V		
R	31	135	1	13002	1	152	455	1	71	905	1	34	178	98	734	146	1	41	818		
N	135	135	4527	30	731	503	314	2656	132	151	272	266	117	508	2027	1271	62	893	76		
D	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		
C	308	1	30	1	261	3495	334	828	88	4	175	1	35	1	319	98	63	74	1		
Q	41	1102	731	261	242	1798	83	3208	72	281	2631	349	100	801	447	524	240	78	1		
E	63	1	503	3495	1	1798	191	306	1	5	1630	1	1	1	54	286	138	1	115	137	
G	727	152	314	334	214	83	191	1	1	40	10	89	5	1	1	651	79	49	1	20	
H	84	455	2656	828	723	3208	306	1	1	43	62	470	13	177	273	341	295	47	3171	1	
I	955	1	132	88	256	72	1	40	43	1875	65	2656	441	117	280	1805	1	200	6030	1	
L	365	71	151	4	247	281	5	10	62	1875	5	10	62	1875	5	10	62	1875	5	10	
K	1	985	2712	175	1	2631	1630	89	470	65	37	1	1	440	52	293	765	808	183	233	6
M	772	1	266	1	1	349	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
F	35	34	117	35	547	160	1	1	177	441	1132	52	546	1	1	1	1	1	1	1	1
P	21	178	508	1	101	801	54	1	273	117	184	293	89	93	1	1	1	1	1	1	1
S	2209	38	2007	319	1788	447	286	651	141	280	388	765	714	310	720	3426	148	218	1	1	1
T	2078	34	1271	98	1046	524	138	73	295	1805	717	808	2547	267	732	3426	125	158	1132	1	1
W	1	146	67	63	222	1	1	49	47	1	190	183	209	57	33	148	125	151	35	1	1
Y	41	4	893	74	897	260	115	1	3171	200	231	239	297	2534	102	118	158	151	36	1	1
V	818	51	76	1	78	137	20	1	6030	465	6	1745	34	48	1	1172	35	36	1	1	1

コドンモデル CTAへ9通りの置換

code	Anticodon	code	Anticodon	code	Anticodon	code	Anticodon
TTT	TCT	TAT	Y Tyr	TGT	C Cys	T	
TTC	TCC	TAC	Y Tyr	TGC	C Cys	A	
TTA	TCA	TAA	--- TGA	---	---	C	
TTG	TCG	TAG	--- TGG	W Trp	G		
CTT	CCT	CAT	CAT	CGT	T		
CTC	CCC	CAA	H His	CGC	C		
CTA	CCA	CAA	CAG	CGA	R Arg	A	
CTG	CCG	CAG	Q Gln	CGG	T		

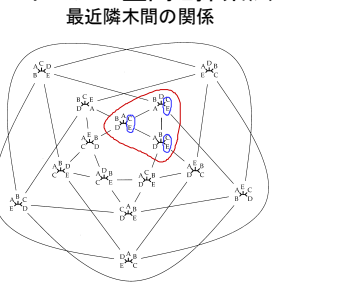
系統樹 = トポロジー + 距離



1億年前の大陸配置と真獣類の進化

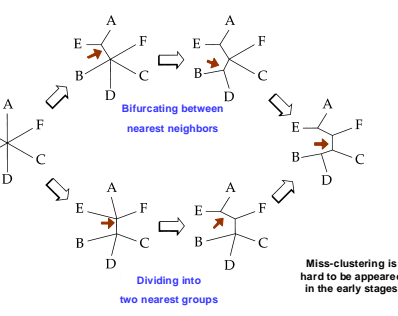


トポロジー空間と探索法



The space of all 15 possible unrooted trees with 5 tips

New star decomposition



遺伝子系統×ゲノム構造比較=進化ダイナミクス

