

生物数学

齐木 里惠

生物数学について・・・

これから話す内容の中では、分子生物学の中で、DNAやタンパク質などを扱いやすくするために数学的に表す。

生物学を学ぶ人にとっては新しい考え方や発見ができる。

数学を学ぶ人にとっては数学が実際にどのようなように活用されているかを知ることができる。

大まかな流れ

DNAのアライメント(整列)



DNA間の距離の推定



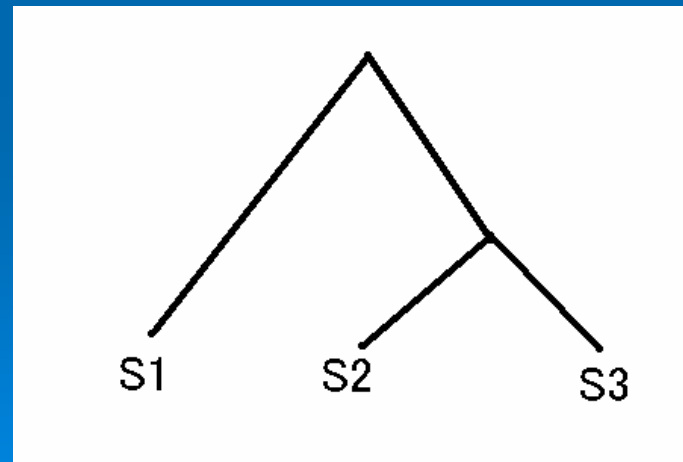
進化系統樹の作成

例えば...

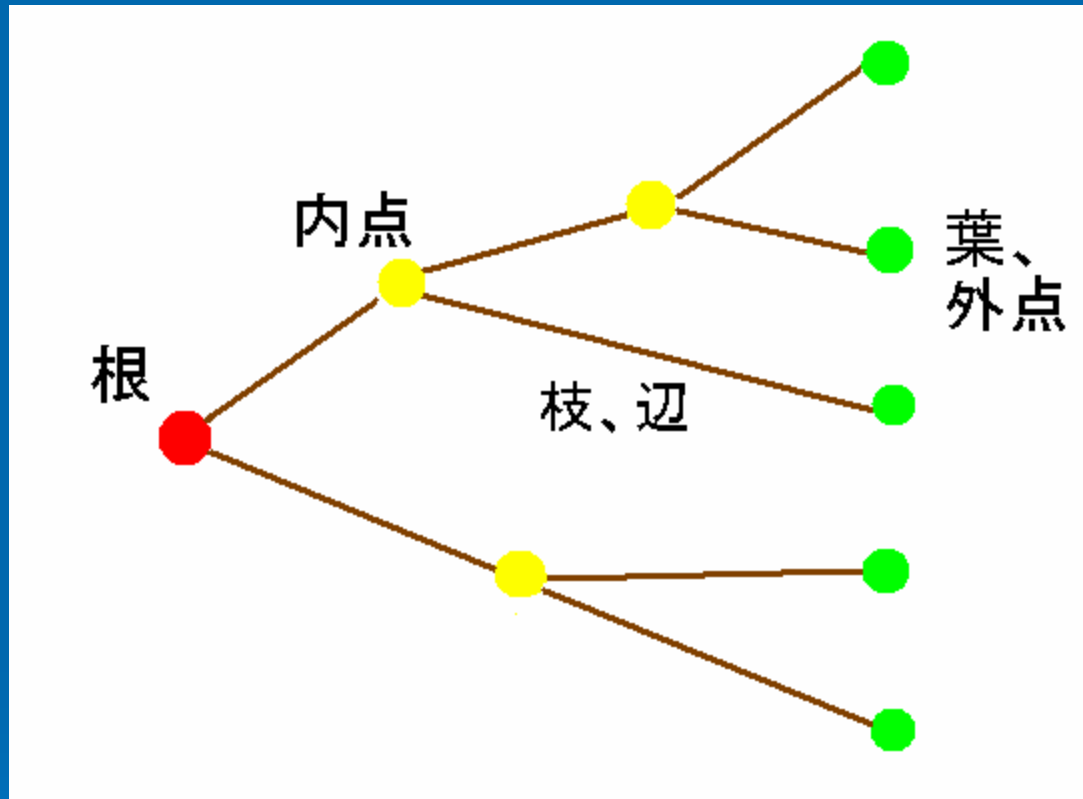
S1:ATTGCGTATGCC

S2:ATGCCTAAGCC

S3:ATGCCTAAGGC



進化系統樹

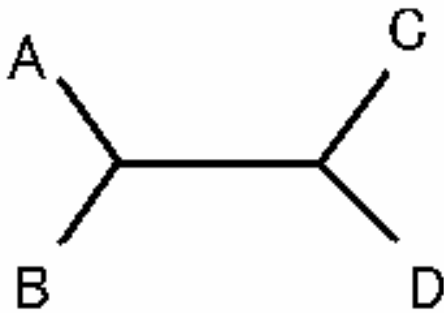


<大原則> Bifurcating (分岐する)

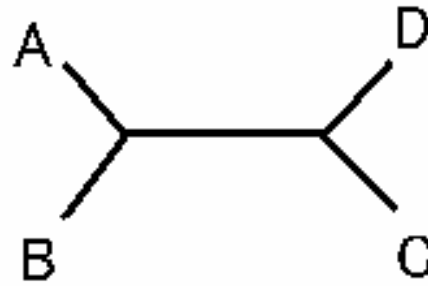
1つの内点に3つのedgeが集結する。
(根には2つのedgeが集結する。)

位相的に同じ

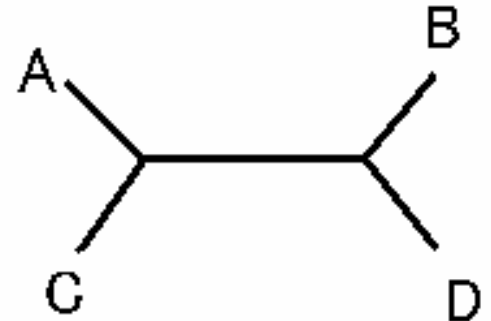
①



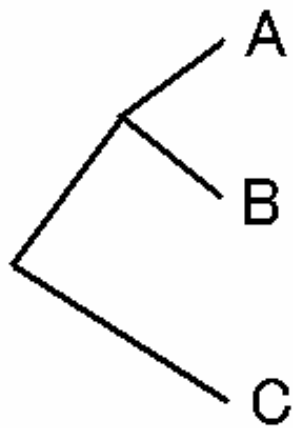
②



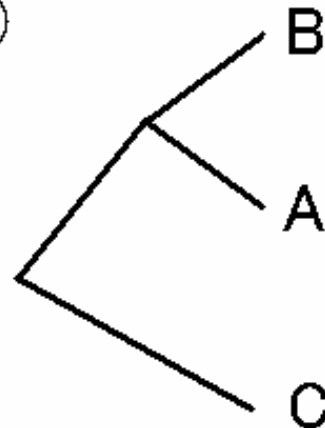
③



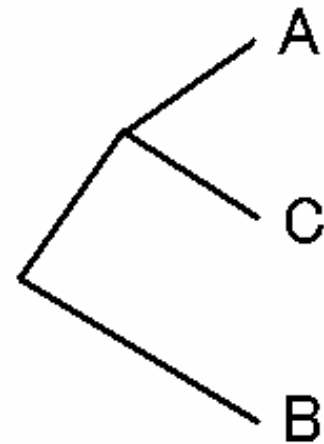
①



②



③



3つの進化系統樹作成法

U P G M A

FitchMargoliashアルゴリズム

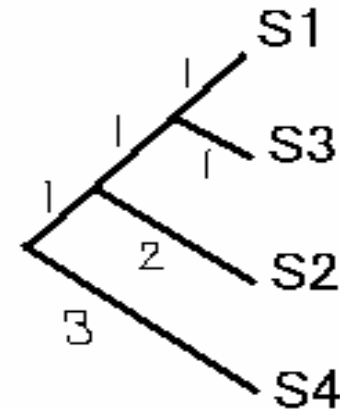
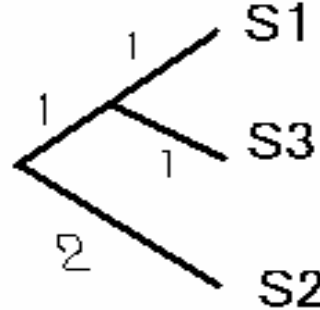
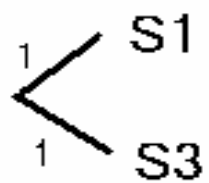
近隣結合法

UPGMA

※分子時計を仮定する。

- ①一番近い2つのtaxaを結合。
- ②その2つのtaxaを1つとみなして、また一番近いtaxaを結合。
- ③これを繰り返す。

	S1	S2	S3	S4
S1		4	2	6
S2			4	6
S3				6
S4				



3点の公式

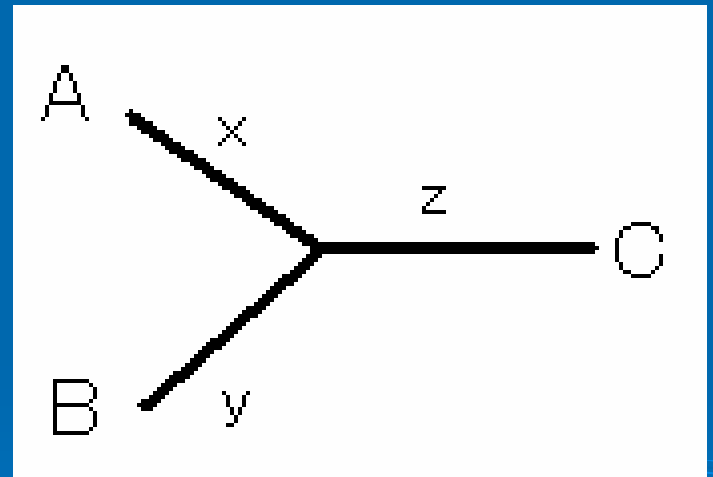
3点A, B, Cについて、それぞれの2点間の距離がわかっているとき、

$$x = (d_{AB} + d_{AC} - d_{BC}) / 2$$

$$y = (d_{AB} + d_{BC} - d_{AC}) / 2$$

$$z = (d_{AC} + d_{BC} - d_{AB}) / 2$$

が成り立つ。

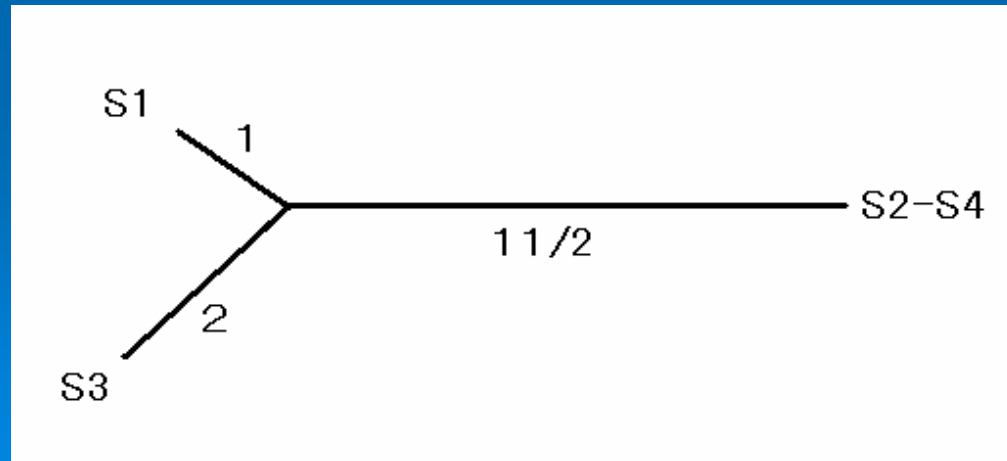


FMアルゴリズム

	S1	S2	S3	S4
S1		6	3	7
S2			7	7
S3				8
S4				

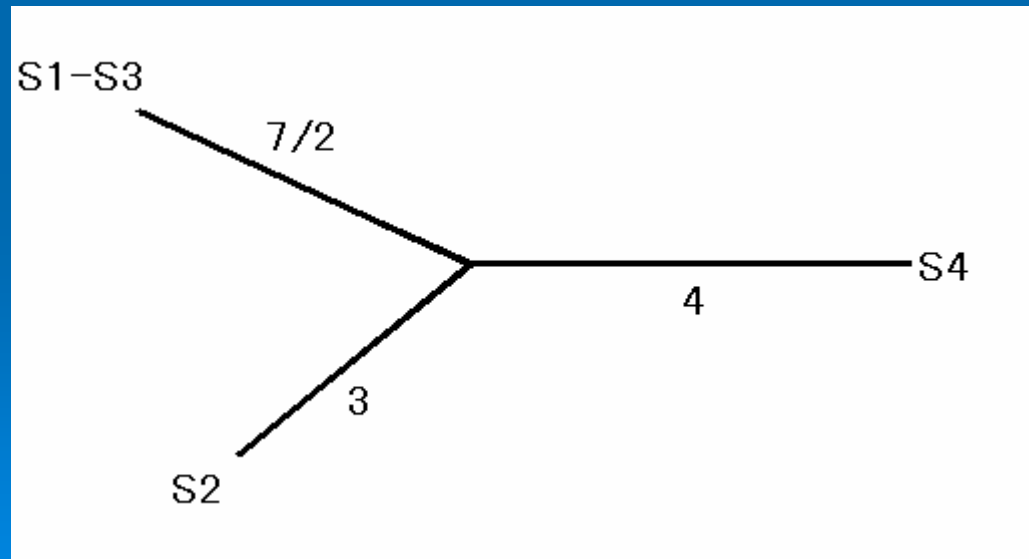
※分子時計を仮定せず、よりデータに近い系統樹を作成。

①1番近い2つのtaxaとそれ以外のすべてのtaxaについて3点の公式を使う。

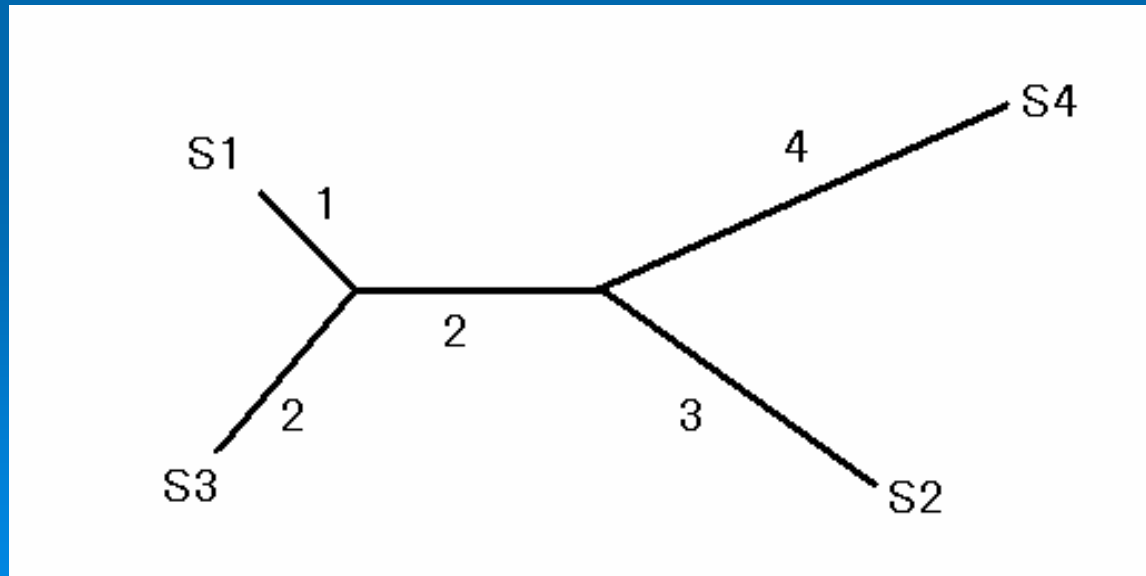
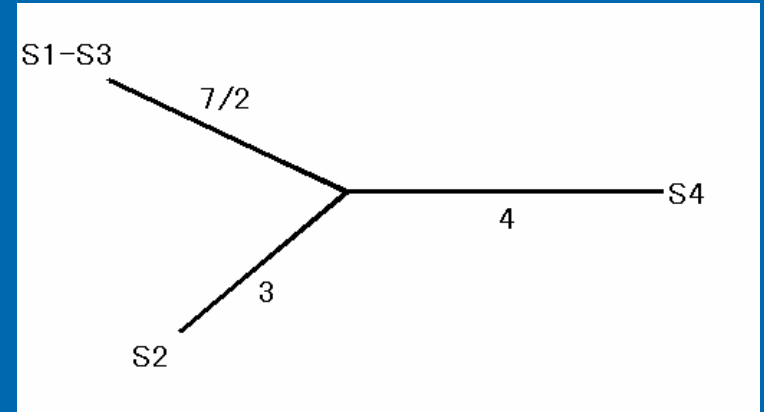
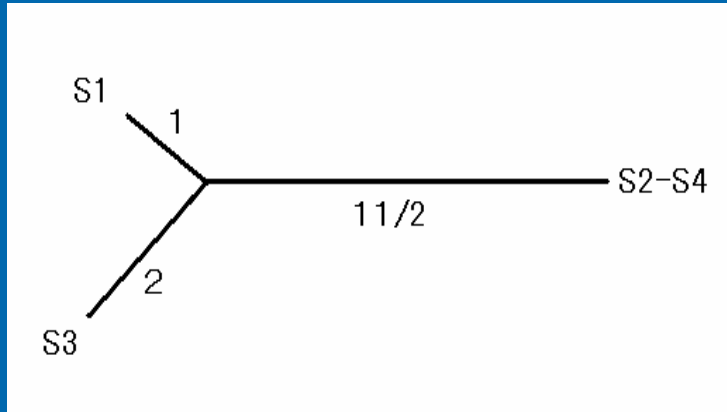


	S1-3	S2	S4
S1-3		13/2	15/2
S2			7
S4			

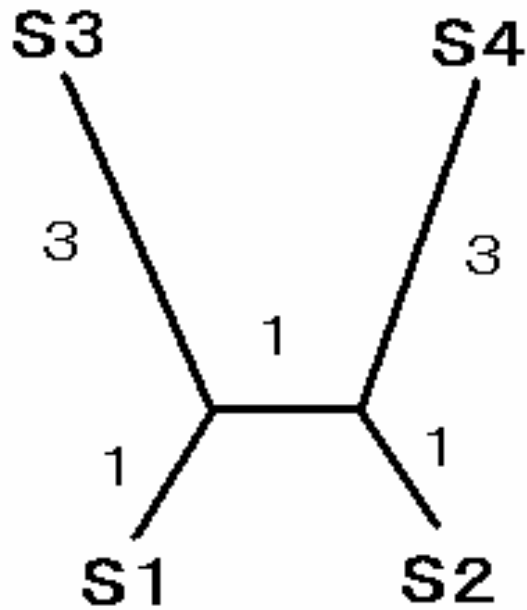
②さっきの2つのtaxaを1つのグループにし、もう1度表を作り直す。そのあとはまた①の手順を繰り返す。



完成



近隣結合法



neighborsではない。

UPGMAやFMアルゴリズムでは距離が1番近いtax aを選んでいた。

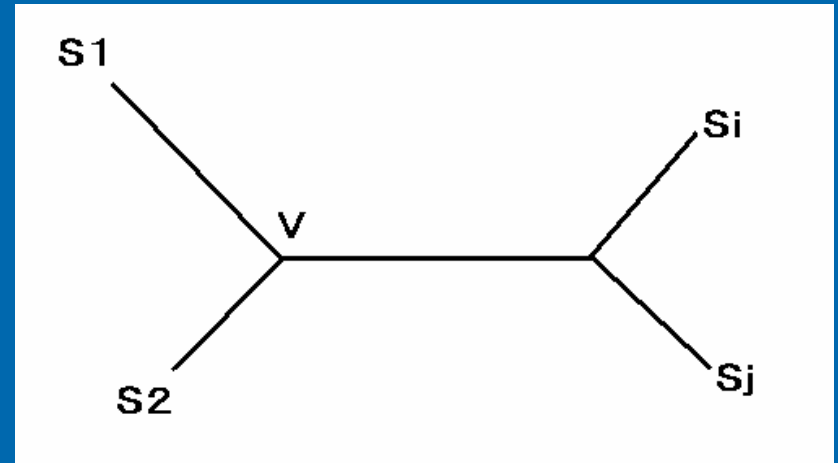
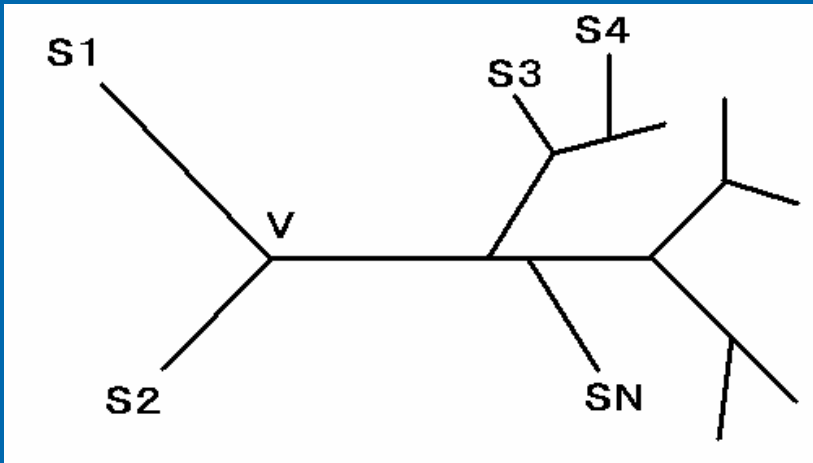


左図の場合S1とS2を結合させてしまう。



neighborsを結合させるような方法が必要。

neighborsなら...



$$d(S1, S2) + d(Si, Sj) < d(S1, Si) + d(S2, Sj)$$

↓

$$(N-3)d(S1, S2) + \sum_{j=3, j \neq i} d(Si, Sj)$$

$$< (N-3)d(S1, Si) + \sum_{j=3, j \neq i} d(S2, Sj)$$

↓

$$(N-2)d(S1, S2) + R_1 < (N-2)d(S1, Si) + R_2$$

$$(N-2)d(S1,S2) + R_1 < (N-2)d(S1,Si) + R_2$$



$$(N-2)d(S1,S2) - R_1 - R_2 < (N-2)d(S1,Si) - R_1 - R_i$$



$$M(S1,S2) < M(S1,Si)$$

つまり、

S_n と S_m がneighborsなら、全ての $k \neq m$ に対して、

$$M(S_n,S_m) < M(S_n,S_k)$$

が成り立つ。

近隣結合法の手順

- ① N 個のtaxaに対し、 M の値の表を作る。 M の値が最も小さくなる S_i 、 S_j を選ぶ。
- ② S_i 、 S_j 以外のtaxaを1つのグループ G とする。3点の公式を使って $d(S_i, V)$ 、 $d(S_j, V)$ を求める。
- ③ $S_k \in G$ に対して3点の公式を使って $d(S_k, V)$ をそれぞれ求める。
それを新しい距離データの表とする。
- ④ これらの操作をtaxaが3つになるまで繰り返す。
3つになったら3点の公式を使って終わり。

大まかな流れ

DNAのアライメント(整列)



DNA間の距離の推定



進化系統樹の作成

塩基置換の行列モデル

初期の塩基配分 $\cdots P_0 = (P_A, P_G, P_C, P_T)$

条件付確率 $\cdots P_{ij} = P(S_1 = i | S_0 = j)$

遷移行列

$$M = \begin{pmatrix} P_{A|A} & P_{A|G} & P_{A|C} & P_{A|T} \\ P_{G|A} & P_{G|G} & P_{G|C} & P_{G|T} \\ P_{C|A} & P_{C|G} & P_{C|C} & P_{C|T} \\ P_{T|A} & P_{T|G} & P_{T|C} & P_{T|T} \end{pmatrix}$$

1 単位時間後の塩基配分

$$MP_0 = P_1$$

マルコフモデル

マルコフ行列とは、成分が全て0以上で、列の和が1の行列。

※成分が正のマルコフ行列はいつも狭義最大固有値 $\lambda = 1$ を持つ。

※ $\lambda = 1$ に対応する固有ベクトルはただ1つ存在する。

∴フロベニウスの定理。

Jukes-Cantorモデル

全ての塩基置換が同じ確率で起こる。

$$P_0 = (1/4, 1/4, 1/4, 1/4)$$

その変異率を α とすると、

$$M = \begin{matrix} 1-\alpha & \alpha/3 & \alpha/3 & \alpha/3 \\ \alpha/3 & 1-\alpha & \alpha/3 & \alpha/3 \\ \alpha/3 & \alpha/3 & 1-\alpha & \alpha/3 \\ \alpha/3 & \alpha/3 & \alpha/3 & 1-\alpha \end{matrix}$$

木村2, 3パラメータモデル

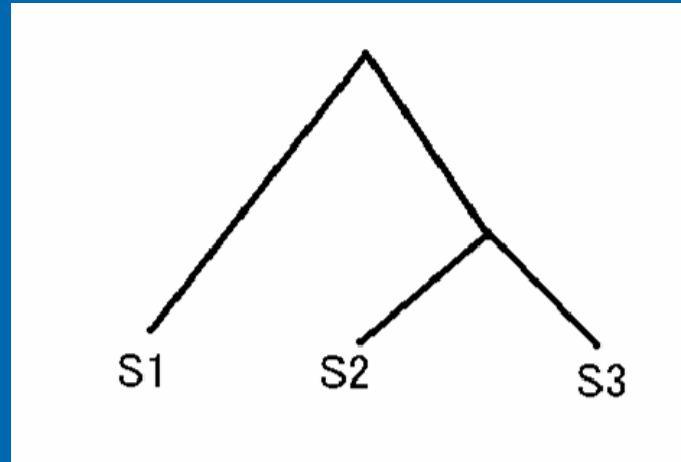
木村2パラメータモデル

$$\begin{array}{cccc} * & \beta & \gamma & \gamma \\ M = \beta & * & \gamma & \gamma \\ \gamma & \gamma & * & \beta \\ \gamma & \gamma & \beta & * \end{array} \quad * = 1 - \beta - \gamma$$

木村3パラメータモデル

$$\begin{array}{cccc} * & \beta & \gamma & \delta \\ M = \beta & * & \delta & \gamma \\ \gamma & \delta & * & \beta \\ \delta & \gamma & \beta & * \end{array} \quad * = 1 - \beta - \gamma - \delta$$

Jukes-Cantor距離



$$d = t \times \alpha$$

$$= (\text{経過時間}) \times (\text{変異率})$$

$$= (\text{経過時間}) \times (\text{siteごとの塩基置換の数})$$

$$= (\text{経過時間内でのsiteごとの塩基置換の数})$$

dは経過時間に比例する。dを進化の距離ととらえる。

いろいろな距離公式

Jukes-Cantor距離

$$d_{JC}(S_0, S_1) = -3/4 \ln(1 - 4p/3)$$

木村2パラメータ距離

$$d_{K2}(S_0, S_1) = -1/2 \ln(1 - 2p_1 - p_2) \\ - 1/4 \ln(1 - 2p_2)$$

木村3パラメータ距離

$$d_{K3}(S_0, S_1) = -1/4 (\ln(1 - 2\beta - 2\gamma) \\ + \ln(1 - 2\beta - 2\delta) + \ln(1 - 2\gamma - 2\delta))$$

Log-det距離

$$d_{LD}(S_0, S_1) = -1/4 (\ln(\det(F)) - 1/2 \ln(g_0 g_1))$$

大まかな流れ

DNAのアライメント(整列)



DNA間の距離の推定



進化系統樹の作成