

## 多重配列アライメントの作成と分子系統樹の推定法

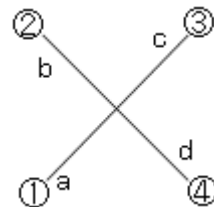
CLUSTALX が計算する多重配列アライメントは、配列間の遺伝子系統樹の推定に利用できます。下の4つの架空アミノ酸配列を元に、多重配列アライメントの作成と遺伝子系統樹の推定法について触れてみましょう。

- ① CHIBAKEN
- ② SHIGAKEN
- ③ SAGAKEN
- ④ SHIMANEKEN

多重配列のアライメント計算も、配列対の場合と同じく動的プログラミングによって計算できます。しかし、 $n$ 本の配列のアライメント作成には  $n$ 次元格子を使用する必要があり、配列数の増加に伴い計算量は指数関数的に増大するため実用的な方法ではありません。CLUSTALX では数百配列の解析を実現するため、累積法といってまず 1) 全ての配列対アライメントを計算し、2) 配列間アライメントスコアを基に近隣結合(NJ)法で系統関係を作成し、3) その系統関係に従って動的プログラミングによるアライメントを累積的に進めるといった近似的な解析方法で、多重配列アライメントを推定しています。

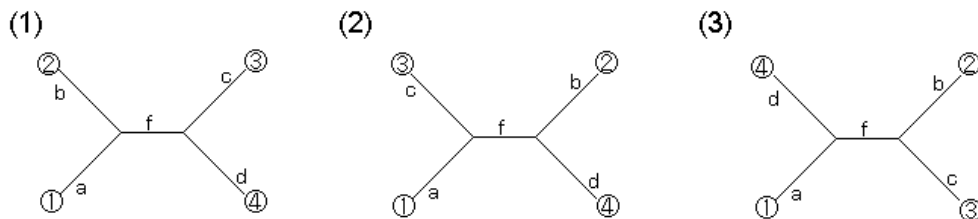
- 1) 上の4つの配列について配列対毎のアライメント計算から配列対の進化距離が計算できます(この際 CLUSTALX ではデフォルトとして PAM250 というスコア行列を用います。PAM250 は進化に伴うアミノ酸置換頻度を見積もったスコア行列です)。

	①	②	③	④
CHIBAKEN	① -			
SHIGAKEN	② 0.360	-		
SAGAKEN	③ 0.626	0.159	-	
SHIMANEKEN	④ 0.360	0.159	0.360	-



- 2) この進化距離を基に案内木(guide tree)を作成します。NJ法ではまず、右図のような星状の無根案内木から出発します。①～④は各配列に相当し、a,b,c,d は中心からの分岐の枝の長さを表します。よって  $S_0 = a + b + c + d$  は案内木の枝長の合計となります。例えば①と②の距離  $a + b$  は上の進化距離行列の  $d_{①②} = 0.360$  に相当します。つまり、 $S_0 = (0.360 + 0.626 + 0.360 + 0.159 + 0.159 + 0.360)/3$  になります。一般に、 $N$ 配列のとき  $S_0 = \sum d_{ij} / (N-1)$  で、 $d_{ij}$  は配列  $i$  と  $j$  ( $i < j$ ) の間の距離です。

次に星状案内木の分解(変形)を考えます。どれか対の配列(近隣)を取り出し樹形を変形させます。上例の4配列の場合、近隣の取り出し方は下の(1)～(3)になります(4配列の場合は、2つを取り出すと残りの2つも自動的に近隣になってしまう)。



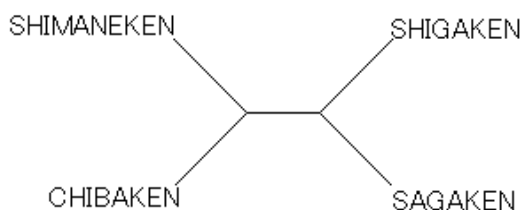
(1)~(3)の場合について枝長の合計を考えます。

$$(1) S_{12} = S_{34} = (d_{13} + d_{14} + d_{23} + d_{24})/4 + d_{12}/2 + d_{34}/2$$

$$(2) S_{13} = S_{24} = (d_{12} + d_{14} + d_{32} + d_{34})/4 + d_{13}/2 + d_{24}/2$$

$$(3) S_{14} = S_{23} = (d_{12} + d_{13} + d_{42} + d_{43})/4 + d_{14}/2 + d_{23}/2$$

(1)~(3)で枝長の合計が最小になるのは(3)でその値は\_\_です。(3)を案内木(guide tree)として採択します。



3) NJ 法によって得た案内木に従って、近隣の順に累積的に配列対アライメントを作成し、多重配列アライメントを推定します。

					*			*	*	*
①	C	H	I	B	A	-	-	K	E	N
②	S	H	I	G	A	-	-	K	E	N
③	S	-	A	G	A	-	-	K	E	N
④	S	H	I	M	A	N	E	K	E	N
	1	2	3	4	5	6	7	8	9	10

この多重配列アライメントは最大を 10 アミノ酸とし、5,8,9,10 の位置のアミノ酸は全配列で保存されていることを示しています。

CLUSTALW では推定した多重配列アライメントを基に、NJ 法による遺伝子系統樹を作成できます。

(注意) CLUSTALX は多重配列アライメント作成時に拡張子 dnd のファイルとして出力します。これは上述したアライメント計算用の案内木であり、アライメント結果から推定した遺伝子系統樹(拡張子 ph や phb がつく)とは別物です!!

NJ 法や累積法などのより詳しい計算方法については以下の文献を参照してください。

『分子生物学のためのバイオインフォマティクス入門』(共立出版)

『分子進化と分子系統学』(培風館)、『分子進化遺伝学』(培風館)