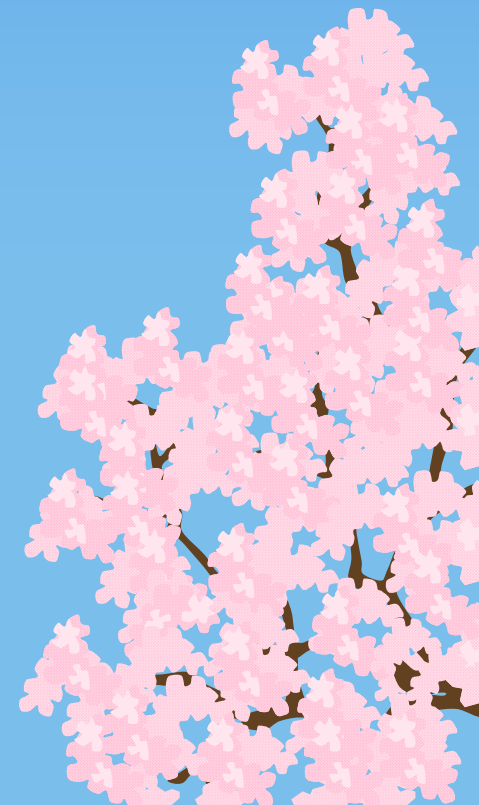


動的計画法を用いたアライメント

0240501 小菅孝史



アラインメントとは何か？

- ❖ 配列を要素ごとに対応づける事、比較
- ❖ 配列の類似度を計算する
- ❖ 2本 → ペアワイズアラインメント
- ❖ 3本以上 → マルチプルアラインメント
- ❖ アラインメントでは、配列要素をそのまま対応づけるだけでなく、進化の過程で生じ得る配列要素の挿入・欠損を扱うため、ギャップを対応づける

グローバル(全体的)アラインメントと ローカル(局所的)アラインメント

❖ 配列1: MIGMMIT

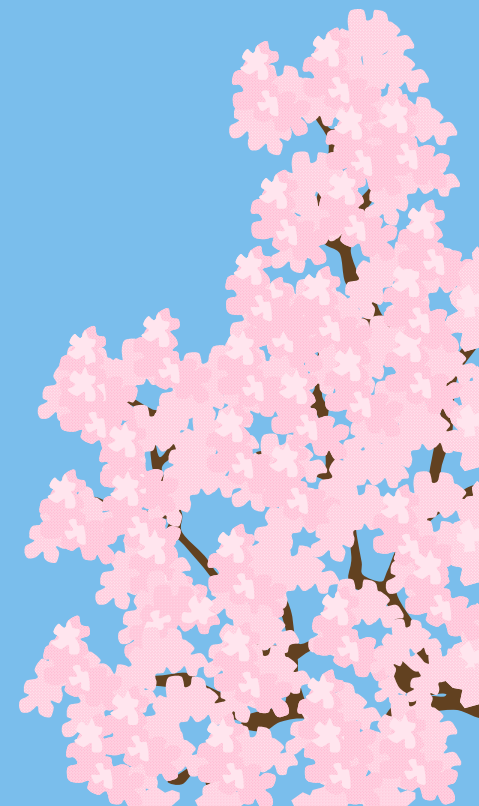
❖ 配列2: MMIGPIT

❖ 全体的に見ると・・・M-IGMMIT

MMIGP-IT

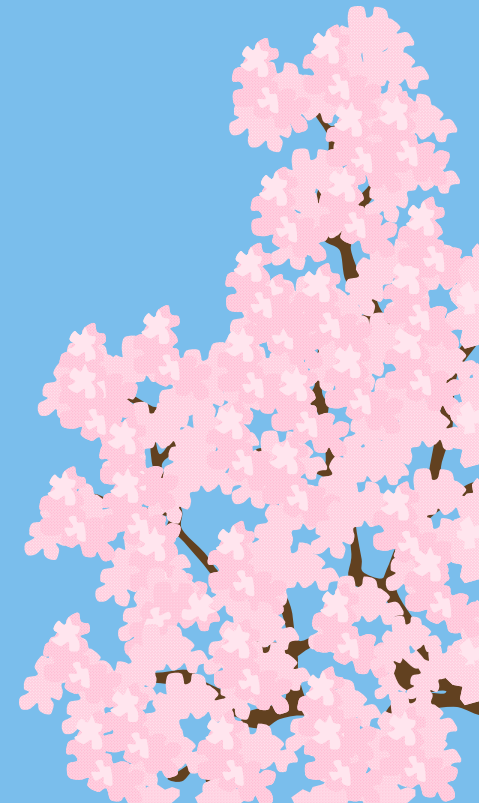
局所的に見ると・・・MIGMMIT

---MMIGPIT



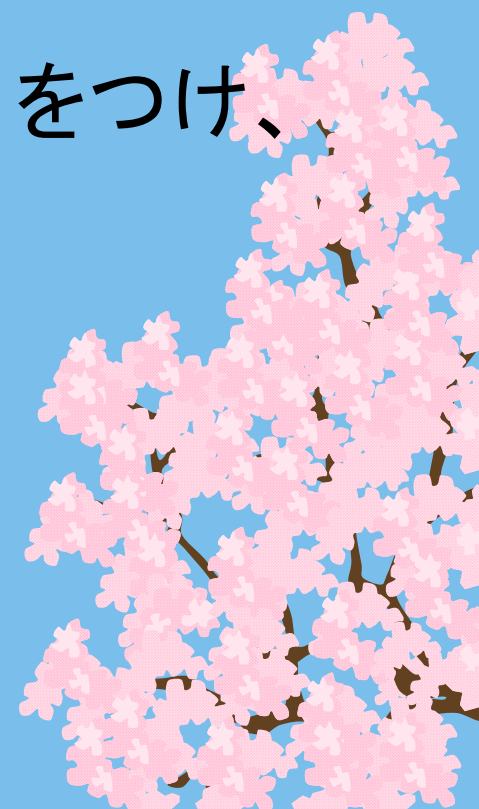
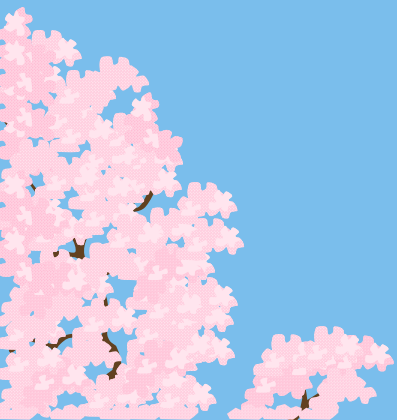
アラインメントから分かる事

- ❖ 類似性から進化系統樹(生物がどのような順番で枝分かれしてきたか)を作成する事が可能



動的計画法を用いた計算法

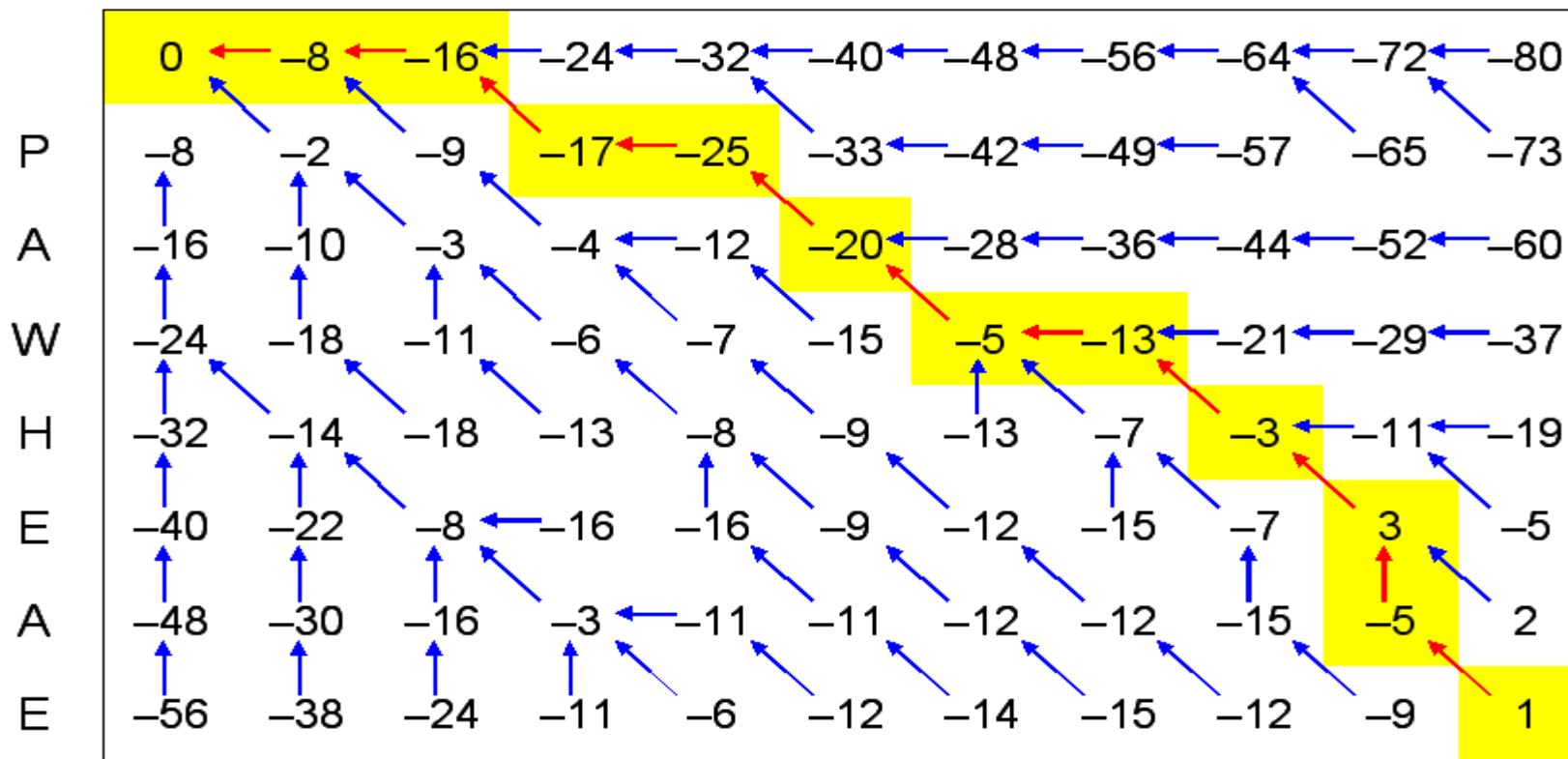
- ❖ 動的計画法・・・ある段階で得られた最適解（最大or最小）をもとに次の段階の最適解を求める
- ❖ アライメントに対して点数（スコア）をつけ、スコアを最大にする事を目標とする



スコアについて

- ❖ 配列要素1つ1つのペアに定義される類似度
- ❖ 配列要素とギャップとを対応づけるときは、スコアペナルティー「 d 」を適用する
- ❖ 最も単純なスコアは、 $s(a,b)$ の値を $a=b$ のとき「1」、 $a \neq b$ でないとき「0」とする定義である。塩基配列には、このスコアが一般に用いられる
- ❖ アミノ酸配列については統計的に関連性のある要素が対になる確率を整数化したものを用いる

H E A G A W G H E E



x = HEAGAWGHEE-E

y = --P-AW-HEAE

2つの配列 $x=x_1x_2\dots x_m$ と $y=y_1y_2\dots y_n$ の アラインメントを求める (グローバル)

F を $(m+1) \times (n+1)$ 行列

$$F(0,0)=0, F(i,0)=id, F(0,j)=jd,$$

$$F(i,j)=\max(A,B,C)$$

$$A=F(i-1,j-1)+s(x_i, x_j)$$

$$B=F(i-1,j)+d$$

$$C=F(i,j-1)+d$$

続き

❖ $F(i-1, j-1) \rightarrow F(i, j)$

X_i と y_i のアラインメントをとる (Aの場合)

❖ $F(i-1, j-1) \rightarrow F(i-1, j) \rightarrow F(i, j)$

X_{i-1} と y_j が対応。 X_i とギャップを対応づける (Bの場合)

❖ $F(i-1, j-1) \rightarrow F(i, j-1) \rightarrow F(i, j)$

X_i と y_{j-1} が対応。 y_j とギャップを対応付ける (Cの場合)

	X_{i-1}	X_i
y_{j-1}	$F(i-1, j-1)$	$F(i, j-1)$
y_i	$F(i-1, j)$	$F(i, j)$

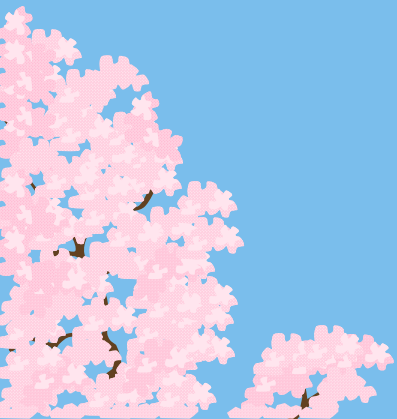
2つのアミノ酸配列の 最適アラインメント

❖ 配列1: HEAGAWGHEE

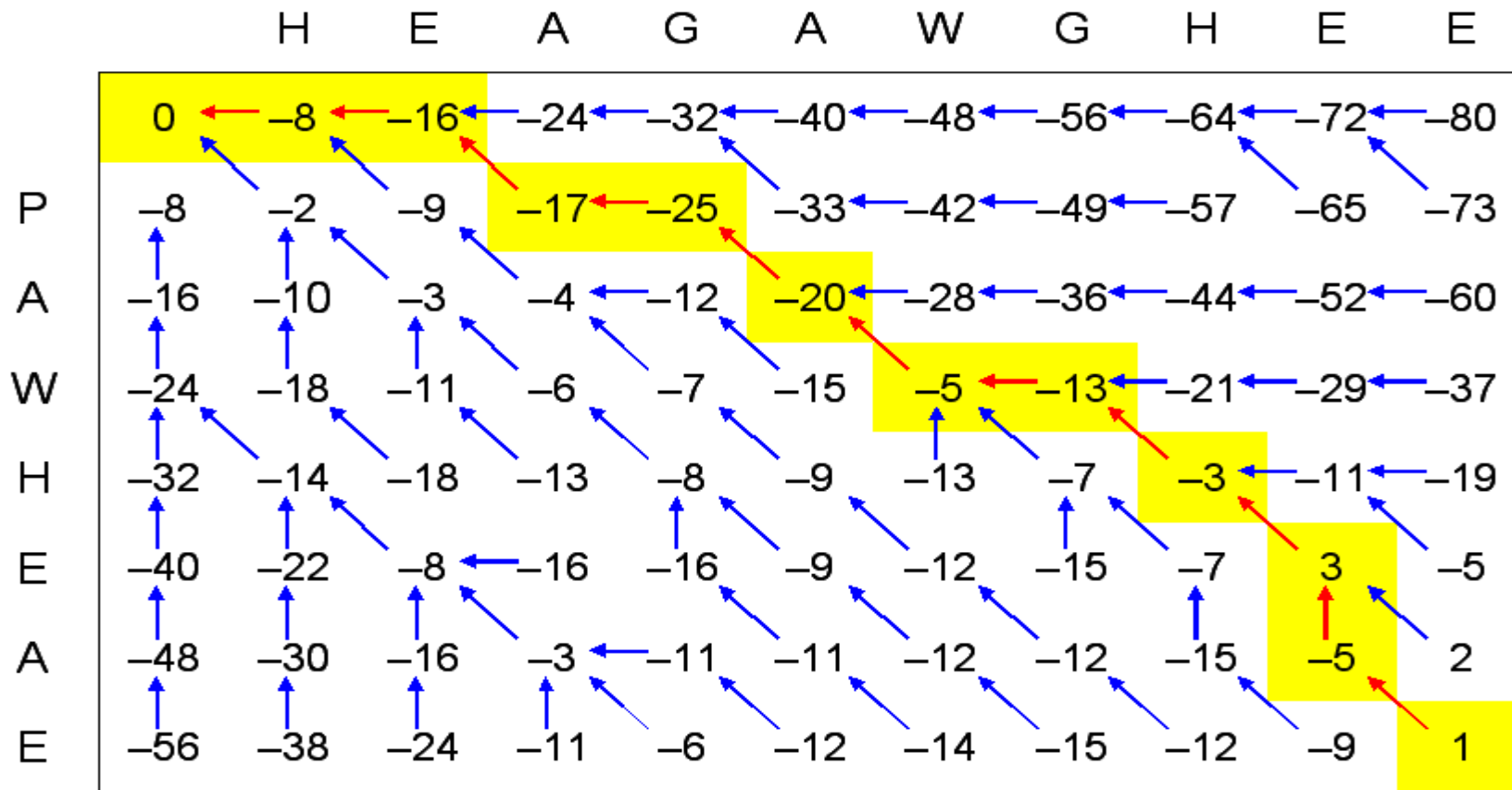
❖ 配列2: PAWHEAE

の最適アラインメントを求める。

ただし、 $d = -8$ (ギャップが起きた場合のスコア
アペナルティー)



グローバルアライメント



x = HEAGAWGHE-E

y = --P-AW-HEAE

2つの配列 $x=x_1x_2\dots x_m$ と $y=y_1y_2\dots y_n$ の アラインメントを求める (ローカル)

$$F(i,j)=\max(A,B,C,0)$$

$$A=F(i-1,j-1)+s(x_i, x_j)$$

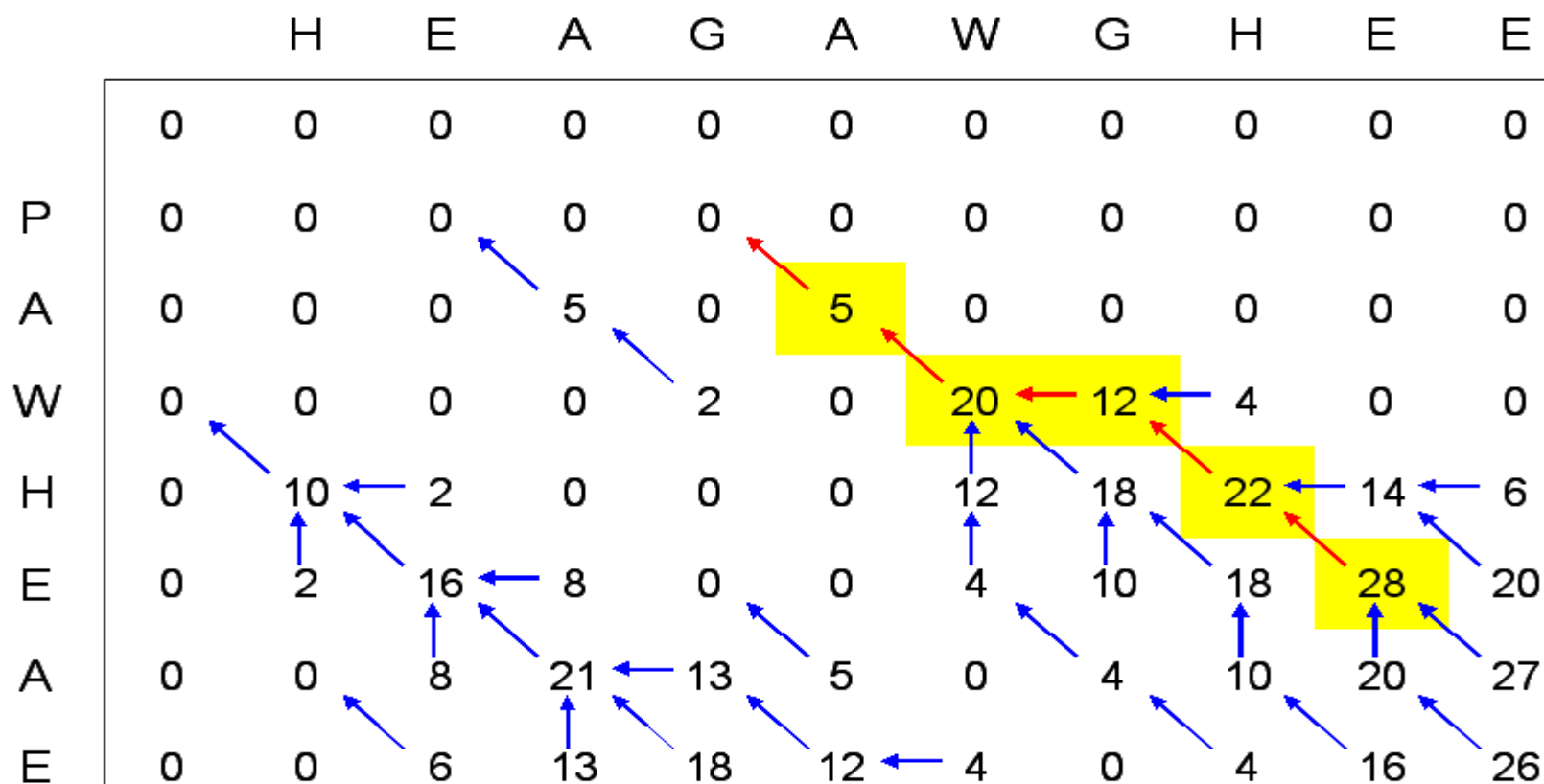
$$B=F(i-1,j)+d$$

$$C=F(i,j-1)+d$$

ただし、 $F(0,0)=F(i,0)=F(0,j)=0$

- ❖ 負のスコア値の場合、アラインメントを新しく始める

ローカルアライメント



x = AWGHE

y = AW-HE