

# MO640 – Biologia Computacional

Zanoni Dias

Instituto de Computação – Unicamp

Segundo Semestre de 2015

# Roteiro

- 1 Alinhamento de Sequências
- 2 Alinhamento Global
- 3 Alinhamento Semi-Global
- 4 Alinhamento Local
- 5 Pontuação para Alinhamento de Caracteres
- 6 Penalidades para Blocos de Buracos
- 7 Alinhamento Global usando Espaço Linear
- 8 Alinhamento Global de Sequências Similares
- 9 Algoritmo Sub-Quadrático para Alinhamento Global
- 10 Banco de Dados de Sequências
- 11 BLAST
- 12 BLAST 2.0

# O que é Alinhamento de Sequências?

- Um alinhamento de duas sequências de caracteres  $\alpha$  e  $\beta$  é obtido inserindo-se espaços nas sequências, e então colocando-se uma sobre a outra de modo que cada caracter ou espaço esteja emparelhado a um único caracter (ou a um espaço) da outra cadeia (não podem haver espaços alinhados com espaços).
- Exemplo:
  - ▶ Sequências:
    - ★  $\alpha = \text{AAACTGCACAATCTTAATGCCCTTTAT}$
    - ★  $\beta = \text{GCGGATCAACTTATTCCATCTCTT}$
  - ▶ Alinhamento:
    - ★  $\alpha' = \text{AAACTGCA-CAATCTTAATGCC--CTTTAT}$
    - ★  $\beta' = \text{--GC-GGATCAA-CTT-ATTCCATCTCTT--}$

# Por que Alinhar Sequências?

- Comparar sequências
- Localizar trechos homólogos
- Predição de função
- Predição de estrutura secundária
- Inferência filogenética

# Como Comparar Alinhamentos?

- Alinhamento 1:
  - ▶  $\alpha' = \text{CAGGATGCTAGCAAAAACCATCGCGGGCGATAA}$
  - ▶  $\beta' = --GGCATGTAGCACACACGACGCTGGGAGAAT-$
- Alinhamento 2:
  - ▶  $\alpha'' = \text{CAGGATGCTAGCAAAAACCATCGC-GGGCGATAA-}$
  - ▶  $\beta'' = -GGCATG-TAGCACACACGA-CGCTGGGAG--AAT$
- Qual é o melhor?
- Critério objetivo: função de pontuação para alinhamentos.
- Exemplo 1: gap =  $-1$ , match =  $2$ , mismatch =  $-4$ 
  - ▶ Alinhamento 1:  $3 \text{ gaps} + 16 \text{ matches} + 14 \text{ mismatches} = -27$
  - ▶ Alinhamento 2:  $7 \text{ gaps} + 22 \text{ matches} + 6 \text{ mismatches} = 13$
- Exemplo 2: gap =  $-10$ , match =  $1$ , mismatch =  $-1$ 
  - ▶ Alinhamento 1:  $3 \text{ gaps} + 16 \text{ matches} + 14 \text{ mismatches} = -28$
  - ▶ Alinhamento 2:  $7 \text{ gaps} + 22 \text{ matches} + 6 \text{ mismatches} = -54$

# Alinhamento Ótimo

## Problema

*Dadas duas sequências  $\alpha$  e  $\beta$  com, respectivamente,  $m$  e  $n$  caracteres, e um critério de pontuação de alinhamentos, deseja-se obter o alinhamento de pontuação máxima entre estas duas sequências.*

- Dado um esquema de pontuação e um alinhamento, é fácil determinar se este é o melhor alinhamento possível entre as duas sequências?
- Como ter certeza que um dado alinhamento é o melhor possível?
  - ▶ Calcular todos os alinhamentos possíveis!

# Tipos de Alinhamentos

- *Global*: alinhamento de pontuação máxima envolvendo as duas sequências completas. Desejável em situações onde as sequências são similares, por exemplo, ao se alinhar genes ou proteínas homólogas.
- *Semi-Global (ou Semi-Local)*: não penaliza buracos criados nas pontas das sequências. Desejável, por exemplo, no caso de montagem de genomas, onde busca-se um alinhamento de pontuação máxima entre o prefixo de uma sequência e o sufixo da outra (ou vice-versa).
- *Local*: alinhamento de pontuação máxima entre qualquer par de subsequências (das sequências originais). Desejável, por exemplo, para se identificar trechos altamente conservados entre dois genomas.

# Tipos de Alinhamentos

## Global

A C T G T G T A A C C C - G C A A A C T G T A G C G C C G
A - T C T G T A C C C C T G C A A A - T G T A C G G C - -

## Semi-Global

A C T G T G T A C C C G - G C
C C C C T G C A A A T T G T A C G G C C G

C G C C G A C G A C C C T G C A A A C T G T T A C A C A T
C G C C C - T G C A A A

## Local

A C T G T G T A A C G C - G C A A A C T G T A G C G C
G A A C G A A A C C T G C A A A G G A A C C T G C C A

# Alinhamento Global - Força Bruta - 1<sup>a</sup> Abordagem

- Onde a primeira base de uma das sequências pode se alinhar?
  - ▶ Opção 1:
    - ★ C [CAGCCGAATCGATCGCATG]
    - ★ - [CATCAGCGATCGATCTTT]
  - ▶ Opção 2:
    - ★ C [CAGCCGAATCGATCGCATG]
    - ★ C [ATCAGCGATCGATCTTT]
  - ▶ Opção 3:
    - ★ -C [CAGCCGAATCGATCGCATG]
    - ★ C- [ATCAGCGATCGATCTTT]
  - ▶ Opção 4:
    - ★ -C [CAGCCGAATCGATCGCATG]
    - ★ CA [TCAGCGATCGATCTTT]
  - ▶ ...

# Alinhamento Global - Força Bruta - 1<sup>a</sup> Abordagem

- Onde a primeira base de uma das sequências pode se alinhar?
  - ▶ ...
  - ▶ Opção  $2n - 2$ :
    - ★ -----C [CAGCCGAATCGATCGCATG]
    - ★ CATCAGCGATCGATCTT [T]
  - ▶ Opção  $2n - 1$ :
    - ★ -----C [CAGCCGAATCGATCGCATG]
    - ★ CATCAGCGATCGATCTT- [T]
  - ▶ Opção  $2n$ :
    - ★ -----C [CAGCCGAATCGATCGCATG]
    - ★ CATCAGCGATCGATCTT[]
  - ▶ Opção  $2n + 1$ :
    - ★ -----C [CAGCCGAATCGATCGCATG]
    - ★ CATCAGCGATCGATCTT- []

# Alinhamento Global - Força Bruta - 1<sup>a</sup> Abordagem

---

## Algoritmo 1: Align

---

**Input:**  $\alpha, a_i, a_f, \beta, b_i, b_f$

**if**  $a_i > a_f$  **then return**  $gap \times (b_f - b_i + 1)$ ;

**if**  $b_i > b_f$  **then return**  $gap \times (a_f - a_i + 1)$ ;

$max \leftarrow gap + Align(\alpha, a_i + 1, a_f, \beta, b_i, b_f);$

**for all**  $b_k \in [b_i..b_f]$  **do**

$aux \leftarrow Align(\alpha, a_i + 1, a_f, \beta, b_k + 1, b_f);$

$aux1 \leftarrow gap \times (b_k - b_i) + \sigma(\alpha[a_i], \beta[b_k]) + aux;$

$aux2 \leftarrow gap \times (b_k - b_i + 2) + aux;$

**if**  $aux1 > max$  **then**  $max \leftarrow aux1;$

**if**  $aux2 > max$  **then**  $max \leftarrow aux2;$

**end**

**return**  $max$

---

# Complexidade e Tempo de Execução

- Complexidade:

- ▶  $T(m, n) = \sum_{i=0}^n T(m - 1, i) + (n + 1)$
- ▶  $T(m, n) \geq \sum_{i=0}^{n-1} T(m - 1, i) + (n + 1)$
- ▶  $T(m, n) \geq \sum_{i=0}^{\min\{m,n\}-1} T(\min\{m, n\} - 1, i) + (\min\{m, n\} + 1)$
- ▶  $T(m, n) \geq \sum_{i=0}^{\min\{m,n\}-1} (\min\{m, n\} - i) \times 2^i = 2^{\min\{m,n\}} - 1$
- ▶  $T(m, n) = \Omega(2^{\min\{m,n\}})$

- Exemplo: computador de 3GHz (1 instrução por ciclo)

- ▶  $m = n = 25$ :

- ▶ Tempo:  $2^{25}/(3 \times 2^{30}) = 3$  centésimos de segundo

- ▶  $m = n = 50$ :

- ▶ Tempo:  $2^{50}/(3 \times 2^{30}) = 12$  dias

- ▶  $m = n = 75$ :

- ▶ Tempo:  $2^{75}/(3 \times 2^{30}) = 1$  milhão de anos

- ▶  $m = n = 100$ :

- ▶ Tempo:  $2^{100}/(3 \times 2^{30}) = 37$  trilhões de anos

## Alinhamento Global - Força Bruta - 2<sup>a</sup> Abordagem

- Quais as possíveis opções para a última coluna do alinhamento?
  - ▶ Opção 1:
    - ★ CCAGCCGAATCGATCGCAT**G**
    - ★ ...CATCAGCGATCGATCT**T**
  - ▶ Opção 2:
    - ★ CCAGCCGAATCGATCGCAT**G**
    - ★ ...CATCAGCGATCGATCT**-**
  - ▶ Opção 3:
    - ★ CCAGCCGAATCGATCGCAT**-**
    - ★ ...CATCAGCGATCGATCT**T**

# Alinhamento Global - Força Bruta - 2<sup>a</sup> Abordagem

---

## Algoritmo 2: Align

---

```
Input:  $\alpha, m, \beta, n$ 
if  $m = 0$  then
    | return  $gap \times n$ 
end
if  $n = 0$  then
    | return  $gap \times m$ 
end
return max {  $Align(\alpha, m - 1, \beta, n - 1) + \sigma(\alpha[m], \beta[n]),$ 
              $Align(\alpha, m - 1, \beta, n) + gap,$ 
              $Align(\alpha, m, \beta, n - 1) + gap$  }
```

---

# Complexidade e Tempo de Execução

- Complexidade:

- ▶  $T(m, n) = T(m - 1, n - 1) + T(m, n - 1) + T(m - 1, n) + \Theta(1)$
- ▶  $T(m, n) \geq 3T(m - 1, n - 1) + \Theta(1)$
- ▶  $T(m, n) = \Omega(3^{\min\{m, n\}})$

- Exemplo: computador de 3GHz (1 instrução por ciclo)

- ▶  $m = n = 10$ :

- ▶ Tempo:  $3^{10}/(3 \times 2^{30}) = 18$  milionésimos de segundo

- ▶  $m = n = 20$ :

- ▶ Tempo:  $3^{20}/(3 \times 2^{30}) = 1$  segundo

- ▶  $m = n = 30$ :

- ▶ Tempo:  $3^{30}/(3 \times 2^{30}) = 18$  horas

- ▶  $m = n = 40$ :

- ▶ Tempo:  $3^{40}/(3 \times 2^{30}) = 120$  anos

- ▶  $m = n = 50$ :

- ▶ Tempo:  $3^{50}/(3 \times 2^{30}) = 7$  milhões de anos

# Contagem de Subproblemas

- O grande problema do método anterior que o alinhamento de dois prefixos de  $\alpha$  e de  $\beta$  podem ser recalculado muitas vezes.
- Existem quantos subproblemas distintos, envolvendo alinhamentos de prefixos não vazios de  $\alpha$  e de  $\beta$ ?
  - ▶ Apenas  $m \times n$  subproblemas!
- Memorização (*memoization*): evita o recálculo de subproblemas, armazenando os valores previamente calculados numa matriz.
- Podemos usar uma matriz para armazena os valores dos alinhamentos ótimos entre todos prefixos de  $\alpha$  e todos os prefixos de  $\beta$ .

# Memorização

---

## Algoritmo 3: Memoization

---

**Input:**  $\alpha, m, \beta, n$   
**for** all  $i \in [1..n]$  **do**  
    **for** all  $j \in [1..m]$  **do**  
         $M[i, j] \leftarrow \infty$   
    **end**  
**end**  
**return** *AlignMemoization*( $\alpha, m, \beta, n$ )

---

## Algoritmo 4: AlignMemoization

---

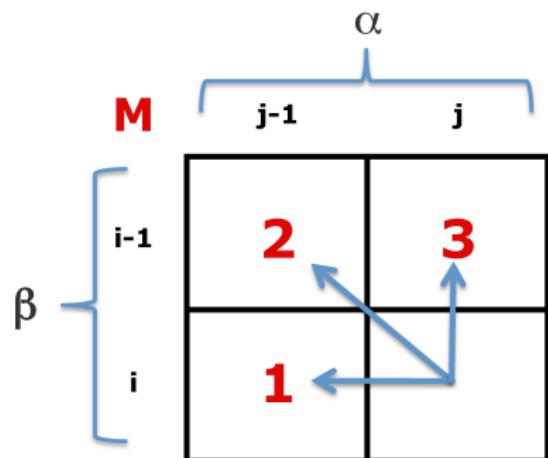
**Input:**  $\alpha, m, \beta, n$   
**if**  $m = 0$  **then return**  $gap \times n$   
**if**  $n = 0$  **then return**  $gap \times m$   
**if**  $M[n - 1, m - 1] = \infty$  **then**  $M[n - 1, m - 1] \leftarrow \text{AlignMemoization}(\alpha, m - 1, \beta, n - 1)$   
**if**  $M[n - 1, m] = \infty$  **then**  $M[n - 1, m] \leftarrow \text{AlignMemoization}(\alpha, m, \beta, n - 1)$   
**if**  $M[n, m - 1] = \infty$  **then**  $M[n, m - 1] \leftarrow \text{AlignMemoization}(\alpha, m - 1, \beta, n)$   
**return**  $\max \left\{ \begin{array}{l} M[n - 1, m - 1] + \sigma(\alpha[m], \beta[n]), \\ M[n - 1, m] + gap, \\ M[n, m - 1] + gap \end{array} \right\}$

---

# Programação Dinâmica

- Ideias básicas:
  - ▶ Matriz M armazena os valores dos alinhamentos ótimos entre todos prefixos de  $\alpha$  e todos os prefixos de  $\beta$ .
  - ▶ O valor do alinhamento ótimo entre as duas sequências deve estar armazenado na posição  $M[n,m]$ .
  - ▶ A matriz M pode ser preenchida, numa ordem adequada, sem a necessidade de nenhuma chamada recursiva.
  - ▶ Para simplificar o algoritmo, podemos armazenar na matriz M os valores dos alinhamentos ótimos de prefixos de  $\alpha$  ou de  $\beta$  com a sequência vazia.
- Em 1970, Saul Needleman e Christian Wunsch propuseram o algoritmo de programação dinâmica para alinhamento de sequências.

# Matriz de Programação Dinâmica



1	$M[i,j-1] + \text{gap}$	$\alpha$	$j-1$	$\alpha_j$	-
2	$M[i-1,j-1] + \text{match/mismatch}$	$\alpha$	$j-1$	$\alpha_j$	$\beta_i$
3	$M[i-1,j] + \text{gap}$	$\alpha$	$j$	-	$\beta_i$

# Alinhamento Global

---

## Algoritmo 5: Global

---

**Input:**  $\alpha, m, \beta, n$

**for** all  $i \in [0..n]$  **do**  $M[i, 0] \leftarrow \text{gap} \times i$

**for** all  $j \in [1..m]$  **do**  $M[0, j] \leftarrow \text{gap} \times j$

**for** all  $i \in [1..n]$  **do**

**for** all  $j \in [1..m]$  **do**

$M[i, j] \leftarrow \max \left\{ \begin{array}{l} M[i, j - 1] + \text{gap}, \\ M[i - 1, j - 1] + \sigma(\alpha[j], \beta[i]), \\ M[i - 1, j] + \text{gap} \end{array} \right\}$

**end**

**end**

**return**  $M[n, m]$

---

# Complexidade e Tempo de Execução

- Complexidade:
  - ▶  $T(m, n) = \Theta(mn)$
- Exemplo: computador de 3GHz (1 instrução por ciclo)
  - ▶  $m = n = 10$ :
    - ★ Tempo:  $(10 \times 10)/(3 \times 2^{30}) = 31$  bilionésimos de segundo
  - ▶  $m = n = 100$ :
    - ★ Tempo:  $(100 \times 100)/(3 \times 2^{30}) = 3$  milionésimos de segundo
  - ▶  $m = n = 1000$ :
    - ★ Tempo:  $(1000 \times 1000)/(3 \times 2^{30}) = 310$  milionésimos de segundo
  - ▶  $m = n = 10000$ :
    - ★ Tempo:  $(10000 \times 10000)/(3 \times 2^{30}) = 31$  milésimos de segundo
  - ▶  $m = n = 100000$ :
    - ★ Tempo:  $(100000 \times 100000)/(3 \times 2^{30}) = 3$  segundos
  - ▶  $m = n = 1000000$ :
    - ★ Tempo:  $(1000000 \times 1000000)/(3 \times 2^{30}) = 5$  minutos

# Alinhamento Global

x	$\alpha$	A	C	T	G	G	G	T	C	A	A	C
$\beta$	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55
A	-5											
T	-10											
T	-15											
G	-20											
G	-25											
C	-30											
C	-35											
A	-40											
C	-45											

Gap = -5

Match = +3

Mismatch = -2

# Alinhamento Global

$\times$	$\alpha$	<b>A</b>	<b>C</b>	<b>T</b>	<b>G</b>	<b>G</b>	<b>G</b>	<b>T</b>	<b>C</b>	<b>A</b>	<b>A</b>	<b>C</b>
$\beta$	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55
<b>A</b>	-5	3										
<b>T</b>	-10											
<b>T</b>	-15											
<b>G</b>	-20											
<b>G</b>	-25											
<b>C</b>	-30											
<b>C</b>	-35											
<b>A</b>	-40											
<b>C</b>	-45											

Gap = -5

Match = +3

Mismatch = -2

# Alinhamento Global

$\times$	$\alpha$	<b>A</b>	<b>C</b>	<b>T</b>	<b>G</b>	<b>G</b>	<b>G</b>	<b>T</b>	<b>C</b>	<b>A</b>	<b>A</b>	<b>C</b>
$\beta$	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55
<b>A</b>	-5	3	<b>-2</b>									
<b>T</b>	-10											
<b>T</b>	-15											
<b>G</b>	-20											
<b>G</b>	-25											
<b>C</b>	-30											
<b>C</b>	-35											
<b>A</b>	-40											
<b>C</b>	-45											

Gap = -5

Match = +3

Mismatch = -2

# Alinhamento Global

$\times$	$\alpha$	<b>A</b>	<b>C</b>	<b>T</b>	<b>G</b>	<b>G</b>	<b>G</b>	<b>T</b>	<b>C</b>	<b>A</b>	<b>A</b>	<b>C</b>
$\beta$	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55
<b>A</b>	-5	3	-2	-7								
<b>T</b>	-10											
<b>T</b>	-15											
<b>G</b>	-20											
<b>G</b>	-25											
<b>C</b>	-30											
<b>C</b>	-35											
<b>A</b>	-40											
<b>C</b>	-45											

Gap = -5

Match = +3

Mismatch = -2

# Alinhamento Global

$\times$	$\alpha$	<b>A</b>	<b>C</b>	<b>T</b>	<b>G</b>	<b>G</b>	<b>G</b>	<b>T</b>	<b>C</b>	<b>A</b>	<b>A</b>	<b>C</b>
$\beta$	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55
<b>A</b>	-5	3	-2	-7	-12	-17	-22	-27	-32	-37	-42	<b>-47</b>
<b>T</b>	-10											
<b>T</b>	-15											
<b>G</b>	-20											
<b>G</b>	-25											
<b>C</b>	-30											
<b>C</b>	-35											
<b>A</b>	-40											
<b>C</b>	-45											

Gap = -5

Match = +3

Mismatch = -2

# Alinhamento Global

$\times$	$\alpha$	<b>A</b>	<b>C</b>	<b>T</b>	<b>G</b>	<b>G</b>	<b>G</b>	<b>T</b>	<b>C</b>	<b>A</b>	<b>A</b>	<b>C</b>
$\beta$	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55
<b>A</b>	-5	3	-2	-7	-12	-17	-22	-27	-32	-37	-42	-47
<b>T</b>	-10	-2	1	1	-4	-9	-14	-19	-24	-29	-34	-39
<b>T</b>	-15	-7	-4	4	-1	-6	-11	-11	-16	-21	-26	-31
<b>G</b>	-20	-12	-9	-1	7	2	-3	-8	-13	-18	-23	-28
<b>G</b>	-25	-17	-14	-6	2	10	5					
<b>C</b>	-30											
<b>C</b>	-35											
<b>A</b>	-40											
<b>C</b>	-45											

Gap = -5

Match = +3

Mismatch = -2

# Alinhamento Global

$\times$	$\alpha$	<b>A</b>	<b>C</b>	<b>T</b>	<b>G</b>	<b>G</b>	<b>G</b>	<b>T</b>	<b>C</b>	<b>A</b>	<b>A</b>	<b>C</b>
$\beta$	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55
<b>A</b>	-5	3	-2	-7	-12	-17	-22	-27	-32	-37	-42	-47
<b>T</b>	-10	-2	1	1	-4	-9	-14	-19	-24	-29	-34	-39
<b>T</b>	-15	-7	-4	4	-1	-6	-11	-11	-16	-21	-26	-31
<b>G</b>	-20	-12	-9	-1	7	2	-3	-8	-13	-18	-23	-28
<b>G</b>	-25	-17	-14	-6	2	10	5	0	-5	-10	-15	-20
<b>C</b>	-30	-22	-14	-11	-3	5	8	3	3	-2	-7	-12
<b>C</b>	-35	-27	-19	-16	-8	0	3	6	6	1	-4	-4
<b>A</b>	-40	-32	-24	-21	-13	-5	-2	1	4	9	4	-1
<b>C</b>	-45	-37	-29	-26	-18	-10	-7	-4	4	4	7	7

Gap = -5

Match = +3

Mismatch = -2

# Alinhamento Global

$\times$	$\alpha$	<b>A</b>	<b>C</b>	<b>T</b>	<b>G</b>	<b>G</b>	<b>G</b>	<b>T</b>	<b>C</b>	<b>A</b>	<b>A</b>	<b>C</b>
$\beta$	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55
<b>A</b>	-5	3	-2	-7	-12	-17	-22	-27	-32	-37	-42	-47
<b>T</b>	-10	-2	1	1	-4	-9	-14	-19	-24	-29	-34	-39
<b>T</b>	-15	-7	-4	4	-1	-6	-11	-11	-16	-21	-26	-31
<b>G</b>	-20	-12	-9	-1	7	2	-3	-8	-13	-18	-23	-28
<b>G</b>	-25	-17	-14	-6	2	10	5	0	-5	-10	-15	-20
<b>C</b>	-30	-22	-14	-11	-3	5	8	3	3	-2	-7	-12
<b>C</b>	-35	-27	-19	-16	-8	0	3	6	6	1	-4	-4
<b>A</b>	-40	-32	-24	-21	-13	-5	-2	1	4	9	4	-1
<b>C</b>	-45	-37	-29	-26	-18	-10	-7	-4	4	4	7	7

$$\alpha =$$

C

$$\beta =$$

C

# Alinhamento Global

$\times$	$\alpha$	<b>A</b>	<b>C</b>	<b>T</b>	<b>G</b>	<b>G</b>	<b>G</b>	<b>T</b>	<b>C</b>	<b>A</b>	<b>A</b>	<b>C</b>
$\beta$	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55
<b>A</b>	-5	3	-2	-7	-12	-17	-22	-27	-32	-37	-42	-47
<b>T</b>	-10	-2	1	1	-4	-9	-14	-19	-24	-29	-34	-39
<b>T</b>	-15	-7	-4	4	-1	-6	-11	-11	-16	-21	-26	-31
<b>G</b>	-20	-12	-9	-1	7	2	-3	-8	-13	-18	-23	-28
<b>G</b>	-25	-17	-14	-6	2	10	5	0	-5	-10	-15	-20
<b>C</b>	-30	-22	-14	-11	-3	5	8	3	3	-2	-7	-12
<b>C</b>	-35	-27	-19	-16	-8	0	3	6	6	1	-4	-4
<b>A</b>	-40	-32	-24	-21	-13	-5	-2	1	4	9	4	-1
<b>C</b>	-45	-37	-29	-26	-18	-10	-7	-4	4	4	7	7

$$\alpha =$$

A C

$$\beta =$$

A C

# Alinhamento Global

$\times$	$\alpha$	<b>A</b>	<b>C</b>	<b>T</b>	<b>G</b>	<b>G</b>	<b>G</b>	<b>T</b>	<b>C</b>	<b>A</b>	<b>A</b>	<b>C</b>
$\beta$	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55
<b>A</b>	-5	3	-2	-7	-12	-17	-22	-27	-32	-37	-42	-47
<b>T</b>	-10	-2	1	1	-4	-9	-14	-19	-24	-29	-34	-39
<b>T</b>	-15	-7	-4	4	-1	-6	-11	-11	-16	-21	-26	-31
<b>G</b>	-20	-12	-9	-1	7	2	-3	-8	-13	-18	-23	-28
<b>G</b>	-25	-17	-14	-6	2	10	5	0	-5	-10	-15	-20
<b>C</b>	-30	-22	-14	-11	-3	5	8	3	3	-2	-7	-12
<b>C</b>	-35	-27	-19	-16	-8	0	3	6	6	1	-4	-4
<b>A</b>	-40	-32	-24	-21	-13	-5	-2	1	4	9	4	-1
<b>C</b>	-45	-37	-29	-26	-18	-10	-7	-4	4	4	7	7

$\alpha =$  A A C

$\beta =$  - A C

# Alinhamento Global

$\times$	$\alpha$	<b>A</b>	<b>C</b>	<b>T</b>	<b>G</b>	<b>G</b>	<b>G</b>	<b>T</b>	<b>C</b>	<b>A</b>	<b>A</b>	<b>C</b>
$\beta$	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55
<b>A</b>	-5	3	-2	-7	-12	-17	-22	-27	-32	-37	-42	-47
<b>T</b>	-10	-2	1	1	-4	-9	-14	-19	-24	-29	-34	-39
<b>T</b>	-15	-7	-4	4	-1	-6	-11	-11	-16	-21	-26	-31
<b>G</b>	-20	-12	-9	-1	7	2	-3	-8	-13	-18	-23	-28
<b>G</b>	-25	-17	-14	-6	2	10	5	0	-5	-10	-15	-20
<b>C</b>	-30	-22	-14	-11	-3	5	8	3	3	-2	-7	-12
<b>C</b>	-35	-27	-19	-16	-8	0	3	6	6	1	-4	-4
<b>A</b>	-40	-32	-24	-21	-13	-5	-2	1	4	9	4	-1
<b>C</b>	-45	-37	-29	-26	-18	-10	-7	-4	4	4	7	7

$$\alpha =$$

C A A C

$$\beta =$$

C - A C

# Alinhamento Global

$\times$	$\alpha$	<b>A</b>	<b>C</b>	<b>T</b>	<b>G</b>	<b>G</b>	<b>G</b>	<b>T</b>	<b>C</b>	<b>A</b>	<b>A</b>	<b>C</b>
$\beta$	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55
<b>A</b>	-5	3	-2	-7	-12	-17	-22	-27	-32	-37	-42	-47
<b>T</b>	-10	-2	1	1	-4	-9	-14	-19	-24	-29	-34	-39
<b>T</b>	-15	-7	-4	4	-1	-6	-11	-11	-16	-21	-26	-31
<b>G</b>	-20	-12	-9	-1	7	2	-3	-8	-13	-18	-23	-28
<b>G</b>	-25	-17	-14	-6	2	10	5	0	-5	-10	-15	-20
<b>C</b>	-30	-22	-14	-11	-3	5	8	3	3	-2	-7	-12
<b>C</b>	-35	-27	-19	-16	-8	0	3	6	6	1	-4	-4
<b>A</b>	-40	-32	-24	-21	-13	-5	-2	1	4	9	4	-1
<b>C</b>	-45	-37	-29	-26	-18	-10	-7	-4	4	4	7	7

$$\alpha =$$

$$\beta =$$

$$\begin{array}{ccccc} \textcolor{red}{T} & \textcolor{orange}{C} & \textcolor{orange}{A} & \textcolor{orange}{A} & \textcolor{orange}{C} \\ - & \textcolor{orange}{C} & - & \textcolor{orange}{A} & \textcolor{orange}{C} \end{array}$$

# Alinhamento Global

$\times$	$\alpha$	<b>A</b>	<b>C</b>	<b>T</b>	<b>G</b>	<b>G</b>	<b>G</b>	<b>T</b>	<b>C</b>	<b>A</b>	<b>A</b>	<b>C</b>
$\beta$	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55
<b>A</b>	-5	3	-2	-7	-12	-17	-22	-27	-32	-37	-42	-47
<b>T</b>	-10	-2	1	1	-4	-9	-14	-19	-24	-29	-34	-39
<b>T</b>	-15	-7	-4	4	-1	-6	-11	-11	-16	-21	-26	-31
<b>G</b>	-20	-12	-9	-1	7	2	-3	-8	-13	-18	-23	-28
<b>G</b>	-25	-17	-14	-6	2	10	5	0	-5	-10	-15	-20
<b>C</b>	-30	-22	-14	-11	-3	5	8	3	3	-2	-7	-12
<b>C</b>	-35	-27	-19	-16	-8	0	3	6	6	1	-4	-4
<b>A</b>	-40	-32	-24	-21	-13	-5	-2	1	4	9	4	-1
<b>C</b>	-45	-37	-29	-26	-18	-10	-7	-4	4	4	7	7

$\alpha = \text{G T C A A C}$

$\beta = \text{C - C - A C}$

# Alinhamento Global

$\times$	$\alpha$	<b>A</b>	<b>C</b>	<b>T</b>	<b>G</b>	<b>G</b>	<b>G</b>	<b>T</b>	<b>C</b>	<b>A</b>	<b>A</b>	<b>C</b>
$\beta$	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55
<b>A</b>	-5	3	-2	-7	-12	-17	-22	-27	-32	-37	-42	-47
<b>T</b>	-10	-2	1	1	-4	-9	-14	-19	-24	-29	-34	-39
<b>T</b>	-15	-7	-4	4	-1	-6	-11	-11	-16	-21	-26	-31
<b>G</b>	-20	-12	-9	-1	7	2	-3	-8	-13	-18	-23	-28
<b>G</b>	-25	-17	-14	-6	2	10	5	0	-5	-10	-15	-20
<b>C</b>	-30	-22	-14	-11	-3	5	8	3	3	-2	-7	-12
<b>C</b>	-35	-27	-19	-16	-8	0	3	6	6	1	-4	-4
<b>A</b>	-40	-32	-24	-21	-13	-5	-2	1	4	9	4	-1
<b>C</b>	-45	-37	-29	-26	-18	-10	-7	-4	4	4	7	7

$\alpha = \text{G G T C A A C}$

$\beta = \text{G C - C - A C}$

# Alinhamento Global

$\times$	$\alpha$	<b>A</b>	<b>C</b>	<b>T</b>	<b>G</b>	<b>G</b>	<b>G</b>	<b>T</b>	<b>C</b>	<b>A</b>	<b>A</b>	<b>C</b>
$\beta$	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55
<b>A</b>	-5	3	-2	-7	-12	-17	-22	-27	-32	-37	-42	-47
<b>T</b>	-10	-2	1	1	-4	-9	-14	-19	-24	-29	-34	-39
<b>T</b>	-15	-7	-4	4	-1	-6	-11	-11	-16	-21	-26	-31
<b>G</b>	-20	-12	-9	-1	7	2	-3	-8	-13	-18	-23	-28
<b>G</b>	-25	-17	-14	-6	2	10	5	0	-5	-10	-15	-20
<b>C</b>	-30	-22	-14	-11	-3	5	8	3	3	-2	-7	-12
<b>C</b>	-35	-27	-19	-16	-8	0	3	6	6	1	-4	-4
<b>A</b>	-40	-32	-24	-21	-13	-5	-2	1	4	9	4	-1
<b>C</b>	-45	-37	-29	-26	-18	-10	-7	-4	4	4	7	7

$\alpha = \text{G G G T C A A C}$

$\beta = \text{G G C - C - A C}$

# Alinhamento Global

$\times$	$\alpha$	<b>A</b>	<b>C</b>	<b>T</b>	<b>G</b>	<b>G</b>	<b>G</b>	<b>T</b>	<b>C</b>	<b>A</b>	<b>A</b>	<b>C</b>
$\beta$	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55
<b>A</b>	-5	3	-2	-7	-12	-17	-22	-27	-32	-37	-42	-47
<b>T</b>	-10	-2	1	1	-4	-9	-14	-19	-24	-29	-34	-39
<b>T</b>	-15	-7	-4	4	-1	-6	-11	-11	-16	-21	-26	-31
<b>G</b>	-20	-12	-9	-1	7	2	-3	-8	-13	-18	-23	-28
<b>G</b>	-25	-17	-14	-6	2	10	5	0	-5	-10	-15	-20
<b>C</b>	-30	-22	-14	-11	-3	5	8	3	3	-2	-7	-12
<b>C</b>	-35	-27	-19	-16	-8	0	3	6	6	1	-4	-4
<b>A</b>	-40	-32	-24	-21	-13	-5	-2	1	4	9	4	-1
<b>C</b>	-45	-37	-29	-26	-18	-10	-7	-4	4	4	7	7

$\alpha = \text{T G G G T C A A C}$

$\beta = \text{T G G C - C - A C}$

# Alinhamento Global

$\times$	$\alpha$	<b>A</b>	<b>C</b>	<b>T</b>	<b>G</b>	<b>G</b>	<b>G</b>	<b>T</b>	<b>C</b>	<b>A</b>	<b>A</b>	<b>C</b>
$\beta$	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55
<b>A</b>	-5	3	-2	-7	-12	-17	-22	-27	-32	-37	-42	-47
<b>T</b>	-10	-2	1	1	-4	-9	-14	-19	-24	-29	-34	-39
<b>T</b>	-15	-7	-4	4	-1	-6	-11	-11	-16	-21	-26	-31
<b>G</b>	-20	-12	-9	-1	7	2	-3	-8	-13	-18	-23	-28
<b>G</b>	-25	-17	-14	-6	2	10	5	0	-5	-10	-15	-20
<b>C</b>	-30	-22	-14	-11	-3	5	8	3	3	-2	-7	-12
<b>C</b>	-35	-27	-19	-16	-8	0	3	6	6	1	-4	-4
<b>A</b>	-40	-32	-24	-21	-13	-5	-2	1	4	9	4	-1
<b>C</b>	-45	-37	-29	-26	-18	-10	-7	-4	4	4	7	7

$\alpha = \text{C T G G G T C A A C}$

$\beta = \text{T T G G C - C - A C}$

# Alinhamento Global

$\times$	$\alpha$	<b>A</b>	<b>C</b>	<b>T</b>	<b>G</b>	<b>G</b>	<b>G</b>	<b>T</b>	<b>C</b>	<b>A</b>	<b>A</b>	<b>C</b>
$\beta$	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55
<b>A</b>	-5	3	-2	-7	-12	-17	-22	-27	-32	-37	-42	-47
<b>T</b>	-10	-2	1	1	-4	-9	-14	-19	-24	-29	-34	-39
<b>T</b>	-15	-7	-4	4	-1	-6	-11	-11	-16	-21	-26	-31
<b>G</b>	-20	-12	-9	-1	7	2	-3	-8	-13	-18	-23	-28
<b>G</b>	-25	-17	-14	-6	2	10	5	0	-5	-10	-15	-20
<b>C</b>	-30	-22	-14	-11	-3	5	8	3	3	-2	-7	-12
<b>C</b>	-35	-27	-19	-16	-8	0	3	6	6	1	-4	-4
<b>A</b>	-40	-32	-24	-21	-13	-5	-2	1	4	9	4	-1
<b>C</b>	-45	-37	-29	-26	-18	-10	-7	-4	4	4	7	7

$\alpha = \text{A C T G G G T C A A C}$

$\beta = \text{A T T G G C - C - A C}$

# Alinhamento Global

$\times$	$\alpha$	<b>A</b>	<b>C</b>	<b>T</b>	<b>G</b>	<b>G</b>	<b>G</b>	<b>T</b>	<b>C</b>	<b>A</b>	<b>A</b>	<b>C</b>
$\beta$	<b>0</b>	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55
<b>A</b>	-5	<b>3</b>	-2	-7	-12	-17	-22	-27	-32	-37	-42	-47
<b>T</b>	-10	-2	<b>1</b>	1	-4	-9	-14	-19	-24	-29	-34	-39
<b>T</b>	-15	-7	-4	<b>4</b>	-1	-6	-11	-11	-16	-21	-26	-31
<b>G</b>	-20	-12	-9	-1	<b>7</b>	2	-3	-8	-13	-18	-23	-28
<b>G</b>	-25	-17	-14	-6	2	<b>10</b>	5	0	-5	-10	-15	-20
<b>C</b>	-30	-22	-14	-11	-3	5	<b>8</b>	<b>3</b>	3	-2	-7	-12
<b>C</b>	-35	-27	-19	-16	-8	0	3	6	<b>6</b>	<b>1</b>	-4	-4
<b>A</b>	-40	-32	-24	-21	-13	-5	-2	1	4	9	<b>4</b>	-1
<b>C</b>	-45	-37	-29	-26	-18	-10	-7	-4	4	4	7	<b>7</b>
$\alpha$	=	<b>A</b>	<b>C</b>	<b>T</b>	<b>G</b>	<b>G</b>	<b>G</b>	<b>T</b>	<b>C</b>	<b>A</b>	<b>A</b>	<b>C</b>
$\beta$	=	<b>A</b>	<b>T</b>	<b>T</b>	<b>G</b>	<b>G</b>	<b>C</b>	-	<b>C</b>	-	<b>A</b>	<b>C</b>

# Exercícios

## Exercício

*Escreva um algoritmo, em pseudocódigo, que dadas duas sequências  $\alpha$  e  $\beta$ , com, respectivamente,  $m$  e  $n$  caracteres, e uma matriz de pontuação de alinhamentos  $M$ , entre todos os pares de prefixos de  $\alpha$  e de  $\beta$ , retorne um alinhamento ótimo entre  $\alpha$  e  $\beta$ .*

## Exercício

*Calcule a complexidade assintótica (usando a notação  $\Theta$ ) dos algoritmos Memoization e AlignMemoization.*

# Alinhamento Global

$\times$	$\alpha$	<b>G</b>	<b>C</b>	<b>G</b>	<b>C</b>	<b>G</b>	<b>T</b>	<b>T</b>	<b>A</b>	<b>G</b>	<b>A</b>	<b>C</b>	<b>T</b>	<b>A</b>	<b>G</b>	<b>C</b>	<b>A</b>	<b>C</b>	<b>C</b>	<b>G</b>
$\beta$	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55	-60	-65	-70	-75	-80	-85	-90	-95
<b>G</b>	-5	3	-2	-7	-12	-17	-22	-27	-32	-37	-42	-47	-52	-57	-62	-67	-72	-77	-82	-87
<b>G</b>	-10	-2	1	1	-4	-9	-14	-19	-24	-29	-34	-39	-44	-49	-54	-59	-64	-69	-74	-79
<b>G</b>	-15	-7	-4	4	-1	-1	-6	-11	-16	-21	-26	-31	-36	-41	-46	-51	-56	-61	-66	-71
<b>T</b>	-20	-12	-9	-1	2	-3	2	-3	-8	-13	-18	-23	-28	-33	-38	-43	-48	-53	-58	-63
<b>T</b>	-25	-17	-14	-6	-3	0	0	5	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55
<b>G</b>	-30	-22	-19	-11	-8	0	-2	0	3	3	-2	-7	-12	-17	-22	-27	-32	-37	-42	-47
<b>C</b>	-35	-27	-19	-16	-8	-5	-2	-4	-2	1	1	1	-4	-9	-14	-19	-24	-29	-34	-39
<b>A</b>	-40	-32	-24	-21	-13	-10	-7	-4	-1	-4	4	-1	-1	-1	-6	-11	-16	-21	-26	-31
<b>C</b>	-45	-37	-29	-26	-18	-15	-12	-9	-6	-3	-1	7	2	-3	-3	-3	-8	-13	-18	-23
<b>C</b>	-50	-42	-34	-31	-23	-20	-17	-14	-11	-8	-5	2	5	0	-5	0	-5	-5	-10	-15
<b>G</b>	-55	-47	-39	-31	-28	-20	-22	-19	-16	-8	-10	-3	0	3	3	-2	-2	-7	-7	-7

Gap = -5

Match = +3

Mismatch = -2

# Alinhamento Global: Alinhamento Downmost

$\times$	$\alpha$	<b>G</b>	<b>C</b>	<b>G</b>	<b>C</b>	<b>G</b>	<b>T</b>	<b>T</b>	<b>A</b>	<b>G</b>	<b>A</b>	<b>C</b>	<b>T</b>	<b>A</b>	<b>G</b>	<b>C</b>	<b>A</b>	<b>C</b>	<b>C</b>	<b>G</b>
$\beta$	<b>0</b>	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55	-60	-65	-70	-75	-80	-85	-90	-95
<b>G</b>	-5	<b>3</b>	<b>-2</b>	-7	-12	-17	-22	-27	-32	-37	-42	-47	-52	-57	-62	-67	-72	-77	-82	-87
<b>G</b>	-10	-2	1	<b>1</b>	<b>-4</b>	-9	-14	-19	-24	-29	-34	-39	-44	-49	-54	-59	-64	-69	-74	-79
<b>G</b>	-15	-7	-4	4	-1	<b>-1</b>	-6	-11	-16	-21	-26	-31	-36	<b>-41</b>	-46	-51	-56	-61	-66	-71
<b>T</b>	-20	-12	-9	-1	2	-3	<b>2</b>	-3	-8	-13	-18	-23	-28	-33	-38	-43	-48	-53	-58	-63
<b>T</b>	-25	-17	-14	-6	-3	0	0	<b>5</b>	<b>0</b>	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55
<b>G</b>	-30	-22	-19	-11	-8	0	-2	0	3	<b>3</b>	<b>-2</b>	-7	-12	-17	-22	-27	-32	-37	-42	-47
<b>C</b>	-35	-27	-19	-16	-8	-5	-2	-4	-2	1	1	<b>1</b>	<b>-4</b>	-9	-14	-19	-24	-29	-34	-39
<b>A</b>	-40	-32	-24	-21	-13	-10	-7	-4	-1	-4	4	-1	-1	<b>-1</b>	<b>-6</b>	-11	-16	-21	-26	-31
<b>C</b>	-45	-37	-29	-26	-18	-15	-12	-9	-6	-3	-1	7	2	-3	-3	<b>-3</b>	<b>-8</b>	-13	-18	-23
<b>C</b>	-50	-42	-34	-31	-23	-20	-17	-14	-11	-8	-5	2	5	0	-5	0	-5	<b>-5</b>	<b>-10</b>	-15
<b>G</b>	-55	-47	-39	-31	-28	-20	-22	-19	-16	-8	-10	-3	0	3	3	-2	-2	-7	-7	<b>-7</b>
$\alpha$	=	<b>G</b>	<b>C</b>	<b>G</b>	<b>C</b>	<b>G</b>	<b>T</b>	<b>T</b>	<b>A</b>	<b>G</b>	<b>A</b>	<b>C</b>	<b>T</b>	<b>A</b>	<b>G</b>	<b>C</b>	<b>A</b>	<b>C</b>	<b>C</b>	<b>G</b>
$\beta$	=	<b>G</b>	-	<b>G</b>	-	<b>G</b>	<b>T</b>	<b>T</b>	-	<b>G</b>	-	<b>C</b>	-	<b>A</b>	-	<b>C</b>	-	<b>C</b>	-	<b>G</b>
-7	=	3	-5	3	-5	3	3	3	-5	3	-5	3	-5	3	-5	3	-5	3	-5	3

# Alinhamento Global: Alinhamento Upmost

$\times$	$\alpha$	<b>G</b>	<b>C</b>	<b>G</b>	<b>C</b>	<b>G</b>	<b>T</b>	<b>T</b>	<b>A</b>	<b>G</b>	<b>A</b>	<b>C</b>	<b>T</b>	<b>A</b>	<b>G</b>	<b>C</b>	<b>A</b>	<b>C</b>	<b>C</b>	<b>G</b>
$\beta$	<b>0</b>	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55	-60	-65	-70	-75	-80	-85	-90	-95
<b>G</b>	-5	<b>3</b>	<b>-2</b>	-7	-12	-17	-22	-27	-32	-37	-42	-47	-52	-57	-62	-67	-72	-77	-82	-87
<b>G</b>	-10	-2	1	<b>1</b>	<b>-4</b>	-9	-14	-19	-24	-29	-34	-39	-44	-49	-54	-59	-64	-69	-74	-79
<b>G</b>	-15	-7	-4	4	-1	<b>-1</b>	<b>-6</b>	-11	-16	-21	-26	-31	-36	-41	-46	-51	-56	-61	-66	-71
<b>T</b>	-20	-12	-9	-1	2	-3	2	<b>-3</b>	<b>-8</b>	<b>-13</b>	<b>-18</b>	<b>-23</b>	-28	-33	-38	-43	-48	-53	-58	-63
<b>T</b>	-25	-17	-14	-6	-3	0	0	5	0	-5	-10	-15	<b>-20</b>	<b>-25</b>	-30	-35	-40	-45	-50	-55
<b>G</b>	-30	-22	-19	-11	-8	0	-2	0	3	3	-2	-7	-12	-17	<b>-22</b>	-27	-32	-37	-42	-47
<b>C</b>	-35	-27	-19	-16	-8	-5	-2	-4	-2	1	1	1	-4	-9	-14	<b>-19</b>	-24	-29	-34	-39
<b>A</b>	-40	-32	-24	-21	-13	-10	-7	-4	-1	-4	4	-1	-1	-1	-6	-11	<b>-16</b>	-21	-26	-31
<b>C</b>	-45	-37	-29	-26	-18	-15	-12	-9	-6	-3	-1	7	2	-3	-3	-3	-8	<b>-13</b>	-18	-23
<b>C</b>	-50	-42	-34	-31	-23	-20	-17	-14	-11	-8	-5	2	5	0	-5	0	-5	-5	<b>-10</b>	-15
<b>G</b>	-55	-47	-39	-31	-28	-20	-22	-19	-16	-8	-10	-3	0	3	3	-2	-2	-7	-7	<b>-7</b>
$\alpha$	=	<b>G</b>	<b>C</b>	<b>G</b>	<b>C</b>	<b>G</b>	<b>T</b>	<b>T</b>	<b>A</b>	<b>G</b>	<b>A</b>	<b>C</b>	<b>T</b>	<b>A</b>	<b>G</b>	<b>C</b>	<b>A</b>	<b>C</b>	<b>C</b>	<b>G</b>
$\beta$	=	<b>G</b>	-	<b>G</b>	-	<b>G</b>	-	<b>T</b>	-	-	-	-	<b>T</b>	-	<b>G</b>	<b>C</b>	<b>A</b>	<b>C</b>	<b>C</b>	<b>G</b>
-7	=	3	-5	3	-5	3	-5	3	-5	-5	-5	-5	3	-5	3	3	3	3	3	3

# Distância de Edição

## Problema

A distância Levenshtein ou distância de edição entre duas sequências de caracteres é dada pelo número mínimo de operações necessárias para transformar uma sequência na outra. Uma operação é definida como uma inserção, uma remoção ou uma substituição de um caracter. Foi definida pela primeira vez em 1965 por Vladimir Levenshtein.

## Exemplo

genoma

gnoma (remoção do “e”)

gnomo (substituição do “a” por um “o”)

gnomos (inserção do “s”)

## Exercício

Dadas duas sequências  $\alpha$  e  $\beta$  com, respectivamente,  $m$  e  $n$  caracteres, mostre como calcular eficientemente a distância de edição entre elas.

# Distância de Edição

- Custos das operações:
  - ▶ Inserção = 1
  - ▶ Remoção = 1
  - ▶ Substituição = 1
- Importante: a distância de edição é um problema de minimização.
- Como usar o algoritmo de Alinhamento Global como uma caixa preta para calcular a distância de edição entre duas sequências?
  - ▶ Match = 0
  - ▶ Mismatch = -1
  - ▶ Gap = -1

# Distância de Edição

$x$	$\alpha$	C	C	T	G	T	G	G	C	A	A	C
$\beta$	0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11
A	-1											
T	-2											
T	-3											
G	-4											
G	-5											
C	-6											
C	-7											
A	-8											
C	-9											

Gap = -1

Match = 0

Mismatch = -1

# Distância de Edição

$\alpha$	C	C	T	G	T	G	G	C	A	A	C	
$\beta$	0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11
A	-1	-1	-2	-3	-4	-5	-6	-7	-8	-8	-9	-10
T	-2	-2	-2	-2	-3	-4	-5	-6	-7	-8	-9	-10
T	-3	-3	-3	-2	-3	-3	-4	-5	-6	-7	-8	-9
G	-4	-4	-4	-3	-2	-3	-3	-4	-5	-6	-7	-8
G	-5	-5	-5	-4	-3	-3	-3	-3	-4	-5	-6	-7
C	-6	-5	-5	-5	-4	-4	-4	-4	-3	-4	-5	-6
C	-7	-6	-5	-6	-5	-5	-5	-5	-4	-4	-5	-5
A	-8	-7	-6	-6	-6	-6	-6	-6	-5	-4	-4	-5
C	-9	-8	-7	-7	-7	-7	-7	-7	-6	-5	-5	-4

Gap = -1

Match = 0

Mismatch = -1

# Distância de Edição

x	$\alpha$	C	C	T	G	T	G	G	C	A	A	C
$\beta$	0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11
A	-1	-1	-2	-3	-4	-5	-6	-7	-8	-8	-9	-10
T	-2	-2	-2	-2	-3	-4	-5	-6	-7	-8	-9	-10
T	-3	-3	-3	-2	-3	-3	-4	-5	-6	-7	-8	-9
G	-4	-4	-4	-3	-2	-3	-3	-4	-5	-6	-7	-8
G	-5	-5	-5	-4	-3	-3	-3	-3	-4	-5	-6	-7
C	-6	-5	-5	-5	-4	-4	-4	-4	-3	-4	-5	-6
C	-7	-6	-5	-6	-5	-5	-5	-5	-4	-4	-5	-5
A	-8	-7	-6	-6	-6	-6	-6	-6	-5	-4	-4	-5
C	-9	-8	-7	-7	-7	-7	-7	-7	-6	-5	-5	-4
$\alpha$	=	C	C	T	G	T	G	G	C	A	A	C
$\beta$	=	-	A	T	-	T	G	G	C	C	A	C

# Alinhamento Semi-Global

- Como alterar o algoritmo de Needleman e Wunsch para produzir alinhamentos semi-globais?
  - ▶ Não penalizar buracos no começo das sequências.
    - ★ Alterar a inicialização da matriz, atribuindo valor zero para o alinhamento de qualquer prefixo com a subsequência vazia.
  - ▶ Não penalizar buracos no final das sequências.
    - ★ Buscar o valor do alinhamento ótimo em todas as posições da última linha ou da última coluna da matriz.

# Alinhamento Semi-Global

---

## Algoritmo 6: Semi-Global

---

**Input:**  $\alpha, m, \beta, n$

**for** all  $i \in [0..n]$  **do**  $M[i, 0] \leftarrow 0$

**for** all  $j \in [1..m]$  **do**  $M[0, j] \leftarrow 0$

**for** all  $i \in [1..n]$  **do**

**for** all  $j \in [1..m]$  **do**

$M[i, j] \leftarrow \max \left\{ \begin{array}{l} M[i, j - 1] + \text{gap}, \\ M[i - 1, j - 1] + \sigma(\alpha[j], \beta[i]), \\ M[i - 1, j] + \text{gap} \end{array} \right\}$

**end**

**end**

$\max \leftarrow -\infty$

**for** all  $i \in [0..n]$  **do**

**if**  $M[i, m] > \max$  **then**  $\max \leftarrow M[i, m]$

**end**

**for** all  $j \in [0..m - 1]$  **do**

**if**  $M[n, j] > \max$  **then**  $\max \leftarrow M[n, j]$

**end**

**return**  $\max$

---

# Alinhamento Semi-Global

x	$\alpha$	A	T	C	T	T	C	G	T	T	A	T	C	A	C	G	C	A	C	T	A
$\beta$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	0																				
T	0																				
T	0																				
G	0																				
G	0																				
C	0																				
C	0																				
A	0																				
A	0																				
T	0																				
C	0																				
C	0																				
C	0																				
G	0																				
C	0																				

Gap = -5

Match = +3

Mismatch = -2

# Alinhamento Semi-Global

$\alpha$	A	T	C	T	T	C	G	T	T	A	T	C	A	C	G	C	A	C	T	A
$\beta$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<b>C</b>	0	-2	-2	3	-2	-2	3	-2	-2	-2	-2	3	-2	3	-2	3	-2	3	-2	-2
<b>T</b>	0	-2	1	-2	6	1	-2	1	1	-4	1	-2	1	-2	1	-2	1	-2	1	6
<b>T</b>	0	-2	1	-1	1	9	4	-1	4	-1	-1	-1	-4	-1	-4	-1	-4	-1	-1	4
<b>G</b>	0	-2	-4	-1	-3	4	7	7	2	2	2	-3	-3	-3	-6	2	-3	-3	-6	-1
<b>G</b>	0	-2	-4	-6	-3	-1	2	10	5	0	0	0	-5	-5	-5	-3	0	-5	-5	-8
<b>C</b>	0	-2	-4	-1	-6	-5	2	5	8	3	-2	-2	3	-2	-2	-7	0	-2	-2	-7
<b>C</b>	0	-2	-4	-1	-3	-8	-2	0	3	6	1	-4	1	1	1	-4	-4	-2	1	-4
<b>A</b>	0	3	-2	-6	-3	-5	-7	-4	-2	1	9	4	-1	4	-1	-1	-6	-1	-4	-1
<b>A</b>	0	3	1	-4	-8	-5	-7	-9	-6	-4	4	7	2	2	2	-3	-3	-3	-6	2
<b>T</b>	0	-2	6	1	-1	-5	-7	-9	-6	-3	-1	7	5	0	0	0	-5	-5	-5	0
<b>C</b>	0	-2	1	9	4	-1	-2	-7	-11	-8	-5	2	10	5	3	-2	3	-2	-2	-5
<b>C</b>	0	-2	-4	4	7	2	2	-3	-8	-13	-10	-3	5	8	8	3	1	1	-4	-7
<b>C</b>	0	-2	-4	-1	2	5	5	0	-5	-10	-15	-8	0	3	11	6	6	1	4	-1
<b>G</b>	0	-2	-4	-6	-3	0	3	8	3	-2	-7	-12	-5	-2	6	14	9	4	-1	2
<b>C</b>	0	-2	-4	-1	-6	-5	3	3	6	1	-4	-9	-9	-7	1	9	17	12	7	2

Gap = -5

Match = +3

Mismatch = -2

# Alinhamento Semi-Global

$\alpha$	A	T	C	T	T	C	G	T	T	A	T	C	A	C	G	C	A	C	T	A
$\beta$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<b>C</b>	0	-2	-2	3	-2	-2	3	-2	-2	-2	-2	3	-2	3	-2	3	-2	3	-2	-2
<b>T</b>	0	-2	1	-2	6	1	-2	1	1	-4	1	-2	1	-2	1	-2	1	-2	1	6
<b>T</b>	0	-2	1	-1	1	9	4	-1	4	-1	-1	-1	-4	-1	-4	-1	-4	-1	-1	4
<b>G</b>	0	-2	-4	-1	-3	4	7	7	2	2	2	-3	-3	-3	-6	2	-3	-3	-6	-1
<b>G</b>	0	-2	-4	-6	-3	-1	2	10	5	0	0	0	-5	-5	-5	-3	0	-5	-5	-8
<b>C</b>	0	-2	-4	-1	-6	-5	2	5	8	3	-2	-2	3	-2	-2	-7	0	-2	-2	-7
<b>C</b>	0	-2	-4	-1	-3	-8	-2	0	3	6	1	-4	1	1	1	-4	-4	-2	1	-9
<b>A</b>	0	3	-2	-6	-3	-5	-7	-4	-2	1	9	4	-1	4	-1	-1	-6	-1	-4	-1
<b>A</b>	0	3	1	-4	-8	-5	-7	-9	-6	-4	4	7	2	2	2	-3	-3	-3	-6	2
<b>T</b>	0	-2	6	1	-1	-5	-7	-9	-6	-3	-1	7	5	0	0	0	-5	-5	-5	0
<b>C</b>	0	-2	1	9	4	-1	-2	-7	-11	-8	-5	2	10	5	3	-2	3	-2	-2	-5
<b>C</b>	0	-2	-4	4	7	2	2	-3	-8	-13	-10	-3	5	8	8	3	1	1	-4	-7
<b>C</b>	0	-2	-4	-1	2	5	5	0	-5	-10	-15	-8	0	3	11	6	6	1	4	-1
<b>G</b>	0	-2	-4	-6	-3	0	3	8	3	-2	-7	-12	-5	-2	6	14	9	4	-1	2
<b>C</b>	0	-2	-4	-1	-6	-5	3	3	6	1	-4	-9	-9	-7	1	9	17	12	7	2

Gap = -5

Match = +3

Mismatch = -2

# Alinhamento Semi-Global

$\alpha$	$\alpha$	A	T	C	T	T	C	G	T	T	A	T	C	A	C	G	C	A	C	T	A
$\beta$	$\beta$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<b>C</b>	0	-2	-2	3	-2	-2	3	-2	-2	-2	-2	-2	3	-2	3	-2	3	-2	3	-2	-2
<b>T</b>	0	-2	1	-2	6	1	-2	1	1	1	-4	1	-2	1	-2	1	-2	1	-2	1	6
<b>T</b>	0	-2	1	-1	1	9	4	-1	4	4	-1	-1	-1	-4	-1	-4	-1	-4	-1	-4	1
<b>G</b>	0	-2	-4	-1	-3	4	7	7	2	2	2	-3	-3	-3	-6	2	-3	-3	-6	-3	-1
<b>G</b>	0	-2	-4	-6	-3	-1	2	10	5	0	0	0	-5	-5	-5	-3	0	-5	-5	-8	-5
<b>C</b>	0	-2	-4	-1	-6	-5	2	5	8	3	-2	-2	3	-2	-2	-7	0	-2	-2	-7	-10
<b>C</b>	0	-2	-4	-1	-3	-8	-2	0	3	6	1	-4	1	1	1	-4	-4	-2	1	-4	-9
<b>A</b>	0	3	-2	-6	-3	-5	-7	-4	-2	1	9	4	-1	4	-1	-1	-6	-1	-4	-1	-1
<b>A</b>	0	3	1	-4	-8	-5	-7	-9	-6	-4	4	7	2	2	2	-3	-3	-3	-3	-6	2
<b>T</b>	0	-2	6	1	-1	-5	-7	-9	-6	-3	-1	7	5	0	0	0	-5	-5	-5	0	-3
<b>C</b>	0	-2	1	9	4	-1	-2	-7	-11	-8	-5	2	10	5	3	-2	3	-2	-2	-5	-2
<b>C</b>	0	-2	-4	4	7	2	2	-3	-8	-13	-10	-3	5	8	8	3	1	1	1	-4	-7
<b>C</b>	0	-2	-4	-1	2	5	5	0	-5	-10	-15	-8	0	3	11	6	6	1	4	-1	-6
<b>G</b>	0	-2	-4	-6	-3	0	3	8	3	-2	-7	-12	-5	-2	6	14	9	4	-1	2	-3
<b>C</b>	0	-2	-4	-1	-6	-5	3	3	6	1	-4	-9	-9	-7	1	9	17	12	7	2	0
$\alpha$																	<b>C</b>	<b>A</b>	<b>C</b>	<b>T</b>	<b>A</b>
$\beta$																	<b>C</b>	-	-	-	-

# Alinhamento Semi-Global

$\alpha$	$\alpha$	A	T	C	T	T	C	G	T	T	A	T	C	A	C	G	C	A	C	T	A
$\beta$	$\beta$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<b>C</b>	0	-2	-2	3	-2	-2	3	-2	-2	-2	-2	-2	3	-2	3	-2	3	-2	3	-2	-2
<b>T</b>	0	-2	1	-2	6	1	-2	1	1	1	-4	1	-2	1	-2	1	-2	1	-2	1	6
<b>T</b>	0	-2	1	-1	1	9	4	-1	4	4	-1	-1	-1	-4	-1	-4	-1	-4	-1	-4	1
<b>G</b>	0	-2	-4	-1	-3	4	7	7	2	2	2	-3	-3	-3	-6	2	-3	-3	-6	-3	-1
<b>G</b>	0	-2	-4	-6	-3	-1	2	10	5	0	0	0	-5	-5	-5	-3	0	-5	-5	-8	-5
<b>C</b>	0	-2	-4	-1	-6	-5	2	5	8	3	-2	-2	3	-2	-2	-7	0	-2	-2	-7	-10
<b>C</b>	0	-2	-4	-1	-3	-8	-2	0	3	6	1	-4	1	1	1	-4	-4	-2	1	-4	-9
<b>A</b>	0	3	-2	-6	-3	-5	-7	-4	-2	1	9	4	-1	4	-1	-1	-6	-1	-4	-1	-1
<b>A</b>	0	3	1	-4	-8	-5	-7	-9	-6	-4	4	7	2	2	2	-3	-3	-3	-3	-6	2
<b>T</b>	0	-2	6	1	-1	-5	-7	-9	-6	-3	-1	7	5	0	0	0	-5	-5	-5	0	-3
<b>C</b>	0	-2	1	9	4	-1	-2	-7	-11	-8	-5	2	10	5	3	-2	3	-2	-2	-5	-2
<b>C</b>	0	-2	-4	4	7	2	2	-3	-8	-13	-10	-3	5	8	8	3	1	1	1	-4	-7
<b>C</b>	0	-2	-4	-1	2	5	5	0	-5	-10	-15	-8	0	3	11	6	6	1	4	-1	-6
<b>G</b>	0	-2	-4	-6	-3	0	3	8	3	-2	-7	-12	-5	-2	6	14	9	4	-1	2	-3
<b>C</b>	0	-2	-4	-1	-6	-5	3	3	6	1	-4	-9	-9	-7	1	9	17	12	7	2	0
$\alpha$											-	A	C	G	C	A	C	T	A		
$\beta$											C	C	C	G	C	-	-	-	-	-	

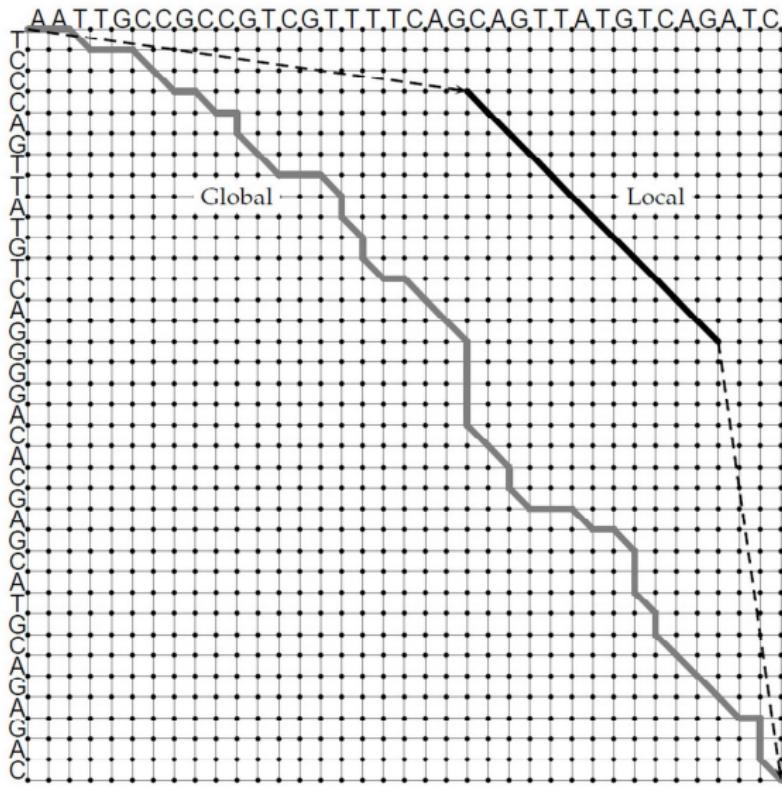
# Alinhamento Semi-Global

$\times$	$\alpha$	A	T	C	T	T	C	G	T	T	A	T	C	A	C	G	C	A	C	T	A
$\beta$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	0	-2	-2	3	-2	-2	3	-2	-2	-2	-2	3	-2	3	-2	3	-2	3	-2	3	-2
T	0	-2	1	-2	6	1	-2	1	1	1	-4	1	-2	1	-2	1	-2	1	-2	1	6
T	0	-2	1	-1	1	9	4	-1	4	4	-1	-1	-1	-4	-1	-4	-1	-4	-1	-4	1
G	0	-2	-4	-1	-3	4	7	7	2	2	2	-3	-3	-3	-6	2	-3	-3	-6	-3	-1
G	0	-2	-4	-6	-3	-1	2	10	5	0	0	0	-5	-5	-5	-3	0	-5	-5	-8	-5
C	0	-2	-4	-1	-6	-5	2	5	8	3	-2	-2	3	-2	-2	-7	0	-2	-2	-7	-10
C	0	-2	-4	-1	-3	-8	-2	0	3	6	1	-4	1	1	1	-4	-4	-2	1	-4	-9
A	0	3	-2	-6	-3	-5	-7	-4	-2	1	9	4	-1	4	-1	-1	-6	-1	-4	-1	-1
A	0	3	1	-4	-8	-5	-7	-9	-6	-4	4	7	2	2	2	-3	-3	-3	-3	-6	2
T	0	-2	6	1	-1	-5	-7	-9	-6	-3	-1	7	5	0	0	0	-5	-5	-5	0	-3
C	0	-2	1	9	4	-1	-2	-7	-11	-8	-5	2	10	5	3	-2	3	-2	-2	-5	-2
C	0	-2	-4	4	7	2	2	-3	-8	-13	-10	-3	5	8	8	3	1	1	1	-4	-7
C	0	-2	-4	-1	2	5	5	0	-5	-10	-15	-8	0	3	11	6	6	1	4	-1	-6
G	0	-2	-4	-6	-3	0	3	8	3	-2	-7	-12	-5	-2	6	14	9	4	-1	2	-3
C	0	-2	-4	-1	-6	-5	3	3	6	1	-4	-9	-9	-7	1	9	17	12	7	2	0
$\alpha$	C	T	T	C	G	T	T	A	-	T	C	A	C	C	G	C	A	C	T	A	
$\beta$	C	T	T	G	G	C	C	A	A	T	C	C	C	G	C	-	-	-	-	-	

# Alinhamento Semi-Global

$\alpha$	A	T	C	T	T	C	G	T	T	A	T	C	A	C	G	C	A	C	T	A
$\beta$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<b>C</b>	0	-2	-2	3	-2	-2	3	-2	-2	-2	-2	3	-2	3	-2	3	-2	3	-2	-2
<b>T</b>	0	-2	1	-2	6	1	-2	1	1	-4	1	-2	1	-2	1	-2	1	-2	6	1
<b>T</b>	0	-2	1	-1	1	9	4	-1	4	-1	-1	-1	-4	-1	-4	-1	-4	-1	1	4
<b>G</b>	0	-2	-4	-1	-3	4	7	7	2	2	2	-3	-3	-3	-6	2	-3	-3	-6	-1
<b>G</b>	0	-2	-4	-6	-3	-1	2	10	5	0	0	0	-5	-5	-5	-3	0	-5	-5	-8
<b>C</b>	0	-2	-4	-1	-6	-5	2	5	8	3	-2	-2	3	-2	-2	-7	0	-2	-2	-10
<b>C</b>	0	-2	-4	-1	-3	-8	-2	0	3	6	1	-4	1	1	1	-4	-4	-2	1	-9
<b>A</b>	0	3	-2	-6	-3	-5	-7	-4	-2	1	9	4	-1	4	-1	-1	-6	-1	-4	-1
<b>A</b>	0	3	1	-4	-8	-5	-7	-9	-6	-4	4	7	2	2	2	-3	-3	-3	-6	2
<b>T</b>	0	-2	6	1	-1	-5	-7	-9	-6	-3	-1	7	5	0	0	0	-5	-5	-5	0
<b>C</b>	0	-2	1	9	4	-1	-2	-7	-11	-8	-5	2	10	5	3	-2	3	-2	-2	-2
<b>C</b>	0	-2	-4	4	7	2	2	-3	-8	-13	-10	-3	5	8	8	3	1	1	-4	-7
<b>C</b>	0	-2	-4	-1	2	5	5	0	-5	-10	-15	-8	0	3	11	6	6	1	4	-1
<b>G</b>	0	-2	-4	-6	-3	0	3	8	3	-2	-7	-12	-5	-2	6	14	9	4	-1	2
<b>C</b>	0	-2	-4	-1	-6	-5	3	3	6	1	-4	-9	-9	-7	1	9	17	12	7	2
$\alpha$	A	T	C	T	T	C	G	T	T	A	-	T	C	A	C	G	C	A	C	A
$\beta$	-	-	C	T	T	G	G	C	C	A	A	T	T	C	C	C	G	C	-	-

# Alinhamento Global x Alinhamento Local



## Alinhamento Local - Abordagem Ingênua

- O alinhamento local ótimo entre duas sequências  $\alpha$  e  $\beta$  é o melhor alinhamento (global) entre uma subsequência de  $\alpha$  e uma subsequência de  $\beta$ .
- É possível usar o algoritmo de alinhamento global de Needleman e Wunsch para produzir alinhamentos locais?
- Quantas subsequências (não nulas) as sequências  $\alpha$  e  $\beta$ , formadas respectivamente por  $m$  e  $n$  caracteres, possuem?
  - ▶ Respectivamente,  $m(m + 1)/2 = \Theta(m^2)$  e  $n(n + 1)/2 = \Theta(n^2)$  subsequências.
- Algoritmo ingênuo: testar todos os pares de subsequências e escolher aquele para com o maior valor de alinhamento global.
- Qual a complexidade do algoritmo ingênuo para alinhamento local?

# Análise de Complexidade

- Complexidade:

$$T(m, n) = \sum_{\alpha'=1}^m \sum_{\alpha''=\alpha'}^m \sum_{\beta'=1}^n \sum_{\beta''=\beta'}^n \Theta((\alpha'' - \alpha' + 1)(\beta'' - \beta' + 1))$$

- Limite superior:

$$T(m, n) \leq \frac{m(m+1)}{2} \frac{n(n+1)}{2} O(mn) = O(m^3 n^3)$$

- Limite inferior:

$$T(m, n) \geq \sum_{\alpha'=1}^{\lfloor \frac{m}{3} \rfloor} \sum_{\alpha''=m-\lfloor \frac{m}{3} \rfloor+1}^m \sum_{\beta'=1}^{\lfloor \frac{n}{3} \rfloor} \sum_{\beta''=n-\lfloor \frac{n}{3} \rfloor+1}^n \Omega(\left\lceil \frac{m}{3} \right\rceil \left\lceil \frac{n}{3} \right\rceil)$$

$$T(m, n) \geq \left\lceil \frac{m}{3} \right\rceil \left\lceil \frac{m}{3} \right\rceil \left\lceil \frac{n}{3} \right\rceil \left\lceil \frac{n}{3} \right\rceil \Omega(\left\lceil \frac{m}{3} \right\rceil \left\lceil \frac{n}{3} \right\rceil) = \Omega(m^3 n^3)$$

# Complexidade e Tempo de Execução

- Complexidade:
  - ▶  $T(m, n) = \Theta(m^3 n^3)$
- Exemplo: computador de 3GHz (1 instrução por ciclo)
  - ▶  $m = n = 10$ :
    - ★ Tempo:  $10^6 / (3 \times 2^{30}) = 310$  bilionésimos de segundo
  - ▶  $m = n = 100$ :
    - ★ Tempo:  $100^6 / (3 \times 2^{30}) = 310$  segundos
  - ▶  $m = n = 1000$ :
    - ★ Tempo:  $1000^6 / (3 \times 2^{30}) = 10$  anos
  - ▶  $m = n = 10000$ :
    - ★ Tempo:  $10000^6 / (3 \times 2^{30}) = 10$  bilhões de anos

# Alinhamento Local

- Como alterar o algoritmo de Needleman e Wunsch para produzir alinhamentos locais?
  - ▶ Redefinir o significado das células da matriz M.
    - ★ A célula  $M[i,j]$  deve representar o valor do alinhamento ótimo entre todos os sufixos das sequências  $\alpha[1..j]$  e  $\beta[1..i]$ .
  - ▶ Qual o valor mínimo que uma célula pode armazenar?
    - ★ O valor correspondente ao alinhamento de um sufixo vazio com qualquer outro, ou seja, zero.
  - ▶ Qual o valor do alinhamento local ótimo entre as duas sequências?
    - ★ O maior valor armazenado em qualquer uma das posições da matriz.
  - ▶ Como recuperar o alinhamento local ótimo?
    - ★ O alinhamento deve começar numa célula com valor zero e terminar na célula de maior valor da matriz (obter alinhamento na ordem inversa).
- Em 1981, Temple Smith e Michael Waterman propuseram o algoritmo de programação dinâmica para alinhamento local de sequências.

# Alinhamento Local

## Algoritmo 7: Local

**Input:**  $\alpha, m, \beta, n$

**for** all  $i \in [0..n]$  **do**  $M[i, 0] \leftarrow 0$

**for** all  $j \in [1..m]$  **do**  $M[0, j] \leftarrow 0$

**for** all  $i \in [1..n]$  **do**

**for** all  $j \in [1..m]$  **do**

$M[i, j] \leftarrow \max \left\{ \begin{array}{l} 0, \\ M[i, j - 1] + gap, \\ M[i - 1, j - 1] + \sigma(\alpha[j], \beta[i]), \\ M[i - 1, j] + gap \end{array} \right\}$

**end**

**end**

$max \leftarrow 0$

**for** all  $i \in [1..n]$  **do**

**for** all  $j \in [1..m]$  **do**

**if**  $M[i, j] > max$  **then**  $max \leftarrow M[i, j]$

**end**

**end**

**return**  $max$

# Alinhamento Local

$\times$	$\alpha$	G	A	C	A	A	C	G	T	T	A	C	T	G	C	T	T	A	C	T	A
$\beta$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	0																				
T	0																				
T	0																				
G	0																				
G	0																				
C	0																				
C	0																				
A	0																				
C	0																				
T	0																				
C	0																				
C	0																				
C	0																				
G	0																				
C	0																				

Gap = -5

Match = +3

Mismatch = -2

# Alinhamento Local

x	α	G	A	C	A	A	C	G	T	T	A	C	T	G	C	T	T	T	A	C	T	A	
β	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	0	0	0	3	0	0	3	0	0	0	0	3	0	0	3	0	0	0	0	3	0	0	0
T	0	0	0	0	1	0	0	1	3	3	0	0	6	1	0	6	3	0	0	6	1		
T	0	0	0	0	0	0	0	0	4	6	1	0	3	4	0	3	9	4	0	3	4		
G	0	3	0	0	0	0	0	3	0	2	4	0	0	6	2	0	4	7	2	0	1		
G	0	3	1	0	0	0	0	3	1	0	0	2	0	3	4	0	0	2	5	0	0		
C	0	0	1	4	0	0	3	0	1	0	0	3	0	0	6	2	0	0	5	3	0		
C	0	0	0	4	2	0	3	1	0	0	0	3	1	0	3	4	0	0	3	3	1		
A	0	0	3	0	7	5	0	1	0	0	3	0	1	0	0	1	2	3	0	1	6		
C	0	0	0	6	2	5	8	3	0	0	0	6	1	0	3	0	0	0	6	1	1		
T	0	0	0	1	4	0	3	6	6	3	0	1	9	4	0	6	3	0	1	9	4		
C	0	0	0	3	0	2	3	1	4	4	1	3	4	7	7	2	4	1	3	4	7		
C	0	0	0	3	1	0	5	1	0	2	2	4	1	2	10	5	0	2	4	1	2		
C	0	0	0	3	1	0	3	3	0	0	0	5	2	0	5	8	3	0	5	2	0		
G	0	3	0	0	1	0	0	6	1	0	0	0	3	5	0	3	6	1	0	3	0		
C	0	0	1	3	0	0	3	1	4	0	0	3	0	1	8	3	1	4	4	0	1		

Gap = -5

Match = +3

Mismatch = -2

# Alinhamento Local

x	α	G	A	C	A	A	C	G	T	T	A	C	T	G	C	T	T	T	A	C	T	A	
β	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	0	0	0	3	0	0	3	0	0	0	0	3	0	0	3	0	0	0	0	3	0	0	0
T	0	0	0	0	1	0	0	1	3	3	0	0	6	1	0	6	3	0	0	6	1		
T	0	0	0	0	0	0	0	0	4	6	1	0	3	4	0	3	9	4	0	3	4		
G	0	3	0	0	0	0	0	3	0	2	4	0	0	6	2	0	4	7	2	0	1		
G	0	3	1	0	0	0	0	3	1	0	0	2	0	3	4	0	0	2	5	0	0		
C	0	0	1	4	0	0	3	0	1	0	0	3	0	0	6	2	0	0	5	3	0		
C	0	0	0	4	2	0	3	1	0	0	0	3	1	0	3	4	0	0	3	3	1		
A	0	0	3	0	7	5	0	1	0	0	3	0	1	0	0	1	2	3	0	1	6		
C	0	0	0	6	2	5	8	3	0	0	0	6	1	0	3	0	0	0	6	1	1		
T	0	0	0	1	4	0	3	6	6	3	0	1	9	4	0	6	3	0	1	9	4		
C	0	0	0	3	0	2	3	1	4	4	1	3	4	7	7	2	4	1	3	4	7		
C	0	0	0	3	1	0	5	1	0	2	2	4	1	2	10	5	0	2	4	1	2		
C	0	0	0	3	1	0	3	3	0	0	0	5	2	0	5	8	3	0	5	2	0		
G	0	3	0	0	1	0	0	6	1	0	0	0	3	5	0	3	6	1	0	3	0		
C	0	0	1	3	0	0	3	1	4	0	0	3	0	1	8	3	1	4	4	0	1		

Gap = -5

Match = +3

Mismatch = -2

# Alinhamento Local

x	α	G	A	C	A	A	C	G	T	T	A	C	T	G	C	T	T	T	A	C	T	A
β		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C		0	0	0	3	0	0	3	0	0	0	3	0	0	3	0	0	0	0	3	0	0
T		0	0	0	0	1	0	0	1	3	3	0	0	6	1	0	6	3	0	0	6	1
T		0	0	0	0	0	0	0	4	6	1	0	3	4	0	3	9	4	0	3	4	
G		0	3	0	0	0	0	0	3	0	2	4	0	0	6	2	0	4	7	2	0	1
G		0	3	1	0	0	0	0	3	1	0	0	2	0	3	4	0	0	2	5	0	0
C		0	0	1	4	0	0	3	0	1	0	0	3	0	0	6	2	0	0	5	3	0
C		0	0	0	4	2	0	3	1	0	0	0	3	1	0	3	4	0	0	3	3	1
A		0	0	3	0	7	5	0	1	0	0	3	0	1	0	0	1	2	3	0	1	6
C		0	0	0	6	2	5	8	3	0	0	0	6	1	0	3	0	0	0	6	1	1
T		0	0	0	1	4	0	3	6	6	3	0	1	9	4	0	6	3	0	1	9	4
C		0	0	0	3	0	2	3	1	4	4	1	3	4	7	7	2	4	1	3	4	7
C		0	0	0	3	1	0	5	1	0	2	2	4	1	2	10	5	0	2	4	1	2
C		0	0	0	3	1	0	3	3	0	0	0	5	2	0	5	8	3	0	5	2	0
G		0	3	0	0	1	0	0	6	1	0	0	0	3	5	0	3	6	1	0	3	0
C		0	0	1	3	0	0	3	1	4	0	0	3	0	1	8	3	1	4	4	0	1

α

β

C

C

# Alinhamento Local

$\times$	$\alpha$	G	A	C	A	A	C	G	T	T	A	C	T	G	C	T	T	A	C	T	A
$\beta$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	0	0	0	3	0	0	3	0	0	0	0	3	0	0	3	0	0	0	3	0	0
T	0	0	0	0	1	0	0	1	3	3	0	0	6	1	0	6	3	0	0	6	1
T	0	0	0	0	0	0	0	0	4	6	1	0	3	4	0	3	9	4	0	3	4
G	0	3	0	0	0	0	0	3	0	2	4	0	0	6	2	0	4	7	2	0	1
G	0	3	1	0	0	0	0	3	1	0	0	2	0	3	4	0	0	2	5	0	0
C	0	0	1	4	0	0	3	0	1	0	0	3	0	0	6	2	0	0	5	3	0
C	0	0	0	4	2	0	3	1	0	0	0	3	1	0	3	4	0	0	3	3	1
A	0	0	3	0	7	5	0	1	0	0	3	0	1	0	0	1	2	3	0	1	6
C	0	0	0	6	2	5	8	3	0	0	0	6	1	0	3	0	0	0	6	1	1
T	0	0	0	1	4	0	3	6	6	3	0	1	9	4	0	6	3	0	1	9	4
C	0	0	0	3	0	2	3	1	4	4	1	3	4	7	7	2	4	1	3	4	7
C	0	0	0	3	1	0	5	1	0	2	2	4	1	2	10	5	0	2	4	1	2
C	0	0	0	3	1	0	3	3	0	0	0	5	2	0	5	8	3	0	5	2	0
G	0	3	0	0	1	0	0	6	1	0	0	0	3	5	0	3	6	1	0	3	0
C	0	0	1	3	0	0	3	1	4	0	0	3	0	1	8	3	1	4	4	0	1

$\alpha$   
 $\beta$

C C  
G C

# Alinhamento Local

x	α	G	A	C	A	A	C	G	T	T	A	C	T	G	C	T	T	T	A	C	T	A
β	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	0	0	0	3	0	0	3	0	0	0	0	3	0	0	3	0	0	0	0	3	0	0
T	0	0	0	0	1	0	0	1	3	3	0	0	6	1	0	6	3	0	0	6	1	
T	0	0	0	0	0	0	0	0	4	6	1	0	3	4	0	3	9	4	0	3	4	
G	0	3	0	0	0	0	0	3	0	2	4	0	0	6	2	0	4	7	2	0	1	
G	0	3	1	0	0	0	0	3	1	0	0	2	0	3	4	0	0	2	5	0	0	
C	0	0	1	4	0	0	3	0	1	0	0	3	0	0	6	2	0	0	5	3	0	
C	0	0	0	4	2	0	3	1	0	0	0	3	1	0	3	4	0	0	3	3	1	
A	0	0	3	0	7	5	0	1	0	0	3	0	1	0	0	1	2	3	0	1	6	
C	0	0	0	6	2	5	8	3	0	0	0	6	1	0	3	0	0	0	6	1	1	
T	0	0	0	1	4	0	3	6	6	3	0	1	9	4	0	6	3	0	1	9	4	
C	0	0	0	3	0	2	3	1	4	4	1	3	4	7	7	2	4	1	3	4	7	
C	0	0	0	3	1	0	5	1	0	2	2	4	1	2	10	5	0	2	4	1	2	
C	0	0	0	3	1	0	3	3	0	0	0	5	2	0	5	8	3	0	5	2	0	
G	0	3	0	0	1	0	0	6	1	0	0	0	3	5	0	3	6	1	0	3	0	
C	0	0	1	3	0	0	3	1	4	0	0	3	0	1	8	3	1	4	4	0	1	

α

β

T C C

T G C

# Alinhamento Local

$\times$	$\alpha$	G	A	C	A	A	C	G	T	T	A	C	T	G	C	T	T	A	C	T	A
$\beta$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	0	0	0	3	0	0	3	0	0	0	0	3	0	0	3	0	0	0	3	0	0
T	0	0	0	0	1	0	0	1	3	3	0	0	6	1	0	6	3	0	0	6	1
T	0	0	0	0	0	0	0	0	4	6	1	0	3	4	0	3	9	4	0	3	4
G	0	3	0	0	0	0	0	3	0	2	4	0	0	6	2	0	4	7	2	0	1
G	0	3	1	0	0	0	0	3	1	0	0	2	0	3	4	0	0	2	5	0	0
C	0	0	1	4	0	0	3	0	1	0	0	3	0	0	6	2	0	0	5	3	0
C	0	0	0	4	2	0	3	1	0	0	0	3	1	0	3	4	0	0	3	3	1
A	0	0	3	0	7	5	0	1	0	0	3	0	1	0	0	1	2	3	0	1	6
C	0	0	0	6	2	5	8	3	0	0	0	6	1	0	3	0	0	0	6	1	1
T	0	0	0	1	4	0	3	6	6	3	0	1	9	4	0	6	3	0	1	9	4
C	0	0	0	3	0	2	3	1	4	4	1	3	4	7	7	2	4	1	3	4	7
C	0	0	0	3	1	0	5	1	0	2	2	4	1	2	10	5	0	2	4	1	2
C	0	0	0	3	1	0	3	3	0	0	0	5	2	0	5	8	3	0	5	2	0
G	0	3	0	0	1	0	0	6	1	0	0	0	3	5	0	3	6	1	0	3	0
C	0	0	1	3	0	0	3	1	4	0	0	3	0	1	8	3	1	4	4	0	1

C   T   C   C  
C   T   G   C

$\alpha$

$\beta$

# Alinhamento Local

$\times$	$\alpha$	G	A	C	A	A	C	G	T	T	A	C	T	G	C	T	T	A	C	T	A
$\beta$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	0	0	0	3	0	0	3	0	0	0	0	3	0	0	3	0	0	0	3	0	0
T	0	0	0	0	1	0	0	1	3	3	0	0	6	1	0	6	3	0	0	6	1
T	0	0	0	0	0	0	0	0	4	6	1	0	3	4	0	3	9	4	0	3	4
G	0	3	0	0	0	0	0	3	0	2	4	0	0	6	2	0	4	7	2	0	1
G	0	3	1	0	0	0	0	3	1	0	0	2	0	3	4	0	0	2	5	0	0
C	0	0	1	4	0	0	3	0	1	0	0	3	0	0	6	2	0	0	5	3	0
C	0	0	0	4	2	0	3	1	0	0	0	3	1	0	3	4	0	0	3	3	1
A	0	0	3	0	7	5	0	1	0	0	3	0	1	0	0	1	2	3	0	1	6
C	0	0	0	6	2	5	8	3	0	0	0	6	1	0	3	0	0	0	6	1	1
T	0	0	0	1	4	0	3	6	6	3	0	1	9	4	0	6	3	0	1	9	4
C	0	0	0	3	0	2	3	1	4	4	1	3	4	7	7	2	4	1	3	4	7
C	0	0	0	3	1	0	5	1	0	2	2	4	1	2	10	5	0	2	4	1	2
C	0	0	0	3	1	0	3	3	0	0	0	5	2	0	5	8	3	0	5	2	0
G	0	3	0	0	1	0	0	6	1	0	0	0	3	5	0	3	6	1	0	3	0
C	0	0	1	3	0	0	3	1	4	0	0	3	0	1	8	3	1	4	4	0	1

A   C   T   C   C  
A   C   T   G   C

$\alpha$

$\beta$

# Alinhamento Local

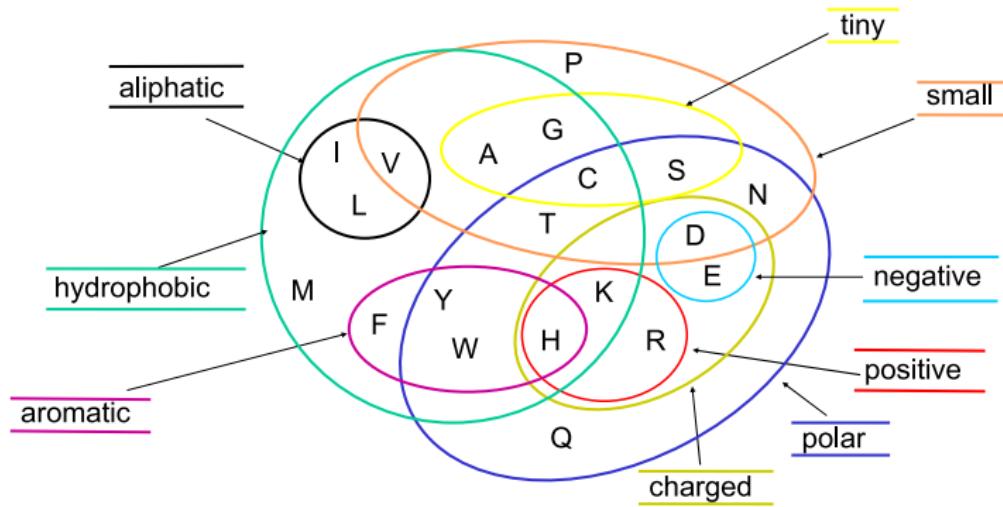
$\times$	$\alpha$	G	A	C	A	A	C	G	T	T	A	C	T	G	C	T	T	A	C	T	A
$\beta$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	0	0	0	3	0	0	3	0	0	0	0	3	0	0	3	0	0	0	3	0	0
T	0	0	0	0	1	0	0	1	3	3	0	0	6	1	0	6	3	0	0	6	1
T	0	0	0	0	0	0	0	0	4	6	1	0	3	4	0	3	9	4	0	3	4
G	0	3	0	0	0	0	0	3	0	2	4	0	0	6	2	0	4	7	2	0	1
G	0	3	1	0	0	0	0	3	1	0	0	2	0	3	4	0	0	2	5	0	0
C	0	0	1	4	0	0	3	0	1	0	0	3	0	0	6	2	0	0	5	3	0
C	0	0	0	4	2	0	3	1	0	0	0	3	1	0	3	4	0	0	3	3	1
A	0	0	3	0	7	5	0	1	0	0	3	0	1	0	0	1	2	3	0	1	6
C	0	0	0	6	2	5	8	3	0	0	0	6	1	0	3	0	0	0	6	1	1
T	0	0	0	1	4	0	3	6	6	3	0	1	9	4	0	6	3	0	1	9	4
C	0	0	0	3	0	2	3	1	4	4	1	3	4	7	7	2	4	1	3	4	7
C	0	0	0	3	1	0	5	1	0	2	2	4	1	2	10	5	0	2	4	1	2
C	0	0	0	3	1	0	3	3	0	0	0	5	2	0	5	8	3	0	5	2	0
G	0	3	0	0	1	0	0	6	1	0	0	0	3	5	0	3	6	1	0	3	0
C	0	0	1	3	0	0	3	1	4	0	0	3	0	1	8	3	1	4	4	0	1

A   C   T   C   C  
A   C   T   G   C

# Pontuação para Alinhamento de Caracteres

- Simples (para nucleotídeos):
  - ▶ Match: alinhamento de dois caracteres iguais
  - ▶ Mismatch: alinhamento de dois caracteres diferentes
- Matrizes de Substituições (para aminoácidos):
  - ▶ Família PAM
  - ▶ Família BLOSUM

# Propriedades dos Aminoácidos



## Família PAM

- Desenvolvida por Margaret Dayhoff (1978).
- PAM: *Percent (or Point) of Accepted Mutations.*
- Distância 1-PAM: uma mutação (substituição) a cada 100 aminoácidos em média.
- Duas proteínas com distância  $k$ -PAM não necessariamente possuem  $k\%$  de diferença entre seus aminoácidos.
- A similariedade esperada entre duas sequências protéicas com distância 200-PAM é de cerca de 25%.
- Seja  $F^k(i,j)$  a probabilidade de uma aminoácido  $a_i$  mutar para o aminoácido  $a_j$  em sequências com distância  $k$ -PAM e seja  $\text{freq}(a_j)$  a frequência do aminoácido  $a_j$ .
- Fórmula geral para matrizes PAM:

$$\text{PAM}_k(i,j) = \left\lceil 10 \times \log \frac{F^k(i,j)}{\text{freq}(a_j)} \right\rceil$$

- Valores mais utilizados:  $k = 40$ ,  $k = 120$  e  $k = 250$ .

## Família BLOSUM

- Desenvolvida por Steven Henikoff e Jorja Henikoff (1992).
- BLOSUM: *BLOcks of Amino Acid SUbstitution Matrix*
- BLOCKS: banco de dados de alinhamentos múltiplos de blocos conservados (sem gaps) de 504 grupos de proteínas.
- $BLOSUM_k$  considera apenas sequências com uma similaridade maior do que  $k\%$  entre suas bases.
- Seja  $freq_k(i, j)$  a frequência de alinhamentos de aminoácidos  $a_i$  com aminoácidos  $a_j$  e seja  $freq_k(a_j)$  a frequência do aminoácido  $a_j$ , em ambos os casos, em sequências com similaridade maior do que  $k\%$  entre suas bases (considerando o banco de dados BLOCKS).
- Fórmula geral para matrizes BLOSUM:

$$BLOSUM_k(i, j) = \left\lceil \log_2 \frac{freq_k(a_i, a_j)}{freq_k(a_i) \times freq_k(a_j)} \right\rceil$$

- $BLOSUM_{62}$ : matriz padrão para alinhamento de proteínas (BLAST).

# BLOCKS

A	A	B	C	D	A	-	-	-	B	B	C	D	A
-	A	B	C	D	A	-	A	-	B	B	C	B	B
B	B	B	C	D	A	B	A	-	B	C	C	A	A
A	A	A	C	D	A	C	-	D	C	B	C	D	-
C	C	B	A	D	A	B	-	D	B	B	D	C	C
A	A	A	C	A	A	-	-	-	B	B	C	C	C

# BLOSUM<sub>62</sub>

C	S	T	P	A	G	N	D	E	Q	H	R	K	M	I	L	V	F	Y	W	
C	9																		C	
S	-1	4																	S	
T	-1	1	5																T	
P	-3	-1	-1	7															P	
A	0	1	0	-1	4														A	
G	-3	0	-2	-2	0	6													G	
N	-3	1	0	-2	-2	0	6												N	
D	-3	0	-1	-1	-2	-1	1	6											D	
E	-4	0	-1	-1	-1	-2	0	2	5										E	
Q	-3	0	-1	-1	-1	-2	0	0	2	5									Q	
H	-3	-1	-2	-2	-2	1	-1	0	0	8									H	
R	-3	-1	-1	-2	-1	-2	0	-2	0	1	0	5							R	
K	-3	0	-1	-1	-1	-2	0	-1	1	1	-1	2	5						K	
M	-1	-1	-1	-2	-1	-3	-2	-3	-2	0	-2	-1	-1	5					M	
I	-1	-2	-1	-3	-1	-4	-3	-3	-3	-3	-3	-3	-3	1	4				I	
L	-1	-2	-1	-3	-1	-4	-3	-4	-3	-2	-3	-2	-2	2	2	4			L	
V	-1	-2	0	-2	0	-3	-3	-3	-2	-2	-3	-3	-2	1	3	1	4		V	
F	-2	-2	-2	-4	-2	-3	-3	-3	-3	-3	-1	-3	-3	0	0	0	-1	6	F	
Y	-2	-2	-2	-3	-2	-3	-2	-3	-2	-1	2	-2	-2	-1	-1	-1	-1	3	7	
W	-2	-3	-2	-4	-3	-2	-4	-4	-3	-2	-2	-3	-3	-1	-3	-2	-3	1	2	11
																			W	

# PAM x BLOSUM

- A família PAM assume um modelo evolucionário baseado em árvores filogenéticas. A família BLOSUM não assume nenhum modelo evolucionário, mas considera blocos conservados de proteínas.
- A Entropia Relativa ( $H$ ) de uma matriz de substituição é dada por:

$$H = \sum_{i=1}^{20} \sum_{j=1}^i freq(a_i) freq(a_j) \sigma(a_i, a_j)$$

- Equivalência entre matrizes (com base em sua Entropia Relativa):

$$\begin{array}{ll} PAM_{100} \iff BLOSUM_{90} & | \quad PAM_{200} \iff BLOSUM_{52} \\ PAM_{120} \iff BLOSUM_{80} & | \quad PAM_{250} \iff BLOSUM_{45} \\ PAM_{160} \iff BLOSUM_{60} & | \quad PAM_{400} \iff BLOSUM_{30} \end{array}$$

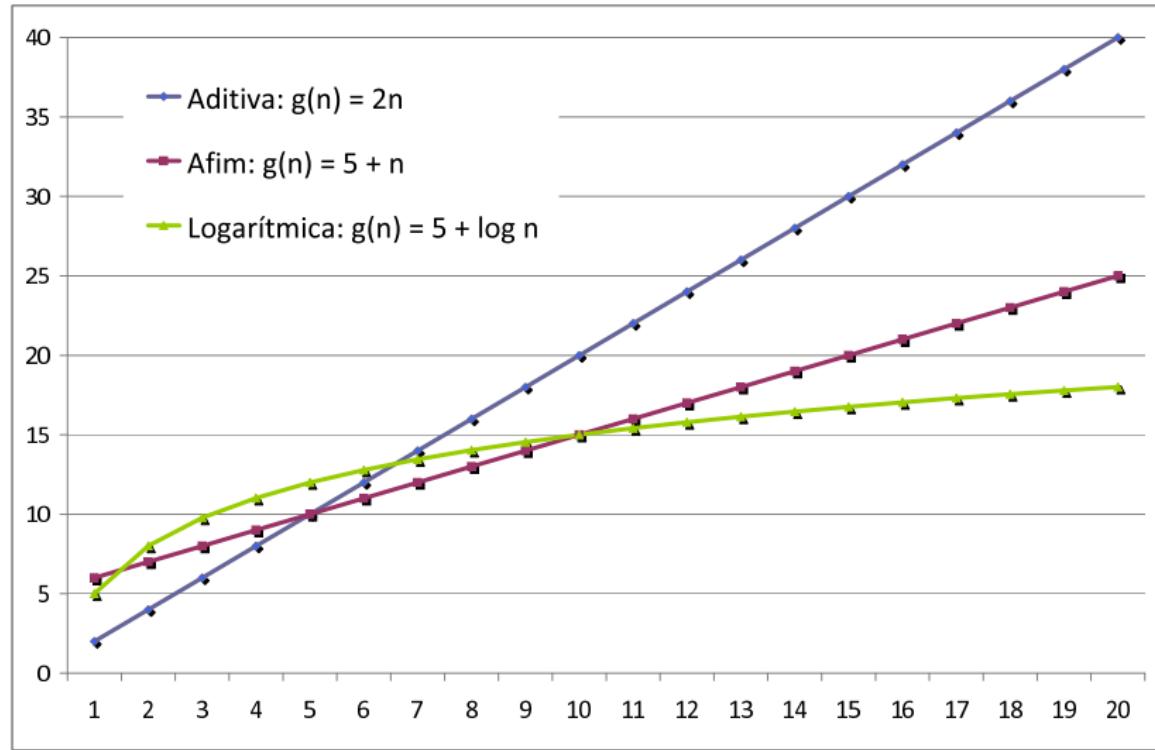
## Erro no Cálculo das Matrizes da Família BLOSUM

- Em 2008, Mark Styczynski e Kyle Jensen mostraram que as matrizes BLOSUM, comumente utilizadas na prática, foram calculadas de forma errada, devido a um erro de software.
- Este erro, que afeta cerca de 15% das posições das matrizes BLOSUM, passou despercebido por mais de 15 anos.
- Testes computacionais, realizados pela mesma dupla de pesquisadores, mostraram que as matrizes “erradas” apresentam melhores resultados na prática do que as matrizes “corretas”.

# Penalidades para Blocos de Buracos

- Ao invés de penalizar buracos individualmente, podemos penalizar blocos de buracos consecutivos num alinhamento.
- Existem várias opções de funções de penalização de blocos de buracos.
- Funções Aditivas:  $g(k_1 + k_2) = g(k_1) + g(k_2)$ 
  - ▶ Linear:  $g(k) = a \times k$
- Funções Sub-Aditivas:  $g(k_1 + k_2) < g(k_1) + g(k_2)$ 
  - ▶ Afim:  $g(k) = a + b \times k$
  - ▶ Funções Convexas:  $g(k) - g(k - 1) > g(k + 1) - g(k)$ 
    - ★ Logarítmica:  $g(k) = a + b \times \log k$

# Exemplos de Funções para Penalidades de Buracos



# Penalidades para Blocos de Buracos

- Diferenças entre alinhamentos com mesma matriz de pontuação de alinhamento de bases, mas com penalidades diferentes para blocos de buracos:
  - ▶ Exemplo de alinhamento global ótimo com pontuação aditiva:

G C G C G T T A G A C T A G C A C C G  
G - G - G T T - G - C - A - C - C - G

- ▶ Exemplo de alinhamento global ótimo com pontuação sub-aditiva:

G C G **C** G T T A G A C T A G C A C C G  
- - **G G** G T T - - - - - G C A C C G

# Alinhamento Global com Função Genérica de Penalidade para Buracos

- Importante: inserir um novo buraco em um alinhamento tem custo diferente caso o alinhamento corrente termine num buraco ou não.
- Para preencher a célula  $M[i, j]$  devemos considerar 3 opções:
  - ▶ Valor do alinhamento ótimo representado pela célula  $M[i - 1, j - 1]$  acrescido do alinhamento dos caracteres  $\alpha[j]$  e  $\beta[i]$ .
  - ▶ Valor do alinhamento ótimo representado por uma célula  $M[i, j - k]$ , acrescido de um alinhamento entre  $\beta[j - k + 1..j]$  e um bloco formado por  $k$  buracos (para  $1 \leq k \leq j$ ).
  - ▶ Valor do alinhamento ótimo representado por uma célula  $M[i - k, j]$ , acrescido de um alinhamento entre  $\alpha[i - k + 1..i]$  e um bloco formado por  $k$  buracos (para  $1 \leq k \leq i$ ).
- Algoritmo proposto por Michael Waterman, Temple Smith e William Beyer em 1976.

# Alinhamento Global com Função Genérica de Penalidade para Buracos

$\times$	$\alpha$	C	C	T	G	T	G	G	C	A	A	C
$\beta$	0	1	2	3	4	5	6	7	8	9	10	11
A	1											
T	2											
T	3											
G	4											
G	5											
C	6						*					
C	7											
A	8											
C	9											

$$M[6,6] + \sigma(G,C)$$

# Alinhamento Global com Função Genérica de Penalidade para Buracos

$\times$	$\alpha$	C	C	T	G	T	G	G	C	A	A	C
$\beta$	0	1	2	3	4	5	6	7	8	9	10	11
A	1											
T	2											
T	3											
G	4											
G	5											
C	6											
C	7						*					
A	8											
C	9											

$$k = 1$$

$$M[7,6] + g(1)$$

# Alinhamento Global com Função Genérica de Penalidade para Buracos

x	$\alpha$	C	C	T	G	T	G	G	C	A	A	C
$\beta$	0	1	2	3	4	5	6	7	8	9	10	11
<b>A</b>	1											
<b>T</b>	2											
<b>T</b>	3											
<b>G</b>	4											
<b>G</b>	5											
<b>C</b>	6											
<b>C</b>	7					*						
<b>A</b>	8											
<b>C</b>	9											

$$k = 2$$

$$M[7,5] + g(2)$$

# Alinhamento Global com Função Genérica de Penalidade para Buracos

x	$\alpha$	C	C	T	G	T	G	G	C	A	A	C
$\beta$	0	1	2	3	4	5	6	7	8	9	10	11
<b>A</b>	1											
<b>T</b>	2											
<b>T</b>	3											
<b>G</b>	4											
<b>G</b>	5											
<b>C</b>	6											
<b>C</b>	7				*							
<b>A</b>	8											
<b>C</b>	9											

$$k = 3$$

$$M[7,4] + g(3)$$

# Alinhamento Global com Função Genérica de Penalidade para Buracos

$\times$	$\alpha$	C	C	T	G	T	G	G	C	A	A	C
$\beta$	0	1	2	3	4	5	6	7	8	9	10	11
A	1											
T	2											
T	3											
G	4											
G	5											
C	6											
C	7		*									
A	8											
C	9											

$$k = 4$$

$$M[7,3] + g(4)$$

# Alinhamento Global com Função Genérica de Penalidade para Buracos

$\times$	$\alpha$	C	C	T	G	T	G	G	C	A	A	C
$\beta$	0	1	2	3	4	5	6	7	8	9	10	11
A	1											
T	2											
T	3											
G	4											
G	5											
C	6											
C	7		*									
A	8											
C	9											

$$k = 5$$

$$M[7,2] + g(5)$$

# Alinhamento Global com Função Genérica de Penalidade para Buracos

$\times$	$\alpha$	C	C	T	G	T	G	G	C	A	A	C
$\beta$	0	1	2	3	4	5	6	7	8	9	10	11
A	1											
T	2											
T	3											
G	4											
G	5											
C	6											
C	7	*										
A	8											
C	9											

$$k = 6$$

$$M[7,1] + g(6)$$

# Alinhamento Global com Função Genérica de Penalidade para Buracos

$\times$	$\alpha$	C	C	T	G	T	G	G	C	A	A	C
$\beta$	0	1	2	3	4	5	6	7	8	9	10	11
A	1											
T	2											
T	3											
G	4											
G	5											
C	6											
C	*											
A	8											
C	9											

$$k = 7$$

$$M[7,0] + g(7)$$

# Alinhamento Global com Função Genérica de Penalidade para Buracos

$\times$	$\alpha$	C	C	T	G	T	G	G	C	A	A	C
$\beta$	0	1	2	3	4	5	6	7	8	9	10	11
A	1											
T	2											
T	3											
G	4											
G	5											
C	6								*			
C	7											
A	8											
C	9											

$$k = 1$$

$$M[6,7] + g(1)$$

# Alinhamento Global com Função Genérica de Penalidade para Buracos

x	$\alpha$	C	C	T	G	T	G	G	C	A	A	C
$\beta$	0	1	2	3	4	5	6	7	8	9	10	11
<b>A</b>	1											
<b>T</b>	2											
<b>T</b>	3											
<b>G</b>	4											
<b>G</b>	5							*				
<b>C</b>	6											
<b>C</b>	7											
<b>A</b>	8											
<b>C</b>	9											

$$k = 2$$

$$M[5,7] + g(2)$$

# Alinhamento Global com Função Genérica de Penalidade para Buracos

x	$\alpha$	C	C	T	G	T	G	G	C	A	A	C
$\beta$	0	1	2	3	4	5	6	7	8	9	10	11
<b>A</b>	1											
<b>T</b>	2											
<b>T</b>	3											
<b>G</b>	4							*				
<b>G</b>	5											
<b>C</b>	6											
<b>C</b>	7											
<b>A</b>	8											
<b>C</b>	9											

$$k = 3$$

$$M[4,7] + g(3)$$

# Alinhamento Global com Função Genérica de Penalidade para Buracos

$\times$	$\alpha$	C	C	T	G	T	G	G	C	A	A	C
$\beta$	0	1	2	3	4	5	6	7	8	9	10	11
A	1											
T	2											
T	3							*				
G	4											
G	5											
C	6											
C	7											
A	8											
C	9											

$$k = 4$$

$$M[3,7] + g(4)$$

# Alinhamento Global com Função Genérica de Penalidade para Buracos

x	$\alpha$	C	C	T	G	T	G	G	C	A	A	C
$\beta$	0	1	2	3	4	5	6	7	8	9	10	11
<b>A</b>	1											
<b>T</b>	2							*				
<b>T</b>	3											
<b>G</b>	4											
<b>G</b>	5											
<b>C</b>	6											
<b>C</b>	7											
<b>A</b>	8											
<b>C</b>	9											

$$k = 5$$

$$M[2,7] + g(5)$$

# Alinhamento Global com Função Genérica de Penalidade para Buracos

x	$\alpha$	C	C	T	G	T	G	G	C	A	A	C
$\beta$	0	1	2	3	4	5	6	7	8	9	10	11
<b>A</b>	1							*				
<b>T</b>	2											
<b>T</b>	3											
<b>G</b>	4											
<b>G</b>	5											
<b>C</b>	6											
<b>C</b>	7											
<b>A</b>	8											
<b>C</b>	9											

$$k = 6$$

$$M[1,7] + g(6)$$

# Alinhamento Global com Função Genérica de Penalidade para Buracos

x	$\alpha$	C	C	T	G	T	G	G	C	A	A	C
$\beta$	0	1	2	3	4	5	6	*	8	9	10	11
<b>A</b>	1											
<b>T</b>	2											
<b>T</b>	3											
<b>G</b>	4											
<b>G</b>	5											
<b>C</b>	6											
<b>C</b>	7											
<b>A</b>	8											
<b>C</b>	9											

$$k = 7$$

$$M[0,7] + g(7)$$

# Alinhamento Global com Função Genérica de Penalidade para Buracos

$\times$	$\alpha$	G	C	G	C	G	T	T	A	G	A	C	T	A	G	C	A	C	C	G
$\beta$	0	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18	-19	-20	-21	-22	-23	-24
<b>G</b>	-6	3	-3	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18	-19	-20
<b>G</b>	-7	-3	1	0	-6	-2	-8	-9	-10	-6	-12	-13	-14	-15	-16	-17	-18	-19	-20	-16
<b>G</b>	-8	-4	-5	4	-2	<b>-3</b>	<b>-9</b>													
<b>T</b>	-9																			
<b>T</b>	-10																			
<b>G</b>	-11																			
<b>C</b>	-12																			
<b>A</b>	-13																			
<b>C</b>	-14																			
<b>C</b>	-15																			
<b>G</b>	-16																			
$\alpha$	=	<b>G</b>	<b>C</b>	<b>G</b>	<b>C</b>	<b>G</b>	<b>T</b>													
$\beta$	=	<b>G</b>	<b>G</b>	<b>G</b>	-	-	-													
-9	=	3	-2	3	-6	-1	-6													

Open Gap = -5    Match = +3

Extend Gap = -1    Mismatch = -2

# Alinhamento Global com Função Genérica de Penalidade para Buracos

$\times$	$\alpha$	G	C	G	C	G	T	T	A	G	A	C	T	A	G	C	A	C	C	G
$\beta$	0	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18	-19	-20	-21	-22	-23	-24
<b>G</b>	-6	3	-3	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18	-19	-20
<b>G</b>	-7	-3	1	0	-6	-2	-8	-9	-10	-6	-12	-13	-14	-15	-16	-17	-18	-19	-20	-16
<b>G</b>	-8	-4	-5	4	<b>-2</b>	-3	<b>-9</b>													
<b>T</b>	-9																			
<b>T</b>	-10																			
<b>G</b>	-11																			
<b>C</b>	-12																			
<b>A</b>	-13																			
<b>C</b>	-14																			
<b>C</b>	-15																			
<b>G</b>	-16																			
$\alpha$	=	<b>G</b>	<b>C</b>	<b>G</b>	<b>C</b>	<b>G</b>	<b>T</b>													
$\beta$	=	<b>G</b>	<b>G</b>	<b>G</b>	-	-	-													
-9	=	3	-2	3	-6	-6	-1													

Open Gap = -5                                  Match = +3

Extend Gap = -1                              Mismatch = -2

# Alinhamento Global com Função Genérica de Penalidade para Buracos

$\alpha$	G	C	G	C	G	T	T	A	G	A	C	T	A	G	C	A	C	C	G	
$\beta$	0	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18	-19	-20	-21	-22	-23	-24
G	-6	3	-3	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18	-19	-20
G	-7	-3	1	0	-6	-2	-8	-9	-10	-6	-12	-13	-14	-15	-11	-17	-18	-19	-20	-16
G	-8	-4	-5	4	-2	-3	-4													
T	-9																			
T	-10																			
G	-11																			
C	-12																			
A	-13																			
C	-14																			
C	-15																			
G	-16																			
$\alpha$	=	G	C	G	C G T			Open Gap = -5											Match = +3	
$\beta$	=	G	G	G	-	-	-	Extend Gap = -1											Mismatch = -2	
-4	=	3	-2	3	-6	-1	-1													

# Alinhamento Global com Função Genérica de Penalidade para Buracos

$\times$	$\alpha$	G	C	G	C	G	T	T	A	G	A	C	T	A	G	C	A	C	C	G
$\beta$	0	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18	-19	-20	-21	-22	-23	-24
G	-6	3	-3	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18	-19	-20
G	-7	-3	1	0	-6	-2	-8	-9	-10	-6	-12	-13	-14	-15	-11	-17	-18	-19	-20	-16
G	-8	-4	-5	4	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17
T	-9	-5	-6	-2	2	-4	0	-1	-7	-8	-9	-10	-6	-12	-13	-14	-15	-16	-17	-18
T	-10	-6	-7	-3	-4	0	-1	3	-3	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14
G	-11	-7	-8	-4	-5	-1	-2	-3	1	0	-6	-7	-8	-9	-5	-11	-12	-13	-14	-10
C	-12	-8	-4	-5	-1	-7	-3	-4	-5	-1	-2	-3	-9	-10	-11	-2	-8	-9	-10	-11
A	-13	-9	-10	-6	-7	-3	-9	-5	-1	-7	2	-4	-5	-6	-7	-8	1	-5	-6	-7
C	-14	-10	-6	-7	-3	-9	-5	-6	-7	-3	-4	5	-1	-2	-3	-4	-5	4	-2	-3
C	-15	-11	-7	-8	-4	-5	-11	-7	-8	-9	-5	-1	3	-3	-4	0	-6	-2	7	1
G	-16	-12	-13	-4	-10	-1	-7	-8	-9	-5	-6	-2	-3	1	0	-6	-2	-3	1	10

Open Gap = -5

Extend Gap = -1

Match = +3

Mismatch = -2

# Alinhamento Global com Função Genérica de Penalidade para Buracos

$\alpha$	G	C	G	C	G	T	T	A	G	A	C	T	A	G	C	A	C	C	G	
$\beta$	0	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18	-19	-20	-21	-22	-23	-24
G	-6	3	-3	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18	-19	-20
G	-7	-3	1	0	-6	-2	-8	-9	-10	-6	-12	-13	-14	-15	-11	-17	-18	-19	-20	-16
G	-8	-4	-5	4	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17
T	-9	-5	-6	-2	2	-4	0	-1	-7	-8	-9	-10	-6	-12	-13	-14	-15	-16	-17	-18
T	-10	-6	-7	-3	-4	0	-1	3	-3	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14
G	-11	-7	-8	-4	-5	-1	-2	-3	1	0	-6	-7	-8	-9	-5	-11	-12	-13	-14	-10
C	-12	-8	-4	-5	-1	-7	-3	-4	-5	-1	-2	-3	-9	-10	-11	-2	-8	-9	-10	-11
A	-13	-9	-10	-6	-7	-3	-9	-5	-1	-7	2	-4	-5	-6	-7	-8	1	-5	-6	-7
C	-14	-10	-6	-7	-3	-9	-5	-6	-7	-3	-4	5	-1	-2	-3	-4	-5	4	-2	-3
C	-15	-11	-7	-8	-4	-5	-11	-7	-8	-9	-5	-1	3	-3	-4	0	-6	-2	7	1
G	-16	-12	-13	-4	-10	-1	-7	-8	-9	-5	-6	-2	-3	1	0	-6	-2	-3	1	10
$\alpha$	=	G	C	G	C	G	T	T	A	G	A	C	T	A	G	C	A	C	G	
$\beta$	=	-	-	G	G	G	T	T	-	-	-	-	-	G	C	A	C	C	G	
10	=	-6	-1	3	-2	3	3	3	-6	-1	-1	-1	-1	-1	3	3	3	3	3	

# Alinhamento Global com Função Genérica de Penalidade para Buracos

---

## Algoritmo 8: Global

---

**Input:**  $\alpha, m, \beta, n$

$M[0, 0] \leftarrow 0$

**for** all  $i \in [1..n]$  **do**  $M[i, 0] \leftarrow g(i)$

**for** all  $j \in [1..m]$  **do**  $M[0, j] \leftarrow g(j)$

**for** all  $i \in [1..n]$  **do**

**for** all  $j \in [1..m]$  **do**

$\max \leftarrow M[i - 1, j - 1] + \sigma(\alpha[j], \beta[i])$

**for** all  $k \in [1..j]$  **do**

**if**  $\max < M[i, j - k] + g(k)$  **then**  $\max \leftarrow M[i, j - k] + g(k)$

**end**

**for** all  $k \in [1..i]$  **do**

**if**  $\max < M[i - k, j] + g(k)$  **then**  $\max \leftarrow M[i - k, j] + g(k)$

**end**

$M[i, j] \leftarrow \max$

**end**

**end**

**return**  $M[n, m]$

---

# Análise de Complexidade

- Complexidade:

$$T(m, n) = \sum_{i=1}^m \sum_{j=1}^n \Theta(i + j + 1)$$

- Limite superior:

$$T(m, n) \leq mnO(m + n + 1) = O(m^2n + mn^2)$$

- Limite inferior:

$$T(m, n) \geq \sum_{i=\lfloor \frac{m}{2} \rfloor + 1}^m \sum_{j=\lfloor \frac{n}{2} \rfloor + 1}^n \Omega(i + j + 1)$$

$$T(m, n) \geq \left\lfloor \frac{m}{2} \right\rfloor \left\lfloor \frac{n}{2} \right\rfloor \Omega\left(\left\lceil \frac{m}{2} \right\rceil + \left\lceil \frac{n}{2} \right\rceil\right) = \Omega(m^2n + mn^2)$$

# Complexidade e Tempo de Execução

- Complexidade:
  - ▶  $T(m, n) = \Theta(mn(m + n)) = \Theta(m^2n + mn^2)$
- Exemplo: computador de 3GHz (1 instrução por ciclo)
  - ▶  $m = n = 10$ :
    - ★ Tempo:  $(10 \times 10 \times 20)/(3 \times 2^{30}) = 31$  bilionésimos de segundo
  - ▶  $m = n = 100$ :
    - ★ Tempo:  $(100 \times 100 \times 200)/(3 \times 2^{30}) = 621$  milionésimos de segundo
  - ▶  $m = n = 1000$ :
    - ★ Tempo:  $(1000 \times 1000) \times 2000/(3 \times 2^{30}) = 621$  milésimos de segundo
  - ▶  $m = n = 10000$ :
    - ★ Tempo:  $(10000 \times 10000) \times 20000/(3 \times 2^{30}) = 10$  minutos
  - ▶  $m = n = 100000$ :
    - ★ Tempo:  $(100000 \times 100000 \times 200000)/(3 \times 2^{30}) = 7$  dias
  - ▶  $m = n = 1000000$ :
    - ★ Tempo:  $(1000000 \times 1000000 \times 2000000)/(3 \times 2^{30}) = 19$  anos

# Alinhamento Global com Função Afim de Penalidade para Buracos

- Para obter um algoritmo de melhor complexidade e ainda assim penalizar um bloco de buracos de forma mais adequada, podemos considerar funções sub-aditivas afins, ao invés de funções genéricas.
- Construção de três matrizes:
  - ▶  $M_{\nwarrow}[i, j]$ : alinhamento ótimo que termina com o alinhamento dos caracteres  $\alpha[j]$  e  $\beta[i]$
  - ▶  $M_{\leftarrow}[i, j]$ : alinhamento ótimo que termina com o alinhamento do caracter  $\alpha[j]$  com um buraco
  - ▶  $M_{\uparrow}[i, j]$ : alinhamento ótimo que termina com o alinhamento de um buraco com o caracter  $\beta[i]$
- Valor do alinhamento global máximo:
  - ▶  $\max\{M_{\nwarrow}[n, m], M_{\leftarrow}[n, m], M_{\uparrow}[n, m]\}$
- Algoritmo proposto por Osamu Gotoh em 1982.

# Alinhamento Global com Função Afim de Penalidade para Buracos

---

## Algoritmo 9: Initialize

---

**Input:** m,n

$M_{\nwarrow}[0, 0] \leftarrow 0$

$M_{\leftarrow}[0, 0] \leftarrow -\infty$

$M_{\uparrow}[0, 0] \leftarrow -\infty$

**for** all  $i \in [1..n]$  **do**

$M_{\nwarrow}[i, 0] \leftarrow -\infty$

$M_{\leftarrow}[i, 0] \leftarrow -\infty$

$M_{\uparrow}[i, 0] \leftarrow -(a + b \times i)$

**end**

**for** all  $j \in [1..m]$  **do**

$M_{\nwarrow}[0, j] \leftarrow -\infty$

$M_{\leftarrow}[0, j] \leftarrow -(a + b \times j)$

$M_{\uparrow}[0, j] \leftarrow -\infty$

**end**

---

# Alinhamento Global com Função Afim de Penalidade para Buracos

---

## Algoritmo 10: GlobalAfim

---

**Input:**  $\alpha, m, \beta, n$

Initialize(m,n)

**for** all  $i \in [1..n]$  **do**

**for** all  $j \in [1..m]$  **do**

$$M_{\nwarrow}[i, j] \leftarrow \sigma(\alpha[j], \beta[i]) + \max \left\{ \begin{array}{l} M_{\nwarrow}[i - 1, j - 1], \\ M_{\leftarrow}[i - 1, j - 1], \\ M_{\uparrow}[i - 1, j - 1] \end{array} \right\}$$

$$M_{\leftarrow}[i, j] \leftarrow \max \left\{ \begin{array}{l} M_{\nwarrow}[i, j - 1] - (a + b), \\ M_{\leftarrow}[i, j - 1] - b, \\ M_{\uparrow}[i, j - 1] - (a + b) \end{array} \right\}$$

$$M_{\uparrow}[i, j] \leftarrow \max \left\{ \begin{array}{l} M_{\nwarrow}[i - 1, j] - (a + b), \\ M_{\leftarrow}[i - 1, j] - (a + b), \\ M_{\uparrow}[i - 1, j] - b \end{array} \right\}$$

end

**end**

**return**  $\max\{M_{\nwarrow}[n, m], M_{\leftarrow}[n, m], M_{\uparrow}[n, m]\}$

---

# Alinhamento Global com Função Afim para Penalidade de Buracos: Matriz $M_K$

$\times$	$\alpha$	G	C	G	C	G	T	T	A	G	A	C	T	A	G	C	A	C	C	G
$\beta$	0	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞
G	-∞	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖
G	-∞	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖
G	-∞	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖
T	-∞	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖
T	-∞	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖
G	-∞	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖
C	-∞	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖
A	-∞	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖
C	-∞	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖
C	-∞	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖
G	-∞	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖	↖

Open Gap = -5

Extend Gap = -1

Match = +3

Mismatch = -2

# Alinhamento Global com Função Afim para Penalidade de Buracos: Matriz $M_{\leftarrow}$

$\times$	$\alpha$	G	C	G	C	G	T	T	A	G	A	C	T	A	G	C	A	C	C	G
$\beta$	-∞	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18	-19	-20	-21	-22	-23	-24
G	-∞	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←
G	-∞	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←
G	-∞	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←
T	-∞	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←
T	-∞	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←
G	-∞	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←
C	-∞	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←
A	-∞	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←
C	-∞	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←
C	-∞	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←
G	-∞	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←

Open Gap = -5

Extend Gap = -1

Match = +3

Mismatch = -2

# Alinhamento Global com Função Afim para Penalidade de Buracos: Matriz $M_{\uparrow}$

$\times$	$\alpha$	G	C	G	C	G	T	T	A	G	A	C	T	A	G	C	A	C	C	G
$\beta$	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞
G	-6	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
G	-7	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
G	-8	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
T	-9	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
T	-10	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
G	-11	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
C	-12	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
A	-13	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
C	-14	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
C	-15	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
G	-16	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑

Open Gap = -5

Extend Gap = -1

Match = +3

Mismatch = -2

# Alinhamento Global com Função Afim para Penalidade de Buracos: Matriz $M_K$

$\alpha$	G	C	G	C	G	T	T	A	G	A	C	T	A	G	C	A	C	C	G	
$\beta$	0	$-\infty$																		
<b>G</b>	$-\infty$	3	-8	-4	-10	-6	-12	-13	-14	-10	-16	-17	-18	-19	-15	-21	-22	-23	-24	-20
<b>G</b>	$-\infty$	-3	1	0	-6	-2	-8	-9	-10	-6	-12	-13	-14	-15	-11	-17	-18	-19	-20	-16
<b>G</b>	$-\infty$	-4	-5	4	-2	-3	-4	-10	-11	-7	-8	-14	-15	-16	-12	-13	-19	-20	-21	-17
<b>T</b>	$-\infty$	-10	-6	-7	2	-4	0	-1	-7	-8	-9	-10	-6	-12	-13	-14	-15	-16	-17	-18
<b>T</b>	$-\infty$	-11	-7	-8	-4	0	-1	3	-3	-9	-10	-11	-7	-8	-14	-15	-16	-17	-18	-19
<b>G</b>	$-\infty$	-7	-8	-4	-5	-1	-2	-3	1	0	-6	-7	-8	-9	-5	-11	-12	-13	-14	-10
<b>C</b>	$-\infty$	-13	-4	-10	-1	-7	-3	-4	-5	-1	-2	-3	-9	-10	-11	-2	-13	-9	-10	-16
<b>A</b>	$-\infty$	-14	-10	-6	-7	-3	-9	-5	-1	-7	2	-4	-5	-6	-12	-13	1	-10	-11	-12
<b>C</b>	$-\infty$	-15	-6	-12	-3	-9	-5	-11	-7	-3	-9	5	-6	-7	-8	-4	-10	4	-2	-8
<b>C</b>	$-\infty$	-16	-7	-8	-4	-5	-11	-7	-8	-9	-5	-1	3	-3	-4	0	-6	-2	7	-4
<b>G</b>	$-\infty$	-12	-13	-4	-10	-1	-7	-13	-9	-5	-11	-7	-3	1	0	-6	-2	-8	-4	10

Open Gap = -5

Extend Gap = -1

Match = +3

Mismatch = -2

# Alinhamento Global com Função Afim para Penalidade de Buracos: Matriz $M_{\leftarrow}$

$\times$	$\alpha$	G	C	G	C	G	T	T	A	G	A	C	T	A	G	C	A	C	C	G
$\beta$	-∞	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18	-19	-20	-21	-22	-23	-24
G	-∞	-12	-3	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18	-19	-20
G	-∞	-13	-9	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18	-19	-20	-21
G	-∞	-14	-10	-11	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17
T	-∞	-15	-11	-12	-8	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18
T	-∞	-16	-12	-13	-9	-10	-6	-7	-3	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14
G	-∞	-17	-13	-14	-10	-11	-7	-8	-9	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15
C	-∞	-18	-14	-10	-11	-7	-8	-9	-10	-11	-7	-8	-9	-10	-11	-12	-8	-9	-10	-11
A	-∞	-19	-15	-16	-12	-13	-9	-10	-11	-7	-8	-4	-5	-6	-7	-8	-9	-5	-6	-7
C	-∞	-20	-16	-12	-13	-9	-10	-11	-12	-13	-9	-10	-1	-2	-3	-4	-5	-6	-2	-3
C	-∞	-21	-17	-13	-14	-10	-11	-12	-13	-14	-15	-11	-7	-3	-4	-5	-6	-7	-8	1
G	-∞	-22	-18	-19	-10	-11	-7	-8	-9	-10	-11	-12	-8	-9	-5	-6	-7	-8	-9	-5

Open Gap = -5

Extend Gap = -1

Match = +3

Mismatch = -2

# Alinhamento Global com Função Afim para Penalidade de Buracos: Matriz $M_{\uparrow}$

$\times$	$\alpha$	G	C	G	C	G	T	T	A	G	A	C	T	A	G	C	A	C	C	G
$\beta$	- $\infty$																			
<b>G</b>	-6	-12	-13	-14	-15	-16	-17	-18	-19	-20	-21	-22	-23	-24	-25	-26	-27	-28	-29	-30
<b>G</b>	-7	-3	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18	-19	-20	-21	-22	-23	-24	-25	-26
<b>G</b>	-8	-4	-5	-6	-12	-8	-14	-15	-16	-12	-18	-19	-20	-21	-17	-23	-24	-25	-26	-22
<b>T</b>	-9	-5	-6	-2	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18	-19	-20	-21	-22	-23
<b>T</b>	-10	-6	-7	-3	-4	-10	-6	-7	-13	-14	-15	-16	-12	-18	-19	-20	-21	-22	-23	-24
<b>G</b>	-11	-7	-8	-4	-5	-6	-7	-3	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18	-19	-20
<b>C</b>	-12	-8	-9	-5	-6	-7	-8	-4	-5	-6	-12	-13	-14	-15	-11	-17	-18	-19	-20	-16
<b>A</b>	-13	-9	-10	-6	-7	-8	-9	-5	-6	-7	-8	-9	-15	-16	-12	-8	-14	-15	-16	-17
<b>C</b>	-14	-10	-11	-7	-8	-9	-10	-6	-7	-8	-4	-10	-11	-12	-13	-9	-5	-11	-12	-13
<b>C</b>	-15	-11	-12	-8	-9	-10	-11	-7	-8	-9	-5	-1	-7	-8	-9	-10	-6	-2	-8	-9
<b>G</b>	-16	-12	-13	-9	-10	-11	-12	-8	-9	-10	-6	-2	-3	-9	-10	-6	-7	-3	1	-5

Open Gap = -5

Extend Gap = -1

Match = +3

Mismatch = -2

# Alinhamento Global com Função Afim para Penalidade de Buracos: Alinhamento Upmost (Matriz $M_{\nwarrow}$ )

$\times$	$\alpha$	<b>G</b>	<b>C</b>	<b>G</b>	<b>C</b>	<b>G</b>	<b>T</b>	<b>T</b>	<b>A</b>	<b>G</b>	<b>A</b>	<b>C</b>	<b>T</b>	<b>A</b>	<b>G</b>	<b>C</b>	<b>A</b>	<b>C</b>	<b>C</b>	<b>G</b>
$\beta$	0	$-\infty$																		
<b>G</b>	$-\infty$	3	-8	-4	-10	-6	-12	-13	-14	-10	-16	-17	-18	-19	-15	-21	-22	-23	-24	-20
<b>G</b>	$-\infty$	-3	1	0	-6	-2	-8	-9	-10	-6	-12	-13	-14	-15	-11	-17	-18	-19	-20	-16
<b>G</b>	$-\infty$	-4	-5	4	-2	-3	-4	-10	-11	-7	-8	-14	-15	-16	-12	-13	-19	-20	-21	-17
<b>T</b>	$-\infty$	-10	-6	-7	2	-4	0	-1	-7	-8	-9	-10	<b>-6</b>	-12	-13	-14	-15	-16	-17	-18
<b>T</b>	$-\infty$	-11	-7	-8	-4	0	-1	3	-3	-9	-10	-11	-7	<b>-8</b>	-14	-15	-16	-17	-18	-19
<b>G</b>	$-\infty$	-7	-8	-4	-5	-1	-2	-3	1	0	-6	-7	-8	-9	<b>-5</b>	-11	-12	-13	-14	-10
<b>C</b>	$-\infty$	-13	-4	-10	-1	-7	-3	-4	-5	-1	-2	-3	-9	-10	-11	<b>-2</b>	-13	-9	-10	-16
<b>A</b>	$-\infty$	-14	-10	-6	-7	-3	-9	-5	-1	-7	2	-4	-5	-6	-12	-13	<b>1</b>	-10	-11	-12
<b>C</b>	$-\infty$	-15	-6	-12	-3	-9	-5	-11	-7	-3	-9	5	-6	-7	-8	-4	-10	<b>4</b>	-2	-8
<b>C</b>	$-\infty$	-16	-7	-8	-4	-5	-11	-7	-8	-9	-5	-1	3	-3	-4	0	-6	-2	<b>7</b>	-4
<b>G</b>	$-\infty$	-12	-13	-4	-10	-1	-7	-13	-9	-5	-11	-7	-3	1	0	-6	-2	-8	-4	<b>10</b>
$\alpha$	=												<b>T</b>	<b>A</b>	<b>G</b>	<b>C</b>	<b>A</b>	<b>C</b>	<b>C</b>	<b>G</b>
$\beta$	=												<b>T</b>	<b>T</b>	<b>G</b>	<b>C</b>	<b>A</b>	<b>C</b>	<b>C</b>	<b>G</b>
	=												3	-2	3	3	3	3	3	3

# Alinhamento Global com Função Afim para Penalidade de Buracos: Alinhamento Upmost (Matriz $M_{\leftarrow}$ )

$\times$	$\alpha$	G	C	G	C	G	T	T	A	G	A	C	T	A	G	C	A	C	C	G
$\beta$	$-\infty$	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18	-19	-20	-21	-22	-23	-24
G	$-\infty$	-12	-3	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18	-19	-20
G	$-\infty$	-13	-9	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18	-19	-20	-21
G	$-\infty$	-14	-10	-11	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17
T	$-\infty$	-15	-11	-12	-8	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18
T	$-\infty$	-16	-12	-13	-9	-10	-6	-7	-3	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14
G	$-\infty$	-17	-13	-14	-10	-11	-7	-8	-9	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15
C	$-\infty$	-18	-14	-10	-11	-7	-8	-9	-10	-11	-7	-8	-9	-10	-11	-12	-8	-9	-10	-11
A	$-\infty$	-19	-15	-16	-12	-13	-9	-10	-11	-7	-8	-4	-5	-6	-7	-8	-9	-5	-6	-7
C	$-\infty$	-20	-16	-12	-13	-9	-10	-11	-12	-13	-9	-10	-1	-2	-3	-4	-5	-6	-2	-3
C	$-\infty$	-21	-17	-13	-14	-10	-11	-12	-13	-14	-15	-11	-7	-3	-4	-5	-6	-7	-8	1
G	$-\infty$	-22	-18	-19	-10	-11	-7	-8	-9	-10	-11	-12	-8	-9	-5	-6	-7	-8	-9	-5
$\alpha$	=		C	G	T	T	A	G	A	C	T	A	G	C	A	C	C	G		
$\beta$	=		-	-	-	-	-	-	-	-	T	T	G	C	A	C	C	G		
	=		-6	-1	-1	-1	-1	-1	-1	-1	3	-2	3	3	3	3	3	3	3	

# Alinhamento Global com Função Afim para Penalidade de Buracos: Alinhamento Upmost (Matriz $M_{\nwarrow}$ )

$\times$	$\alpha$	G	C	G	C	G	T	T	A	G	A	C	T	A	G	C	A	C	C	G
$\beta$	0	$-\infty$																		
G	$-\infty$	3	-8	-4	-10	-6	-12	-13	-14	-10	-16	-17	-18	-19	-15	-21	-22	-23	-24	-20
G	$-\infty$	-3	1	0	-6	-2	-8	-9	-10	-6	-12	-13	-14	-15	-11	-17	-18	-19	-20	-16
G	$-\infty$	-4	-5	4	-2	-3	-4	-10	-11	-7	-8	-14	-15	-16	-12	-13	-19	-20	-21	-17
T	$-\infty$	-10	-6	-7	2	-4	0	-1	-7	-8	-9	-10	-6	-12	-13	-14	-15	-16	-17	-18
T	$-\infty$	-11	-7	-8	-4	0	-1	3	-3	-9	-10	-11	-7	8	-14	-15	-16	-17	-18	-19
G	$-\infty$	-7	-8	-4	-5	-1	-2	-3	1	0	-6	-7	-8	-9	5	-11	-12	-13	-14	-10
C	$-\infty$	-13	-4	-10	-1	-7	-3	-4	-5	-1	-2	-3	-9	-10	-11	2	-13	-9	-10	-16
A	$-\infty$	-14	-10	-6	-7	-3	-9	-5	-1	-7	2	-4	-5	-6	-12	-13	1	-10	-11	-12
C	$-\infty$	-15	-6	-12	-3	-9	-5	-11	-7	-3	-9	5	-6	-7	-8	-4	-10	4	-2	-8
C	$-\infty$	-16	-7	-8	-4	-5	-11	-7	-8	-9	-5	-1	3	-3	-4	0	-6	-2	7	-4
G	$-\infty$	-12	-13	-4	-10	-1	-7	-13	-9	-5	-11	-7	-3	1	0	-6	-2	-8	-4	10
$\alpha$	=	G	C	G	C	G	T	T	A	G	A	C	T	A	G	C	A	C	C	G
$\beta$	=	G	G	G	-	-	-	-	-	-	-	-	T	T	G	C	A	C	C	G
10	=	3	-2	3	-6	-1	-1	-1	-1	-1	-1	-1	3	-2	3	3	3	3	3	3

# Alinhamento Global com Função Afim para Penalidade de Buracos: Alinhamento Downmost (Matriz $M_{\nwarrow}$ )

$\times$	$\alpha$	<b>G</b>	<b>C</b>	<b>G</b>	<b>C</b>	<b>G</b>	<b>T</b>	<b>T</b>	<b>A</b>	<b>G</b>	<b>A</b>	<b>C</b>	<b>T</b>	<b>A</b>	<b>G</b>	<b>C</b>	<b>A</b>	<b>C</b>	<b>C</b>	<b>G</b>
$\beta$	0	$-\infty$																		
<b>G</b>	$-\infty$	3	-8	-4	-10	-6	-12	-13	-14	-10	-16	-17	-18	-19	-15	-21	-22	-23	-24	-20
<b>G</b>	$-\infty$	-3	1	0	-6	-2	-8	-9	-10	-6	-12	-13	-14	-15	-11	-17	-18	-19	-20	-16
<b>G</b>	$-\infty$	-4	-5	4	-2	-3	-4	-10	-11	-7	-8	-14	-15	-16	-12	-13	-19	-20	-21	-17
<b>T</b>	$-\infty$	-10	-6	-7	2	-4	0	-1	-7	-8	-9	-10	-6	-12	-13	-14	-15	-16	-17	-18
<b>T</b>	$-\infty$	-11	-7	-8	-4	0	-1	3	-3	-9	-10	-11	-7	-8	-14	-15	-16	-17	-18	-19
<b>G</b>	$-\infty$	-7	-8	-4	-5	-1	-2	-3	1	0	-6	-7	-8	-9	-5	-11	-12	-13	-14	-10
<b>C</b>	$-\infty$	-13	-4	-10	-1	-7	-3	-4	-5	-1	-2	-3	-9	-10	-11	-2	-13	-9	-10	-16
<b>A</b>	$-\infty$	-14	-10	-6	-7	-3	-9	-5	-1	-7	2	-4	-5	-6	-12	-13	1	-10	-11	-12
<b>C</b>	$-\infty$	-15	-6	-12	-3	-9	-5	-11	-7	-3	-9	5	-6	-7	-8	-4	-10	4	-2	-8
<b>C</b>	$-\infty$	-16	-7	-8	-4	-5	-11	-7	-8	-9	-5	-1	3	-3	-4	0	-6	-2	7	-4
<b>G</b>	$-\infty$	-12	-13	-4	-10	-1	-7	-13	-9	-5	-11	-7	-3	1	0	-6	-2	-8	-4	10
$\alpha$	=														<b>G</b>	<b>C</b>	<b>A</b>	<b>C</b>	<b>C</b>	<b>G</b>
$\beta$	=														<b>G</b>	<b>C</b>	<b>A</b>	<b>C</b>	<b>C</b>	<b>G</b>
	=														3	3	3	3	3	3

# Alinhamento Global com Função Afim para Penalidade de Buracos: Alinhamento Downmost (Matriz $M_{\leftarrow}$ )

$\times$	$\alpha$	G	C	G	C	G	T	T	A	G	A	C	T	A	G	C	A	C	C	G
$\beta$	$-\infty$	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18	-19	-20	-21	-22	-23	-24
G	$-\infty$	-12	-3	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18	-19	-20
G	$-\infty$	-13	-9	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18	-19	-20	-21
G	$-\infty$	-14	-10	-11	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17
T	$-\infty$	-15	-11	-12	-8	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18
T	$-\infty$	-16	-12	-13	-9	-10	-6	-7	-3	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14
G	$-\infty$	-17	-13	-14	-10	-11	-7	-8	-9	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15
C	$-\infty$	-18	-14	-10	-11	-7	-8	-9	-10	-11	-7	-8	-9	-10	-11	-12	-8	-9	-10	-11
A	$-\infty$	-19	-15	-16	-12	-13	-9	-10	-11	-7	-8	-4	-5	-6	-7	-8	-9	-5	-6	-7
C	$-\infty$	-20	-16	-12	-13	-9	-10	-11	-12	-13	-9	-10	-1	-2	-3	-4	-5	-6	-2	-3
C	$-\infty$	-21	-17	-13	-14	-10	-11	-12	-13	-14	-15	-11	-7	-3	-4	-5	-6	-7	-8	1
G	$-\infty$	-22	-18	-19	-10	-11	-7	-8	-9	-10	-11	-12	-8	-9	-5	-6	-7	-8	-9	-5
$\alpha$	=								A	G	A	C	T	A	G	C	A	C	C	G
$\beta$	=								-	-	-	-	-	-	G	C	A	C	C	G
	=								-6	-1	-1	-1	-1	-1	3	3	3	3	3	3

# Alinhamento Global com Função Afim para Penalidade de Buracos: Alinhamento Downmost (Matriz $M_{\nwarrow}$ )

$\times$	$\alpha$	G	C	G	C	G	T	T	A	G	A	C	T	A	G	C	A	C	C	G	
$\beta$	0	$-\infty$																			
G	$-\infty$	3	-8	-4	-10	-6	-12	-13	-14	-10	-16	-17	-18	-19	-15	-21	-22	-23	-24	-20	
G	$-\infty$	-3	1	0	-6	-2	-8	-9	-10	-6	-12	-13	-14	-15	-11	-17	-18	-19	-20	-16	
G	$-\infty$	-4	-5	4	-2	-3	-4	-10	-11	-7	-8	-14	-15	-16	-12	-13	-19	-20	-21	-17	
T	$-\infty$	-10	-6	-7	2	-4	0	-1	-7	-8	-9	-10	-11	-6	-12	-13	-14	-15	-16	-17	-18
T	$-\infty$	-11	-7	-8	-4	0	-1	3	-3	-9	-10	-11	-7	-8	-14	-15	-16	-17	-18	-19	
G	$-\infty$	-7	-8	-4	-5	-1	-2	-3	1	0	-6	-7	-8	-9	-5	-11	-12	-13	-14	-10	
C	$-\infty$	-13	-4	-10	-1	-7	-3	-4	-5	-1	-2	-3	-9	-10	-11	-2	-13	-9	-10	-16	
A	$-\infty$	-14	-10	-6	-7	-3	-9	-5	-1	-7	2	-4	-5	-6	-12	-13	1	-10	-11	-12	
C	$-\infty$	-15	-6	-12	-3	-9	-5	-11	-7	-3	-9	5	-6	-7	-8	-4	-10	4	-2	-8	
C	$-\infty$	-16	-7	-8	-4	-5	-11	-7	-8	-9	-5	-1	3	-3	-4	0	-6	-2	7	-4	
G	$-\infty$	-12	-13	-4	-10	-1	-7	-13	-9	-5	-11	-7	-3	1	0	-6	-2	-8	-4	10	
$\alpha$	=					T	T	A	G	A	C	T	A	G	C	A	C	C	G		
$\beta$	=					T	T	-	-	-	-	-	-	-	G	C	A	C	G		
	=					3	3	-6	-1	-1	-1	-1	-1	3	3	3	3	3	3		

# Alinhamento Global com Função Afim para Penalidade de Buracos: Alinhamento Downmost (Matriz $M_{\leftarrow}$ )

$\times$	$\alpha$	G	C	G	C	G	T	T	A	G	A	C	T	A	G	C	A	C	C	G
$\beta$	$-\infty$	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18	-19	-20	-21	-22	-23	-24
<b>G</b>	$-\infty$	-12	-3	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18	-19	-20
<b>G</b>	$-\infty$	-13	-9	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18	-19	-20	-21
<b>G</b>	$-\infty$	-14	-10	-11	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17
<b>T</b>	$-\infty$	-15	-11	-12	-8	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18
<b>T</b>	$-\infty$	-16	-12	-13	-9	-10	-6	-7	-3	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14
<b>G</b>	$-\infty$	-17	-13	-14	-10	-11	-7	-8	-9	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15
<b>C</b>	$-\infty$	-18	-14	-10	-11	-7	-8	-9	-10	-11	-7	-8	-9	-10	-11	-12	-8	-9	-10	-11
<b>A</b>	$-\infty$	-19	-15	-16	-12	-13	-9	-10	-11	-7	-8	-4	-5	-6	-7	-8	-9	-5	-6	-7
<b>C</b>	$-\infty$	-20	-16	-12	-13	-9	-10	-11	-12	-13	-9	-10	-1	-2	-3	-4	-5	-6	-2	-3
<b>C</b>	$-\infty$	-21	-17	-13	-14	-10	-11	-12	-13	-14	-15	-11	-7	-3	-4	-5	-6	-7	-8	1
<b>G</b>	$-\infty$	-22	-18	-19	-10	-11	-7	-8	-9	-10	-11	-12	-8	-9	-5	-6	-7	-8	-9	-5
$\alpha$	=		C	G	T	T	A	G	A	C	T	A	G	C	A	C	C	G		
$\beta$	=		-	-	T	T	-	-	-	-	-	-	G	C	A	C	C	G		
	=		-6	-1	3	3	-6	-1	-1	-1	-1	-1	3	3	3	3	3	3	3	

# Alinhamento Global com Função Afim para Penalidade de Buracos: Alinhamento Downmost (Matriz uma $M_K$ )

$\alpha$	G	C	G	C	G	T	T	A	G	A	C	T	A	G	C	A	C	C	G		
$\beta$	0	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞			
G	-∞	3	-8	-4	-10	-6	-12	-13	-14	-10	-16	-17	-18	-19	-15	-21	-22	-23	-24	-20	
G	-∞	-3	1	0	-6	-2	-8	-9	-10	-6	-12	-13	-14	-15	-11	-17	-18	-19	-20	-16	
G	-∞	-4	-5	4	-2	-3	-4	-10	-11	-7	-8	-14	-15	-16	-12	-13	-19	-20	-21	-17	
T	-∞	-10	-6	-7	2	-4	0	-1	-7	-8	-9	-10	-11	-6	-12	-13	-14	-15	-16	-17	-18
T	-∞	-11	-7	-8	-4	0	-1	3	-3	-9	-10	-11	-7	-8	-14	-15	-16	-17	-18	-19	
G	-∞	-7	-8	-4	-5	-1	-2	-3	1	0	-6	-7	-8	-9	-5	-11	-12	-13	-14	-10	
C	-∞	-13	-4	-10	-1	-7	-3	-4	-5	-1	-2	-3	-9	-10	-11	-2	-13	-9	-10	-16	
A	-∞	-14	-10	-6	-7	-3	-9	-5	-1	-7	2	-4	-5	-6	-12	-13	1	-10	-11	-12	
C	-∞	-15	-6	-12	-3	-9	-5	-11	-7	-3	-9	5	-6	-7	-8	-4	-10	4	-2	-8	
C	-∞	-16	-7	-8	-4	-5	-11	-7	-8	-9	-5	-1	3	-3	-4	0	-6	-2	7	-4	
G	-∞	-12	-13	-4	-10	-1	-7	-13	-9	-5	-11	-7	-3	1	0	-6	-2	-8	-4	10	
$\alpha$	=	G	C	G	C	G	T	T	A	G	A	C	T	A	G	C	A	C	C	G	
$\beta$	=	G	G	G	-	-	T	T	-	-	-	-	-	-	G	C	A	C	C	G	
10	=	3	-2	3	-6	-1	3	3	-6	-1	-1	-1	-1	-1	3	3	3	3	3	3	

# Exercícios

## Exercício

*Calcule a complexidade assintótica (usando a notação  $\Theta$ ) do algoritmo de alinhamento global com função afim para penalidade de blocos de buracos proposto por Osamu Gotoh.*

## Exercício

*Escreva um algoritmo, em pseudocódigo, que dadas duas sequências  $\alpha$  e  $\beta$ , com respectivamente  $m$  e  $n$  caracteres, e três matrizes de pontuação de alinhamentos entre  $\alpha$  e  $\beta$  usando função afim para penalidade de blocos de buracos ( $M_{\leftarrow}$ ,  $M_{\nwarrow}$  e  $M_{\uparrow}$ ), retorne um alinhamento ótimo entre  $\alpha$  e  $\beta$ .*

# Alinhamento Global usando Espaço Linear

- Todos os algoritmos de alinhamento que vimos até agora usam espaço quadrático em relação ao tamanho das sequências ( $\Theta(mn)$ ).
- Este requisito de memória pode tornar estes algoritmos proibitivos.
- Por exemplo, ao tentar alinhar duas sequências com um milhão de caracteres seria necessário aproximadamente 4TB de memória, considerando que um inteiro ocupe 4 bytes (32 bits) de memória.
- Note que é possível obter o valor do alinhamento ótimo usando apenas uma linha da matriz de programação dinâmica.
- Como obter o valor ótimo e o alinhamento usando apenas espaço linear em relação ao tamanho das sequências de entrada?

# Alinhamento Global usando Espaço Linear

- Usar o paradigma de Divisão e Conquista em cima da matriz de Programação Dinâmica:
  - ▶ Calcular o vetor  $V_1$  com os valores dos alinhamentos ótimos entre todos os prefixos da sequência  $\alpha$  com a sequência  $\beta[1.. \lfloor \frac{n}{2} \rfloor]$ .
  - ▶ Calcular o vetor  $V_2$  com os valores dos alinhamentos ótimos entre todos os prefixos da sequência  $\bar{\alpha}$  com a sequência  $\beta[\lfloor \frac{n}{2} \rfloor + 1..n]$ .
    - ★ *Observação:*  $\bar{x}$  significa a sequência reversa de  $x$ .
  - ▶ Encontrar  $k$  tal que  $V_1[k] + V_2[m - k]$  seja o máximo possível.
  - ▶ O valor  $k$  indica que o alinhamento ótimo entre as sequências  $\alpha$  e  $\beta$  passa pela posição  $M[\lfloor \frac{n}{2} \rfloor, k]$ .
  - ▶ Resolver recursivamente o alinhamento entre  $\alpha[1..k]$  e  $\beta[1.. \lfloor \frac{n}{2} \rfloor]$ , assim como o alinhamento entre  $\overline{\alpha[k + 1..m]}$  e  $\overline{\beta[\lfloor \frac{n}{2} \rfloor + 1..n]}$ .
- Algoritmo proposto por Daniel Hirschberg em 1975.

# Alinhamento Global usando Espaço Linear

x	$\alpha$	A	C	T	G	G	T	C	A	A	C	C	G	T	C	T	G	C	G	
$\beta$	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55	-60	-65	-70	-75	-80	-85	-90	-95
<b>A</b>	-5	3	-2	-7	-12	-17	-22	-27	-32	-37	-42	-47	-52	-57	-62	-67	-72	-77	-82	-87
<b>T</b>	-10	-2	1	1	-4	-9	-14	-19	-24	-29	-34	-39	-44	-49	-54	-59	-64	-69	-74	-79
<b>T</b>	-15	-7	-4	4	-1	-6	-11	-11	-16	-21	-26	-31	-36	-41	-46	-51	-56	-61	-66	-71
<b>G</b>	-20	-12	-9	-1	7	2	-3	-8	-13	-18	-23	-28	-33	-33	-38	-43	-48	-53	-58	-63
<b>G</b>	-25	-17	-14	-6	2	10	5	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-45	-50	-55
<b>C</b>	-30	-22	-14	-11	-3	5	8	3	3	-2	-7	-12	-17	-22	-27	-32	-37	-42	-42	-47
<b>C</b>																				
<b>A</b>																				
<b>C</b>																				
<b>A</b>																				
<b>G</b>																				
<b>G</b>																				
<b>C</b>																				
<b>A</b>																				
<b>T</b>																				

Gap = -5      Match = +3      Mismatch = -2

# Alinhamento Global usando Espaço Linear

x	$\alpha$	A	C	T	G	G	T	C	A	A	C	C	G	T	C	T	G	C	G	
$\beta$	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55	-60	-65	-70	-75	-80	-85	-90	-95
<b>A</b>	-5	3	-2	-7	-12	-17	-22	-27	-32	-37	-42	-47	-52	-57	-62	-67	-72	-77	-82	-87
<b>T</b>	-10	-2	1	1	-4	-9	-14	-19	-24	-29	-34	-39	-44	-49	-54	-59	-64	-69	-74	-79
<b>T</b>	-15	-7	-4	4	-1	-6	-11	-11	-16	-21	-26	-31	-36	-41	-46	-51	-56	-61	-66	-71
<b>G</b>	-20	-12	-9	-1	7	2	-3	-8	-13	-18	-23	-28	-33	-33	-38	-43	-48	-53	-58	-63
<b>G</b>	-25	-17	-14	-6	2	10	5	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-45	-50	-55
<b>C</b>	-30	-22	-14	-11	-3	5	8	3	3	-2	-7	-12	-17	-22	-27	-32	-37	-42	-42	-47
<b>C</b>	-35	-27																		
<b>A</b>																				
<b>C</b>																				
<b>A</b>																				
<b>G</b>																				
<b>G</b>																				
<b>C</b>																				
<b>A</b>																				
<b>T</b>																				

Gap = -5      Match = +3      Mismatch = -2

# Alinhamento Global usando Espaço Linear

x	$\alpha$	A	C	T	G	G	T	C	A	A	C	C	G	T	C	T	G	C	G	
$\beta$	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55	-60	-65	-70	-75	-80	-85	-90	-95
<b>A</b>	-5	3	-2	-7	-12	-17	-22	-27	-32	-37	-42	-47	-52	-57	-62	-67	-72	-77	-82	-87
<b>T</b>	-10	-2	1	1	-4	-9	-14	-19	-24	-29	-34	-39	-44	-49	-54	-59	-64	-69	-74	-79
<b>T</b>	-15	-7	-4	4	-1	-6	-11	-11	-16	-21	-26	-31	-36	-41	-46	-51	-56	-61	-66	-71
<b>G</b>	-20	-12	-9	-1	7	2	-3	-8	-13	-18	-23	-28	-33	-33	-38	-43	-48	-53	-58	-63
<b>G</b>	-25	-17	-14	-6	2	10	5	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-45	-50	-55
<b>C</b>	-30	-22	-14	-11	-3	5	8	3	3	-2	-7	-12	-17	-22	-27	-32	-37	-42	-42	-47
<b>C</b>	-35	-27	-19																	
<b>A</b>																				
<b>C</b>																				
<b>A</b>																				
<b>G</b>																				
<b>G</b>																				
<b>C</b>																				
<b>A</b>																				
<b>T</b>																				

Gap = -5      Match = +3      Mismatch = -2

# Alinhamento Global usando Espaço Linear

x	$\alpha$	A	C	T	G	G	T	C	A	A	C	C	G	T	C	T	G	C	G	
$\beta$	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55	-60	-65	-70	-75	-80	-85	-90	-95
<b>A</b>	-5	3	-2	-7	-12	-17	-22	-27	-32	-37	-42	-47	-52	-57	-62	-67	-72	-77	-82	-87
<b>T</b>	-10	-2	1	1	-4	-9	-14	-19	-24	-29	-34	-39	-44	-49	-54	-59	-64	-69	-74	-79
<b>T</b>	-15	-7	-4	4	-1	-6	-11	-11	-16	-21	-26	-31	-36	-41	-46	-51	-56	-61	-66	-71
<b>G</b>	-20	-12	-9	-1	7	2	-3	-8	-13	-18	-23	-28	-33	-33	-38	-43	-48	-53	-58	-63
<b>G</b>	-25	-17	-14	-6	2	10	5	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-45	-50	-55
<b>C</b>	-30	-22	-14	-11	-3	5	8	3	3	-2	-7	-12	-17	-22	-27	-32	-37	-42	-42	-47
<b>C</b>	-35	-27	-19	-16	-8	0	3	6	6	1										
<b>A</b>																				
<b>C</b>																				
<b>A</b>																				
<b>G</b>																				
<b>G</b>																				
<b>C</b>																				
<b>A</b>																				
<b>T</b>																				

Gap = -5      Match = +3      Mismatch = -2

# Alinhamento Global usando Espaço Linear

x	$\alpha$	A	C	T	G	G	T	C	A	A	C	C	G	T	C	T	G	C	G	
$\beta$	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55	-60	-65	-70	-75	-80	-85	-90	-95
<b>A</b>	-5	3	-2	-7	-12	-17	-22	-27	-32	-37	-42	-47	-52	-57	-62	-67	-72	-77	-82	-87
<b>T</b>	-10	-2	1	1	-4	-9	-14	-19	-24	-29	-34	-39	-44	-49	-54	-59	-64	-69	-74	-79
<b>T</b>	-15	-7	-4	4	-1	-6	-11	-11	-16	-21	-26	-31	-36	-41	-46	-51	-56	-61	-66	-71
<b>G</b>	-20	-12	-9	-1	7	2	-3	-8	-13	-18	-23	-28	-33	-33	-38	-43	-48	-53	-58	-63
<b>G</b>	-25	-17	-14	-6	2	10	5	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-45	-50	-55
<b>C</b>	-30	-22	-14	-11	-3	5	8	3	3	-2	-7	-12	-17	-22	-27	-32	-37	-42	-42	-47
<b>C</b>	-35	-27	-19	-16	-8	0	3	6	6	1	-4	-4	-9	-14	-19	-24	-29	-34	-39	-44
<b>A</b>																				
<b>C</b>																				
<b>A</b>																				
<b>G</b>																				
<b>G</b>																				
<b>C</b>																				
<b>A</b>																				
<b>T</b>																				

Gap = -5      Match = +3      Mismatch = -2

# Alinhamento Global usando Espaço Linear

x	$\alpha$	A	C	T	G	G	T	C	A	A	C	C	G	T	C	T	G	C	G	
$\beta$	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55	-60	-65	-70	-75	-80	-85	-90	-95
<b>A</b>	-5	3	-2	-7	-12	-17	-22	-27	-32	-37	-42	-47	-52	-57	-62	-67	-72	-77	-82	-87
<b>T</b>	-10	-2	1	1	-4	-9	-14	-19	-24	-29	-34	-39	-44	-49	-54	-59	-64	-69	-74	-79
<b>T</b>	-15	-7	-4	4	-1	-6	-11	-11	-16	-21	-26	-31	-36	-41	-46	-51	-56	-61	-66	-71
<b>G</b>	-20	-12	-9	-1	7	2	-3	-8	-13	-18	-23	-28	-33	-33	-38	-43	-48	-53	-58	-63
<b>G</b>	-25	-17	-14	-6	2	10	5	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-45	-50	-55
<b>C</b>	-30	-22	-14	-11	-3	5	8	3	3	-2	-7	-12	-17	-22	-27	-32	-37	-42	-42	-47
<b>C</b>	-35	-27	-19	-16	-8	0	3	6	6	1	-4	-4	-9	-14	-19	-24	-29	-34	-39	-44
<b>A</b>																				
<b>C</b>																				
<b>A</b>																				
<b>G</b>																				
<b>G</b>																				
<b>C</b>																				
<b>A</b>																				
<b>T</b>																				

Gap = -5      Match = +3      Mismatch = -2

# Alinhamento Global usando Espaço Linear

$\times$	$\alpha$	A	C	T	G	G	T	C	A	A	C	C	G	T	C	T	G	C	G	
$\beta$	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55	-60	-65	-70	-75	-80	-85	-90	-95
<b>A</b>	-5	3	-2	-7	-12	-17	-22	-27	-32	-37	-42	-47	-52	-57	-62	-67	-72	-77	-82	-87
<b>T</b>	-10	-2	1	1	-4	-9	-14	-19	-24	-29	-34	-39	-44	-49	-54	-59	-64	-69	-74	-79
<b>T</b>	-15	-7	-4	4	-1	-6	-11	-11	-16	-21	-26	-31	-36	-41	-46	-51	-56	-61	-66	-71
<b>G</b>	-20	-12	-9	-1	7	2	-3	-8	-13	-18	-23	-28	-33	-33	-38	-43	-48	-53	-58	-63
<b>G</b>	-25	-17	-14	-6	2	10	5	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-45	-50	-55
<b>C</b>	-30	-22	-14	-11	-3	5	8	3	3	-2	-7	-12	-17	-22	-27	-32	-37	-42	-42	-47
<b>C</b>	-35	-27	-19	-16	-8	0	3	6	6	1	-4	-4	-9	-14	-19	-24	-29	-34	-39	-44
<b>A</b>	-40	-32	-24	-21	-13	-5	-2	1	4	9	4	-1	-6	-11	-16	-21	-26	-31	-36	-41
<b>C</b>	-45	-37	-29	-26	-18	-10	-7	-4	4	4	7	7	2	-3	-8	-13	-18	-23	-28	-33
<b>A</b>	-50	-42	-34	-31	-23	-15	-12	-9	-1	7	7	5	5	0	-5	-10	-15	-20	-25	-30
<b>G</b>	-55	-47	-39	-36	-28	-20	-12	-14	-6	2	5	5	3	8	3	-2	-7	-12	-17	-22
<b>G</b>	-60	-52	-44	-41	-33	-25	-17	-14	-11	-3	0	3	3	6	6	1	-4	-4	-9	-14
<b>C</b>	-65	-57	-49	-46	-38	-30	-22	-19	-11	-8	-5	3	6	1	4	9	4	-1	-1	-6
<b>A</b>	-70	-62	-54	-51	-43	-35	-27	-24	-16	-8	-5	-2	1	4	-1	4	7	2	-3	-3
<b>T</b>	<b>-75</b>	<b>-67</b>	<b>-59</b>	<b>-51</b>	<b>-48</b>	<b>-40</b>	<b>-32</b>	<b>-24</b>	<b>-21</b>	<b>-13</b>	<b>-10</b>	<b>-7</b>	<b>-4</b>	<b>-1</b>	<b>7</b>	<b>2</b>	<b>7</b>	<b>5</b>	<b>0</b>	<b>-5</b>

Gap = -5

Match = +3

Mismatch = -2

# Alinhamento Global usando Espaço Linear

x	$\alpha$	A	C	T	G	G	T	C	A	A	C	C	G	T	C	T	G	C	G	
$\beta$	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55	-60	-65	-70	-75	-80	-85	-90	-95
<b>A</b>	-5	3	-2	-7	-12	-17	-22	-27	-32	-37	-42	-47	-52	-57	-62	-67	-72	-77	-82	-87
<b>T</b>	-10	-2	1	-4	-9	-14	-19	-24	-29	-34	-39	-44	-49	-54	-59	-64	-69	-74	-79	
<b>T</b>	-15	-7	-4	4	-1	-6	-11	-11	-16	-21	-26	-31	-36	-41	-46	-51	-56	-61	-66	-71
<b>G</b>	-20	-12	-9	-1	7	2	-3	-8	-13	-18	-23	-28	-33	-33	-38	-43	-48	-53	-58	-63
<b>G</b>	-25	-17	-14	-6	2	10	5	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-45	-50	-55
<b>C</b>	-30	-22	-14	-11	-3	5	8	3	3	-2	-7	-12	-17	-22	-27	-32	-37	-42	-42	-47
<b>C</b>	-35	-27	-19	-16	-8	0	3	6	6	1	-4	-4	-9	-14	-19	-24	-29	-34	-39	-44
<b>A</b>	-40	-32	-24	-21	-13	-5	-2	1	4	9	4	-1	-6	-11	-16	-21	-26	-31	-36	-41
<b>C</b>	-45	-37	-29	-26	-18	-10	-7	-4	4	4	7	7	2	-3	-8	-13	-18	-23	-28	-33
<b>A</b>	-50	-42	-34	-31	-23	-15	-12	-9	-1	7	7	5	5	0	-5	-10	-15	-20	-25	-30
<b>G</b>	-55	-47	-39	-36	-28	-20	-12	-14	-6	2	5	5	3	8	3	-2	-7	-12	-17	-22
<b>G</b>	-60	-52	-44	-41	-33	-25	-17	-14	-11	-3	0	3	3	6	6	1	-4	-4	-9	-14
<b>C</b>	-65	-57	-49	-46	-38	-30	-22	-19	-11	-8	-5	3	6	1	4	9	4	-1	-1	-6
<b>A</b>	-70	-62	-54	-51	-43	-35	-27	-24	-16	-8	-5	-2	1	4	-1	4	7	2	-3	-3
<b>T</b>	-75	-67	-59	-51	-48	-40	-32	-24	-21	-13	-10	-7	-4	-1	7	2	7	5	0	-5

Gap = -5

Match = +3

Mismatch = -2

# Alinhamento Global usando Espaço Linear

$\times$	$\alpha$	A	C	T	G	G	T	C	A	A	C	C	G	T	C	T	G	C	G	
$\beta$	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55	-60	-65	-70	-75	-80	-85	-90	-95
<b>A</b>	-5	3	-2	-7	-12	-17	-22	-27	-32	-37	-42	-47	-52	-57	-62	-67	-72	-77	-82	-87
<b>T</b>	-10	-2	1	-4	-9	-14	-19	-24	-29	-34	-39	-44	-49	-54	-59	-64	-69	-74	-79	
<b>T</b>	-15	-7	-4	4	-1	-6	-11	-11	-16	-21	-26	-31	-36	-41	-46	-51	-56	-61	-66	-71
<b>G</b>	-20	-12	-9	-1	7	2	-3	-8	-13	-18	-23	-28	-33	-33	-38	-43	-48	-53	-58	-63
<b>G</b>	-25	-17	-14	-6	2	10	5	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-45	-50	-55
<b>C</b>	-30	-22	-14	-11	-3	5	8	3	3	-2	-7	-12	-17	-22	-27	-32	-37	-42	-42	-47
<b>C</b>	-35	-27	-19	-16	-8	0	3	6	6	1	-4	-4	-9	-14	-19	-24	-29	-34	-39	-44
<b>A</b>	-40	-32	-24	-21	-13	-5	-2	1	4	9	4	-1	-6	-11	-16	-21	-26	-31	-36	-41
<b>C</b>	-45	-37	-29	-26	-18	-10	-7	-4	4	4	7	7	2	-3	-8	-13	-18	-23	-28	-33
<b>A</b>	-50	-42	-34	-31	-23	-15	-12	-9	-1	7	7	5	5	0	-5	-10	-15	-20	-25	-30
<b>G</b>	-55	-47	-39	-36	-28	-20	-12	-14	-6	2	5	5	3	8	3	-2	-7	-12	-17	-22
<b>G</b>	-60	-52	-44	-41	-33	-25	-17	-14	-11	-3	0	3	3	6	6	1	-4	-4	-9	-14
<b>C</b>	-65	-57	-49	-46	-38	-30	-22	-19	-11	-8	-5	3	6	1	4	9	4	-1	-1	-6
<b>A</b>	-70	-62	-54	-51	-43	-35	-27	-24	-16	-8	-5	-2	1	4	-1	4	7	2	-3	-3
<b>T</b>	-75	-67	-59	-51	-48	-40	-32	-24	-21	-13	-10	-7	-4	-1	7	2	7	5	0	-5

-5	=	3	-2	3	3	3	-5	-5	3	-2	3	3	-2	3	-2	3	-5	-2	-5	-2
$\alpha$	=	A	C	T	G	G	G	T	C	A	A	C	C	G	T	C	T	G	C	G
$\beta$	=	A	T	T	G	G	-	-	C	C	A	C	A	G	G	C	-	A	-	T

# Alinhamento Global usando Espaço Linear

$\times$	$\alpha$	A	C	T	G	G	T	C	A	A	C	C	G	T	C	T	G	C	G	
$\beta$	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55	-60	-65	-70	-75	-80	-85	-90	-95
<b>A</b>	-5	3	-2	-7	-12	-17	-22	-27	-32	-37	-42	-47	-52	-57	-62	-67	-72	-77	-82	-87
<b>T</b>	-10	-2	1	1	-4	-9	-14	-19	-24	-29	-34	-39	-44	-49	-54	-59	-64	-69	-74	-79
<b>T</b>	-15	-7	-4	4	-1	-6	-11	-11	-16	-21	-26	-31	-36	-41	-46	-51	-56	-61	-66	-71
<b>G</b>	-20	-12	-9	-1	7	2	-3	-8	-13	-18	-23	-28	-33	-33	-38	-43	-48	-53	-58	-63
<b>G</b>	-25	-17	-14	-6	2	10	5	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-45	-50	-55
<b>C</b>	-30	-22	-14	-11	-3	5	8	3	3	-2	-7	-12	-17	-22	-27	-32	-37	-42	-42	-47
<b>C</b>	-35	-27	-19	-16	-8	0	3	6	6	1	-4	-4	-9	-14	-19	-24	-29	-34	-39	-44
	-36	-41	-41	-36	-31	-26	-21	-16	-11	-6	-14	-14	-9	-7	-10	-18	-21	-24	-32	<b>A</b>
	-44	-39	-39	-39	-34	-29	-24	-19	-19	-14	-9	-12	-12	-7	-5	-13	-16	-19	-27	<b>C</b>
	-47	-47	-42	-37	-37	-37	-32	-27	-22	-17	-17	-12	-15	-10	-5	-8	-11	-19	-22	<b>A</b>
	-55	-50	-45	-40	-35	-35	-40	-35	-30	-25	-20	-15	-10	-13	-8	-3	-6	-14	-17	<b>G</b>
	-63	-58	-53	-48	-43	-38	-38	-38	-33	-28	-23	-18	-13	-13	-11	-6	-1	-9	-12	<b>G</b>
	-71	-66	-61	-56	-51	-46	-41	-36	-36	-31	-26	-21	-21	-16	-11	-11	-6	-4	-12	<b>C</b>
	-79	-74	-69	-64	-59	-54	-49	-44	-39	-34	-34	-29	-24	-19	-14	-14	-9	-4	-7	<b>A</b>
	-87	-82	-77	-72	-67	-62	-57	-52	-47	-42	-37	-32	-27	-22	-17	-12	-12	-7	-2	<b>T</b>
	-95	-90	-85	-80	-75	-70	-65	-60	-55	-50	-45	-40	-35	-30	-25	-20	-15	-10	-5	0
	<b>A</b>	<b>C</b>	<b>T</b>	<b>G</b>	<b>G</b>	<b>G</b>	<b>T</b>	<b>C</b>	<b>A</b>	<b>A</b>	<b>C</b>	<b>C</b>	<b>G</b>	<b>T</b>	<b>C</b>	<b>T</b>	<b>G</b>	<b>C</b>	<b>G</b>	$\alpha$
																				$\times$

Gap = -5

Match = +3

Mismatch = -2

# Alinhamento Global usando Espaço Linear

$\times$	$\alpha$	A	C	T	G	G	T	C	A	A	C	C	G	T	C	T	G	C	G	
$\beta$	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55	-60	-65	-70	-75	-80	-85	-90	-95
<b>A</b>	-5	3	-2	-7	-12	-17	-22	-27	-32	-37	-42	-47	-52	-57	-62	-67	-72	-77	-82	-87
<b>T</b>	-10	-2	1	1	-4	-9	-14	-19	-24	-29	-34	-39	-44	-49	-54	-59	-64	-69	-74	-79
<b>T</b>	-15	-7	-4	4	-1	-6	-11	-11	-16	-21	-26	-31	-36	-41	-46	-51	-56	-61	-66	-71
<b>G</b>	-20	-12	-9	-1	7	2	-3	-8	-13	-18	-23	-28	-33	-33	-38	-43	-48	-53	-58	-63
<b>G</b>	-25	-17	-14	-6	2	10	5	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-45	-50	-55
<b>C</b>	-30	-22	-14	-11	-3	5	8	3	3	-2	-7	-12	-17	-22	-27	-32	-37	-42	-42	-47
<b>C</b>	-35	-27	-19	-16	-8	0	3	6	6	1	-4	-4	-9	-14	-19	-24	-29	-34	-39	-44
	-36	-41	-41	-36	-31	-26	-21	-16	-11	-6	-14	-14	-9	-7	-10	-18	-21	-24	-32	<b>A</b>
	-44	-39	-39	-39	-34	-29	-24	-19	-19	-14	-9	-12	-12	-7	-5	-13	-16	-19	-27	<b>C</b>
	-47	-47	-42	-37	-37	-37	-32	-27	-22	-17	-17	-12	-15	-10	-5	-8	-11	-19	-22	<b>A</b>
	-55	-50	-45	-40	-35	-35	-40	-35	-30	-25	-20	-15	-10	-13	-8	-3	-6	-14	-17	<b>G</b>
	-63	-58	-53	-48	-43	-38	-38	-38	-33	-28	-23	-18	-13	-13	-11	-6	-1	-9	-12	<b>G</b>
	-71	-66	-61	-56	-51	-46	-41	-36	-36	-31	-26	-21	-21	-16	-11	-11	-6	-4	-12	<b>C</b>
	-79	-74	-69	-64	-59	-54	-49	-44	-39	-34	-34	-29	-24	-19	-14	-14	-9	-4	-7	<b>A</b>
	-87	-82	-77	-72	-67	-62	-57	-52	-47	-42	-37	-32	-27	-22	-17	-12	-12	-7	-2	<b>T</b>
	-95	-90	-85	-80	-75	-70	-65	-60	-55	-50	-45	-40	-35	-30	-25	-20	-15	-10	-5	0
	<b>A</b>	<b>C</b>	<b>T</b>	<b>G</b>	<b>G</b>	<b>G</b>	<b>T</b>	<b>C</b>	<b>A</b>	<b>A</b>	<b>C</b>	<b>C</b>	<b>G</b>	<b>T</b>	<b>C</b>	<b>T</b>	<b>G</b>	<b>C</b>	<b>G</b>	$\alpha$
																				$\times$

Gap = -5

Match = +3

Mismatch = -2

# Alinhamento Global usando Espaço Linear

$\times$	$\alpha$	A	C	T	G	G	T	C	A	A	C	C	G	T	C	T	G	C	G	
$\beta$	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55	-60	-65	-70	-75	-80	-85	-90	-95
A	-5	3	-2	-7	-12	-17	-22	-27	-32	-37	-42	-47	-52	-57	-62	-67	-72	-77	-82	-87
T	-10	-2	1	1	-4	-9	-14	-19	-24	-29	-34	-39	-44	-49	-54	-59	-64	-69	-74	-79
T	-15	-7	-4	4	-1	-6	-11	-11	-16	-21	-26	-31	-36	-41	-46	-51	-56	-61	-66	-71
G	-20	-12	-9	-1	7	2	-3	-8	-13	-18	-23	-28	-33	-33	-38	-43	-48	-53	-58	-63
G	-25	-17	-14	-6	2	10	5	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-45	-50	-55
C	-30	-22	-14	-11	-3	5	8	3	3	-2	-7	-12	-17	-22	-27	-32	-37	-42	-42	-47
C	-35	-27	-19	-16	-8	0	3	6	6	1	-4	-4	-9	-14	-19	-24	-29	-34	-39	-44
	-36	-41	-41	-36	-31	-26	-21	-16	-11	-6	-14	-14	-9	-7	-10	-18	-21	-24	-32	A
	-44	-39	-39	-39	-34	-29	-24	-19	-19	-14	-9	-12	-12	-7	-5	-13	-16	-19	-27	C
	-47	-47	-42	-37	-37	-37	-32	-27	-22	-17	-17	-12	-15	-10	-5	-8	-11	-19	-22	A
	-55	-50	-45	-40	-35	-35	-40	-35	-30	-25	-20	-15	-10	-13	-8	-3	-6	-14	-17	G
	-63	-58	-53	-48	-43	-38	-38	-38	-33	-28	-23	-18	-13	-13	-11	-6	-1	-9	-12	G
	-71	-66	-61	-56	-51	-46	-41	-36	-36	-31	-26	-21	-21	-16	-11	-11	-6	-4	-12	C
	-79	-74	-69	-64	-59	-54	-49	-44	-39	-34	-34	-29	-24	-19	-14	-14	-9	-4	-7	A
	-87	-82	-77	-72	-67	-62	-57	-52	-47	-42	-37	-32	-27	-22	-17	-12	-12	-7	-2	T
	-95	-90	-85	-80	-75	-70	-65	-60	-55	-50	-45	-40	-35	-30	-25	-20	-15	-10	-5	0
	A	C	T	G	G	G	T	C	A	A	C	C	G	T	C	T	G	C	G	
																			$\alpha \times$	

Gap = -5

Match = +3

Mismatch = -2

# Alinhamento Global usando Espaço Linear

$\times$	$\alpha$	<b>A</b>	<b>C</b>	<b>T</b>	<b>G</b>	<b>G</b>	<b>T</b>	<b>C</b>	<b>A</b>	<b>A</b>	<b>C</b>	<b>C</b>	<b>G</b>	<b>T</b>	<b>C</b>	<b>T</b>	<b>G</b>	<b>C</b>	<b>G</b>	
$\beta$	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55	-60	-65	-70	-75	-80	-85	-90	-95
<b>A</b>	-5	3	-2	-7	-12	-17	-22	-27	-32	-37	-42	-47	-52	-57	-62	-67	-72	-77	-82	-87
<b>T</b>	-10	-2	1	1	-4	-9	-14	-19	-24	-29	-34	-39	-44	-49	-54	-59	-64	-69	-74	-79
<b>T</b>	-15	-7	-4	4	-1	-6	-11	-11	-16	-21	-26	-31	-36	-41	-46	-51	-56	-61	-66	-71
<b>G</b>	-20	-12	-9	-1	7	2	-3	-8	-13	-18	-23	-28	-33	-33	-38	-43	-48	-53	-58	-63
<b>G</b>	-25	-17	-14	-6	2	10	5	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-45	-50	-55
<b>C</b>	-30	-22	-14	-11	-3	5	8	3	3	-2	-7	-12	-17	-22	-27	-32	-37	-42	-42	-47
<b>C</b>	-35	-27	-19	-16	-8	0	3	6	6	1	-4	-4	-9	-14	-19	-24	-29	-34	-39	-44
	-36	-41	-41	-36	-31	-26	-21	-16	-11	-6	-14	-14	-9	-7	-10	-18	-21	-24	-32	<b>A</b>
	-44	-39	-39	-39	-34	-29	-24	-19	-19	-14	-9	-12	-12	-7	-5	-13	-16	-19	-27	<b>C</b>
	-47	-47	-42	-37	-37	-37	-32	-27	-22	-17	-17	-12	-15	-10	-5	-8	-11	-19	-22	<b>A</b>
	-55	-50	-45	-40	-35	-35	-40	-35	-30	-25	-20	-15	-10	-13	-8	-3	-6	-14	-17	<b>G</b>
	-63	-58	-53	-48	-43	-38	-38	-38	-33	-28	-23	-18	-13	-13	-11	-6	-1	-9	-12	<b>G</b>
	-71	-66	-61	-56	-51	-46	-41	-36	-36	-31	-26	-21	-21	-16	-11	-11	-6	-4	-12	<b>C</b>
	-79	-74	-69	-64	-59	-54	-49	-44	-39	-34	-34	-29	-24	-19	-14	-14	-9	-4	-7	<b>A</b>
	-87	-82	-77	-72	-67	-62	-57	-52	-47	-42	-37	-32	-27	-22	-17	-12	-12	-7	-2	<b>T</b>
	-95	-90	-85	-80	-75	-70	-65	-60	-55	-50	-45	-40	-35	-30	-25	-20	-15	-10	-5	0
	<b>A</b>	<b>C</b>	<b>T</b>	<b>G</b>	<b>G</b>	<b>G</b>	<b>T</b>	<b>C</b>	<b>A</b>	<b>A</b>	<b>C</b>	<b>C</b>	<b>G</b>	<b>T</b>	<b>C</b>	<b>T</b>	<b>G</b>	<b>C</b>	<b>G</b>	$\alpha$
										<b>A</b>	<b>A</b>									$\times$

# Alinhamento Global usando Espaço Linear

$\times$	$\alpha$	<b>A</b>	<b>C</b>	<b>T</b>	<b>G</b>	<b>G</b>	<b>T</b>	<b>C</b>	<b>A</b>	<b>A</b>	<b>C</b>	<b>C</b>	<b>G</b>	<b>T</b>	<b>C</b>	<b>T</b>	<b>G</b>	<b>C</b>	<b>G</b>	
$\beta$	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55	-60	-65	-70	-75	-80	-85	-90	-95
<b>A</b>	-5	3	-2	-7	-12	-17	-22	-27	-32	-37	-42	-47	-52	-57	-62	-67	-72	-77	-82	-87
<b>T</b>	-10	-2	1	1	-4	-9	-14	-19	-24	-29	-34	-39	-44	-49	-54	-59	-64	-69	-74	-79
<b>T</b>	-15	-7	-4	4	-1	-6	-11	-11	-16	-21	-26	-31	-36	-41	-46	-51	-56	-61	-66	-71
<b>G</b>	-20	-12	-9	-1	7	2	-3	-8	-13	-18	-23	-28	-33	-33	-38	-43	-48	-53	-58	-63
<b>G</b>	-25	-17	-14	-6	2	10	5	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-45	-50	-55
<b>C</b>	-30	-22	-14	-11	-3	5	8	3	3	-2	-7	-12	-17	-22	-27	-32	-37	-42	-42	-47
<b>C</b>	-35	-27	-19	-16	-8	0	3	6	6	1	-4	-4	-9	-14	-19	-24	-29	-34	-39	-44
	-36	-41	-41	-36	-31	-26	-21	-16	-11	-6	-14	-14	-9	-7	-10	-18	-21	-24	-32	-40
	-44	-39	-39	-39	-34	-29	-24	-19	-19	-14	-9	-12	-12	-7	-5	-13	-16	-19	-27	-35
	-47	-47	-42	-37	-37	-37	-32	-27	-22	-17	-17	-12	-15	-10	-5	-8	-11	-19	-22	-30
	-55	-50	-45	-40	-35	-35	-40	-35	-30	-25	-20	-15	-10	-13	-8	-3	-6	-14	-17	-25
	-63	-58	-53	-48	-43	-38	-38	-38	-33	-28	-23	-18	-13	-13	-11	-6	-1	-9	-12	-20
	-71	-66	-61	-56	-51	-46	-41	-36	-36	-31	-26	-21	-21	-16	-11	-11	-6	-4	-12	-15
	-79	-74	-69	-64	-59	-54	-49	-44	-39	-34	-34	-29	-24	-19	-14	-14	-9	-4	-7	-10
	-87	-82	-77	-72	-67	-62	-57	-52	-47	-42	-37	-32	-27	-22	-17	-12	-12	-7	-2	-5
	-95	-90	-85	-80	-75	-70	-65	-60	-55	-50	-45	-40	-35	-30	-25	-20	-15	-10	-5	0
	<b>A</b>	<b>C</b>	<b>T</b>	<b>G</b>	<b>G</b>	<b>G</b>	<b>T</b>	<b>C</b>	<b>A</b>	<b>A</b>	<b>C</b>	<b>C</b>	<b>G</b>	<b>T</b>	<b>C</b>	<b>T</b>	<b>G</b>	<b>C</b>	<b>G</b>	$\alpha$
																				$\times$

A      A  
C      A

# Alinhamento Global usando Espaço Linear

x	$\alpha$	A	C	T	G	G	G	T	C
$\beta$	0	-5	-10	-15	-20	-25	-30	-35	-40
A	-5	3	-2	-7	-12	-17	-22	-27	-32
T	-10	-2	1	1	-4	-9	-14	-19	-24
T	-15	-7	-4	4	-1	-6	-11	-11	-16
		-16	-11	-6	-1	4	4	-4	-7
		-24	-19	-14	-9	-4	1	1	-2
		-32	-27	-22	-17	-12	-7	-2	3
		-40	-35	-30	-25	-20	-15	-10	-5
		A	C	T	G	G	G	T	C

A A  
C A

# Alinhamento Global usando Espaço Linear

x	$\alpha$	A	C	T	G	G	G	T	C
$\beta$	0	-5	-10	-15	-20	-25	-30	-35	-40
A	-5	3	-2	-7	-12	-17	-22	-27	-32
T	-10	-2	1	1	-4	-9	-14	-19	-24
T	-15	-7	-4	4	-1	-6	-11	-11	-16
		-16	-11	-6	-1	4	4	-4	-7
		-24	-19	-14	-9	-4	1	1	-2
		-32	-27	-22	-17	-12	-7	-2	3
		-40	-35	-30	-25	-20	-15	-10	-5
		A	C	T	G	G	G	T	C

A A  
C A

# Alinhamento Global usando Espaço Linear

x	$\alpha$	A	C	T	G	G	G	T	C
$\beta$	0	-5	-10	-15	-20	-25	-30	-35	-40
A	-5	3	-2	-7	-12	-17	-22	-27	-32
T	-10	-2	1	1	-4	-9	-14	-19	-24
T	-15	-7	-4	4	-1	-6	-11	-11	-16

-16	-11	-6	-1	4	4	-4	-7	-15	G
-24	-19	-14	-9	-4	1	1	-2	-10	G
-32	-27	-22	-17	-12	-7	-2	3	-5	C
-40	-35	-30	-25	-20	-15	-10	-5	0	
A	C	T	G	G	G	T	C		

A A  
C A

# Alinhamento Global usando Espaço Linear

x	$\alpha$	A	C	T	G	G	G	T	C
$\beta$	0	-5	-10	-15	-20	-25	-30	-35	-40
A	-5	3	-2	-7	-12	-17	-22	-27	-32
T	-10	-2	1	1	-4	-9	-14	-19	-24
T	-15	-7	-4	4	-1	-6	-11	-11	-16

-16	-11	-6	-1	4	4	-4	-7	-15	G
-24	-19	-14	-9	-4	1	1	-2	-10	G
-32	-27	-22	-17	-12	-7	-2	3	-5	C
-40	-35	-30	-25	-20	-15	-10	-5	0	
A	C	T	G	G	G	T	C		

T G  
T G

A A  
C A

# Alinhamento Global usando Espaço Linear

x	$\alpha$	A	C	T	G	G	G	T	C
$\beta$	0	-5	-10	-15	-20	-25	-30	-35	-40
A	-5	3	-2	-7	-12	-17	-22	-27	-32
T	-10	-2	1	1	-4	-9	-14	-19	-24
T	-15	-7	-4	4	-1	-6	-11	-11	-16

-16	-11	-6	-1	4	4	-4	-7	-15	G
-24	-19	-14	-9	-4	1	1	-2	-10	G
-32	-27	-22	-17	-12	-7	-2	3	-5	C
-40	-35	-30	-25	-20	-15	-10	-5	0	
A	C	T	G	G	G	T	C		

T G  
T G

A A  
C A

# Alinhamento Global usando Espaço Linear

x	$\alpha$	A	C	
$\beta$	0	-5	-10	
A	-5	3	-2	
	-7	-2	-5	T
	-10	-5	0	
	A	C		

T G  
T G

A A  
C A

# Alinhamento Global usando Espaço Linear

x	$\alpha$	A	C	
$\beta$	0	-5	-10	
A	-5	3	-2	
	-7	-2	-5	T
	-10	-5	0	
	A	C		

T G  
T G

A A  
C A

# Alinhamento Global usando Espaço Linear

x	$\alpha$	A	C	
$\beta$	0	-5	-10	
A	-5	3	-2	
	-7	-2	-5	T
	-10	-5	0	
A	C			

T G  
T G

A A  
C A

# Alinhamento Global usando Espaço Linear

x	$\alpha$	<b>A</b>	<b>C</b>	
$\beta$	0	-5	-10	
<b>A</b>	-5	<b>3</b>	-2	
	-7	<b>-2</b>	-5	<b>T</b>
	-10	-5	0	
	<b>A</b>	<b>C</b>		

A C T G  
A T T G

A A  
C A

# Alinhamento Global usando Espaço Linear

	G	G	T	C
0	-5	-10	-15	-20
G	-5	3	-2	-7
	-12	-7	-2	3
	-20	-15	-10	-5
G	G	T	C	

A C T G  
A T T G

A A  
C A

# Alinhamento Global usando Espaço Linear

	G	G	T	C	
0	-5	-10	-15	-20	
G	-5	3	-2	-7	-12
-12	-7	-2	3	-5	C
-20	-15	-10	-5	0	
G	G	T	C		

A C T G  
A T T G

A A  
C A

# Alinhamento Global usando Espaço Linear

		G	G	T	C	
		0	-5	-10	-15	-20
G		-5	3	-2	-7	-12
		-12	-7	-2	3	-5
		-20	-15	-10	-5	0
G		G	T	C		

A C T G  
A T T G

A A  
C A

# Alinhamento Global usando Espaço Linear

		G	G	T	C	
		0	-5	-10	-15	-20
G		-5	3	-2	-7	-12
		-12	-7	-2	3	-5
		-20	-15	-10	-5	0
G		G	T	C		

A C T G G G  
A T T G G -

A A  
C A

# Alinhamento Global usando Espaço Linear

		T	C
		0	-5
C	-5	-2	-2

A C T G G G  
A T T G G -

A A  
C A

# Alinhamento Global usando Espaço Linear

		T	C
		0	-5
C	0	-5	-10
	-5	-2	-2

A C T G G G T C A A  
A T T G G - - C C A

# Alinhamento Global usando Espaço Linear

	C	C	G	T	C	T	G	C	G		
0	-5	-10	-15	-20	-25	-30	-35	-40	-45		
C	-5	3	-2	-7	-12	-17	-22	-27	-32	-37	
A	-10	-2	1	-4	-9	-14	-19	-24	-29	-34	
G	-15	-7	-4	4	-1	-6	-11	-16	-21	-26	
	-23	-18	-13	-13	-11	-6	-1	-9	-12	-20	<b>G</b>
	-26	-21	-21	-16	-11	-11	-6	-4	-12	-15	<b>C</b>
	-34	-29	-24	-19	-14	-14	-9	-4	-7	-10	<b>A</b>
	-37	-32	-27	-22	-17	-12	-12	-7	-2	-5	<b>T</b>
	-45	-40	-35	-30	-25	-20	-15	-10	-5	0	
	C	C	G	T	C	T	G	C	G		
A	C	T	G	G	T	C	A	A			
A	T	T	G	G	-	-	C	C	A		

# Alinhamento Global usando Espaço Linear

	C	C	G	T	C	T	G	C	G
0	-5	-10	-15	-20	-25	-30	-35	-40	-45
C	-5	3	-2	-7	-12	-17	-22	-27	-32
A	-10	-2	1	-4	-9	-14	-19	-24	-29
G	-15	-7	-4	4	-1	-6	-11	-16	-21
	-23	-18	-13	-13	-11	-6	-1	-9	-12
	-26	-21	-21	-16	-11	-11	-6	-4	-12
	-34	-29	-24	-19	-14	-14	-9	-4	-7
	-37	-32	-27	-22	-17	-12	-12	-7	-2
	-45	-40	-35	-30	-25	-20	-15	-10	-5
	C	C	G	T	C	T	G	C	G
A	C	T	G	G	G	C	C	A	A
A	T	T	G	G	-	-	C	C	A

# Alinhamento Global usando Espaço Linear

	C	C	G	T	C	T	G	C	G		
0	-5	-10	-15	-20	-25	-30	-35	-40	-45		
C	-5	3	-2	-7	-12	-17	-22	-27	-32	-37	
A	-10	-2	1	-4	-9	-14	-19	-24	-29	-34	
G	-15	-7	-4	4	-1	-6	-11	-16	-21	-26	
	-23	-18	-13	-13	-11	-6	-1	-9	-12	-20	<b>G</b>
	-26	-21	-21	-16	-11	-11	-6	-4	-12	-15	<b>C</b>
	-34	-29	-24	-19	-14	-14	-9	-4	-7	-10	<b>A</b>
	-37	-32	-27	-22	-17	-12	-12	-7	-2	-5	<b>T</b>
	-45	-40	-35	-30	-25	-20	-15	-10	-5	0	
	C	C	G	T	C	T	G	C	G		
A	C	T	G	G	T	C	A	A			
A	T	T	G	G	-	-	C	C	A		

# Alinhamento Global usando Espaço Linear

	C	C	G	T	C	T	G	C	G		
0	-5	-10	-15	-20	-25	-30	-35	-40	-45		
C	-5	3	-2	-7	-12	-17	-22	-27	-32	-37	
A	-10	-2	1	-4	-9	-14	-19	-24	-29	-34	
G	-15	-7	-4	4	-1	-6	-11	-16	-21	-26	
	-23	-18	-13	-13	-11	-6	-1	-9	-12	-20	<b>G</b>
	-26	-21	-21	-16	-11	-11	-6	-4	-12	-15	<b>C</b>
	-34	-29	-24	-19	-14	-14	-9	-4	-7	-10	<b>A</b>
	-37	-32	-27	-22	-17	-12	-12	-7	-2	-5	<b>T</b>
	-45	-40	-35	-30	-25	-20	-15	-10	-5	0	
	C	C	G	T	C	T	G	C	G		
A	C	T	G	G	T	C	A	A			
A	T	T	G	G	-	-	C	C	A		
					G	T					
					G	G					

# Alinhamento Global usando Espaço Linear

	C	C	G	T	C	T	G	C	G
0	-5	-10	-15	-20	-25	-30	-35	-40	-45
C	3	-2	-7	-12	-17	-22	-27	-32	-37
A	-10	-2	1	-4	-9	-14	-19	-24	-34
G	-15	-7	-4	4	-1	-6	-11	-16	-21
			-23	-18	-13	-13	-11	-6	-1
			-26	-21	-21	-16	-11	-11	-6
			-34	-29	-24	-19	-14	-14	-9
			-37	-32	-27	-22	-17	-12	-12
			-45	-40	-35	-30	-25	-20	-15
			C	C	G	T	C	T	G
A	C	T	G	G	G	T	C	A	A
A	T	T	G	G	-	-	C	C	A
							G	T	
							G	G	

# Alinhamento Global usando Espaço Linear

		C	C
	0	-5	-10
C	-5	3	-2
	-7	-2	-5
	-10	-5	0
C	C		A

A C T G G G T C A A                    G T  
A T T G G - - C C A                    G G

# Alinhamento Global usando Espaço Linear

		C	C
	0	-5	-10
C	-5	3	-2
	-7	-2	-5
	-10	-5	0
C	C		A

A C T G G G T C A A                    G T  
A T T G G - - C C A                    G G

# Alinhamento Global usando Espaço Linear

		<b>C</b>	<b>C</b>
	0	-5	-10
<b>C</b>	-5	<b>3</b>	-2
	-7	<b>-2</b>	-5
	-10	-5	0
	<b>C</b>	<b>C</b>	

A C T G G G T C A A                    G T  
A T T G G - - C C A                    G G

# Alinhamento Global usando Espaço Linear

		<b>C</b>	<b>C</b>
	0	-5	-10
<b>C</b>	-5	<b>3</b>	-2
	-7	<b>-2</b>	-5
	-10	-5	0
	<b>C</b>	<b>C</b>	

A C T G G G T C A A C C G T  
A T T G G - - C C A C A G G

# Alinhamento Global usando Espaço Linear

		C	T	G	C	G		
		0	-5	-10	-15	-20	-25	
C		-5	3	-2	-7	-12	-17	
		-14	-14	-9	-4	-7	-10	A
		-17	-12	-12	-7	-2	-5	T
		-25	-20	-15	-10	-5	0	
C		T	G	C	G			
A	C	T	G	G	G	T		
A	T	T	G	G	-	-		
		C	C	A	C	A		
		G	G	A	G	G		

Sequence 1: A C T G G G T C A A C C G T  
Sequence 2: A T T G G - - C C A C A G G

# Alinhamento Global usando Espaço Linear

		C	T	G	C	G								
C	0	-5	-10	-15	-20	-25								
	-5	3	-2	-7	-12	-17								
	-14	-14	-9	-4	-7	-10	A							
	-17	-12	-12	-7	-2	-5	T							
	-25	-20	-15	-10	-5	0								
C T G C G														
A	C	T	G	G	T	C	A	A	C	C	G	T		
A	T	T	G	G	-	-	C	C	A	C	A	G	G	

# Alinhamento Global usando Espaço Linear

		C	T	G	C	G		
		0	-5	-10	-15	-20	-25	
C		-5	3	-2	-7	-12	-17	
		-14	-14	-9	-4	-7	-10	A
		-17	-12	-12	-7	-2	-5	T
		-25	-20	-15	-10	-5	0	
C		T	G	C	G			
A	C	T	G	G	G	T		
A	T	T	G	G	-	-		
				C	C	A		
				A	C	A		
				G	G	G		

# Alinhamento Global usando Espaço Linear

		<b>C</b>	T	G	C	G
		0	-5	-10	-15	-20
<b>C</b>	-5	<b>3</b>	-2	-7	-12	-17
		-14	<b>-14</b>	-9	-4	-7
		-17	-12	-12	-7	-2
		-25	-20	-15	-10	-5
		<b>C</b>	T	G	C	G
A	C	T	G	G	T	C
A	T	T	G	G	-	-
				C	C	A
				A	C	A
				C	G	G
				A	G	G
				C	C	-

# Alinhamento Global usando Espaço Linear

		C	T	G	C	G
	0	-5	-10	-15	-20	-25
C	-5	3	-2	-7	-12	-17
	-14	-14	-9	-4	-7	-10
	-17	-12	-12	-7	-2	-5
	-25	-20	-15	-10	-5	0
	C	T	G	C	G	
A	C	T	G	G	T	C
A	T	T	G	G	-	-
				C	C	A
				A	C	A
				C	G	G
				G	G	C
				T	C	T
				C	T	

# Alinhamento Global usando Espaço Linear

			G	C	G			
			0	-5	-10	-15		
A	-5		-2	-7	-12			
			-12	-7	-2	-5	T	
			-15	-10	-5	0		
			G	C	G			
A	C	T	G	G	G	T	C	A
A	T	T	G	G	-	-	C	C
							A	A
							C	C
							G	G
							C	T
							-	-

# Alinhamento Global usando Espaço Linear

		G	C	G		
	A	0	-5	-10	-15	
A	-5	-2	<b>-7</b>	-12		
		-12	-7	<b>-2</b>	-5	T
		-15	-10	-5	0	
		G	C	G		
A	C	T	G	G	T	
A	T	T	G	G	-	
		C	C	A	C	
		C	A	C	A	
		G	G	G	G	
		C	C	C	C	
		T	-	T	T	

# Alinhamento Global usando Espaço Linear

			G	C	G			
			0	-5	-10	-15		
A	-5	-2	<b>-7</b>	-12				
			-12	-7	<b>-2</b>	-5	T	
			-15	-10	-5	0		
			G	C	G			
A	C	T	G	G	A	C	C	T
A	T	T	G	G	-	C	C	-

# Alinhamento Global usando Espaço Linear

			<b>G</b>	<b>C</b>	<b>G</b>			
			0	-5	-10	-15		
<b>A</b>	-5	-2	<b>-7</b>	-12				
			-12	-7	<b>-2</b>	-5	<b>T</b>	
			-15	-10	-5	0		
			<b>G</b>	<b>C</b>	<b>G</b>			
A	C	T	G	G	A	A	C	T
A	T	T	G	G	-	-	C	C
					C	C	A	A
					G	G	G	T
					C	G	C	T
					-	-	T	

# Alinhamento Global usando Espaço Linear

			<b>G</b>	<b>C</b>	<b>G</b>			
			0	-5	-10	-15		
<b>A</b>	-5	-2	-7	-12				
			-12	-7	-2	-5	T	
			-15	-10	-5	0		
			<b>G</b>	<b>C</b>	<b>G</b>			
A	C	T	G	G	A	A	C	T
A	T	T	G	G	-	-	C	C
					C	C	A	A
					C	A	G	G
					G	G	C	T
					C	G		
					-	-	T	

# Alinhamento Global usando Espaço Linear

		G
	0	-5
A	-5	-2

A C T G G G T C A A C C C G T C T G C - A - T  
A T T G G - - C C A C A G G C C - A - T

# Alinhamento Global usando Espaço Linear

x	$\alpha$	A	C	T	G	G	T	C	A	A	C	C	G	T	C	T	G	C	G	
$\beta$	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55	-60	-65	-70	-75	-80	-85	-90	-95
A	-5	3	-2	-7	-12	-17	-22	-27	-32	-37	-42	-47	-52	-57	-62	-67	-72	-77	-82	-87
T	-10	-2	1	-4	-9	-14	-19	-24	-29	-34	-39	-44	-49	-54	-59	-64	-69	-74	-79	
T	-15	-7	-4	4	-1	-6	-11	-11	-16	-21	-26	-31	-36	-41	-46	-51	-56	-61	-66	-71
G	-20	-12	-9	-1	7	2	-3	-8	-13	-18	-23	-28	-33	-33	-38	-43	-48	-53	-58	-63
G	-25	-17	-14	-6	2	10	5	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-45	-50	-55
C	-30	-22	-14	-11	-3	5	8	3	3	-2	-7	-12	-17	-22	-27	-32	-37	-42	-42	-47
C	-35	-27	-19	-16	-8	0	3	6	6	1	-4	-4	-9	-14	-19	-24	-29	-34	-39	-44
A	-40	-32	-24	-21	-13	-5	-2	1	4	9	4	-1	-6	-11	-16	-21	-26	-31	-36	-41
C	-45	-37	-29	-26	-18	-10	-7	-4	4	4	7	7	2	-3	-8	-13	-18	-23	-28	-33
A	-50	-42	-34	-31	-23	-15	-12	-9	-1	7	7	5	5	0	-5	-10	-15	-20	-25	-30
G	-55	-47	-39	-36	-28	-20	-12	-14	-6	2	5	5	3	8	3	-2	-7	-12	-17	-22
G	-60	-52	-44	-41	-33	-25	-17	-14	-11	-3	0	3	3	6	6	1	-4	-4	-9	-14
C	-65	-57	-49	-46	-38	-30	-22	-19	-11	-8	-5	3	6	1	4	9	4	-1	-1	-6
A	-70	-62	-54	-51	-43	-35	-27	-24	-16	-8	-5	-2	1	4	-1	4	7	2	-3	-3
T	-75	-67	-59	-51	-48	-40	-32	-24	-21	-13	-10	-7	-4	-1	7	2	7	5	0	-5

-5	=	3	-2	3	3	3	-5	-5	3	-2	3	3	-2	3	-2	3	-5	-2	-5	-2
$\alpha$	=	A	C	T	G	G	G	T	C	A	A	C	C	G	T	C	T	G	C	G
$\beta$	=	A	T	T	G	G	-	-	C	C	A	C	A	G	G	C	-	A	-	T

# Complexidade de Tempo e de Espaço

- Complexidade de Espaço:
  - ▶  $E(m, n) = \Theta(m + n)$
- Complexidade de Tempo:
  - ▶ Primeira observação: o número de chamadas recursivas é no máximo proporcional aos tamanhos das sequências  $\alpha$  e  $\beta$ , já que conseguimos alinhar pelo menos um caractere a cada chamada recursiva:
    - ★  $T(m, n) = O(m + n) \times O(mn) = O(m^2n + mn^2)$
  - ▶ Segunda observação: a cada passo do algoritmo, metade da matriz  $M$  é descartada, logo o número de chamadas recursivas deve ser no máximo proporcional ao logaritmo do tamanho da matriz:
    - ★  $T(m, n) = O(\lceil \log_2 mn \rceil) \times O(mn) = O(mn \log mn)$
  - ▶ Terceira observação: podemos obter um resultado melhor fazendo uma análise assintótica agregada (soma de todos os passos do algoritmo):
    - ★  $T(m, n) = \sum_{k=0}^{\lceil \log_2 mn \rceil} \Theta\left(\frac{mn}{2^k}\right)$
    - ★  $T(m, n) = \Theta\left(\sum_{k=0}^{\lceil \log_2 mn \rceil} \frac{mn}{2^k}\right)$
    - ★  $T(m, n) = \Theta\left(mn \sum_{k=0}^{\lceil \log_2 mn \rceil} \frac{1}{2^k}\right) = \Theta\left(mn \sum_{k=0}^{\infty} \frac{1}{2^k}\right)$
    - ★  $T(m, n) = \Theta(mn)$

# Alinhamento com Pontuação Afim e Espaço Linear

- Algoritmo de Daniel Hirschberg (1975):
  - ▶ Pontuação Aditiva
  - ▶ Divisão e Conquista + Programação Dinâmica (1 matriz)
  - ▶ Complexidade de Espaço:  $\Theta(m + n)$
  - ▶ Complexidade de Tempo:  $\Theta(mn)$
- Algoritmo de Osamu Gotoh (1982):
  - ▶ Pontuação Afim
  - ▶ Programação Dinâmica (3 matrizes)
  - ▶ Complexidade de Espaço:  $\Theta(mn)$
  - ▶ Complexidade de Tempo:  $\Theta(mn)$
- Algoritmo de Eugene Myers e Webb Miller (1988):
  - ▶ Pontuação Afim
  - ▶ Divisão e Conquista + Programação Dinâmica (3 matrizes)
  - ▶ Complexidade de Espaço:  $\Theta(m + n)$
  - ▶ Complexidade de Tempo:  $\Theta(mn)$

# Alinhamento de Sequências com Função Convexa para Penalidade de Buracos

- Algoritmo proposto por Webb Miller e Eugene Myers (1988).
- Complexidade:
  - ▶ Tempo (pior caso):  $\Theta(mn \log mn)$ .
  - ▶ Espaço (esperado):  $\Theta(m + n)$ .
- Pouco utilizado na prática.

# Exercícios

## Exercício

*Escreva algoritmos, em pseudocódigo, que dadas duas sequências  $\alpha$  e  $\beta$ , com respectivamente  $m$  e  $n$  caracteres, calcule:*

- *O valor de um alinhamento global ótimo entre  $\alpha$  e  $\beta$ .*
- *O valor de um alinhamento semi-global ótimo entre  $\alpha$  e  $\beta$ .*
- *O valor de um alinhamento local ótimo entre  $\alpha$  e  $\beta$ .*

*Seus algoritmos devem considerar função aditiva para penalidade de blocos de buracos e usar espaço linear em relação ao tamanho da entrada.*

# Alinhamento Global de Sequências Similares

- Suponha que queremos alinhar duas sequências  $\alpha$  e  $\beta$  de mesmo tamanho  $n$  (é fácil adaptar o algoritmo para funcionar com duas sequências de tamanho diferentes).
- Se as duas sequências são similares, é razoável supor que existam poucos buracos no alinhamento ótimo entre as duas sequências.
- Seja  $2k$  o número de buracos no alinhamento ótimo.
  - ▶ Note que no caso das duas sequências terem o mesmo tamanho, o número de buracos será sempre par.
- O que podemos afirmar em relação as células da matriz que representam alinhamentos ótimos entre  $\alpha$  e  $\beta$ ?
  - ▶ Elas estão, no máximo, a  $k$  células de distância da diagonal principal.
- Ideia: preencher apenas as células que estão a  $k$  células de distância da diagonal principal da matriz de Programação Dinâmica.

# Alinhamento Global de Sequências Similares

- Como determininar o valor de  $k$ , sem conhecer as sequências?
- Seja  $M$  a pontuação para *match*,  $m$  para *mismatch* e  $g$  para *gap*.
- Qual a maior pontuação possível para um alinhamento que não esteja completamente contido na faixa de  $k$  células de distância da diagonal principal?
  - ▶  $M \times (n - k - 1) + 2 \times (k + 1) \times g$
- Algoritmo:
  - ① Inicialização:  $k = 1$
  - ② Preencha a matriz de Programação Dinâmica considerando apenas as células com distância no máximo  $k$  para a diagonal principal.
  - ③ Se  $M[n, n] > M \times (n - k - 1) + 2 \times (k + 1) \times g$ , pare e retorne o alinhamento máximo.
  - ④ Caso contrário, incremente o valor de  $k$  e volte ao passo 2.
- Algoritmo proposto por James Fickett (1983).

# Alinhamento Global de Sequências Similares

$\times$	$\alpha$	G	C	G	A	A	A	G	T	G	A	T	T	T	C	C	T	C	C	
$\beta$	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55	-60	-65	-70	-75	-80	-85	-90	-95
C	-5	-2	-2	-7	-12	-17	-22	-27	-32	-37	-42	-47	-52	-57	-62	-67	-72	-77	-82	-87
A	-10	-7	-4	-4	-4	-9	-14	-19	-24	-29	-34	-39	-44	-49	-54	-59	-64	-69	-74	-79
G	-15	-7	-9	-1	-6	-6	-11	-16	-16	-21	-26	-31	-36	-41	-46	-51	-56	-61	-66	-71
G	-20	-12	-9	-6	-3	-8	-8	-13	-13	-18	-18	-23	-28	-33	-38	-43	-48	-53	-58	-63
G	-25	-17	-14	-6	-8	-5	-10	-10	-10	-15	-15	-20	-25	-30	-35	-40	-45	-50	-55	-60
C	-30	-22	-14	-11	-8	-10	-7	-12	-12	-12	-17	-17	-22	-27	-32	-32	-37	-42	-47	-52
G	-35	-27	-19	-11	-13	-10	-12	-9	-9	-14	-9	-14	-19	-24	-29	-34	-34	-39	-44	-49
A	-40	-32	-24	-16	-8	-10	-7	-9	-11	-11	-14	-6	-11	-16	-21	-26	-31	-36	-41	-46
A	-45	-37	-29	-21	-13	-5	-7	-4	-9	-13	-13	-11	-8	-13	-18	-23	-28	-33	-38	-43
A	-50	-42	-34	-26	-18	-10	-2	-4	-6	-11	-15	-10	-13	-10	-15	-20	-25	-30	-35	-40
A	-55	-47	-39	-31	-23	-15	-7	1	-4	-8	-13	-12	-12	-15	-12	-17	-22	-27	-32	-37
G	-60	-52	-44	-36	-28	-20	-12	-4	4	-1	-5	-10	-14	-14	-17	-14	-19	-24	-29	-34
T	-65	-57	-49	-41	-33	-25	-17	-9	-1	7	2	-3	-7	-11	-11	-16	-16	-21	-26	
G	-70	-62	-54	-46	-38	-30	-22	-14	-6	2	10	5	0	-5	-10	-13	-18	-18	-23	
A	-75	-67	-59	-51	-43	-35	-27	-19	-11	-3	5	13	8	3	-2	-7	-12	-17	-20	-20
T	-80	-72	-64	-56	-48	-40	-32	-24	-16	-8	0	8	16	11	6	1	-4	-9	-14	-19
G	-85	-77	-69	-61	-53	-45	-37	-29	-21	-13	-5	3	11	14	9	4	-1	-6	-11	-16
T	-90	-82	-74	-66	-58	-50	-42	-34	-26	-18	-10	-2	6	14	17	12	7	2	-3	-8
A	-95	-87	-79	-71	-63	-55	-47	-39	-31	-23	-15	-7	1	9	12	15	10	5	0	-5

n = 19	M = 3	m = -2	g = -5
--------	-------	--------	--------

# Alinhamento Global de Sequências Similares

$\alpha$	G	C	G	A	A	A	A	G	T	G	A	T	T	T	C	C	T	C	C
$\beta$	0	-5																	
C	-5	-2	-2																
A		-7	-4	-4															
G			-9	-1	-6														
G				-6	-3	-8													
G					-8	-5	-10												
C						-10	-7	-12											
G							-12	-9	-9										
A								-9	-11	-11									
A									-11	-13	-13								
A										-13	-15	-10							
A											-15	-12	-12						
G												-17	-14	-14					
T													-14	-11	-11				
G													-16	-13	-13				
A														-18	-15	-15			
T															-20	-17	-12		
G																-22	-17	-14	
T																	-19	-19	-16
A																		-21	-21

n = 19 M = 3 m = -2 g = -5 k = 1

máximo = 31

# Alinhamento Global de Sequências Similares

$\alpha$	G	C	G	A	A	A	A	G	T	G	A	T	T	T	C	C	T	C	C	
$\beta$	0	-5	-10																	
C	-5	-2	-2	-7																
A	-10	-7	-4	-4	-4															
G		-7	-9	-1	-6	-6														
G			-9	-6	-3	-8	-8													
G				-6	-8	-5	-10	-10												
C					-8	-10	-7	-12	-12											
G						-10	-12	-9	-9	-14										
A							-7	-9	-11	-11	-16									
A								-4	-9	-13	-13	-13								
A									-6	-11	-15	-10	-15							
A										-8	-13	-12	-12	-17						
G											-5	-10	-14	-14	-19					
T												-7	-7	-11	-11	-16				
G													-9	-9	-13	-13	-18			
A														-11	-11	-15	-15	-20		
T															-8	-13	-17	-12	-17	
G																-10	-15	-17	-14	-19
T																	-12	-12	-17	-16
A																		-14	-14	-19

n = 19 M = 3 m = -2 g = -5 k = 2

máximo = 18

# Alinhamento Global de Sequências Similares

$\times$	$\alpha$	G	C	G	A	A	A	G	T	G	A	T	T	T	C	C	T	C	C	
$\beta$	0	-5	-10	-15																
C	-5	-2	-2	-7	-12															
A	-10	-7	-4	-4	-4	-9														
G	-15	-7	-9	-1	-6	-6	-11													
G		-12	-9	-6	-3	-8	-8	-13												
G			-14	-6	-8	-5	-10	-10	-10	-10										
C				-11	-8	-10	-7	-12	-12	-12										
G					-13	-10	-12	-9	-9	-14	-9									
A						-10	-7	-9	-11	-11	-14	-6								
A							-7	-4	-9	-13	-13	-11	-8							
A								-4	-6	-11	-15	-10	-13	-10						
A									-6	-8	-13	-12	-12	-15	-12					
G										-8	-5	-10	-14	-14	-17	-14				
T											-10	-7	-7	-11	-11	-16	-16			
G												-12	-9	-9	-13	-13	-18	-18		
A													-14	-11	-11	-15	-15	-20	-20	
T														-11	-8	-13	-17	-12	-17	-22
G															-13	-10	-15	-17	-14	-19
T																-15	-12	-12	-17	-16
A																	-17	-14	-14	-19

n = 19 M = 3 m = -2 g = -5 k = 3

máximo = 5

# Alinhamento Global de Sequências Similares

$\alpha$	G	C	G	A	A	A	G	T	G	A	T	T	T	C	C	T	C	C	
$\beta$	0	-5	-10	-15	-20														
C	-5	-2	-2	-7	-12	-17													
A	-10	-7	-4	-4	-4	-9	-14												
G	-15	-7	-9	-1	-6	-6	-11	-16											
G	-20	-12	-9	-6	-3	-8	-8	-13	-13										
G		-17	-14	-6	-8	-5	-10	-10	-10	-15									
C			-14	-11	-8	-10	-7	-12	-12	-12	-17								
G				-11	-13	-10	-12	-9	-9	-14	-9	-14							
A					-8	-10	-7	-9	-11	-11	-14	-6	-11						
A						-5	-7	-4	-9	-13	-13	-11	-8	-13					
A							-2	-4	-6	-11	-15	-10	-13	-10	-15				
A								1	-4	-8	-13	-12	-12	-15	-12	-17			
G									4	-1	-5	-10	-14	-14	-17	-14	-19		
T										7	2	-3	-7	-11	-11	-16	-16	-16	
G											10	5	0	-5	-10	-13	-18	-18	
A												13	8	3	-2	-7	-12	-17	
T													16	11	6	1	-4	-9	
G														14	9	4	-1	-6	
T															17	12	7	-3	
A																15	10	5	0
																		-5	

n = 19 M = 3 m = -2 g = -5 k = 4

máximo = -8

# Alinhamento Global de Sequências Similares

$\alpha$	G	C	G	A	A	A	G	T	G	A	T	T	T	C	C	T	C	C	
$\beta$	0	-5	-10	-15	-20	-25													
C	-5	-2	-2	-7	-12	-17	-22												
A	-10	-7	-4	-4	-4	-9	-14	-19											
G	-15	-7	-9	-1	-6	-6	-11	-16	-16										
G	-20	-12	-9	-6	-3	-8	-8	-13	-13	-18									
G	-25	-17	-14	-6	-8	-5	-10	-10	-10	-15	-15								
C		-22	-14	-11	-8	-10	-7	-12	-12	-12	-17	-17							
G			-19	-11	-13	-10	-12	-9	-9	-14	-9	-14	-19						
A				-16	-8	-10	-7	-9	-11	-11	-14	-6	-11	-16					
A					-13	-5	-7	-4	-9	-13	-13	-11	-8	-13	-18				
A						-10	-2	-4	-6	-11	-15	-10	-13	-10	-15	-20			
A							-7	1	-4	-8	-13	-12	-12	-15	-12	-17	-22		
G								-4	4	-1	-5	-10	-14	-14	-17	-14	-19	-24	
T									-1	7	2	-3	-7	-11	-11	-16	-16	-21	
G										2	10	5	0	-5	-10	-13	-18	-18	
A											5	13	8	3	-2	-7	-12	-17	-20
T												8	16	11	6	1	-4	-9	-14
G													11	14	9	4	-1	-6	-11
T														14	17	12	7	2	-3
A															12	15	10	5	0
																		-5	

n = 19 M = 3 m = -2 g = -5 k = 5

máximo = -21

# Alinhamento Global de Sequências Similares

$\times$	$\alpha$	G	C	G	A	A	A	G	T	G	A	T	T	T	C	C	T	C	C	
$\beta$	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55	-60	-65	-70	-75	-80	-85	-90	-95
C	-5	-2	-2	-7	-12	-17	-22	-27	-32	-37	-42	-47	-52	-57	-62	-67	-72	-77	-82	-87
A	-10	-7	-4	-4	-4	-9	-14	-19	-24	-29	-34	-39	-44	-49	-54	-59	-64	-69	-74	-79
G	-15	-7	-9	-1	-6	-6	-11	-16	-16	-21	-26	-31	-36	-41	-46	-51	-56	-61	-66	-71
G	-20	-12	-9	-6	-3	-8	-8	-13	-13	-18	-18	-23	-28	-33	-38	-43	-48	-53	-58	-63
G	-25	-17	-14	-6	-8	-5	-10	-10	-10	-15	-15	-20	-25	-30	-35	-40	-45	-50	-55	-60
C	-30	-22	-14	-11	-8	-10	-7	-12	-12	-12	-17	-17	-22	-27	-32	-32	-37	-42	-47	-52
G	-35	-27	-19	-11	-13	-10	-12	-9	-9	-14	-9	-14	-19	-24	-29	-34	-34	-39	-44	-49
A	-40	-32	-24	-16	-8	-10	-7	-9	-11	-11	-14	-6	-11	-16	-21	-26	-31	-36	-41	-46
A	-45	-37	-29	-21	-13	-5	-7	-4	-9	-13	-13	-11	-8	-13	-18	-23	-28	-33	-38	-43
A	-50	-42	-34	-26	-18	-10	-2	-4	-6	-11	-15	-10	-13	-10	-15	-20	-25	-30	-35	-40
A	-55	-47	-39	-31	-23	-15	-7	1	-4	-8	-13	-12	-12	-15	-12	-17	-22	-27	-32	-37
G	-60	-52	-44	-36	-28	-20	-12	-4	4	-1	-5	-10	-14	-14	-17	-14	-19	-24	-29	-34
T	-65	-57	-49	-41	-33	-25	-17	-9	-1	7	2	-3	-7	-11	-11	-16	-16	-21	-26	
G	-70	-62	-54	-46	-38	-30	-22	-14	-6	2	10	5	0	-5	-10	-13	-18	-18	-23	
A	-75	-67	-59	-51	-43	-35	-27	-19	-11	-3	5	13	8	3	-2	-7	-12	-17	-20	-20
T	-80	-72	-64	-56	-48	-40	-32	-24	-16	-8	0	8	16	11	6	1	-4	-9	-14	-19
G	-85	-77	-69	-61	-53	-45	-37	-29	-21	-13	-5	3	11	14	9	4	-1	-6	-11	-16
T	-90	-82	-74	-66	-58	-50	-42	-34	-26	-18	-10	-2	6	14	17	12	7	2	-3	-8
A	-95	-87	-79	-71	-63	-55	-47	-39	-31	-23	-15	-7	1	9	12	15	10	5	0	-5

n = 19	M = 3	m = -2	g = -5
--------	-------	--------	--------

# Complexidade de Tempo e de Espaço

- Complexidade de Espaço:

- ▶  $E(n) = \Theta(kn)$
- ▶ Pior caso  $k = \Theta(n)$ , logo:  $E(n) = \Theta(n^2)$
- ▶ Usando as ideias de Hirschberg, é possível implementar o algoritmo usando apenas espaço linear no tamanho das sequências ( $E(n) = \Theta(n)$ ).

- Complexidade de Tempo:

- ▶ Primeira abordagem:

- ★ Incremento (passo 4):  $k = k + 1$
- ★  $T(n) = \sum_{i=1}^k \Theta(in) = \Theta(n) \sum_{i=1}^k i$
- ★  $T(n) = \Theta(n) \frac{(k+1)k}{2} = \Theta(nk^2)$
- ★ Pior caso  $k = \Theta(n)$ , logo:  $T(n) = \Theta(n^3)$

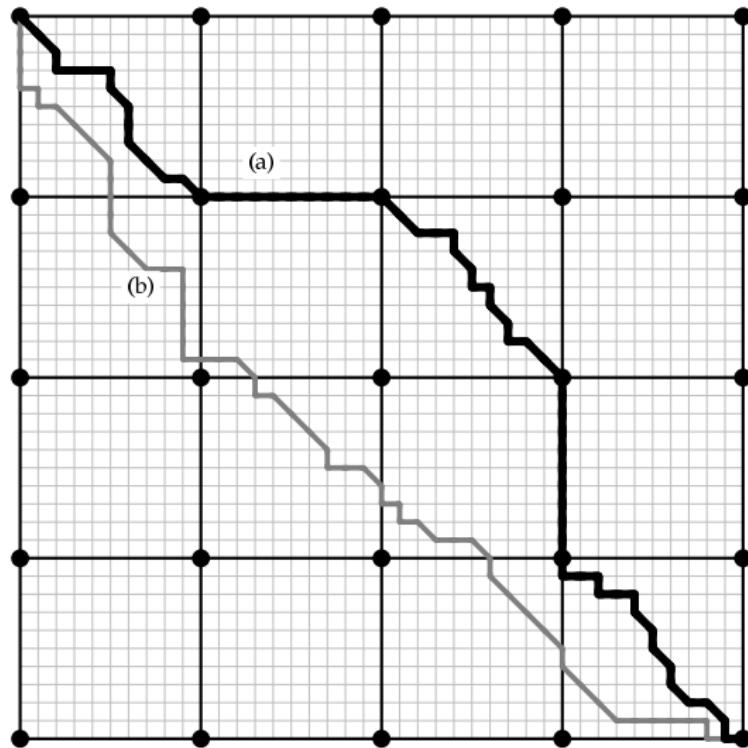
- ▶ Segunda abordagem:

- ★ Incremento (passo 4):  $k = 2 \times k$
- ★  $T(n) = \sum_{i=0}^{\log_2 k} \Theta(2^i n) = \Theta(n) \sum_{i=0}^{\log_2 k} 2^i$
- ★  $T(n) = \Theta(n)(2^{1+\log_2 k} - 1) = \Theta(n)(2k - 1) = \Theta(nk)$
- ★ Pior caso  $k = \Theta(n)$ , logo:  $T(n) = \Theta(n^2)$

# Algoritmo Sub-Quadrático para Alinhamento Global

- É possível obter um algoritmo de tempo sub-quadrático para alinhamento global considerando pontuação aditiva para buracos?
- Por simplificação, vamos considerar a seguinte variação do problema de alinhamento global chamada de alinhamento de blocos:
  - ▶ Dadas duas sequências  $\alpha$  e  $\beta$  de tamanhos, respectivamente,  $m = m' \times t$  e  $n = n' \times t$ , ambas divididas em blocos de  $t$  caracteres consecutivos, encontrar o melhor alinhamento entre os blocos da sequência  $\alpha$  com os blocos da sequência  $\beta$ .
  - ▶ Note que cada um dos blocos das sequências deve ser alinhado com um bloco da outra sequência ou com um bloco de buracos de tamanho  $t$ .
- Na figura do próximo slide é mostrado (a) um alinhamento de blocos e (b) um alinhamento global entre um par de sequências.

# Alinhamento de Blocos × Alinhamento Global



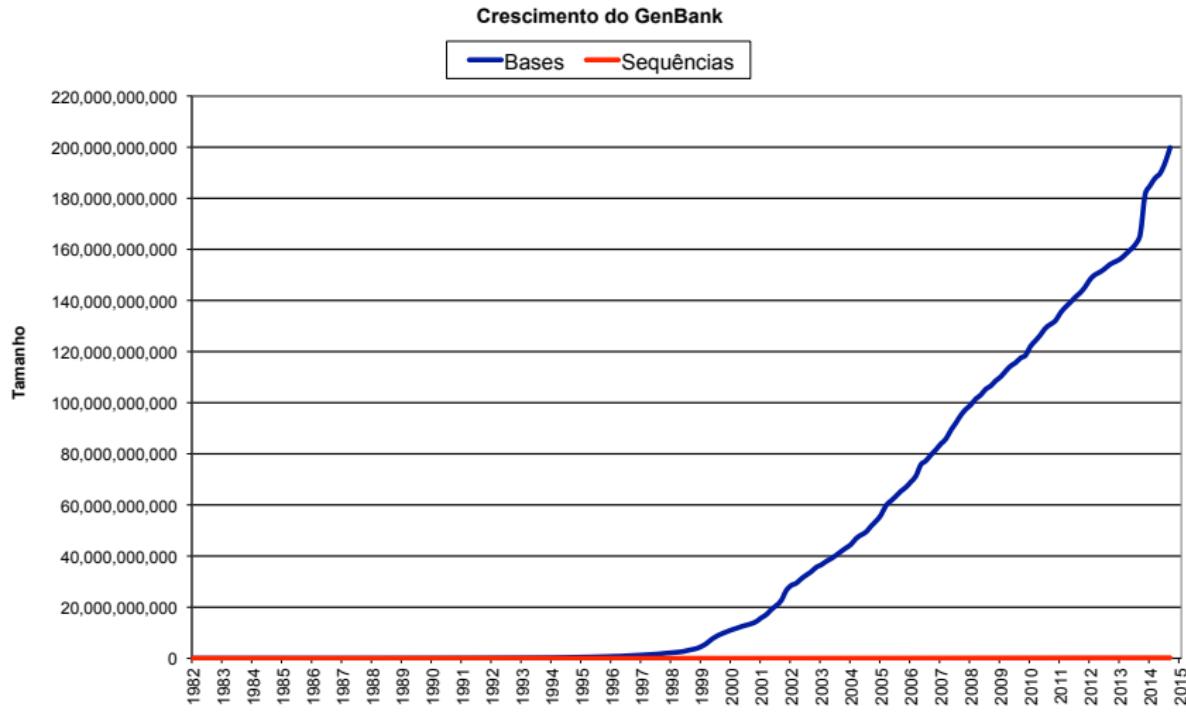
# Algoritmo Sub-Quadrático para Alinhamento de Blocos

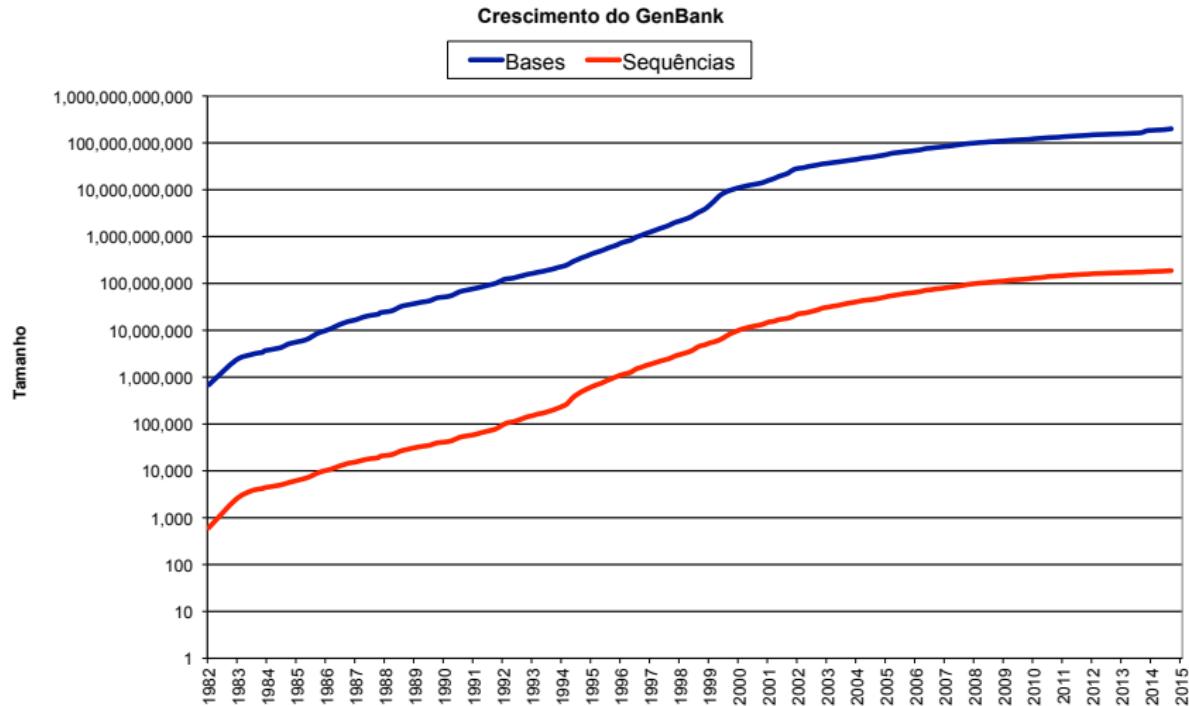
- Four-Russian Algorithm (Arlazarov, Dinic, Kronrod e Faradzev), 1970.
- O método foi desenvolvido apenas por Kronrod, único russo do grupo.
- Considere que ambas as sequências são formadas por caracteres de um alfabeto de quatro letras, por exemplo,  $\{A, C, G, T\}$ .
- Passos do algoritmo:
  - ▶ Dividir a matriz de programação dinâmica em quadrados de dimensão  $t \times t$ .
  - ▶ Pré-computar os alinhamentos de todos os pares de sequências de tamanho  $t$ , em tempo  $\Theta(4^t 4^t t^2) = \Theta(16^t t^2)$ .
  - ▶ Preencher a matriz de Programação Dinâmica, usando os valores pré-computados, em tempo  $\Theta(\frac{m}{t} \frac{n}{t} t) = \Theta(\frac{mn}{t})$ .
- Complexidade:  $T(n) = \Theta(16^t t^2 + \frac{mn}{t})$

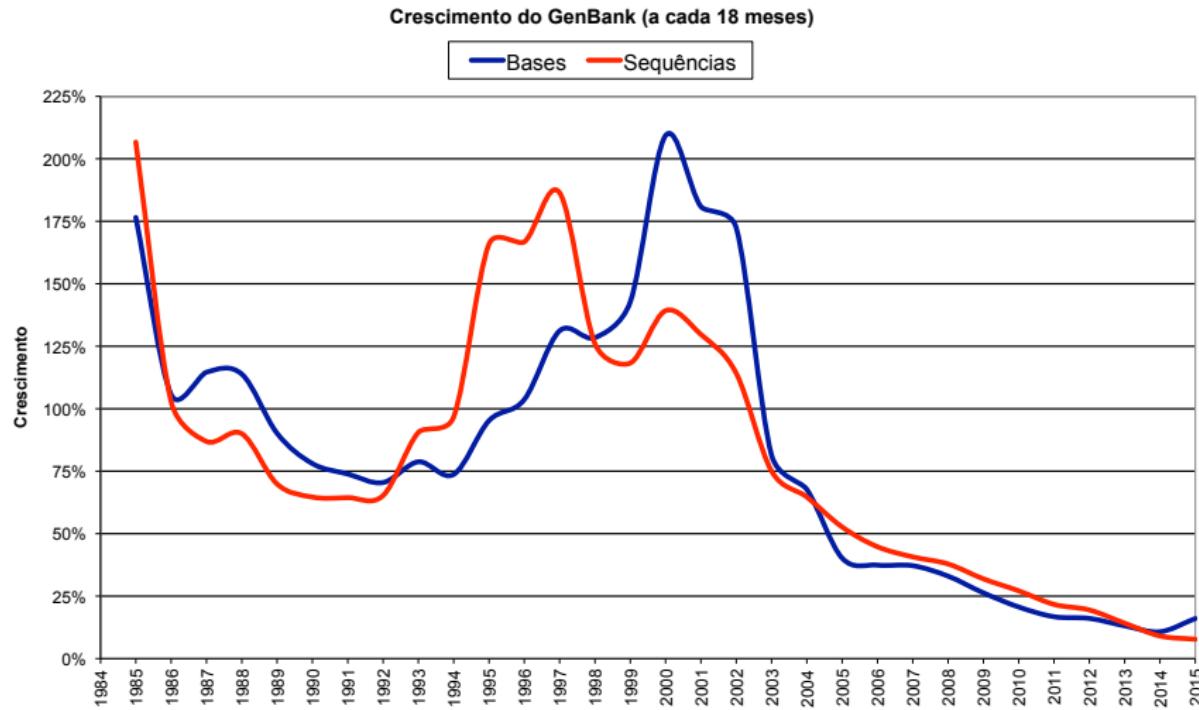
# Algoritmo Sub-Quadrático para Alinhamento de Blocos

- Complexidade:  $T(n) = \Theta(16^t t^2 + \frac{mn}{t})$
- Valor de  $t$  que minimiza a complexidade do algoritmo:
  - ▶  $t = \log_{16}(m + n)$ .
- Complexidade:
  - ▶  $T(n) = \Theta((m + n) \log^2(m + n) + \frac{mn}{\log(m+n)}) = \Theta(\frac{mn}{\log(m+n)})$
- William Masek e Michael Paterson (1980) mostraram como calcular o alinhamento global em tempo  $\Theta(\frac{mn}{\log(m+n)})$ , usando uma adaptação do Four-Russian Algorithm.
- Algoritmo não utilizado na prática.

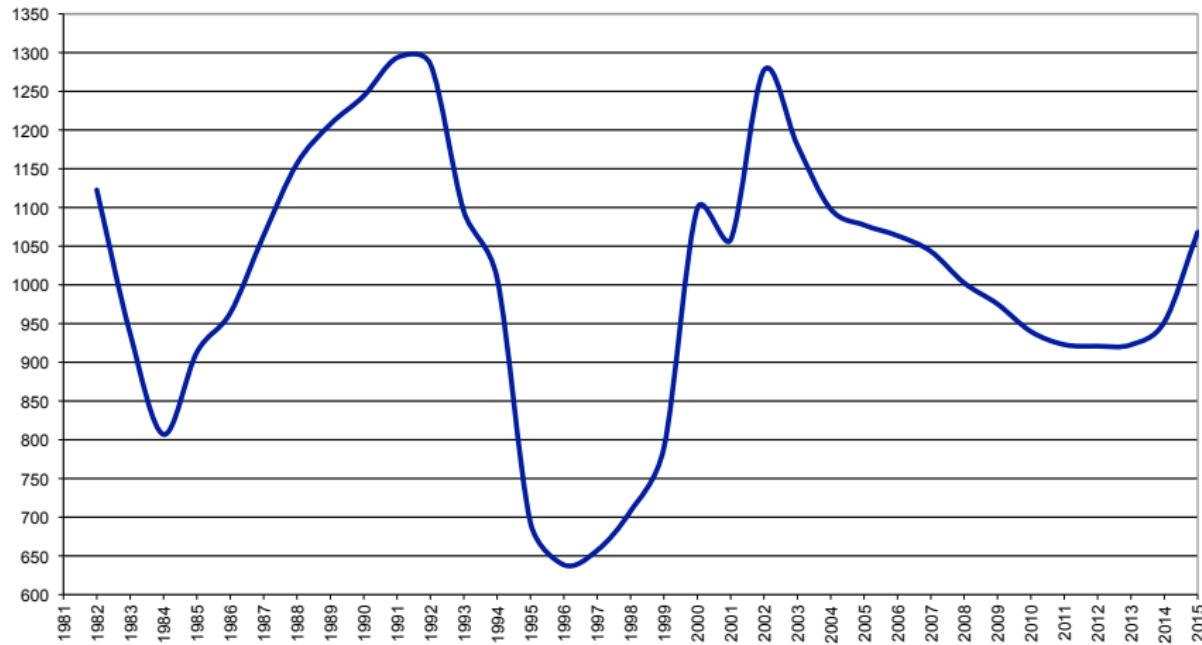
- Maior banco público de sequências de nucleotídeos e de proteínas do mundo.
- Fundado por Walter Goad em 1982.
- Desenvolvida pelo National Center for Biotechnology Information (NCBI) e financiada pelo National Institutes of Health (NIH).
- Crescimento do GenBank:
  - ▶ *From its inception, GenBank has doubled in size about every 18 months (or less). [GenBank, 1997-2008]*
  - ▶ *From its inception, GenBank has grown exponentially, and continues to do so with a current doubling time of 30 months. [GenBank, 2009]*
  - ▶ *From its inception, GenBank has grown exponentially, and continues to do so with the number of sequence records doubling approximately every 35 months. [GenBank, 2010]*







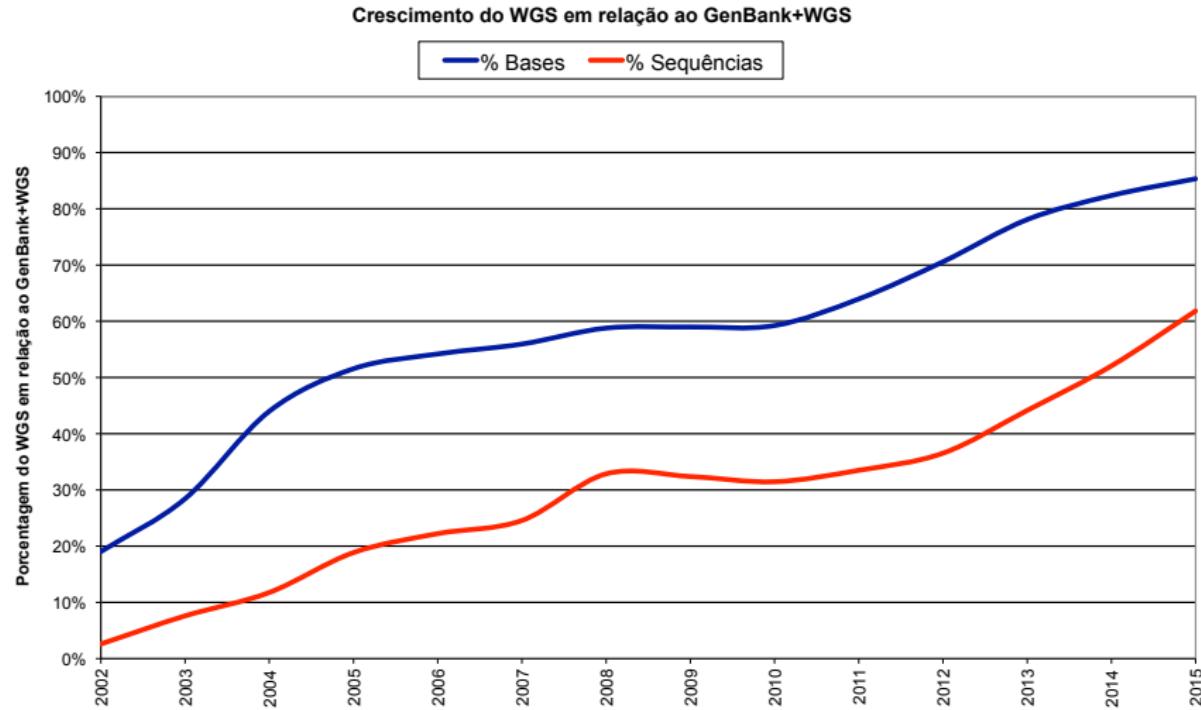
## Tamanho Médio das Sequências do GenBank

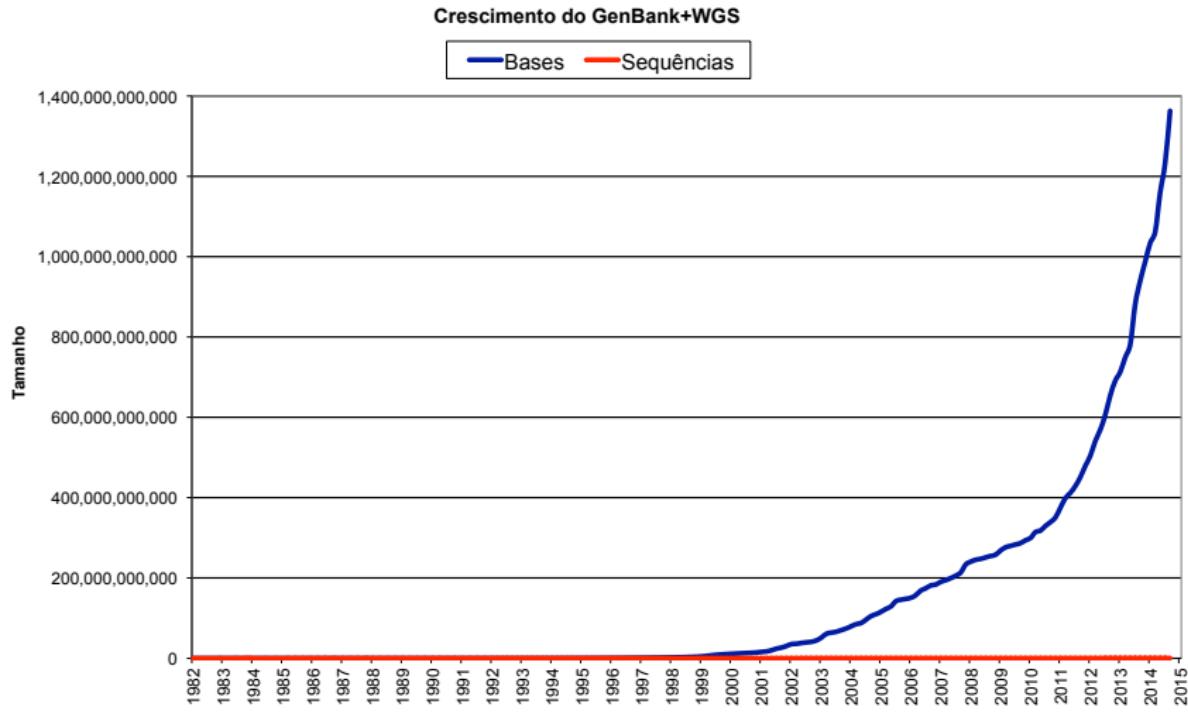


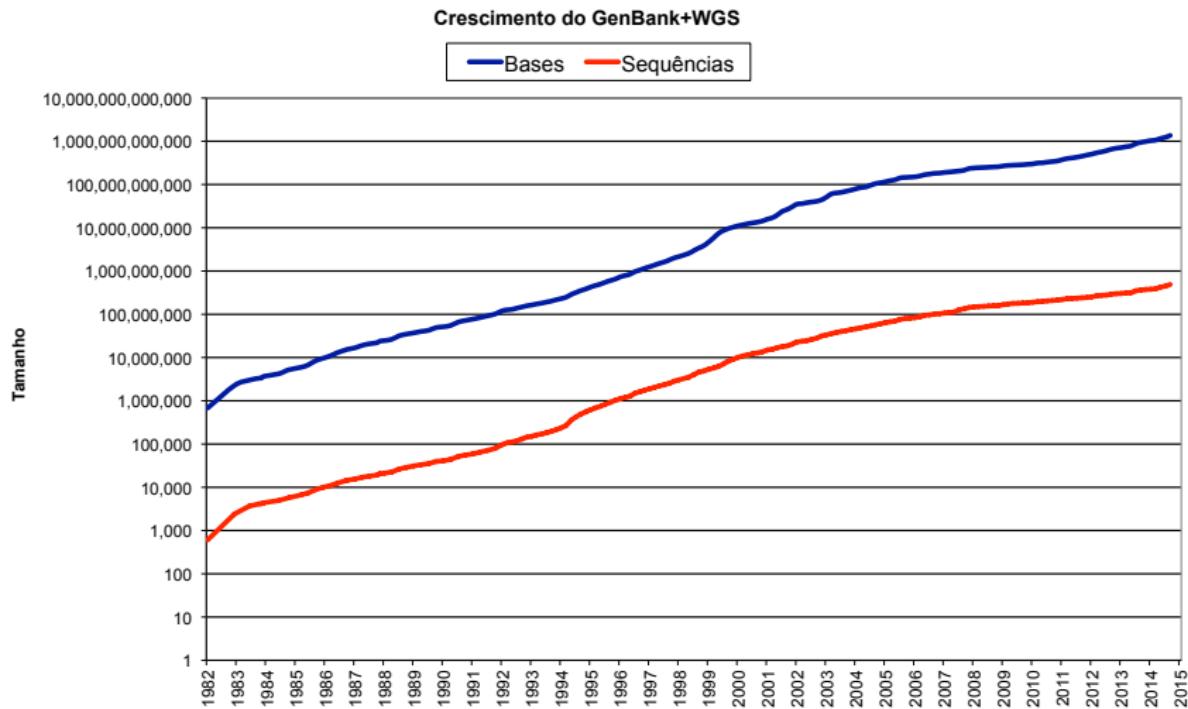
## GenBank - WGS (Whole Genome Shotgun)

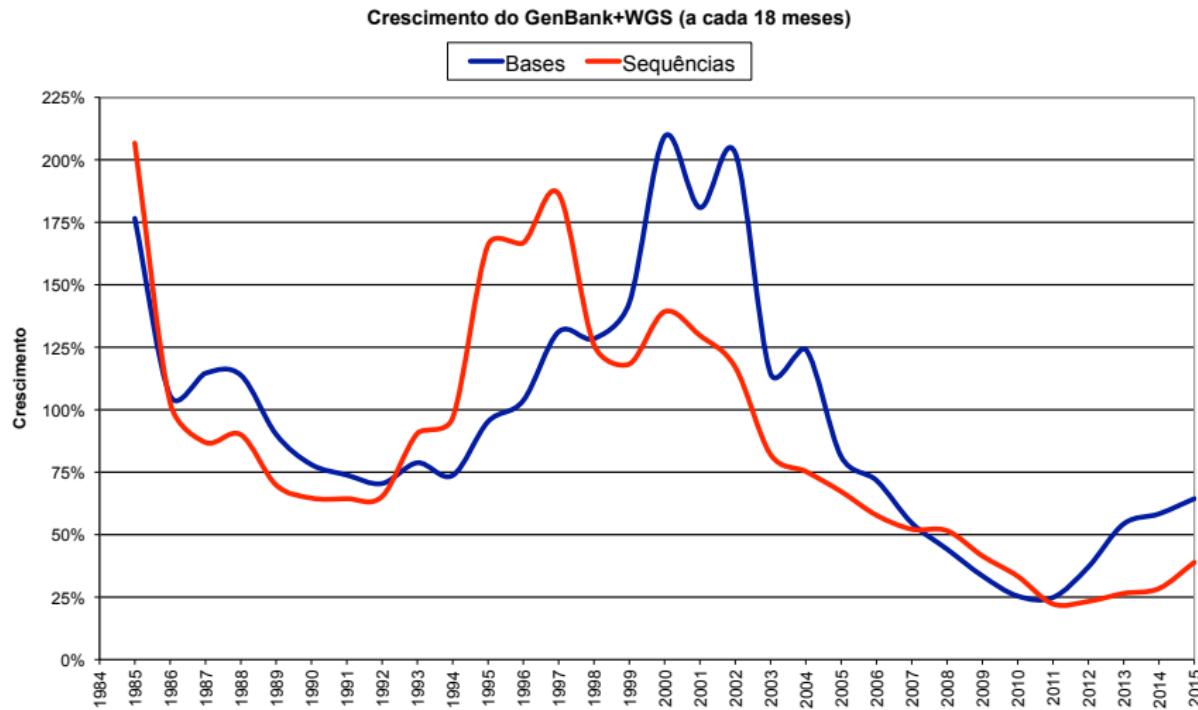
- Com a simplificação, automatização e barateamento do processo de sequenciamento, o número de genomas completamente ou parcialmente sequenciados cresceu muito nos últimos anos.
- Dois importantes marcos da genômica:
  - ▶ Nacional: em 2000, sequenciamento da bactéria *Xylella fastidiosa* (2.7Mbp, 3 mil genes), causadora da doença “amarelinho” que afeta laranjeiras.
  - ▶ Internacional: em 2001, sequenciamento do genoma humano (3Gbp, 30 mil genes).
- Desde 2002 o GenBank possui uma divisão (WGS - Whole Genome Shotgun) dedicado a montagens completas ou parciais de genomas.
- A proporção de bases e sequências do WGS em relação ao GenBank vem crescendo muito nos últimos anos.

# WGS × (GenBank+WGS)

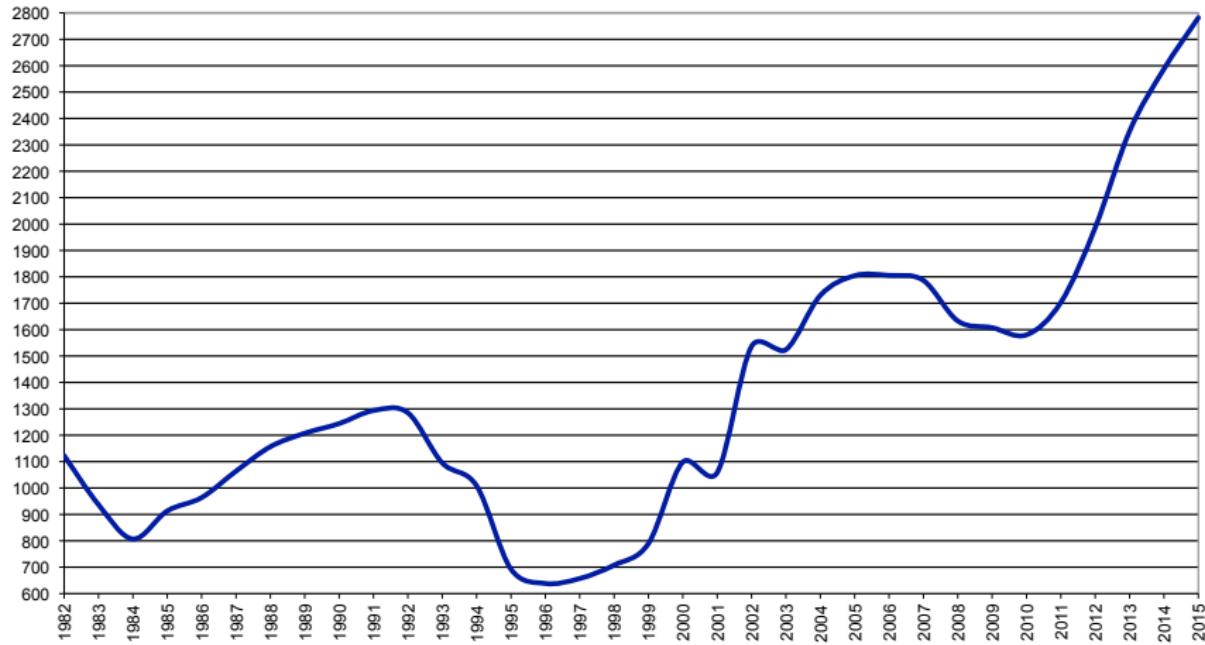




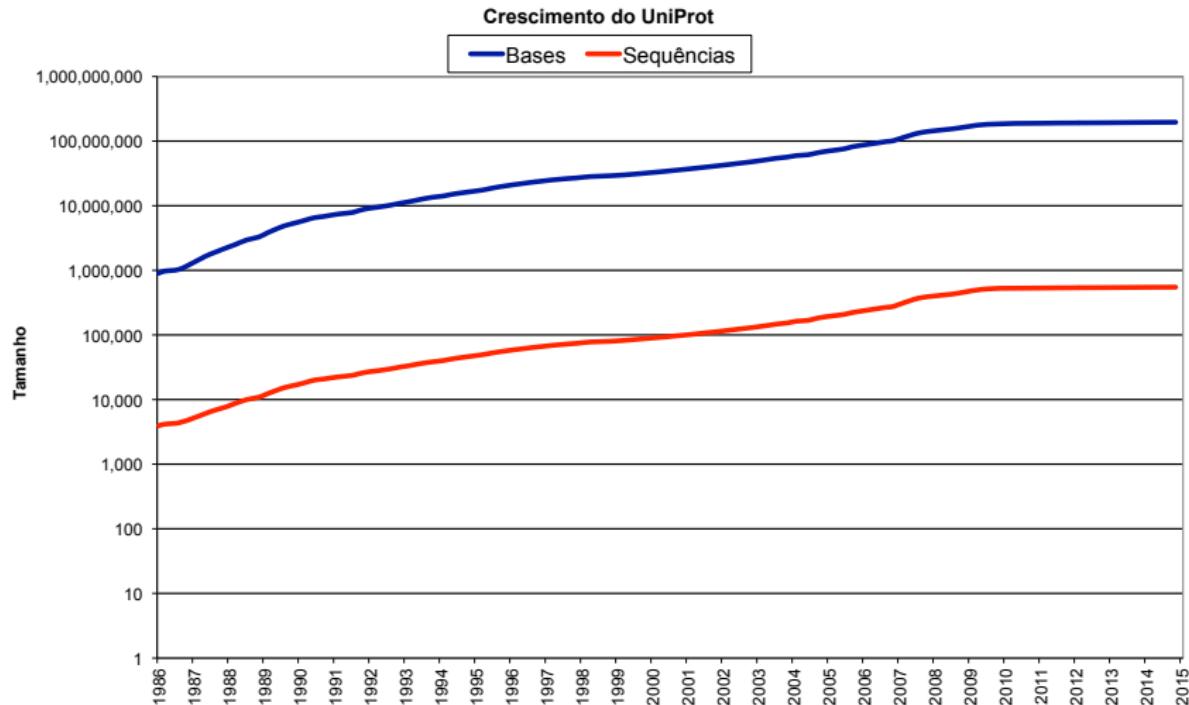




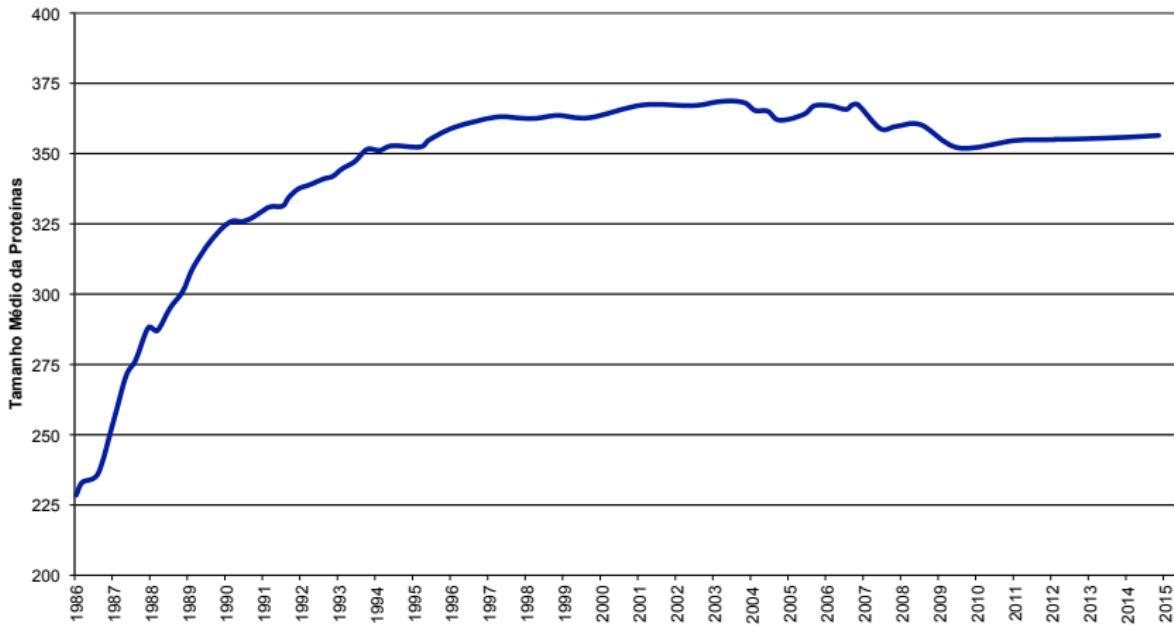
Tamanho Médio das Sequências do GenBank+WGS



- UniProt: Universal Protein Resource.
- Consórcio criado em 2002 envolvendo:
  - ▶ Swiss-Prot: Swiss Institute of Bioinformatics (SIB) e European Bioinformatics Institute (EBI). Maior banco manualmente curado de proteínas do mundo.
  - ▶ TrEMBL: Swiss Institute of Bioinformatics (SIB) e European Molecular Biology Laboratory (EMBL-EBI). Banco de proteínas gerado computacionalmente pela tradução dos dados do EMBL Nucleotide Sequence Database.
  - ▶ PIR: Georgetown University Medical Center (GUMC). Conjunto de banco de dados de proteínas criados para auxiliar a análise genômica e proteômica.



## Tamanho das Sequências do UniProt



# Alinhamento de uma Sequência contra todas as Sequências do GenBank

- Exemplo: computador de 3GHz (1 instrução por ciclo)

- ▶  $n = 10$ :

- ★ Tempo:  $(10 \times 10^{12}) / (3 \times 2^{30}) = 50$  segundos

- ▶  $n = 100$ :

- ★ Tempo:  $(100 \times 10^{12}) / (3 \times 2^{30}) = 9$  minutos

- ▶  $n = 1000$ :

- ★ Tempo:  $(1000 \times 10^{12}) / (3 \times 2^{30}) = 4$  dias

- ▶  $n = 10000$ :

- ★ Tempo:  $(10000 \times 10^{12}) / (3 \times 2^{30}) = 36$  dias

- ▶  $n = 100000$ :

- ★ Tempo:  $(100000 \times 10^{12}) / (3 \times 2^{30}) = 1$  ano

- ▶  $n = 1000000$ :

- ★ Tempo:  $(1000000 \times 10^{12}) / (3 \times 2^{30}) = 10$  anos

- ▶  $n = 10000000$ :

- ★ Tempo:  $(10000000 \times 10^{12}) / (3 \times 2^{30}) = 1$  século

# BLAST

- BLAST: Basic Local Alignment Search Tool.
- Ferramenta proposta por Stephen Altschul, Warren Gish, Webb Miller, Eugene Myers e David Lipman em 1990.
- Desenvolvido pelo National Center for Biotechnology Information (NCBI) e financiado pelo National Institutes of Health (NIH).
- Heurística para alinhamento local: não garante a obtenção do alinhamento local ótimo.
- Possui uma forte base estatística.
- Site oficial:
  - ▶ <http://blast.ncbi.nlm.nih.gov/>
- O artigo original do BLAST foi o artigo mais citado da década de 1990 e hoje em dia possui mais de 50000 citações.

- Nomenclatura:
  - ▶ *query*: sequência que será comparada.
  - ▶ *database*: banco de sequências.
  - ▶ *HSP*: high-scoring sequence pair, par de subsequências com alta similariedade.
  - ▶ *seed*: sequência curta utilizada para iniciar um alinhamento.
  - ▶ *hit*: alinhamento com similariedade maior que a mínima.
- Passos básicos:
  - ▶ Obter uma lista de *seeds*.
  - ▶ Procurar *hits* de *seeds* com sequências do banco de dados.
  - ▶ Estender os *hits* para obter os alinhamentos.

# BLAST

- Remover regiões de baixa complexidade da *query* (regiões com poucos tipos de elementos). Estas regiões são marcadas como subsequências de Xs (para sequências protéicas) e de Ns (para sequências de DNA).
- Construir uma lista com todas as sequências de tamanho  $w$  que possuam pontuação pelo menos  $T$  quando alinhadas com a *query*.
- Geralmente os parâmetros  $w$  e  $T$ , sob algum esquema de pontuação específico, são ajustados para se obter uma lista de sementes cerca de 50x maior que o tamanho da *query*.
- Em geral,  $w \geq 3$  para proteínas e  $w \geq 11$  para sequências de DNA.
- A escolha de uma matriz de pontuação adequada ( $PAM_{120}$ ,  $BLOSUM_{62}$ , etc) é fundamental nesta fase.

# BLAST - Criação da Lista de Seeds

```
V H R E M A A R T S P L R P L V A T A G P A L S P V P P C V H L T L R
```

# BLAST - Criação da Lista de Seeds

```
V H R E M A A R T S P L R P L V A T A G P A L S P V P P C V H L T L R  
V H R E
```

# BLAST - Criação da Lista de Seeds

```
V H R E M A A R T S P L R P L V A T A G P A L S P V P P C V H L T L R  
V H R E  
H R E M
```

# BLAST - Criação da Lista de Seeds

```
V H R E M A A R T S P L R P L V A T A G P A L S P V P P C V H L T L R  
V H R E E  
H R E E M  
R E M A
```

# BLAST - Criação da Lista de Seeds

```
V H R E M A A R T S P L R P L V A T A G P A L S P V P P C V H L T L R  
V H R E  
H R E M  
R E M A  
E M A A  
M A A R  
A A R T  
A R T S  
R T S P
```

# BLAST - Criação da Lista de Seeds

```
V H R E M A A R T S P L R P L V A T A G P A L S P V P P C V H L T L R  
V H R E           T S P L  
H R E M           S P L R  
R E M A           P L R P  
E M A A           L R P L  
M A A R           R P L V  
A A R T           P L V A  
A R T S           L V A T  
R T S P           V A T A
```

# BLAST - Criação da Lista de Seeds

V H R E M A A R T S P L R P L V A T A G P A L	<b>S P V P P C V H L T L R</b>	
V H R E	T S P L	A T A G
H R E M	S P L R	T A G P
R E M A	P L R P	A G P A
E M A A	L R P L	G P A L
M A A R	R P L V	P A L S
A A R T	P L V A	A L S P
A R T S	L V A T	L S P V
R T S P	V A T A	<b>S P V P</b>

# BLAST - Criação da Lista de Seeds

V	H	R	E	M	A	A	R	T	S	P	L	R	P	L	V	A	T	A	G	P	A	L	S	P	V	P	P	C	V	H	L	T	L	R									
V	H	R	E						T	S	P	L					A	T	A	G					P	V	P	P															
H	R	E	M						S	P	L	R					T	A	G	P					V	P	P	C															
R	E	M	A						P	L	R	P					A	G	P	A					P	P	C	V															
E	M	A	A						L	R	P	L					G	P	A					P	C	V	H																
M	A	A	R						R	P	L	V					P	A	L	S					C	V	H	L															
A	A	R	T						P	L	V	A					A	L	S	P					V	H	L	T															
A	R	T	S						L	V	A	T					L	S	P	V					H	L	T	L															
R	T	S	P						V	A	T	A					S	P	V	P					L	T	L	R															

# BLAST - Criação da Lista de Seeds

V	H	R	E	M	A	A	R	T	S	P	L	R	P	L	V	A	T	A	G	P	A	L	S	P	V	P	P	C	V	H	L	T	L	R				
V	H	R	E						T	S	P	L						A	T	A	G					P	V	P	P									
H	R	E	M						S	P	L	R						T	A	G	P					V	P	P	C									
R	E	M	A						P	L	R	P						A	G	P	A					P	P	C	V									
E	M	A	A						L	R	P	L						G	P	A					P	C	V	H										
M	A	A	R						R	P	L	V						P	A	L					C	V	H	L										
A	A	R	T						P	L	V	A						A	L	S	P					V	H	L	T									
A	R	T	S						L	V	A	T						L	S	P	V					H	L	T	L									
R	T	S	P						V	A	T	A						S	P	V	P					L	T	L	R									

# BLAST - Criação da Lista de Seeds

V H R E M A A R T S P L R P L V A T A G P A L S P V P P C V H L T L R

# BLAST - Criação da Lista de Seeds

V H R E M A A R T S P L R P L V A T A G P A L S P V P P C V H L T L R  
E M A A = 18

# BLAST - Criação da Lista de Seeds

V H R E M A A R T S P L R P L V A T A G P A L S P V P P C V H L T L R

E M A A = 18

A A A C = 6

# BLAST - Criação da Lista de Seeds

V H R E M A A R T S P L R P L V A T A G P A L S P V P P C V H L T L R

E M A A = 18

A A A C = 6

A A A D = 3

# BLAST - Criação da Lista de Seeds

V	H	R	E	M	A	A	R	T	S	P	L	R	P	L	V	A	T	A	G	P	A	L	S	P	V	P	P	C	V	H	L	T	L	R
E	M	A	A	=	18																													
A	A	A	C	=	6																													
A	A	A	D	=	3																													
•	•	•																																
E	H	A	I	=	14																													
E	H	A	K	=	9																													
•	•	•																																
Y	Y	Y	Y	=	-15																													

# BLAST - Criação da Lista de Seeds

V H R E M A A R T S P L R P L V A T A G P A L S P V P P C V H L T L R

E M A A = 18

A A A C = 6

A A A D = 3

• • •

E H A I = 14

E H A K = 9

• • •

Y Y Y Y = -15

Seeds:  
 $(T \geq 11)$

# BLAST - Criação da Lista de Seeds

V H R E M A A R T S P L R P L V A T A G P A L S P V P P C V H L T L R	
E M A A = 18	Seeds: (T ≥ 11)
A A A C = 6	
A A A D = 3	
• • •	• • •
E H A I = 14	E M A A
E H A K = 9	E H A I
• • •	• • •
Y Y Y Y = -15	

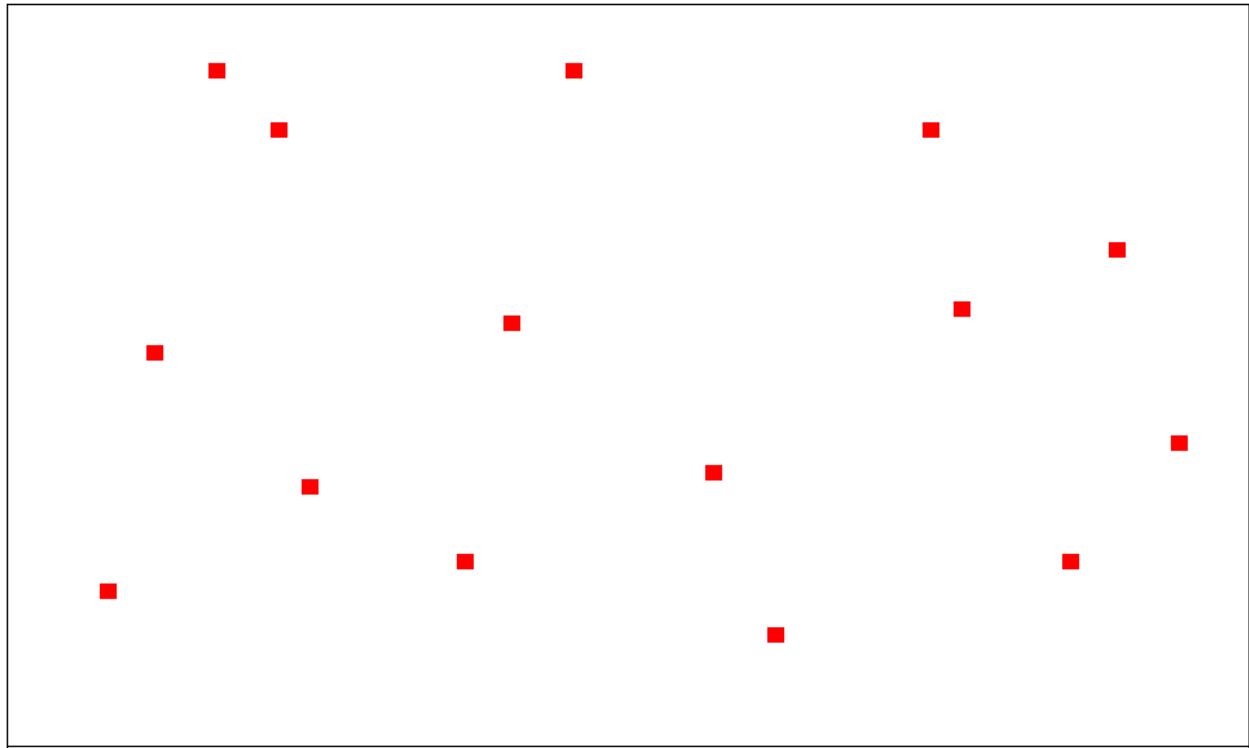
# BLAST - Busca por *Hits*

- Duas opções de busca:
  - ▶ Para cada *seed*, buscar *hits* em cada uma das sequências do banco de dados.
  - ▶ Para cada sequência do banco de dados, buscar *hits* com cada um dos *seeds*.
- Duas opções de estrutura de dados para auxiliar a busca:
  - ▶ Construção de um vetor, onde cada posição representa uma sequência protéica de tamanho  $w$ . A  $i$ -ésima posição do vetor armazena uma lista com todas as ocorrências da  $i$ -ésima sequência na *query*.
    - ★ Poucas posições deste vetor armazenam informações úteis.
    - ★ Alternativa: armazenar as informações num *hash*.
  - ▶ Construção de uma máquina de estados, usando autômatos finitos, onde cada estado representa a última palavra lida, e as transições de estados ocorrem a cada leitura de uma nova base da sequência do banco onde se está buscado por *hits*.
- Geralmente usa-se autômatos finitos para buscar todos os *seeds* em cada uma das sequências do banco (uma por uma).

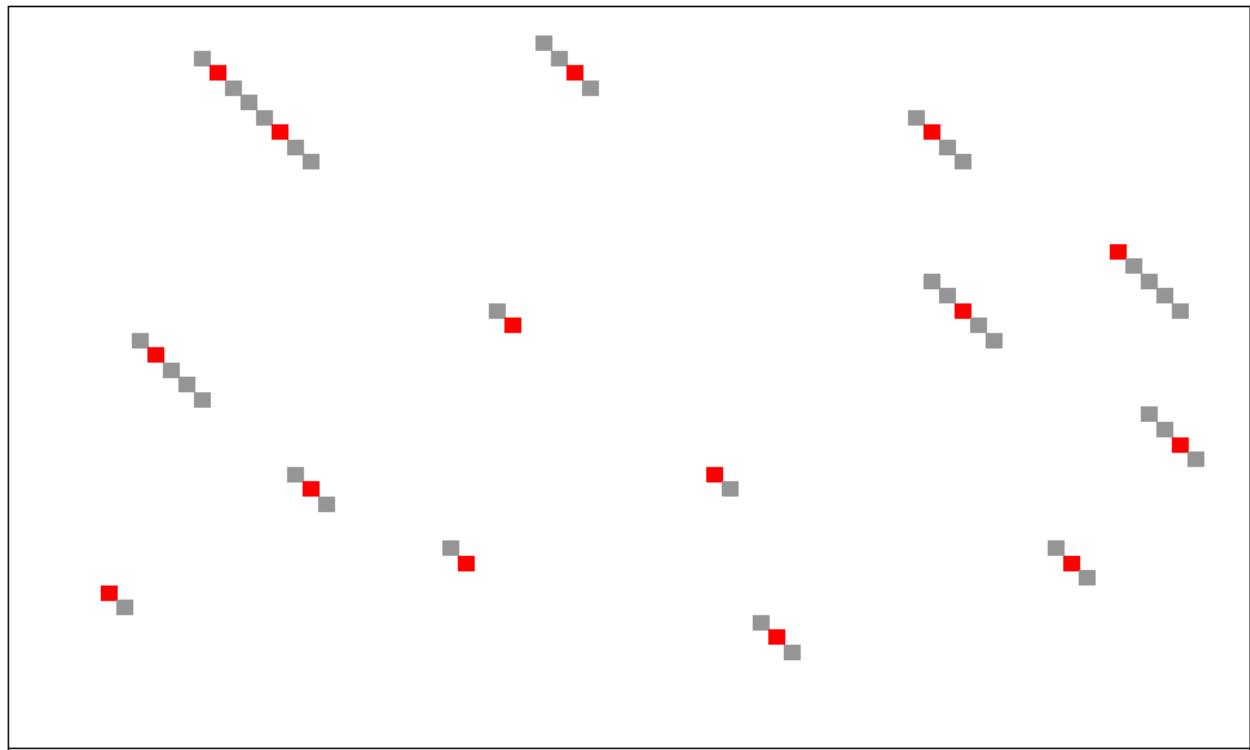
## BLAST - Obtenção dos *HSPs*

- Estende-se o *hit* em ambas as direções, apenas considerando alinhamento sem buracos.
- A extensão é interrompida após se distanciar muito do melhor alinhamento obtido até então.
- Por exemplo, para proteínas, o valor da distância máxima é 20. Este valor garante que a probabilidade deste método perder um alinhamento de maior pontuação é de cerca de 0,1%.
- Apenas *HSPs* com pontuação maior ou igual a um limiar  $S$  são apresentados como respostas.
- Estima-se que 90% do tempo de processamento é gasto nesta etapa.
- A performance do algoritmo nesta fase está intimamente relacionada a escolha dos parâmetros  $w$  e  $T$ .
  - ▶ Quanto maior for o valor de  $w$ , maior o número de *seeds* a se considerar.
  - ▶ Quanto maior for o valor de  $T$ , mais restrita será a busca por *HSPs*.

# BLAST - Obtenção dos *HSPs*



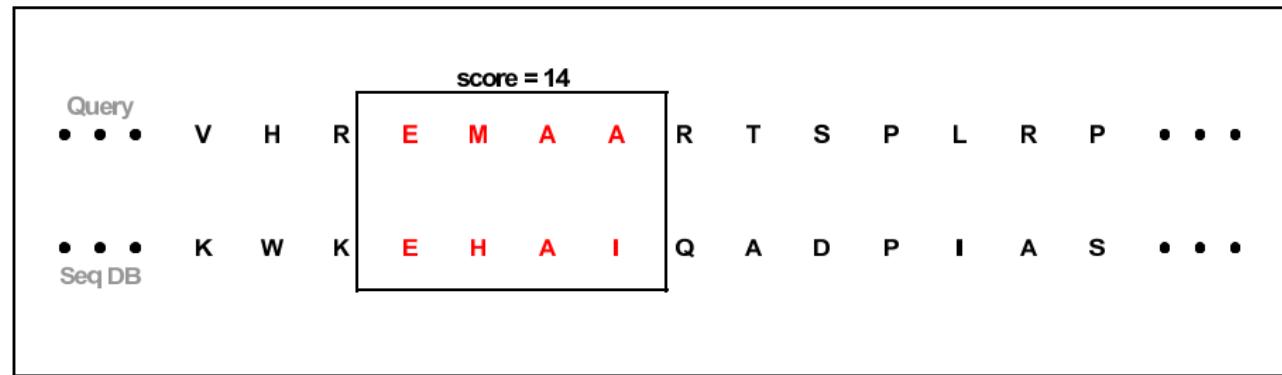
## BLAST - Obtenção dos *HSPs*



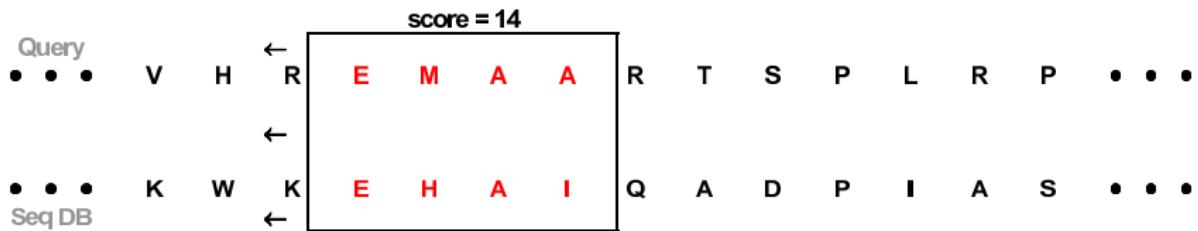
# BLAST - Obtenção dos *HSPs*

Query	• • • V H R E M A A R T S P L R P • • •
Seq DB	• • • K W K E H A I Q A D P I A S • • •

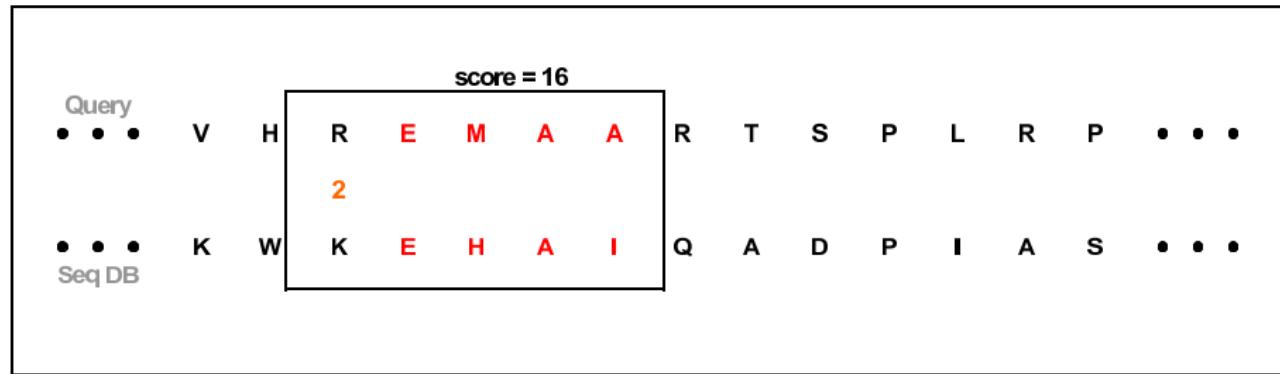
# BLAST - Obtenção dos *HSPs*



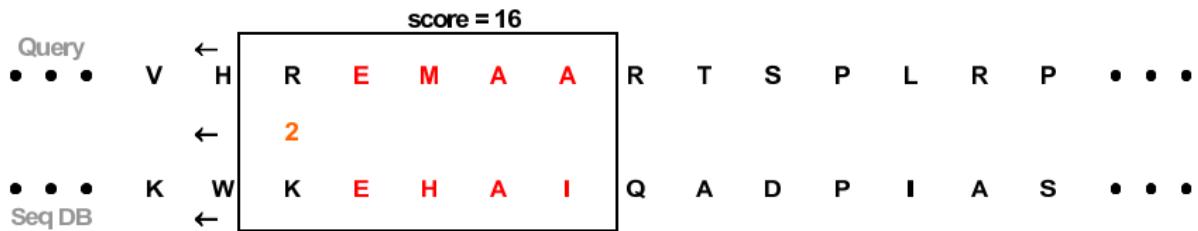
# BLAST - Obtenção dos *HSPs*



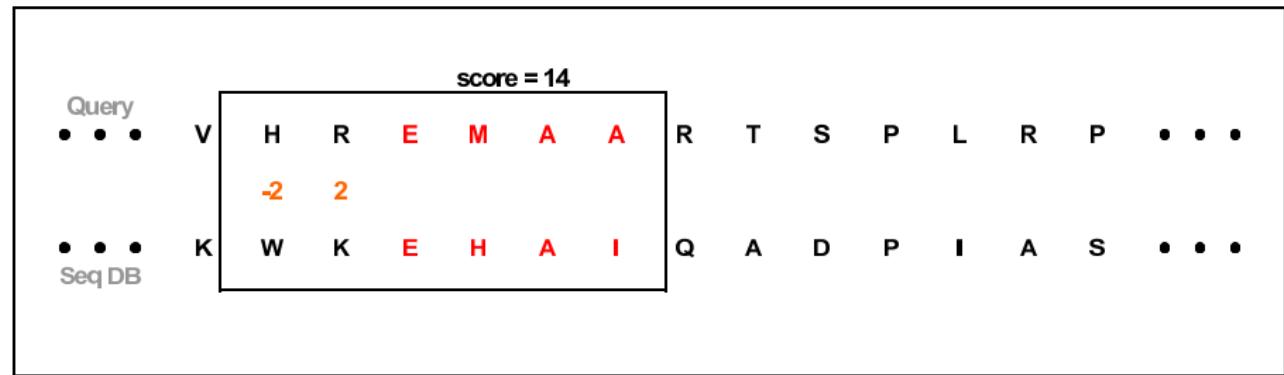
# BLAST - Obtenção dos *HSPs*



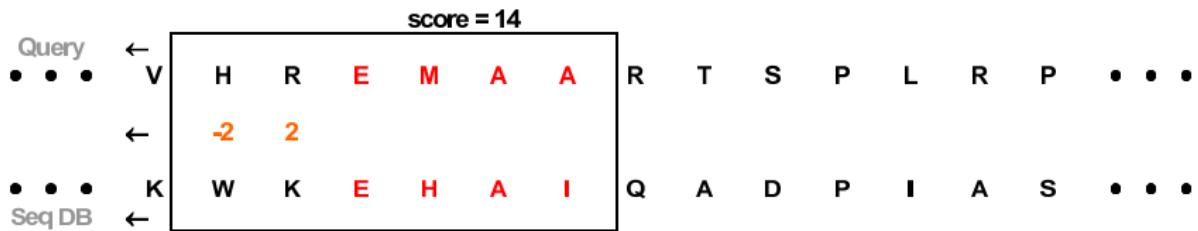
# BLAST - Obtenção dos *HSPs*



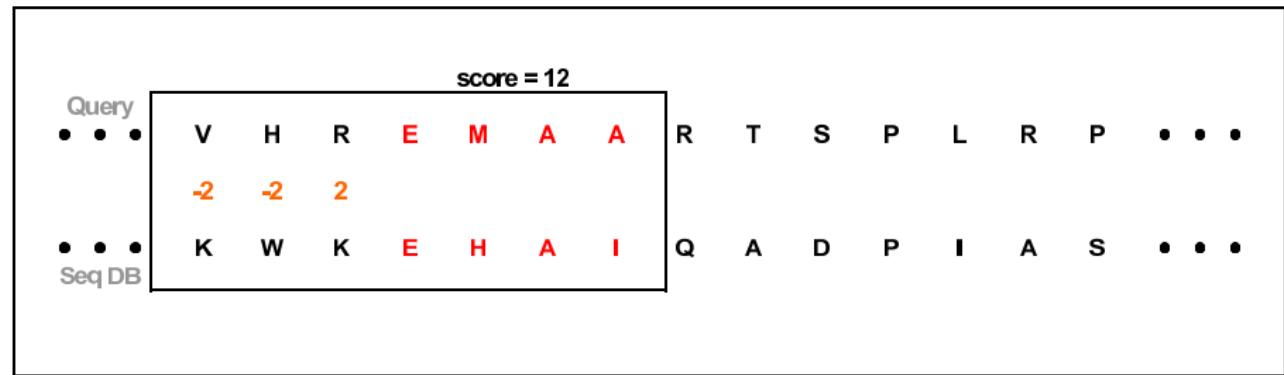
# BLAST - Obtenção dos *HSPs*



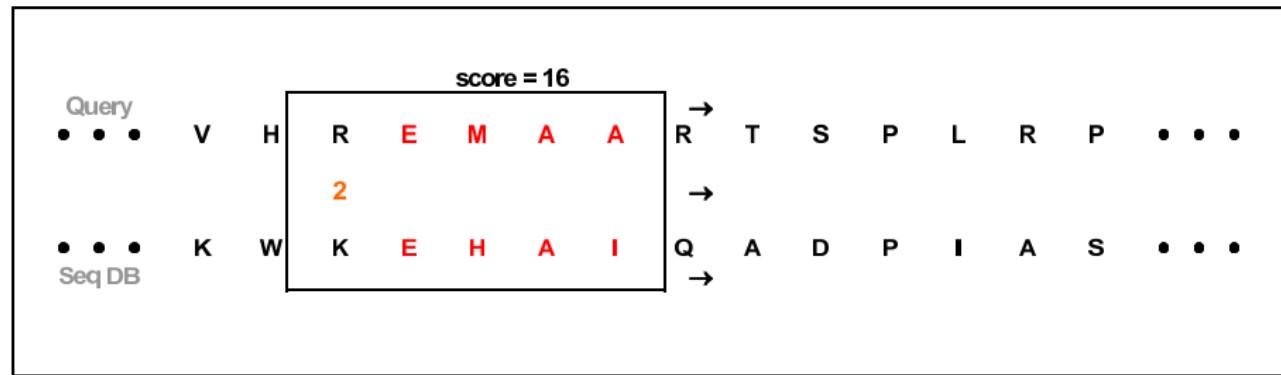
# BLAST - Obtenção dos *HSPs*



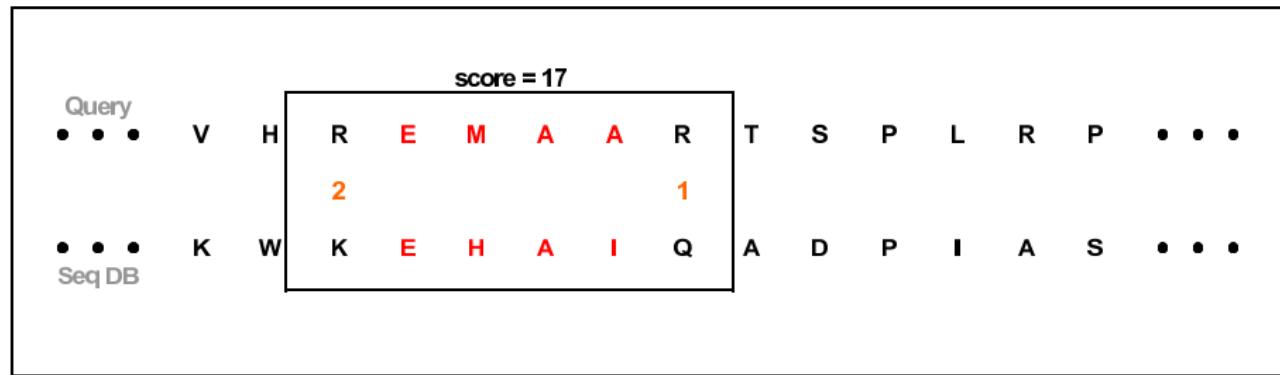
# BLAST - Obtenção dos *HSPs*



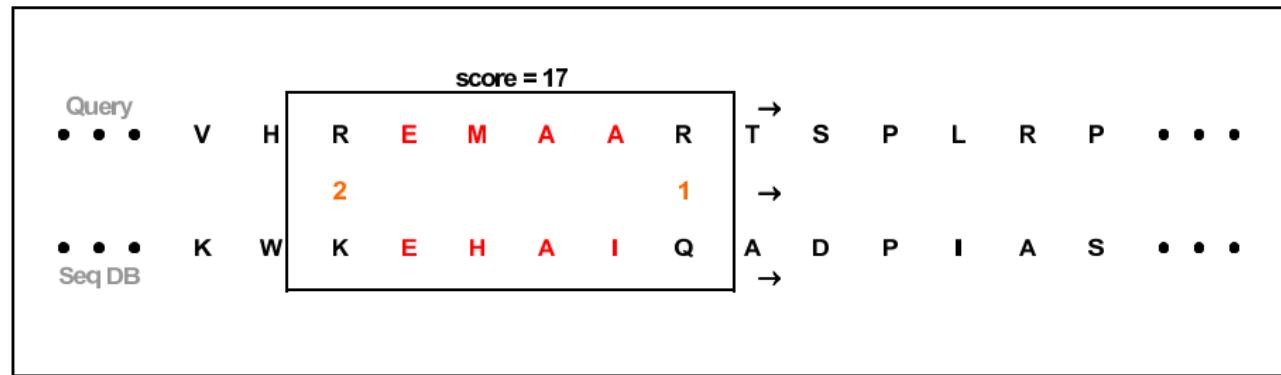
# BLAST - Obtenção dos *HSPs*



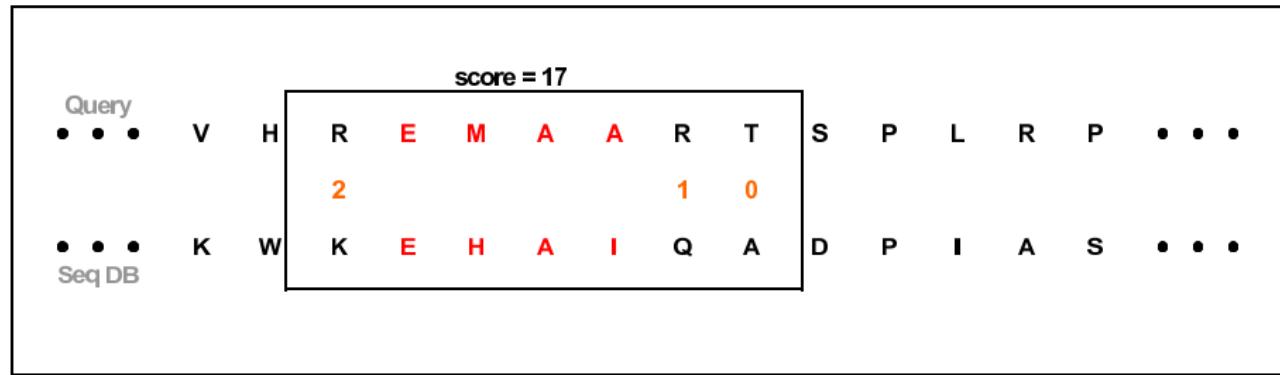
# BLAST - Obtenção dos *HSPs*



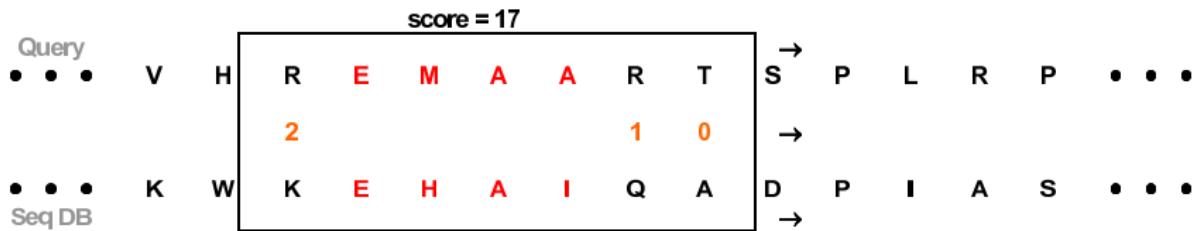
# BLAST - Obtenção dos *HSPs*



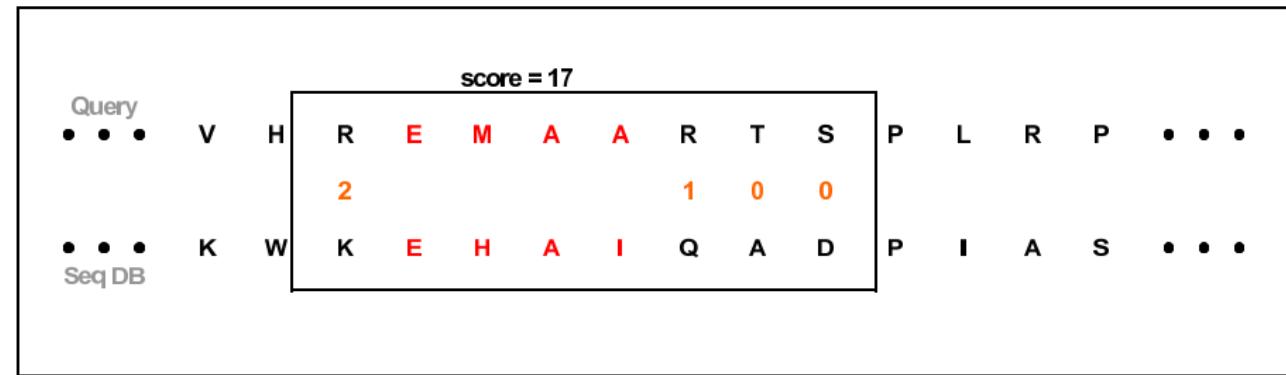
# BLAST - Obtenção dos *HSPs*



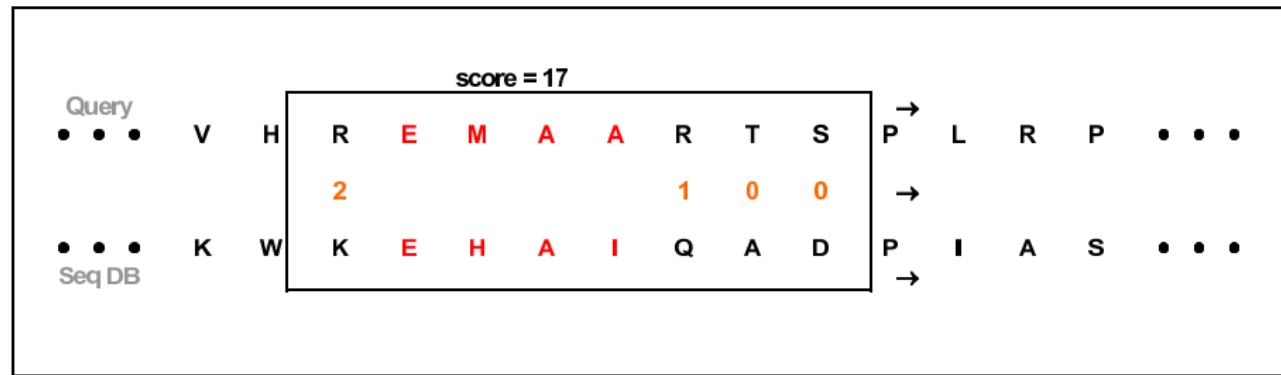
# BLAST - Obtenção dos *HSPs*



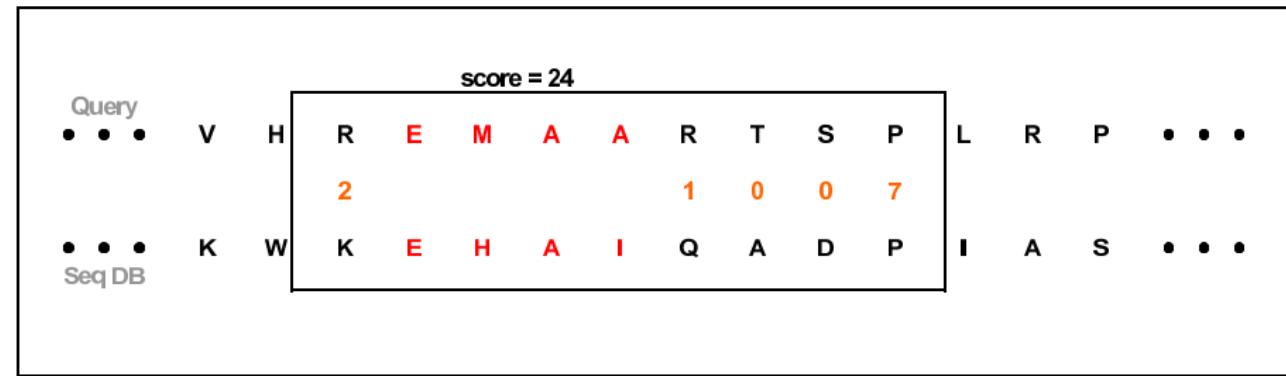
# BLAST - Obtenção dos *HSPs*



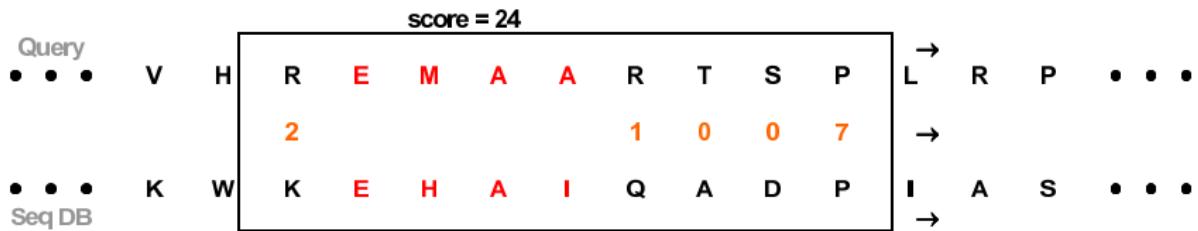
# BLAST - Obtenção dos *HSPs*



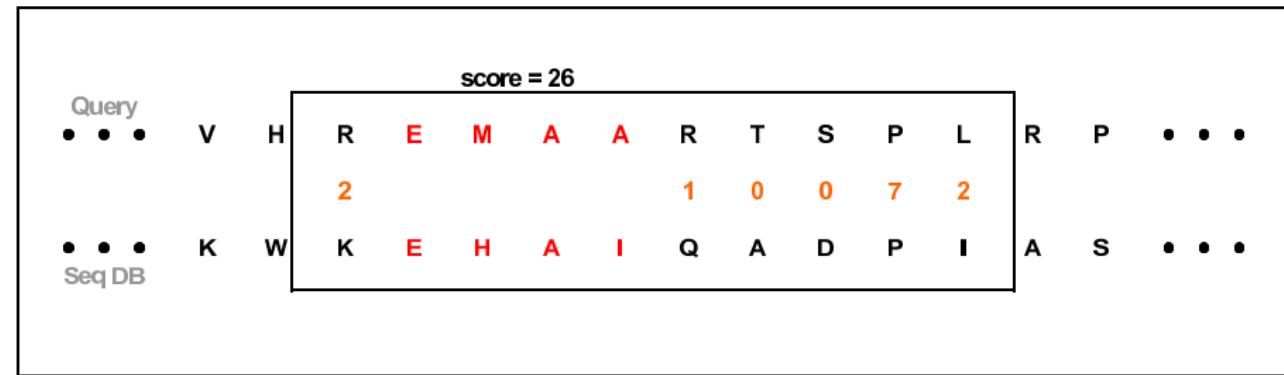
# BLAST - Obtenção dos *HSPs*



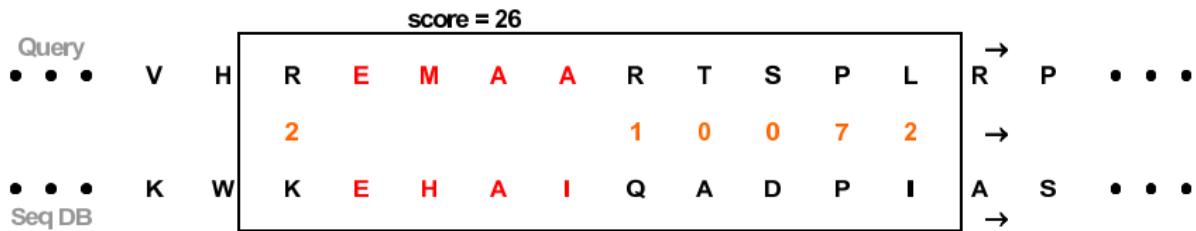
# BLAST - Obtenção dos *HSPs*



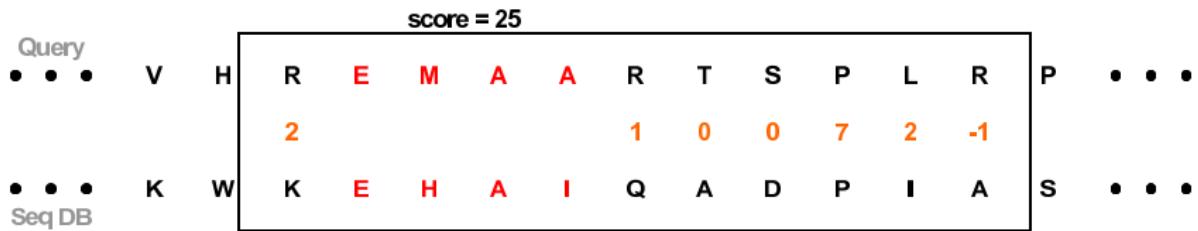
# BLAST - Obtenção dos *HSPs*



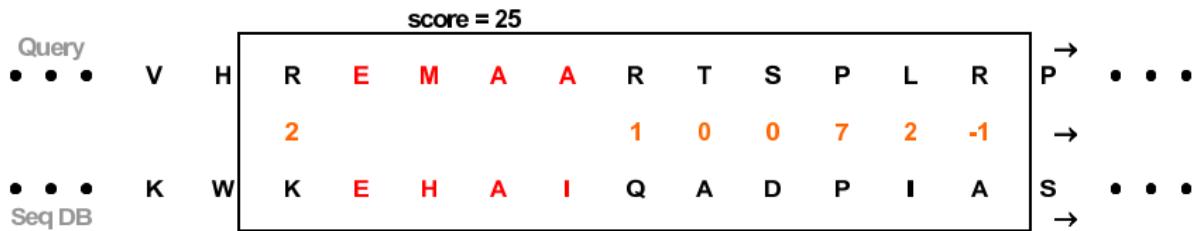
# BLAST - Obtenção dos *HSPs*



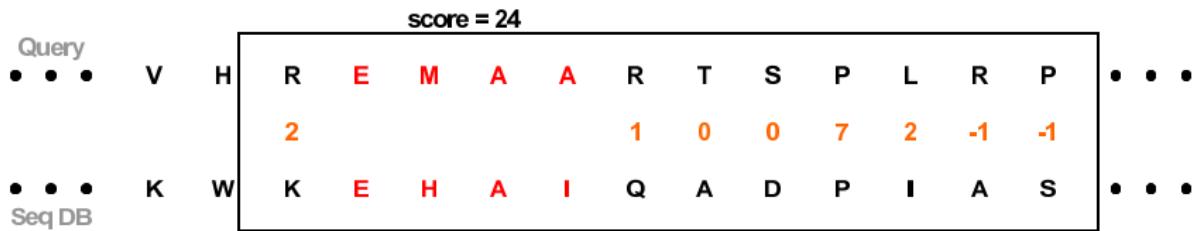
# BLAST - Obtenção dos *HSPs*



# BLAST - Obtenção dos HSPs



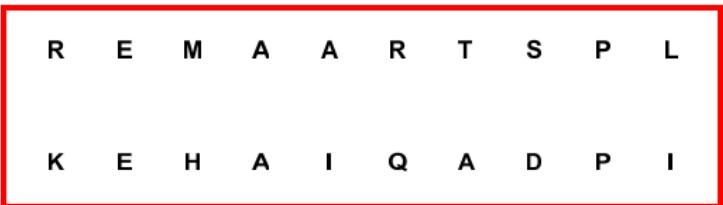
# BLAST - Obtenção dos *HSPs*



# BLAST - Obtenção dos *HSPs*

score = 26 HSP

Query • • • V H R E M A A R T S P L R P • • •	<b>score = 26</b>	Seq DB • • • K W K E H A I Q A D P I A S • • •
--	-------------------	---



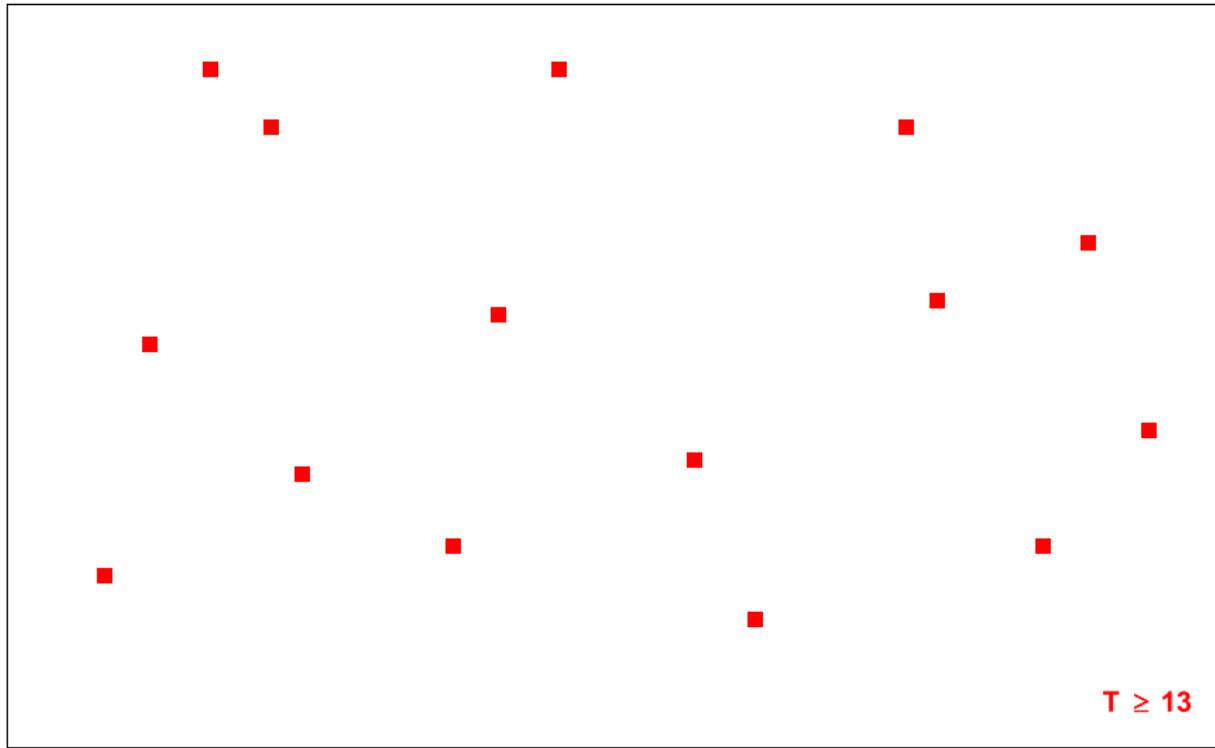
# BLAST 2.0

- Extensão apresentada por Stephen Altschul, Thomas Madden, Alejandro Schaeffer, Jinghui Zhang, Zheng Zhang, Webb Miller e David Lipman em 1997.
- Duas principais inovações:
  - ▶ The Two-Hit Method
  - ▶ Gapped BLAST

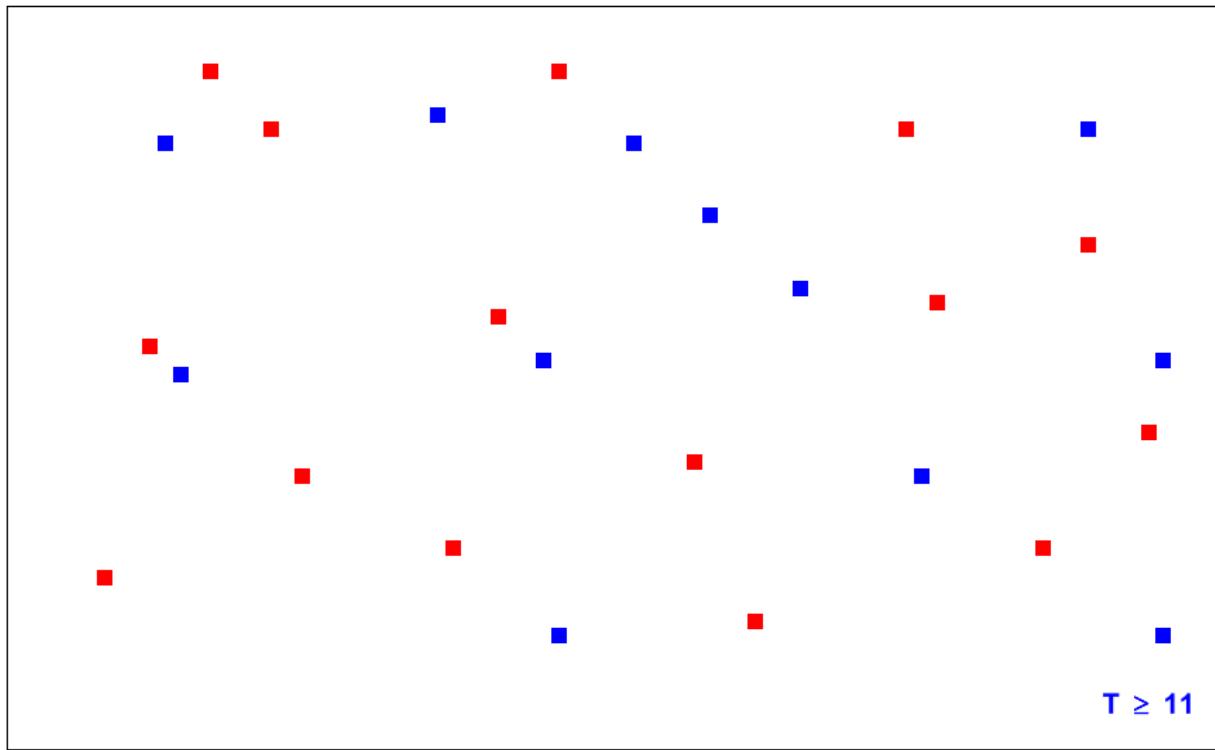
## BLAST 2.0 - The Two-Hit Method

- Objetivo: acelerar o tempo de processamento do algoritmo original.
- Reduz o número de extensões.
- Observação: *HSPs* são muito maiores que  $w$ .
- Um *HSP* frequentemente contém dois ou mais *hits*.
- Apenas procurar um *HSP* se existirem dois *hits* na mesma diagonal.
- Como implementar:
  - ▶ Se os *hits* se sobrepõem, ignorar.
  - ▶ Se os *hits* estiverem a uma distância menor do que um certo valor  $A$ , estender.
- O valor de  $T$  deve ser reduzido para se manter a mesma sensibilidade.

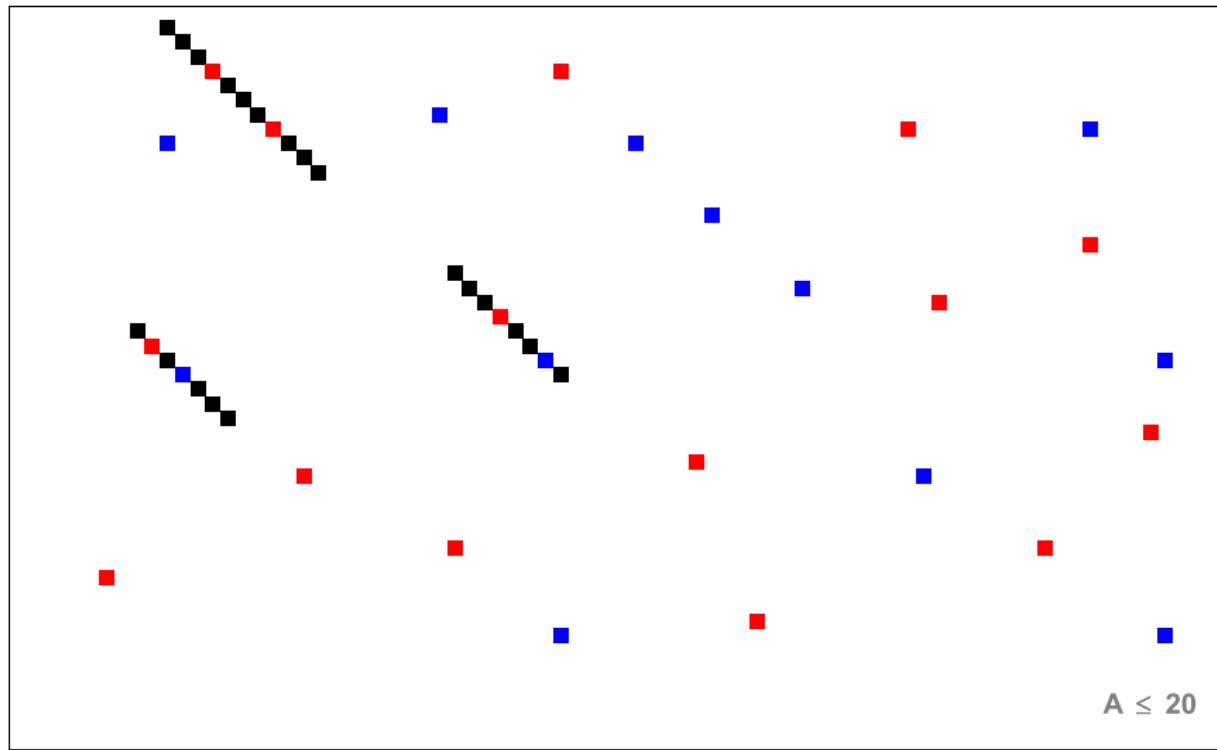
# BLAST 2.0 - The Two-Hit Method



# BLAST 2.0 - The Two-Hit Method



# BLAST 2.0 - The Two-Hit Method



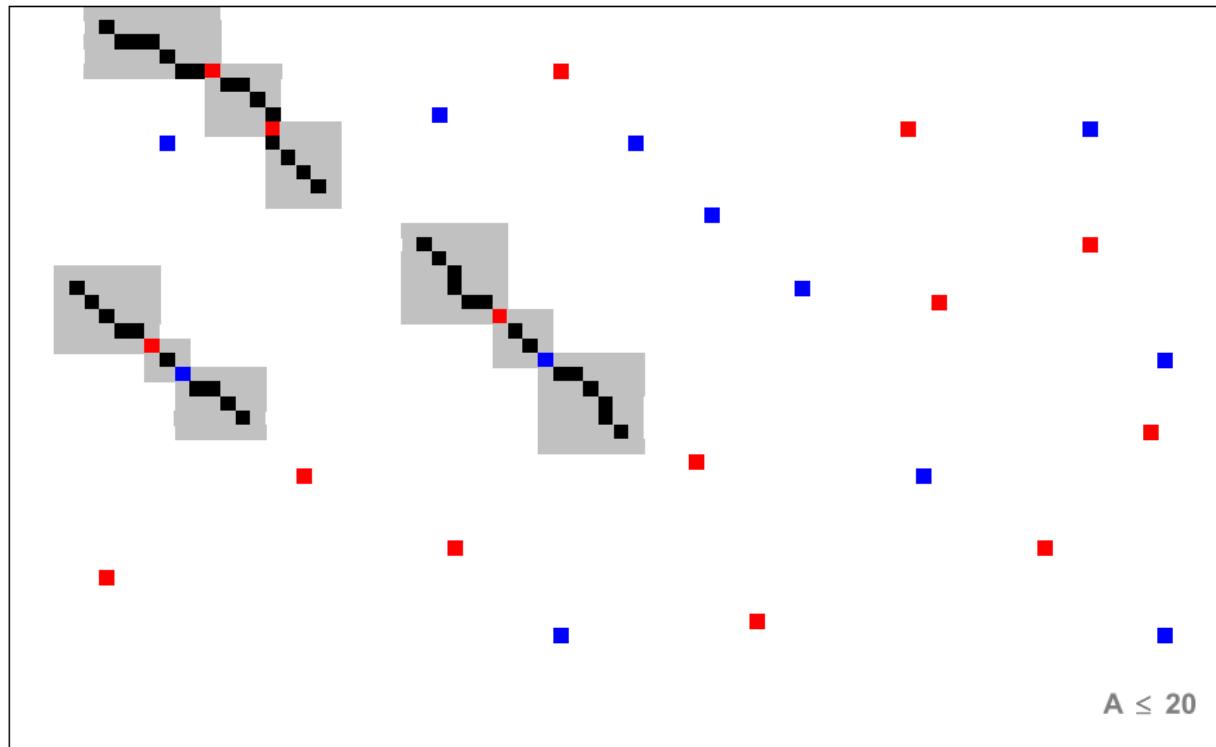
## BLAST 2.0 - The Two-Hit Method

- Métodos para extensão de hits (valores padrão para proteínas):
  - ▶ *One-Hit*:  $w = 3$  e  $T = 13$ .
  - ▶ *Two-Hits*:  $w = 3$  e  $T = 11$ .
- Comparação entre os métodos:
  - ▶ *Two-Hits* gera aproximadamente 3.2x mais *hits*.
  - ▶ *Two-Hits* faz aproximadamente 86% menos extensões.

## BLAST 2.0 - Gapped BLAST

- Extensões de *hits* não são mais limitados a diagonais da matriz de Programação Dinâmica, permitindo alinhamento com buracos.
- A extensão é interrompida quando o alinhamento cai abaixo de um limiar pré-estabelecido ( $X_G$ ).
- Se a pontuação do *HSP* for maior que um parâmetro  $S$ , então o HSP é apresentado na lista de respostas.
- O BLAST 2.0 é cerca 3x mais rápido do que a versão original.

# BLAST 2.0 - Gapped BLAST



# Complexidade

- Seja  $w$  o tamanho dos *seeds*,  $\mathcal{A}$  o alfabeto ( $|\mathcal{A}| = 4$  para DNA,  $|\mathcal{A}| = 20$  para proteínas),  $M$  o tamanho do banco de dados (número total de bases),  $h$  o número de *hits* encontrados e  $n$  o tamanho da *query*.
  - ▶  $O(nw|\mathcal{A}|^w)$
- Fase 1: Obter uma lista de *seeds*:
  - ▶  $O(M)$
- Fase 2: Procurar *hits* de *seeds* nas sequências do banco de dados:
  - ▶  $O(hn^2)$
- Fase 3: Estender os *hits* para obter os alinhamentos (*HSPs*):
  - ▶  $O(nw|\mathcal{A}|^w + M + hn^2)$
- Total:
  - ▶  $O(nw|\mathcal{A}|^w + M + hn^2)$

## Bit Score e E-Value

- O BLAST indica para cada HSP retornado um *bit score* e um *E-value*.
- O *bit score* representa a pontuação normalizada do alinhamento, e é dada pela fórmula:

$$S' = \frac{\lambda S - \ln K}{\ln 2}$$

- O *E-value* representa o número esperado de HSPs com score maior ou igual a  $S$ , e é dado pela fórmula:

$$E = \frac{KMn}{e^{\lambda S}}$$

- As constantes  $\lambda$  e  $K$  são calculadas considerando a matriz de pontuação e a distribuição das bases no banco de dados.

## E-Value x P-Value

- Não confundir E-value com P-value:
  - ▶ E-Value: número esperado de alinhamentos aleatórios com pontuação maior ou igual a  $S$ .
  - ▶ P-Value: probabilidade de se obter aleatoriamente um alinhamento com uma pontuação maior ou igual a  $S$ .
- E-value pode ser um número maior que 1, enquanto o P-value é sempre um valor entre 0 e 1.
- Dado um E-value é possível obter o P-value (e vice-versa) através da seguinte fórmula:

$$P = 1 - e^{-E}$$

# Família BLAST

- O BLAST é composto por uma família de programas, todos acessíveis através do executável *blastall*:
  - ▶ blastn: [*query*: DNA] x [*database*: DNA].
  - ▶ blastp: [*query*: proteína] x [*database*: proteína].
  - ▶ blastx: [*query*: DNA] x [*database*: proteína] (nos 6 frames da *query*).
  - ▶ tblastx: [*query*: DNA] x [*database*: DNA] (nos 6 frames da *query* e de cada sequência do banco de dados).
  - ▶ tblastn: [*query*: proteína] x [*database*: DNA] (nos 6 frames de cada sequência do banco de dados).
  - ▶ megablast: ideal para comparar várias sequências contra um banco de sequências. Concatena todas as sequências de entrada em uma única, e depois faz um pós-processamento para obter os alinhamentos corretos.