Alinhamento de Sequências

Zanoni Dias

Instituto de Computação – Unicamp

3 de maio de 2009

O que é Alinhamento de Sequências?

- Um alinhamento de duas sequências de caracteres α e β 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.
- Exemplo:
  - Sequências:
    - α = AAACTGCACAATCTTAATGCCCTTTTAT
    - β = GCGGATCAACTTATTCCATCTCTT
  - Alinhamento:
    - α' = AAACTGCA-CAATCTTAATGCC--CTTTTAT
    - β' = --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:
  - α' = CAGGGTGCTAGCAAAAACCATCGCGGGCGATAA
  - β' = --GGCATGTAGCACACACGACGCTGGGAGAAT-
- Alinhamento 2:
  - α'' = CAGGATGCTAGCAAAAACCATCGC-GGGCGATAA-
  - β'' = -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 gaps + 17 matches + 13 mismatches = -21
  - Alinhamento 2: 7 gaps + 22 matches + 6 mismatches = 13
- Exemplo 2: gap = -10, match = 1, mismatch = -1
  - Alinhamento 1: 3 gaps + 17 matches + 13 mismatches = -26
  - Alinhamento 2: 7 gaps + 22 matches + 6 mismatches = -54

Alinhamento Ótimo

Problema

Dadas duas sequências α e β 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 altamente 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.

Alinhamento Global
Força Bruta - 1ª Abordagem

- Onde a primeira base pode se alinhar?
  - Opção 1:
    - `CCAGCCGAATCGATCGCATG`
    - `-CATCAGCGATCGATCTT...`
  - Opção 2:
    - `CCAGCCGAATCGATCGCATG`
    - `CATCAGCGATCGATCTT...`
  - Opção 3:
    - `-CCAGCCGAATCGATCGCATG`
    - `CATCAGCGATCGATCTT...`
  - ...
  - Opção n+1:
    - `-----------------CCAGCCGAATCGATCGCATG`
    - `CATCAGCGATCGATCTTT...`

Alinhamento Global
Força Bruta - 1ª Abordagem

Algoritmo 1: Align

Input: $\alpha, a_i, a_j, \beta, b_i, b_j$
if $a_i > a_j$ **then**
  return $gap * (b_j - b_i + 1)$
else
  max $\leftarrow$ gap + Align($\alpha, a_{i+1}, a_j, \beta, b_i, b_j$);
  for *all* $b_k \in [b_i..b_j]$ **do**
    aux $\leftarrow$ $gap.(b_k - b_i) + \sigma(\alpha[a_i], \beta[b_k])$ + Align($\alpha, a_{i+1}, a_j, \beta, b_{k+1}, b_j$);
    if $aux > max$ **then**
      max $\leftarrow$ aux
    end
  end
  return max
end

Força Bruta - 1ª Abordagem
Complexidade e Tempo de Execução

- Complexidade:
  - $T(m,n) = \sum_{i=0}^{n} T(m-1, n-i) + \Theta(n)$
  - $T(m,n) \geq \sum_{i=0}^{n-1}((n-i)(i+1)^m) \geq n^m/4$
  - $T(m,n) = \Omega(n^m)$
- Exemplo: computador de 3GHz (1 instrução por ciclo)
  - m = n = 05:
    - Tempo: $5^5/(3*2^{30}) = 1$ milionésimo de segundo
  - m = n = 10:
    - Tempo: $10^{10}/(3*2^{30}) = 3$ segundos
  - m = n = 15:
    - Tempo: $15^{15}/(3*2^{30}) = 1500$ dias
  - m = n = 20:
    - Tempo: $20^{20}/(3*2^{30}) = 1$ bilhão de anos
  - m = n = 25:
    - Tempo: $25^{25}/(3*2^{30}) = 874$ quatrilhões de anos

Alinhamento Global
Força Bruta - 2ª Abordagem

- Como as últimas bases das sequências se alinham?
  - Opção 1:
    - `CCAGCCGAATCGATCGCATG`
    - `...CATCAGCGATCGATCT`
  - Opção 2:
    - `CCAGCCGAATCGATCGCATG`
    - `...CATCAGCGATCGATCT-`
  - Opção 3:
    - `CCAGCCGAATCGATCGCATG-`
    - `...CATCAGCGATCGATCT`

Alinhamento GLobal
Força Bruta - 2ª Abordagem

Algoritmo 2: Align

Input: α, m, β, n
if $m = 0$ **then**
  return $gap * n$
end
if $n = 0$ **then**
  return $gap * m$
end

$$\textbf{return } max \left\{ \begin{array}{l} Align(\alpha, m-1, \beta, n-1) + \sigma(\alpha[m], \beta[n]), \\ Align(\alpha, m, \beta, n-1) + gap, \\ Align(\alpha, m-1, \beta, n) + gap \end{array} \right\}$$

Força Bruta - 2ª Abordagem
Complexidade e Tempo de Execução

- Complexidade:
  - $T(m,n) = T(m-1, n-1) + T(m, n-1) + T(m-1, n) + \Theta(1)$
  - $T(m,n) \geq 3T(m-1, n-1) + \Theta(1)$
  - $T(m,n) = \Omega(3^{min(n,m)})$
- Exemplo: computador de 3GHz (1 instrução por ciclo)
  - m = n = 10:
    - Tempo: $(3^{10})/(3*2^{30}) = 18$ milionésimos de segundo
  - m = n = 20:
    - Tempo: $(3^{20})/(3*2^{30}) = 1$ segundo
  - m = n = 30:
    - Tempo: $(3^{30})/(3*2^{30}) = 18$ horas
  - m = n = 40:
    - Tempo: $(3^{40})/(3*2^{30}) = 120$ anos
  - m = n = 50:
    - Tempo: $(3^{50})/(3*2^{30}) = 7$ milhões de anos
  - m = n = 60:
    - Tempo: $(3^{60})/(3*2^{30}) = 417$ bilhões de anos

Contagem de Subproblemas

- Existem quantos subproblemas distintos, envolvendo alinhamentos de prefixos não vazios de α e de β?
  - Apenas $m.n$!
- Memorização (*memoization*): evita o recálculo de subproblemas, armazenando os valores previamente calculados numa matriz.

Memorização

Algoritmo 3: Memoization

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

Algoritmo 4: AlignMemoization

Input: α, m, β, n
if $m = 0$ **then return** $gap * n$
if $n = 0$ **then return** $gap * m$
if $M[n-1, m-1] = \infty$ **then** $M[n-1, m-1] \leftarrow$ AlignMemoization$(\alpha, m-1, \beta, n-1)$
if $M[n-1, m] = \infty$ **then** $M[n-1, m] \leftarrow$ AlignMemoization$(\alpha, m, \beta, n-1)$
if $M[n, m-1] = \infty$ **then** $M[n, m-1] \leftarrow$ AlignMemoization$(\alpha, m-1, \beta, n)$

$$\textbf{return } max \left\{ \begin{array}{l} M[n-1, m-1] + \sigma(\alpha[m], \beta[n]), \\ M[n, m-1] + gap, \\ M[n-1, m] + gap \end{array} \right\}$$

Programação Dinâmica

- Idéias básicas:
  - Matriz M armazena os valores dos alinhamentos ótimos entre todos prefixos de α e todos os prefixos de β.
  - O valor do alinhamento ótimo entre as duas sequências estaria 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 α ou de β com a sequência vazia.
- Em 1970, Saul Needleman e Christian Wunsch propuseram o algoritmo de programação dinâmica para alinhamento de sequências.

Alinhamento Global

Algoritmo 5: Global

Input: α, m, β, n
for *all* $i \in [0..n]$ **do** $M[i, 0] \leftarrow gap * i$
for *all* $j \in [1..m]$ **do** $M[0, j] \leftarrow gap * j$
for *all* $i \in [1..n]$ **do**
  for *all* $j \in [1..m]$ **do**

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

  end
end
return $M[n, m]$

Complexidade e Tempo de Execução - Alinhamento Global

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

Alinhamento Global

×	α	A	C	T	G	G	G	T	C	A	A	C
β	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

×	α	A	C	T	G	G	G	T	C	A	A	C
β	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55
A	-5	3										
T	-10											
T	-15											
G	-20											
G	-25											
C	-30											
C	-35											
A	-40											
C	-45											

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

Alinhamento Global

×	α	A	C	T	G	G	G	T	C	A	A	C
β	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55
A	-5	3	-2									
T	-10											
T	-15											
G	-20											
G	-25											
C	-30											
C	-35											
A	-40											
C	-45											

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

Alinhamento Global

×	α	A	C	T	G	G	G	T	C	A	A	C
β	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55
A	-5	3	-2	-7								
T	-10											
T	-15											
G	-20											
G	-25											
C	-30											
C	-35											
A	-40											
C	-45											

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

Alinhamento Global

×	α	A	C	T	G	G	G	T	C	A	A	C
β	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55
A	-5	3	-2	-7	-12	-17	-22	-27	-32	-37	-42	-47
T	-10											
T	-15											
G	-20											
G	-25											
C	-30											
C	-35											
A	-40											
C	-45											

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

Alinhamento Global

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

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

Alinhamento Global

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

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

Alinhamento Global

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

Alinhamento Global

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

Alinhamento Global

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

Alinhamento Global

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

Alinhamento Global

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

Alinhamento Global

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

Alinhamento Global

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

Alinhamento Global

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

Alinhamento Global

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

Alinhamento Global

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

Alinhamento Global

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

Alinhamento Global

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

Exercícios

Exercício

Escreva um algoritmo, em pseudcódigo, que dadas duas sequências α e β, com, respectivamente, m e n caracteres, e uma matriz de pontuação de alinhamentos M, entre todos os pares de prefixos de α e de β, retorne um alinhamento ótimo entre α e β.

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.

Exercício

Suponha que uma inserção e uma remoção têm o mesmo custo, e que uma substituição pode ser realizada por uma remoção seguida de uma inserção, descreva como resolver eficientemente o problema da distância de edição.

Alinhamento Semi-Glogal

- Como alterar o algoritmo de Needleman-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 início no alinhamento ótimo em todas as posições da última linha ou da última coluna da matriz.

Alinhamento Semi-Global

Algoritmo 6: Semi-Glocal

Input: α, m, β, 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**

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

  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

×	α	A	T	C	T	T	C	G	T	T	A	T	C	A	C	G	C	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																				
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

×	α	A	T	C	T	T	C	G	T	T	A	T	C	A	C	G	C	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	-2	-2	3	-2	-2	3	-2	-2	-2	-2	-2	3	-2	3	-2	3	-2	3	-2	-2
T	0	-2	1	-2	6	1	-2	1	1	1	-4	1	-2	1	-2	1	-2	1	-2	6	1
T	0	-2	1	-1	1	9	4	-1	4	4	-1	-1	-1	-4	-1	-4	-1	-4	-1	1	4
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

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

Alinhamento Semi-Global

×	α	A	T	C	T	T	C	G	T	T	A	T	C	A	C	G	C	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	-2	-2	3	-2	-2	3	-2	-2	-2	-2	-2	3	-2	3	-2	3	-2	3	-2	-2
T	0	-2	1	-2	6	1	-2	1	1	1	-4	1	-2	1	-2	1	-2	1	-2	6	1
T	0	-2	1	-1	1	9	4	-1	4	4	-1	-1	-1	-4	-1	-4	-1	-4	-1	1	4
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

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

Alinhamento Semi-Global

×	α	A	T	C	T	T	C	G	T	T	A	T	C	A	C	G	C	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	-2	-2	3	-2	-2	3	-2	-2	-2	-2	-2	3	-2	3	-2	3	-2	3	-2	-2
T	0	-2	1	-2	6	1	-2	1	1	1	-4	1	-2	1	-2	1	-2	1	-2	6	1
T	0	-2	1	-1	1	9	4	-1	4	4	-1	-1	-1	-4	-1	-4	-1	-4	-1	1	4
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
α																	**C**	A	C	T	A
β																	**C**	-	-	-	-

Alinhamento Semi-Global

×	α	A	T	C	T	T	C	G	T	T	A	T	C	A	C	G	C	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	-2	-2	3	-2	-2	3	-2	-2	-2	-2	-2	3	-2	3	-2	3	-2	3	-2	-2
T	0	-2	1	-2	6	1	-2	1	1	1	-4	1	-2	1	-2	1	-2	1	-2	6	1
T	0	-2	1	-1	1	9	4	-1	4	4	-1	-1	-1	-4	-1	-4	-1	-4	-1	1	4
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
α													-	A	C	G	C	A	C	T	A
β													C	C	C	G	C	-	-	-	-

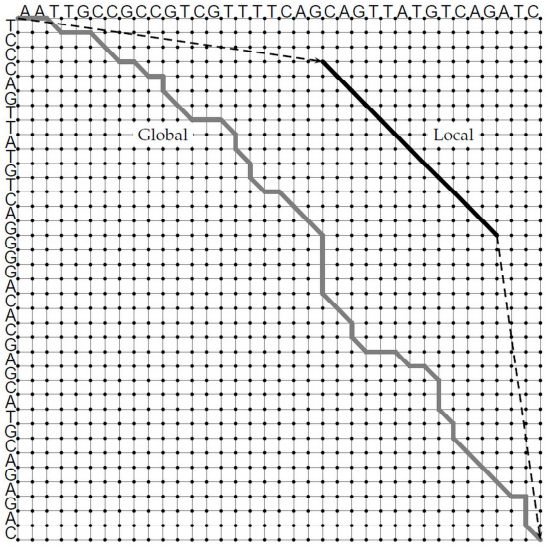
Alinhamento Semi-Global

×	α	A	T	C	T	T	C	G	T	T	A	T	C	A	C	G	C	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
C	0	-2	-2	3	-2	-2	3	-2	-2	-2	-2	-2	3	-2	3	-2	3	-2	3	-2	-2
T	0	-2	1	-2	6	1	-2	1	1	1	-4	1	-2	1	-2	1	-2	1	-2	6	1
T	0	-2	1	-1	1	9	4	-1	4	4	-1	-1	-1	-4	-1	-4	-1	-4	-1	1	4
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
α			C	T	T	C	G	T	T	A	-	T	C	A	C	G	C	A	C	T	A
β			C	T	T	G	G	C	C	A	A	T	C	C	C	G	C	-	-	-	-

Alinhamento Semi-Global

×	α	A	T	C	T	T	C	G	T	T	A	T	C	A	C	G	C	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
C	0	-2	-2	3	-2	-2	3	-2	-2	-2	-2	-2	3	-2	3	-2	3	-2	3	-2	-2
T	0	-2	1	-2	6	1	-2	1	1	1	-4	1	-2	1	-2	1	-2	1	-2	6	1
T	0	-2	1	-1	1	9	4	-1	4	4	-1	-1	-1	-4	-1	-4	-1	-4	-1	1	4
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
α		A	T	C	T	T	C	G	T	T	A	-	T	C	A	C	G	C	A	C	T	A
β		-	-	C	T	T	G	G	C	C	A	A	T	C	C	C	G	C	-	-	-	-

Alinhamento Global x Local



Alinhamento Local

- Como alterar o algoritmo de Needleman-Wunsch para produzir alinhamentos locais?
- Qual o ponto inicial dos alinhamentos ótimos de subsequências?
  - O valor mínimo de um alinhamento a se considerar é zero (alinhamento da subsequência vazia com qualquer outra).
- Qual o ponto final dos alinhamentos ótimos de subsequências?
  - Buscar o início no alinhamento ótimo em todas as posições da matriz. O alinhamento começa da célula de maior valor.
- Em 1981, Temple Smith e Michael Waterman propuseram o algoritmo de programação dinâmica para alinhamento local de sequências.

Alinhamento Local

Algoritmo 7: Local

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

×	α	G	A	C	A	A	C	G	T	T	A	C	T	G	C	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																				
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

×	α	G	A	C	A	A	C	G	T	T	A	C	T	G	C	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	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

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

Alinhamento Local

×	α	G	A	C	A	A	C	G	T	T	A	C	T	G	C	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	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

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

Alinhamento Local

×	α	G	A	C	A	A	C	G	T	T	A	C	T	G	C	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	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
α															C						
β															C						

Alinhamento Local

×	α	G	A	C	A	A	C	G	T	T	A	C	T	G	C	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	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
α														C	C						
β														G	C						

Alinhamento Local

×	α	G	A	C	A	A	C	G	T	T	A	C	T	G	C	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	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
α													T	C	C						
β													T	G	C						

Alinhamento Local

×	α	G	A	C	A	A	C	G	T	T	A	C	T	G	C	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	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
α													C	T	C	C					
β													C	T	G	C					

Alinhamento Local

×	α	G	A	C	A	A	C	G	T	T	A	C	T	G	C	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
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
	α											A	C	T	C	C						
	β											A	C	T	G	C						

Alinhamento Local

×	α	G	A	C	A	A	C	G	T	T	A	C	T	G	C	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
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
	α											A	C	T	C	C						
	β											A	C	T	G	C						

Distância de Edição

- Custos das operações:
  - Inserção = 1
  - Remoção = 1
  - Substituição = 2
- Calcular 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 = -2
  - Gap = -1

Distância de Edição

×	α	C	C	T	G	T	G	G	C	A	A	C
	β	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 = -2

Distância de Edição

×	α	C	C	T	G	T	G	G	C	A	A	C
β	0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11
A	-1	-2	-3	-4	-5	-6	-7	-8	-9	-8	-9	-10
T	-2	-3	-4	-3	-4	-5	-6	-7	-8	-9	-10	-11
T	-3	-4	-5	-4	-5	-4	-5	-6	-7	-8	-9	-10
G	-4	-5	-6	-5	-4	-5	-4	-5	-6	-7	-8	-9
G	-5	-6	-7	-6	-5	-6	-5	-4	-5	-6	-7	-8
C	-6	-5	-6	-7	-6	-7	-6	-5	-4	-5	-6	-7
C	-7	-6	-5	-6	-7	-8	-7	-6	-5	-6	-7	-6
A	-8	-7	-6	-7	-8	-9	-8	-7	-6	-5	-6	-7
C	-9	-8	-7	-8	-9	-10	-9	-8	-7	-6	-7	-6

Gap = -1 Match = 0 Mismatch = -2

Distância de Edição

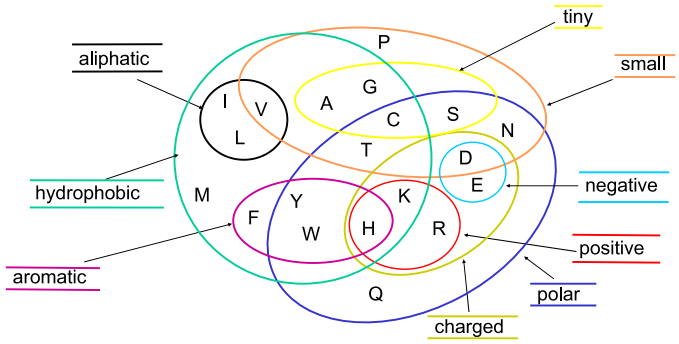
×	α	C	C	T	G	T	G	G	C	A	A	C
β	0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11
A	-1	-2	-3	-4	-5	-6	-7	-8	-9	-8	-9	-10
T	-2	-3	-4	-3	-4	-5	-6	-7	-8	-9	-10	-11
T	-3	-4	-5	-4	-5	-4	-5	-6	-7	-8	-9	-10
G	-4	-5	-6	-5	-4	-5	-4	-5	-6	-7	-8	-9
G	-5	-6	-7	-6	-5	-6	-5	-4	-5	-6	-7	-8
C	-6	-5	-6	-7	-6	-7	-6	-5	-4	-5	-6	-7
C	-7	-6	-5	-6	-7	-8	-7	-6	-5	-6	-7	-6
A	-8	-7	-6	-7	-8	-9	-8	-7	-6	-5	-6	-7
C	-9	-8	-7	-8	-9	-10	-9	-8	-7	-6	-7	-6

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

Pontuação para Alinhamento de Caracteres

- Simples:
  - Match: alinhamento de dois caracteres iguais
  - Mismatch: alinhamento de dois caracteres diferentes
- Matrizes de Substituições:
  - 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 suas bases.
- A similariedade esperada entre duas sequências protéicas com distância 200-PAM é de cerca de 25%.
- Seja $F(i,j)$ a probabilidade de uma aminoácido a_i mutar para o aminoácido a_j em sequências com distância 1-PAM e $freq(a_j)$ a frequência do aminoácido a_j.
- Fórmula geral para matrizes PAM:

$$PAM_k(i,j) = \left\lceil 10.\log \frac{F^k(i,j)}{freq(a_j)} \right\rceil$$

- 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 de 504 grupos de proteínas.
- $BLOSUM_K$ considera apenas sequências com não menos do que K% de divergência entre suas bases.
- Fórmula geral para matrizes BLOSUM:

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

- $BLOSUM_{62}$: 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

$BLOSUM_{62}$

	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	-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		Y
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.
- BLOSUM é mais tolerante a substituições hidrofóbicas, mas menos tolerante a substituições hidrofílicas.
- A Entropia Relativa (H) de uma matriz de substituição é dada por:

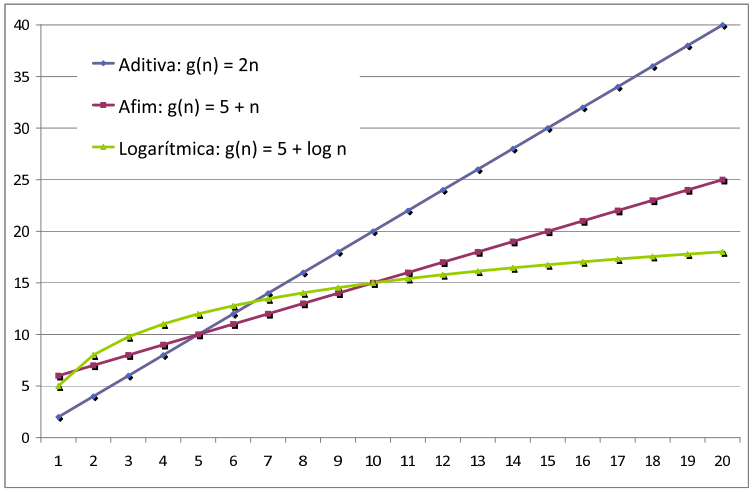
$$H = \sum_{i=1}^{20} \sum_{j=1}^{i} freq(a_i) freq(a_j) \sigma(a_i, a_j)$$

- Equivalência entre matrizes (com base em sua Entropia Relativa):
  - $PAM_{100} \Longleftrightarrow BLOSUM_{90}$
  - $PAM_{120} \Longleftrightarrow BLOSUM_{80}$
  - $PAM_{160} \Longleftrightarrow BLOSUM_{60}$
  - $PAM_{200} \Longleftrightarrow BLOSUM_{50}$
  - $PAM_{250} \Longleftrightarrow BLOSUM_{45}$
  - $PAM_{400} \Longleftrightarrow BLOSUM_{30}$

Penalidades para Blocos de Buracos

- Funções Aditivas: $g(k_1 + k_2) = g(k_1) + g(k_2)$
  - Linear: $g(k) = a.k$
- Funções Sub-Aditivas: $g(k_1 + k_2) < g(k_1) + g(k_2)$
  - Afim: $g(k) = a + b.k$
  - Funções Convexas: $g(k) - g(k-1) > g(k+1) - g(k)$
    - Logarítmica: $g(k) = a + b.\log k$

Exemplos de Funções para Penalidades de Buracos



Penalidades para Buracos

- Diferenças entre alinhamentos com mesma matriz de pontuação de alinhamento de bases, mas com penalidades diferentes para 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 Aditiva para Penalidade de Buracos

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

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

Alinhamento Global com Função Aditiva para Penalidade de Buracos: Alinhamento Upmost

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

Alinhamento Global com Função Sub-Aditiva 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 $M[i-k, j]$, acrescido de um alinhamento entre $\alpha[i-k+1..i]$ e um bloco formado por k buracos (para $1 \leq k \leq i$).
  - 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 $1 \leq k \leq j$).
- Algoritmo proposto por Michael Waterman, Temple Smith e William Beyer em 1976.

Alinhamento Global
com Função Sub-Aditiva de Penalidade para Buracos

×	α	C	C	T	G	T	G	G	C	A	A	C
	β											
A												
T												
T												
G												
G												
C												
C												
A												
C												

Alinhamento Global
com Função Sub-Aditiva de Penalidade para Buracos

×	α	G	C	G	C	G	T	T	A	G	A	C	T	A	G	C	A	C	C	G
	β	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	-9													
T		-9																			
T		-10																			
G		-11																			
C		-12																			
A		-13																			
C		-14																			
C		-15																			
G		-16																			

α = G C G C G T — Open Gap = -5 — Match = +3
β = G G G - - -
–9 = 3 –2 3 –6 –1 –6 — Extend Gap = -1 — Mismatch = -2

Alinhamento Global
com Função Sub-Aditiva de Penalidade para Buracos

×	α	G	C	G	C	G	T	T	A	G	A	C	T	A	G	C	A	C	C	G
	β	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	-9													
T		-9																			
T		-10																			
G		-11																			
C		-12																			
A		-13																			
C		-14																			
C		-15																			
G		-16																			

α = G C G C G T — Open Gap = -5 — Match = +3
β = G G G - - -
–9 = 3 –2 3 –6 –1 –6 — Extend Gap = -1 — Mismatch = -2

Alinhamento Global
com Função Sub-Aditiva de Penalidade para Buracos

×	α	G	C	G	C	G	T	T	A	G	A	C	T	A	G	C	A	C	C	G
	β	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																			

α = G C G C G T — Open Gap = -5 — Match = +3
β = G G G - - -
–4 = 3 –2 3 –6 –1 –1 — Extend Gap = -1 — Mismatch = -2

Alinhamento Global
com Função Sub-Aditiva de Penalidade para Buracos

Algoritmo 8: Global

Input: α, m, β, 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 $max < M[i,k-1] + gap(j-k+1)$ **then** $\max \leftarrow M[i,k-1] + gap(j-k+1)$
    end
    for *all* $k \in [1..i]$ **do**
      if $max < M[k-1,j] + gap(i-k+1)$ **then** $\max \leftarrow M[k-1,j] + gap(i-k+1)$
    end
    $M[i,j] \leftarrow \max$
  end
end
return $M[n,m]$

Complexidade e Tempo de Execução - Alinhamento Global

- Complexidade:
  - $T(m,n) = \Theta(mn(m+n)) = \Theta(m^2n + nm^2)$
- Exemplo: computador de 3GHz (1 instrução por ciclo)
  - m = n = 10:
    - Tempo: $(10 * 10 * 20)/(3 * 2^{30}) = 31$ bilionésimos de segundo
  - m = n = 100:
    - Tempo: $(100 * 100 * 200)/(3 * 2^{30}) = 621$ milionésimos de segundo
  - m = n = 1000:
    - Tempo: $(1000 * 1000) * 2000/(3 * 2^{30}) = 621$ milésimos de segundo
  - m = n = 10000:
    - Tempo: $(10000 * 10000) * 20000/(3 * 2^{30}) = 10$ minutos
  - m = n = 100000:
    - Tempo: $(100000 * 100000 * 200000)/(3 * 2^{30}) = 172$ horas
  - m = n = 1000000:
    - Tempo: $(1000000 * 1000000 * 2000000)/(3 * 2^{30}) = 19$ anos

Alinhamento Global com Função Afim
para Penalidade de Buracos

×	α	G	C	G	C	G	T	T	A	G	A	C	T	A	G	C	A	C	C	G
β	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 Afim
para Penalidade de Buracos

×	α	G	C	G	C	G	T	T	A	G	A	C	T	A	G	C	A	C	C	G
β	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
α	=	G	C	G	C	G	T	T	A	G	A	C	T	A	G	C	A	C	C	G
β	=	-	-	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	3

Alinhamento Global com Função Afim de Penalidade para Buracos

- 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**
  $M_{\nwarrow}[i,0] \leftarrow -\infty$
  $M_{\leftarrow}[i,0] \leftarrow -\infty$
  $M_{\uparrow}[i,0] \leftarrow -(a+b.i)$
end
for *all* $j \in [1..m]$ **do**
  $M_{\nwarrow}[0,j] \leftarrow -\infty$
  $M_{\leftarrow}[0,j] \leftarrow -(a+b.j)$
  $M_{\uparrow}[0,j] \leftarrow -\infty$
end

Alinhamento Global com Função Afim de Penalidade para Buracos

Algoritmo 10: GlobalAfim

Input: α, m, β, 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 $M_{\nwarrow}$

×	α	G	C	G	C	G	T	T	A	G	A	C	T	A	G	C	A	C	C	G
β	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 $M_{\leftarrow}$

×	α	G	C	G	C	G	T	T	A	G	A	C	T	A	G	C	A	C	C	G
β	-∞	-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 $M_{\uparrow}$

×	α	G	C	G	C	G	T	T	A	G	A	C	T	A	G	C	A	C	C	G
β	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞
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 $M_{\nwarrow}$

×	α	G	C	G	C	G	T	T	A	G	A	C	T	A	G	C	A	C	C	G
β	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	-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

Open Gap = -5 Extend Gap = -1 Match = +3 Mismatch = -2

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

×	α	G	C	G	C	G	T	T	A	G	A	C	T	A	G	C	A	C	C	G
β	-∞	-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 $M_{\uparrow}$

×	α	G	C	G	C	G	T	T	A	G	A	C	T	A	G	C	A	C	C	G
β		-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞	-∞
G		-6	-12	-13	-14	-15	-16	-17	-18	-19	-20	-21	-22	-23	-24	-25	-26	-27	-28	-29	-30
G		-7	-3	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18	-19	-20	-21	-22	-23	-24	-25	-26
G		-8	-4	-5	-6	-12	-8	-14	-15	-16	-12	-18	-19	-20	-21	-17	-23	-24	-25	-26	-22
T		-9	-5	-6	-2	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18	-19	-20	-21	-22	-23
T		-10	-6	-7	-3	-4	-10	-6	-7	-13	-14	-15	-16	-12	-18	-19	-20	-21	-22	-23	-24
G		-11	-7	-8	-4	-5	-6	-7	-3	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18	-19	-20
C		-12	-8	-9	-5	-6	-7	-8	-4	-5	-6	-12	-13	-14	-15	-11	-17	-18	-19	-20	-16
A		-13	-9	-10	-6	-7	-8	-9	-5	-6	-7	-8	-9	-15	-16	-12	-8	-14	-15	-16	-17
C		-14	-10	-11	-7	-8	-9	-10	-6	-7	-8	-4	-10	-11	-12	-13	-9	-5	-11	-12	-13
C		-15	-11	-12	-8	-9	-10	-11	-7	-8	-9	-5	-1	-7	-8	-9	-10	-6	-2	-8	-9
G		-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 Downmost (Matriz $M_{\nwarrow}$)

×	α	G	C	G	C	G	T	T	A	G	A	C	T	A	G	C	A	C	C	G
β		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	-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
α	=															**G**	**C**	**A**	**C**	**C**	**G**
β	=															**G**	**C**	**A**	**C**	**C**	**G**
	=															3	3	3	3	3	3

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

×	α	G	C	G	C	G	T	T	A	G	A	C	T	A	G	C	A	C	C	G
β		-∞	-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
α	=								**A**	**G**	**A**	**C**	**T**	**A**	**G**	**C**	**A**	**C**	**C**	**G**
β	=								-	-	-	-	-	-	**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 $M_{\nwarrow}$)

×	α	G	C	G	C	G	T	T	A	G	A	C	T	A	G	C	A	C	C	G
β		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	-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
α	=							**T**	**T**	**A**	**G**	**A**	**C**	**T**	**A**	**G**	**C**	**A**	**C**	**G**
β	=							**T**	**T**	-	-	-	-	-	-	**G**	**C**	**A**	**C**	**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 $M_{\leftarrow}$)

×	α	G	C	G	C	G	T	T	A	G	A	C	T	A	G	C	A	C	C	G
β	-∞	-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
α	=				C	G	T	T	A	G	A	C	T	A	G	C	A	C	C	G
β	=				-	-	T	T	-	-	-	-	-	-	G	C	A	C	C	G
	=				−6	−1	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 $M_{\nwarrow}$)

×	α	G	C	G	C	G	T	T	A	G	A	C	T	A	G	C	A	C	C	G
β	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	-∞	-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	-8	-4	-5	-6	-12	-13	1	-10	-11	-12
C	-∞	-16	-7	-8	-4	-5	-4	-7	-8	-9	-5	-1	-6	-7	-8	-9	-5	-1	3	-4
C	-∞	-17	-8	-4	-5	-11	-7	-8	-9	-5	-11	-7	-3	-4	-5	-6	-2	-8	-4	-4
G	-∞	-12	-13	-4	-10	-1	-7	-13	-9	-5	-11	-7	-3	1	0	-6	-2	-8	-4	10
α	=	G	C	G	C	G	T	T	A	G	A	C	T	A	G	C	A	C	C	G
β	=	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

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

×	α	G	C	G	C	G	T	T	A	G	A	C	T	A	G	C	A	C	C	G
β	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	-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
α	=												T	A	G	C	A	C	C	G
β	=												T	T	G	C	A	C	C	G
	=												3	−2	3	3	3	3	3	3

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

×	α	G	C	G	C	G	T	T	A	G	A	C	T	A	G	C	A	C	C	G
β	-∞	-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
α	=				C	G	T	T	A	G	A	C	T	A	G	C	A	C	C	G
β	=				-	-	-	-	-	-	-	-	T	T	G	C	A	C	C	G
	=				−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 Upmost (Matriz $M_{\nwarrow}$)

×	α	G	C	G	C	G	T	T	A	G	A	C	T	A	G	C	A	C	C	G
β	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	-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

α	=	G	C	G	C	G	T	T	A	G	A	C	T	A	G	C	A	C	C	G
β	=	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 usando Espaço Linear

- 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 as sequências de entrada?
- 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 α 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 $\overline{\alpha}$ com a sequência $\overline{\beta[\lfloor \frac{n}{2} \rfloor + 1..n]}$.
    - ⋆ *Observação*: $\overline{x}$ 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 α e β 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

×	α	A	C	T	G	G	G	T	C	A	A	C	C	G	T	C	T	G	C	G
β	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																				
A																				
C																				
A																				
G																				
G																				
C																				
A																				
T																				

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

Alinhamento Global usando Espaço Linear

×	α	A	C	T	G	G	G	T	C	A	A	C	C	G	T	C	T	G	C	G
β	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																		
A																				
C																				
A																				
G																				
G																				
C																				
A																				
T																				

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

Alinhamento Global usando Espaço Linear

×	α	A	C	T	G	G	G	T	C	A	A	C	C	G	T	C	T	G	C	G
β	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																	
A																				
C																				
A																				
G																				
G																				
C																				
A																				
T																				

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

Alinhamento Global usando Espaço Linear

×	α	A	C	T	G	G	G	T	C	A	A	C	C	G	T	C	T	G	C	G
β	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										
A																				
C																				
A																				
G																				
G																				
C																				
A																				
T																				

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

Alinhamento Global usando Espaço Linear

×	α	A	C	T	G	G	G	T	C	A	A	C	C	G	T	C	T	G	C	G
β	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
A																				
C																				
A																				
G																				
G																				
C																				
A																				
T																				

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

Alinhamento Global usando Espaço Linear

×	α	A	C	T	G	G	G	T	C	A	A	C	C	G	T	C	T	G	C	G
β	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
A																				
C																				
A																				
G																				
G																				
C																				
A																				
T																				

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

Alinhamento Global usando Espaço Linear

×	α	A	C	T	G	G	G	T	C	A	A	C	C	G	T	C	T	G	C	G
β	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
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

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

Alinhamento Global usando Espaço Linear

×	α	A	C	T	G	G	G	T	C	A	A	C	C	G	T	C	T	G	C	G
β	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
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

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

Alinhamento Global usando Espaço Linear

×	α	A	C	T	G	G	G	T	C	A	A	C	C	G	T	C	T	G	C	G
β	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
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
α	=	A	C	T	G	G	G	T	C	A	A	C	C	G	T	C	T	G	C	G
β	=	A	T	T	G	G	-	-	C	C	A	C	A	G	G	C	-	A	-	T

Alinhamento Global usando Espaço Linear

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

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

Alinhamento Global usando Espaço Linear

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

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

Alinhamento Global usando Espaço Linear

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

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

Alinhamento Global usando Espaço Linear

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

Alinhamento Global usando Espaço Linear

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

Alinhamento Global usando Espaço Linear

×	α	A	C	T	G	G	G	T	C		
β	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

×	α	A	C	T	G	G	G	T	C		
β	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

×	α	A	C	T	G	G	G	T	C		
β	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

×	α	A	C	T	G	G	G	T	C		
β	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

×	α	A	C	T	G	G	G	T	C
β	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

×	α	A	C
β	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

×	α	A	C
β	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

×	α	A	C
β	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

×	α	A	C				
β	0	-5	-10				
A	-5	3	-2				
		-7	-2	-5	T		
		-10	-5	0			
		A	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	C
		-20	-15	-10	-5	0	
		G	G	T	C		

A C T G
A T T G

A A
C A

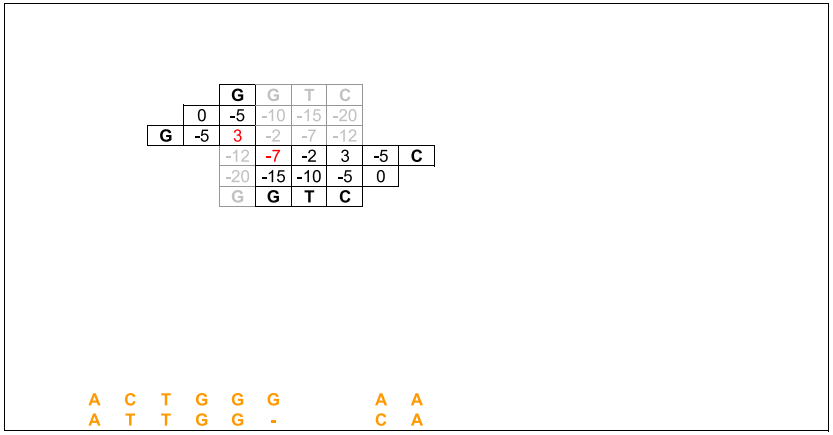
Alinhamento Global usando Espaço Linear

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

A C T G
A T T G

A A
C A

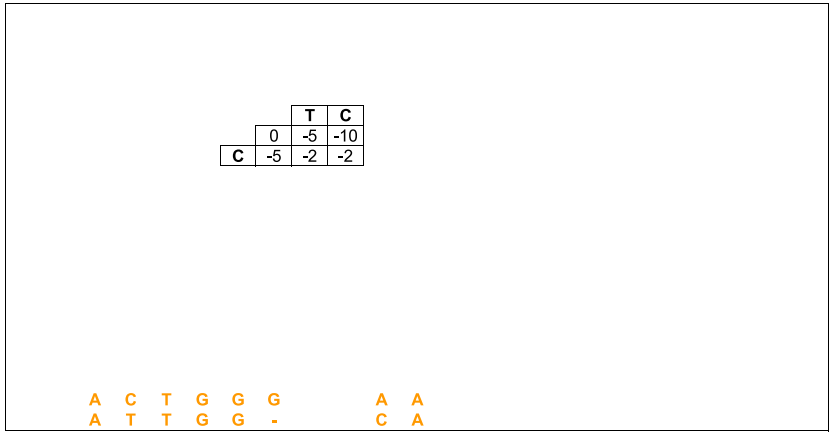
Alinhamento Global usando Espaço Linear



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

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

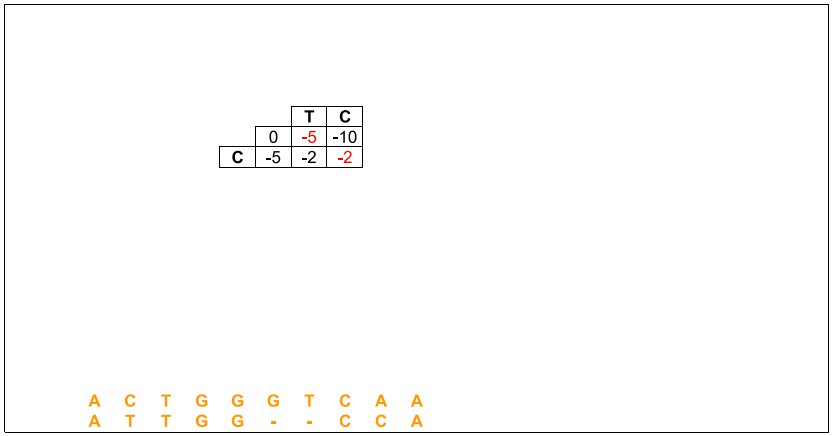
Alinhamento Global usando Espaço Linear



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

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

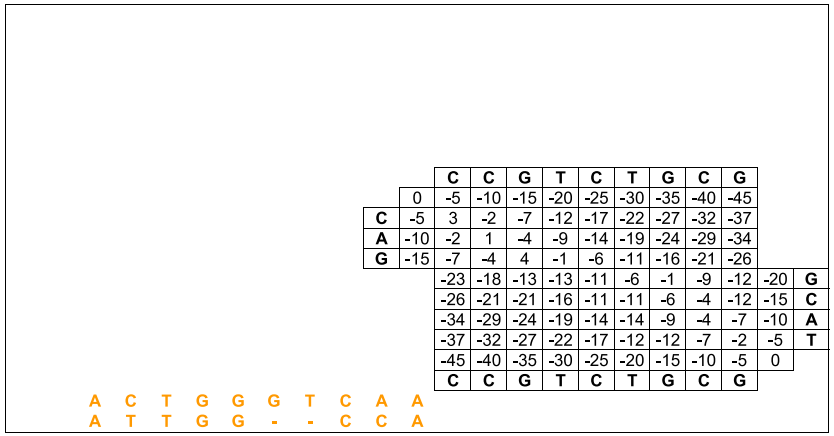
Alinhamento Global usando Espaço Linear



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

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

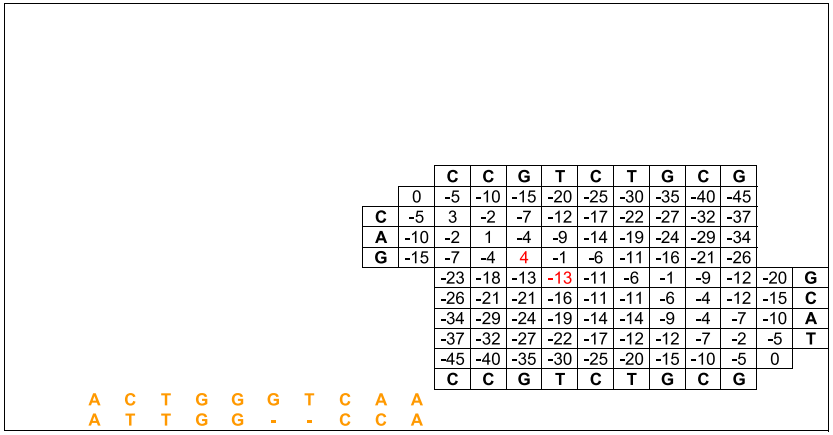
Alinhamento Global usando Espaço Linear



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

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

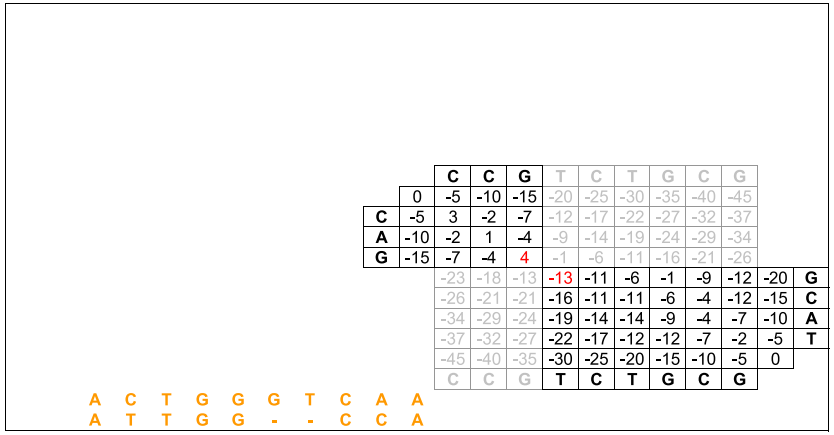
Alinhamento Global usando Espaço Linear



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

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

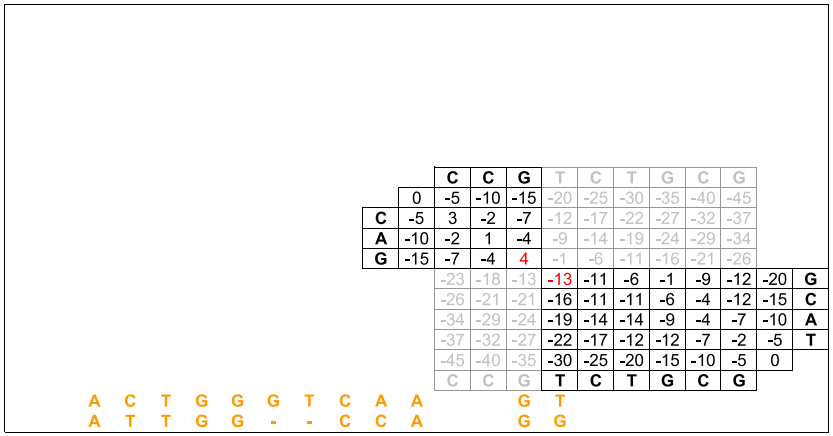
Alinhamento Global usando Espaço Linear



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

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

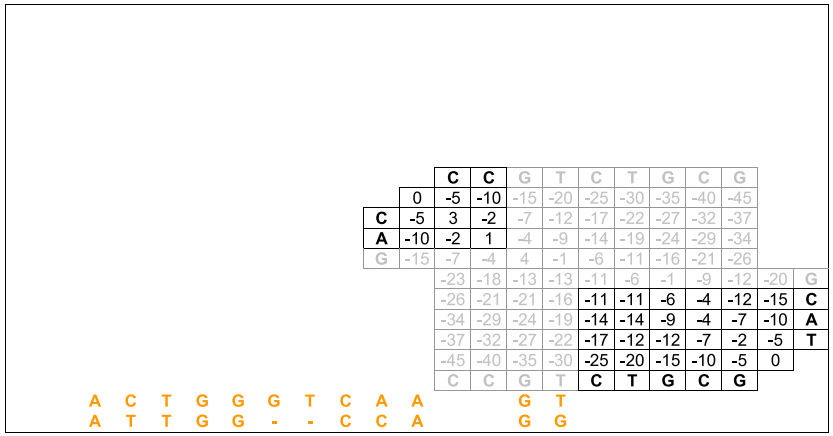
Alinhamento Global usando Espaço Linear



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

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

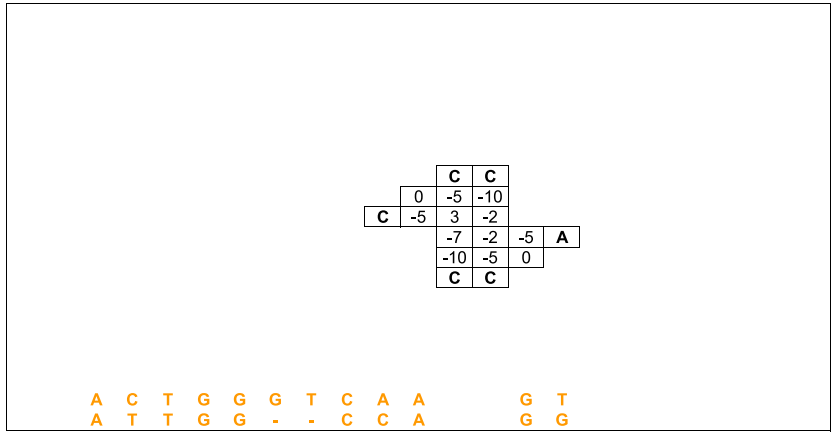
Alinhamento Global usando Espaço Linear



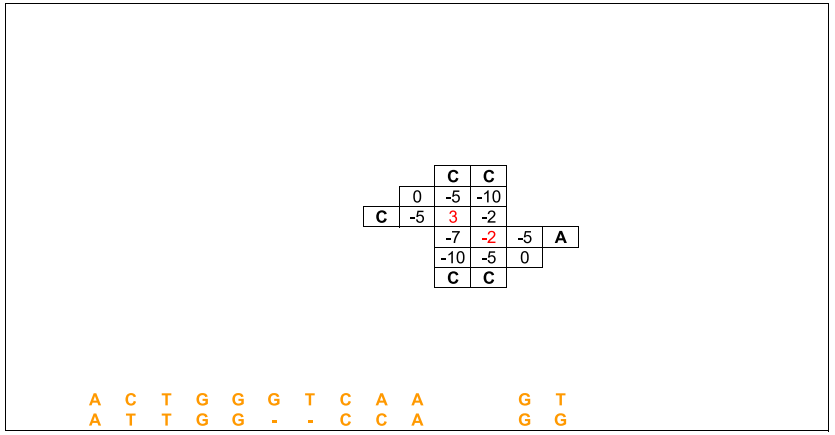
		C	C	G	T	C	T	G	C	G		
	0	-5	-10	-15	-20	-25	-30	-35	-40	-45		
C	-5	3	-2	-7	-12	-17	-22	-27	-32	-37		
A	-10	-2	1	-4	-9	-14	-19	-24	-29	-34		
G	-15	-7	-4	4	-1	-6	-11	-16	-21	-26		
		-23	-18	-13	-13	-11	-6	-1	-9	-12	-20	G
		-26	-21	-21	-16	-11	-11	-6	-4	-12	-15	C
		-34	-29	-24	-19	-14	-14	-9	-4	-7	-10	A
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

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

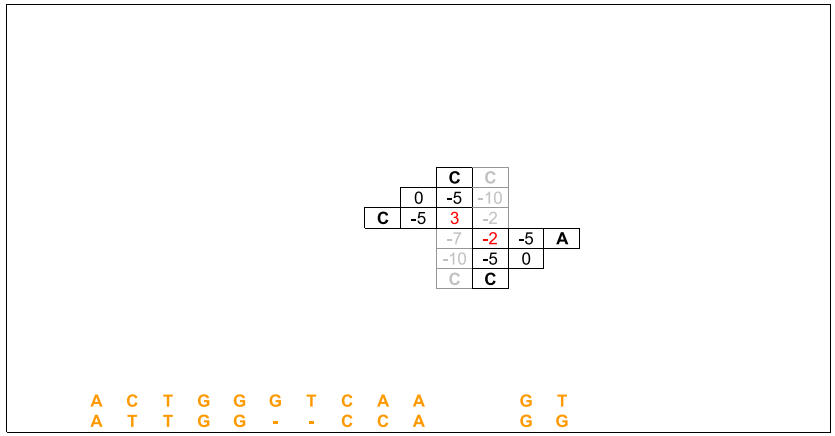
Alinhamento Global usando Espaço Linear



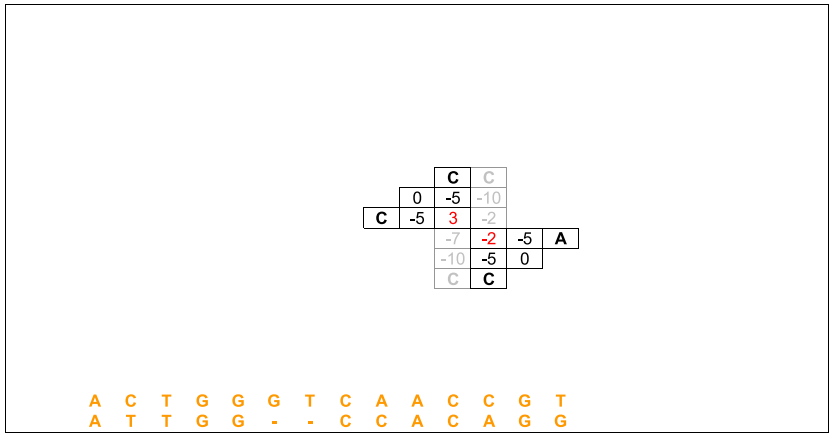
Alinhamento Global usando Espaço Linear



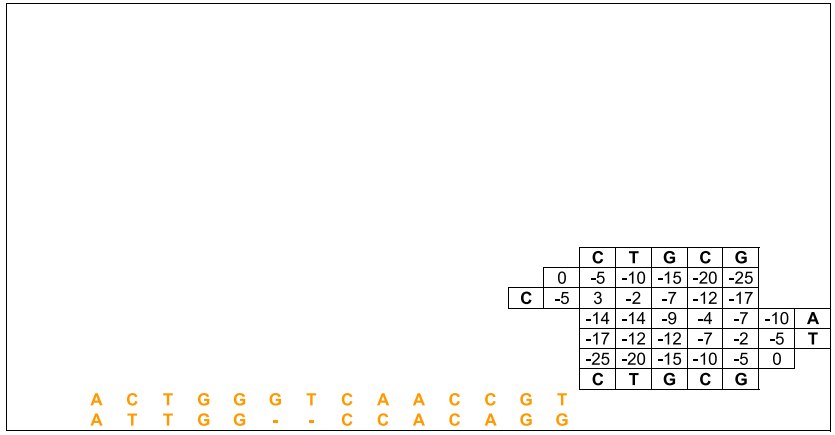
Alinhamento Global usando Espaço Linear



Alinhamento Global usando Espaço Linear



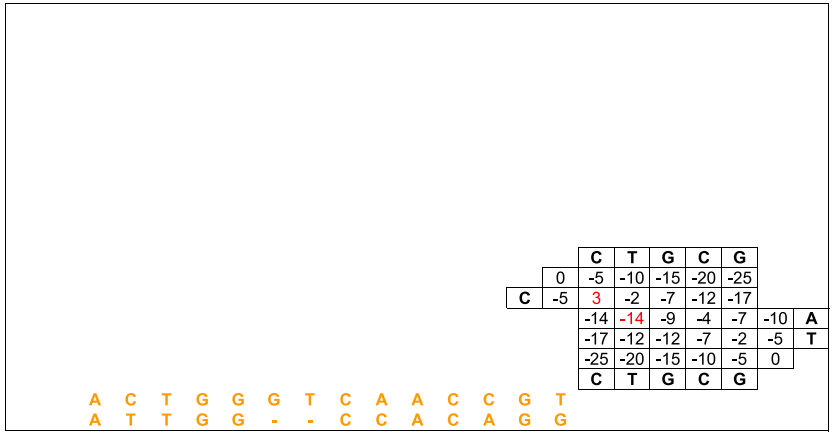
Alinhamento Global usando Espaço Linear



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

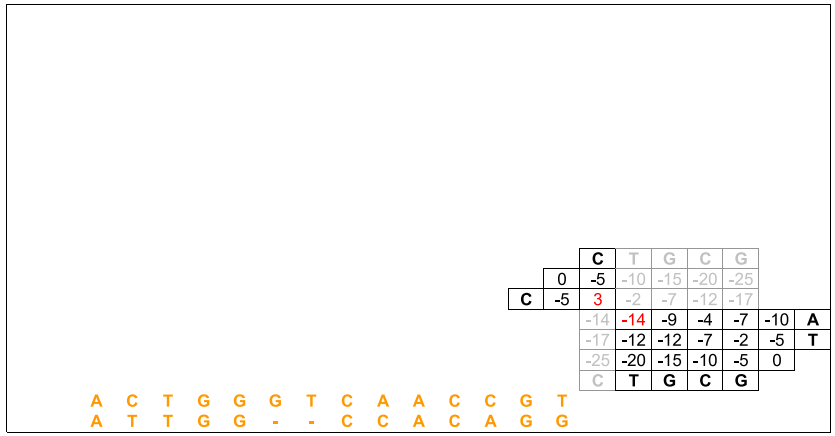
Alinhamento Global usando Espaço Linear



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

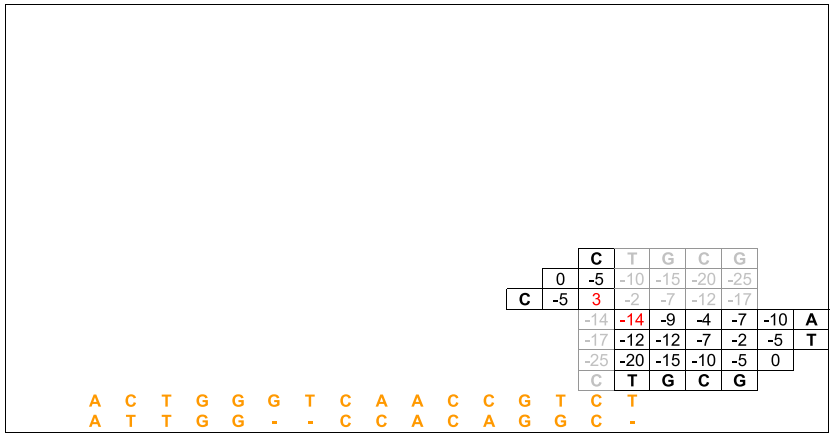
Alinhamento Global usando Espaço Linear



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

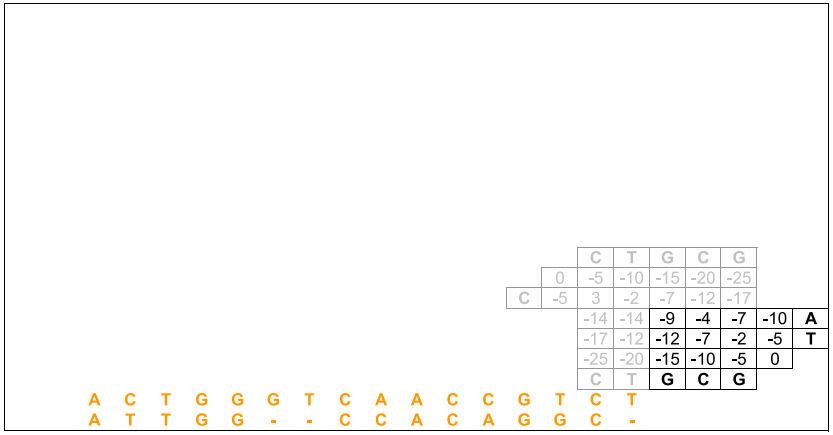
Alinhamento Global usando Espaço Linear



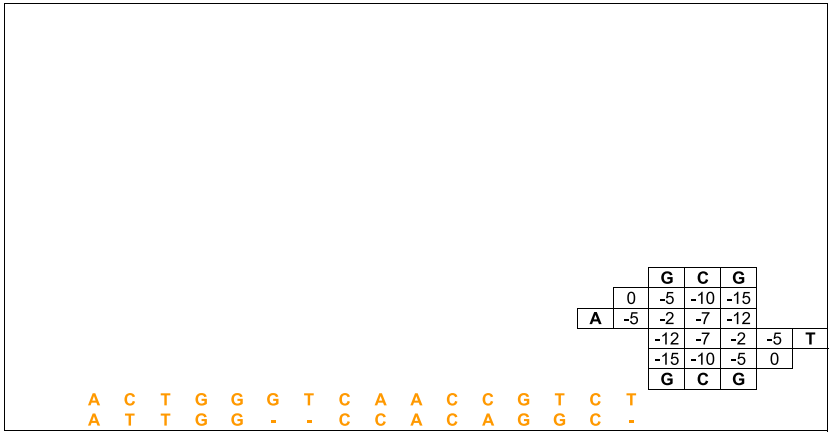
		C	T	G	C	G		
	0	-5	-10	-15	-20	-25		
C	-5	3	-2	-7	-12	-17		
		-14	-14	-9	-4	-7	-10	A
		-17	-12	-12	-7	-2	-5	T
		-25	-20	-15	-10	-5	0	
		C	T	G	C	G		

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

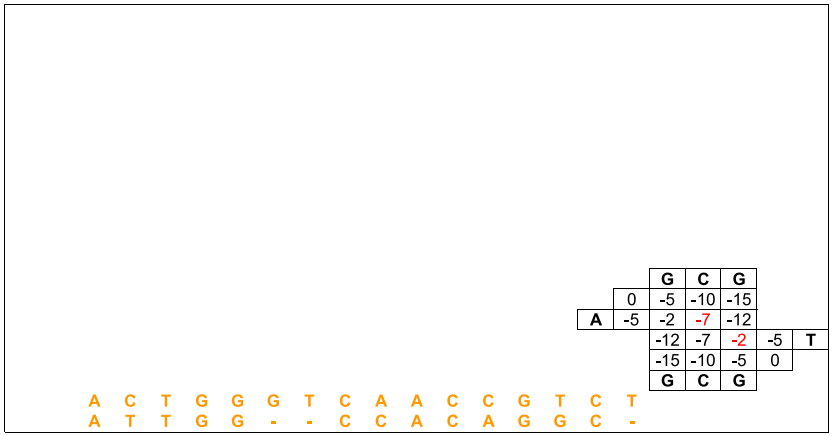
Alinhamento Global usando Espaço Linear



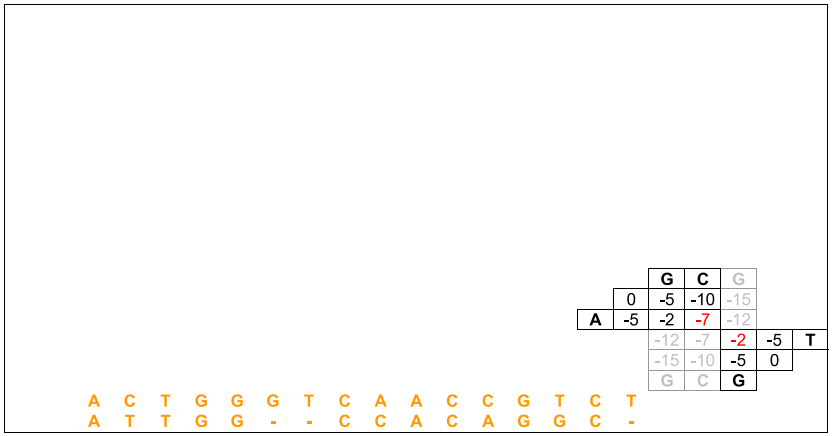
Alinhamento Global usando Espaço Linear



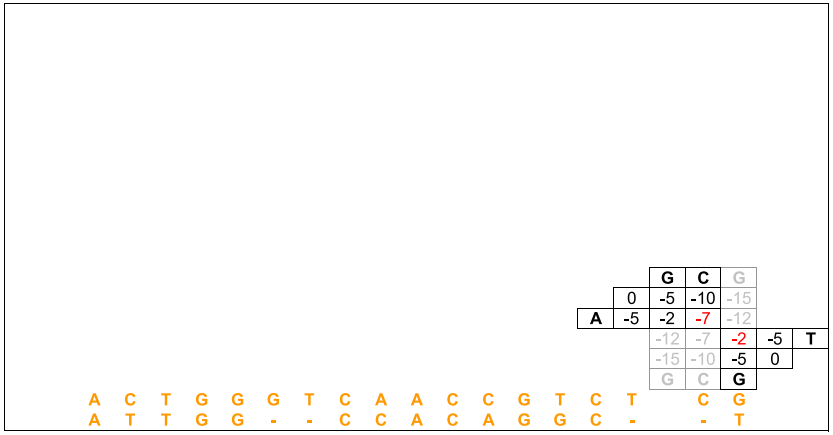
Alinhamento Global usando Espaço Linear



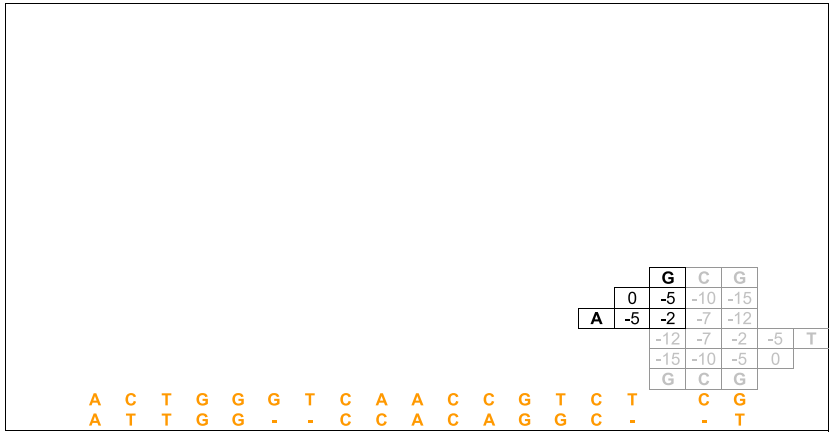
Alinhamento Global usando Espaço Linear



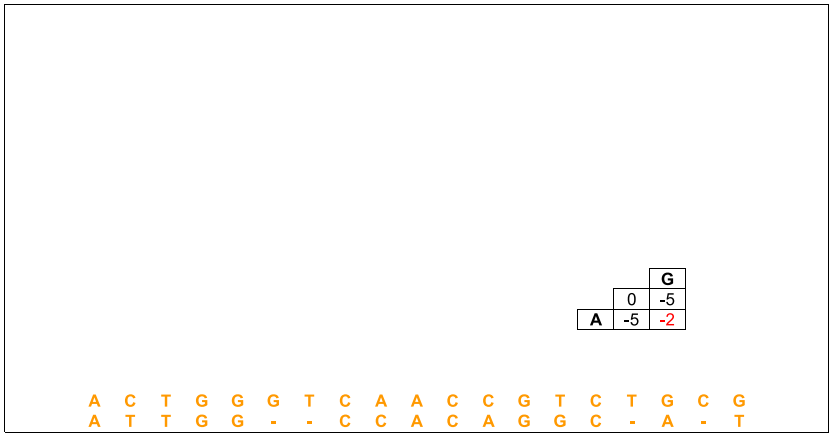
Alinhamento Global usando Espaço Linear



Alinhamento Global usando Espaço Linear



Alinhamento Global usando Espaço Linear



Alinhamento Global usando Espaço Linear

×	α	A	C	T	G	G	G	T	C	A	A	C	C	G	T	C	T	G	C	G
β	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
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
α	=	A	C	T	G	G	G	T	C	A	A	C	C	G	T	C	T	G	C	G
β	=	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 sequência α e β, já que conseguimos alinhar pelo menos um caracter a cada chamada recursiva:
    - $T(m,n) = O(m+n).O(mn) = O(m^2n + mn^2)$
  - 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:
    - $T(m,n) = O(\log mn).O(mn) = O(mn \log mn)$
  - Terceira observação: podemos obter um resultado melhor fazendo uma análise assintótica agregada (soma de todos os passos do algoritmo):
    - $T(m,n) = \sum_{k=0}^{\log mn} \Theta(\frac{mn}{2^k})$
    - $T(m,n) = \Theta(\sum_{k=0}^{\log mn} \frac{mn}{2^k})$
    - $T(m,n) = \Theta(\sum_{k=0}^{\infty} \frac{mn}{2^k})$
    - $T(m,n) = \Theta(mn)$

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
  - Uso de 3 matrizes para armazenar os alinhamentos
  - 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 Global de Sequências Similares

- Suponha que queremos alinhar duas sequências α e β 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 k o número de buracos no alinhamento ótimo.
- O que podemos afirmar em relação as células da matriz que representam alinhamentos ótimos entre α e β?
  - Elas estão, no máximo, a k células de distância da diagonal principal.
- Idéia: 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?
  - $M(n-k-1) + 2(k+1)g$
- 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 $M[n,n] > M(n-k-1) + 2(k+1)g$, 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

×	α	G	C	G	A	A	A	A	G	T	G	A	T	T	T	C	C	T	C	C
β	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	-16	-21	-26
G	-70	-62	-54	-46	-38	-30	-22	-14	-6	2	10	5	0	-5	-10	-13	-18	-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

×	α	G	C	G	A	A	A	A	G	T	G	A	T	T	T	C	C	T	C	C
β	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 = 37

Alinhamento Global de Sequências Similares

×	α	G	C	G	A	A	A	A	G	T	G	A	T	T	T	C	C	T	C	C
β	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	-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 = 24

Alinhamento Global de Sequências Similares

×	α	G	C	G	A	A	A	A	G	T	G	A	T	T	T	C	C	T	C	C
β	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											
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 = 11

Alinhamento Global de Sequências Similares

×	α	G	C	G	A	A	A	A	G	T	G	A	T	T	T	C	C	T	C	C
β	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	-18	
A												13	8	3	-2	-7	-12	-17	-20	-20
T													16	11	6	1	-4	-9	-14	-19
G														14	9	4	-1	-6	-11	-16
T															17	12	7	2	-3	-8
A																15	10	5	0	-5

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

Alinhamento Global de Sequências Similares

×	α	G	C	G	A	A	A	A	G	T	G	A	T	T	T	C	C	T	C	C
β	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	-16	-21	
G										2	10	5	0	-5	-10	-13	-18	-18	-18	-23
A											5	13	8	3	-2	-7	-12	-17	-20	-20
T												8	16	11	6	1	-4	-9	-14	-19
G													11	14	9	4	-1	-6	-11	-16
T														14	17	12	7	2	-3	-8
A															12	15	10	5	0	-5

n = 19	M = 3	m = -2	g = -5	k = 5	máximo = -15

Alinhamento Global de Sequências Similares
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 idéias 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): $k = 2k$
    - ★ $T(n) = \sum_{i=0}^{\log_2 k} \Theta(2^i n) = \Theta(n) \sum_{i=1}^{\log_2 k} 2^i$
    - ★ $T(n) = \Theta(n)(2k - 1) = \Theta(nk)$
    - ★ Pior caso $k = \Theta(n)$, logo: $T(n) = \Theta(n^2)$

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.

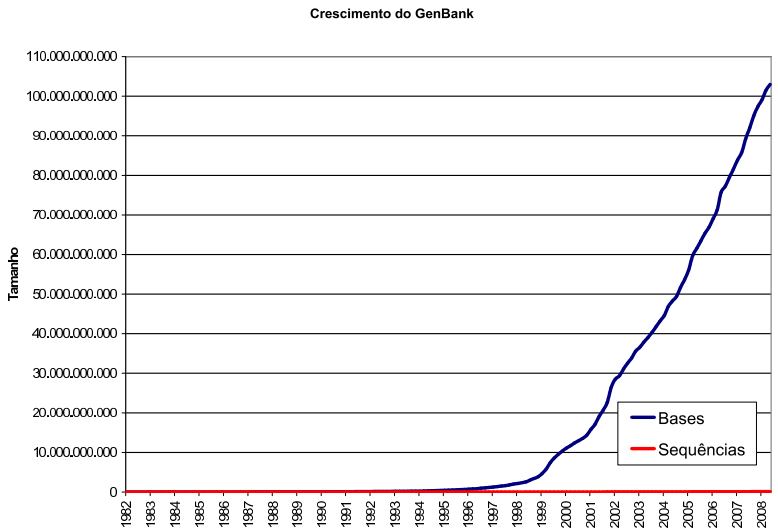
Algoritmo Sub-Quadrático para Alinhamento Global

- É possível obter um algoritmo de tempo sub-quadrático para alinhamento global considerando pontuação aditiva para buracos?
- Four-Russian Algorithm (Arlazarov, Dinic, Kronrod e Faradzev), 1970.
  - Divide a matriz de programação dinâmica em quadrados de dimensão t.
  - Pré-computa os alinhamentos de todos os pares de subsequências de tamanho t em tempo $\Theta(12^t t^2)$.
  - Preenche a matriz de Programação Dinâmica, usando os valores pré-computados, em tempo $\Theta((m/t)(n/t)t) = \Theta(mn/t)$.
- Complexidade: $T(n) = \Theta(12^t t^2 + mn/t)$
- Valor de t que minimiza a complexidade do algoritmo:
  - $t = \log_{12}(m+n)$.
- Complexidade:
  - $T(n) = \Theta((m+n)\log^2(m+n) + mn/\log(m+n)) = \Theta(mn/\log(m+n)))$
- Algoritmo não utilizado na prática.

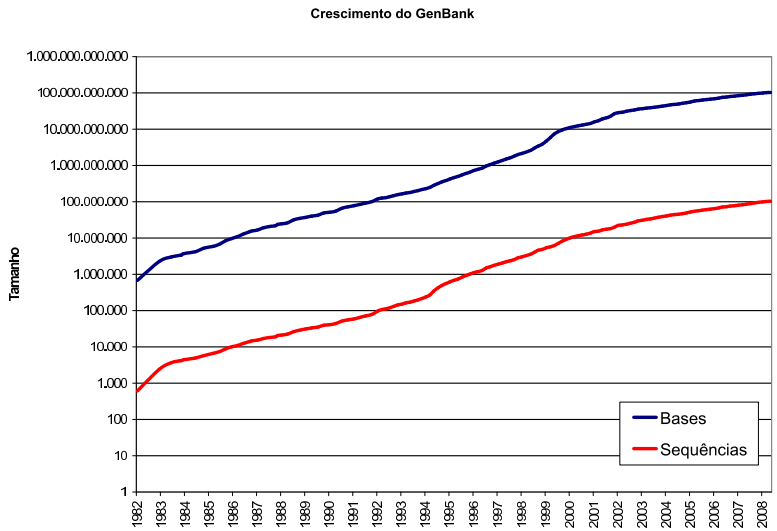
GenBank

- Maior banco público de sequências de nuceotídeos 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).
- Bioinformática x Lei de Moore:
  - *"From 1982 to the present, the number of bases in GenBank has doubled approximately every 18 months."*

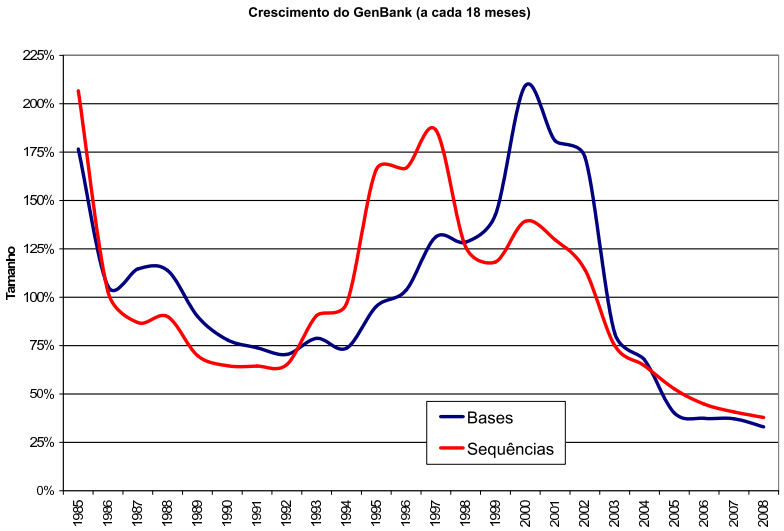
GenBank



GenBank



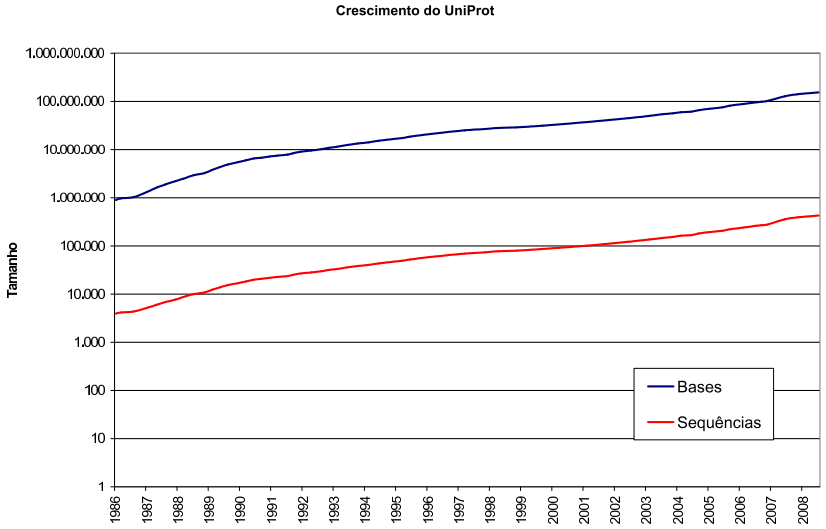
GenBank



UniProt

- 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.

UniProt



Alinhamento de uma Sequência contra todas as Sequências do GenBank

- Exemplo: computador de 3GHz (1 instrução por ciclo)
  - n = 10:
    - ★ Tempo: $(10 * 1000 * 100000000)/(3 * 2^{30}) = 5$ minutos
  - n = 100:
    - ★ Tempo: $(100 * 1000 * 100000000)/(3 * 2^{30}) = 1$ hora
  - n = 1000:
    - ★ Tempo: $(1000 * 1000 * 100000000)/(3 * 2^{30}) = 10$ horas
  - n = 10000:
    - ★ Tempo: $(10000 * 1000 * 100000000)/(3 * 2^{30}) = 3$ dias
  - n = 100000:
    - ★ Tempo: $(100000 * 1000 * 100000000)/(3 * 2^{30}) = 1$ mês
  - n = 1000000:
    - ★ Tempo: $(1000000 * 1000 * 100000000)/(3 * 2^{30}) = 1$ ano
  - n = 10000000:
    - ★ Tempo: $(10000000 * 1000 * 100000000)/(3 * 2^{30}) = 1$ década
  - n = 100000000:
    - ★ Tempo: $(100000000 * 1000 * 100000000)/(3 * 2^{30}) = 1$ século
  - n = 1000000000:
    - ★ Tempo: $(1000000000 * 1000 * 100000000)/(3 * 2^{30}) = 1$ milênio

BLAST

- BLAST: Basic Local Alignment Serch Tool.
- Ferramenta proposta por Stephen Altschul, Warren Gish, Webb Miller, Eugene Myers e David Lipman em 1990.
- Desenvolvida pelo National Center for Biotechnology Information (NCBI) e financiada 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 possivelmente é ainda o artigo mais citado do mundo).

BLAST

- 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 - Criação da Lista de Seeds

- 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 ajustado para se obter uma lista de sementes até 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 para 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

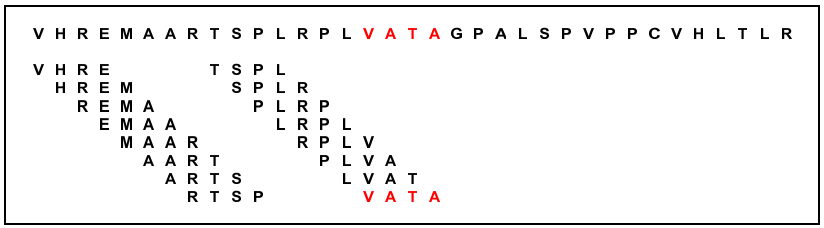
V H R E
  H R E M
    R E M A

BLAST - Criação da Lista de Seeds

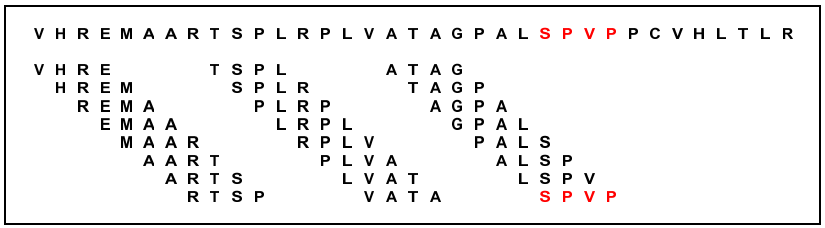
V H R E M A A R T S P L R P L V A T A G P A L S P V P P C V H L T L R

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

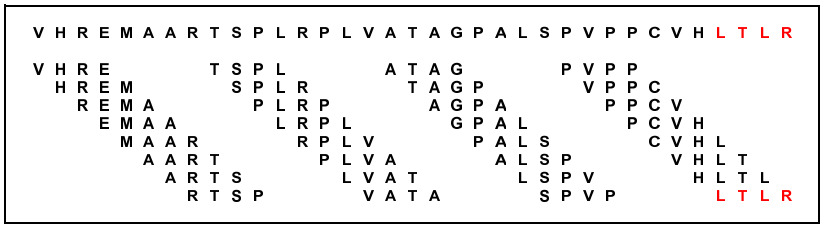
BLAST - Criação da Lista de Seeds



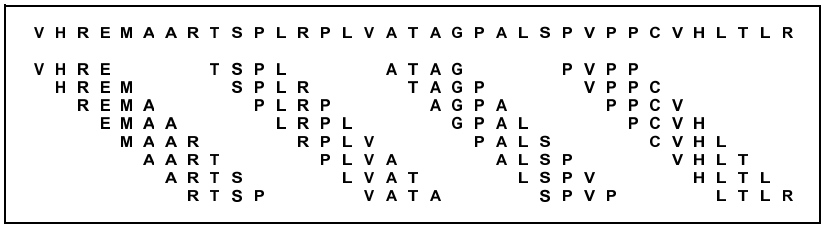
BLAST - Criação da Lista de Seeds



BLAST - Criação da Lista de Seeds



BLAST - Criação da Lista de Seeds



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

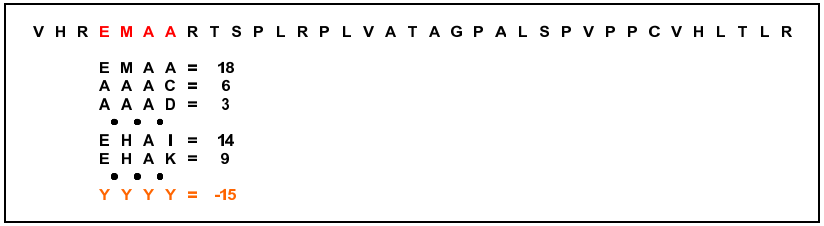
E M A A = 18
A A A C = 6

BLAST - Criação da Lista de Seeds

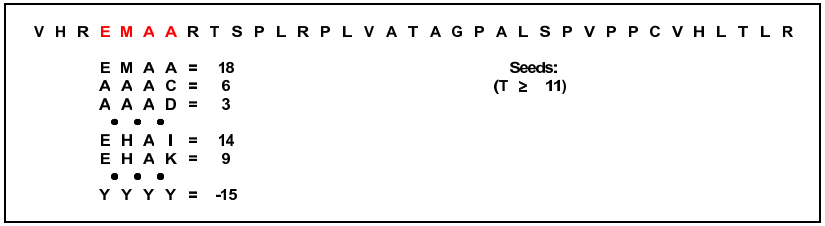
V H R E M A A R T S P L R P L V A T A G P A L S P V P P C V H L T L R

E M A A = 18
A A A C = 6
A A A D = 3

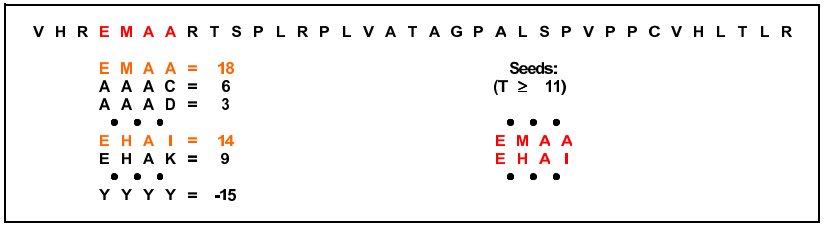
BLAST - Criação da Lista de Seeds



BLAST - Criação da Lista de Seeds



BLAST - Criação da Lista de Seeds



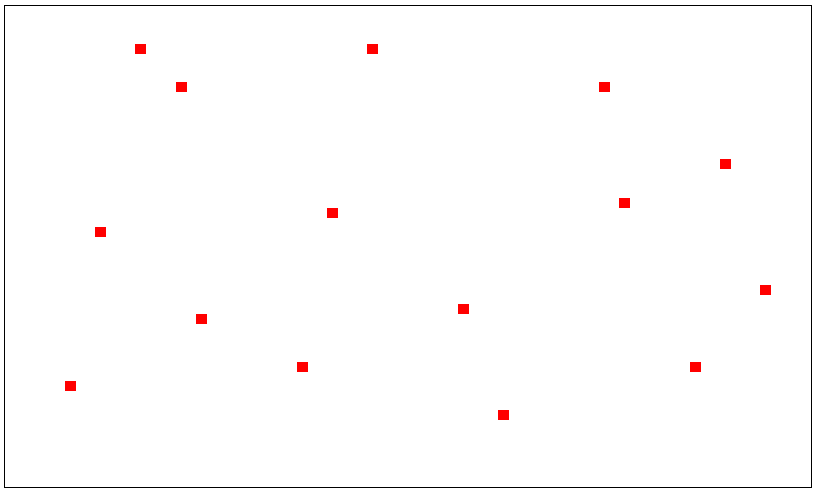
BLAST - Busca por Hits

- 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 sequência do banco (uma por uma).

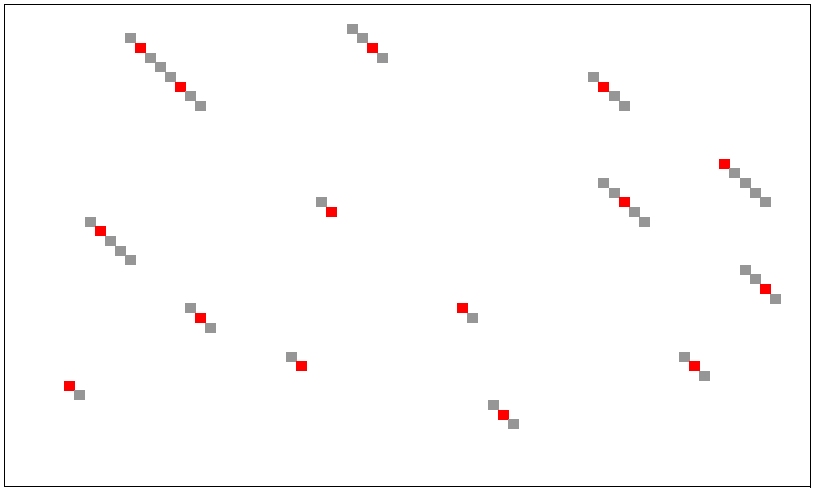
BLAST - Obtenção dos HSPs

- 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.
- Estimasse 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*.

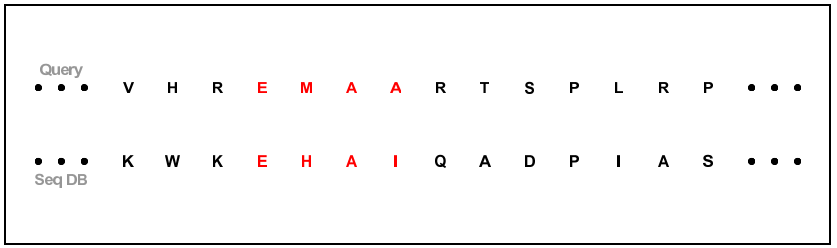
BLAST - Obtenção dos HSPs



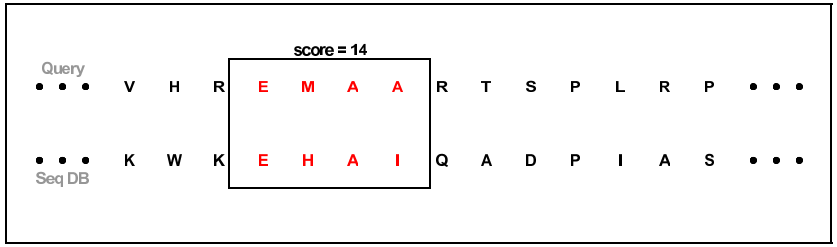
BLAST - Obtenção dos HSPs



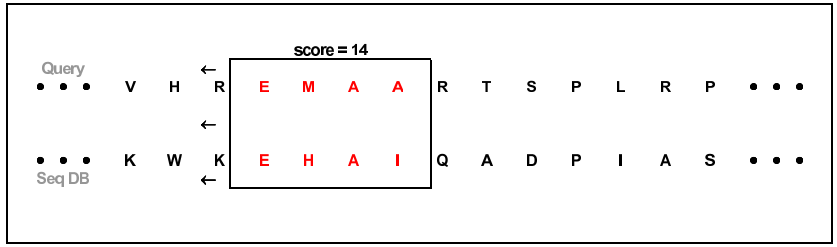
BLAST - Obtenção dos HSPs



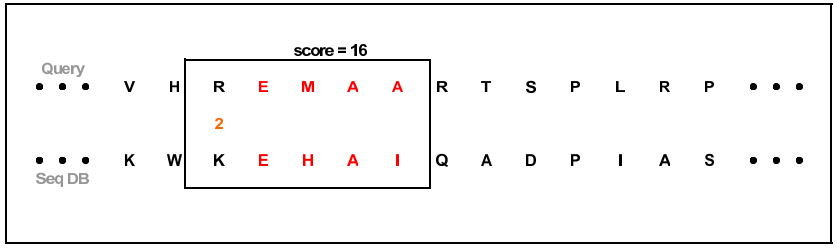
BLAST - Obtenção dos HSPs



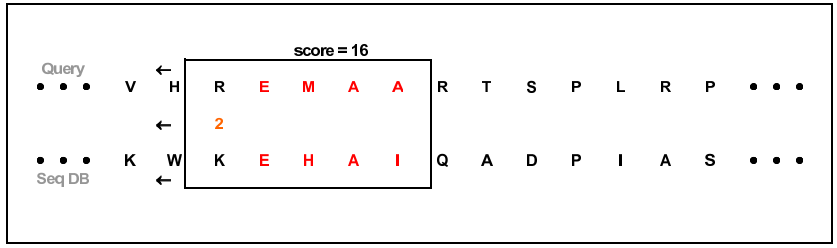
BLAST - Obtenção dos HSPs



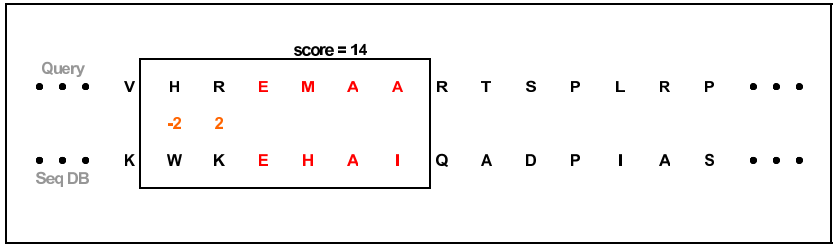
BLAST - Obtenção dos HSPs



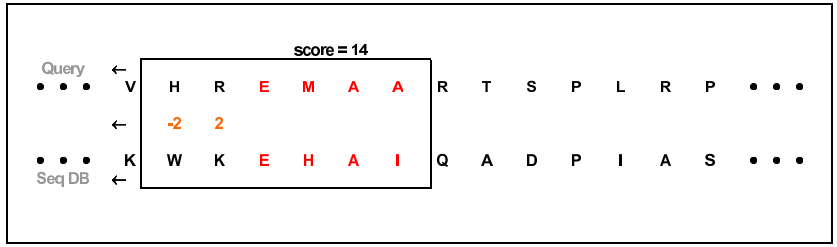
BLAST - Obtenção dos HSPs



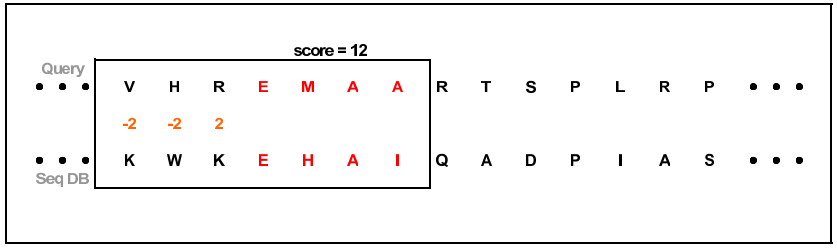
BLAST - Obtenção dos HSPs



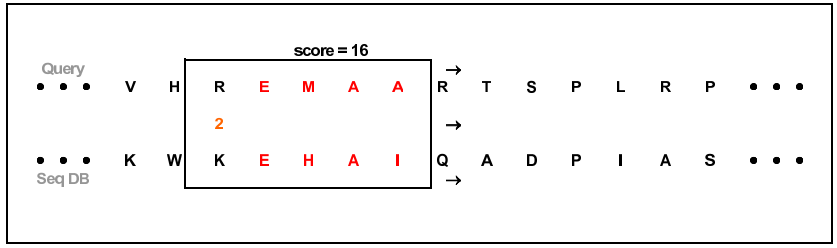
BLAST - Obtenção dos HSPs



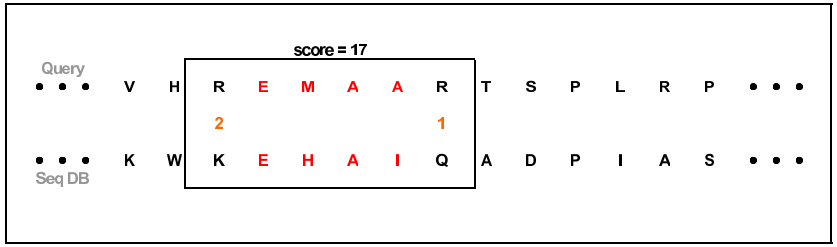
BLAST - Obtenção dos HSPs



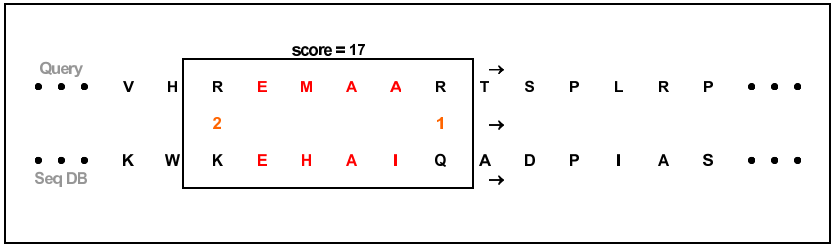
BLAST - Obtenção dos HSPs



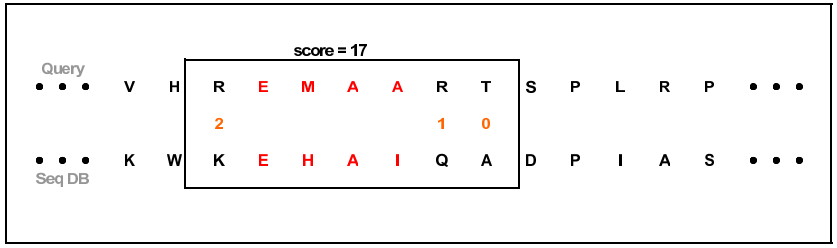
BLAST - Obtenção dos HSPs



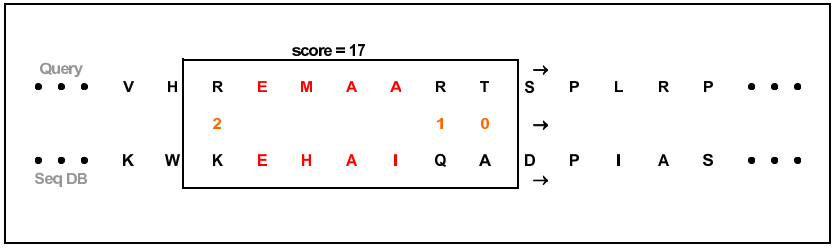
BLAST - Obtenção dos HSPs



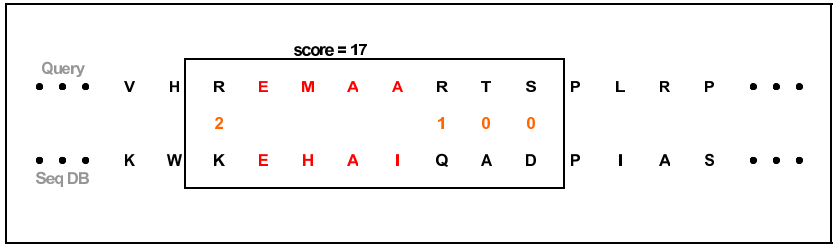
BLAST - Obtenção dos HSPs



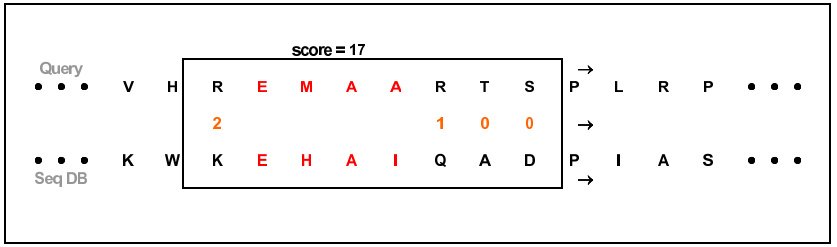
BLAST - Obtenção dos HSPs



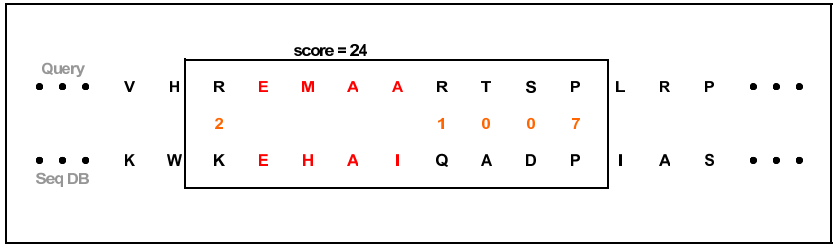
BLAST - Obtenção dos HSPs



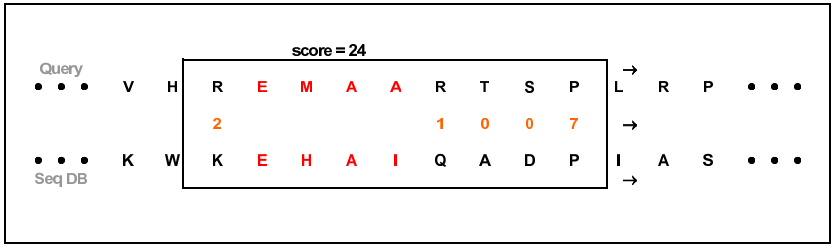
BLAST - Obtenção dos HSPs



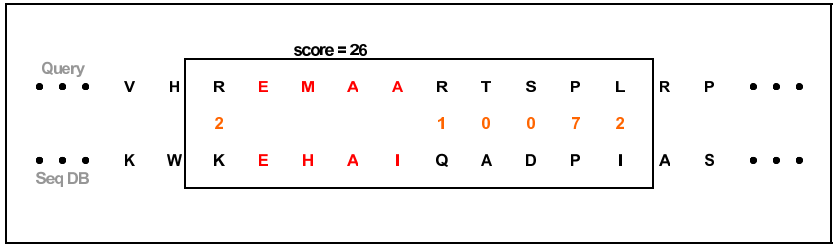
BLAST - Obtenção dos HSPs



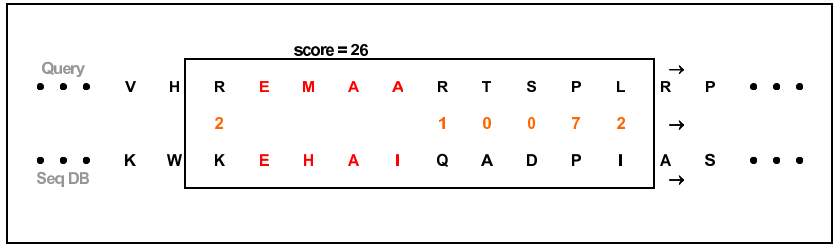
BLAST - Obtenção dos HSPs



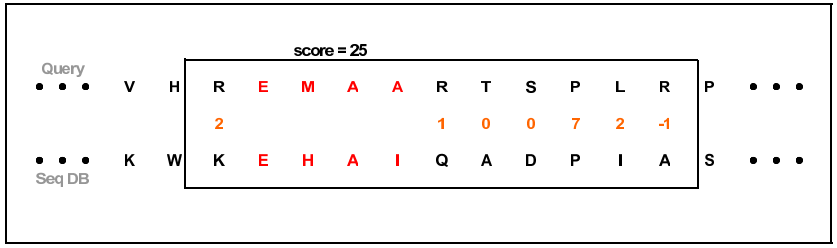
BLAST - Obtenção dos HSPs



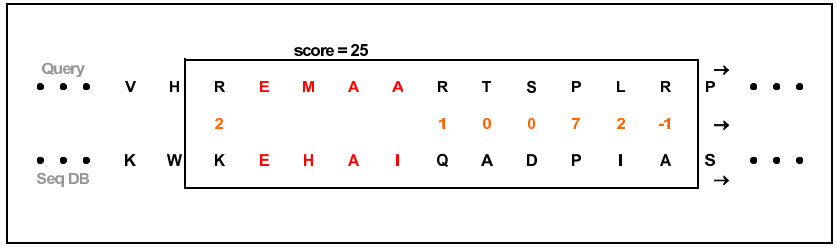
BLAST - Obtenção dos HSPs



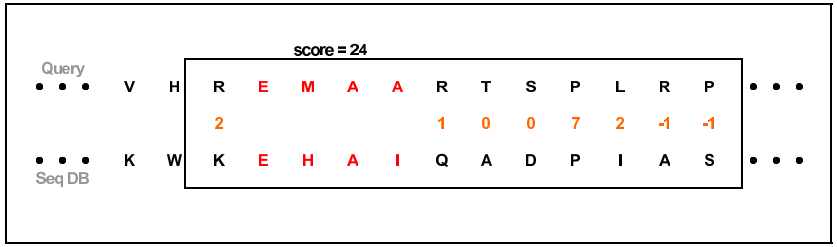
BLAST - Obtenção dos HSPs



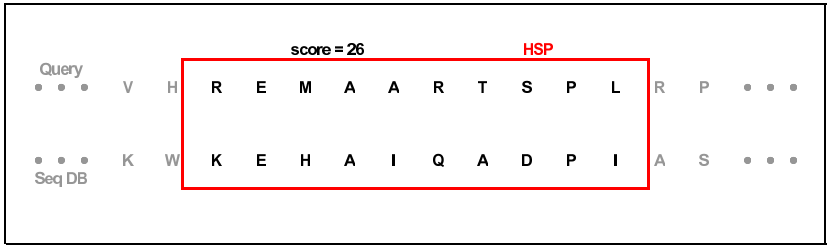
BLAST - Obtenção dos HSPs



BLAST - Obtenção dos HSPs



BLAST - Obtenção dos HSPs



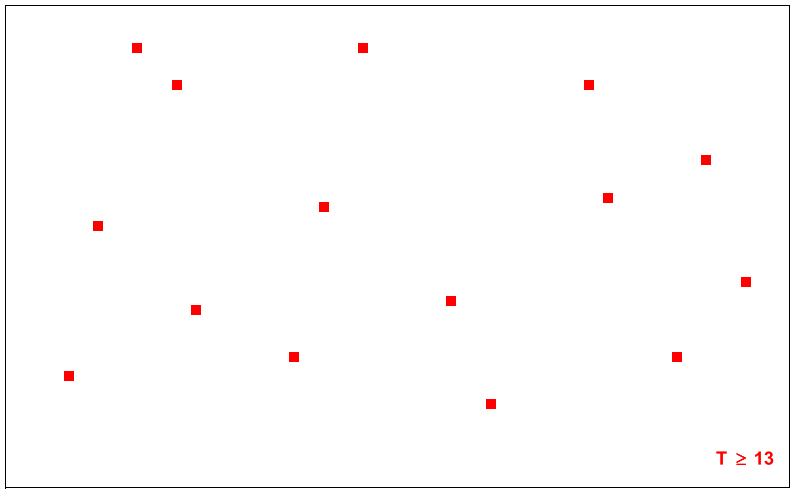
BLAST 2.0

- 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

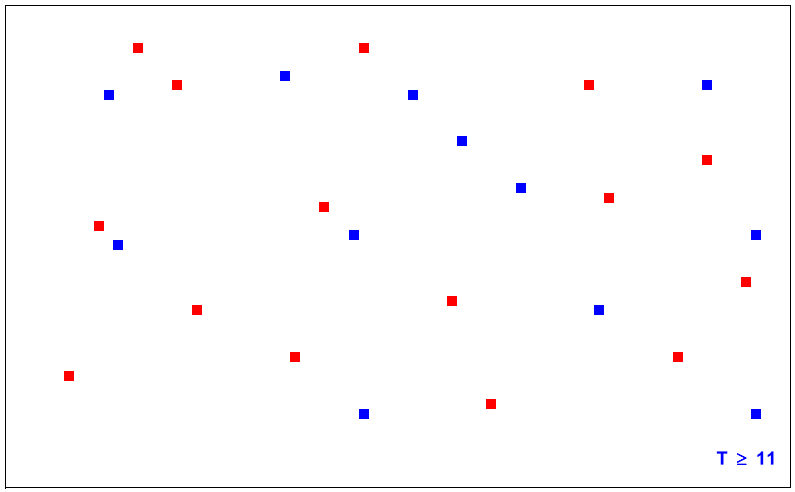
BLAST 2.0 - The Two-Hit Method

- 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 sobrepõe, 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.

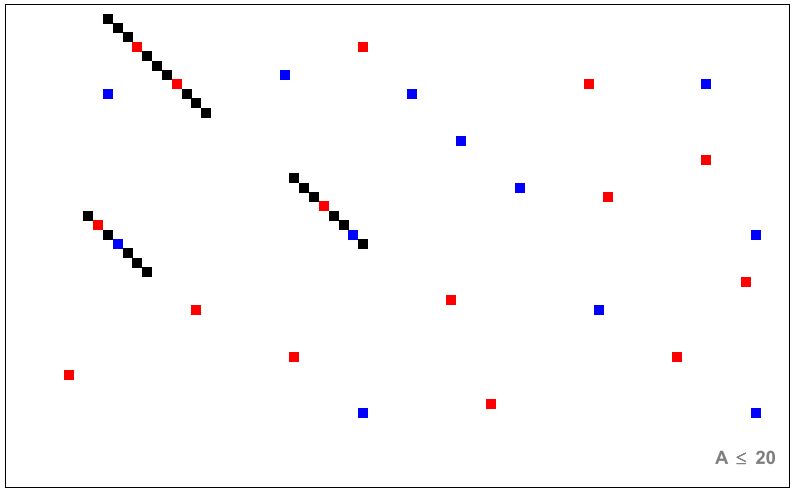
BLAST 2.0 - The Two-Hit Method



BLAST 2.0 - The Two-Hit Method



BLAST 2.0 - The Two-Hit Method



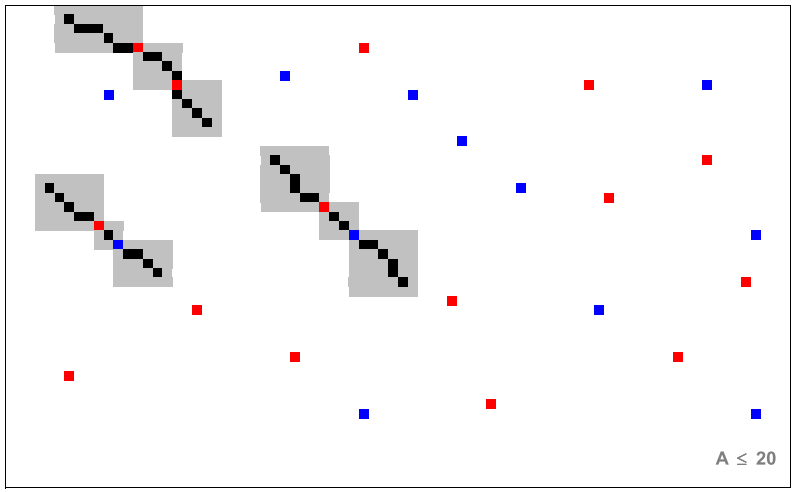
BLAST 2.0 - The Two-Hit Method

- 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



BLAST 2.0 - Complexidade

- Seja w o tamanho dos *seeds*, β é o tamanho do alfabeto ($\beta = 4$ para DNA, $\beta = 20$ 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*.
- Fase 1: Obter uma lista de *seeds*:
  - $O(nw\beta^w)$
- Fase 2: Procurar *hits* de *seeds* nas sequências do banco de dados:
  - $O(M)$
- Fase 3: Estender os *hits* para obter os alinhamentos (*HSPs*):
  - $O(hn^2)$
- Total:
  - $O(nw\beta^w + M + hn^2)$

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 λ e k são calculadas considerando a matriz de pontuação e a distribuição das bases no banco de dados.
- Não confundir E-value com P-value (probabilidade de se obter um alinhamento com uma dada pontuação). E-value pode ser um número maior que 1, enquanto o P-value é sempre um valor entre 0 e 1.

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.