

Comparaison de sequences

Dr. Bousbaa Fatima Zohra

f.bousbaaf@cu-aflou.edu.dz
Centre universitaire d'Aflou
Institut des sciences
Département d'Informatique

2024-2025



Plan

- 1 Alignement de séquences biologiques
 - Alignement de séquences biologiques : Pourquoi?
 - Activité de recherche en Bio-informatique
 - Principe d'alignement de séquences biologiques
- 2 Alignements de paires de séquences
 - Substitution
 - Matrice de points (Matrice de pixels, Dot-plot)
 - Programmation dynamique
 - Alignement global (Algorithme Needleman-Wunsch)
 - Alignement local
 - Alignement local (Algorithme Smith-Waterman)
- 3 Alignements multiples : Algorithmes d'alignement heuristiques

Alignement de séquences biologiques : Pourquoi?

Aligner=comparer

Buts de la comparaison en général :

Tout ce qui peut concerner la comparaison de textes, ou plus généralement de séquences linéaires

- Fouille de données (de grep à Google).
- Comparaison de fichiers.
- Linguistique comparée.
- Détection de plagiat.

Alignement de séquences biologiques : Pourquoi?

Aligner=comparer

Alignement

- Identifier les points communs ou les différences entre des séquences
- Retracer l'histoire évolutive des séquences en simulant les mutations:
 - Substitution : remplacement d'un nucléotide par un autre.
 - Insertion / Délétion : (ajout / suppression) d'un fragment d'ADN.
 - Duplication : doublement d'un fragment d'ADN.




Buts de la comparaison en Bio-informatique

Comparaisons de séquences protéiques ou nucléiques pour la détection de **l'homologie** et **la similarité**.

- Identification de séquences / prédiction de gènes.
- Prédiction de fonction ou domaines communs.
- Phylogénie, évolution.

Alignement de séquences biologiques : Pourquoi?

- La protéine de l'insuline de l'homme, de la souris et le dègue du Chili (Octodon degus) exercent la même fonction biologique mais ont des structures primaires qui ne sont pas totalement identiques !

<p><i>Homo sapiens</i></p> <p>AAA59172</p> <p>110 acides aminés</p>		<p>MALWMRLLPLLALLALWGPDPAAAF VNQHLCGSHLVEALYLVCGERGFFY TPKTRREAEDLQVGQVELGGGPGAG SLQPLALEGSLQKRGIVEQCCTSICSL YQLENYCN</p>
<p><i>Mus musculus</i></p> <p>AAI45871</p> <p>108 acides aminés</p>		<p>MALLVHFLPLLALLALWEPKPTQAF VKQHLCGPHLVEALYLVCGERGFFY TPKSRREVEDPQVEQLELGGSPGDL QTLALEVARQKRGIVDQCCTSICSL YQLENYCN</p>
<p><i>Octodon degus</i></p> <p>AAA40590</p> <p>109 acides aminés</p>		<p>MAPWMHLLTVLALLALWGPNSVQAY SSQHLCGSNLVEALYMTCSRSGFYR PHDRRELEDLQVEQAELGLEAGGLQ PSALEMILQKRGIVDQCCNNICTFN</p>

Activité quotidienne de milliers de Bio-informaticiens

- Domaine de recherche en informatique toujours actif, toujours demandeur de nouvelles fonctionnalités et performances, publications, conférences, etc.
- Découverte de nouvelles méthodes Bio-informatiques, en termes de rapidité de calcul et de consommation de ressources.

Principe d'alignement de séquences biologiques

Définition

Une **séquence** contient les informations sur le rôle biologique d'une macromolécule (ADN/ARN/Protéine) : Fonction, relation avec les autres cellules, etc.

- Du point de vue Bio-informaticien, une **séquence** biologique est un **mot**.
- Un **mot** (ou bien une chaîne) défini sur un alphabet \mathcal{A} est une suite finie de symboles juxtaposés de \mathcal{A} .

Principe d'alignement de séquences biologiques

- Du point de vue Bio-informaticien, une **séquence** biologique est un **mot**.
- Un **mot** (ou bien une chaîne) défini sur un alphabet \mathcal{A} est une suite finie de symboles juxtaposés de \mathcal{A} .
- Séquence nucléotidique (ADN) : 4 acides nucléiques

Exemple :

ADN : ATGCGATGCGTAGTATCTAGTGC
 $\mathcal{A} = \{A, T, G, C\}$

- Séquence protéique (protéine): 20 acides aminés

Exemple :

Protéine : ADGFHIKLPRSFGHYSRADKLMEWS
 $\mathcal{A} = \{A, D, G, F, H, I, K, L, P, R, S, Y, M, E, W\}$

Format d'une séquence

- La séquence est représentée sous un format donnée.

Format fasta

```
>sp|P05231|IL6_HUMAN Interleukin-6 precursor (IL-6) - Homo sapiens (Human).
atgtctcaacccgctggctcacttagcctg ggcgggcgccctgcacacgatgggtcacgccatcatggtgc
cgttgactaggcgagtgctcgccctgctcgtcctcttccattacctgtggtcgccgctgttcggcgcgactc
cgggtatctcgtggacgttagccatcgctctgctgacgattttcattcgtttgctgctggtgccgtgttcg
```

Format « raw ou brut »

```
atgtctcaacccgctggctcacttagcctg ggcgggcgccctgcacacgatgggtcacgccatcatggtgc
cgttgactaggcgagtgctcgccctgctcgtcctcttccattacctgtggtcgccgctgttcggcgcgactc
cgggtatctcgtggacgttagccatcgctctgctgacgattttcattcgtttgctgctggtgccgtgttcg
```

Format EMBL

```
ID  MMVASPHOS  standard; RNA; EST; 140 BP.
AC  X97897;
DE  M.musculus mRNA for protein homologous to vasodilator-stimulated
DE  phosphoprotein
SQ  Sequence 140 BP; 25 A; 58 C; 39 G; 17 T; 1 other;
    ttctccaga agctgactct atgngaccc cgagagagac tgagcagaac      60
    cccgcaccc ctgcacttcc atcagagggc gcccagggag cactcccgct      120
```

Principe d'alignement de séquences biologiques

En Bio-informatique, la comparaison des séquences repose essentiellement sur la notion de l'alignement, et permet de déterminer le degré de ressemblance entre celles-ci (similitude ou identité en révélant des régions proches dans leurs séquences primaires). Cela peut alors indiquer que :

- La structure (primaire, secondaire ou tertiaire) des deux séquences est semblable.
- La fonction biologique est proche ou différente (dans le cas de la dissémination).
- L'origine des séquences alignées est commune ou éloignée (notion d'homologie).

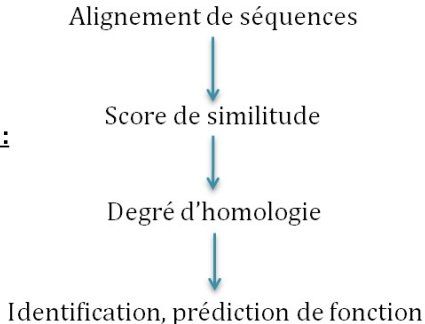
Cependant, la comparaison pour l'obtention d'un alignement optimal entre deux séquences biologiques, nécessite néanmoins la mise en oeuvre des **algorithmes** permettant de quantifier la notion de ressemblance entre ces séquences.

Principe d'alignement de séquences biologiques

La comparaison de séquences est l'outil central en Bio-informatique :

- Repose sur des calculs matriciels ou des algorithmes complexes qui rendent des résultats sous forme de données statistiques (score).
- Logiciel d'alignement le plus connu = BLAST (Basic Local Alignment Search Tool).

Démarche globale :



Traitement des séquences

Il existe différents algorithmes pour la détermination de segments identiques entre deux séquences biologiques (fenêtres ou de mots) sur lesquelles une similitude significative peut exister.

Notion de score

Le score élémentaire (noté \mathcal{S}) est une entité numérique que l'on attribue à chaque couple de nucléotides des deux séquences à comparer. Il peut prendre la valeur de **1** lorsque les deux nucléotides des deux séquences sont identiques, et la valeur de **0** sinon.

Exemple :

Séquence 1	A	C	A	C	G	T	C	A	G	G	C	T
Séquence 2	A	A	C	C	T	T	G	A	G	G	A	T
	1	0	0	1	0	1	0	1	1	1	0	1

$$\mathcal{S} = \sum_{i=1}^n \mathcal{S}_i = 7$$

Traitement des séquences

Il existe différents algorithmes pour la détermination de segments identiques entre deux séquences biologiques (fenêtres ou de mots) sur lesquelles une similitude significative peut exister.

Notion de score

Le score élémentaire (noté S) est une entité numérique que l'on attribue à chaque couple de nucléotides des deux séquences à comparer. Il peut prendre la valeur de **1** lorsque les deux nucléotides des deux séquences sont identiques, et la valeur de **0** sinon.

Evaluation de score

En pratique, plus le score d'alignement est élevé, plus les séquences sont similaires et présenteront des propriétés et des fonctions proches.

- Plus de **70%** de similarité permettent d'affirmer qu'il y a **homologie**.

Matrices de scores

Matrice d'identité nucléique

	A	T	G	C
A	1	0	0	0
T	0	1	0	0
G	0	0	1	0
C	0	0	0	1

Il existe une autre matrice de score, qui tient compte de l'analogie structurale entre purines (A et G) et pyrimidines (C, T et U) et affecte des scores en fonction de cette ressemblance.

Matrice de transition/transversion

	A	T	G	C
A	3	0	1	0
T	0	3	0	1
G	1	0	3	0
C	0	1	0	3

En Bio-informatique, on utilise beaucoup plus la matrice d'identité.

Matrices de scores

La matrice BLOSUM (Protéique)

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

Principe du calcul des scores d'alignement : Méthode 1

Séquence de Référence : S_R = ACACGTCAGGCT

Séquence 1 à analyser : S_1 = AACCTTGAGGAT

S_R	A	C	A	C	G	T	C	A	G	G	C	T
S_1	A	A	C	C	T	T	G	A	G	G	A	T

match **mismatch**
(identique) (Non-identique)

match = 1

mismatch = 0

Principe du calcul des scores d'alignement : Méthode 1

Séquence de Référence : S_R = ACACGTCAGGCT

Séquence 1 à analyser : S_1 = AACCTTGAGGAT

S_R	A	C	A	C	G	T	C	A	G	G	C	T
S_1	A	A	C	C	T	T	G	A	G	G	A	T

match **mismatch**
(identique) (Non-identique)

1	0	0	1	0	1	0	1	1	1	0	1
----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------

Score de S_1 = 7

Principe du calcul des scores d'alignement : Méthode 2

Séquence de Référence : S_R = ACACGTCAGGCT

Séquence 2 à analyser : S_2 = AACCTTGAT

Proposition 1 : S_2' = AACCTT_ _GAT

S_R	A	C	A	C	G	T	C	A	G	G	C	T
S_2'	A	A	C	C	T	T	-	-	-	G	A	T

match

mismatch

gap

(identique)

(Non-identique)

match = 1

mismatch = -1

gap = -2

		-	A	T	G	C
-	0	-2	-2	-2	-2	-2
A	-2	1	-1	-1	-1	-1
T	-2	-1	1	-1	-1	-1
G	-2	-1	-1	1	-1	-1
C	-2	-1	-1	-1	1	-1

Principe du calcul des scores d'alignement : Méthode 2

Séquence de Référence : S_R = ACACGTCAGGCT

Séquence 2 à analyser : S_2 = AACCTTGAT

Proposition 1 : S_2' = AACCTT_ _GAT

S_R	A	C	A	C	G	T	C	A	G	G	C	T
S_2'	A	A	C	C	T	T	-	-	-	G	A	T

match

mismatch

gap

(identique)

(Non-identique)

1	-1	-1	1	-1	1	-2	-2	-2	1	-1	1
----------	-----------	-----------	----------	-----------	----------	-----------	-----------	-----------	----------	-----------	----------

Score de S_2' = -5

Principe du calcul des scores d'alignement : Méthode 2

Séquence de Référence : S_R = ACACGTCAGGCT

Séquence 2 à analyser : S_2 = AACCTTGAT

Proposition 2 : S_2'' = AACCTTGAT_ _ _

S_R	A	C	A	C	G	T	C	A	G	G	C	T
S_2''	A	A	C	C	T	T	G	A	T	-	-	-

match

mismatch

gap

(identique)

(Non-identique)

1	-1	-1	1	-1	1	-1	1	-1	-2	-2	-2
----------	-----------	-----------	----------	-----------	----------	-----------	----------	-----------	-----------	-----------	-----------

Score de S_2'' = -7

Principe du calcul des scores d'alignement : Méthode 3

Séquence de Référence : S_R = ACACGTCAG

Séquence 3 à analyser : S_3 = ACACGTC

Proposition 1 : S_3' = ACACGTC_ _

S_R	A	C	A	C	G	T	C	A	G
S_3'	A	C	A	C	G	T	C	-	-

gap_open

gap_extend

match = 2

mismatch = 0

gap_open = -5

gap_extend = -1

	A	T	G	C
A	2	0	0	0
T	0	2	0	0
G	0	0	2	0
C	0	0	0	2

Principe du calcul des scores d'alignement : Méthode 3

Séquence de Référence : S_R = ACACGTCAG

Séquence 3 à analyser : S_3 = ACACGTC

Proposition 1 : S_3' = ACACGTC_ _

S_R	A	C	A	C	G	T	C	A	G
S_3'	A	C	A	C	G	T	C	-	-
								gap-open	gap-extend
								-5	-1

Score de S_3' = 8

Principe du calcul des scores d'alignement : Méthode 3

Séquence de Référence : S_R = ACACGTCAG

Séquence 3 à analyser : S_3 = ACACGTC

Proposition 1 : S_3'' = ACACGT_ C_

S_R	A	C	A	C	G	T	C	A	G
S_3''	A	C	A	C	G	T	-	C	-

							gap_open		gap_open
2	2	2	2	2	2		-5	0	-5

Score de S_3'' = 2

Problème : Comment nous pouvons trouver le meilleur score ?

Matrice de points (Matrice de pixels, Dot-plot)

Le Dot Plot est une représentation graphique simple pour la comparaison de séquences


Principe :

- Mettre les séquences le long des axes d'une matrice.
- Mettre un point (dot) là où il y a un match.
- Les diagonales relèvent les régions alignables (similaires) entre deux séquences.

Matrice de points (Matrice de pixels, Dot-plot) : Exemple

	A	D	S	Q	S	D	Q	S	A	Q	S
A											
D											
S											
Q											
S											
A											
Q											
D											
Q											
S											

match 

mismatch 

Fenêtre de taille 1

Matrice de points (Matrice de pixels, Dot-plot) : Exemple

	A	D	S	Q	S	D	Q	S	A	Q	S
A											
D											
S											
Q											
S											
A											
Q											
D											
Q											
S											

match 


mismatch 

Fenêtre de taille 2

Matrice de points (Matrice de pixels, Dot-plot) : Exemple

	A	D	S	Q	S	D	Q	S	A	Q	S
A											
D											
S											
Q											
S											
A											
Q											
D											
Q											
S											

match 


mismatch 

Fenêtre de taille 3

Matrice de points (Matrice de pixels, Dot-plot) : Exemple

	A	D	S	Q	S	D	Q	S	A	Q	S
A											
D											
S											
Q											
S											
A											
Q											
D											
Q											
S											

match 


mismatch 

Fenêtre de taille 4

Matrice de points (Matrice de pixels, Dot-plot) : Exemple

	A	D	S	Q	S	D	Q	S	A	Q	S
A											
D											
S											
Q											
S											
A											
Q											
D											
Q											
S											

match 

mismatch 

Fenêtre de taille 5

Avantage et inconvénients Dot-plot

- Avantages :
 - Simple et visuel.
 - Peu de perte d'information.
- Inconvénients :
 - Identification : pas de méthode de détection automatique
 - Interprétation : pas de mesure objective et d'évaluation de la pertinence de l'information.

Programmation dynamique

Programmation dynamique

La programmation dynamique résout les problèmes en combinant les solutions de sous-problèmes.

- Résoudre chaque sous-problème une seule fois.
- Mémoriser sa solution dans une matrice de la programmation dynamique (épargnant ainsi le recalcul de la solution chaque fois que le sous-problème est rencontré).

Calculs intermédiaires / résolution sous-problèmes

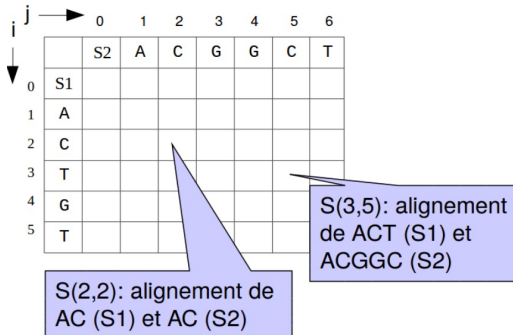
=

Calculs scores d'alignements entre préfixes

⇒ combinaison des sous-alignements précédents

Alignement global (Algorithme Needleman-Wunsch)

⇒ Application du principe de la programmation dynamique = combinaison des sous-alignements précédents.



Case (i, j) : score alignement entre les i premières bases de ACGGCT et les j premières bases de ACTGT.

Alignement global (Algorithme Needleman-Wunsch)

- S : une matrice de score ; \mathcal{G} : pénalité associée à un gap (indel).
- Initialisation :
 - $\mathcal{M}[0,0]=0$
 - $\mathcal{M}[i,0]=\mathcal{G} \times i$
 - $\mathcal{M}[0,j]=\mathcal{G} \times j$
- Remplissage :

$$\mathcal{M}[i,j] = \text{Max} \begin{cases} \mathcal{M}[i-1,j-1] + S[x_i, y_j] & \text{match/mismatch} \\ \mathcal{M}[i-1,j] + \mathcal{G} \\ \mathcal{M}[i,j-1] + \mathcal{G} \end{cases}$$

		j-1	j
i-1		$\mathcal{M}[i-1,j-1] + S[x_i, y_j]$	$\mathcal{M}[i-1,j] + \mathcal{G}$
i		$\mathcal{M}[i,j-1] + \mathcal{G}$	$\mathcal{M}[i,j]$

Alignement global (Algorithme Needleman-Wunsch) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -1

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

Alignement global (Algorithme Needleman-Wunsch) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -1

		A	C	G	G	C	T	A	T	C
	0									
A										
C										
T										
G										
T										
A										
A										
T										

Alignement global (Algorithme Needleman-Wunsch) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -1

		A	C	G	G	C	T	A	T	C
	0	-1								
A										
C										
T										
G										
T										
A										
A										
T										

Alignement global (Algorithme Needleman-Wunsch) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -1

		A	C	G	G	C	T	A	T	C
	0	-1	-2							
A										
C										
T										
G										
T										
A										
A										
T										

Alignement global (Algorithme Needleman-Wunsch) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -1

		A	C	G	G	C	T	A	T	C
	0	-1	-2	-3						
A										
C										
T										
G										
T										
A										
A										
T										

Alignement global (Algorithme Needleman-Wunsch) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -1

		A	C	G	G	C	T	A	T	C
	0	-1	-2	-3	-4					
A										
C										
T										
G										
T										
A										
A										
T										

Alignement global (Algorithme Needleman-Wunsch) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -1

		A	C	G	G	C	T	A	T	C
	0	-1	-2	-3	-4	-5				
A										
C										
T										
G										
T										
A										
A										
T										

Alignement global (Algorithme Needleman-Wunsch) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -1

		A	C	G	G	C	T	A	T	C
	0	-1	-2	-3	-4	-5	-6			
A										
C										
T										
G										
T										
A										
A										
T										

Alignement global (Algorithme Needleman-Wunsch) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -1

		A	C	G	G	C	T	A	T	C
	0	-1	-2	-3	-4	-5	-6	-7		
A										
C										
T										
G										
T										
A										
A										
T										

Alignement global (Algorithme Needleman-Wunsch) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -1

		A	C	G	G	C	T	A	T	C
	0	-1	-2	-3	-4	-5	-6	-7	-8	
A										
C										
T										
G										
T										
A										
A										
T										

Alignement global (Algorithme Needleman-Wunsch) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -1

		A	C	G	G	C	T	A	T	C
	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
A										
C										
T										
G										
T										
A										
A										
T										

Alignement global (Algorithme Needleman-Wunsch) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -1

		A	C	G	G	C	T	A	T	C
	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
A	-1									
C										
T										
G										
T										
A										
A										
T										

Alignement global (Algorithme Needleman-Wunsch) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -1

		A	C	G	G	C	T	A	T	C
	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
A	-1									
C	-2									
T										
G										
T										
A										
A										
T										

Alignement global (Algorithme Needleman-Wunsch) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -1

		A	C	G	G	C	T	A	T	C
	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
A	-1									
C	-2									
T	-3									
G										
T										
A										
A										
T										

Alignement global (Algorithme Needleman-Wunsch) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -1

		A	C	G	G	C	T	A	T	C
	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
A	-1									
C	-2									
T	-3									
G	-4									
T										
A										
A										
T										

Alignement global (Algorithme Needleman-Wunsch) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -1

		A	C	G	G	C	T	A	T	C
	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
A	-1									
C	-2									
T	-3									
G	-4									
T	-5									
A										
A										
T										

Alignement global (Algorithme Needleman-Wunsch) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -1

		A	C	G	G	C	T	A	T	C
	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
A	-1									
C	-2									
T	-3									
G	-4									
T	-5									
A	-6									
A										
T										

Alignement global (Algorithme Needleman-Wunsch) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -1

		A	C	G	G	C	T	A	T	C
	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
A	-1									
C	-2									
T	-3									
G	-4									
T	-5									
A	-6									
A	-7									
T										

Alignement global (Algorithme Needleman-Wunsch) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -1

		A	C	G	G	C	T	A	T	C
	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
A	-1									
C	-2									
T	-3									
G	-4									
T	-5									
A	-6									
A	-7									
T	-8									

Alignement global (Algorithme Needleman-Wunsch) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -1

		A	C	G	G	C	T	A	T	C
	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
A	-1	2								
C	-2									
T	-3									
G	-4									
T	-5									
A	-6									
A	-7									
T	-8									

Alignement global (Algorithme Needleman-Wunsch) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -1

		A	C	G	G	C	T	A	T	C
	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
A	-1	2	1							
C	-2									
T	-3									
G	-4									
T	-5									
A	-6									
A	-7									
T	-8									

Alignement global (Algorithme Needleman-Wunsch) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -1

		A	C	G	G	C	T	A	T	C
	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
A	-1	2	1	0						
C	-2									
T	-3									
G	-4									
T	-5									
A	-6									
A	-7									
T	-8									

Alignement global (Algorithme Needleman-Wunsch) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -1

		A	C	G	G	C	T	A	T	C
	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
A	-1	2	1	0	-1					
C	-2									
T	-3									
G	-4									
T	-5									
A	-6									
A	-7									
T	-8									

Alignement global (Algorithme Needleman-Wunsch) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -1

		A	C	G	G	C	T	A	T	C
	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
A	-1	2	1	0	-1	-2	-3	-4	-5	-6
C	-2									
T	-3									
G	-4									
T	-5									
A	-6									
A	-7									
T	-8									

Alignement global (Algorithme Needleman-Wunsch) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -1

		A	C	G	G	C	T	A	T	C
	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
A	-1	2	1	0	-1	-2	-3	-4	-5	-6
C	-2	1								
T	-3									
G	-4									
T	-5									
A	-6									
A	-7									
T	-8									

Alignement global (Algorithme Needleman-Wunsch) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -1

		A	C	G	G	C	T	A	T	C
	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
A	-1	2	1	0	-1	-2	-3	-4	-5	-6
C	-2	1	4							
T	-3									
G	-4									
T	-5									
A	-6									
A	-7									
T	-8									

Alignement global (Algorithme Needleman-Wunsch) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -1

		A	C	G	G	C	T	A	T	C
	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
A	-1	2	1	0	-1	-2	-3	-4	-5	-6
C	-2	1	4	3						
T	-3									
G	-4									
T	-5									
A	-6									
A	-7									
T	-8									

Alignement global (Algorithme Needleman-Wunsch) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -1

		A	C	G	G	C	T	A	T	C
	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
A	-1	2	1	0	-1	-2	-3	-4	-5	-6
C	-2	1	4	3	2					
T	-3									
G	-4									
T	-5									
A	-6									
A	-7									
T	-8									

Alignement global (Algorithme Needleman-Wunsch) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -1

		A	C	G	G	C	T	A	T	C
	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
A	-1	2	1	0	-1	-2	-3	-4	-5	-6
C	-2	1	4	3	2	1	0	-1	-2	-3
T	-3									
G	-4									
T	-5									
A	-6									
A	-7									
T	-8									

Alignement global (Algorithme Needleman-Wunsch) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -1

		A	C	G	G	C	T	A	T	C
	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
A	-1	2	1	0	-1	-2	-3	-4	-5	-6
C	-2	1	4	3	2	1	0	-1	-2	-3
T	-3	0								
G	-4									
T	-5									
A	-6									
A	-7									
T	-8									

Alignement global (Algorithme Needleman-Wunsch) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -1

		A	C	G	G	C	T	A	T	C
	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
A	-1	2	1	0	-1	-2	-3	-4	-5	-6
C	-2	1	4	3	2	1	0	-1	-2	-3
T	-3	0	3							
G	-4									
T	-5									
A	-6									
A	-7									
T	-8									

Alignement global (Algorithme Needleman-Wunsch) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -1

		A	C	G	G	C	T	A	T	C
	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
A	-1	2	1	0	-1	-2	-3	-4	-5	-6
C	-2	1	4	3	2	1	0	-1	-2	-3
T	-3	0	3	3						
G	-4									
T	-5									
A	-6									
A	-7									
T	-8									

Alignement global (Algorithme Needleman-Wunsch) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -1

		A	C	G	G	C	T	A	T	C
	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
A	-1	2	1	0	-1	-2	-3	-4	-5	-6
C	-2	1	4	3	2	1	0	-1	-2	-3
T	-3	0	3	3	2	1	3	2	1	0
G	-4									
T	-5									
A	-6									
A	-7									
T	-8									

Alignement global (Algorithme Needleman-Wunsch) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -1

		A	C	G	G	C	T	A	T	C
	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
A	-1	2	1	0	-1	-2	-3	-4	-5	-6
C	-2	1	4	3	2	1	0	-1	-2	-3
T	-3	0	3	3	2	1	3	2	1	0
G	-4	-1	2	5	5	4	3	2	1	0
T	-5									
A	-6									
A	-7									
T	-8									

Alignement global (Algorithme Needleman-Wunsch) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -1

		A	C	G	G	C	T	A	T	C
	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
A	-1	2	1	0	-1	-2	-3	-4	-5	-6
C	-2	1	4	3	2	1	0	-1	-2	-3
T	-3	0	3	3	2	1	3	2	1	0
G	-4	-1	2	5	5	4	3	2	1	0
T	-5	-2	1	4	4	4	6	5	4	3
A	-6	-3	0	3	3	3	5	8	7	6
A	-7	-4	-1	2	2	2	4	7	7	6
T	-8	-5	-2	1	1	1	4	6	9	8

$\mathcal{M}[8,9]$: score de l'alignement global entre ACGGCTATC et ACTGTAAT

Alignement global (Algorithme Needleman-Wunsch)

Procédure

Permet de trouver l'alignement en fonction de la matrice.

Score du meilleur alignement global \Rightarrow valeur de la case en bas à droite dans la matrice.

Fonctionnement

- 1 A partir de la cellule d'arrivée, remonter vers la(les) cellule(s) voisine(s) de score maximal et telle que :
Son score + la mutation correspondante = le score de la cellule courante
- 2 Itérer jusqu'à arriver à la cellule initiale.

Remarque

Si en une cellule, on peut revenir vers **plusieurs cellules voisines**, alors il existe **plusieurs chemins optimaux**.

Alignement global (Algorithme Needleman-Wunsch) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -1

		A	C	G	G	C	T	A	T	C
	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
A	-1	2	1	0	-1	-2	-3	-4	-5	-6
C	-2	1	4	3	2	1	0	-1	-2	-3
T	-3	0	3	3	2	1	3	2	1	0
G	-4	-1	2	5	5	4	3	2	1	0
T	-5	-2	1	4	4	4	6	5	4	3
A	-6	-3	0	3	3	3	5	8	7	6
A	-7	-4	-1	2	2	2	4	7	7	6
T	-8	-5	-2	1	1	1	4	6	9	8

C

Alignement global (Algorithme Needleman-Wunsch) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -1

		A	C	G	G	C	T	A	T	C
	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
A	-1	2	1	0	-1	-2	-3	-4	-5	-6
C	-2	1	4	3	2	1	0	-1	-2	-3
T	-3	0	3	3	2	1	3	2	1	0
G	-4	-1	2	5	5	4	3	2	1	0
T	-5	-2	1	4	4	4	6	5	4	3
A	-6	-3	0	3	3	3	5	8	7	6
A	-7	-4	-1	2	2	2	4	7	7	6
T	-8	-5	-2	1	1	1	4	6	9	8

T C
T

Alignement global (Algorithme Needleman-Wunsch) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -1

		A	C	G	G	C	T	A	T	C
	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
A	-1	2	1	0	-1	-2	-3	-4	-5	-6
C	-2	1	4	3	2	1	0	-1	-2	-3
T	-3	0	3	3	2	1	3	2	1	0
G	-4	-1	2	5	5	4	3	2	1	0
T	-5	-2	1	4	4	4	6	5	4	3
A	-6	-3	0	3	3	3	5	8	7	6
A	-7	-4	-1	2	2	2	4	7	7	6
T	-8	-5	-2	1	1	1	4	6	9	8

- T C
A T

Alignement global (Algorithme Needleman-Wunsch) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -1

		A	C	G	G	C	T	A	T	C
	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
A	-1	2	1	0	-1	-2	-3	-4	-5	-6
C	-2	1	4	3	2	1	0	-1	-2	-3
T	-3	0	3	3	2	1	3	2	1	0
G	-4	-1	2	5	5	4	3	2	1	0
T	-5	-2	1	4	4	4	6	5	4	3
A	-6	-3	0	3	3	3	5	8	7	6
A	-7	-4	-1	2	2	2	4	7	7	6
T	-8	-5	-2	1	1	1	4	6	9	8

A - T C

A A T

Alignement global (Algorithme Needleman-Wunsch) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -1

		A	C	G	G	C	T	A	T	C
	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
A	-1	2	1	0	-1	-2	-3	-4	-5	-6
C	-2	1	4	3	2	1	0	-1	-2	-3
T	-3	0	3	3	2	1	3	2	1	0
G	-4	-1	2	5	5	4	3	2	1	0
T	-5	-2	1	4	4	4	6	5	4	3
A	-6	-3	0	3	3	3	5	8	7	6
A	-7	-4	-1	2	2	2	4	7	7	6
T	-8	-5	-2	1	1	1	4	6	9	8

T A - T C
T A A T

Alignement global (Algorithme Needleman-Wunsch) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -1

		A	C	G	G	C	T	A	T	C
	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
A	-1	2	1	0	-1	-2	-3	-4	-5	-6
C	-2	1	4	3	2	1	0	-1	-2	-3
T	-3	0	3	3	2	1	3	2	1	0
G	-4	-1	2	5	5	4	3	2	1	0
T	-5	-2	1	4	4	4	6	5	4	3
A	-6	-3	0	3	3	3	5	8	7	6
A	-7	-4	-1	2	2	2	4	7	7	6
T	-8	-5	-2	1	1	1	4	6	9	8

C T A _ T C
T A A T

Alignement global (Algorithme Needleman-Wunsch) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -1

		A	C	G	G	C	T	A	T	C
	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
A	-1	2	1	0	-1	-2	-3	-4	-5	-6
C	-2	1	4	3	2	1	0	-1	-2	-3
T	-3	0	3	3	2	1	3	2	1	0
G	-4	-1	2	5	5	4	3	2	1	0
T	-5	-2	1	4	4	4	6	5	4	3
A	-6	-3	0	3	3	3	5	8	7	6
A	-7	-4	-1	2	2	2	4	7	7	6
T	-8	-5	-2	1	1	1	4	6	9	8

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

Alignement global (Algorithme Needleman-Wunsch) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -1

		A	C	G	G	C	T	A	T	C
	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
A	-1	2	1	0	-1	-2	-3	-4	-5	-6
C	-2	1	4	3	2	1	0	-1	-2	-3
T	-3	0	3	3	2	1	3	2	1	0
G	-4	-1	2	5	5	4	3	2	1	0
T	-5	-2	1	4	4	4	6	5	4	3
A	-6	-3	0	3	3	3	5	8	7	6
A	-7	-4	-1	2	2	2	4	7	7	6
T	-8	-5	-2	1	1	1	4	6	9	8

G G C T A _ T C
T G T A A T

Alignement global (Algorithme Needleman-Wunsch) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -1

		A	C	G	G	C	T	A	T	C
	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
A	-1	2	1	0	-1	-2	-3	-4	-5	-6
C	-2	1	4	3	2	1	0	-1	-2	-3
T	-3	0	3	3	2	1	3	2	1	0
G	-4	-1	2	5	5	4	3	2	1	0
T	-5	-2	1	4	4	4	6	5	4	3
A	-6	-3	0	3	3	3	5	8	7	6
A	-7	-4	-1	2	2	2	4	7	7	6
T	-8	-5	-2	1	1	1	4	6	9	8

C G G C T A - T C

C T G T A A T

Alignement global (Algorithme Needleman-Wunsch) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -1

		A	C	G	G	C	T	A	T	C
	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
A	-1	2	1	0	-1	-2	-3	-4	-5	-6
C	-2	1	4	3	2	1	0	-1	-2	-3
T	-3	0	3	3	2	1	3	2	1	0
G	-4	-1	2	5	5	4	3	2	1	0
T	-5	-2	1	4	4	4	6	5	4	3
A	-6	-3	0	3	3	3	5	8	7	6
A	-7	-4	-1	2	2	2	4	7	7	6
T	-8	-5	-2	1	1	1	4	6	9	8

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

Alignement global (Algorithme Needleman-Wunsch) : Exercice

match = 1

mismatch = 0

gap (\mathcal{G}) = 0

- Initialisation :
 - $\mathcal{M}[0,0]=0$
 - $\mathcal{M}[i,0]=\mathcal{G} \times i$
 - $\mathcal{M}[0,j]=\mathcal{G} \times j$
- Remplissage :

$$\mathcal{M}[i,j] = \text{Max} \begin{cases} \mathcal{M}[i-1,j-1] + \mathcal{S}[x_i,y_j] & \text{match/mismatch} \\ \mathcal{M}[i-1,j] + \mathcal{G} \\ \mathcal{M}[i,j-1] + \mathcal{G} \end{cases}$$

		G	G	C	T	G	A	C
G								
A								
T								
C								

Difficultés de l'alignement global (Algorithme Needleman-Wunsch)

- Nombre d'alignements possibles : *faramineux*.
- Mémoire et temps utilisés $\mathcal{O}(n \times m)$ c'est-à-dire proportionnels au produit de la longueur des séquences.

Alignement local

Problème

- Quelles sont les régions de forte similarité entre les deux séquences?
- Quel est le segment de séquence de score maximal?

Alignement local

match = 2

mismatch = -1

gap (\mathcal{G}) = -2

Deux séquences : GGCTGACCACCTT et GATCACTTCCATG

- Alignement global.

1 GGCTGACCACC-TT 13

| | | | | | Score = 5

1 GA-TCACTTCCATG 13

- Alignement local.

5 GACCACCTT 13

| | | | | | Score = 11

1 GATCAC-TT 8

Alignement local (Algorithme Smith-Waterman)

Principe

Alignement des meilleures sous-séquences.

... GQVARYAG	<table border="1"> <tr> <td>E</td><td>K</td><td>L</td><td>F</td><td>H</td><td>S</td><td>I</td><td>F</td><td>V</td><td>E</td> </tr> <tr> <td> </td><td> </td><td> </td><td> </td><td></td><td></td><td></td><td></td><td> </td><td> </td> </tr> <tr> <td>E</td><td>K</td><td>L</td><td>F</td><td>V</td><td>-</td><td>L</td><td>R</td><td>V</td><td>E</td> </tr> </table>	E	K	L	F	H	S	I	F	V	E											E	K	L	F	V	-	L	R	V	E	QNIFSLT ...
E	K	L	F	H	S	I	F	V	E																							
E	K	L	F	V	-	L	R	V	E																							
... TEXLINYI		LAESAS ...																														

- Evaluation d'une *ressemblance locale* entre deux séquences.
- Recherche de la région de plus forte similarité.

Algorithme

- L'algorithme d'alignement local de Smith-Waterman est basé sur l'algorithme introduit par Needleman-Wunsch.
- Score max avec possibilité de remise à zéro.

Alignement local (Algorithme Smith-Waterman)

- S : une matrice de score ; \mathcal{G} : pénalité associée à un gap (indel).
- Initialisation :
 - $\mathcal{M}[i,0]=0$
 - $\mathcal{M}[0,j]=0$
- Remplissage :

$$\mathcal{M}[i,j] = \text{Max} \begin{cases} \mathcal{M}[i-1,j-1] + S[xi,yj] & \text{match/mismatch} \\ \mathcal{M}[i-1,j] + \mathcal{G} \\ \mathcal{M}[i,j-1] + \mathcal{G} \\ \mathbf{0} \end{cases}$$

		j-1	j
i-1		$\mathcal{M}[i-1,j-1] + S[xi,yj]$	$\mathcal{M}[i-1,j] + \mathcal{G}$
i		$\mathcal{M}[i,j-1] + \mathcal{G}$	$\mathcal{M}[i,j]$

Alignement local (Algorithme Smith-Waterman) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -2

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

Alignement local (Algorithme Smith-Waterman) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -2

		A	C	G	G	C	T	A	T
	0	0	0	0	0	0	0	0	0
A	0								
G	0								
C	0								
T	0								
T	0								
T	0								
C	0								

Alignement local (Algorithme Smith-Waterman) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -2

		A	C	G	G	C	T	A	T
	0	0	0	0	0	0	0	0	0
A	0	2	0	0	0	0	0	2	0
G	0								
C	0								
T	0								
T	0								
T	0								
C	0								

Alignement local (Algorithme Smith-Waterman) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -2

		A	C	G	G	C	T	A	T
	0	0	0	0	0	0	0	0	0
A	0	2	0	0	0	0	0	2	0
G	0	0	1	2	2	0	0	0	1
C	0	0	2	0	1	4	2	0	0
T	0	0	0	1	0	2	6	4	2
T	0	0	0	0	0	0	4	5	6
T	0	0	0	0	0	0	2	3	7
C	0	0	2	0	0	2	0	1	5

Alignement local (Algorithme Smith-Waterman) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -2

		A	C	G	G	C	T	A	T
	0	0	0	0	0	0	0	0	0
A	0	2	0	0	0	0	0	2	0
G	0	0	1	2	2	0	0	0	1
C	0	0	2	0	1	4	2	0	0
T	0	0	0	1	0	2	6	4	2
T	0	0	0	0	0	0	4	5	6
T	0	0	0	0	0	0	2	3	7
C	0	0	2	0	0	2	0	1	5

- Traceback à partir du *meilleur score dans toute la matrice = valeur max dans la matrice*

Alignement local (Algorithme Smith-Waterman) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -2

		A	C	G	G	C	T	A	T
	0	0	0	0	0	0	0	0	0
A	0	2	0	0	0	0	0	2	0
G	0	0	1	2	2	0	0	0	1
C	0	0	2	0	1	4	2	0	0
T	0	0	0	1	0	2	6	4	2
T	0	0	0	0	0	0	4	5	6
T	0	0	0	0	0	0	2	3	7
C	0	0	2	0	0	2	0	1	5

T
T

Alignement local (Algorithme Smith-Waterman) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -2

		A	C	G	G	C	T	A	T
	0	0	0	0	0	0	0	0	0
A	0	2	0	0	0	0	0	2	0
G	0	0	1	2	2	0	0	0	1
C	0	0	2	0	1	4	2	0	0
T	0	0	0	1	0	2	6	4	2
T	0	0	0	0	0	0	4	5	6
T	0	0	0	0	0	0	2	3	7
C	0	0	2	0	0	2	0	1	5

A T
T T

Alignement local (Algorithme Smith-Waterman) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -2

		A	C	G	G	C	T	A	T
	0	0	0	0	0	0	0	0	0
A	0	2	0	0	0	0	0	2	0
G	0	0	1	2	2	0	0	0	1
C	0	0	2	0	1	4	2	0	0
T	0	0	0	1	0	2	6	4	2
T	0	0	0	0	0	0	4	5	6
T	0	0	0	0	0	0	2	3	7
C	0	0	2	0	0	2	0	1	5

T A T
T T T

Alignement local (Algorithme Smith-Waterman) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -2

		A	C	G	G	C	T	A	T
	0	0	0	0	0	0	0	0	0
A	0	2	0	0	0	0	0	2	0
G	0	0	1	2	2	0	0	0	1
C	0	0	2	0	1	4	2	0	0
T	0	0	0	1	0	2	6	4	2
T	0	0	0	0	0	0	4	5	6
T	0	0	0	0	0	0	2	3	7
C	0	0	2	0	0	2	0	1	5

C T A T
C T T T

Alignement local (Algorithme Smith-Waterman) : Exemple

match = 2

mismatch = -1

gap (\mathcal{G}) = -2

		A	C	G	G	C	T	A	T
	0	0	0	0	0	0	0	0	0
A	0	2	0	0	0	0	0	2	0
G	0	0	1	2	2	0	0	0	1
C	0	0	2	0	1	4	2	0	0
T	0	0	0	1	0	2	6	4	2
T	0	0	0	0	0	0	4	5	6
T	0	0	0	0	0	0	2	3	7
C	0	0	2	0	0	2	0	1	5

G C T A T
G C T T T

Difficultés des alignements global et local

- Les algorithmes de Needleman-Wunsch et Smith-Waterman donnent des solutions exactes mais ils ne sont pas très rapides (complexité $\mathcal{O}(n \times m)$)
- Comparer une séquence de 1000pb à une BD $\Rightarrow \sim 10^{11}$ cellules à évaluer
Calcul 10 millions de cellules par sec. $\Rightarrow 10^4$ sec = ≈ 3 heures.
Nécessité d'algorithmes plus rapides
- Algorithmes **heuristiques**, les plus connus : **BLAST** et **FASTA**.

Alignements multiples : Algorithmes d'alignement heuristiques

Algorithmes **heuristiques**, les plus connus : **BLAST** [Altschul et al] et **FASTA** [Pearson et Lipman].

- Comparaison d'une séquence requête avec toutes les séquences d'une BD.
- Score (normalisé) = mesure de similarité entre les séquences.
- Idée : les séquences similaires ont des segments communs de taille K quasiment identiques, ces petits segments vont servir de graines.

Alignements multiples : Algorithmes d'alignement heuristiques