


Comparaison de 2 séquences

UE ASM1

O. Lecompte
Laboratoire de Bioinformatique et de Génomique Intégratives – IGBMC

<http://www-bio3d-igbmc.u-strasbg.fr/~lecompte/enseignement.html>



Comparaison de 2 séquences

- Introduction :
 - notions de base
 - systèmes de scores (matrices de scores, pénalités de gaps)
- Comparaison par la matrice de points (dotplot)
- Alignements optimaux

Odile Lecompte -IGBMC ASM1

Comparaison – Pour quoi faire ?

- localiser un gène sur un génome
- recherche de similarité dans une banque
transférer des informations connues sur une nouvelle séquence
- recherche de domaines ou motifs conservés
identification de résidus importants pour la structure ou la fonction
- reconstruire l'histoire évolutive d'une famille de gènes...


information sur la fonction, la structure et l'évolution des molécules

Odile Leconte -IGBMC ASMI

Similarité et homologie

Similarité

- mesurée en % d'identité ou en % de similarité
- la similarité n'a pas directement de connotation évolutive

□ 2 séquences peuvent être similaires

- ⊗ *par hasard*
ex: région de faible complexité
- ⊗ *par évolution convergente*
ex: subtilisine et chymotrypsine sont 2 serine protéases avec la même triade catalytique (Ser, His, Asp)
mais séquences et structures 3D différentes
- ⊗ *par évolution divergente*
elles dérivent d'une même séquence ancestrale

Odile Leconte -IGBMC ASMI

Similarité et homologie

Homologie : 2 séquences sont homologues si elles descendent d'un ancêtre commun



il n'existe pas de pourcentage d'homologie !!!



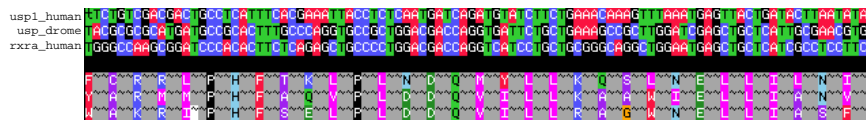
- Une similarité significative (>20% d'identité) est généralement le signe d'une homologie
- Une similarité non significative ne veut pas dire que les séquences ne sont pas homologues

Ex: myoglobine des mammifères et leghémoglobine des plantes
séquences : moins de 20% d'identité
même structure

Séquences nucléiques ou protéiques ?

- alphabet réduit dans le cas des séquences nucléiques
 - distinction entre le signal et le bruit plus difficile
- redondance du code génétique
 - mutations silencieuses → similarité plus longtemps observable au niveau des séquences protéiques

Alignement d'une région conservée de 3 récepteurs nucléaires au niveau nucléotidique et protéique



Les différents types d'alignements

Alignement deux à deux (pairwise alignment):

ex : recherche de similarité dans une banque
=> Fasta, Blast

Alignement multiple :

ex : alignement d'une famille de protéines
=> clustalx, DbClustal, mafft, multiZ

Alignement global

sur la totalité de la longueur des séquences

Alignement local

alignement de la ou des régions les plus fortement conservées
=> Intéressant si les séquences ne sont pas colinéaires

Odile Leconte -IGBMC

ASMI

Alignement et scores

mésappariement (mismatch)

```
taccagatcgggcggtat
||| ||| |||
taccagcatcgggcggtta
```

appariement (match)

9 bases identiques

lacune (gap)

```
taccag-atcgggcggtat
||| ||| ||| ||| ||| |||
taccagcatcgggcggtta-
```

17 bases identiques

- introduction de gap(s) pour augmenter les appariements
- évaluation de la similarité entre les séquences
⇒ calcul du **score** de l'alignement

$$\text{Score} = \sum se - \sum sp$$

score élémentaire
⇒ matrice de scores

pénalité liée à l'introduction de gap
(*gap penalty*)

Odile Leconte -IGBMC

ASMI

Les matrices nucléiques

Matrice d'identité utilisée dans blastn

La valeur du score élémentaire est de :

- 1 entre bases identiques
- 2 entre bases différentes (« mismatch »)

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

Autres matrices

D'autres matrices privilégient certaines mutations.

ex : transitions moins pénalisées que les transversions

	A	C	G	T
A	1.36	-1.60	-0.37	-1.60
C	-1.60	1.36	-1.60	-0.37
G	-0.37	-1.60	1.36	-1.60
T	-1.60	-0.37	-1.60	1.36

Transitions : substitution purine ↔ purine (A ↔ G) ou pyrimidine ↔ pyrimidine (C ↔ T)

Transversions : substitution purine ↔ pyrimidine (A ↔ C ou A ↔ T ou G ↔ C ou G ↔ T)

Les matrices protéiques

Les matrices de substitution :

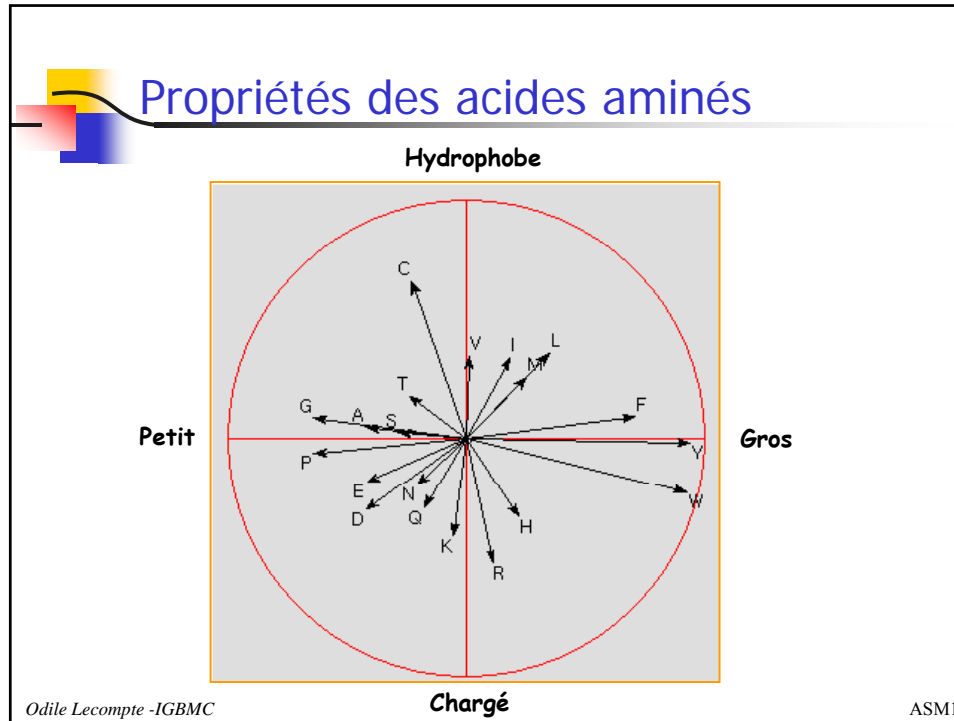
valeurs proportionnelles aux probabilités qu'un acide aminé i soit remplacé par un acide aminé j

score élevé pour les substitutions

- fréquemment observées
- ne modifiant pas la structure ni la fonction de la protéine (substitutions **conservatives**)



Exemple de matrices : PAM, BLOSUM, GONNET



Les matrices protéiques

basées sur :

- **le code génétique (Minimum Mutation Matrix)**
 nb de changements nécessaires au niveau nucléique pour passer d'un acide aminé à un autre
 ex : matrice de Fitch, 1966
- **les propriétés physico-chimiques des acides aminés**
 ex : matrice de Levitt (1976) basée sur des mesures d'hydrophobicité des acides aminés
- **les substitutions « observées » au cours de l'évolution**
 ex : les matrices PAM, BLOSUM, GONNET
 => **les plus couramment utilisées**

Odile Lecompte -IGBMC ASMI

Matrices de Dayhoff (matrices PAM)

■ Référence

Dayhoff *et al.* *A Model of Evolutionary Change in Proteins.*
In Atlas of Protein Sequence and Structure. Volume 5 (sup. 3):345-352 (1978)

■ Alignements utilisés

- 71 familles de protéines très proches : - **de 15 % de différence**
- Alignement global pour chaque famille
- Arbre phylogénétique pour chaque famille
- Reconstitution des séquences ancestrales selon le principe de parcimonie (minimum de changement)
- Décompte des substitutions permettant de passer des séquences ancestrales aux séquences observées

Odile Lecompte -IGBMC

ASMI

La série des PAM

■ basées sur :

- les substitutions observées entre aa dans le set de référence
- la mutabilité et fréquence relatives des résidus

Relative Mutabilities of the Amino Acids			
Asn	134	His	66
Ser	120	Arg	65
Asp	106	Lys	56
Glu	102	Pro	56
Ala	100	Gly	49
Thr	97	Tyr	41
Ile	96	Phe	41
Met	94	Leu	40
Gln	93	Cys	20
Val	74	Trp	18

The value for Ala has been arbitrarily set at 100

f_i : fréquence relative des aa dans les données utilisées

Gly	0.089	Arg	0.041
Ala	0.087	Asn	0.040
Leu	0.085	Phe	0.040
Lys	0.081	Gln	0.038
Ser	0.070	Ile	0.037
Val	0.065	His	0.034
Thr	0.058	Cys	0.033
Pro	0.051	Tyr	0.030
Glu	0.050	Met	0.015
Asp	0.047	Trp	0.010

- PAM1 : la probabilité globale de mutation est ~1% pour chaque aa
- Obtention de la série : $PAM-N = (PAM-1)^N$ avec N, nb de mutations pour 100 résidus

Odile Lecompte -IGBMC

ASMI

Les matrices de Dayhoff

La PAM-250 constitue un optimum pour des séquences divergentes

Pour N=250 (250 mutations pour une séquence de 100 résidus) :
seul un acide aminé sur 5 reste inchangé (divergence ~ 80 %)

Pam value

Odile Lecompte -IGBMC ASMI

Les matrices de Dayhoff

Log odds matrix for 250 PAMs.

C Cys	12																				
S Ser	0	2																			
T Thr	-2	1	3																		
P Pro	-3	1	0	6																	
A Ala	-2	1	1	1	2																
G Gly	-3	1	0	-1	1	5															
K Asn	-4	1	0	-1	0	0	2														
D Asp	-5	0	0	-1	0	1	2	4													
E Glu	-5	0	0	-1	0	0	1	3	4												
Q Gln	-5	-1	-1	0	0	-1	1	2	2	4											
H His	-3	-1	-1	0	-1	-2	2	1	1	3	6										
R Arg	-4	0	-1	0	-2	-3	0	-1	-1	1	2	6									
K Lys	-5	0	0	-1	-1	-2	1	0	0	1	0	3	5								
M Met	-5	-2	-1	-2	-1	-3	-2	-3	-2	-1	-2	0	0	6							
I Ile	-2	-1	0	-2	-1	-3	-2	-2	-2	-2	-2	-2	-2	2	5						
L Leu	-6	-3	-2	-3	-2	-4	-3	-4	-3	-2	-2	-3	-3	4	2	6					
V Val	-2	-1	0	-1	0	-1	-2	-2	-2	-2	-2	-2	-2	2	4	2	4				
F Phe	-4	-3	-3	-5	-4	-5	-4	-6	-5	-5	-2	-4	-5	0	1	2	-1	9			
Y Tyr	0	-3	-3	-5	-3	-5	-2	-4	-4	-4	0	-4	-4	-2	-1	-1	-2	7	10		
W Trp	-8	-2	-5	-6	-6	-7	-4	-7	-7	-5	-3	2	-3	-4	-5	-2	-6	0	0	12	
	C	S	T	P	A	G	N	D	E	Q	H	R	K	M	I	L	V	F	Y	W	
	Cys	Ser	Thr	Pro	Ala	Gly	Asn	Asp	Glu	Gln	His	Arg	Lys	Met	Ile	Leu	Val	Phe	Tyr	Trp	

Elements are shown multiplied by 10

Odile Lecompte -IGBMC ASMI

Les matrices BLOSUM

BLOSUM = BLOcks SUBstitution Matrix

- **Référence**
Henikoff and Henikoff. *Amino acid substitution matrices from protein blocks.*
PNAS 89:10915-10919 (1992)
- **Alignements utilisés**
BLOCKS : banques d'alignements multiples de régions conservées (sans gap)

Plus de 2000 « blocs » ont été utilisés

```

WNAGMIPVPYV
WCEAGWVPSNYI
WVKGMFPRNYV
WVRAGYIPSNYV
WMYGMLPANYV
WVRGLWVPSNYV
WLNAGDFPGTYV

```

Exemple de segments similaires

Odile Lecompte -IGBMC ASMI

Les matrices BLOSUM

Matrice logarithmique de probabilités :

$$s_{ij} = \log_2 (q_{ij}/e_{ij})$$

↑

fréquence de substitution observée

↑

fréquence de substitution attendue

$s_{ij} = 0$	fréquences observées identiques aux fréquences attendues
$s_{ij} > 0$	- supérieures -
$s_{ij} < 0$	- inférieures -

Odile Lecompte -IGBMC ASMI

La série des matrices BLOSUM

Les matrices BLOSUM diffèrent par le **pourcentage d'identité** entre les séquences utilisées.

Ex : la matrice **BLOSUM 62**
calculée sur des blocs de séquences ayant **au moins 62%** d'identité

☞ matrices les plus utilisées dans les alignements de paires de séquences

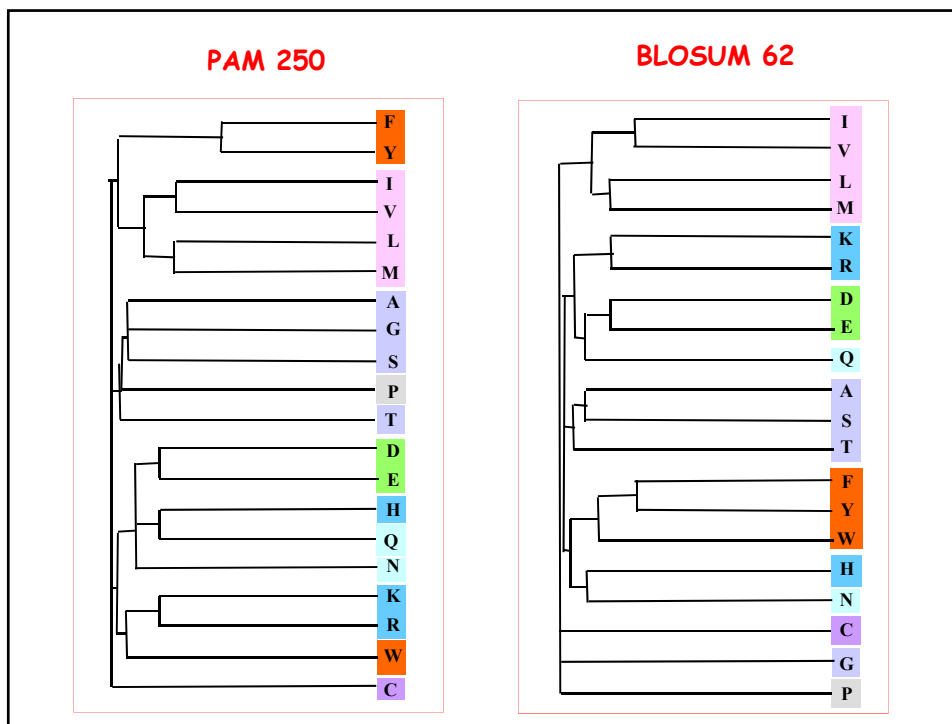
Odile Lecompte -IGBMCASMI

Les matrices Blosum

The log odds matrix for BLOSUM 62

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

Odile Lecompte -IGBMCASMI



Choix d'une matrice protéique

- **Etudes comparatives**

Il n'existe **pas de matrice idéale** ! Le choix dépend du programme utilisé et des séquences à comparer.

Les matrices basées directement sur la comparaison de séquences (BLOSUM, GONET) semblent globalement donner de meilleurs résultats que les matrices issues du modèle de Dayhoff.

Séquences proches et courtes : matrices BLOSUM élevées ou PAM faibles
Séquences divergentes et longues : matrices BLOSUM faibles ou PAM élevées

BLOSUM 80	BLOSUM 62	BLOSUM 45
PAM 1	PAM 120	PAM 250
Less divergent	←————→	More divergent

Odile Lecompte -IGBMC ASMI

Pénalisation des gaps

L'introduction d'un gap dans un alignement représente une insertion ou une délétion
=> Coût qu'il faut estimer

Systèmes de pénalités :

- 1) Pénalité fixe par gap
- 2) Pénalité variable selon la longueur du gap (affine)

$P = x + yL$ avec L longueur du gap

x : pénalité d'ouverture de gap (gap opening penalty=GOP)

y : pénalité d'extension de gap (gap extension penalty=GEP)

$x \gg y \Rightarrow$ favorise le fait d'avoir peu d'insertions (même longues)

```
taccagttgcgggcggtta
||||||| | || |||||
taccagt-g-gg-cggta
```

```
taccagttgcgggcggtta
||||||| | |||||
taccagt---ggcggtta
```

une insertion ou une délétion de plusieurs bases est plus probable que de nombreuses insertions ou délétions indépendantes d'une seule base.

Odile Leconte -IGBMC ASMI

Comparaison de 2 séquences

- Introduction :
 - notions de base
 - systèmes de scores (matrices de scores, pénalités de gaps)

- Comparaison par la matrice de points (dotplot)

- Alignements optimaux

Odile Leconte -IGBMC ASMI

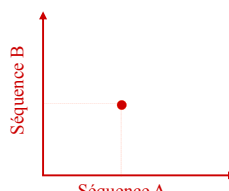
Matrice de points (dotplot)

➤ **Référence :**
Gibbs and McIntyre. The diagram, a method for comparing sequences. Its use with amino acid and nucleotide sequences. *European Journal of biochemistry* 1970

➤ **Principe de base :**

Les 2 séquences à comparer sont placées le long des axes d'un graphique

L'intersection de chaque colonne et de chaque ligne est marquée d'un point si la lettre est la même dans les 2 séquences

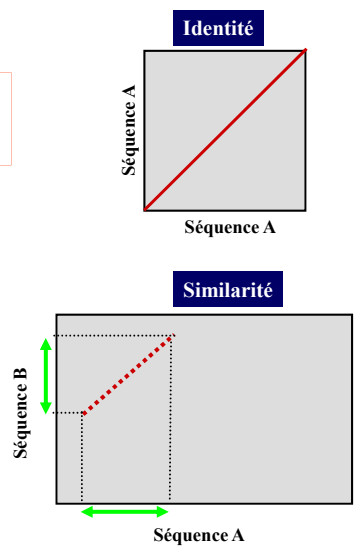


Odile Leconte -IGBMC ASMI

Le signal : les diagonales

Une grande diagonale en cas d'identité parfaite : la séquence contre elle-même

Les régions de similarité apparaissent comme des suites de points alignés ⇒ **diagonales**



Odile Leconte -IGBMC ASMI

Le signal : les diagonales

Un **décalage** par rapport à la diagonale indique une **insertion** ou une **délétion** dans l'une des séquences.

Une inversion de l'orientation d'une diagonale traduit une **inversion** d'une région d'ADN.

Odile Leconte -IGBMC ASMI

Le signal : les diagonales

Une séquence avec elle-même :
diagonales **parallèles** => présence de **régions répétées**

Une séquence avec sa séquence complémentaire :
Détection de régions complémentaires dans une séquence (ex: structure secondaire d'ARN)

Odile Leconte -IGBMC ASMI

Le bruit

Bruits = points résultant d'une correspondance fortuite (hasard)

Nombre de points attendus :
 a_i : nombre de résidus de type i dans la séquence A
 b_i : - i - B

 Nombre total de points : $N = \sum_i^n a_i b_i$
 avec n le nombre de lettres dans l'alphabet considéré

⇒ Pour les séquences nucléiques, le bruit sera très important

Odile Lecompte -IGBMC ASMI

Le bruit

Bruit = points résultant d'une correspondance fortuite

⇒ 7 points significatifs sur 18

⇒ 7 points significatifs sur 9

La distinction signal/ bruit difficile :
 - pour les séquences nucléiques (bruit important)
 - pour des séquences protéiques éloignées (signal faible)

Odile Lecompte -IGBMC ASMI

Distinction signal/bruit

Nécessité d'un filtre :
 Comparaison sur des **fenêtres** de longueur fixe que l'on déplace le long des 2 séquences (*sliding windows*).

On met un point si :

- nb de résidus (ou bases) identiques > seuil
- ou*
- score obtenu sur la fenêtre > score seuil (stringence)

Longueur de la fenêtre : w
 Stringence : s
 Incrément (déplacement de la fenêtre) : i

w = 10
 s = 8, i = 1

Odile Leconte -IGBMC ASM1

Exemple de dotplot – 2 séquences

Grande sous-unité de l'ADN polymerase II

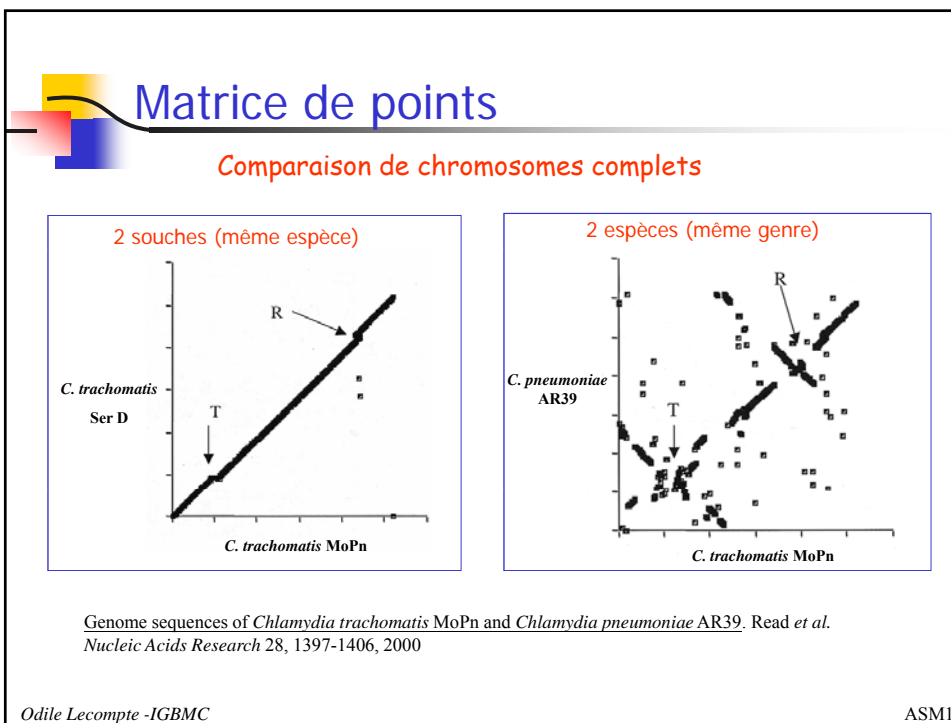
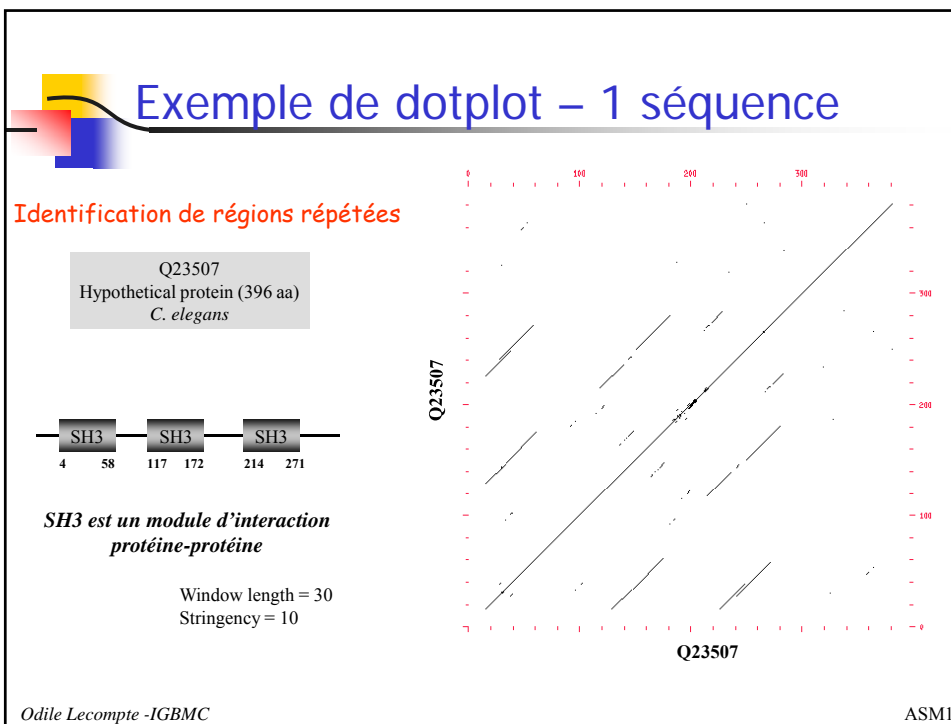
- *P. abyssi* (1455 aa)
- *M. jannaschii* (1139 aa)

dp2l_metja

dp2l_pyrab

insertion

Odile Leconte -IGBMC ASM1





La matrice de points (dotplot)

➤ Avantages :

- vision globale de similarité entre 2 séquences
- toutes les zones de similarité sont visibles
- détection rapide :
 - des insertions / délétions
 - des inversions
 - des régions répétées
 - des zones d'appariements potentiels de l'ARN

➤ Inconvénients :

- méthode visuelle
- aucun alignement n'est fourni

Programmes : Dotter, GCG (Compare et Dotplot), MUMmer...



Comparaison de 2 séquences

- Introduction
- Comparaison par la matrice de points (dotplot)

■ Alignements optimaux

- Alignement **global** => Needleman et Wunsch
Needleman and Wunsch. J. Mol. Biol. 488:443-453 (1970)
- Alignement **local** => Smith & Waterman
Smith and Waterman. J. Mol. Biol. 25:195-7 (1981)

Needleman & Wunsch

Matrice initiale
Scores élémentaires $se(i,j)$

	L	N	A	W	M	E	T	R	C
Y	-1	-2	-2	2	-1	-2	-2	-2	-2
Q	-2	0	-1	-2	0	2	-1	1	-3
A	-1	-2	4	-3	-1	-1	0	-1	0
W	-2	-4	-3	11	-1	-3	-2	-3	-2
I	2	-3	-1	-3	1	-3	-1	-3	-1
V	1	-3	0	-3	1	-2	0	-3	-1
D	-4	1	-2	-4	-3	2	-1	-2	-3
S	-2	1	1	-3	-1	0	1	-1	-1

Blosum62

Matrice de comparaison
Scores $S(i,j)$

	L	N	A	W	M	E	T	R	C
Y	-1	-2	-2	2	-1	-2	-2	-2	-2
Q	-2	0							
A	-1								
W	-2								
I	2								
V	1								
D	-4								
S	-2								

$$S(i,j) = se(i,j) + \text{MAX} \begin{cases} |S(i-1,j-1) \\ |S(i-2 \leq x \leq 1, j-1) - P \\ |S(i-1, j-2 \leq y \leq 1) - P \end{cases}$$

avec P pénalité d'insertion de gaps
 Exemple : $P = a + b \cdot l$ avec
 a pénalité d'ouverture de gap (ici $a=2$)
 b pénalité d'extension de gap (ici $b=1$)
 l longueur du gap

Odile Lecompte -IGBMC ASMI

Needleman & Wunsch

Matrice initiale
Scores élémentaires $se(i,j)$

	L	N	A	W	M	E	T	R	C
Y	-1	-2	-2	2	-1	-2	-2	-2	-2
Q	-2	0	-1	-2	0	2	-1	1	-3
A	-1	-2	4	-3	-1	-1	0	-1	0
W	-2	-4	-3	11	-1	-3	-2	-3	-2
I	2	-3	-1	-3	1	-3	-1	-3	-1
V	1	-3	0	-3	1	-2	0	-3	-1
D	-4	1	-2	-4	-3	2	-1	-2	-3
S	-2	1	1	-3	-1	0	1	-1	-1

Blosum62

Matrice de comparaison
Scores $S(i,j)$

	L	N	A	W	M	E	T	R	C
Y	-1	-2	-2	2	-1	-2	-2	-2	-2
Q	-2	-1	-3	-4	2				
A	-1	-4	3	-6	-2				
W	-2	-5	-7	14	-1				
I	2	-5	-6	-3	15				
V	1	-1	-1	-4	1				
D	-4	2	-3	-5					
S	-2	-1	3	-4					

$$S(i,j) = se(i,j) + \text{MAX} \begin{cases} |S(i-1,j-1) \\ |S(i-2 \leq x \leq 1, j-1) - P \\ |S(i-1, j-2 \leq y \leq 1) - P \end{cases}$$

avec P pénalité d'insertion de gaps
 Exemple : $P = a + b \cdot l$ avec
 a pénalité d'ouverture de gap (ici $a=2$)
 b pénalité d'extension de gap (ici $b=1$)
 l longueur du gap

Odile Lecompte -IGBMC ASMI

Needleman & Wunsch

Recherche du chemin de score maximum

départ
score max dans dernière ligne ou dernière colonne

chemin inverse
(méorisé à l'étape précédente)

Arrêt
première ligne et/ou première colonne

	L	N	A	W	M	E	T	R	C
Y	-1	-2	-2	2	-1	-2	-2	-2	-2
Q	-2	-1	-3	-4	2	1	-3	-1	-5
A	-1	-4	3	-6	-2	1	1	-3	-1
W	-2	-5	-7	14	-1	-4	-1	-2	-4
I	2	-5	-6	-3	15	8	9	6	7
V	1	-1	-1	-4	12	13	12	8	9
D	-4	2	-3	-5	7	14	12	10	6
S	-2	-1	3	-4	8	11	15	11	9

TRC
:
S--

Odile Lecompte -IGBMC ASMI

Needleman & Wunsch

Scores identiques pour différents chemins

Il peut exister plusieurs chemins avec un même score, en particulier dans le cas de séquences nucléiques.
=> choix du chemin avec le minimum de gap généralement

Implémentation

ex : programme `gap` dans GCG, `needle` dans Emboss

Problèmes

la similarité peut concerner une partie seulement des séquences
(ex: cas des protéines multidomaines)

Dans ce cas, l'alignement global risque de conduire à des erreurs.

Odile Lecompte -IGBMC ASMI

Smith & Waterman

Matrice initiale
Scores élémentaires $se(i,j)$

	L	N	A	W	M	E	T	R	C
Y	-1	-2	-2	2	-1	-2	-2	-2	-2
Q	-2	0	-1	-2	0	2	-1	1	-3
A	-1	-2	4	-3	-1	-1	0	-1	0
W	-2	-4	-3	11	-1	-3	-2	-3	-2
I	2	-3	-1	-3	1	-3	-1	-3	-1
V	1	-3	0	-3	1	-2	0	-3	-1
D	-4	1	-2	-4	-3	2	-1	-2	-3
S	-2	1	1	-3	-1	0	1	-1	-1

Blosum62

Matrice de comparaison
Scores $S(i,j)$

	L	N	A	W	M	E	T	R	C
Y	0	0	0	2	0	0	0	0	0
Q	0								
A	0								
W	0								
I	2								
V	1								
D	0								
S	0								

$S(i, j) = \text{MAX} (0, se(i, j) + \text{MAX} (S(i-1, j-1), S(i-2 \leq x \leq 1, j-1) - P, S(i-1, j-2 \leq y \leq 1) - P))$
 avec P pénalité d'insertion de gaps
 Exemple : $P = a + b \cdot l$ avec
 a pénalité d'ouverture de gap (ici $a=2$)
 b pénalité d'extension de gap (ici $b=1$)
 l longueur du gap

Odile Lecompte -IGBMC ASMI

Smith & Waterman

Recherche du chemin de score maximum

départ
score max dans **toute** la matrice

chemin inverse
(méorisé à l'étape précédente)

Arrêt
score nul


	L	N	A	W	M	E	T	R	C
Y	0	0	0	2	0	0	0	0	0
Q	0	0	0	0	2	2	0	1	0
A	0	0	4	0	0	1	2	0	1
W	0	0	0	15	0	0	0	0	0
I	2	0	0	0	16	9	10	7	8
V	1	0	0	0	13	14	13	9	10
D	0	2	0	0	8	15	13	11	7
S	0	1	3	0	9	12	16	12	10

AW-MET AWM

|| ::: ou || :

AWIVDS AWI

Odile Lecompte -IGBMC ASMI



Implémentation

Comparaison de 2 séquences
Bestfit dans GCG, water dans emboss

Recherche de similarité dans une banque
Paralign <http://dna.uio.no/search/>
SSEARCH <http://www.ddbj.nig.ac.jp/E-mail/homology.html>
MPsrch http://www.dna.affrc.go.jp/htdocs/MPsrch/index_PP.html
(Bioaccelerator de l'EMBL)
sw n, sw p <http://eta.embl-heidelberg.de:8000/misc/>

Problème
Smith & Waterman est considéré comme sensible mais long !!!

Odile Leconte -IGBMC ASMI