

Alignement multiple

Thierry Lecroq

Université de Rouen
FRANCE

Méthode progressive

- ① Calcul des scores d'alignement deux à deux entre les séquences ;
- ② Conversion des scores en distances ;
- ③ Construction d'un arbre guide ;
- ④ Alignement des nœuds de l'arbre par ordre croissant des distances.

Exemple

5 séquences

- $x_0 = \text{ATCTCGAGA}$
- $x_1 = \text{ATCCGAGA}$
- $x_2 = \text{ATGTCGACGA}$
- $x_3 = \text{ATGTCGACAGA}$
- $x_4 = \text{ATTCAACGA}$

Alignement des séquences deux à deux

Par programmation dynamique avec

- $s(a, a) = 1$;
- $s(a, b) = -1$ avec $a \neq b$;
- $ins(a) = sup(a) = -2$.

Conversion des scores en distances

$$1 - \frac{\text{nombre d'identités de l'alignement}}{\text{longueur de l'alignement}}$$

x_0 et x_1

		A	T	C	C	G	A	G	A
	0	-2	-4	-6	-8	-10	-12	-14	-16
A	-2	1	-1	-3	-5	-7	-9	-11	-13
T	-4	-1	2	0	-2	-4	-6	-8	-10
C	-6	-3	0	3	1	-1	-3	-5	-7
T	-8	-5	-2	1	2	0	-2	-4	-6
C	-10	-7	-4	-1	2	1	-1	-3	-5
G	-12	-9	-6	-3	0	3	1	0	-2
A	-14	-11	-8	-5	-2	1	4	2	1
G	-16	-13	-10	-7	-4	-1	2	5	3
A	-18	-15	-12	-9	-6	-3	0	3	6

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

$$d(x_0, x_1) = 1 - \frac{8}{9} = 0,11$$

x_0 et x_2

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

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

$$d(x_0, x_2) = 1 - \frac{8}{10} = 0,2$$

x_0 et x_3

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

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

$$d(x_0, x_3) = 1 - \frac{8}{11} = 0,27$$

x_0 et x_4

		A	T	T	C	A	A	C	G	A
	0	-2	-4	-6	-8	-10	-12	-14	-16	-18
A	-2	1	-1	-3	-5	-7	-9	-11	-13	-15
T	-4	-1	2	0	-2	-4	-6	-8	-10	-12
C	-6	-3	0	1	1	-1	-3	-5	-7	-9
T	-8	-5	-2	1	0	0	-2	-4	-6	-8
C	-10	-7	-4	-1	2	0	-1	-1	-3	-5
G	-12	-9	-6	-3	0	1	-1	-2	0	-2
A	-14	-11	-8	-5	-2	1	2	0	-2	1
G	-16	-13	-10	-7	-4	-1	0	1	1	-1
A	-18	-15	-12	-9	-6	-3	0	-1	0	2

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

$$d(x_0, x_4) = 1 - \frac{7}{10} = 0,3$$

x_1 et x_2

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

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

$$d(x_1, x_2) = 1 - \frac{7}{10} = 0,3$$

x_1 et x_3

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

AT-CCG-AGA ATGTCGACAGA

$$d(x_1, x_3) = 1 - \frac{7}{11} = 0,36$$

x_1 et x_4

		A	T	T	C	A	A	C	G	A
	0	-2	-4	-6	-8	-10	-12	-14	-16	-18
A	-2	1	-1	-3	-5	-7	-9	-11	-13	-15
T	-4	-1	2	0	-2	-4	-6	-8	-10	-12
C	-6	-3	0	1	1	-1	-3	-5	-7	-9
C	-8	-5	-2	-1	2	0	-2	-2	-4	-6
G	-10	-7	-4	-3	0	1	-1	-3	-1	-3
A	-12	-9	-6	-5	-2	1	2	0	-2	0
G	-14	-11	-8	-7	-4	-1	0	1	1	-1
A	-16	-13	-10	-9	-6	-3	0	-1	0	2

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

$$d(x_1, x_4) = 1 - \frac{6}{9} = 0,33$$

x_2 et x_3

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

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

$$d(x_2, x_3) = 1 - \frac{10}{11} = 0,09$$

x_2 et x_4

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

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

$$d(x_2, x_4) = 1 - \frac{8}{10} = 0,2$$

x_3 et x_4

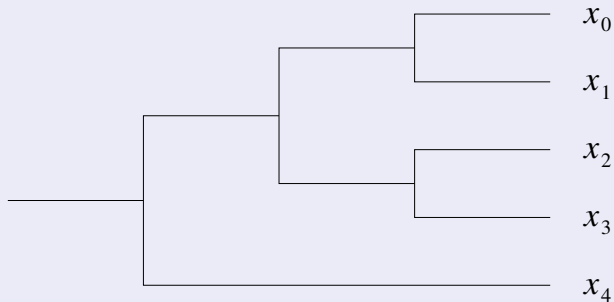
		A	T	T	C	A	A	C	G	A
	0	-2	-4	-6	-8	-10	-12	-14	-16	-18
A	-2	1	-1	-3	-5	-7	-9	-11	-13	-15
T	-4	-1	2	0	-2	-4	-6	-8	-10	-12
G	-6	-3	0	1	-1	-3	-5	-7	-7	-9
T	-8	-5	-2	1	0	-2	-4	-6	-8	-8
C	-10	-7	-4	-1	2	0	-2	-3	-5	-7
G	-12	-9	-6	-3	0	1	-1	-3	-2	-4
A	-14	-11	-8	-5	-2	1	2	0	-2	-1
C	-16	-13	-10	-7	-4	-1	0	3	1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

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

$$d(x_3, x_4) = 1 - \frac{8}{11} = 0,27$$

x_1	0, 11			
x_2	0, 2	0, 3		
x_3	0, 27	0, 36	0, 9	
x_4	0, 3	0, 33	0, 2	0, 27
	x_0	x_1	x_2	x_3

L'arbre guide



Un dendrogramme.

$a(x_0, x_1)$

x_0	A	T	C	T	C	G	A	G	A
x_1	A	T	C	-	C	G	A	G	A

$a(x_2, x_3)$

x_2	A	T	G	T	C	G	A	C	-	G	A
x_3	A	T	G	T	C	G	A	C	A	G	A

$a(a(x_0, x_1), a(x_2, x_3))$

x_0	A	T	C	T	C	G	A	-	-	G	A
x_1	A	T	C	-	C	G	A	-	-	G	A
x_2	A	T	G	T	C	G	A	C	-	G	A
x_3	A	T	G	T	C	G	A	C	A	G	A

$$a(a(a(x_0, x_1), a(x_2, x_3)), x_4)$$

x_0	A	T	C	T	C	G	A	-	-	G	A
x_1	A	T	C	-	C	G	A	-	-	G	A
x_2	A	T	G	T	C	G	A	C	-	G	A
x_3	A	T	G	T	C	G	A	C	A	G	A
x_4	A	T	-	T	C	A	A	C	-	G	A