

Contrôle continu  
Models, methods and algorithms for bioinformatics  
mardi 17 décembre 2019  
16H00 - 17H30

documents de cours, TD et TP autorisés  
tout autre document interdit

Nombre d'intercalaires supplémentaires utilisés :

Vous devez répondre sur le présent document. En cas de manque de place, veuillez indiquer le numéro d'intercalaires supplémentaires utilisés en débutant à 1. Le présent document n'est pas compté comme intercalaire. Si aucune référence ne figure, la réponse ne sera pas lue.

Les copies illisibles ou comportant des fautes d'orthographe seront sanctionnées.

Toute réponse doit être justifiée (éventuellement par un schéma). Une réponse oui/non ne rapportera aucun point.

### Question 1 : alignement global de deux séquences

**convention : La séquence de référence est celle décrite **verticalement**.**

**1.1.** Ecrivez les trois alignements correspondant aux chemins indiqués sur la matrice d'alignement présentée ci-dessous. Calculez-en les coûts en donnant le détail des calculs.

Coûts de délétion

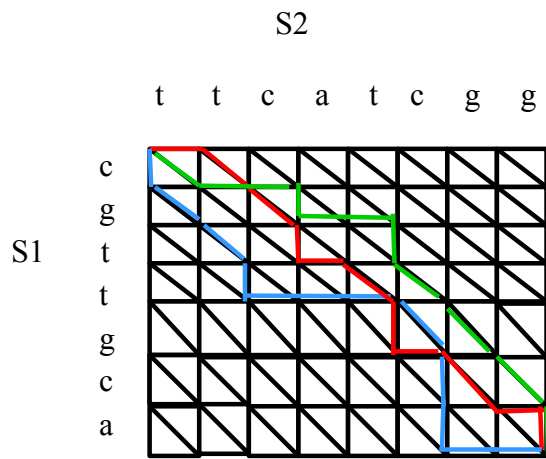
a	t	c	g
3	3	2	4

Coûts d'insertion

a	t	c	g
2	1	2	1

Coûts de substitution

		Caractère substituant			
		a	t	c	g
Caractère substitué	a	0	1	2	1
	t	1	0	4	2
	c	3	2	0	2
	g	3	1	2	0



alignement rouge : S1  
 S2  
 calcul coût :

alignement vert : S1  
 S2  
 calcul coût :

alignement bleu : S1  
 S2  
 calcul coût :

**1.2.** Quel est le meilleur des trois alignements ?

**1.3.** Que doit-on modifier dans l'algorithme d'alignement global de deux séquences présenté en cours pour aligner deux séquences protéiques ?

## Question 2 : Phylogénie – Méthode par parcimonie

**2.1.** Afin d'utiliser la méthode par parcimonie avec les données ci-dessous, identifiez les sites informatifs en complétant les schémas, comme vu en Travaux Dirigés, avec une pondération identique entre transitions et transversions (égale à 1).

**2.2.** Même question qu'en 1.1. avec une pondération de 1 pour chaque transition et une pondération de 2 pour chaque transversion.

Transition :  $a \leftrightarrow g$ ,  $t \leftrightarrow c$

Transversion :  $(a/g) \leftrightarrow (t/c)$

Réponse question 2.1

	sites				
séquences	1	2	3	4	5
A	a	g	c	a	t
B	c	a	t	t	g
C	c	g	c	c	t
D	a	a	c	t	g

site 1

arbre I

A

B

C

D

score =

site 2

arbre II

A

C

B

D

score =

site 3

arbre III

A

D

B

C

score =

site 4

arbre I

A

B

C

D

score =

site 5

arbre II

A

C

B

D

score =

site 6

arbre III

A

D

B

C

score =

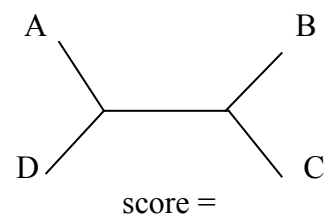
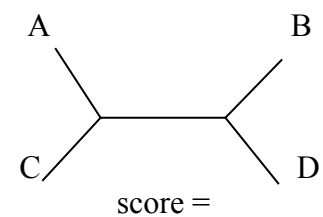
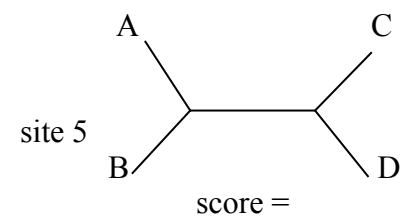
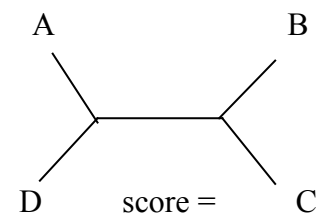
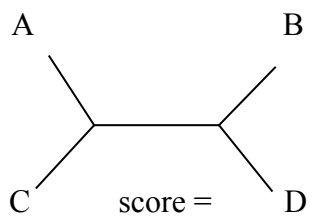
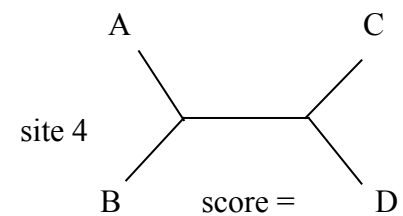
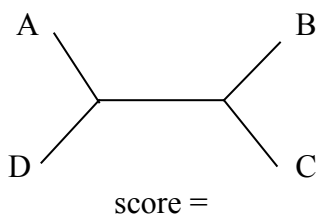
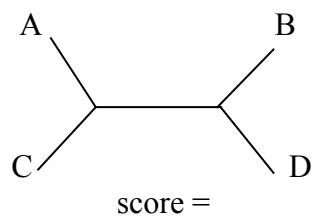
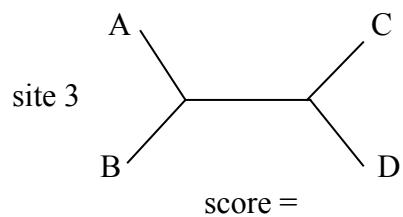
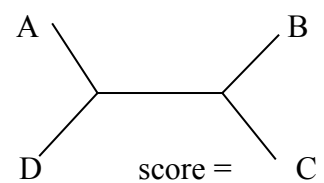
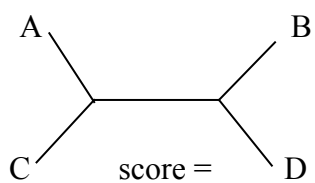
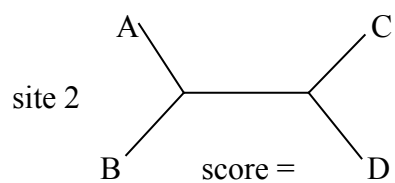
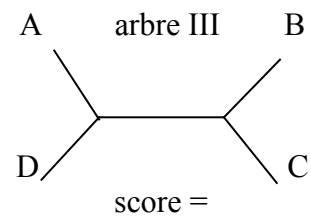
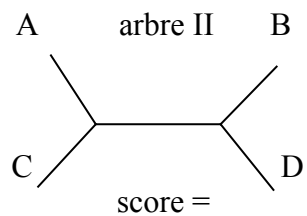
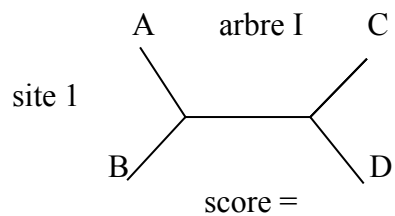
(Barrez les sites non informatifs dans le tableau, le cas échéant)

	Site 1	Site 2	Site 3	Site 4	Site 5	Total sites
Arbre I						
Arbre II						
Arbre III						

Arbre retenu :  
Justification :

## Réponse question 2.2

	sites				
séquences	1	2	3	4	5
A	a	g	c	a	t
B	c	a	t	t	g
C	c	g	c	c	t
D	a	a	c	t	g



(Barrez les sites non informatifs dans le tableau, le cas échéant)

	Site 1	Site 2	Site 3	Site 4	Site 5	Total sites
Arbre I						
Arbre II						
Arbre III						

Arbre retenu :

Justification :

### **Question 3 : Algorithme d'alignement multiple de plusieurs séquences**

Un algorithme d'alignement multiple de plusieurs séquences a été présenté en cours.

**3.1.** A quoi servent les arbres élaborés successivement au cours cet algorithme ?

**3.2.** Comment chacun de ces arbres est-il généré ?

**3.3.** En quoi la première itération et les itérations suivantes de l'algorithme d'alignement multiple vu en cours diffèrent-elles ?

**3.4.** Dans l'alignement multiple de plusieurs séquences, une paire de projections prises au hasard correspond-elle à l'alignement global des deux séquences correspondantes ?

**3.5.** Est-ce qu'on peut utiliser la distance de Levenshtein dans un algorithme d'alignement multiple ?

**3.6.** Est-ce que l'alignement multiple vu en cours, appliqué à deux séquences, revient à un alignement global des deux séquences ?

#### **Question 4 : Reconstruction phylogénétique - Condition des quatre points**

**4.1.** Expliquez à l'aide de schémas ce que signifie cette condition.

**4.2.** Comment la condition des quatre points peut-elle être utilisée en reconstruction phylogénétique ?