

## Exercices sur les graines

### Exercice 1: Automates

- Question 1.** Donnez un automate déterministe pour la graine  $k$ -mer 111111
- Question 2.** Donnez un automate déterministe pour la graine espacée 1\*11\*\*1.
- Question 3.** Donnez un automate déterministe pour l'expression Prosite W-L-[AKP]-x(2,3)-L.
- Question 4.** Donnez un automate déterministe pour l'ensemble des mots qui ont au plus une erreur avec le mot AGGCT.

### Exercice 2: Conception de graines

- Question 1.** Donnez un exemple d'alignement de longueur 30 entre deux séquences nucléiques qui soit accepté avec la graine par défaut de BlastN, mais pas accepté avec la graine par défaut de mégablast.
- Question 2.** Donnez une graine espacée de longueur 7 et de poids 5, qui reconnaisse l'alignement 1 mais pas l'alignement 2.

alignement 1	alignement 2
ACGTACGTACGT	ACGTACGTACGT
ACTTGC-TGCGT	AAGTTGAAC-T

### Exercice 3: Graines et graines espacées

Dans la première partie de l'exercice, vous allez travailler avec la graine  $k$ -mer 111

- Question 1:** Construisez deux séquences d'ADN de longueur 9 qui donnent la même séquence protéique, mais dont l'alignement deux à deux n'est pas reconnu par 111. Le code génétique est rappelé ci-dessous.

		Second letter											
		U		C		A		G					
First letter	U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U			
	UUC	UCC		UAC		UGC		C					
	UUA	UCA		UAA		UGA		A					
	UUG	UCG	UAG	UGG		G							
						STOP		STOP					
						STOP							
First letter	C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	U			
	CUC	CCC		CAC		CGC		C					
	CUA	CCA		CAA		CGA		A					
	CUG	CCG	CAG	CGG		G							
First letter	A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	U			
	AUC	ACC		AAC		AGC		C					
	AUA	ACA		AAA		AGA		A					
	AUG	ACG	AAG	AGG		G							
First letter	G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U			
	GUC	GCC		GAC		GGC		C					
	GUA	GCA		GAA		GGA		A					
	GUG	GCG	GAG	GGG		G							

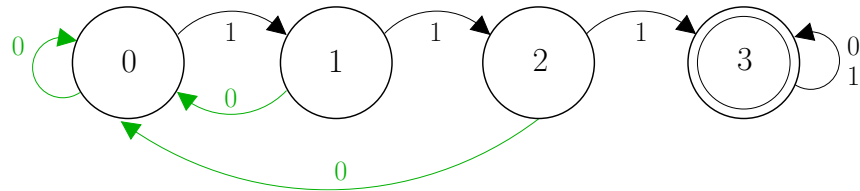
**Question 2:** Montrez que tous les alignements de longueur 50 et de pourcentage d'identité supérieur ou égal à 70% sont attrapés par cette graine 3-mer.

**Question 3:** Quel est le pourcentage d'identité le plus faible pour lequel 111 est une graine sans perte pour des alignements de longueur 50 ?

Vous avez vu en cours que l'on pouvait estimer la sensibilité d'une graine  $k$ -mer en calculant les formules suivantes:

- $P(i, q, x)$  : probabilité d'atteindre l'état  $q$  après avoir lu un mot  $u$  de longueur  $i$  dont la dernière lettre est  $x$
- $P(i, q) = P(i, q, 0) + P(i, q, 1)$ : probabilité d'atteindre l'état  $q$  après  $i$  lettres

Ces formules peuvent s'exprimer de manière récursive en considérant l'automate déterministe qui reconnaît les alignement contenant la graine. Dans le cas 111, il s'agit de l'automate ci-dessous.



**Question 4:** Ecrivez les formules pour la graine 111 et le pourcentage d'identité  $p = 50\%$ .

En repensant à la question 1, on considère maintenant la graine espacée 11\*1, qui semble particulièrement bien adaptée à la comparaison de séquences codantes.

**Question 5:** Construisez un automate non déterministe pour cette graine espacée.

**Question 6:** A partir de l'automate non déterministe, donnez un automate déterministe pour 11\*1.

**Question 7:** Déduisez-en les formules de récurrence analogues à celle de la question 3, toujours pour un pourcentage d'identité égal à 50%.