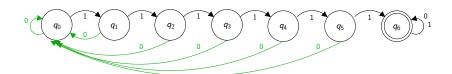
Exercices sur les graines

Exercice 1 – automates

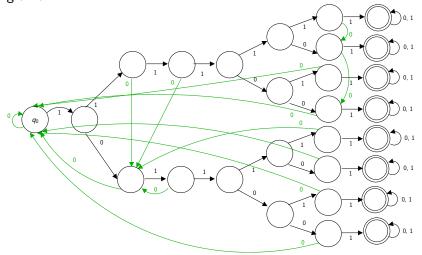
Question 1: Automate pour la graine k-mer 111111

Ensemble des mots sur l'alphabet $\{0,1\}$ qui contiennent le facteur 111111



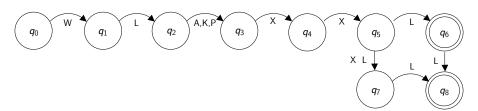
Question 2: Automate pour la graine espacée 1*11**1.

Ensemble des mots sur l'alphabet $\{0,1\}$ qu isont reconnus par la graine 1*11**1

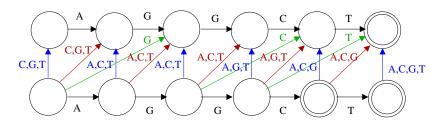


Question 3: Automate déterministe pour l'expression Prosite W-L-[AKP]-x(2,3)-L

Ensemble des mots représentés par cette expression



Question 4: Automate pour l'ensemble des mots qui ont au plus une erreur avec le mot AGGCT



arc bleus: insertion arcs rouges: substitution arcs verts: délétion

Exercice 2 Conception de graines

Question 1: 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

Graine par défaut pour BlastN:

Graine par défaut pour megablast:

Question 2: Graine espacée de longueur 7 et de poids 5, qui reconnait l'alignement 1 mais pas l'alignement 2

 alignement 1 alignement 2 ACGTACGTACGT ACGTACGTACGT 1 11 11 11 1 ACTTGC-TGCGT AAGTTCGAAC-T 1*1*111 1*1*111 1*1*111 1*1*111 1*1*111 1*1*111 1*1*111

Exercice 3 – Graines et graines espacées

Question 1: 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

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

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

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

- on a au plus 15 erreurs
- on découpe la séquence en 16 blocs: 15 blocs de longueur 3 et 1 bloc de longueur 5
- principe du pigeonnier: au moins un bloc ne contient pas d'erreur: il est attrapé par la graine 111

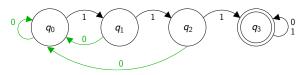
Question 3: pourcentage d'identité le plus faible pour lequel 111 est une graine sans perte pour des alignements de longueur 50 ?

- 15 erreurs: ok
- 16 erreurs ?

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

- 15 erreurs: ok
- 16 erreurs ? L'alignement $(110)^{16}11$ n'est pas reconnu par la graine 111

Question 4: formules pour la graine 111 et le pourcentage d'identité p=50%



$$P(i, q_0, 0) = 0.5[P(i - 1, q_0) + P(i - 1, q_1) + P(i - 1, q_2)]$$

$$P(i, q_0, 1) = 0$$

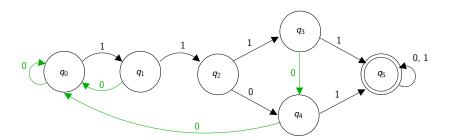
$$P(i, q, 0) = 0 \quad \text{si } q_0 < q < q_3$$

$$P(i, q, 1) = 0.5 P(i - 1, q - 1)$$

$$P(i, q_3, 0) = 0.5P(i - 1, q_3)$$

$$P(i, q_3, 1) = 0.5 [P(i - 1, q_2) + P(i - 1, q_3)]$$

Question 5,6: automate pour la graine 11*1



Question 7: formules pour la sensibilité

$$P(i, q_0, 0) = 0.5[P(i - 1, q_0) + P(i - 1, q_1) + P(i - 1, q_4)]$$

$$P(i, q_0, 1) = 0$$

$$P(i, q_1, 0) = 0$$

$$P(i, q_1, 1) = 0.5 P(i - 1, q_0)$$

$$P(i, q_2, 0) = 0$$

$$P(i, q_2, 1) = 0.5 P(i - 1, q_1)$$

$$P(i, q_3, 0) = 0P(i - 1, q_3)$$

$$P(i, q_3, 1) = 0.5 P(i - 1, q_2)$$

$$P(i, q_4, 0) = 0.5 [P(i - 1, q_2) + P(i - 1, q_3)]$$

$$P(i, q_4, 1) = 0$$

$$P(i, q_5, 0) = 0.5 P(i - 1, q_5)$$

$$P(i, q_5, 1) = 0.5 [P(i - 1, q_3) + P(i - 1, q_4) + P(i - 1, q_5)]$$