

IFT6291/BIN6000 – Bioinformatique génomique
Devoir 3, session Automne 2020

À remettre le jeudi 26 novembre 2020

1. Prétraitement de P par un automate

Soit le mot $P = ACAC$ sur l'alphabet $\{A, C\}$.

- a) Faites le prétraitement d'Ukkonen de P consistant à construire l'automate fini déterministe M_P dont les états sont les colonnes possibles d'une table de programmation dynamique pour ce mot. Laissez les traces de vos calculs. (Votre automate devrait avoir 12 états. Pour simplifier la visualisation, mettre toutes les transitions "A" en rouge et les transitions "C" en bleu.)
- b) Soit le texte $T = CAACCACACAAA$. Trouvez les occurrences exactes ou à une erreur de P dans T . Inscrive le chemin suivi dans l'automate à la lecture de ce texte et les positions des occurrences. (Par exemple, l'état initial étant 01234, on aura comme début de chemin $01234 \xrightarrow{C} 01123 \xrightarrow{A} \dots$)

2. Calcul de diagonales

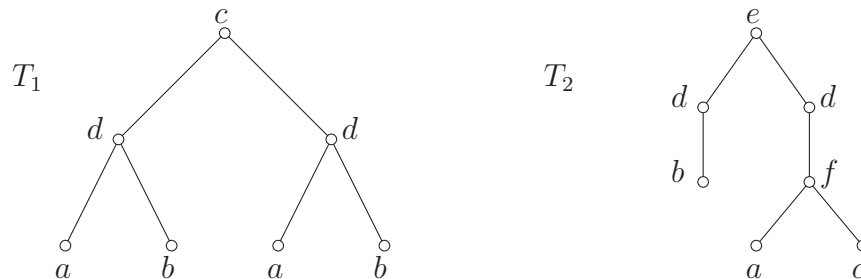
Soit la table de programmation dynamique suivante pour le calcul des occurrences approximatives de $P = AACC$ dans $T = GTTAACTGAACAACC$ dans laquelle les 0-extrémaux pour les diagonales 0 à 11 ont été calculés. Servez-vous de cette information pour calculer les 1-extrémaux des diagonales 0 à 10. (Vous pouvez le faire sans la construction de l'arbre ou automate des suffixes mais seulement en utilisant la boucle tant que vu en classe.). Indiquez les positions où se terminent une occurrence exacte ou à une erreur de P dans T .

	λ	G	T	T	A	A	C	T	G	A	A	C	A	A	C	C
λ	0	0	0			0	0	0			0					
A						0					0					
A																
C							0					0				
C																0

3. Algorithmes vectoriels

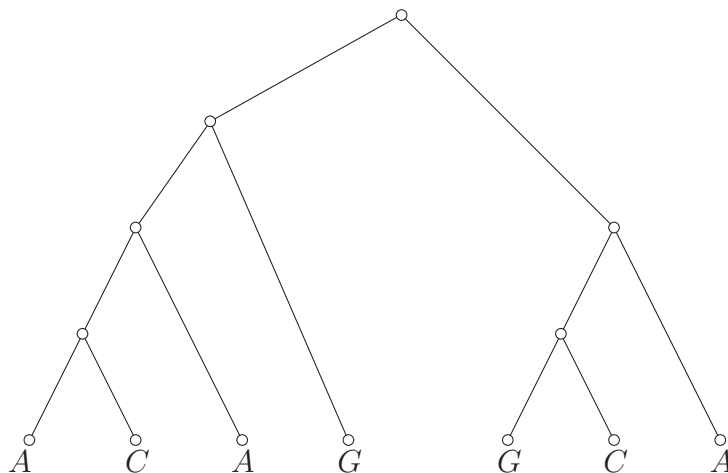
Soit $P = GATC$ et $T = GATTCTTGATC$. Calculez les vecteurs de bits R_j^0 et R_j^1 , pour $0 \leq j \leq 11$, de l'algorithme de Wu et Manber pour la recherche approximative de P dans T et indiquez les positions où se terminent une occurrence exacte ou à une erreur de P dans T .

4. Trouver la distance d'édition entre les arbres T_1 et T_2 suivants en calculant toutes les tables de distances nécessaires de l'algorithme de Zhang-Shasha.



5. Soient les séquences $S_1 = ACAG$, $S_2 = ACGG$, $S_3 = AGAC$ et $S_4 = AGAG$. Calculer les différents arbres de phylogénie que vous pouvez obtenir en utilisant l'algorithme UPGMA.

6. Donnez les différents arbres de phylogénie obtenus en utilisant l'algorithme de Fitch sur l'arbre ci-bas. Indiquer pour chacun d'eux les branches sur lesquelles les mutations ont lieu.



BON TRAVAIL!!