

BIN3002 automne 2012 — Devoir 1a

Miklós Csűrös

29 novembre 2012

À remettre avant 23 :59 le 10 décembre. Remettez un rapport écrit par courriel (à csuros@iro...) en format PDF. Travaillez seul.

1a. D'ailleurs, la pondération la plus simple n'est pas aussi simple que ça (10 points)

En parlant de la pondération de substitutions, on travaille souvent avec la pondération simple de +1 pour match et -1 pour mismatch. Dans le contexte de pondération LODS, quelles distributions de nucléotides et probabilités de substitutions mènent à un tel système de scores ?

► Donnez les solutions pour l'échelle α , et les probabilités π_a , $\mathbf{M}[a, b]$ quand

$$C_{a,b} = \log_{\alpha} \frac{\mathbf{M}[a,b]}{\pi_b} = \begin{cases} -1 & \{a \neq b\} \\ 1 & \{a = b\} \end{cases}$$

pour tout $a, b \in \{\mathbf{A}, \mathbf{C}, \mathbf{G}, \mathbf{T}\}$.

Indice: Il faut résoudre un système d'équations avec 21 inconnus (π, \mathbf{M}, α). On a les pondérations (16 équations), et les 5 contraintes sur les probabilités :

$$\begin{aligned} \pi_{\mathbf{A}} + \pi_{\mathbf{C}} + \pi_{\mathbf{G}} + \pi_{\mathbf{T}} &= 1; \\ \sum_{b \in \{\mathbf{A}, \mathbf{C}, \mathbf{G}, \mathbf{T}\}} \mathbf{M}_{a,b} &= 1 \quad \{a = \mathbf{A}, \mathbf{C}, \mathbf{G}, \mathbf{T}\}. \end{aligned}$$

Pondérations moins simples (5 points boni). ► À quelles probabilités de substitution correspondent les scores +2 et -1 ? ► À quelles probabilités de substitution correspondent les scores +3 et -1 ?