Frameshift aware alignment with affine gap panelty

source: Frameshift alignment: statistics and post-genomic applications Sergey L. Sheetlin1,y, Yonil Park1,y, Martin C. Frith2,*,y and John L. Spouge1,*

Notations

$A = a_1, ...a_n$ DNA sequence,
$B = b_1, ...b_m$ amino acid sequence,
$a_i \in \{a, c, g, t\}$,
$b_i \in \{A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W\}$
$T(a_{i-2}, a_{i-1}, a_i) = A_1$ (amino acid translated from codon) $a_{i-2}...a_i$
$w_g = d_0 + d_1 \cdot g$ affine gap panelty with $d_0, d_1 \geq 0$, d_0 gap open and d_1 gap extension panelty, g > 0 length of amino acid gap
γ frameshift panelty

Recursion linear gap panelty

$$M(i,j) = f(x) = \begin{cases} M(i-3, j-1) + s(A_i, B_j) & \\ M(i-3, j) - d, & \\ M(i, j-1) - d & \\ M(i-2, j) - \gamma & forwardframeshift \\ M(i-1, j) - \gamma & backwardframeshift \end{cases}$$

Recursion affine gap panelty

Needed: three arrays S (for substitutions), I (for insertions) and D (for deletions).

Init:

$S(0,0) = 0$,

$D(3g,0) = I(0,g) = -d_0 - d_1 \cdot g$

$D(3g-1,0) = D(3g+1,0) = -d_0 - d_1 \cdot g - \gamma$,

$S(-1,0) = S(-1,g) = S(0,g) = S(1,g) = S(2,g) = S(g,0) = -\infty$,

$D(-1,0) = D(-1,g) = D(0,0) = D(0,g) = D(1,0) = D(1,g) = D(2,g) = -\infty$

$I(-1,0) = I(-1,g) = I(0,0) = I(g,0) = -\infty$ for $g > 0$

$$S(i,j) = max \begin{cases} max\{S(i-3,j-1), D(i-3,j-1), I(i-3,j-1)\} \\ max\{S(i-2,j-1), D(i-2,j-1), I(i-2,j-1), S(i-4,j-1), D(i-4,j-1), I(i-4,j-1)\} - \gamma \end{cases} +\mathrm{s}(A_i, B_j)$$

$$I(i,j) = max \begin{cases} S(i,j-1) - d_0, I(i,j-1) \end{cases} \quad \text{-}\mathrm{d}_1$$

$$D(i,j) = max \begin{cases} max\{S(i-3,j) - d_0, D(i-3,j)\}, \\ max\{S(i-2,j) - d_0, D(i-2,j), S(i-4,j) - d_0, D(i-4,j)\} - \gamma \end{cases} \quad \text{-}\mathrm{d}_1$$

The default frameshift alignment scoring system in FALP is BLOSUM80 with w $w_g = 11 + 2g$ and $\gamma = 15$.

Question1:

How chose g? > 0, but how great exactly? We don't now the size of the gap in advance?

Question2:

What does

"The preceding calculation disallows adjacent insertions and deletions. If desired, they can be allowed by replacing the recursion for I(i,j) with

$I(i,j) = max\{S(i,j-1) - d_0, I(i,j-1), D(i,j-1) - d_0\} - d_1$"

mean?