Студент: Гераськин Ярослав
Группа: МОАД
Дата: 26 января 2021 г.

---

1. Из таблицы:

$$w(D, E) = 2$$
$$w(I, V) = 3$$
$$w(D, V) = -3$$

Чем больше число, тем больше вероятность замены. Связан этот эффект с тем, что пары аминокислот D, E и I, V обладают похожими свойствами, а D, V отличаются сильнее.

2. Рассмотрим 3 множества выравниваний строк a, b, M, A, B. M – выравнивание заканчивается двумя символами, A – выравнивание заканчивается пропуском в a, B – выравнивание заканчивается пропуском в b. Будем рассматривать выравнивание с точностью до перестановки соседних гэпов (которые идут непосредственно друг за другом). Пусть $f(n, m)$ – кол-во таких выравниваний для строк длины n и m.

Начальные условия: $f(0, 0) = 1$, $f(0, 1) = 1$, $f(1, 0) = 1$.

$$f(n, m) = f(n-1, m-1) + f(n-1, m) + f(n, m-1) - f(n-1, m-1)$$
$$|M| + |A| + |B| - |C|$$

C – множество выравниваний, которые заканчиваются сменой гэпа, мы посчитали их в A и в B, т.к. хотим считать с точностью до перестановки, нужно вычесть кол-во выравниваний вида

$$\ldots\_x$$
$$\ldots y\_$$

это и есть множество C.

$$f(n, m) = f(n-1, m) + f(n, m-1)$$
$$f(n, m) = C_{n+m}^{n}$$

Начальные условия выполняются.

Т.к. $C_n^k = C_{n-1}^{k-1} + C_{n-1}^{k}$, то по индукции $f(n, m) = C_{n+m-1}^{n-1} + C_{n+m-1}^{n} = C_{n+m}^{n}$

При $m \sim n$:

$$C_{2n}^{n} = \frac{(2n)!}{n!^2} \sim \frac{\sqrt{2\pi 2n}(2n)^{2n}}{(\sqrt{2\pi n} n^n)^2} = \frac{2^{2n}}{\sqrt{\pi n}}$$