

# Алгоритмы в биоинформатике: Домашнее задание № 2, теоретическая часть

Выполнила: Костина Ю.М., НИУ ВШЭ, группа МВБ 221

12 февраля 2023 г.

## 1. Количество выравниваний

Выведите рекуррентную формулу количества всех возможных выравниваний последовательностей длины  $n$  и  $m$  пользуясь разбиением всех выравниваний на непересекающиеся блоки. (1.5 балл) Получите точную формулу, основываясь на начальные условия и рекуррентную формулу. (1.5 балл) Воспользуйтесь приближением Стирлинга чтобы получить приближенную формулу количества выравниваний. (1)

Решение: Пусть количество всех возможных выравниваний последовательностей  $a_1 \dots a_n$  и  $b_1 \dots b_m$  обозначается как  $f(n, m)$ .

Пусть  $|F| = f(n, m)$  - множество всех возможных выравниваний. Тогда:

- (a)  $|F_1| = f(n-1, m-1)$  - все выравнивания для символов  $a_i = b_i$
- (b)  $|F_2| = f(n, m-1)$  - все выравнивания где  $a_i = \emptyset$
- (c)  $|F_3| = f(n-1, m)$  - все выравнивания где  $b_i = \emptyset$
- (d)  $|F_4| = |F_2 \cap F_3| = f(n-1, m-1)$  - пересечение  $V_2$  и  $V_3$ , так как выравнивание вида  $a_i = \emptyset \cap b_i = \emptyset$  эквивалентны

Отсюда:

$$f(n, m) = f(n-1, m-1) + f(n, m-1) + f(n-1, m) - f(n-1, m-1) = f(n, m-1) + f(n-1, m)$$

Если в последовательности в одной последовательности 1 элемент и в другой - нет элементов, то справедливо  $f(1, 0) = C_{1+0}^0 = 1, f(0, 1) = C_{1+0}^1 = 1$ . Допустим, условия выполняются для  $f(n, m-1)$  и  $f(n-1, m)$ , тогда

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