

# Алгоритмы в биоинформатике

## Теоретическая часть

Иван Щукин

13 февраля 2023 г.

### Количество выравниваний (4 балла)

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

1. Пусть даны две последовательности длины  $n$  и  $m$ . Количество всевозможных выравниваний этих последовательностей обозначим за  $V(n, m)$ .

Тогда возможны 4 случая:

- равны последние столбцы выравнивания ( $a_n = b_n$ ) - это  $V(n-1, m-1)$
- $gap \cup b_n$  - это  $V(n, m-1)$
- $a_n \cup gap$  - это  $V(n-1, m)$
- пересечение предыдущих двух случаев - это  $V(n-1, m-1)$

Т.к. итоговое  $V(n, m)$  - это есть сумма всех четырех пунктов (только последний пункт идет со знаком минус, чтобы убрать повторяющиеся выравнивания), то в формульном виде количество всех возможных выравниваний запишется как  $V(n, m) = V(n, m-1) + V(n-1, m) - V(n-1, m-1)$

2. Очевидно, что  $V(1, 0) = V(0, 1) = 1$  и тогда  $V(n, m-1) + V(n-1, m) - V(n-1, m-1) = C_{n+m-1}^n + C_{n-1+m}^{n-1} - C_{n+m-1}^{n-1} = C_{n+m}^n = \frac{(n+m)!}{n!m!}$

3. Формула Стирлинга для приближенного вычисления значения  $n! = \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$   
Тогда подставим ее в найденное значение для  $V(n, m)$ :

$$V(n, m) = C_{n+m}^n = \frac{(n+m)!}{n!m!} = \frac{\sqrt{2\pi(n+m)} \left(\frac{n+m}{e}\right)^{n+m}}{2\pi\sqrt{nm} \left(\frac{n}{e}\right)^n \left(\frac{m}{e}\right)^m}$$