#5

aatcatgac

taacatggc

aaccatggt

aagaatggc

caccatggt

cagcatgga

aaacatgga

caacatgat

cacaatggt

cagcatggc

cacaatggc

**Алгоритм подсчета IC выравнивания.**

Для начала составляем таблицу с подсчетом количества букв: N (b, j), где b – буква, j – позиция, по которой идет расчет.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| A | 4 | 11 | 3 | 3 | 11 | 0 | 0 | 2 | 2 |
| G | 0 | 0 | 3 | 0 | 0 | 0 | 11 | 9 | 0 |
| T | 1 | 0 | 1 | 0 | 0 | 11 | 0 | 0 | 4 |
| C | 6 | 0 | 4 | 8 | 0 | 0 | 0 | 0 | 5 |
| Сумма | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 |

Далее считаем частоты букв f (b, j).

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| A | 0,36 | 1,00 | 0,27 | 0,27 | 1,00 | 0,00 | 0,00 | 0,18 | 0,18 |
| G | 0,00 | 0,00 | 0,27 | 0,00 | 0,00 | 0,00 | 1,00 | 0,82 | 0,00 |
| T | 0,09 | 0,00 | 0,09 | 0,00 | 0,00 | 1,00 | 0,00 | 0,00 | 0,36 |
| C | 0,55 | 0,00 | 0,36 | 0,73 | 0,00 | 0,00 | 0,00 | 0,00 | 0,45 |
| Сумма | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

Теперь считаем отношение частот букв к их частотам в геноме. При этом частота G и C – 0.386, а A и T – 0.127. Пользуемся уравнением IC (b, j) = f (b, j) \* log2[f (b, j) / p(b)] =

= f (b, j) \* w (b, j).

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| A | 2,863 | 7,874 | 2,147 | 2,147 | 7,874 | 0,000 | 0,000 | 1,432 | 1,432 |
| G | 0,000 | 0,000 | 0,707 | 0,000 | 0,000 | 0,000 | 2,591 | 2,120 | 0,000 |
| T | 0,716 | 0,000 | 0,716 | 0,000 | 0,000 | 7,874 | 0,000 | 0,000 | 2,863 |
| C | 1,413 | 0,000 | 0,942 | 1,884 | 0,000 | 0,000 | 0,000 | 0,000 | 1,178 |

Дальше считаем логарифм с основанием 2 по каждой клетке последней таблицы. Тем самым получаем веса для каждой буквы.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| A | 1,52 | 2,98 | 1,10 | 1,10 | 2,98 | 0,00 | 0,00 | 0,52 | 0,52 |
| G | 0,00 | 0,00 | -0,50 | 0,00 | 0,00 | 0,00 | 1,37 | 1,08 | 0,00 |
| T | -0,48 | 0,00 | -0,48 | 0,00 | 0,00 | 2,98 | 0,00 | 0,00 | 1,52 |
| C | 0,50 | 0,00 | -0,09 | 0,91 | 0,00 | 0,00 | 0,00 | 0,00 | 0,24 |

Теперь умножаем f (b, j) на соответствующий вес из последней таблицы и получаем

IC (b, j).

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| A | 0,55 | 2,98 | 0,30 | 0,30 | 2,98 | 0,00 | 0,00 | 0,09 | 0,09 |
| G | 0,00 | 0,00 | -0,14 | 0,00 | 0,00 | 0,00 | 1,37 | 0,89 | 0,00 |
| T | -0,04 | 0,00 | -0,04 | 0,00 | 0,00 | 2,98 | 0,00 | 0,00 | 0,55 |
| C | 0,27 | 0,00 | -0,03 | 0,66 | 0,00 | 0,00 | 0,00 | 0,00 | 0,11 |
| IC(j) | 0,78 | 2,98 | 0,09 | 0,97 | 2,98 | 2,98 | 1,37 | 0,98 | 0,75 |

Так же в последней строке посчитано IC для колонки.

Итоговое информационное содержание сигнала есть сумма IC(j), что ровняется 13,87.

Картинка LOGO представлена ниже.

