Фамилия *Мельникова* имя *Ксения*

Дано выравнивание. GC генома равен 0.6. p(b) – частота буквы b. (b) = псевдоотсчёт для буквы b.

1. Напишите консенсус (Под каждой колонкой самую частую букву БОЛЬШОЙ, если такая не одна любую из них маленькой)
2. Напишите паттерн используя таблицу “ Ambiguous nucleotide codes”
3. Напишите тот же паттерн на комплементарной цепи
4. Постройте матрицу PWM для данного выравнивания
5. Вычислите вес W данной последовательности относительно полученной матрицы PWM

**Выравнивание 17**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| C | C | C | A | C | G | C | A | A | A | C | G | T | T | T | T |
| C | A | C | A | C | G | C | A | A | A | C | G | T | T | T | T |
| T | C | C | A | C | G | C | A | A | A | C | G | G | T | T | T |
| G | C | C | A | C | G | C | T | A | C | C | G | T | T | T | T |
| G | A | T | A | C | G | C | A | A | A | C | G | T | G | T | G |
| C | C | G | A | C | G | C | A | A | T | C | G | G | T | T | A |
| G | T | T | G | C | G | C | A | A | A | C | G | T | T | T | T |

Число букв по столбцам

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| **A** | **0** | 2 | 0 | 6 | 0 | 0 | 0 | 6 | 7 | 5 | 0 | 0 | 0 | 0 | 0 | 1 |
| **T** | **1** | 1 | 2 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 5 | 6 | 7 | 5 |
| **G** | **3** | 0 | 1 | 1 | 0 | 7 | 0 | 0 | 0 | 0 | 0 | 7 | 2 | 1 | 0 | 1 |
| **C** | **3** | 4 | 4 | 0 | 7 | 0 | 7 | 0 | 0 | 1 | 7 | 0 | 0 | 0 | 0 | 0 |
| **consensus** | **G** | C | C | A | C | G | C | A | A | A | C | G | T | T | T | T |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **pattern** | **V** | H | B | V | C | G | C | W | A | H | C | G | K | K | T | D |
| **complement** | **c** | D | V | B | G | C | G | W | T | D | G | C | M | M | A | H |

Матрица PWM **и в в**ес последовательности относительно PWM

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | p(b) | (b) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| A | 0.2 | 1 | -1.14 | 0.51 | -1.14 | 1.51 | -1.14 | -1.14 | -1.14 | 1.51 | 1.51 | 1.67 | 0.51 | -1.14 | -1.14 | -1.14 | -1.14 | -0.14 |
| T | 0.2 | 1 | -0.14 | -0.14 | 0.28 | -1.14 | -1.14 | -1.14 | -1.14 | -1.14 | -0.14 | -1.14 | -0.14 | -1.14 | -1.14 | 1.51 | 1.51 | 1.51 |
| G | 0.3 | 1 | 0.28 | -1.14 | -0.52 | -0.52 | -1.14 | 1.51 | -1.14 | -1.14 | -1.14 | -1.14 | -1.14 | 1.51 | 0.51 | -0.14 | -0.52 | -1.14 |
| C | 0.3 | 1 | 0.28 | 0.51 | 0.51 | -1.14 | 1.51 | -1.14 | 1.51 | -1.14 | -1.14 | -1.14 | 0.28 | -1.14 | -1.14 | -1.14 | -1.14 | -1.14 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| W= | 8,61 | G | T | C | T | C | G | C | A | A | A | C | G | T | c | T | G |

Частоту буквы в позиции вычисляла по принципу Лапласса

PWM = log2 (частота буквы в позиции/частота буквы в геноме)

W = суммирую значения PWM для соответствующих букв в каждой позиции