

Фамилия _____ Тарасов _____
 имя _____ Алексей _____

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

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

Выравнивание 14

									1	1	1	1	1	1	1
1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6
C	C	T	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	A	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	7	7	5	0	0	0	0	0	1
T	1	1	3	0	0	0	0	0	0	1	7	0	0	0	0	0
G	3	0	1	1	0	7	0	0	0	0	0	7	2	1	0	1
C	3	4	3	0	7	0	7	0	0	1	7	0	0	0	0	0
consensus	c	C	t	A	C	G	C	A	A	A	C	G	T	T	T	T
pattern	S	C	Y	A	C	G	C	A	A	A	C	G	T	T	T	T
complement	S	G	R	T	G	C	G	T	T	T	G	C	A	A	A	A

Матрица PWM и **в** вес последовательности относительно PWM

	p(b)	ε(b)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
A	0.2	1	0.09	0.27	0.09	0.64	0.09	0.09	0.09	0.73	0.73	0.55	0.09	0.09	0.09	0.09	0.09	0.18
T	0.2	1	0.18	0.18	0.36	0.09	0.09	0.09	0.09	0.09	0.09	0.18	0.09	0.09	0.55	0.64	0.73	0.55
G	0.3	1	0.36	0.09	0.18	0.18	0.09	0.73	0.09	0.09	0.09	0.09	0.09	0.73	0.27	0.18	0.09	0.18
C	0.3	1	0.36	0.45	0.36	0.09	0.73	0.09	0.73	0.09	0.09	0.18	0.73	0.09	0.09	0.09	0.09	0.09
			0,192372	-0,095310	0,192372	-0,788457	0,885519	0,885519	0,885519	1,290984	1,290984	1,003302	0,885519	0,885519	1,003302	1,157453	1,290984	-0,500775
W = 10.464806			G	T	C	T	C	G	C	A	A	A	C	G	T	T	T	G

$$F(b,j) = [N(b,j) + \text{eps}(b)] / (N + \text{eps})$$

Здесь eps=1, а N = 7

eps = 4

$$F(b,j) = [N(b,j) + 1] / (7 + 4)$$