

Демидов Иван

Дано выравнивание. GC генома равен 0.6. $p(b)$ – частота буквы b . $e(b)$ = 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 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	c	C	C	A	C	G	C	A	A	A	C	G	T	T	T	T
pattern	B	H	B	R	C	G	C	W	A	H	C	G	K	K	T	D
complement	V	D	V	Y	G	C	G	S	T	D	G	C	M	M	A	H

Матрица PWM и в вес последовательности относительно PWM

	p(b)	e(b)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
A	0.2	0.01	-4.94	0.36	-4.94	1.46	-4.94	-4.94	-4.94	1.46	1.61	1.27	-4.94	-4.94	-4.94	-4.94	-4.94	-0.33
T	0.2	0.01	-0.33	-0.33	0.36	-4.94	-4.94	-4.94	-4.94	-0.33	-4.94	-0.33	-4.94	-4.94	1.27	1.46	1.61	1.27
G	0.3	0.01	0.36	-5.35	-0.73	-0.73	-5.35	1.20	-5.35	-5.35	-5.35	-5.35	-5.35	1.20	-0.05	-0.73	-5.35	-0.73
C	0.3	0.01	0.36	0.65	0.65	-5.35	1.20	-5.35	1.20	-5.35	-5.35	-0.73	1.20	-5.35	-5.35	-5.35	-5.35	-5.35
		0.04	0.36	-0.33	0.65	-4.94	1.20	1.20	1.20	1.46	1.61	1.27	1.20	1.20	1.27	-5.35	1.61	-0.73
W=	2.89		G	T	C	T	C	G	C	A	A	A	C	G	T	C	T	G