Вариант 5:

aatcatgac

taacatggc

aaccatggt

aagaatggc

caccatggt

cagcatgga

aaacatgga

caacatgat

cacaatggt

cagcatggc

cacaatggc

Для данного выравнивания было построено LOGO с помощью онлайн-сервиса из презентации:



Для построения матрицы информационного содержания букв нужны матрицы с частотами и весами, они были созданы в Excel.

Вот матрица частот:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|   | частота |   |   |   |   |   |   |   |   |
| буква |  -4 | -3 | -2 | -1 | a | t | g | 4 | 5 |
| a | 0,36 | 1,00 | 0,27 | 0,27 | 1,00 | 0,00 | 0,00 | 0,18 | 0,18 |
| t | 0,09 | 0,00 | 0,09 | 0,00 | 0,00 | 1,00 | 0,00 | 0,00 | 0,36 |
| g | 0,00 | 0,00 | 0,27 | 0,00 | 0,00 | 0,00 | 1,00 | 0,82 | 0,00 |
| c | 0,55 | 0,00 | 0,36 | 0,73 | 0,00 | 0,00 | 0,00 | 0,00 | 0,45 |

Вот матрица весов:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|   |   | вес |   |   |   |   |   |   |   |   |
| буква | базовая частота | -4 | -3 | -2 | -1 | a | t | g | 4 | 5 |
| a | 0,15 | 0,88 | 1,90 | 0,59 | 0,59 | 1,90 | -2,77 | -2,77 | 0,18 | 0,18 |
| t | 0,35 | -1,35 | -3,58 | -1,35 | -3,58 | -3,58 | 1,05 | -3,58 | -3,58 | 0,03 |
| g | 0,15 | -0,47 | -2,77 | 0,59 | -2,77 | -2,77 | -2,77 | 1,90 | 1,70 | -2,77 |
| c | 0,35 | 0,45 | -3,58 | 0,03 | 0,74 | -3,58 | -3,58 | -3,58 | -3,58 | 0,25 |

Матрица информационных содержаний для каждой буквы считалась по формуле: частота данной буквы умножить на вес данной буквы, и вот, что получилось:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|   | IC |   |   |   |   |   |   |   |   |
| буква | -4 | -3 | -2 | -1 | a | t | g | 4 | 5 |
| a | 0,32 | 1,90 | 0,16 | 0,16 | 1,90 | 0,00 | 0,00 | 0,03 | 0,03 |
| t | -0,12 | 0,00 | -0,12 | 0,00 | 0,00 | 1,05 | 0,00 | 0,00 | 0,01 |
| g | 0,00 | 0,00 | 0,16 | 0,00 | 0,00 | 0,00 | 1,90 | 1,39 | 0,00 |
| c | 0,25 | 0,00 | 0,01 | 0,54 | 0,00 | 0,00 | 0,00 | 0,00 | 0,11 |

Суммарное информационное содержание = сумме информационных содержаний для каждой буквы:

9,68

 (посчитано в Excel)