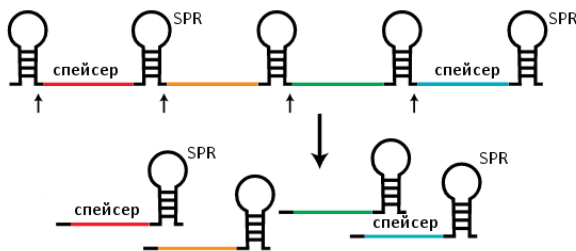




**Математика для школьников 7 – 11 класса (заочный тур)**

**Задача 7. Поиск CRISPR в геноме *E. Coli***



Нобелевская премия по химии 2020 года была присуждена за развитие метода редактирования генома, базирующегося на использовании системы адаптивного иммунитета бактерий, направленной на борьбу с вирусами.

Основа этой системы – особые последовательности нуклеотидов<sup>1</sup> – короткие палиндромные<sup>2</sup> кластерные повторы или CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats), впервые обнаруженные в 1987 году в геноме бактерии кишечной палочки *Escherichia coli*. Короткие палиндромные повторы (SPR) в системе CRISPR отделены друг от друга так называемыми спейсерами – последовательностями нуклеотидов, являющихся фрагментами (до 80 нуклеотидов длиной) генома вирусов, с которыми бактерия сталкивалась ранее.

В тексте генома<sup>3</sup> *E. Coli* найдите нуклеотидные последовательности, отвечающие

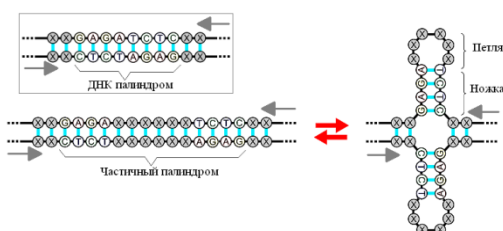
- а) SPR вида .GGTTTATCCCCGCT..CGCGGGAAC.C (здесь точка обозначает, что на данной позиции может находиться любой символ) **(6 баллов)**
- б) и разделяющим их спейсерам **(5 баллов)**,

а также их длины и порядковые номера первых нуклеотидов этих последовательностей.

Для выполнения задания напишите программу (программы) (на любом языке программирования, к решению приложите исходный код), которая выполнит нужные действия и выведет на экран найденную информацию.

<sup>1</sup> Молекула ДНК состоит всего из четырех «букв»-нуклеотидов – **A, C, G, T**. «Буквы» ДНК из одной цепочки способны связываться попарно (**A ↔ T, G ↔ C**) с буквами из противоположной цепочки, называемой комплементарной.

<sup>2</sup> ДНК-палиндромом называется такая последовательность ДНК, прочтение которой совпадает с прочтением в обратном направлении по комплементарной цепочке. Например, последовательность **АТТА** – «обычный» палиндром, а последовательность **ААТТ** – ДНК-палиндром.



<sup>3</sup> Бактерия *E. Coli* (кишечная палочка) является одним из удобных модельных организмов в биологии, а геном ее лабораторного штамма K-12 был расшифрован одним из первых. Для выполнения этого задания сохраните по ссылке <http://enanos.nanometer.ru/uploads/archive/ecoli.zip> со страницы задачи (~1,3 МВ) файл генома штамма K-12 *E.Coli*, который состоит из одной непрерывной строки, содержащей только буквы **A, G, T, C**.

**Всего – 11 баллов**