



Математика для школьников 7 – 11 класса (заочный тур) Решение задачи 7. Поиск CRISPR в геноме *E. Coli*

Программа на языке Pascal

```
type
mas = array[1..100] of integer;

var
f: Text;
str, spr, subStr, spacer: String;
char: Char;
len, n, a, i: Longint;
sprList, spsList, spsLenList: mas;

begin
spr := '.GGTTTATCCCCGCT..CGCGGGGAAC.C';
len := length(spr);
writeln(spr, ' ', len);
writeln;

Assign(f, 'ecoli.txt');
Reset(f);
writeln('Position and SPR:');
{посимвольное чтение файла}
while (not Eof(f)) do {пока не достигнут конец файла}
begin
Read(f, char);
n := n + 1; {счетчик прочитанных нуклеотидов}
str := str + char; {добавляем прочитанный нуклеотид в строку}
if n >= len then
begin
if (length(str.MatchValue(spr, RegexOptions.None)) > 0) then
begin
a := a + 1;
sprList[a] := n - len + 1;
write(sprList[a]);
writeln(' ', str);
end;
str := copy(str, 2, length(str)); {отбрасываем первый символ str, чтобы
начало строки на следующем шаге приходилось на следующий нуклеотид}
end;
end;
Close(f);
writeln();
writeln('The total number of SPR: ', a);
writeln();

{расчет номера позиций всех спейсеров, а также их длины}
for i := 2 to a do
begin
spsList[i] := sprList[i - 1] + len; {расчет позиции спейсера}
spsLenList[i] := sprList[i] - sprList[i - 1] - len; {расчет длины
спейсера}
end;

spsList[12] := 0;
spsLenList[12] := 0;

Reset(f);
n := 0;
```

```

{посимвольное считывание файла с целью формирования подстроки от первого
нуклеотида первого спейсера до последнего нуклеотида последнего}
while n < (sprList[a] - 1 + len) do {пока не достигнут конец последнего
SPR}
begin
Read(f, char);
n := n + 1;
subStr := subStr + char; {добавляем прочитанный нуклеотид в строку}
{ограничение длины считываемой подстроки}
if n > (sprList[a] - sprList[1] + len) then subStr := copy(subStr, 2,
length(subStr)); {удаление первого символа подстроки}
end;
Close(f);

writeln('Spacers:');

for i := 1 to a do
begin
if spsLenList[i] <> 0 then
begin
spacer := copy(subStr, (spsList[i] - sprList[1] + 1), spsLenList[i]);
{выделяем подстроку, отвечающую спейсеру}
writeln(i, ' ', spsList[i], ' ', spacer, ' ', length(spacer));
end
else writeln(i, ' - - 0');
end;
end.
    
```

Результат работы программы (вывод на экран):

.GGTTTATCCCCGCT..CGCGGGGAAC.C 29

Position and SPR:

```

2877823 CGGTTTATCCCCGCTAACGCGGGGAACTC
2877884 CGGTTTATCCCCGCTGGCGCGGGGAACTC
2877945 CGGTTTATCCCCGCTGGCGCGGGGAACTC
2878006 CGGTTTATCCCCGCTGGCGCGGGGAACTC
2878067 CGGTTTATCCCCGCTGGCGCGGGGAACTC
2878128 CGGTTTATCCCCGCTGGCGCGGGGAACTC
2878190 CGGTTTATCCCCGCTGGCGCGGGGAACTC
2878252 CGGTTTATCCCCGCTGGCGCGGGGAACTC
2878313 CGGTTTATCCCCGCTGGCGCGGGGAACTC
2878374 CGGTTTATCCCCGCTGGCGCGGGGAACTC
2878435 CGGTTTATCCCCGCTGGCGCGGGGAACTC
2904013 TGGTTTATCCCCGCTGGCGCGGGGAACTC
2904074 CGGTTTATCCCCGCTGGCGCGGGGAACAC
2904135 CGGTTTATCCCCGCTGGCGCGGGGAACAC
2904196 CGGTTTATCCCCGCTGGCGCGGGGAACAC
2904257 CGGTTTATCCCCGCTGGCGCGGGGAACAC
2904318 AGGTTTATCCCCGCTGGCGCGGGGAACAC
2904379 CGGTTTATCCCCGCTGGCGCGGGGAACAC
    
```

The total number of SPR: 18

Spacers:

```

1 - - 0
2 2877852 GTAGTCCATCATTCACSTATGTCTGAACTCC 32
3 2877913 CCGGGGGATAATGTTTACGGTCATGCCCCCC 32
4 2877974 TGGGCGGCTTGCCCTGCAGCCAGCTCCAGCAG 32
5 2878035 AAGCTGGCTGGCAATCTCTTTTCGGGTGAGTC 32
6 2878096 TAGTTTCCGTATCTCCGGATTTATAAAGCTGA 32
7 2878157 GCAGGCGGCGACGCGCAGGGTATGCGCGATTCG 33
8 2878219 GCGACCGCTCAGAAATTCAGACCCGATCCAAA 33
    
```

9 2878281 TCAACATTATCAATTACAACCGACAGGGAGCC 32
10 2878342 AGCGTGTTTCGGCATCACCTTTGGCTTCGGCTG 32
11 2878403 TGC GTGAGCGTATCGCCGCGCGTCTGCGAAAG 32
12 - - 0
13 2904042 GACAGAACGGCCTCAGTAGTCTCGTCAGGCTC 32
14 2904103 CTGTTTTTCGCAAATCTATGGACTATTGCTATT 32
15 2904164 GGGCGCACGGAATACAAAGCCGTGTATCTGCT 32
16 2904225 TGGCTCTGCAACAGCAGCACCCATGACCACGT 32
17 2904286 GAAATGCTGGTGAGCGTTAATGCCGCAAACAC 32
18 2904347 ATTACGCCTTTTTGCGATTGCCCGTTTTTGC 32