Algorithms in Bioinformatics

Andrey Prjibelski (andrewprzh@gmail.com),
Ira Vasilinetc (vasilinetc.ira@gmail.com)

March 7, 2014

Alignment 2

Task 6

For given set of strings $\{s_i\}$ construct pairwise alignments and guide tree. Output tree in Newick-like format.

Input:

Fasta file with several sequences.

Output:

Guide tree in Newick like format.

Example:

Input:

>seq1

ac

>seq2

aa

>seq3

tt

Output:

((seq1 : 1 : seq2) : 2 : seq3)

Task 7

Given sequence s and n ranges $\{[l_i, r_i]\}$. For each range calculate the number of all palindromic substrings in $s[l_i, r_i]$.

Input:

Fasta file with sequences and file with ranges.

Output:

For each range calculate the number of all palindromic substrings in $s[l_i, r_i]$.

Example:

Input:

>seq1

acacaca

File 2:

```
3
0 0
0 1
0 2
0 3
```
Output:
```
1
2
4
6
```

Task 8

For a given sequence S and an integer $k > 0$ construct Bloom filter that contains all k-mers from this sequence. Find the false positive rate for your Bloom filter. Bloom filter answers a query whether given k-mer is substring of S.

Input:
Fasta file with sequence S, file with an integer $k > 0$, and a set of k-mers (one per line).

Output:
False positive rate for all 4^k k-mers; for every k-mer from the second file output 1 if the k-mer is a substring of S, 0 otherwise.

Example:

Input file 1:
>seq1
acgtacgt
Input file 2:
2
ac
aa
Output:
FP = 10 %
1
0