

Algorithms in Bioinformatics

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

April 9, 2014

Task 13

Implement Nussinov algorithm for predicting secondary RNA structure with the following constraints:

- Complimentary base pairs should be at least 4 non-complimentary base pairs apart
- One pair of complimentary sequences is preferred of two pairs of complimentary sequences with the same total length

Input:

RNA sequence in FASTA format.

Output:

RNA secondary structure. Output all pairs of complimentary sequences with their positions in the RNA sequence.

Example:

Input:

>seq

CCCCCCACGUUUUUCCCUUUUGGGUUUUUACGUCCCCCCC

Output:

6,ACGU-29,ACGU

14,CCC-21,GGG

Task 14

Implement Zuker algorithm for predicting secondary RNA structure.

Input:

RNA sequence in FASTA format.

Output: RNA secondary structure. Output all pairs of complimentary sequences with their positions in the RNA sequence.

Example:

Input:

>seq

CCCCACGUUUUUACGUCCCC

Output:
4,ACGU-12,ACGU

Task 15

We define an exact tandem repeat as a subsequence that appears consecutively at least twice. For example, AATT is a tandem repeat in CCAATTAATTAATTCC. Design an efficient algorithm for finding the longest exact tandem repeat within a given sequence.

Input:

Sequence in FASTA format.

Output:

Longest tandem repeat in FASTA format.

Example:

Input:
>seq
AAAAAACGTACGTACGTAAAAAA
Output:
>repeat
ACGT

Task 16

Design an efficient algorithm for finding the longest exact repeat with at most one mismatch in a given sequence.

Input:

Sequence in FASTA format.

Output:

Two instances of the longest repeat with maximum 1 mismatch in FASTA format.

Example:

Input:
>seq
ACGTAAAAAAAAAACCAAATAAAACGG
Output:
>repeat1
AAAATAAAAC
>repeat2
AAAAAAAAC