

Stem-loop structure search

Scalabrin S (1), Policriti A (1), Morgante M (2)

(1) Dipartimento di Matematica ed Informatica, Universita' di Udine, Udine

(2) Dipartimento di Scienze Agrarie ed Ambientali, Universita' di Udine, Udine

Motivation

The base-pairing of a nucleic acid secondary structure is a sort of biological palindrome. The base pairs of nucleic acid stems nest in a palindromic fashion with complementary base pairings rather than identical letters. In addition, nucleic acid stems are usually separated by a loop, i.e. a non palindromic sequence. Biological examples of such a structure can be found in hairpins at the 3' of helitrons, MITEs, microRNAs and tRNAs. At the moment, the problem of finding such a structure is solved through dynamic programming or in linear time with the use of the lowest common ancestor preprocessing.

Methods

We propose an algorithm which makes use of matching statistics on a suffix tree, is still linear on the size of the input string S , does not use the lowest common ancestor preprocessing and builds the suffix tree only for S and not for its complement reversed string, resulting in a great saving of memory, highly valuable in long sequence scan. Stems are easily detected using a suffix tree. Successively, the loop constraint (maximum distance between two stems) is solved linearly, on the number of putative stems, sorting the lists of occurrences in the input string S and its complement reversed string, and exploiting the linearity of the constraint.

Availability: Contact the authors via email to receive a copy of the software

Contact email: scalabrin@dimi.uniud.it