# An Algorithm for Finding Common Secondary Structure Motifs in a Set of Unaligned RNA Sequences - (session: Novel Algorithms for Bioinformatics)

Giulio Pavesi, Giancarlo Mauri, Graziano Pesole

Università Milano Bicocca

We present an algorithm for finding conserved secondary structure motifs in a set of RNA sequences, that is, secondary structure elements that appear in all or most of the secondary structures formed by the sequences of the set.
Differently from the methods introduced so far for this problem, the approach we present does not compute an alignment of the sequences beforehand, nor takes into account sequence similarity in any way, but looks directly for structural similarities. Thus, it can be applied also to cases when RNA sequences do not present significant similarity in their nucleotide sequence. The algorithm takes as input the secondary structure of the sequences, exhaustively enumerates all pattern representing feasible secondary structure elements up to a maximum size (that can equal the length of the sequences), searches for each one in the structures, and finally reports those patterns that appear in all or most of the sequences of the set.
Occurrences of patterns can be approximate, that is, can differ in the size of a stem, of an internal loop, in the presence or not of a bulge, and so on: the type and degree of approximation can be chosen at freedom by the user.
The input structures can be either determined experimentally, or predicted by one of the existing methods. In the latter case, we show how the algorithm can deal with the uncertainty deriving from predictions, by considering different alternative secondary structures for each sequence.
Experiments have shown that the algorithm, coupled with existing secondary structure prediction methods, is able to discover efficiently known RNA structural motifs, such as histone and IRE stem-loop motifs in RNA untranslated regions, as well as structural motifs shared by the members of different virus families.