

# **EFFICIENT ANNOTATION OF NON-CODING RNA STRUCTURES INCLUDING PSEUDOKNOTS VIA AUTOMATED FILTERS**

Chunmei Liu and Yinglei Song  
Department of Computer Science, University of Georgia  
Athens, Georgia 30602, USA

Ping Hu  
Department of Genetics, University of Georgia  
Athens, GA 30602, USA

Russell L. Malmberg  
Department of Plant Biology, University of Georgia  
Athens, GA 30602, USA

Liming Cai  
Department of Computer Science, University of Georgia  
Athens, Georgia 30602, USA

Computational search of genomes for RNA secondary structure is an important approach to the annotation of non-coding RNAs. The bottleneck of the search is sequence-structure alignment, which is often computationally intensive. A plausible solution is to devise effective filters that can efficiently remove segments unlikely to contain the desired structure patterns in the genome and to apply search only on the remaining portions. Since filters can be substructures of the RNA to be searched, the strategy to select which substructures to use as filters is critical to the overall search speed up. Such an issue becomes more involved when the structure contains pseudoknots; approaches that can filter pseudoknots are yet available. In this paper, a new effective filtration scheme is introduced to filter RNA pseudoknots. Based upon the authors' earlier work in tree-decomposable graph model for RNA pseudoknots, the new scheme can automatically derive a set of filters with the overall optimal filtration ratio. Search experiments on both synthetic and biological genomes showed that, with this filtration approach, RNA structure search can speed up 11 to 60 folds while maintaining the same search sensitivity and specificity of without the filtration. In some cases, the filtration even improves the specificity that is already high.