

# **Discovery of Regularities in Biological Sequences and Application in Sequence Analysis**

**Katerina Perdikuri**

*Computer Engineering and Informatics Department, University of Patras,  
26500 Patras, Greece  
E-mail: [perdikur@ceid.upatras.g](mailto:perdikur@ceid.upatras.g)*

A deciding factor in sequence analysis is that biomolecules, i.e.: DNA, RNA and proteins that are responsible for storing and utilizing the genetic material in an organism, can be described as strings over finite alphabets. The string representation of biomolecules allows for a wide range of algorithmic techniques concerned with strings to be applied for analyzing and comparing biological data.

Looking for regularities (i.e.: periods, covers, seeds, tandem repeats, etc) in one or more biological sequences is a well-known algorithmic technique with many applications. For example looking for regularities in a single biological sequence can be used to reconstruct part of the evolutionary history of the sequence or to identify the sequence among other sequences. Moreover known regularities such as covers or tandem repeats can be used in biological sequence compression techniques to efficiently store or analyze billions of nucleotides. Another interesting application is comparing sequences and producing sequence families based on common regularities.

In this contribution we describe existing algorithms for finding regularities in biological sequences and present the corresponding applications in sequence analysis.