

# Genetic algorithm learning: A new approach for the discovery of small, diffuse patterns in genomic sequences

Genomic sequence analysis has centered on the identification of critical signals associated with genes and regulatory regions. A variety of methods, based on statistical inference and machine learning, have been used to detect the presence of genes, promoter signals and associated patterns. However, such approaches have limitations. In particular, they generally require the pattern of interest to be easily discernible from random sequence noise by at least one characteristic, e.g. distinctive sequence composition, consistent length or well-defined locations within the genome. For example, the TATA box promoter sequence has a distinctive composition, TATAA, a generally invariable length of five nucleotides and is almost always located upstream of protein-coding eukaryotic genes. However, not all signals involved with the regulation of genes are characterized by such distinctive features. The signals associated with alternative splicing tend to be very short patterns (as few as 2 nucleotides), have weak sequence conservation and their locations within genes can be highly variable. Yet it is precisely these types of signals that we must uncover as we learn how genes are manipulated to generate organismal complexity.

Given the limitations of pattern finding algorithms, a different approach has been proposed that uses genetic algorithms to better define patterns of interest. Genetic algorithms mimic biological evolution and have been applied with great success to solve a wide variety of search and optimization tasks. By transforming the pattern finding challenge in genomes into a search and optimization task, genetic algorithms offer a powerful new way to identify the features associated with a given regulatory signal. We present the genetic algorithm approach and describe one application to the discovery of the signals associated with an unusual regulatory process known as RNA editing.

James Thompson and Shuba Gopal. (2006). "Genetic algorithm learning as a robust approach to RNA editing site prediction," *BMC Bioinformatics*, **7**:145. Correction in *BMC Bioinformatics*, **7**:406.