

---

**ARTICLES**

## **GeneMark: Parallel Gene Recognition for Both DNA Strands**

**Mark Borodovsky and James McIninch**

School of Biology and Schools of Biology and Mathematics, Georgia Institute of Technology, Atlanta, GA 30332-0230, USA.

The problem of predicting gene locations in newly sequenced DNA is well known but still far from being successfully resolved. A novel approach to the problem based on the frame dependent (non-homogeneous) Markov chain models of protein-coding regions was previously suggested. This approach is, apparently, one of the most powerful "search by content" methods. The initial idea of the method combines the specific Markov models of coding and non-coding region together with Bayes' decision making function and allows easy generalization for employing of higher order Markov chain models. Another generalization which is described in this article allows the analysis of both DNA strands simultaneously. Currently known gene searching methods perform the analysis of the two DNA strands in turn, one after another. In doing this all the known methods fail in the sense that they generate false (artificial) prediction signals for the given strand when the real coding region is located on the complementary DNA strand. This common drawback is avoided by employing the Bayesian algorithm which uses an additional non-homogeneous Markov chain model of the "shadow" of the coding region -- the sequence which is complementary to the protein-coding sequence.