Utilização de pair-HMM e GPHMM em alinhamentos de sequências e predição de genes comparativa.

A proposta desta resenha é fazer uma revisão sobre os modelos de pair-HMM e pair-HMMs generalizados., assim como suas aplicações em alinhamentos de pares sequências, alinhamentos múltiplos, e predição de genes comparativa. Discutiremos ainda sobre arquiteturas de pair-HMM que permitem a utilização de diferentes modelos de alinhamento para regiões com maior ou menor pressão seletiva.

Referências

[1] W. H. Majoros, M. Pertea and S. L. Salzberg. Efficient implementation of a generalized pair hidden Markov model for comparative gene finding. Bioinformatics **21** (2005), 1782-1788

[2] A. Löytynoja, N. Goldman, A model of evolution and structure for multiple sequence alignment. Philos Trans R Soc Lond B Biol Sci. 2008 Dec 27;363(1512):3913-9.

Abstracts

[1]

Motivation: The increased availability of genome sequences of closely related organisms has generated much interest in utilizing homology to improve the accuracy of gene prediction programs. Generalized pair hidden Markov models (GPHMMs) have been proposed as one means to address this need. However, all GPHMM implement- ations currently available are either closed-source or the details of their operation are not fully described in the literature, leaving a significant hurdle for others wishing to advance the state of the art in GPHMM design. Results: We have developed an open-source GPHMM gene finder, TWAIN, which performs very well on two related Aspergillus species, A.fumigatus and A.nidulans, finding 89% of the exons and predicting 74% of the gene models exactly correctly in a test set of 147 con- served gene pairs. We describe the implementation of this GPHMM and we explicitly address the assumptions and limitations of the sys- tem. We suggest possible ways of relaxing those assumptions to improve the utility of the system without sacrificing efficiency beyond what is practical. Availability: Available at http://www.tigr.org/software/pirate/twain/ twain.html under the open-source Artistic License.

[2]

We have developed a phylogeny-aware progressive alignment method that recognizes insertions and deletions as distinct evolutionary events and thus avoids systematic errors created by traditional alignment methods. We now extend this method to simultaneously model regional heterogeneity and evolution. This novel method can be flexibly adapted to alignment of nucleotide or amino acid sequences evolving under processes that vary over genomic regions and, being fully probabilistic, provides an estimate of regional heterogeneity of the evolutionary process along the alignment and a measure of local reliability of the solution. Furthermore, the evolutionary modelling of substitution process permits adjusting the sensitivity and specificity of the alignment and, if high specificity is aimed at, leaving sequences unaligned when their divergence is beyond a meaningful detection of homology.