Beoung Hun Lee

CS 199

March 23, 2019

Edwin Solares

In this paper, the researcher compares Genemark program, which utilizes Viterbi algorithm, with Genemark.fba, which utilizes Forward-Backward algorithm(FB algorithm). The original Genemark algorithm uses heuristic approach instead of rigorous approach as the sequence is assigned into single functional category and average aposteriori possibility. FB algorithm assigns aposteriori probability of a functional state to each nucleotide instead. This rigorous approach should result in higher accuracy of gene prediction compared to the original Genemark program. Genemar.fba is also compared with GenemarkS, which uses more developed HMM architecture. Although the original Genemark algorithm has a accuracy over 90%, the goal is to find algorithm that has higher accuracy for prokaryote genome.

Result show that Genemark.fba is capable of identifying all genes that were identified by the Genemark and showed improvement in accuracy of gene prediction. Plot of posteriori probabilities of genomic sequences containing nhaR, insB1, insA1 and rpsT show that both Genemark.fba and Genemark correctly identified all the genes in the sequence, with Genemark.fba showing less fluctuation in aposteriori possibility plot. Results of complete genome analysis of Genemark.fba were compared with GenemarkS and Genemark. Genemark.fba outperformed Genemark on all prokaryote genome except for one genome, while GenemarkS’s performance varied depending on the species, as Genemark.fba outperformed GenemarkS in E.coli, M.jannaschii and Synechocytis but underperformed in B.subitlis, A.fulgidius and H.influenzae. Genemark.fba took more computational time when compared to GenemarkS and Genemark, as both approximates aposteriori possibilities, while Genemark.fba takes rigorous approach to obtain aposteriori possibilities. This result shows that there is still a room to improve accuracy of Genemark using a different algorithm.

Reference

1. Azad, R. and Borodovsky, M. (2004). Probabilistic methods of identifying genes in prokaryotic genomes: Connections to the HMM theory. Briefings in Bioinformatics. Vol 5. 118-130.