Identification of Pathogen and Anti-Ebola Drug Targets Using Bayes' Theorem and Information Entropy

Cho, Min Jean

To develop simple methods for identifying pathogens and anti-Ebola drug targets based on their DNA, RNA, and protein sequences, Bayes' theorem and information entropy was applied to sequence analysis. It was hypothesized that the conditional probability of a DNA/RNA sequence (16S rRNA gene sequence for bacterial pathogens and RNA sequences of RNA-dependent RNA polymerase [RdRp] for viral pathogens) from an unknown pathogen being a member of a particular species could be the posterior probability, which could be estimated from prior probability and likelihood function using Bayes' theorem. The use of highly informative region of DNA sequences defined using information entropy was effective for the species identification method. The overall performance of the developed method was evaluated by simulation tests using selected DNA/RNA sequences, and all tested sequences were correctly identified. In addition, anti-Ebola drug targets were identified from the genome sequences of Ebola virus using information entropy. A total of 15 anti-Ebola miRNA targets were identified from three low entropy regions of Ebola virus RdRp RNA sequences, and three miRNA drugs were designed. Also, among highly conserved regions of RdRp protein sequences, one region was identified as a candidate target for ligand-based drugs. For preventing Ebola infection, three mimotope-based peptide vaccines were designed from the protein sequences of Ebola glycoprotein. The target sites of these anti-Ebola peptide vaccines were conserved in all known subtypes of Ebola virus and were predicted to be surface-exposed regions. The entropy-based method for predicting drug targets and designing miRNA drugs and peptide vaccines will be a valuable tool for developing effective drugs against life-threatening pathogen.