Evaluation of predictor programs of genomic islands

Antonio Camilo da Silva Filho, Diônata Willian Augusto, Izabella Castilhos Ribeiro dos Santos Weiss, Paulo Afonso Bracarense Costa, Jeroniza Nunes Marchaukoski

Universidade Federal do Paraná (UFPR), Setor de Educação Profissional e Tecnológica, Programa de Pós Graduação em Bioinformática, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)

Genomic islands (GI) are segments of DNA characterized by gene content and are associated with a function or adaptive capacity of medical and environmental interest, being directly related to bacterial evolution. The GIs are present in different taxonomic clades and are often obtained by horizontal gene transfer. The classification of a GI is given accordingly to the function of encoded genes, it can be classified as pathogenic, symbiotic, metabolic, saprophagic, ecological or resistance islands. Among these, stands out the pathogenicity islands (PAIs) that has the ability to transmit pathogenic properties in a single genetic event thus allowing the evolution and the emergence of new pathogens. In this work, we evaluated predictors programs of genomic islands using only complete genomes. We analyzed the similarity between the results generated by the predictive programs and the gold standard classification. For testing programs we selected: Alien Hunter, Zisland Explorer, IslandViewer3, GI Hunter, Predict Bias and Gipsy. It was considered tree points of criteria to choose the predictors: 1- the availability (the predictor is local or web); 2 – article published less than three years and 3 - that has more than 20 citations in the last three years. We chose Escherichia coli CFT073 as a test organism. For the gold standard we consider the published pathogenic islands of Escherichia coli CFT073 for that it was researched and proved *in vitro* and this class of pathogenic islands is the most studied. In the analysis we observed differences in the results of the compared predictors to the 13 pathogenic islands which was determined as gold standard. Determination of these islands was analyzed *in vitro* in previous research. Each predictor showed different results, among them are the start and end positions of the island, size, number of coding sequences and the number of predicted islands. Alien Hunter identified 13 PAIs of the gold standard in 86 GIs predicted, IslandViewer3 identified 11 PAIs of 78, Predict Bias identified 11 PAIs of 76, Gipsy identified 12 PAIs of 26, GI Hunter identified 7 PAIs of 18 and Zisland Explorer identified 3 PAIs out of 11GIs. In order to evaluate the differences between the results, we need to identify what tools and methodologies have high and low sensitivity, specificity and accuracy determining the positive and false negative results, as well as real positive and negative islands. This research will help to find the best strategies to prediction, being constructive and direct for researchers to obtain promising results in studies related to genomic islands.