CONSIDER FOR TALK

2022 SEA Symposium Abstract

LeTourneau University

Longview TX

Corresponding Faculty Member: Fred Baliraine (fredbaliraine@letu.edu)

How Reliable Are DNABIND and DNAbinder Software for Identifying Actinobacteriophage DNA-binding Proteins?

Christina A Holder, Christina E Spencer, Angela L Salazar, Frederick N Baliraine

DNA-binding proteins (DNA-BPs) are an important class of proteins with DNA-binding domains that enable them to have general or specific affinity for single- or double-stranded DNA molecules. Because of their ability to target specific DNA sequences, DNA-BPs play key roles in activities such as transcription regulation and prokaryotic host defense. Therefore, identifying DNA-BPs is important in genome annotation. Many annotated phage genomes in the phagesDB and NCBI GenBank databases have open reading frames with unknown functions, also known as hypothetical proteins (HPs). Some of these HPs may be previously unidentified DNA-BPs. The purpose of this study was to access the reliability of the DNABIND and DNAbinder software for determining DNA-BP gene function. While annotating *Mycobacterium* phage Dynamo, these two DNA-BPs detection software were used to check HPs for DNA-binding activity. Dynamo gp 52 had a strong DNA-binding prediction in both programs and strong alignments in HHpred to a TATA-box binding protein. However, a function could not be confidently determined for lack of significant alignments with known functions in phagesDB and NCBI. DNABIND and DNAbinder were further investigated for their sensitivity and specificity by utilizing published phages from various clusters. Seven known non-DNA-BPs and eight DNA-BPs were analyzed from each phage. DNABIND was run at false positive rate values ranging from 1% to the default 15% rate. DNAbinder was run using an amino acid composition with main, realistic, and alternate datasets. Preliminary results indicate that DNABIND is fairly reliable, while DNAbinder's reliability depends on which of the three datasets was used in the analysis. The utility and implications of the collected data for using these two software to call DNA-BPs will be discussed.