SEQUENCE FINGERPRINTS OF HOST PROTEIN MIMICRY IN BACTERIAL PATHOGENS

Andrew C. Doxey, Trevor Charles, Brendan J McConkey

Department of Biology, University of Waterloo, Canada

Many bacterial pathogens are thought to interfere with host processes through molecular mimicry of host proteins. However, no studies have been done to quantitatively determine whether host protein mimicry is a distinguishing feature of microbial pathogens. We conducted a large-scale analysis of human-microbe protein similarities for 189 microbial genomes including both human pathogenic and non-pathogenic bacteria. Assessments were done separately for different functional categories of human proteins according to their gene ontology (GO) functions. It was found that sequence mimicry of host proteins is significantly elevated in human pathogens, with the highest detected levels of mimicry occurring in key structural components of the human extracellular matrix (e.g., mimicry of collagen). In addition, elevated patterns of host protein mimicry were found in virulent strains/species of *E. coli*, *Vibrio*, and *Bacillus* relative to their non-virulent forms. The detected mimics in pathogenic bacterial genomes consist largely of known virulence factors, while many are likely novel proteins involved in bacterial pathogenesis. This work provides a means for prediction of human pathogenicity from genome sequence, and represents a promising new avenue in pathogenomics research.