Contribution of horizontal gene transfer to microbial evolution and pathogenicity

Rajeev K. Azad (Rajeev.Azad@unt.edu)

Department of Biological Sciences, Department of Mathematics
University of North Texas, Denton, Texas 76203, USA

It is now well recognized that microbial genomes are chimeras of genes with different ancestries. The process of horizontal gene transfer makes microbes the most versatile and dynamic organisms. The antibiotic resistance and virulence traits in microorganisms have often been attributed to horizontal gene transfer. Detection and characterization of genes undergoing horizontal flux is central to understanding microbial genome evolution. Of immense interest is deconstructing the evolution of antibiotic resistance pattern and virulence. The first step in achieving this goal is to identify robustly the horizontally acquired genes in microbial genomes. Because the composition of horizontally acquired genes reflects the mutational proclivities of donor genomes, they appear 'atypical' in their recipient's genome. This is exploited by most methods which assess their atypicality by measuring their compositional biases against the genome background. Yet these methods incur high false positive and false negative rates due to the significant overlap of compositional spectrum of native and alien genes. To identify alien genes robustly and infer their sources, we have developed a multipronged strategy. First, the genes that are similar to each other are grouped together using a novel gene clustering method based on entropic divergence measures. Misclassification of compositionally ambiguous genes is circumvented by invoking gene context and operon structural information. This procedure not only minimizes the Type I and Type II errors of misclassification but also identifies the donor clades initiating gene transfers. We have also developed a recursive segmentation method that can simultaneously analyze multiple genes in large genomic islands, allowing identification of weakly atypical islands. At the core of this method is a highly sensitive, generalized divergence measure to exploit the predictive power of higher order models in analyzing large genomic regions. Our method performed consistently well in identifying transfer events in both artificial chimeric genomes and genuine bacterial genomes. In application to genomes of well studied pathogens *Salmonella enterica typhi* CT18 which causes typhoid fever and *Pseudomonas aeruginosa* LESB58 implicated in chronic lung infections, our method could decipher robustly both the span and boundaries of the known pathogenicity islands validated using microarray and PCR experiments. We identified 30 novel pathogenicity islands in *S. enterica typhi* CT18 harboring genes known to be upregulated during pathogenesis and 10 novel islands in *Pseudomonas aeruginosa* LESB58 supported by evidence from literature and phylogenetic studies. In addition, our method could localize the antibiotic resistance islands in the genomes of hospital- and community-associated methicillin-resistant *Staphylococcus aureus*, thus facilitating the understanding of evolution of antibiotic resistance in related strains. Furthermore, we inferred the roadmap of gene flow by assessing the similarity of each gene cluster from a genome against the gene clusters from all other genomes. This helps determine unambiguously the direction of gene flow. We used this strategy to decipher the gene sharing network in over 700 prokaryotic genomes, representing the sampling of prokaryotic genomes in the databases. Our results show that the core genes are relatively recalcitrant to transfer; in contrast, alien gene clusters merged frequently displaying a significant fraction of genetic elements in motion across species boundaries.