Automated detection of prokaryotic genomic islands using a comparative genomics approach

Morgan G.I. Langille, William W.L. Hsiao, and Fiona S.L. Brinkman
Dept. of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, B.C., Canada

Genomic islands (GIs) are defined as clusters of genes of potential horizontal origin in bacterial and archaeal genomes. GIs, including pathogenicity islands, are frequently associated with particular adaptations of microbes that are of medical, agricultural or environmental importance. There is a growing interest in identifying and analyzing these regions in pathogenic bacteria, as part of efforts to identify genes involved in antimicrobial resistance and virulence that warrant further study.

One general approach to identify GIs involves using genome sequence composition and is based on the observation that organisms often exhibit distinct sequence compositions in their genome, and a bias in one sub-region compared to the rest of the genome may suggest horizontal gene transfer. However, GI predictors based on sequence composition may lead to false positive GI predictions due to other factors causing a bias in sequence composition, such as high gene expression level. Further, these approaches may miss the identification of GIs that have been acquired from genomes with similar sequence composition or ancient GIs events that may have ameliorated to the host genome over time.

We are interested in testing the accuracy of the numerous GI predictors that currently exist. We have therefore developed an additional high-throughput comparative genomics approach that is independent of sequence composition to identify GIs. In addition to this positive dataset of GIs, we have constructed a negative dataset (regions not containing GIs), by finding conserved regions across multiple genomes. These datasets have allowed us to test the accuracy of several sequence based GI predictors that will lead to improved detection of these GI regions that are of particular interest.