

Identification of Genomic Islands by Pattern Discovery

Nita Parekh

1. Center for Computational Natural Sciences and Bioinformatics (CCNSB), International Institute of Information Technology (IIIT-H) Gachibowli, Hyderabad 500032, India,

Book: Pattern Discovery Using Sequence Data Mining: Applications and Studies

Abstract: Pattern discovery is at the heart of bioinformatics, and algorithms from computer science have been widely used for identifying biological patterns. The assumption behind pattern discovery approaches is that a pattern that occurs often enough in biological sequences/structures or is conserved across organisms is expected to play a role in defining the respective sequence's or structure's functional behavior and/or evolutionary relationships. The pattern recognition problem addressed here is at the genomic level and involves identifying horizontally transferred regions, called genomic islands. A horizontally transferred event is defined as the movement of genetic material between phylogenetically unrelated organisms by mechanisms other than parent to progeny inheritance. Increasing evidence suggests the importance of horizontal transfer events in the evolution of bacteria, influencing traits such as antibiotic resistance, symbiosis and fitness, virulence, and adaptation in general. In the genomic era, with the availability of large number of bacterial genomes, the identification of genomic islands also form the first step in the annotation of the newly sequenced genomes and in identifying the differences between virulent and non-virulent strains of a species. Considerable effort is being made in their identification and analysis and in this chapter a brief summary of various approaches used in the identification and validation of horizontally acquired regions is discussed.