Abstract

Large amount of space is required to store biological sequences in DNA database like GenBank sequence database. The data storage for biological sequences has become very essential in today’s current situation. Standard compression algorithms are not competent enough to compress biological sequences. In recent times, special algorithms have been introduced specifically for the purpose of compressing the biological sequences like DNA and protein sequences. In this paper, the Burrows-Wheeler Transform (BWT) based approaches are explored to compress the biological sequences. In comparison with the existing general purpose compression algorithms, the proposed BWT based method compresses these types of sequences better and at the same time the cost of Burrows-Wheeler Transform is almost insignificant.

References


Index Terms
Computer Science
Information Sciences

Keywords
DNA sequence compression, Burrows-Wheeler Transform, BWT and genome.