Abstract

Genome resequencing produces enormous amount of data daily. Biologists need to frequently mine this data with the provided processing and storage resources. Therefore, it becomes very critical to professionally store this data in order to efficiently browse it in a frequent manner. Reference-based Compression algorithms (RbCs) showed significant genome compression results compared to the traditional text compression algorithms. By avoiding the complete decompression of the compressed genomes, they can be browsed by performing partial decompressions at specific regions, taking lower runtime and storage resources. This paper introduces the inCompressi algorithm that is designed and implemented to efficiently pick sequences from genomes, that are compressed by an existing Reference-based Compression algorithm (RbC), through partial decompressions. Moreover, inCompressi performs a more efficient complete genome decompression compared to the original decompression algorithm. The experimental results showed a significant reduction in both runtime and memory consumption compared to the original algorithm.
References


Index Terms

Computer Science

Keywords
Reference-based Genome Compression; Partial Genome Decompression; Browsing Genome Sequences; inCompressi.