

Department of Biological Sciences

Carolina Gabbai

Mentor: Brian Morton

Development of a Java Class to Analyze Structural Heterogeneity in DNA
Sequence: Application with Chloroplast Genomes

Plastids are thought to be the result of an endosymbiotic event leading to the formation of a novel organism, composed of both the original eukaryotic symbiont and the prokaryote it engulfed. Sequenced prokaryotic genomes have been shown to have several structural features such as codon adaptation, clustering of genes on the leading strand, and compositional heterogeneity in strand skew and G+C content. Since plastid genomes are prokaryotic in origin, this paper analyzes certain structural characteristics in plastid genomes that are found in bacterial chromosomes. The chloroplast genome *Euglena gracilis* has been shown to have a similar skew pattern to sequenced bacterial genomes such that the leading strand displays equivalent bias all around the genome. Another structural phenomenon seen in prokaryotes is a relative increase and then decrease of the G+C composition at third codon positions (G+C3) across the genome. This paper is an attempt to determine if certain structural features observed in bacteria are also found in plastid genomes. For the analysis of G+C3 heterogeneity along a sequence, we developed a Java class and applied it in an analysis of fully sequenced plastid genomes.