

Poster H-29
Network Complexity and
GC-asimetry in bacterial genomes



Authors:

Ney Lemke (*Departamento de Física e Biofísica, Instituto de Biociências de Botucatu, Unesp*)
Gilberto Corso (*Departamento de Biofísica e Farmacologia, Centro de Biociências, UFRN*)
Gunther Gerhardt (*Pós-Graduação em Ciência Médica, Universidade Federal do Rio Grande do Sul*)
Sergio Echeverrigarray (*Instituto de Biotecnologia, UCS*)

Short Abstract: In this work we present a method to analyse DNA sequences based on graph theory: the DNA Sequence Network (DSN). We measure the clustering coefficient C to describe DSN. The intervall around $0.5 < 0.6$ is unfavorable for sequences. On the same region C shows a maximum.

Long Abstract:

In this work we further develop the DNA Sequence Network DSN, a method to extract information from DNA sequences. We apply this tool to a set of 152 bacterial genomes obtained from GenBank. The network is constructed as follows: vertices are nucleotide triplets and a link between two triplets is established if two triplets are juxtaposed somewhere in the DNA sequence. This network incorporates information on base pair frequency, triplet adjacency and general structure organization of the sequence. We choose the normalized clustering coefficient C as complexity measure to describe DSN, we interpret this quantity as a measure of drift from randomness. We discuss this point considering GC assymetric histogram. The intervall around $0.5 < \text{GC content} < 0.6$ is not favorable for sequences. For the same region C shows a local maximum which indicates the existence of an unstable GC content region.

Poster H-29
Gnomon - NCBI gene prediction
tool for eukaryotic genomes



Authors:

Alexander Souvorov (*NCBI*)
Yuri Kapustin (*NCBI*)
Boris Kiryutin (*NCBI*)
Vyacheslav Chetvernin (*NCBI*)
Tatiana Tatusova (*NCBI*)
David Lipman (*NCBI*)

Short Abstract: We have developed a new gene prediction tool, Gnomon, for eukaryotic genome annotation. Gnomon constrains its core ab initio gene-finding algorithm using transcript and protein alignment data with a goal to minimize instances where a biological

expert could generate an obviously improved gene model.

Long Abstract:

We have developed a new gene prediction tool, Gnomon, for eukaryotic genome annotation. Gnomon constrains its core ab initio gene-finding algorithm using transcript and protein alignment data with a goal to minimize instances where a biological expert could generate an obviously improved gene model. Similar to GenScan, Gnomon uses a 3-periodic fifth-order Hidden Markov Model for the coding propensity score and incorporates descriptions of the basic transcriptional, translational and splicing signals, as well as length distributions and compositional features of exons, introns and intergenic regions. Depending on the amount of evidence (alignments), a model can be completely supported (no ab initio prediction at all), partially supported (some parts were reconstructed using ab initio) or pure ab initio prediction. Gnomon has been used in every recent NCBI annotation of eukaryotic genomes, ranging from Human to *Tribolium castaneum* (red flour beetle).