

Poster H-10

Successful clustering of ortholog groups by Bidirectional Best Hit (BBH) using organisms modeled from a single ancestral via stepwise mutation



Authors:

Francisco Prosdocimi (UFMG)
Maurício A Mudado (UFMG)
J Miguel Ortega (UFMG)

Short Abstract: We set up for modeling synthetic CDS sequences aside their respective deduced proteins and an ancestral organism containing was created. Afterward, ten different organisms were modeled from the single ancestral via stepwise mutation. Results suggest that BBH driven clustering of ortholog groups seems to properly stand for homology searches.

Long Abstract:

Perhaps the most favorite point of divergence amongst bioinformaticians is the biological meaning of the BLAST results, being similarity searches understood by some researchers as homology searches for ortholog and paralog genes. To approach this debate we set up for modeling synthetic CDS sequences aside their respective deduced proteins. The random proteins composition respected the proportion of occurrence of the codons, thus presenting six R per each M residue. An ancestral organism containing one thousand proteins of 100 amino acids long was created. Afterward, ten different organisms were modeled from the single ancestral via stepwise mutation, ranging from 50 up to 150 mutations per coding strand. Deduced proteins were aligned by BLAST searches using a given organism proteome as query and the remaining nine as formatted database. Remarkably, from the around 9 thousand expected hits, only thirteen were encountered under a reasonable E-value cutoff of $1E-10$, when 150 mutations had been introduced per molecule. The resultant alignments averaged around 35% similarity (from 22% up to 80%). However, when the BBH approach was applied with no restrictive E-value cutoff, clusters of orthologs have been formed with over 93.1% accuracy under 150 mutations per molecule, and 100% when 50 mutations per molecule have had applied. No evidence was detected so far for an event able to be classified as convergent evolution. Modeled ortholog groups provide an excellent opportunity for the development of clustering algorithms. We are currently automating the clustering process based on BBH and improving the elimination of paralogs bearing stop codons during synthetic evolution. In conclusion, BBH driven clustering of ortholog groups seems to properly stand for homology searches.