Lawrence Berkeley National Laboratory
Recent Work

Title
Tree based orthology and paralogy determination: Phylogenetically inferred groups (PHIGS)

Permalink
https://escholarship.org/uc/item/06x6g5h3

Authors
Dehal, Paramvir
Huang, Wayne
Boore, Jeffrey L.

Publication Date
2005-05-06

TREE BASED ORTHOLOGY AND PARALOGY DETERMINATION:
PHYLOGENETICALLY INFERRED GROUPS (PHIGS)

Paramvir Dehal, Wayne Huang and Jeffrey L. Boore
DOE Joint Genome Institute

The determination of orthology versus paralogy relationships amongst
genes is of great importance to both comparative, functional and
evolutionary genomics, as well as genome annotation. Our process
involves hierarchical clustering of the genes from the available whole
genome datasets such that gene clusters are consistent with the known
evolutionary relationships of the organisms. For each cluster of genes that
shares descent from a single common ancestor, a multiple sequence
alignment and a maximum likelihood phylogenetic tree is created. By
reconstructing the gene tree, not only can the tree be reconciled with the
known evolutionary tree in order to identify orthology and paralogy
relationships, but we also obtain a much more clear picture for rates of
sequence change along the tree and potential positive selection across the
lineages. In our process we are able to annotate genes using their
evolutionary history instead of relying on similarity measures such as top
or reciprocal best BLAST hits, which can be in error for genomes with
large amounts of duplication or rate variation. We present results from
clustering genes from over 30 Eukaryotic genomes as well as a significant
number of Bacterial genomes. Additionally, we demonstrate the utility of
this resource [http://PhIGs.jgi-psf.org] for transferring gene annotation
from well annotated genomes to poorly annotated genomes, construction
of phylogenetic trees and comparative genomics in general.