

Poster C-18
reconstructing prokaryotic
phylogeny applying a novel
integrative approach



Authors:

naryttza n. diaz (*center for biotechnology (cebitec), universität bielefeld*)
lutz krause (*center for biotechnology (cebitec), universität bielefeld*)
alfred pühler (*lehrstuhl für genetik, universität bielefeld*)
kasten niehaus (*lehrstuhl für genetik, universität bielefeld*)
timm w. nattkemper (*technische fakultät, universität bielefeld*)

Short Abstract: we present a novel method for reconstructing phylogenetic trees, based on a novel integrative distance measurement of whole genomes and proteomes. an exhaustive analysis using 350 prokaryotic genomes not only show the validity of the method, but also suggest novel phylogenetic insights and its application to environmental sample analysis.

Long Abstract:

rapid sequence technologies have produced over 350 prokaryotic genomic sequences and almost 1000 are forthcoming. whole genomes sequences comprise a universe of profitable information that should be exploit in constructing phylogenies. in particular, new knowledge gained during the past decades on distinctive characteristics of genomic sequences should also be taken into account. however, these enlightenments are not yet reflected in the field of phylogenetics. In fact, the number of published works based on a single gene or protein greatly overcomes those based on whole genome information (e.g. gene order). using information contained in the entire genome will undoubtedly take into account the complexity of evolutionary processes affecting the complete genome which otherwise cannot be captured in single gene based phylogenies. therefore, it seems valuable to make use as much genetic information as available for the reconstruction of phylogenies.

we present a novel method for reconstructing phylogenetic trees, based on a new integrative distance measurement of whole genomes and proteomes. the core of our method is to make use of different genomic features (e.g. gc content, codon usage, oligonucleotide frequencies, gene content, encoded metabolic pathways, etc.).

an integrated distance matrix is generated and subsequently used to reconstruct the phylogenies with the neighbor-joining algorithm. species trees were then generated by 1) computing each feature on each genome, 2) building a feature vector, for each genome, 3) deriving an integrated distance values for each pair of organisms, and 4) using neighbor-joining method to generate trees from the resulting distance values.

in general, the results evince good agreement with traditional phylogenies and “taxonomic truth” (i.e. ssu-rna and Bergey’s Manual). in cases where misplacements were encountered for many there is documented support for the alternative placement. others represent interesting hypothesis well worth for further explorations.

additionally, a fast adjusted integrative measurement was implemented that only takes into account compositional features. eventhough gc content, up to now, has a dominant effect on

this measurement many bacterial groups can be recovered. within these trees, alpha proteobacteria, gamma proteobacteria, actinobacteria firmicutes, chlamydiae, aquificales, deinococcales, cyanobacteria, and archae were well grouped and reasonably arranged.

the adequacy of the novel integrative approach presented here will mature in time when enough experience is gained. we believed that the adjusted version of our distance measurement exploiting only the compositional genomic features should be feasible for rapid classification of environmental samples, as is being currently in evaluation.