
Origin and Evolution of Bacterial Genomes
Origine et évolution des génomes bactériens
(Org: **Paul Higgs** and/et **Ralph Pudritz** (McMaster))

DANIEL BROWN, University of Waterloo

Fast algorithms for phylogenetic reconstruction of aligned sequences

We present LSHtree and LSHplace, two fast methods for phylogenetics problems.

In LSHtree, we build phylogenies by starting with each taxon in its own tree, and merging trees (not necessarily at their roots, which is the practice in algorithms like Neighbour-Joining) until a single tree remains. The algorithm is sped up by using locality-sensitive hashing to find pairs of nearby sequences in the tree. We also approximately reconstruct the sequences at ancestral positions of the tree, to enable joins of trees where the proper edge is far from the leaves of the tree. The algorithm is extremely fast in theory, and can be shown to yield high-quality trees in Markov models of evolution. It also gives excellent trees in practice, and scales very well for large data sets.

We also present LSHplace, which adapts the idea of LSHtree to phylogenetic placement, giving a very fast method for putting new taxa onto an existing phylogeny. We discuss the implications of this work for applications such as metagenomics.

Both algorithms are joint work with PhD student Jakub Truszkowski.

ERIC COLLINS, University of Alaska

TAL DAGAN, Christian-Albrechts University Kiel

Phylogenomic networks reveal trends and barriers to lateral gene transfer during microbial evolution

Gene acquisition by lateral gene transfer (LGT) is an important mechanism for natural variation among prokaryotes. Laboratory experiments show that protein-coding genes can be laterally transferred extremely fast among microbial cells, inherited to most of their descendants, and adapt to a new regulatory regime within a short time. Recent advance in the phylogenetic analysis of microbial genomes using networks approach reveals a substantial impact of LGT during microbial genome evolution. Phylogenomic networks of LGT among prokaryotes reconstructed from completely sequenced genomes uncover barriers to LGT in multiple levels including (i) barriers to gene acquisition in nature including physical barriers for gene transfer between cells, (ii) genomic barriers for the integration of acquired DNA, and (iii) functional barriers for the acquisition of new genes.

RADHEY GUPTA, McMaster University

A Signature Protein Based In silico Microbial Identification Tool Using Next Generation Sequence Data

Rapid and reliable interrogation of the next generation sequence (NGS) data for the presence or absence of different organisms poses a significant challenge which limits the routine use of NGS techniques for clinical diagnostic and metagenomic investigations. A new approach/tool is described here based on Conserved Signature Proteins (CSPs), which are sets of proteins that are uniquely found in specific groups of organisms, for rapid interrogation of the NGS data for the presence or absence of different organisms. A large database of validated CSPs has been created that are specific for different prokaryotic and some eukaryotic organisms at multiple taxonomic levels ranging from phylum to species/strain levels. All significant blast hits for these CSPs are for the indicated group(s) of organisms. Due to the predicted presence of these CSPs in the indicated groups of organisms, Blast searches with their sequences (amino acid or nucleotide) provide a highly specific mean for rapidly and reliably determining the presence or absence of organisms from these groups in the metagenomic sequences. Using these CSPs, an in silico Web-based Microbial Identification Tool has been developed for rapidly determining the presence or absence of either specific organisms, or comprehensive taxonomic profiling of different organisms, in metagenomic sequences.

WEILONG HAO, Wayne State University

Estimating evolutionary rates of discrete characters, and its application on genome evolution

The study of non-DNA discrete characters is crucial for the understanding of evolutionary processes. Discrete characters often have different transition rate matrices, variable rates among sites and sometimes contain unobservable states. To obtain accurate estimation, we implement sophisticated maximum likelihood methodologies and flexible transition rate matrices capable of analyzing a variety of discrete characters. We then show application examples on gene family data and on intron presence/absence data.

PAUL HIGGS, McMaster University

Phylogenetic models of bacterial genome evolution incorporating gene insertion and deletion and horizontal gene transfer

The gene content of bacterial genomes differ significantly, even for closely related genomes. This illustrates that non-essential genes have high rates of insertion and deletion. Nevertheless, other genes can be found that have arisen only once in a phylogenetic tree and are signatures of monophyletic groups of genomes. There is thus a wide range of time scales involved in gene gain and loss. We analyse the presence-absence patterns of all genes in a specified group of related genomes using maximum likelihood methods. Each gene is assigned to one of three different scenarios. Scenario 0 genes are inferred to be present at the root and may have been deleted subsequently in some species. Scenario 1 genes are inferred to be absent at the root, have arisen only once within the tree, and may have been subsequently deleted. Scenario 2 genes have arisen more than once. Scenario 2 requires the occurrence of horizontal transfer, whereas scenario 1 can be explained either by origin of a new gene within the group studied or by horizontal transfer from outside the group. Preliminary results using Cyanobacteria and Archaea indicate that a majority of genes fall into scenarios 0 and 1, which means that their presence-absence pattern is consistent with the underlying genome tree. A significant number of scenario 2 genes are observed, but these do not obscure the strong tree-like signature in the evolution of the complete sets of genes.

YURI WOLF, NCBI/NLM/NIH

Turbulent genomes: quantification of gene acquisition, loss and displacement in prokaryotes

Genomes of bacteria and archaea exist in incessant flux, constantly acquiring and losing genes. We performed a comprehensive analysis of 36 groups of closely related microbial genomes to quantify relative contribution of different genome dynamic events. The results suggest an extremely high rate of both gene loss and acquisition, in large part, coupled through the phenomenon known as xenologous gene displacement whereby a recently acquired homologous gene rapidly replaces the vertically inherited copy. Acquisition of homologous genes, not intragenomic duplication, also dominates expansions of multi-gene families. The relative rates of acquisition and loss are not precisely balanced, indicating that the prevailing mode of evolution in bacteria and archaea is genome contraction, probably punctuated by bursts of gene gain.