A Global Gene Evolution Analysis on Vibrionaceae Family Using Phylogenetic Profile

Vitulo N (1),Vezzi A (1), Campanaro S (1), Romualdi C (1), Lauro F (2), Valle G (1)

(1) CRIBI biotechnology centre, Department of Biology, University of Padova
(2) SCRIPPS institution of oceanography, University of California, San Diego

Motivation

In the past ten years, thanks to the technology advance, a great number of microbial genomes have been sequenced and a huge amount of genes have been stored in the public databases. This gave the possibility to identify the gene core shared by all the organisms, the genes characteristic of particular group of microbes and laterally transferred ORFs (alien DNA). The phylogenetic profile is based on the observation that genes involved in the same metabolism or structural complex tend to be both present or absent within genomes (Pellegrini et al. 1999). This allowed the prediction of hypothetical proteins function on the basis of phylogenetic profile shared with known genes. Moreover this approach is useful to identify cluster of genes shared by bacteria that are not phylogenetically related, suggesting possible laterally transferred elements. The Vibrionaceae family represents a significant portion of the culturable eterotrophic bacteria of the sea; they strongly influence nutrient cycling and various species are also devastating pathogens. In this work we propose a phylogenetic profile analysis performed on Vibrionaceae sequenced genomes using a gene distance calculation method based on substitution matrix of all orthologous genes. We applied this approach to study the evolution of the Vibrionaceae family on the basis of the gene content, identifying genes that are specific to the Vibrionaceae family and genes that are laterally transferred.

Methods

Firstly we recovered all the predicted ORFs of the completed Vibrionaceae genomes, generating a redundant list. In order to reduce the redundancy we grouped all the ORFs according to COG annotation. The remaining proteins not related to any COG entry were clustered based on similarity search. We aligned all the ORFs against all the available bacterial proteomes, using BLAST. For each ORF we generated a phylogenetic profile encoded by an array of values representing the distance calculated as the number of amino acid substitutions between the gene and the orthologue, weighted using a substitution matrix. With this phylogenetic profile we produced a distance matrix of the clusters where each element is the median distance value of the ORFs belonging to each cluster. The matrix generated underwent a cluster and a pairwise correlation analysis that produced groups of ORFs sharing similar phylogenetic profiles (see figure attached).

Results

Results shown in the figure were obtained using Photobacterium profundum as a reference and calculating the distances from the orthologous ORFs from other bacteria in order to generate the distance matrix. Obviously the analysis was restricted to ORFs present in the reference organism, while we are now performing a study considering all the Vibrionaceae ORFs as described in the "method" section. The figure shows the pairwise correlation matrix calculated using the distance measured between phylogenetic profiles. Colors varying from red to green, blue and black represent increasing phylogenetic profile distance. With this analysis we could answer many biological question on the Vibrionaceae genomic features. First of all it is possible to suggest a function for genes annotated as unknown on the basis of their phylogenetic distance from well characterized ORFs. Another interesting aspect is the identification of laterally transferred genes which have an important role on the pathogenicity of this group of bacteria.

Contact email: nicolav@cribi.unipd.it