Network analysis of plasmids encoded proteins

Brilli M, Mengoni A, Fani R

Dipartimento di Biologia Animale e Genetica dell'Università di Firenze,
Via Romana 17/19, 50125 Firenze, Italia

Motivation

Plasmids are widespread in prokaryotes and harbour genes coding for a number of activities that can vary between different organisms groups. Plasmids are often transferred among microorganisms, and this permit the spread of new metabolic activities within natural bacterial communities. A clearcut example is the transfer of antibiotic resistances, often coded by plasmidborne genes. The evolution of plasmids has been studied using classical bioinformatics tools, such as phylogenesis and gene order/function comparisons. However, large scale analyses of plasmid encoded proteins have not been performed. Network analyses have been applied both to protein interaction and gene expression data, and we decided to use networks to represent and analyse the plasmid content and the homology relationships linking the proteins they code for, using dedicated software.

Methods

We downloaded the entire NCBI plasmid encoded protein dataset, and we divided it into subdatasets in conformity to different grouping rules (e.g. all proteins from Archaea, or those from a single genus or specie). Each subset was used to generate a database by using the formatdb utility of the stand-alone blast program (Altschul et al., 1997), and the very same dataset was then used for an all-against-all blast search. Dedicated java classes were written to: 1. search information concerning plasmid source on the web; 2. calculate coordinates of each node (protein) so that each plasmid is arranged circularly with the encoded proteins on its circonference and separated from other plasmids; 3. read the blast output to write down links among the different nodes, storing the percent identity value corresponding to each alignment (link). The java class writes a file in vsn format, readable by the network representation and analysis software Visone (www.visone.de). The user is prompted at start for E-value and alignment length cut-off values, permitting to explore networks at different detail levels.

Results

The network representation and analysis of the homologies among plasmid encoded proteins revealed differences when comparing the results obtained for organisms of different taxonomic groups. For example, the average number of links for different plasmid networks is generally lower for free-living organisms, in agreement with an at least partial genetic isolation as the analysis of Sulfolobus plasmids clearly showed, having a little average number of links for node, in agreement with the work of Whitaker et al. (2003). Moreover, the visual inspection of these networks can reveal the presence of large duplicated regions inside plasmids, and the rapid identification of fused proteins. Figure 1 is an example of the uniform network obtained for proteins encoded by Escherichia coli plasmids against a database containing Escherichia coli (squares), Salmonella (circles) and Shigella (diamonds) plasmids encoded proteins (E-value cut-off=0.0001; alignment length cut-off=200; alignment identity cut-off=70% identity); the function, if known, of proteins in the most important clusters is also indicated. In addition of proteins involved in plasmid replication and partition, transposases are the most represented proteins, followed by those involved in conjugative transfer and antibiotic resistanc. Resistance genes are shared by Escherichia coli and Salmonella strains, but not always by Shigella, and this corresponds to the presence of antibiotic resistance containing regions in the chromosome of the last. At this level of filtering, it appears that haemolysin-related genes are peculiar of Escherichia coli plasmids. Moreover, plasmid replication (Rep) proteins appear to be more conserved between Escherichia and Shigella, in agreement with taxonomic relationships.

Contact email: r_fani@dbag.unifi.it

Supplementary informations

After complete testing and refinement the java classes written for this work will be released free for academic users.