Poster H-1

Analysis of the Conservation of Regulation Systems in Firmicutes

Authors:
Nicolas Sierro (*The Institute of Medical Science, The University of Tokyo*)
Kenta Nakai (*The Institute of Medical Science, The University of Tokyo*)

Short Abstract: Recently, the number of sequenced bacterial genome has considerably increased, so that comparative analysis of bacterial promoters can provide new insight in the evolution of regulatory networks. Investigation of firmicute promoters was therefore carried out using the available sequenced genomes and the database of transcriptional regulation in Bacillus subtilis (DBTBS).

Long Abstract:

A significant increase in the number of published complete bacterial genome sequence occurred in the last years, and new genomes are regularly released. With this rapid increase in the number of bacterial genome entirely sequenced, systematic function analysis projects have started to decipher the total gene activity of these organisms. Due to the probable co-regulation by a common transcription factor of genes showing a similar expression profile, investigation of their promoter regions is an important step towards the understanding of global cell regulation networks. By coupling the current knowledge about experimentally proven transcriptional regulation with the currently available raw genetic data, a better understanding of the similarities and differences between various species could be obtained. Most of the bacterial transcriptional regulation data have however been obtained in two specific organisms, Escherichia coli and Bacillus subtilis, and comparative genomics is therefore necessary to evaluate to which extent the acquired knowledge is applicable to other bacterial species. The previously constructed database of transcriptional regulation in Bacillus subtilis (DBTBS) focuses on known transcription factors, their recognition sequences and the genes they control in B. subtilis. However, in the recent years, more than 60 other firmicute genomes, including medically and industrially important species such as Streptococcus pyogenes, Staphylococcus aureus or Lactococcus lactis, have been completely sequenced. Starting from the DBTBS database, which contains transcriptional regulation data specific to B. subtilis, and using evidences obtained experimentally in related strains and published in peer-reviewed journals, will likely lead to the identification of subgroups of bacteria in which the commonly accepted gene regulation networks are not present, and a new insight in the evolution of regulatory networks could be obtained.

In order to extend the DBTBS database, work published in peer-reviewed journals concerning transcription regulation in bacterial species phylogenetically closely related to B. subtilis was collected. Unfortunately, a large majority of these publications report protein-DNA interactions based only on sequence similarities with homologous proteins and promoters of B. subtilis, for which an interaction has been experimentally proven. In addition, only few of the experimentally proven interactions were obtained in bacterial strains for which the complete genome sequence is available. To overcome this limitation, the annotated proteins of 66 complete firmicutes genomes, including that of B. subtilis, were compared to each other in order to build clusters of homologous proteins and their upstream intergenic

regions. Two approaches were then used to analyze these upstream intergenic regions: the mapping of known B. subtilis transcription factor binding sites on every genome, and the analysis of the upstream intergenic region conservation for each cluster of homologous genes.

DBTBS provides position specific weight matrices for more than 30 transcriptional regulators based on the sequences of experimentally proven binding sites. These matrices were therefore used to investigate the presence or absence of a specific motif in the intergenic regions upstream of the genes encoding homologous proteins. To provide a comprehensive yet easy to understand and interpret representation of the motif conservation pattern, new tools were developed that can generate a graphic for each cluster indicating on one side in which stains homologous proteins are found, and on the other side whether or not these proteins possess a certain transcription factor binding site in their upstream intergenic region.

With this method, the existence of different regulation systems for the CtsR and HrcA heat shock response regulons within the firmicutes could for instance be shown. Although these two regulons have been extensively studied in B. subtilis only little experimental data have been obtained regarding their regulatory mechanism in other firmicutes. Our data suggest for example that the mollicutes, which are characterised by a very small genome size and lack CtsR, have placed the regulation of genes typically regulated by CtsR under the control of HrcA, or that in the Staphylococcus strains the HrcA regulons is not only regulated by HrcA itself, but also by CtsR. This example clearly illustrates the importance of extending the comprehensive B. subtilis information available with that obtained from other related bacteria by comparative analysis in order to provide a more accurate picture of bacterial gene regulation.

The analysis of the upstream intergenic region conservation for each cluster of homologous genes is also of interest because it is expected that regions involved in gene regulation are more conserved than regions with no particular function. To investigate this conservation, each cluster was therefore divided in subclusters based both on the bacterial genus and the size of the intergenic regions and aligned by ClustalW. The aligned sequences were used to calculate the sequence conservation and the stretches of sequences yielding a conservation of 80 % or more were collected to generate position specific weight matrices that will be compared both across subclusters of a same cluster and across clusters to identify novel binding sites and regulatory structures. Preliminary results indeed showed that these conserved regions do not always match known transcription factor binding sites, possibly indicating the presence of new cis regulatory elements.

By carrying out comparative analysis of a large number of related genomes, concentrating particularly on the conservation of the promoter regions of homologous genes, and using the experimental data available in literature, significant differences in the regulatory networks of not only a single strain, but of whole genus could be highlighted. This approach will therefore not only allow a refinement of the current understanding of bacterial regulation networks, but also provide new input for experimental research by helping in the design of experiments and the interpretation of their results.