Expression evolution of prokaryotic organisms

Pieter Meysman, Aminael Sanchez-Rodriguez, Qiang Fu, Kathleen Marchal and Kristof Engelen

Introduction

Regulation of gene expression is one of the primary adaption techniques used by prokaryotic species to survive in a changing environment. It is thus no surprise that not only does the genetic content of prokaryotic organisms vary greatly between species, but also the timing and manner of the regulation of this content. It has been shown that massive rewiring can occur in the regulatory network between relatively closely related species. In this work, we tried to assess how much the gene expression varies between a model organism, namely *Escherichia coli*, a closely related species, namely *Salmonella enterica* serovar Typhimurium and a more distant species, namely *Bacillus subtilis*.

Methods and materials

The gene expression data was collected from the second release of the COLOMBO5 database. This data consisted of microarray contrasts for 1570 conditions for *E. coli*, 999 conditions for *S. enterica* and 340 conditions for *B. subtilis*. To compare changes in gene expression, we compared these three microarray compendia based on the expression convergence measure as defined by Tirosh and Barkia. This method does not directly compare the expression values between orthologs as they were measured for different species and under different conditions, but instead defines the expression divergence as the variation between the expression correlation of orthologous genes. Ortholog mapping was accomplished using OrthoMCL on the protein sequences for the three organisms.

Results

Despite the conservation of a large amount of genetic content between *E. coli* and *S. enterica*, we found that a substantial number of genes had different expression behaviors. These genes were mostly enriched for processes involved in the first steps of compound metabolism, such as transport proteins and the carbohydrate catabolism. In one specific case, the genes involved in the resistance against the antibiotic polymyxin were found to be highly divergently expressed between *E. coli* and *S. enterica*, which corresponds well to prior observations of the regulatory rewiring of these genes between the two species. In contrast, the genes with conserved expression seemed to mostly involve key cellular processes, such as ribosomal assembly or ATP synthesis. This category was therefore also strongly enriched for genes which had been found to be essential in both species. This observation was only amplified in the comparison between *E. coli* and *B. subtilis*, with a larger fraction of genes showing very divergent expression behavior and only the
most critical genes showing conserved expression. Furthermore in both cases, we found a significant correlation between the conservation of the protein sequence and the conservation of the gene expression. Also for in-paralogs, where a gene duplication occurred after the speciation event, we found that in most cases the gene with the most conserved protein sequence was also the gene with the best corresponding gene expression pattern.

Conclusions

The variation in the gene expression behavior corresponded well with the general genomic variation: more distant organisms had a higher rate of gene expression divergence and genes with lower sequence conservation also had a more divergent gene expression. This could be the result of similar evolutionary pressure on both the expression and the sequence of a gene.

References

