Comparative analysis of transcriptional regulation in *E.coli* and relative genomes: FNR, ArcA, NarP, NadR, and ModE regulons.

Anna Gerasimova[1], Mikhail Gelfand[2]

[1] Laboratory of Bioinformatics, State Scientific Center GOSNIIGenetika
1-iy Dorozhny proezd 1, Moscow, 113545, Russia,
a_gerasimova@yahoo.com
[2] Institute for Information Transmission Problems, RAS
Bolshoy Karetny per. 19, Moscow, GSP-4, 127994, Russia,
gelfand@iitp.ru

Background: Comparative genomics is a powerful approach to the study of transcriptional regulation in bacteria. Availability of many sequenced genomes of gamma-proteobacteria allow for detailed analysis of multiple interacting regulatory systems and taxon-specific changes in regulatory patterns.
Respiration in gamma-proteobacteria is controlled by several transcription factors forming complex regulators cascades. This, together with the importance of the aerobic/anaerobic switch for the physiology of bacteria makes, it an ideal object for the comparative genomics analysis.
Methods: In this study we analyzed the FNR, ArcA, NarP and ModE regulation in organisms from three bacterial families: Enterobacteriaceae (*Yersinia pestis, Y. entercolitica*), Pasteurellaceae (*Pasteurella multocida, Actinobacillus actinomycetemcomitans, Haemophilus influenzae, H. ducreyi*), and Vibrionaceae (*Vibrio cholerae, V. parahaemolyticus, V. vulnificus, V. fischeri*). We also analyzed the NadR regulon in *Enreobacreriaceae* (*E. coli, Shigella flexneri, Salmonella typhi, Klebsiella pneumoniae, Erwinia carotovora, Serratia marcescens, Y. pestis, Y. enterocolitica* and *Photorhabdus luminescens*). To describe these interacting regulatory systems we applied the comparative genomics in particular, phylogenetic footprinting, consistency altering of candidate sites, positional clustering of genes.
Results: We identified changes in regulatory cascade, compared to *E. coli*. In *H. influenzae* the *fnr* gene has candidate FNR, ArcA and NarP sites, *arcA* has NarP and FNR sites, and *narP* has FNR ArcA and NarP sites.
In all studied genomes, the *nuo* operon encoding the proton translocating NADH-dehydrogenase has no predicted sites (*Enterobacteriaceae*) or is absent (*Pasteurellaceae* and *Vibrionaceae*). However, all these genomes contain the *nqr* operon encoding the sodium translocating NADH-dehydrogenase. In all genomes, this operon is preceded by candidate sites of at least one respiration-regulating transcriptional factor. Thus, in this case we observed non-homologous gene displacenemt with partially conserved function.
Analysis of the NadR system demonstrated that even a very simple regulon covering an essential methabolic pathway could be different in closely related genomes. In particular, *nadR* seems to be autoregulated in *E. carotovora, S. marcescens, Y. pestis* and *Y. enterocolitica* but not in *E. coli*.
In *E.coli* we have identified new candidate regulon members. In particular, we observed conserved binding sites for FNR and ArcA upstream of the *b2503* gene encoding putative cytochrome C-type biogenesis protein. The gene *b1674* encoding putative oxidoreductase, Fe-S subunit, has strong ModE, FNR and NarP candidate binding sites, and *moaA* has a candidate NarP site.
Conclusion: Using comparative genomic analysis, we described five regulons in *Enterobacteriaceae*, identified new candidate regulon members, and demonstrated that both simple and complex regulatory cascades are subject to evolutionary changes even in rather closely related genomes.