On the origin and evolution of biosynthetic pathways: integrating microarray data with gene structure and organization

Fondi M, Brilli M, Fani R

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

Motivation

The availability of the nucleotide sequence of complete genomes from an increasing number of organisms belonging to the three cell domains, Archaea, Bacteria and Eucarya is providing an enormous body of data concerning the structure and the organization of genes and genomes. As a result it is now possible to shed some light on the mechanisms involved in their evolution and responsible for the shaping of metabolic pathways. The emergence of basic biosynthetic pathways represent one of the major and crucial events during the early evolution of life. Their appearance and refinement allowed primitive organisms to become increasingly less dependent on exogenous sources of amino acids, purines, and other compounds. Among the theories proposed to explain how metabolic pathways have assembled and evolved, the patchwork hypothesis (Jensen 1976) is the most accepted one. Schematically it predicts that the extant pathways have been assembled starting from a restricted core of ancient genes that underwent gene duplication events (generating paralogous genes) followed by evolutionary divergence. By this “two-step” mechanism, ancient genomes may have increased their dimensions and/or gained novel metabolic abilities. In some cases, the divergence between paralogous genes (or set of genes) might be due to divergence in the regulation mechanisms controlling their expression and\or the regulation of enzymatic activity rather than mutations affecting catalytic sites or sites involved in the bindingof specific ligands. This is the case of genes involved in the so-called Common Pathway (CP) of lysine, threonine and methionine. These three biosynthetic routes share the first two steps, i.e. the phosphorylation of an aspartate molecule and the subsequent oxidation that lead to the formation of aspartyl phosphate. In the γ-proteobacterium Escherichia coli three different aspartokinases (AKI, AKII, AKIII, the products of thrA, metL and lysC, respectively) can perform the first step of the CP. Moreover, two of them (metL and thrA) are bifunctional, carrying also homoserine dehydrogenasic (hsd) activity, the final branching point of threonine and methionine routes. The second step of the CP is catalyzed by a single aspartate semialdeyde dehydrogenase (ASDH, the product of asd). Thus, in the CP of E. coli while a single copy of ASDH perform the same reaction for three different metabolic routes, three different AKs have evolved and been maintained to perfom a unique step. Why such a situation emerged and maintained? How is it correlated to the different regulatory mechanisms acting on these genes? The integration of data concerning gene structure, organization, and phylogenetic distribution, with information coming from the analysis of microarray experimental data represents a very powerful tool to elucidate the mechanisms and forces driving the assembly and the evolution of metabolic pathways. For this reason, we used such an approach to answer those questions mentioned above, focusing the attention on proteobacteria whose genome was fully sequenced.

Methods

Microarray data were downloaded as supplemental material to published papers. Database search was performed using BLAST software (Altschul et al. 1997). Phylogenetic trees were constructed using MEGA3 software (Kumar et al.2004). Microarray experiments data statistic analysis were conducted using R software (ver 2.1.1 R Development Core Team 2005).

Results

Structure and phylogenetic distribution of AK and HD coding genes in proteobacteria

Data obtained revealed that the presence of multiple copies of the AK coding gene and their fusion with HD domains are restricted to the Y-subdivision of proteobacteria. A model explaining the origin and evolution of AK and HD coding genes was depicted, which was also supported by a phylogenetic analysis of both AK and HD sequences. According to this model, the genome of the

proteobacterial ancestor harboured a monofunctional copy of both ask and hsd genes. Then a paralogous duplication of these genes concomitant to the fusion of the copies occurred within the γ-proteobacteria giving raise to a bifunctional gene that, in turn, duplicated generating the ancestors of the extant metL and thrA.

Organization of AK and HD coding genes in proteobacteria

The analysis of the organization of lysine, threonine, and methionine biosynthetic genes revealed that the appearance of fused genes was paralleled to the assembly of operons of different sizes, suggesting a strong correlation between the structure and organization of these genes.

Analysis on microarray data

In order to check the existence of a correlation between the structure, the organization and the coexpession of genes belonging to the same metabolic route, a statistic analysis of microarray data retrieved from experiments conducted on E.coli and Pseudomonas aeruginosa, showing different gene structure and organization, was carried out and data obtained will be discussed.

Contact email: r_fani@dbag.unifi.it

Supplementary informations

Microarray data references are available and not cited for shortness (10 publications, 79 conditions)