International Journal of Applied Research in Bioinformatics
Volume 9 • Issue 1 • January-June 2019

Looking Into the Binary Interactome of *Enterobacteriaceae* Family of Bacteria

Saritha Namboodiri, University of Calicut, Malappuram, India
Alessandro Giuliani, Istituto Superiore di Sanità, Roma, Italy

ABSTRACT

Protein-protein interactions (PPIs) regulate most of the biological activities within a cell. A set of pairwise PPIs in seven genera of bacterial pathogens (Salmonella, Escherichia, Shigella, Yersinia, Klebsiella, Photorhabdus, and Pantoea) of Enterobacteriaceae family was analysed. At genotypic level, the correlation coefficient analysis of the mutation spectra of the ten sets of directly interacting protein partners in Escherichia coli recognised all the 'interacting partners' in Escherichia coli. Extending the correlation analysis to include strains from the rest of the bacterial genera decreased the recognition efficiency providing quantitative evidence that binary interactome have incomplete superposition across species. At phenotype level, a reliable classification of bacterial pathogens was obtained by measuring PPI variations in terms of between phylogenetic distance correlation distances among ten sets of proteins partners. This forces us to rethink upon the possibility of PPI rewiring with a consequent change in physiological role of the same protein.

KEYWORDS

Bacterial Classification, Enterobacteriaceae, Genotype-Phenotype Relations, Kernel K-Means Clustering, Phylogeny, PPI

INTRODUCTION

Proteins are tools that carry out most of the biological process in a cell by their very specific interaction with other proteins. Enzymes involved in the same metabolic pathway mutually interact and aggregate to build molecular machines. This specificity in interactions is essential to maintain a viable metabolism. Likewise, signal propagation within tissue micro-environment depends on highly organized cytoskeleton protein in meshes. The presence of an organized 'interactome' is thus a key prerequisite for all cellular processes. Consequently, determining the Protein Protein Interaction (PPI) network, discovering associations among proteins involved in metabolic pathways and determining pathway kinetics are of prime importance in understanding how a cell works.

In this work, we examined the pairwise protein interactions within species strains of *Escherichia col*i and between species strains from *Salmonella, Yersinia, Klebsiella, Photorhabdus, Shigella and Pantoea* genera) of *Enterobacteriaceae* family, based on the correlation between the phylogenetic trees of the interacting protein pairs, to elucidate the evolutionary relationships in terms of variations in the PPI patterns. We hypothesis that PPI's are strictly preserved within species whilst their degree of superposition inversely scale with phylogenetic distance which implies that PPI's is not conserved across species and that orthologs of proteins have different interacting patterns across species and consequently play a different physiological role.

DOI: 10.4018/IJARB.2019010104

To prove our hypothesis, we took ten set of binary interacting proteins from different strains of *Escherichia coli* and its orthologs from species of *Salmonella, Yersinia, Klebsiella, Photorhabdus, Shigella and Pantoea* genera of *Enterobacteriaceae* family. Global molecular phylogenetic protein profile was constructed based on the distance matrix obtained from Multiple Sequence Alignment of the entire protein set. Scoring of a strong correlation between distance matrices for an interacting protein pair is an indicator that the two proteins interact, given they undergo co-evolution (Goh, Bogan, Joachimiak, Walther & Cohen, 2000). High values of Pearson's Correlation coefficient of Global molecular phylogenetic protein profile of *Escherichia coli* strains were observed indicating strong binary interaction among the ten protein pairs. However the Global molecular phylogenetic protein profile between species strains of *Salmonella, Yersinia, Klebsiella, Photorhabdus, Shigella and Pantoea* genera of *Enterobacteriaceae* family only allows for a partial recognition of binary interactions which led us to conclude that binary protein interactions are preserved within the Escherichia coli species whereas it is not seen conserved across *species of Enterobacteriaceae* family. Based on the variation in the correlation between the phylogenetic trees of the interacting protein pairs, we could obtain a reliable classification of the seventy-five strains of bacteria from *Enterobacteriaceae* family that is largely in concordant with the currently accepted phylogenetic theory.

BACKGROUND

At present, protein–protein interaction studies focus on the combinatorial complexity of functional interactome with an intention to build a viable cell from them. The bewildering combinatorial complexity of possible interactomes raises many questions on development as well as rewiring of PPIs along evolution. Peter Tompa and George Rose sketched a minimalistic combinatorial model ($7.9\times10^{7.9\times10^{9}}$ possible protein-protein interaction patterns) for yeast and concluded that a prior framework of existing interactions is essential, as emergent interactome cannot self-organise spontaneously (Tompa & Rose, 2011). In evolutionary terms, this implies that a particular PPI pattern (an interactome) is highly conserved physical entity and small variations are important phenotypic markers of species separation (Tompa & Rose, 2011; Kyaw, Pawan, Maria & Alessandro. 2006; Goh, Bogan, Joachimiak, Walther & Cohen, 2000). The analysis of eventual variation in the PPI patterns from the orchestrated protein interactions is thus of crucial importance in taxonomy.

The variation of PPI pattern was also analysed by Reuveni E. et al., in terms of species specificity of synonymous and non-synonymous mutation in yeast genome (Reuveni & Giuliani 2012). Mutation in single gene occurs in two manners: synonymous mutation, wherein no change occurs in primary protein sequence; non-synonymous mutation that produces change in the primary sequence of the protein thereby potentially changing its functionality with consequent phenotype variation. The authors observed that, in proteins, that play a crucial role in the cellular process, the ratio of synonymous and non-synonymous mutation (dN/dS) ratio computed on a gene-by gene basis were conserved across proteomes of inter-breeding populations. Strong correlation between two genomes in terms of their dN/dS ratio profiles implies that vital proteins and their interaction are preserved across genomes. However, on changing the focus from a single gene to a genomic scale, it was observed that the dN/dS ratio varied across species. This lack of correlation implies that the same protein changes its physiological role in different species. This led them to propose that proteins interact differently in different species and form different PPI patterns, given that the proteins (with only rare exceptions) do not work in isolation.

Kirill Evlampiev et al. demonstrated that network motifs, unlike individual proteins, cannot be indefinitely conserved under general duplication-divergence evolution, regardless of any network-rewiring dynamics (Kirill & Hervé, 2008). In this light, we looked into the PPI's of *Enterobacteriaceae* family of bacteria. We hypothesis that the binary protein interactions are conserved within the *Escherichia coli* species whereas it is not conserved across the species of *Enterobacteriaceae* family.