Diversity of Transcription Factors in Escherichia coli

The fitness of an organism is defined in terms of its environment wherein it must process information thereof optimally to survive and propagate. Central to this goal is control of gene expression which, in bacteria, is largely achieved by transcriptional regulation. Bacteria inhabiting diverse niches are therefore, expected to have diverse regulation. Indeed, distantly-related species of Bacteria have distinct families of transcription factors even when they do not differ significantly in their metabolic capabilities. Similarly, one expects to observe a higher diversity in transcription factors (TFs) than their target genes within a bacterial population that occupy diverse niches. This study aims to formally test the above hypothesis and is intended toward understanding if the evolutionary mechanisms operating at the population-level can account for the large scale bacterial diversity. *Escherichia coli* serve as an appropriate model for such a study due to the fluctuating nature of their environment, different lifestyles -commensal or pathogenic and their ability to survive outside their mammalian hosts. Years of genome sequencing efforts targeted to *E coli* have made it the best sampled bacterial population, which is crucial for a study aimed at understanding population diversity. We make use of these publicly available WGS datasets of *E coli* strains and perform variant calling to test our hypothesis. SNP matrices generated using the genome of *E coli* K12 MG1655 as the reference were used to test if nucleotide diversity of TFs is greater than all other genes, all regulated genes or only to their target genes. From the datasets tested so far, TFs do not exhibit a higher diversity, in general. However, some TFs appear to have higher diversity than their target genes and this diversity show a negative relationship with the number of regulated transcriptional units (TUs). CRP and Fis are two global regulators with higher diversity than their target genes, even though this is against the general trend. Going further, we aim to identify the regions in TFs where most of the diversity is localised and find the extent to which this diversity of TFs has been shaped by natural selection.