MAR17-2016-020323

Abstract for an Invited Paper
for the MAR17 Meeting of
the American Physical Society

Theory of microbial genome evolution

EUGENE KOONIN, NCBI, NIH

Bacteria and archaea have small genomes tightly packed with protein-coding genes. This compactness is commonly perceived as evidence of adaptive genome streamlining caused by strong purifying selection in large microbial populations. In such populations, even the small cost incurred by nonfunctional DNA because of extra energy and time expenditure is thought to be sufficient for this extra genetic material to be eliminated by selection. However, contrary to the predictions of this model, there exists a consistent, positive correlation between the strength of selection at the protein sequence level, measured as the ratio of nonsynonymous to synonymous substitution rates, and microbial genome size. By fitting the genome size distributions in multiple groups of prokaryotes to predictions of mathematical models of population evolution, we show that only models in which acquisition of additional genes is, on average, slightly beneficial yield a good fit to genomic data. Thus, the number of genes in prokaryotic genomes seems to reflect the equilibrium between the benefit of additional genes that diminishes as the genome grows and deletion bias. New genes acquired by microbial genomes, on average, appear to be adaptive. Evolution of bacterial and archaeal genomes involves extensive horizontal gene transfer and gene loss. Many microbes have open pangenomes, where each newly sequenced genome contains more than 10% 'ORFans', genes without detectable homologues in other species. A simple, steady-state evolutionary model reveals two sharply distinct classes of microbial genes, one of which (ORFans) is characterized by effectively instantaneous gene replacement, whereas the other consists of genes with finite, distributed replacement rates. These findings imply a conservative estimate of at least a billion distinct genes in the prokaryotic genomic universe.