Title
Selection for efficiency at the transcriptional and translational levels are correlated across bacteria

Permalink
https://escholarship.org/uc/item/1q85c613

Authors
Froula, Jeff
Francino, M. Pilar

Publication Date
2006-04-01
Selection for efficiency at the transcriptional and translational levels are correlated across bacteria

Jeff Feinla, M. Pilar Francino
Department of Evolutionary Genomics, Joint Genome Institute, Walnut Creek, CA.

Background

It has been shown that selection acts to remove the promoter’s –10 and –35 consensus sequences in both coding and noncoding regions, implying that it is disadvantageous to maintain misplaced sites that can strongly bind σ70 and interfere with proper gene expression. Here we analyze 56 bacterial genomes and show that the numbers of non-consensus, potential σ70-binding sites in both regulatory and non-regularatory noncoding regions deviate significantly from the random expectations when compensating for base composition, di- and tri-nucleotidic bias in a majority of cultivated bacteria.

Not only do we expect that selection is maintaining high densities of potential σ70 binding sites in regulatory DNA, but that there is selection against these sites in non-regulatory DNA. The often overlapping binding sites in regulatory DNA likely confer some subtle survival advantage, even though experimental evidence suggests only one or a few of these sites are actual transcription initiation sites. Remarkably, we find that the degree of selection against potential σ70 binding sites in non-regulatory DNA correlates positively with rate of growth, adaptive codon bias and number of tRNA genes. This evidence that the efficiency needed for faster growing bacteria can only be achieved by reducing spurious RNA polymerase binding to false sites, and that transcription and translation efficiencies are both optimized at a genome-wide level to permit faster growth.

Getting list of potential sigma-70 binding sites

The three steps in identifying different density patterns of potential σ70-binding sites were to obtain a list of short sequences representing the -10 box from 522 experimentally proven E. coli promoters.

I) From 522 experimentally proven promoters from E. coli

II) Parse the genome into three biologically relevant steps:
   a) coding
   b) noncoding & regulatory
   c) noncoding & nonregulatory

III) Count the number of hexamers that occur in a genomic partition. This count is obtained from shuffled sequence to get the expected value and from the unshuffled sequence to get the observed value. To get statistical significance, the observed value is compared to a histogram from 1000 expected values from 1000 shuffled sequences.

Bacteria Across Multiple Phyla Have Similar Patterns

Analysis of over/under-representation of sigma-70 binding sites on 56 bacterial species representing multiple phyla

Results:
A) 63% of the species had under-representation of sigma-70 sites in non-regulatory DNA.
B) 63% of the species had over-representation of sigma-70 sites in regulatory DNA.
C) There was no significant bias in the coding sequence when all species were taken together.

Selection Against Sigma-70 Sites is Correlated with Optimized Growth

It is widely known that growth rate is strongly correlated to the number of tRNA genes in a genome. We can therefore correlate the number of tRNA genes with the amount of selection against spurious sigma-70 binding sites. We suspect such selection is important for optimizing transcription, an important requirement in fast growing bacteria.

Selection against spurious sigma-70 binding sites was measured as how far the observed value was from the expected mean. Thus: multiples of stdev = (obs - mean)/stdev where stdev (standard deviations), obs (observed value), and mean (expected mean).

Nonfunctional DNA

Histogram of tRNA gene counts

r = -0.417 *  
$P < 0.002$