Text S1 1: the justification for using the tAI and the RFM as an predictor of the co-adaptation between codon bias and tRNA pool

The tAI and the RFM are based on the genomic tRNA copy number (tGCN; when the expression levels of the tRNAs is unknown) as a surrogate measure for the cellular abundances of tRNAs; it is justified by several observations. First, in the past, in many organisms, it has been observed that the in vivo concentration of a tRNA bearing a certain anticodon is highly proportional to the number of gene copies coding for this tRNA type. Specifically, in S. cerevisiae a correlation of r=0.91 [1] was reported. In B. subtilis, a correlation of 0.86 between tRNA copy number and tRNA abundance was reported [2]. Similarly, previous papers reported about significant correlation between genomic tRNA copy number and tRNA abundance in E. coli [3,4]. A related interesting result is the analysis of [5] who measured the translation rate of two glutamate codons: GAA and GAG. They found them to have a threefold difference in translation rate (21.6 and 6.4 codons per second, respectively). Remarkably, the $w_i$ of these codons, which is based on the tRNA pool and affinity of codon-anti-codon coupling and is the basis for the tAI calculation, captures the ratio of translation rate between the two codons. Calculating $w_i$ values for E. coli we found that the ratio between the $w_i$ of GAA and GAG is 3.125 (0.5/0.16) as compared to the 3.34 reported in the experiments (21.4/6.4). This result suggests that there is a direct relation between the adaptation of a codon to the tRNA pool, based on the genomic tRNA copy number, and the time it takes to translate it.

Second, a recent study showed that in S. cerevisiae the promotors of many of the tRNA genes have a low predicted affinity to the nucleosome, suggesting a constitutive expression with little transcriptional regulation capacity [6]. Thus, for fully sequenced genomes, the relative concentrations of the various tRNAs in the cell, and therefore the optimality of the various codons in terms of translation, can be approximated using the respective tRNA gene copy numbers in the genome. Additionally, as we show in this paper, measures that are based on tRNA copy number highly correlated ro protein expression levels (see also [7,8]). It was found that even among genes with similar transcript levels, higher tAI often corresponds to higher protein abundance [7].