Please cite this article in press as: Tuller et al., An Evolutionarily Conserved Mechanism for Controlling the Efficiency of Protein Translation, Cell (2010), doi:10.1016/j.cell.2010.03.031

Cell

An Evolutionarily Conserved Mechanism for Controlling the Efficiency of Protein Translation

Tamir Tuller,[1,2,5] Asaf Carmi,[1,5] Kalin Vestsigian,[3] Sivan Navon,[1] Yuval Dorfan,[1] John Zaborske,[4] Tao Pan,[4] Orna Dahan,[1] Itay Furman,[1] and Yitzhak Pilpel[1]
[1]Department of Molecular Genetics
[2]Faculty of Mathematics and Computer Science
Weizmann Institute of Science, Rehovot 76100, Israel
[3]Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA
[4]Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, IL 60637, USA
[5]These authors contributed equally to this work

SUMMARY

Recent years have seen intensive progress in measuring protein translation. However, the contributions of coding sequences to the efficiency of the process remain unclear. Here, we identify a universally conserved profile of translation efficiency along mRNAs computed based on adaptation between coding sequences and the tRNA pool. In this profile, the first ~30–50 codons are, on average, translated with a low efficiency. Additionally, in eukaryotes, the last ~50 codons show the highest efficiency over the full coding sequence. The profile accurately predicts position-dependent ribosomal density along yeast genes. These data suggest that translation speed and, as a consequence, ribosomal density are encoded by coding sequences and the tRNA pool. We suggest that the slow "ramp" at the beginning of mRNAs serves as a late stage of translation initiation, forming an optimal and robust means to reduce ribosomal traffic jams, thus minimizing the cost of protein expression.

INTRODUCTION

mRNA translation is controlled at multiple stages and by diverse mechanisms. A major part of the control occurs at the stage of initiation, where ribosomes are recruited to and assembled on the mRNA, typically on the 5′ untranslated region (UTR) (Ingolia et al., 2009). The elongation phase is governed by both the mRNA secondary structure (Gray and Hentze, 1994) and the extent of adaptation of the coding sequence to the cellular tRNA pool (dos Reis et al., 2004; Sharp and Li, 1987). The abundance of tRNAs that correspond to the different codons in a gene was suggested to determine the speed (Akashi, 2003; Man and Pilpel, 2007) and accuracy (Drummond and Wilke, 2008) of translation. Thus, codons that are recognized by abundant or rare tRNAs will be respectively referred to here as codons with high and low efficiency (or as codons that are respectively highly or lowly adapted to the tRNA pool).

Transcripts whose codons are biased toward the more abundant tRNAs were found to be more highly expressed (Man and Pilpel, 2007; Qin et al., 2004). Indeed protein expression levels can be artificially increased by designed mutations that increase their codon-tRNA adaptation (Arava et al., 2003; DeRisi et al., 1997; Percudani et al., 1997; Tuller et al., 2007), pointing to a causal relationship between codon usage and expression level. Accordingly, the extent of adaptation between genes and the tRNA pool in different species was found to vary in evolution according to organisms' lifestyle needs (Man and Pilpel, 2007).

So far, studies that gauge translation efficiency have mostly considered average codon usage over entire genes (dos Reis et al., 2004; Man and Pilpel, 2007; Sharp and Li, 1987; Tuller et al., 2007). Such studies typically do not consider the order in which codons with low and high translation efficiency appear along the transcript. Although it was shown previously that the extent of codon bias changes along transcripts (Qin et al., 2004), it is not known whether gene sequences are arranged so as to determine specific levels of speed and accuracy of translation at various positions along transcripts. The order of high-efficiency and low-efficiency codons along transcripts could govern the process of translation, especially given that multiple ribosomes are often simultaneously loaded on a given transcript (Arava et al., 2003). Such instructions could affect speed and processivity of translation as well as the overall cost of protein production in cells. Such "traffic rules" of translation may thus be selected for during evolution.

To investigate a role for codon selection in modulating translation efficiency, we inspected open reading frame sequences and tRNA repertoires in dozens of fully sequenced genomes. The input to our computed profile of translation efficiency consists of coding sequences and a measure of adaptation between codons and the various tRNAs, which was previously shown to represent translation efficiency (dos Reis et al., 2004; Man and

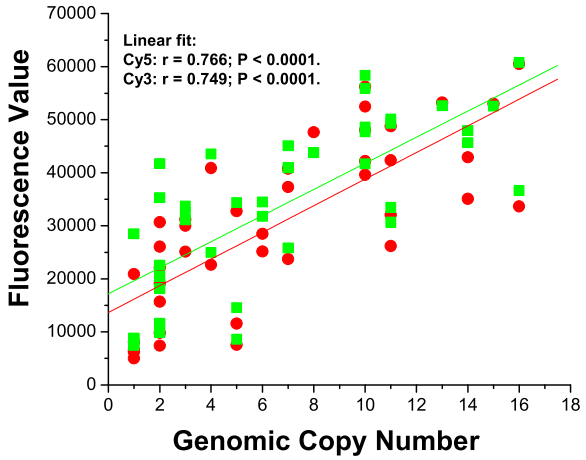


Figure 1. The tRNA Gene Copy Correlates with Levels of tRNA Genes in *S. cerevisiae*

tRNA gene copy numbers versus the expression levels of tRNA genes in *S. cerevisiae* measured by a microarray dedicated to the tRNAs in this species (Dittmar et al., 2004). tRNA levels were measured independently with two alternative dyes (Cy5 and Cy3), each producing similar correlations with the gene copy numbers.

See also Table S1 and Figure S1.

Pilpel, 2007; Tuller et al., 2007). We describe a universally conserved translation efficiency profile that features low translation efficiency over the first ~30–50 codons of mRNAs. This feature is conserved in species that represent the three domains of life. Translation efficiency profiles of individual genes are often very noisy, yet selections of subsets of the genes in each genome display this pattern. The observed profile in the yeast *S. cerevisiae* accurately predicts the recently measured ribosome density profiles of mRNAs in this species (Ingolia et al., 2009). This correlation provides an indication that codon-tRNA adaptation approximates well the speed of translation at various positions along mRNAs. This result also indicates that ribosomal speed and hence density are encoded into genes' sequences and the tRNA pool. We propose that the conserved translation efficiency profile may have been selected for in diverse species as it minimizes ribosome traffic jams and abortive protein synthesis and, as a consequence, the cost of protein expression.

RESULTS

A Universally Conserved Translation Efficiency Profile

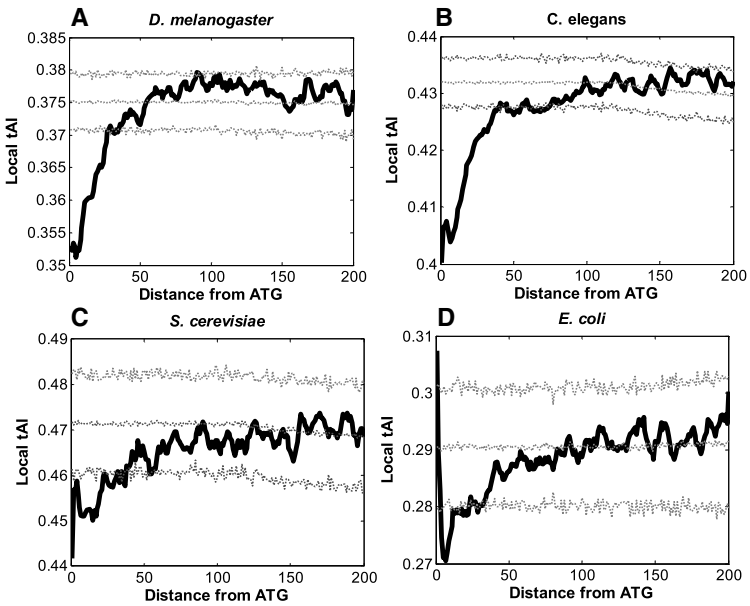
The translation efficiency profile of a gene is defined, for each codon position, as the estimated availability of the tRNAs that participate in translating that codon. The profile is high at codons that correspond to abundant tRNAs and low at codons that correspond to rare tRNAs. In particular, we used the tRNA-adaptation index (tAI) to evaluate translation efficiency (dos Reis et al., 2004 and Experimental Procedures) at each codon. The tAI measure of an entire gene, developed following the classical Codon Adaptation Index (Sharp and Li, 1987), is defined as the (geometric) average of tRNA availability values over all the codons in the gene (see Experimental Procedures). For each codon, the tAI considers the availability of the tRNA with the perfectly matched anti-codon along with weighted contributions from imperfect codon-anticodon pairs, reflecting wobble interactions. Whereas the original tAI measure is defined as an averaged value over the entire gene (Experimental Procedures, Equation 2), here we consider separately each codon along the sequence in what we define as the "local tAI" of a codon (see Experimental Procedures, Equation 1). Table S1 (available online) contains the local tAI value of each of the 61 types of codons in a diversity of species.

Typically, the tAI uses the copy number of tRNA genes in the genome as a proxy for their abundance in the cytoplasm. Although this is a common assumption (see Extended Experimental Procedures 1 and 2, Table S1, and Man and Pilpel, 2007; Percudani et al., 1997; Tuller et al., 2007), we examined it in *S. cerevisiae* using a microarray dedicated to the tRNAs in this species (Dittmar et al., 2004, 2006; Pavon-Eternod et al., 2009; Zaborske et al., 2009). Specifically, we examined the correlation between tRNA abundance and their gene copy number in yeast cells growing on a rich medium. tRNA abundance measurements were based on the Cy3 and Cy5 fluorescence values of each tRNA on the array and came from a labeling method that relies only on the single-stranded 3′NCCA in every tRNA (demonstration of the feasibility of the method was done by quantitative comparison to 2D PAGE analysis of tRNAs; Dittmar et al., 2004, 2006; Pavon-Eternod et al., 2009; Zaborske et al., 2009).

We found that tRNA gene copy numbers are relatively highly correlated with their expression levels in rich medium conditions ($r = 0.76$ over 39 tRNA species, see Figure 1 and Figure S1). We also found that this correlation remains relatively high even when yeast undergo a major metabolic shift, termed "diauxic shift" (DeRisi et al., 1997), from fermenting to respiratory conditions (correlation between 0.65 and 0.71; see also Table S1). The array analysis indicates that the tRNA gene copy number provides a reasonable proxy for tRNA abundance, and we thus use it, in *S. cerevisiae* and in the rest of the species too, in all subsequent tAI calculations.

We started by inspecting the averaged translation efficiency profiles of all the genes in a given genome. To accomplish this analysis, all genes were lined up according to their start codon, and an average local tAI value at each position was calculated (see Experimental Procedures). In parallel, we also aligned all the genes in each genome relative to the stop codon and computed the average across all genes in the last position, the penultimate position, etc. Altogether we analyzed genomes of 27 organisms with representatives from all three domains of life (see Table S2).

As seen in Figure 2 and Table S2 (see also Figure S1, Figure S2, and Figure S3), in almost all species examined, the averaged translation efficiency profile reveals several remarkably conserved features. Translation starts with relatively low-efficiency codons for about the first 30–50 positions. We term this part the "low-efficiency ramp" or the "ramp" for short. The ramp region is then followed by a plateau with ~5%–10% higher translation efficiency on a genome average. A clear outlier in the ramp is the second codon position, which follows the initiating methionine that shows high efficiency compared to its neighboring codons in the majority of the species (Figure S2). This



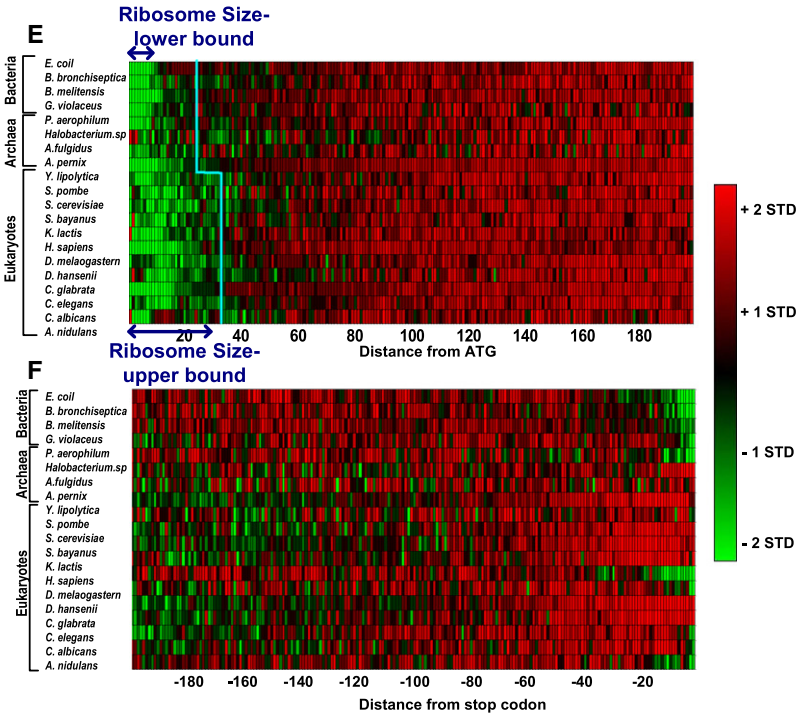


Figure 2. Selected Genome-Averaged Translation Efficiency Profiles

(A–D) Averaged translation efficiency profile (for the start line-up, see Experimental Procedures) for the first 200 codons in *D. melanogaster* (A), *C. elegans* (B), *S. cerevisiae* (C), and *E. coli* (D). Note the different span of values in each subplot. Each figure contains the averaged tAI profile (black) and the randomized profile ±3 standard deviations (gray; see details in the Experimental Procedures).

(E and F) The translation efficiency profiles in various organisms for the start/end codon line-up (see Experimental Procedures). Each row describes the translation efficiency profile of a different organism and each pixel describes a codon. Green denotes lower tAI, whereas red denotes higher tAI (see color bar on the right). The blue vertical line (E) denotes the means of the length of the ramp (Figure S2, Experimental Procedures) in prokaryotes/eukaryotes; the ratio between the means of these regions in prokaryotes/eukaryotes (34.5/24 = 1.43) may correspond to a difference in the size of the footprinted region of the eukaryotic and prokaryotic ribosomes on transcripts.

See also Table S2 and Figure S2.

design might support a fast release and recycling of the initiating methionine tRNA. In most of the examined eukaryotic species, predominantly in fungi, the profile shows an increase in efficiency toward the last ~50 codons of the genes, which in fungi are higher by up to 5% more than the value at the plateau in the middle section of genes (Figure 2F). Properly randomized sequences (and also some particular gene sets, see below) do not give rise to such signals (Figure 2, Table S2).

Although the averaged profile over all genes in a genome is relatively smooth, the profile of single genes is often noisy. Still it was important to inspect single genes, identify those that contribute the most to the observed averaged signal, and examine the possibility that other weaker signals may have been missed at the genome-average level. We thus defined and identified the "bottleneck" region in each individual gene—a sequence window of 15 codons in length (that represents the length of the ribosome footprint region on mRNAs (Alberts et al., 2002; Ingolia et al., 2009; Kaczanowska and Ryden-Aulin, 2007; Menetret et al., 2000; Milo et al., 2009; Zhang et al., 1994; Zhu et al., 1997); very similar results were observed for windows of 10–20 codons in length), with the highest averaged values of 1/(local tAI) (i.e., 15 codons with the longest dwell time in a gene). Figure 3 shows the distribution of locations of the bottleneck regions along all genes in two distantly related yeast species, *S. cerevisiae* and *S. pombe*. Both distributions show a consistent picture—a clear tendency to have the bottleneck relatively early along the genes. Other than this region there are no other regions that show any pronounced preference to contain bottlenecks. This picture shows that the ramp is a superposition of the translation efficiency profile of a relatively high number of genes (e.g., the bottleneck of 1330 genes is within the first 54 codons).

The analyses presented so far have addressed individual genes on one hand and averages of entire genomes on the other. An intermediate level is that of sets of genes that share a biological function. We considered genes that share the same Gene Ontology (GO) slim categories (Hirschman et al., 2006). Figures 4A and 4B and Table S3 show the averaged profiles of genes from representative categories. We found that the genes from

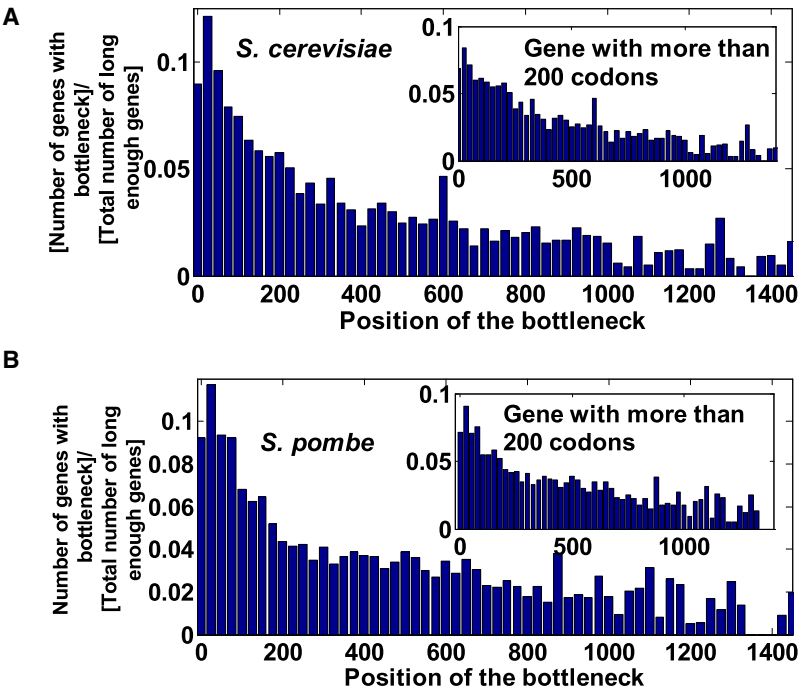


Figure 3. Bottlenecks in Translation Efficiency Tend to Be Localized Close to mRNAs 5′ Ends

The distribution of the positions of the bottlenecks in *S. cerevisiae* (A) and *S. pombe* (B). For each bottleneck position, the number of genes with a bottleneck in that position was normalized by dividing it by the number of genes whose length extends beyond that position. The distribution is similar also when considering only genes with more than 200 codons (inset).

many of the GO slim categories show evidence of containing a ramp. For example, genes that share the GO categories "cellular carbohydrate metabolic process" and "transport" demonstrate a very clear ramp. However, other gene sets, including those that share the GO categories "transcription" and "nucleus organization" do not have a ramp (see more examples in Figure 4B and Table S3). Interestingly, the presence of the ramp is seen even among categories with very different absolute translation efficiency levels. For example, Figure 4C shows the local translation efficiency profile of cytosolic and mitochondrial ribosomal proteins in *S. cerevisiae.* Interestingly, although selection acted to enhance overall efficiency of the cytosolic ribosomal proteins, the initial region shows lower efficiency relative to the rest of the genes.

Beyond inspection of single genes, gene sets, and genome average, the highest level of averaging is from multiple genomes. We averaged all the eukaryotic profiles and all the prokaryotic ones. Notably, the length of the ramp (see Figure S2 for an illustration of how the length of the ramp was computed) in eukaryotes and prokaryotes is around 1–3 ribosomes (depending on the definition of the actual number of nucleotides that are covered by a single ribosme; Alberts et al., 2002; Ingolia et al., 2009; Kaczanowska and Ryden-Aulin, 2007; Menetret et al., 2000; Milo et al., 2009; Zhang et al., 1994; Zhu et al., 1997). The ratio between the length of the ramp in eukaryotes and prokaryotes (mean ramp length in prokaryotes is 24 codons; mean ramp length in eukaryotes is 34.5 codons; the ratio between these lengths is 1.43) may correspond to a difference in the size of the footprinted region of the eukaryotic and prokaryotic ribosomes on transcripts.

We next wanted to examine if the ramp is maintained during an environmental change. For that we returned to the diauxic shift experiment in which we found some changes in the relative representation of the various tRNAs in the tRNA pool (Figures S1 and S2). We computed the ramp for all genes but measured tRNA levels at each time point instead of the static gene copy numbers, and we found that the ramp is largely maintained genome-wide (Figures S1 and S2).

The Universal Translation Efficiency Profile Is under Selection

The translation efficiency profile is highly conserved in evolution, but this fact by itself is not a guarantee that the profile is under direct selection. An alternative might be that the profile is conserved as a by-product of selection acting on other features. We have examined and excluded several specific alternatives.

We started by examining the possibility that the observed profile is conserved merely because the tRNA pool and codon biases are sufficiently conserved. According to this null hypothesis, the interspecies differences in the tRNA pool and the coding sequences are small enough to maintain the translation efficiency even if the profile is not under direct selection. We thus computed the translation efficiency profile of all the genes from one species using the tRNA pool of another, repeating this procedure for various pairs of species. This is a simple computational resemblance of true species hybridization experiments that are used to tell apart the contribution of *cis*- and *trans*-acting factors used in transcription research (Tirosh et al., 2009; Wang et al., 2007; Wittkopp et al., 2004). In one such hybrid analysis, the *S. cerevisiae* genes were translated using the *Y. lipolytica* tRNA pool, and in a reciprocal analysis, the coding sequences of *Y. lipolytica* were translated using the tRNA pool from *S. cerevisiae*. We have chosen these two species because their tRNA pools and their codon biases have diverged quite significantly (Man and Pilpel, 2007), yet both species display the conserved translation efficiency profile. Figure 5 shows that in these hybrids the ramp region is much shorter and shallower. More generally, the ramp region becomes shorter and shallower when hybridizing *S. cerevisiae* with tRNA pools from organisms with increasing evolutionary distances, and also when examining additional pairs of species (see also Table S4). These results indicate that the tRNA pools and the codon preferences have sufficiently diverged between species so as to eliminate the translation efficiency profile if it were not directly selected for. We thus conclude that coevolution of the tRNA pool and the coding sequences took place in

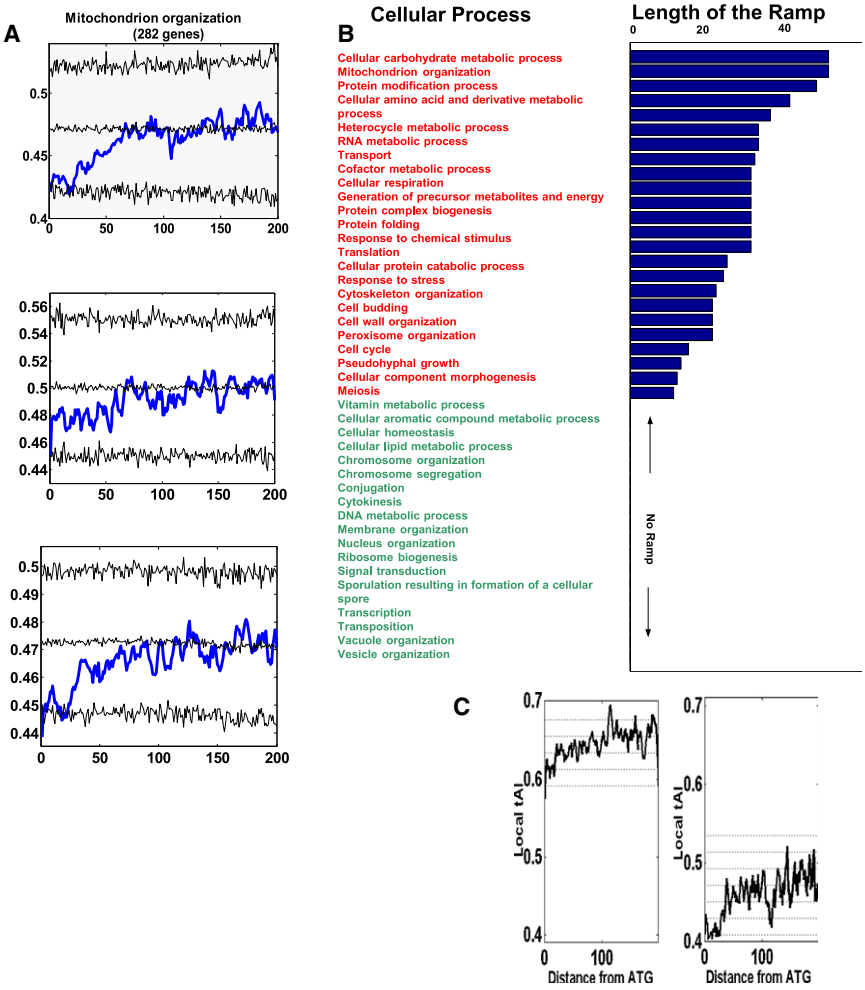


Figure 4. The Profile of Local Translation Efficiency of Selected Gene Groups

(A) The profile of local translation efficiency of three GO slim categories that have a ramp.

(B) The length of the ramp (L2P) of all the GO slim categories.

(C) The profile of local translation efficiency of cytosolic ribosomal proteins (left) and mitochondrial ribosomal proteins (right).

See also Table S3.

each species so as to conserve the translation efficiency profile. This suggestion is in line with the general indications of the emergence of codon bias from coevolution (Vetsigian and Goldenfeld, 2009).

The second null hypothesis relates to the possibility that the observed translation efficiency profile results from selection acting at the amino-acid sequence level. In contrast to this possibility, we found that in the region of the initial ramp the actual codon chosen from all possible codons of the given amino acid is often the one with low efficiency (see the red plots corresponding to the AAtAI profile in Table S2; see Experimental Procedures for explanations about the AAtAI profile). Beyond the ramp region, codon choice does not show this bias. This result excludes the possibility that the observed profile is a by-product of constraints at the protein sequence level.

Likewise, we have excluded another potential alternative reason for the observed profile, that it is a by-product of a putative position-dependent variation in the GC content along genes (Figure S3).

We thus conclude that the translation efficiency profiles are not only universally conserved but are also likely under direct selection, presumably due to direct effects on fitness.

Codon-tRNA Adaptation May Determine Translation Speed and Ribosome Density along Transcripts

A central question thus is what actual physical or biochemical quantity is encoded by the translation efficiency profile. One interesting possibility is that the values of the local tAI determine the local speed of movement of the translating ribosome through each codon along mRNAs. According to this hypothesis, the ribosomes are moving on average more slowly at the ramp region. The hypothesis that the observed profile determines the speed of the ribosome at each position generates a clear prediction about the density of ribosomes at any given position. Assuming that ribosome distributions on mRNAs are at steady state (e.g., assuming little or no premature abortions of translation), the flux of ribosomes through a codon position x is given by

$$j(x) = v(x)^{*}f(x),$$

where v(x) is the speed of translation at the position x, and f(x) is the density of ribosomes at that position. In other words, if the translation efficiency profile is a speed profile, we expect it to be inversely correlated with a ribosome density profile along genes. In particular, our profile predicts a high density of ribosomes at the first 30–50 codon positions.

The averaged ribosome density on transcripts (number of ribosomes divided by the length of the transcript) had been previously measured for most of the genes in the *S. cerevisiae* genome (Arava et al., 2003). Recently, ribosome densities at a single base resolution were measured genome-wide for thousands of transcripts in the *S. cerevisiae* genome (Ingolia et al., 2009). The measured distribution features a high density of ribosome at the 5′ most 50 codons and a plateau from that point on (dos Reis et al., 2004; Ingolia et al., 2009). Thus, the low-efficiency ramp that we observed computationally coincides well with the experimentally observed region of high ribosomal density. In general, comparing the experimentally measured

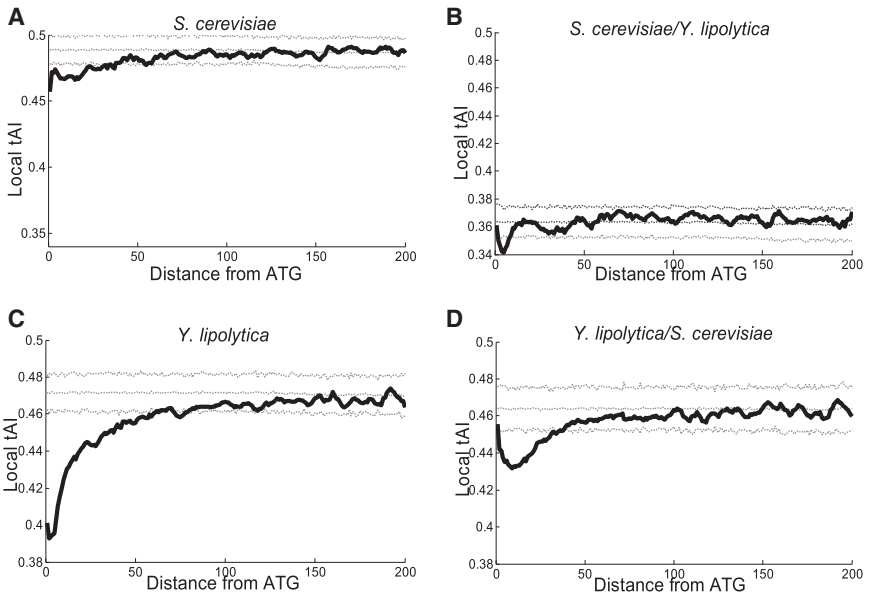


Figure 5. Hybrid Analysis Indicates Selection for Coevolution of tRNA Pools and Genes Sequences to Preserve the Ramp

Translation efficiency profiles with native and nonnative tRNA pools for start codon line-up.

(A) The translation efficiency profile of *S. cerevisiae*.

(B) The translation efficiency profile of *S. cerevisiae* using *Y. lipolytica* tRNA pool.

(C) The translation efficiency profile of *Y. lipolytica*.

(D) The translation efficiency profile of *Y. lipolytica* using *S. cerevisiae* tRNA pool.

The black bolded line represents the actual calculated tAI profile; the gray lines represent the mean ± 3 standard deviations of the tAI profiles of randomized sets of gene. See also Table S4 and Figure S3.

ribosomal density with the reciprocal of the translation efficiency profile reveals relatively high similarity (Pearson correlation $r = 0.5749$; $p < 10^{-28}$).

We further realized that imperfections in the correlation between density and the reciprocal of the translation efficiency profile might reflect a discrepancy between the translation efficiency profile and the actual speed of ribosomes, e.g., due to ribosome traffic jams. We will term the local tAI-based speed profile the "nominal speed profile." At low translation initiation rates, the ribosome may indeed move according to this speed profile. Yet at higher initiation rates, ribosomes may start to jam and hence might move with a different local speed, which we term the "effective speed profile." To estimate the effective speed profile and its deviation from the nominal speed profile at each position, we simulated ribosome movement on transcripts (Zhang et al., 1994, Figure S4, Experimental Procedures, and Extended Experimental Procedures 3–5). The basic rule that governs the movement in the simulation is that a ribosome proceeds through a given codon position according to the nominal translation speed profile unless it collides with the ribosome in front of it, in which case it halts until that ribosome proceeds forward. We ran the simulation for each *S. cerevisiae* gene whose ribosome density profile was measured experimentally by Ingolia et al. (2009) separately and inferred the effective speed profiles and the simulated profiles of ribosome densities. Strikingly, when we averaged the single-gene effective speed profiles (with the same averaging as done for the density profiles; Ingolia et al., 2009), we found that the reciprocal of the obtained effective speed profile highly correlates ($r = 0.93$; $p < 10^{-75}$; Figure 6A) with the experimental density profile. Similarly, when we averaged the profiles of ribosome densities we found that the computed density profile highly resembles the experimental one (up to $r = 0.96$; Figure 6B). The main free parameter in the simulation is the translation initiation rate – the inverse of the time required for a ribosome to attach to, and assemble on, an mRNA (see Figure S4 and Experimental Procedures). We thus experimented with a range of feasible initiation rates. As can be seen in Figure 6B, the high correlation between the experimental and computed densities is maintained throughout a broad range of initiation rates. Only at very low initiation rates (i.e., long initiation times) where traffic jams are not formed does the simulated density lose resemblance to the experimental one (Figure 6B).

Together these results provide an indication that codon-tRNA adaptation may serve as a code that determines ribosomal translation speed. These results also suggest that translation speed, and hence ribosome density, may be encoded in gene sequences and the tRNA pool. The agreement between the computational and the experimental profiles also indirectly lends support to the assumptions of the codon-tRNA adaptation model, e.g., that tRNA gene copy numbers are a good predictor of the tRNA abundance (dos Reis et al., 2004). A corollary of this conclusion is that the computed profile is a simple means to predict the shape of the ribosomal density function in other species, and that the density function seen in yeast is likely conserved. This result also suggests how potential changes in the relative amounts of particular tRNAs (Dittmar et al., 2006; Graslund et al., 2008) might modify the profile of ribosome density in a given species or a given condition. For instance, if the concentration of a rare tRNA, whose matched codons mainly concentrate close to 5′ ends of genes, is elevated at a given condition the result might be a more flat density function.

Finally, under conditions of amino-acid starvation, the first 40–50 codons showed high ribosomal density (Ingolia et al., 2009). This increased ribosomal density seems more pronounced than the density that is observed in rich media. Based on the correlation we demonstrated between ribosome density and reciprocal of speed, we suggest that the speed of the ribosome is slower at the beginning of genes also in this condition. That is, in *S. cerevisiae*, the ramp is maintained (and is probably more pronounced) under a starvation condition (although, the tRNA pool may change in starvation; see, for example, Dittmar et al., 2005).

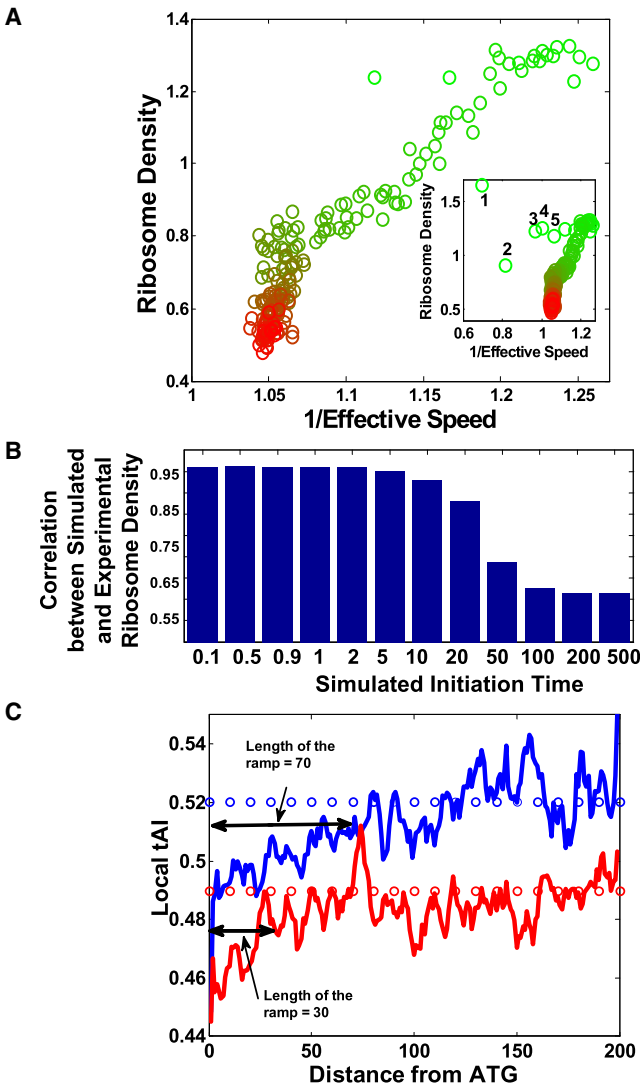


Figure 6. Experimentally Measured Ribosome Density Negatively Correlates with Computed Translation Efficiency

(A) Correlation between experimentally measured ribosomal density (Ingolia et al., 2009) and the reciprocal of the simulated speed profile when not considering the first five codons (which are outliers) and when considering all the codons (the subfigure at the lower right corner). Dots are color coded according to codon location along genes, with the greenest dots representing codons that are close to the ATG, and codons that are farthest away in red. The density and speed profiles were obtained by averaging the profiles at each position of the genes in the *S. cerevisiae* genome. The speed profile was obtained by simulating ribosomal scan of all the transcripts in this species. The Pearson correlation between density and 1/speed is 0.93 ($p < 10^{-75}$). The correlation between density and the reciprocal of original "nominal" translation efficiency profile is lower, $r = 0.5749$ ($p < 10^{-28}$).

(B) The correlation between the mean profile of ribosome density (Ingolia et al., 2009) and the mean profile of simulated ribosome density (Experimental Procedures) at a resolution of single codons for different simulated ribosome binding initiation time (units of the translation time of the slowest codon; see Experimental Procedures for definition).

(C) The translation efficiency profile of genes with the top and the lowest ribosomal density distribution. As can be seen, the extent of ramping decreases at lowly dense genes.

We do note that additional factors are likely involved in setting the speed and density of translating ribosomes. In particular, the folding energy of the mRNA secondary structure appears to be highly relevant (Kudla et al., 2009; Tuller et al., 2010), and it remains to be investigated how tRNA availability and mRNA structure interact in determining the final density profile.

The Potential Fitness Advantages of the Translation Efficiency Profile

The conclusion from the density and speed analysis is that the ramp limits the speed of ribosomes over the first dozens of codons on transcripts and generates as a consequence a high-density area. The generation of a short high-density section at the 5′ region of mRNAs may give rise to a jam-free region over the rest of the transcript because ribosomes that pass the bottleneck are less likely to jam. Under this assumption, ramping would be needed mainly for genes with high overall ribosome density as these genes would be more prone to jamming. In addition, it is conceivable that a jam on a gene with high mRNA copy number is likely to be more detrimental than a comparable jam that occurs on a low mRNA copy number gene. We thus returned to the ribosomal density (Ingolia et al., 2009) and mRNA copy number data (Ghaemmaghami et al., 2003) and characterized each gene by its mRNA copy number or ribosome density or by multiplying its mRNA copy number by its ribosome density, and we looked separately on genes at the top and bottom 10% of the distribution of these features. We found that genes with highest ribosomal density, mRNA level, or the product of ribosome density and mRNA levels display a stronger ramp than genes with the lowest levels of these features (Figure 6C). We verified that this signal is not a result of difference in protein length between the two gene sets (Figure 6C). We note that the length of the slow ramp is a meaningful parameter given that across the various species it is significantly correlated with the extent of selection for translation efficiency (Figure S4). In addition, a similar analysis demonstrated that the ramp signal is stronger in genes with higher protein-to-mRNA abundance ratios (see Figure S4). In that respect, we note that the ramp is "universal" as it is observed in all analyzed species, yet we expect that only a strategic portion of the genes in each genome will feature this design—primarily genes that need to operate at a high production level.

A recent study (Kudla et al., 2009) looking at the influence of mRNA structure on translation initiation provides an opportunity to measure in a more direct way the effect of obeying the translation efficiency profile on the organism's fitness. In that study, 154 versions of the gene encoding the green fluorescent protein (GFP) were synthesized such that the 3rd codon positions were randomized. Each version was driven by a strong promoter, expressed in *E. coli*, and cells' fitness was measured for each

The results remain significant also after controlling for the length of the genes (genes with higher ribosome density tend to be shorter). Specifically, the group of the 20% genes with the lowest ribosome density after removing the 50% longest genes (the final mean length is 1433; ramp length 30) have longer ramp than the group of the 20% genes with the highest ribosome density after removing the 50% shortest genes (the final mean length is 1447; ramp length 89). See also Figure S4.

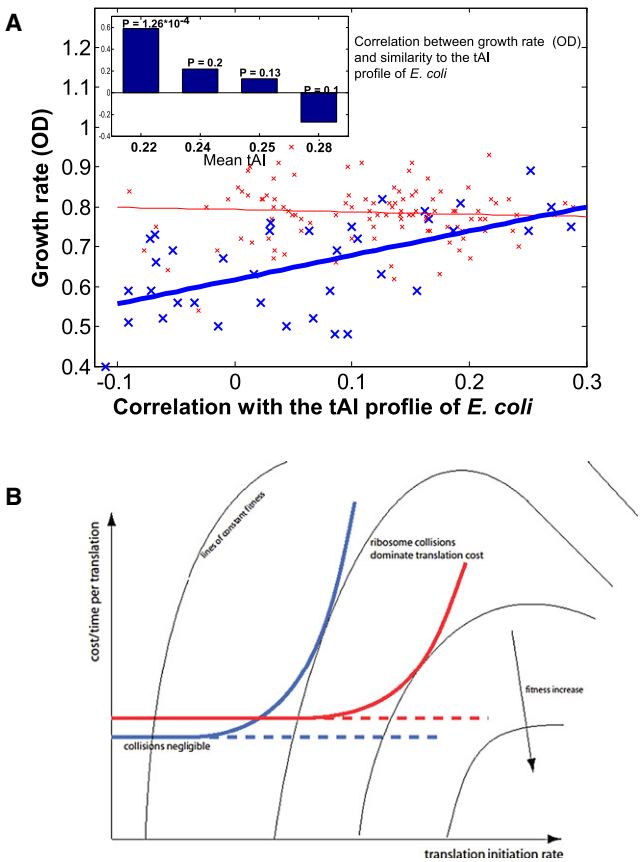


Figure 7. The Effect of Ramping on Fitness, Production, and Expression Cost

(A) Growth rate (measured by OD) of each GFP variant (Kudla et al., 2009) versus similarity (measured by Spearman correlation) between the translation efficiency profile of the variant and the averaged profile of all endogenous genes in *E. coli*. Upper left corner: the above correlation computed separately for quadrants of the GFP library binned according to their tAI values. Main figure: dot plot of the growth rate versus similarity to the genomic translation efficiency profile of *E. coli* for the different variants in the GFP library (correlation coefficient $r = 0.2$; $p < 0.014$); points that are related to the lowest tAI quadrant bin in the subfigure are colored blue, other points are red.

(B) A conceptual model depicting the value in selection for ramping in transcripts with high translation rates. We compare the relationship between translation initiation rate and cost of translation per protein for transcripts with (red) and without (blue) ramp designs. At low translation initiation rates, the ribosomes move independently of one another, thus the cost of translation per protein is independent of the initiation rate. The ramp design incurs a cost because it slows down the ribosomes. At high translation initiation rates, however, ribosome traffic jams increasingly dominate the cost of translation. In this regime, ramping reduces the cost of protein production at a given production level and increases the production capacity at a given cost. The translation initiation rate and the degree of ramping are two knobs that evolution can tune to maximize fitness, which, in the case shown, favor the ramp. The iso-fitness lines reflect an increase in fitness with protein production rate and decreases with the total cost of translation.

See also Table S5.

strain (Kudla et al., 2009). That study presented experimental evidence that average codon bias might minimize the burden of protein expression on the cell (Kudla et al., 2009). Although this GFP library was not designed for the purpose of testing the effects of the ramp, we still found that some of the GFP proteins have a profile that somewhat resembles the ramp, whereas others do not show this feature. Consistently, we find a modest, yet significant correlation between the extent to which a GFP variant obeys the translation efficiency profile and the fitness of the strain that expresses it ($r = 0.2$; p value < 0.014; Figure 7A). Further, we partitioned the genes in the library according to their average tAI value and focused on 25% of the variants with the lowest tAI values. These genes could generate the highest burden on fitness due to potential overall low translation efficiency (Kudla et al., 2009). Among those genes, there exists a higher correlation ($r = 0.6$, $n = 37$, p value $< 1.26*10^{-4}$) between the extent of agreement with the translation efficiency profile and bacterial fitness (Figure 7A). Thus, especially among genes with lower translation efficiency (and hence higher ribosome density) obeying the translation efficiency profile is crucial for minimizing fitness reduction. This observation thus suggests that the ramp in translation efficiency profiles found in endogenous genes may have been selected for minimizing the cost of protein expression due to translation.

The data of Kudla et al. (2009) can be used for understanding the relation between ramp, fitness, and the component of the initiation rate (relative to the elongation rate) that is determined by the properties of the 5′UTRs. All the GFP variants have identical 5′UTR sequences and thus identical (absolute) initiation rates; however, as the coding sequences were randomized, the elongation rate (measured by the tAI) is different for different GFP variants. Thus, the different GFP variants have different *relative* initiation-to-elongation rates. The relative initiation rate is higher when the tAI is lower. Figure 7A (inset), thus, demonstrates that the ramp is more important when the *relative* initiation rate is higher.

DISCUSSION

The Potential Role of the Ramp in the Context of Initiation Control

Translation initiation is a prime point of control as it governs the binding and assembly of the ribosome and the initiation machinery on a transcript. These processes mainly take place at the 5′UTRs of genes. The ramp we describe here may represent an important next stage of translational control that modulates the parameters set by the previous initiation stage. The ramp may thus couple between initiation and elongation and add unique regulatory features, stemming from the fact that it is "written" on a translated sequence.

Robustness and Economy—Reduction of Expression Cost by Filtering out Randomness

The 5′UTR of each mRNA serves as a control region by modulating translation initiation through, for example, influence over ribosome binding rate. The secondary structure and the nucleotide sequence in this region affect both the average spacing between ribosomes (and hence the protein production rate at steady state) and the statistical deviations around that average. As we turn the initiation rate knob to increase the protein production rate, these fluctuations become increasingly more likely to generate instances of too closely spaced ribosomes that can jam and, potentially, also abort translation. Thus, fluctuations

become increasingly more costly. The ramp in the coding sequence, however, may provide a second and independent knob that can tune down the variance set by initiation rate in the spacing between ribosomes. The ramp consists of dozens of codons, each determining a separate random event of tRNA binding, and thus serves as a very effective noise filter; the probability that with the passage of a given ribosome, all the codons in the ramp will allow fast movement of the ribosome, and potentially collision with the ribosome ahead of it, is practically zero. Of course such a design is not needed for proteins with a very low binding rate; these are unlikely to jam anyway, and indeed the extent of ramping decreases at lowly dense genes in the yeast genome (Figure 6C).

Potential for Gene-Specific and Condition-Specific Control

The ramp may encode an interesting sensing capability—the low-efficiency codons at the beginning of transcripts may allow high sensitivity to the abundance of amino acid-loaded tRNAs in the cell. As such, it may provide a simple mechanism for early, thus low-cost, abortion of translation in the case of a paucity of raw material. In more general terms, compared to control through the UTR, the ramp has a potential to control differentially individual genes under different conditions. Indeed we see that different functional gene sets may represent different designs of the ramp. We also predict that the shape of the ramp may change for particular genes across conditions (e.g., if the concentration of some of the tRNAs is modulated relative to others). The ramp may thus encode for gene-specific and condition-specific dynamic control of early translation elongation.

A Range of Potential Physiological Roles for the Ramp

The ramp in translation efficiency has the potential of reducing traffic jamming of ribosomes. Reduction in jamming is, in turn, desired for several reasons. First, it reduces the total amount of time that ribosomes are sequestered on a given transcript. Second, jammed ribosomes, which halt at slow codons (Li et al., 2006), spend more time on the transcript, increasing the probability of spontaneous fall off. Abortive protein synthesis may also occur due to collisions between jammed ribosomes; thus ramps may be advantageous in preventing spontaneous and collision-dependent abortions. On top of that, the ramp may limit most abortions that do occur to the beginning of the transcripts. This may be desired because in these regions fall off is least wasteful in terms of energy (ATPs) and raw materials (e.g., charged tRNAs). Thus low speed at the beginning may reflect reduced purifying pressure against early abortion, or a pressure to concentrate abortions at early stages if they reduce late, costly ones. In this respect, the elevation in translation speed toward the end of transcripts (seen mainly among the fungi) may reflect a deliberate selective pressure to avoid late abortions.

We assume that at a given level of protein expression the cost of production increases with the total time ribosomes spend on mRNAs. Although the initial ramp may increase that time (and hence the cost) over the initial section of mRNA transcripts, it may result in an overall shorter duration of sequestering of translating ribosomes. Thus if ramping decreases the probability of jamming it may reduce the cost of gene expression at a given production level and increase the production capacity at a given cost (Figure 7B; see also Table S5 and Extended Experimental Procedures 3–5). Note, however, that when initiation rate is low, jamming is less likely as matter of course and hence ramping may only incur a slowing-down cost with no subsequent gain (Figure 7B), explaining why genes with low ribosome densities have a shorter ramp (Figure 6C).

A profile of translation speed may also correspond to a position-specific profile of translation errors—longer dwell times at slow codons may result not only in abortion but also in higher translation error probabilities (Drummond and Wilke, 2008; Kurland, 1992). According to this notion we predict that the beginnings of proteins may accumulate more translation mutations compared to other regions in proteins. The speed profile may also constitute an essential code for proper protein folding (Kimchi-Sarfaty et al., 2007; Widmann et al., 2008).

Some of these considerations may also apply to other biological processes. For example, it would be interesting if transcription, and the movement of motor proteins such as kinesin, implement a similar design of "slow start" to alleviate some of the "process costs" and the effects of stochasticity (Schnitzer and Block, 1995; Svoboda et al., 1994). Our work may have direct implications not only for understanding evolutionary processes underlying translational control but also for the technology of heterologous gene expression. This more applied area is another domain where the question "what is the optimal level of expression of genes?" must be asked. The science and art of expressing a gene from one species in another often amounts to modifying the codons of the gene and supplementing the host with specific tRNAs (Graslund et al., 2008). Yet the full challenge of heterologous expression is not only to maximize expression per host cell but also to minimize the burden on the host or maximize fitness of the entire "factory." We suggest that implementation of appropriate ramping in heterologous proteins, given the host's tRNA pool, might improve the yield and success rate with this technology for the same reasons it was selected evolutionarily.

EXPERIMENTAL PROCEDURES

Data Sources of Information

The tRNA copy numbers and the coding sequences of the nine yeasts were downloaded from the work of Man and Pilpel (Man and Pilpel, 2007). The tRNA copy numbers of the other organisms were downloaded from the Genomic tRNA Database (http://lowelab.ucsc.edu/GtRNAdb/) (Lowe and Eddy, 1997). The coding sequences of the fly and the worm were downloaded from BioMart (Durinck et al., 2005) on August 2008; the coding sequences of other eukaryotes were downloaded from NCBI on May 2009. The coding regions of all the archaea and bacteria were downloaded from NCBI (http://www.ncbi.nlm.nih.gov/Ftp/) on August 2008. The mRNA levels and protein abundance were downloaded from the work of Ghaemmaghami (Ghaemmaghami et al., 2003).

Per nucleotide ribosome density of 1525 genes was obtained from the work of Ingolia et al. (2009); when we compared density to speed (tAI) or simulated density we used the same set of genes as was used in Ingolia et al. (2009).

The version of the GFP protein with synthetic random codon bias and corresponding measurements of growth rate (optical density, OD) were obtained from the work of Kudla et al. (2009). Information about the length of the eukaryotic/prokaryotic ribosome footprinted mRNA segment was based on Alberts et al. (2002), Ingolia et al. (2009), Kaczanowska and Ryden-Aulin (2007),

Menetret et al. (2000), Milo et al. (2009), Zhang et al. (1994), and Zhu et al. (1997). The lists of ribosomal proteins were downloaded from Hirschman et al. (2006).

Computing tAI, AAtAI

In this subsection we will briefly describe the different measures for translation efficiency used in this work. We used two measures for translation efficiency: tAI and AAtAI. As we explain later, the latter was used as a control for a potential amino-acid sequence bias.

The tAI

We computed the tAI similarly to the way it was computed in the work of dos Reis et al. (2004). This measure gauges the availability of tRNAs for each codon along an mRNA. As codon-anticodon coupling is not unique due to wobble interactions, several anticodons can recognize the same codon, with different efficiency weights (see dos Reis et al. for all the relations between codon-anti-codons).

Let n_i be the number of tRNA isoacceptors recognizing codon i. Let $tGCNij$ be the copy number of the jth tRNA that recognizes the ith codon, and let S_{ij} be the selective constraint on the efficiency of the codon-anticodon coupling. We define the *absolute adaptiveness*, W_i, for each codon i as

$$W_i = \sum_{j=1}^{n_i} (1 - S_{ij}) tGCN_{ij}. \quad (1a)$$

From W_i we obtain w_i, which is the *relative adaptiveness value of codon i*, by normalizing the W_i's values (dividing them by the maximal of all 61W_i).

$$w_i = W_i / (\max W_i). \quad (1b)$$

The final tAI of a gene, g, is the following geometric mean:

$$tAIg = \left(\prod_{k=1}^{lg} wikg \right)^{1/lg}, \quad (2)$$

where i_{kg} is the codon defined by the k'th triplet on gene g, and lg is the length of the gene (excluding stop codons).

We made one change compared to the computations of dos Reis et al.; we re-inferred the S_{ij} values by performing hill-climbing optimization of the Spearman correlation between protein abundance and translation efficiency in *S. cerevisae*. For this purpose we used the protein abundance measurements of Ghaemmaghami et al. (2003). The S_{ij} values can be organized in a vector (S vector) as described in dos Reis et al. (2004); each component in this vector is related to one wobble nucleoside-nucleoside paring: I:U, G:U, G:C, I:C, U:A, I:A, etc. The final S vector obtained by our optimization was

[0 0 0 0 0.561 0.28 0.9999 0.68 0.89].

The AAtAI

The *a*mino *a*cid tAI (AAtAI) was computed similarly to the tAI. The only change is that each w_i is obtained from W_i by dividing it by the maximal W_i of all codons coding the same amino acid that codon i codes for. Thus, the AAtAI reflects normalization by the maximal possible tAI of a given protein sequence.

Computing Local tAI and AAtAI

In the case of the tAI, the local profile of a gene was defined as the vector of the tAI values assigned to the gene's codons (omitting the first ATG), i.e.:

$$Local\_tAI_{Gene_i} = \left(tAI_{c_2}, tAI_{c_3}, \ldots, tAI_{c_n} \right),$$

where c_i is the codon at position i in the gene (c_n is the codon before the stop codon). For a particular species, all the genes in the genome were lined up once according to their start codon, and once according to their stop codon, and averaged head and tail profiles were calculated as

$$\overline{Local\_tAI_{start}} = \left(\overline{tAI_2}, \overline{tAI_3}, \overline{tAI_4}, \ldots \right)$$
$$\overline{Local\_tAI_{end}} = \left(\overline{tAI_n}, \overline{tAI_{n-1}}, \overline{tAI_{n-2}}, \ldots \right)$$

where

$$\overline{tAI_i} = \sum_{Genes_i} tAI_{c_i} / |Genes_i|$$

and $Genes_i$ is the number of genes with at least $i+1$ codons.

The local values for AAtAI were computed in a similar way—we considered AAtAI of codons instead of codons' tAI.

Profiles of AAtAI describe tAI after controlling for amino-acid bias. Thus, we expect these profiles to be similar to the profiles of the tAI if indeed the observed tAI profile is related to translation efficiency and not amino-acid bias.

Randomized Profiles of Translation Efficiency

To verify that the observed translation efficiency profile is not a result of the fact that we registered the genes by the start/stop codon, we performed the following control.

Each coding sequence was randomly shuffled, and the average genome profile was calculated. This process was repeated 100 times. The mean and standard deviations of the 100 sets of profile were then calculated for each position. The randomized profiles were compared to the original profile.

Simulation of Ribosomal Movement

To explore the movement of ribosomes along the mRNA sequences we used a simulation, based on the model of Zhang et al. (1994) (see Figure S4 and more details in Extended Experimental Procedures 3–5). By this model, a single codon translation time is determined per ribosome by the translation time of that codon (i.e., the tAI of the codon) and the potential presence of a ribosome in front of it: if there is no ribosome in front of the given ribosome its velocity is solely governed by the translation efficiency profile, yet to maintain a required minimal distance between the subsequent ribosomes, if there is a ribosome in front of the given one, it is delayed until the ribosome in front of it has proceeded on. Other parameters of the simulations are as follows: the minimum distance between two consecutive ribosomes, the ribosome binding time, and the termination time—the time required for the ribosome to release the mRNA.

Measurement of the tRNA Pool in *S. cerevisiae*

Logarithmic culture (2.5 × 10^6 cells/ml) of *S. cerevisae* cells (Strain 4741: *MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0*) was grown on YPD medium (2% yeast extract, 1% peptone, 1% dextrose) at 30°C until reaching stationary phase. During growth, glucose concentration in the media was measured using UV test kit (Boehringer Mannheim catalog number 716251) and the diauxic shift was identified. At 1.5 hr intervals during growth, samples were taken and frozen in liquid nitrogen. RNA was extracted using MasterPure (EPICENTER Biotechnologies) and hybridized to tRNA microarray as described in Dittmar et al. (2006), Pavon-Eternod et al. (2009), and Zaborske et al. (2009). Briefly, the basic protocol consists of four steps starting from total RNA: (1) deacylation to remove remaining amino acids attached to the tRNA, (2) selective Cy3/Cy5 labeling of tRNA, (3) hybridization on commercially printed arrays, and (4) data analysis.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, four figures, and five tables and can be found with this article online at doi:10.1016/j.cell.2010.03.031.

ACKNOWLEDGMENTS

We thank the Pilpel lab and Chaim Kahana, Ran Kafri, Ron Milo, Uri Alon, Michael Springer, Jeremy Gunawardena, Eytan Ruppin, and Martin Kupiec for stimulating discussions and Orna Man for critically reading the manuscript. We thank the "Ideas" program of the European Research Council, EMBRACE

Network of Excellence grant of the European Commission within its FP6 Programme, and the Ben May Charitable Trust for grant support. T.T. is a Koshland Scholar at the Weizmann Institute of Science.

Received: July 7, 2009
Revised: December 8, 2009
Accepted: March 19, 2010
Published: April 15, 2010

REFERENCES

Akashi, H. (2003). Translational selection and yeast proteome evolution. Genetics *164*, 1291–1303.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002). Molecular Biology of the Cell (New York: Garland Publishing).

Arava, Y., Wang, Y., Storey, J.D., Liu, C.L., Brown, P.O., and Herschlag, D. (2003). Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. USA *100*, 3889–3894.

DeRisi, J.L., Iyer, V.R., and Brown, P.O. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. Science *278*, 680–686.

Dittmar, K.A., Mobley, E.M., Radek, A.J., and Pan, T. (2004). Exploring the regulation of tRNA distribution on the genomic scale. J. Mol. Biol. *337*, 31–47.

Dittmar, K.A., Sorensen, M.A., Elf, J., Ehrenberg, M., and Pan, T. (2005). Selective charging of tRNA isoacceptors induced by amino-acid starvation. EMBO Rep. *6*, 151–157.

Dittmar, K.A., Goodenbour, J.M., and Pan, T. (2006). Tissue-specific differences in human transfer RNA expression. PLoS Genet. *2*, e221.

dos Reis, M., Savva, R., and Wernisch, L. (2004). Solving the riddle of codon usage preferences: a test for translational selection. Nucleic Acids Res. *32*, 5036–5044.

Drummond, D.A., and Wilke, C.O. (2008). Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. Cell *134*, 341–352.

Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., De Moor, B., Brazma, A., and Huber, W. (2005). BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics *21*, 3439–3440.

Ghaemmaghami, S., Huh, W.K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K., and Weissman, J.S. (2003). Global analysis of protein expression in yeast. Nature *425*, 737–741.

Graslund, S., Nordlund, P., Weigelt, J., Hallberg, B.M., Bray, J., Gileadi, O., Knapp, S., Oppermann, U., Arrowsmith, C., Hui, R., et al. (2008). Protein production and purification. Nat. Methods *5*, 135–146.

Gray, N.K., and Hentze, M.W. (1994). Regulation of protein synthesis by mRNA structure. Mol. Biol. Rep. *19*, 195–200.

Hirschman, J.E., Balakrishnan, R., Christie, K.R., Costanzo, M.C., Dwight, S.S., Engel, S.R., Fisk, D.G., Hong, E.L., Livstone, M.S., Nash, R., et al. (2006). Genome Snapshot: a new resource at the Saccharomyces Genome Database (SGD) presenting an overview of the Saccharomyces cerevisiae genome. Nucleic Acids Res. *34*, D442–D445.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science *324*, 218–223.

Kaczanowska, M., and Ryden-Aulin, M. (2007). Ribosome biogenesis and the translation process in Escherichia coli. Microbiol. Mol. Biol. Rev. *71*, 477–494.

Kimchi-Sarfaty, C., Oh, J.M., Kim, I.-W., Sauna, Z.E., Calcagno, A.M., Ambudkar, S.V., and Gottesman, M.M. (2007). A "Silent" polymorphism in the MDR1 gene changes substrate specificity. Science *315*, 525–528.

Kudla, G., Murray, A.W., Tollervey, D., and Plotkin, J.B. (2009). Coding-sequence determinants of gene expression in Escherichia coli. Science *324*, 255–258.

Kurland, C.G. (1992). Translational accuracy and the fitness of bacteria. Annu. Rev. Genet. *26*, 29–50.

Li, X., Hirano, R., Tagami, H., and Aiba, H. (2006). Protein tagging at rare codons is caused by tmRNA action at the 3′ end of nonstop mRNA generated in response to ribosome stalling. RNA *12*, 248–255.

Lowe, T.M., and Eddy, S.R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. *25*, 955–964.

Man, O., and Pilpel, Y. (2007). Differential translation efficiency of orthologous genes is involved in phenotypic divergence of yeast species. Nat. Genet. *39*, 415–421.

Menetret, J.F., Neuhof, A., Morgan, D.G., Plath, K., Radermacher, M., Rapoport, T.A., and Akey, C.W. (2000). The structure of ribosome-channel complexes engaged in protein translocation. Mol. Cell *6*, 1219–1232.

Milo, R., Jorgensen, P., Moran, U., Weber, G., and Springer, M. (2009). BioNumbers–the database of key numbers in molecular and cell biology. Nucleic Acids Res. *23*, 23.

Pavon-Eternod, M., Wei, M., Pan, T., and Kleiman, L. (2009). Profiling non-lysyl tRNAs in HIV-1. RNA *16*, 267–273.

Percudani, R., Pavesi, A., and Ottonello, S. (1997). Transfer RNA gene redundancy and translational selection in Saccharomyces cerevisiae. J. Mol. Biol. *268*, 322–330.

Qin, H., Wu, W.B., Comeron, J.M., Kreitman, M., and Li, W.H. (2004). Intragenic spatial patterns of codon usage bias in prokaryotic and eukaryotic genomes. Genetics *168*, 2245–2260.

Schnitzer, M.J., and Block, S.M. (1995). Statistical kinetics of processive enzymes. Cold Spring Harb. Symp. Quant. Biol. *60*, 793–802.

Sharp, P.M., and Li, W.H. (1987). The codon Adaptation Index–a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. *15*, 1281–1295.

Svoboda, K., Mitra, P.P., and Block, S.M. (1994). Fluctuation analysis of motor protein movement and single enzyme kinetics. Proc. Natl. Acad. Sci. USA *91*, 11782–11786.

Tirosh, I., Reikhav, S., Levy, A.A., and Barkai, N. (2009). A yeast hybrid provides insight into the evolution of gene expression regulation. Science *324*, 659–662.

Tuller, T., Kupiec, M., and Ruppin, E. (2007). Determinants of protein abundance and translation efficiency in S. cerevisiae. PLoS Comput. Biol. *3*, e248.

Tuller, T., Waldman, Y.Y., Kupiec, M., and Ruppin, E. (2010). Translation efficiency is determined by both codon bias and folding energy. Proc. Natl. Acad. Sci. USA *107*, 3645–3650.

Vetsigian, K., and Goldenfeld, N. (2009). Genome rhetoric and the emergence of compositional bias. Proc. Natl. Acad. Sci. USA *106*, 215–220.

Wang, D., Sung, H.M., Wang, T.Y., Huang, C.J., Yang, P., Chang, T., Wang, Y.C., Tseng, D.L., Wu, J.P., Lee, T.C., et al. (2007). Expression evolution in yeast genes of single-input modules is mainly due to changes in trans-acting factors. Genome Res. *17*, 1161–1169.

Widmann, M., Clairo, M., Dippon, J., and Pleiss, J. (2008). Analysis of the distribution of functionally relevant rare codons. BMC Genomics *9*, 207.

Wittkopp, P.J., Haerum, B.K., and Clark, A.G. (2004). Evolutionary changes in cis and trans gene regulation. Nature *430*, 85–88.

Zaborske, J.M., Narasimhan, J., Jiang, L., Wek, S.A., Dittmar, K.A., Freimoser, F., Pan, T., and Wek, R.C. (2009). Genome-wide analysis of tRNA charging and activation of the eIF2 kinase Gcn2p. J. Biol. Chem. *284*, 25254–25267.

Zhang, S., Goldman, E., and Zubay, G. (1994). Clustering of low usage codons and ribosome movement. J. Theor. Biol. *170*, 339–354.

Zhu, J., Penczek, P.A., Schroder, R., and Frank, J. (1997). Three-dimensional reconstruction with contrast transfer function correction from energy-filtered cryoelectron micrographs: procedure and application to the 70S Escherichia coli ribosome. J. Struct. Biol. *118*, 197–219.

EXTENDED EXPERIMENTAL PROCEDURES

1. Justification for Using tAI

The tAI is based on the genomic tRNA copy number (tGCN) as a surrogate measure for the cellular abundances of tRNAs; it is justified by several observations.

First we provide here a direct experimental support for the correlation between tRNA gene copy number and expression levels in yeast.

Second, 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 (Percudani et al., 1997) was reported. In *B. subtilis*, a correlation of 0.86 between tRNA copy number and tRNA abundance was reported (Kanaya et al., 1999). Similarly, previous papers reported about significant correlation between genomic tRNA copy number and tRNA abundance in *E. coli* (Dong et al., 1996; Ikemura, 1981). A related interesting result is the analysis of (Sorensen and Pedersen, 1991) 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 *wi* of these codons, which is based on the tRNA pool and affinity of codon-anticodon coupling and is the basis for the tAI calculation, captures the ratio of translation rate between the two codons. Calculating *wi* values for *E. coli* we found that the ratio between the *wi* 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.

Finally, the correlation between tRNA copy number and tRNA abundance in human tissues (brain, liver, vulva, testis, ovary, thymus, lymph node, spleen; Dittmar et al., 2006) is between 0.605 (thymus) and 0.83 (brain) (see Table S1); see Extended Experimental Procedures 5 about how this correlations were computed.

Third, a recent study showed that in *S. cerevisiae* the promoters of many of the tRNA genes have a low predicted affinity to the nucleosome, suggesting a constitutive expression with little transcriptional regulation capacity (Segal et al., 2006). 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, the tAI has been shown to be highly correlated (r = 0.63 for *S. cerevisiae*) to protein expression levels (Man and Pilpel, 2007; Tuller et al., 2007). It was found that even among genes with similar transcript levels, higher tAI often corresponds to higher protein abundance (Man and Pilpel, 2007).

This definition stems from an early observation of a trend of increasing codon usage bias with increasing gene expression levels in a sample of *E. coli* genes (Sharp and Li, 1986), and that tRNA concentrations are rate limiting in the elongation of nascent peptides (Varenne et al., 1984). The translation efficiency, as defined above, has also been shown to be correlated with translation rate and accuracy (Akashi, 2003), phenotypic divergence of yeast species (Man and Pilpel, 2007), evolutionary rate (Wall et al., 2005), and to also play part in protein functionality (Kimchi-Sarfaty et al., 2007).

2. Inference of tRNA Expression Levels in Human Tissues

Dittmar et al. (2006) have characterized the expression levels of tRNA across various human tissues relative to the brain. In order to get an estimate of the absolute tRNA pool in each tissue we employed an optimization procedure.

Briefly, instead of assuming that the tRNA levels in the brain are determined by the tRNA copy numbers, we allowed their levels to vary and, while maintaining the tRNA expression ratios reported in the work of Dittmar et al. as constraints, we maximized the overall correlation between tAI and expression levels across the tissues with both expression data and relative tRNA values.

Formally, let T_i denote the inferred vector of tRNA levels in tissue i (i = 1 denotes the brain); let $tAI(T_i)$ denote the corresponding vector of gene tAI when using T_i as proxy of tRNA levels; let $GE\ i$ demote the vector gene expression in tissue i ; Let $(R\cdot,\cdot)$ denote spearman correlation.

We solved the following optimization problem:

Optimize $R([tAI\ (T_1),tAI\ (T_2),..],[GE(T_1),GE(T_2),..])$,

such that:

(1) For every i each entry in $T\ I$ is larger than 0 if the copy number of the corresponding entry is larger than 0; otherwise the entry remains 0.

(2) The ratios between the expression levels of tRNAs in each tissue (i = 1,2,..) relative to the brain (i = 1) are identical to the results reported in Dittmar et al. (2006).

To solve this problem we employed the optimization approach of Nelder–Mead (Lagarias et al., 1998).

3. Implementing the Model of Zhang et al. on Artificial Genes and on the Genome of S. cerevisae to Show the Optimality of the Nondecreasing Profile

At the first stage, we generated artificial genes with three types of codons (based on the tAI of codons in *S. cerevisae*): fast (tAI = 1), slow (tAI = 0.0271; a weighted average of the three slowest codons in *S. cerevisae*), and medium (tAI = 0.43; a weighted average of all

the codons in *S. cerevisae* considering codon bias). All the results were compared to an artificial gene with an ''optimal'' profile of translation efficiency where the first 50 codns are slow, the last 50 codons were fast, and the others codons are medium (a total of 500 codons). For comparison we generated three six randomized versions of the ''optimal'' profile: (1) Random permutation of the codons of the initial ''optimal'' profile. (2) Random permutation of codons 51–500 of the ''optimal'' profile (codons 1–50 remain slow). (3) Random permutation of codons 1–450 of the ''optimal'' profile (codons 451–500 remain fast). (4) Random cyclic shift of the codons of the ''optimal'' profile. (5) Random cyclic shift of the codons of the ''optimal'' profile while not touching the first 50 codons. (6) Random cyclic shift of the codons of the ''optimal'' profile while not touching the last 50 codons.

We computed nine measures of quality: (1) Number of ribosome initiations before reaching steady state (RISS). (2) Time to reach steady state (TSS). (3) Initiation time at steady state (IT). (4) Translation time at steady state (TT). (5) Number of ribosomes per mRNA at steady state (RPM). (6) Number of codons translated per ''second'' (arbitrary time unit; CPS). (7) Number of ribosome collisions at steady state (RCSS). (8) Number of ribosome collisions till steady state (RCTSS). (9) Translation efficiency, number of codons translated per ''second'' per ribosome: measure 5 divided by measure 4) (CPSPR).

As can be seen, the ''optimal'' profile is not worse than all the random models by all criterions (see Table S5). Red denotes significantly low; green denotes significantly high. The simulations demonstrate that the slow translation at the beginning is significantly important for efficient translation and decreased number of ribosome collisions. On the other hand, the fast translation at the end does not seem to play an important role when measuring these readouts under the Zhang model.

4. Implementing the Model of Zhang et al. on S. cerevisae Genome

In the second stage, we implemented the model of Zhang on all the *S. cerevisae* genes. The local translation speed at each codon and at each gene was derived from the tAI of the codon (see Experimental Procedures). Higher tAI corresponded to higher speed. We compared the profiles of all the genes to two randomized versions of the input: (1) Random permutations of the codons of all the genes, and (2) Random shift of the codons of all the genes. To this end, we used the nine measures of quality mentioned above.

The simulation shows a similar trend as in the previous subsection. The original genome outperforms the two randomized genomes according to all the measures of quality (see Table S5).

5. More Details about the Simulation of Ribosomal Movement

In case where the input to the simulation was a vector of tAI values, we used the reciprocal of such values at each codon position as a proxy for the waiting time of the ribosome at the position. The translation time values were normalized to set the maximum translation time to 1.

The simulation has three main parameters: (1) *The minimum distance between two consecutive ribosomes* (*H*). (2) The *ribosome binding time (the initiation time)*—technically, it was defined as the time required for a ribosome to start translating the mRNA, measured from the moment the ribosome in front has cleared the initiation site (first *H* codons). (3) *The termination time*—the time required for the ribosome to release the mRNA.

The minimum distance between ribosomes was set to be 15 codons (see BioNumbers for the size of the ribosome). We used initiation time of 0.9 and the termination time was set to 0.1. The value of the slowest codon (longest translation time) was also the same as in (Zhang et al., 1994) (time = 1). The reported results were not sensitive to changes of ±100% these parameters.

The steady-state parameters are defined by the translation parameters of a ribosome during the steady state. The system reaches steady state when the following conditions are met: (1) At least one ribosome finished translating the mRNA. (2) The number of ribosomes translating the mRNA is constant (i.e., the initiation rate is equal to the release rate).

At steady state all the translation parameters are constant and one can define the following parameters: (a) Total translation time: the time it takes a single ribosome to translate the mRNA. (b) Number of ribosomes on the transcript: the number of initiations in the period of a single translation time. (c) Average translation rate: length of the message (number of codons) divided by the translation time (i.e., number of codons translated per time unit). (d) Number of ribosomal collisions till steady state and at steady state. This is the number of events (per unit time) in which a ribosome cannot continue due to a ribosome in front of it till steady state and at steady state.

SUPPLEMENTAL REFERENCES

Dong, H., Nilsson, L., and Kurland, C.G. (1996). Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates. J. Mol. Biol. *260*, 649–663.

Ikemura, T. (1981). Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. J. Mol. Biol. *151*, 389–409.

Kanaya, S., Yamada, Y., Kudo, Y., and Ikemura, T. (1999). Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of Bacillus subtilis tRNAs: gene expression level and species-specific diversity of codon usage based on multivariate analysis. Gene *238*, 143–155.

Lagarias, J.C., Reeds, J.A., Wright, M.H., and Wright, P.E. (1998). Convergence properties of the Nelder-Mead simplex method in low dimensions. SIAM J. Optim. *9*, 112–147.

Segal, E., Fondufe-Mittendorf, Y., Chen, L., Thastrom, A., Field, Y., Moore, I.K., Wang, J.P., and Widom, J. (2006). A genomic code for nucleosome positioning. Nature *442*, 772–778.

Sharp, P.M., and Li, W.H. (1986). An evolutionary perspective on synonymous codon usage in unicellular organisms. J. Mol. Evol. *24*, 28–38.

Sorensen, M.A., and Pedersen, S. (1991). Absolute in vivo translation rates of individual codons in *Escherichia coli*. The two glutamic acid codons GAA and GAG are translated with a threefold difference in rate. J. Mol. Biol. *222*, 265–280.

Varenne, S., Buc, J., Lloubes, R., and Lazdunski, C. (1984). Translation is a nonuniform process: Effect of tRNA availability on the rate of elongation of nascent polypeptide chains. J. Mol. Biol. *180*, 549–576.

Wall, D.P., Hirsh, A.E., Fraser, H.B., Kumm, J., Giaever, G., Eisen, M.B., and Feldman, M.W. (2005). Functional genomic analysis of the rates of protein evolution. Proc. Natl. Acad. Sci. USA *102*, 5483–5488.

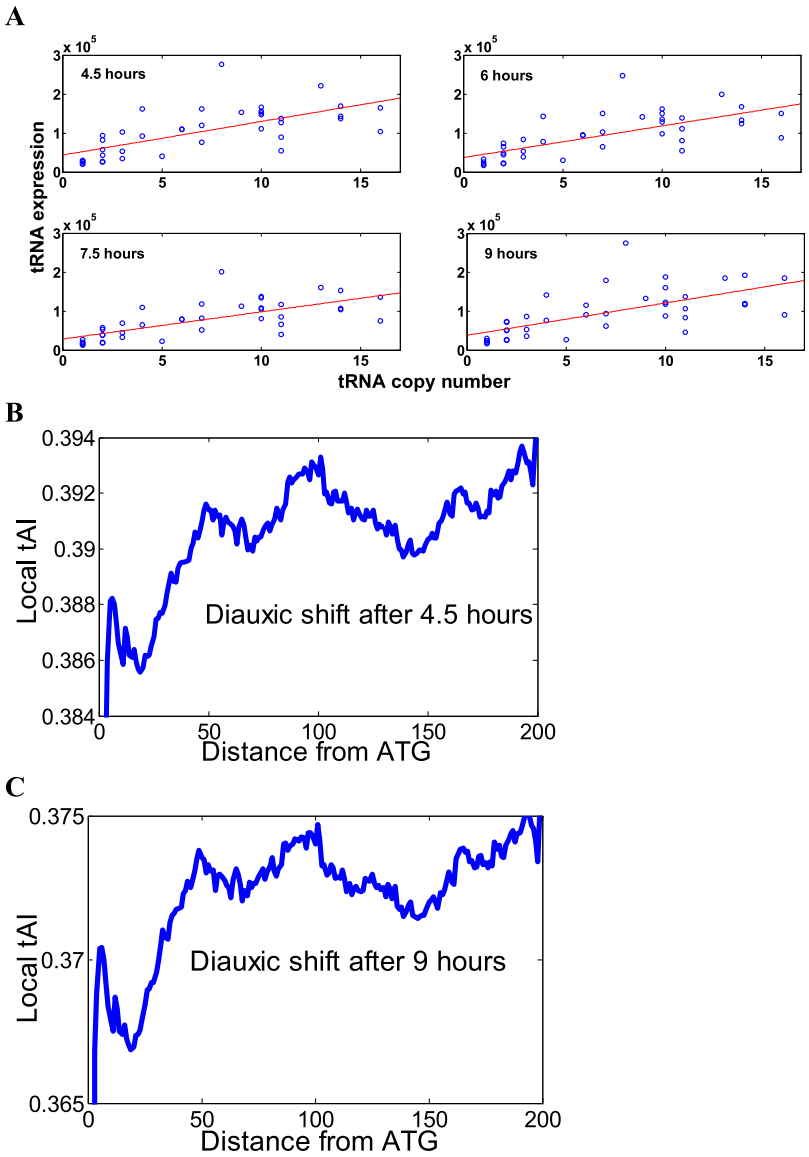


Figure S1. The tRNA Pool and the Ramp in Diauxic Shift, Related to Figure 1

(A) tRNA levels versus copy number in different points of the diauxic shift experiment.

(B and C) Translation efficiency profile (smoothed) based on tRNA expression levels in diauxic shift after 4.5 hr (B) and after 9 hr (C); the other time points have similar profiles. The length of the ramp in different points of the experiment are: point 0 hr: 23 codons; point 4.5 hr: 28 codons; point 6 hr: 30 codons; point 7.5 hr: 30 codons; point 9 hr: 30 codons.

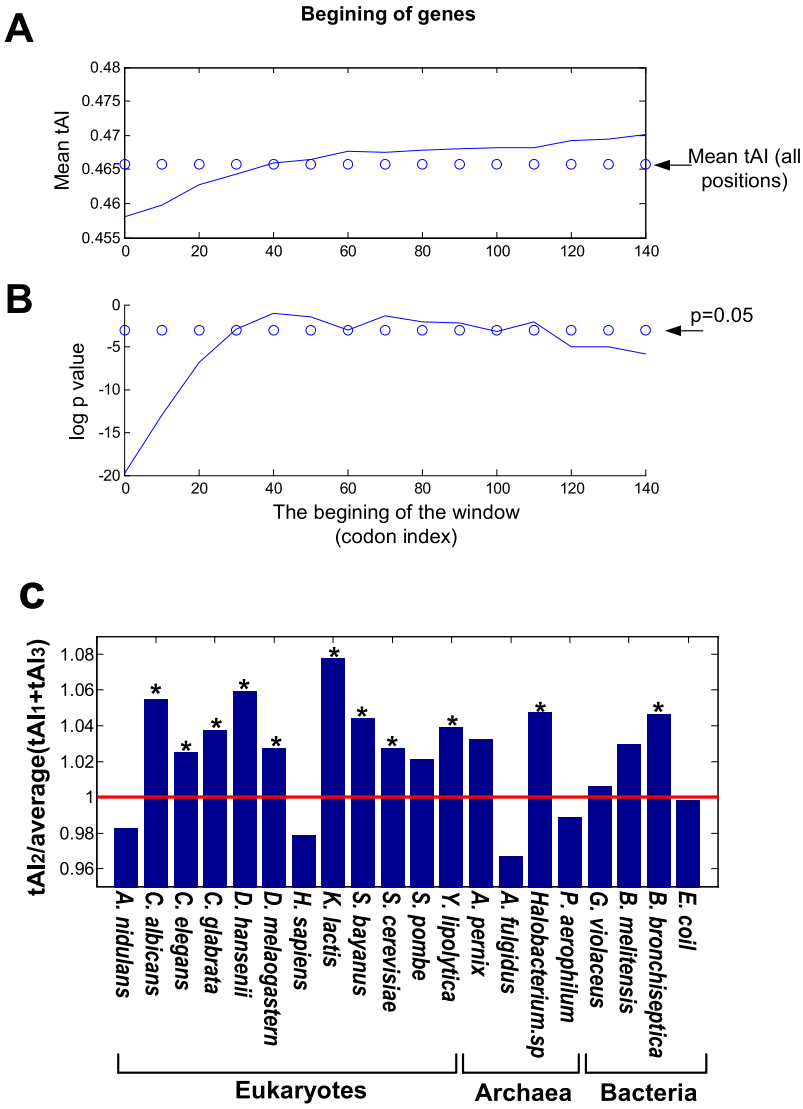


Figure S2. The Ramp Properties, Related to Figure 2

(A and B) Computing the length of the ramp for *S. cerevisiae*. The length of the ramp for a local profile of tAI was computed by comparing the mean (KS-test) of sliding windows of length 20 to the mean of the profile of local tAI from codon 1 to codon 200. The region at the beginning corresponding to windows whose mean is significantly lower than the mean of the entire profile was defined as the length of the ramp. (A) Mean tAI of sliding windows compared to the mean tAI of all positions (dotted). (B) p value for each sliding window compared to the a cut-off p value = 0.05 (dotted).

(C) The local tAI of the second codon from the ATG codon divided by the mean tAI of the first and the third codons for various organisms. Organisms where this ratio is significantly high are marked with an asterisk (empirical p value < 0.05/20; by comparison to the distribution of the ratios between the tAI of codon i and the mean tAI of codons i-1 and i+1 over all the codons along the translation efficiency profile; the p value is the fraction of positions with lower/higher ratio). In most of the organisms (14 out of 20) this ratio is larger than one (the red line), in 11 organisms this ratio was significantly high; the ratio was not significantly low in any of the analyzed organisms.

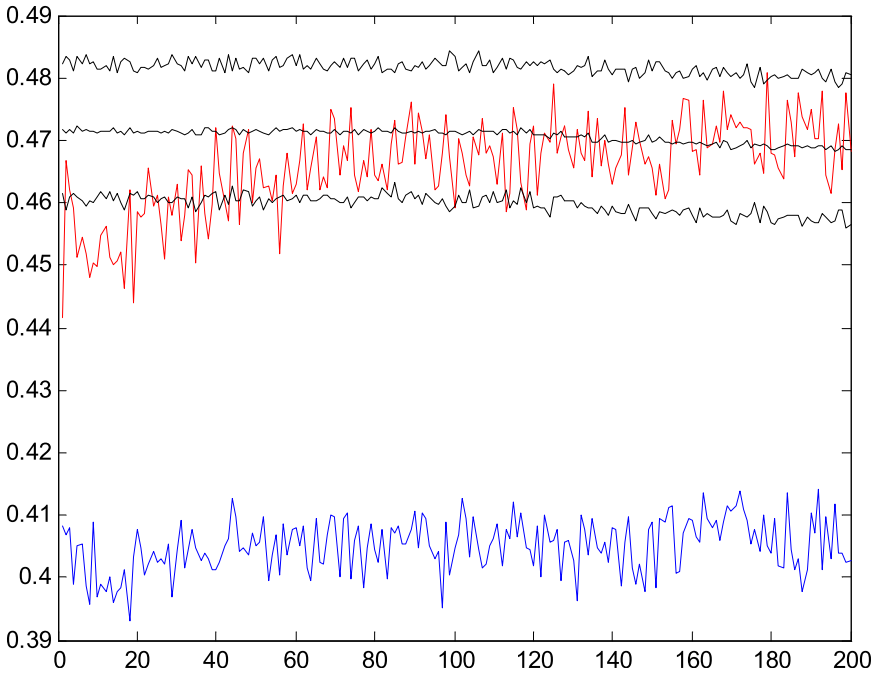


Figure S3. Controlling for Gradients of GC Content, Related to Figure 5

The GC content cannot explain the profile of local tAI in *S. cerevisae.* In order to randomize each sequence but retain the local GC content at each codon, the codons were divided into 10 groups, according to the number of times G or C appear in them. For each sequence, every codon was replaced by a randomly chosen codon from the same group of the original codon. The local tAI profile and the averaged genome profile were then calculated as described in the Experimental Procedures. The figure contains the local tAI profile (red), the randomized (permutated) profile ±3 standard deviations (black) and the averaged profile of the GC-content-preserving randomized genome (blue). The profile clearly shows that keeping the GC content of a gene is not enough to generate the local tAI increasing profile.

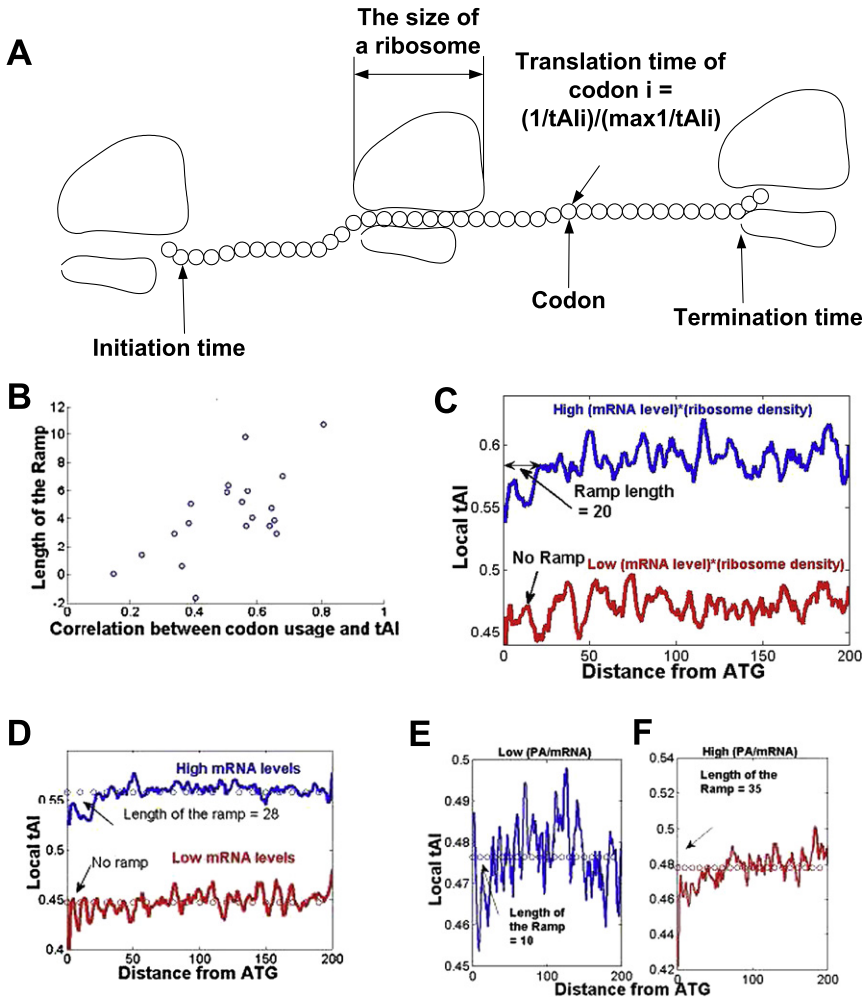


Figure S4. Relation between the Length of the Ramp and Codon Usage, Ribosome Density, mRNA Levels, and (Protein Abundance)/mRNA, Related to Figure 6

(A) The model of Zhang et al. for simulating ribosome movement.

(B) The length of the ramp versus correlation between codon usage (the relative frequency of the codon in the coding sequences) and codon tAI.

(C) Genes with high and low ribosome density*mRNA levels.

(D) Genes with high and low mRNA levels.

(E and F) Genes with high (E) and low (F) (protein abundance)/mRNA levels.