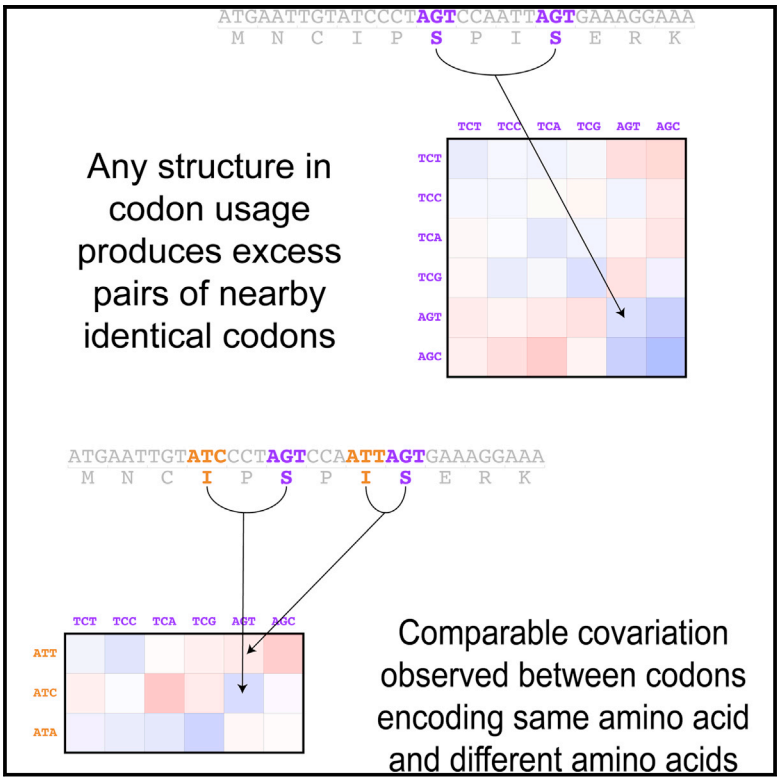
Report

Cell Reports

Local Correlations in Codon Preferences Do Not Support a Model of tRNA Recycling

Graphical Abstract



Highlights

Synonymous codon usage patterns are known to vary across the genome and within genes

Mathematically, this variation implies a diagonal-positive local covariance signal

That signal is thus not evidence for molecular tRNA reuse by the ribosome

Rather, it reflects a complicated covariance structure across 61 codons

Authors

Jeffrey A. Hussmann, William H. Press

Correspondence

jah@ices.utexas.edu

In Brief

Hussmann and Press find that observed excesses in pairs of coding sequence occurrences of the same synonymous codon are a generic consequence of the existence of spatial variation in codon preferences and therefore cannot be interpreted as evidence for the recycling of individual tRNA molecules during translation.

Hussmann & Press, 2014, Cell Reports 8, 1624–1629
September 25, 2014
http://dx.doi.org/10.1016/j.celrep.2014.08.012

OPEN ACCESS
CellPress

Cell Reports
Report

Local Correlations in Codon Preferences Do Not Support a Model of tRNA Recycling

Jeffrey A. Hussmann[1,2,*] and William H. Press[1,2]
[1]Institute for Computational Engineering and Sciences, University of Texas, Austin, TX 78712, USA
[2]Institute for Cellular and Molecular Biology, University of Texas, Austin, TX 78712, USA
*Correspondence: jah@ices.utexas.edu
http://dx.doi.org/10.1016/j.celrep.2014.08.012

SUMMARY

It has been proposed that patterns in the usage of synonymous codons provide evidence that individual tRNA molecules are recycled through the ribosome, translating several occurrences of the same amino acid before diffusing away. The claimed evidence is based on counting the frequency with which pairs of synonymous codons are used at nearby occurrences of the same amino acid, as compared to the frequency expected if each codon were chosen independently from a single genome-wide distribution. We show that such statistics simply measure variation in codon preferences across a genome. As a negative control on the potential contribution of pressure to exploit tRNA recycling on these signals, we examine correlations in the usage of codons that encode different amino acids. We find that these controls are statistically as strong as the claimed evidence and conclude that there is no informatic evidence that tRNA recycling is a force shaping codon usage.

INTRODUCTION

Due to degeneracies in the genetic code, sets of synonymous codons are translated into the same amino acid. Despite the fact that substitutions between synonymous codons in a coding sequence do not change the amino acid sequence of the translated protein, synonymous codons are not used with equal frequencies in the genomes of many organisms (Andersson and Kurland, 1990; Sharp and Li, 1987). The extent and directions of codon usage biases vary between organisms, between genes within an organism's genome, and within genes (Plotkin and Kudla, 2011). Many theories have been advanced that invoke the mechanics of the complex chain of processes that lead from packaged DNA to translated protein to explain the observed trends, including, but not limited to, mutational bias (Bulmer, 1991), bias in repair or heteroduplex mismatch resolution mechanisms (Duret, 2002), selection for enhanced translational elongation speed or translational accuracy via the coupling of codon usage frequencies to tRNA abundance differentials (Lavner and Kotlar, 2005; Hershberg and Petrov, 2008; Drummond and Wilke, 2008; Zhang et al., 2010), selection to enhance mRNA stability (Katz and Burge, 2003) or to minimize mRNA secondary structure in the neighborhood of binding sites for the translation initiation complex (Kudla et al., 2009; Gu et al., 2010), and selection to maintain control over splicing (Pagani et al., 2005; Chamary et al., 2006; Parmley and Hurst, 2007). The relative importance of these mechanisms in shaping the structure of codon usage biases remains poorly understood.

Just as existing biological knowledge can be used to make sense of patterns in codon usage, the detection of patterns in codon usage across and between genomes can be used to make inferences about biological processes. In a recent paper, Cannarozzi et al. (2010) make such an inference about the dynamics of translation. They examine all coding sequences of the genomes of several organisms and measure several related statistics, which are based on counting the frequency with which a given pair of codons is used to encode pairs of occurrences of the same amino acid that are located close to each other in a coding sequence. They observe that the same codon is used for two nearby occurrences more often than would be expected if every codon choice was drawn independently from a single genome-wide distribution. Furthermore, they observe that nearby pairs consisting of two distinct codons that occur more often than expected tend to be codons that are translated by the same isoaccepting tRNA species. They interpret these results as evidence for the intriguing hypothesis that consecutive codon choices are not made independently but instead experience selective pressure to use codons from the same isoaccepting class. They speculate that such reuse allows a single tRNA molecule to translate multiple codons before diffusing away from the ribosome, perhaps via a physical association between the ribosome and aminoacyl-tRNA synthetases. When making an inference such as this, care must be taken to disentangle other potential sources of the observed supporting evidence. It is important to determine whether the statistical evidence presented by Cannarozzi et al. offers specific support for their proposed tRNA recycling hypothesis over other previously established mechanisms influencing codon usage.

RESULTS

Positive Diagonal Entries Are a Generic Indicator of Nonuniform Codon Preferences

The main line of Cannarozzi et al.'s informatic evidence consists of a set of statistics that we will call the local covariances

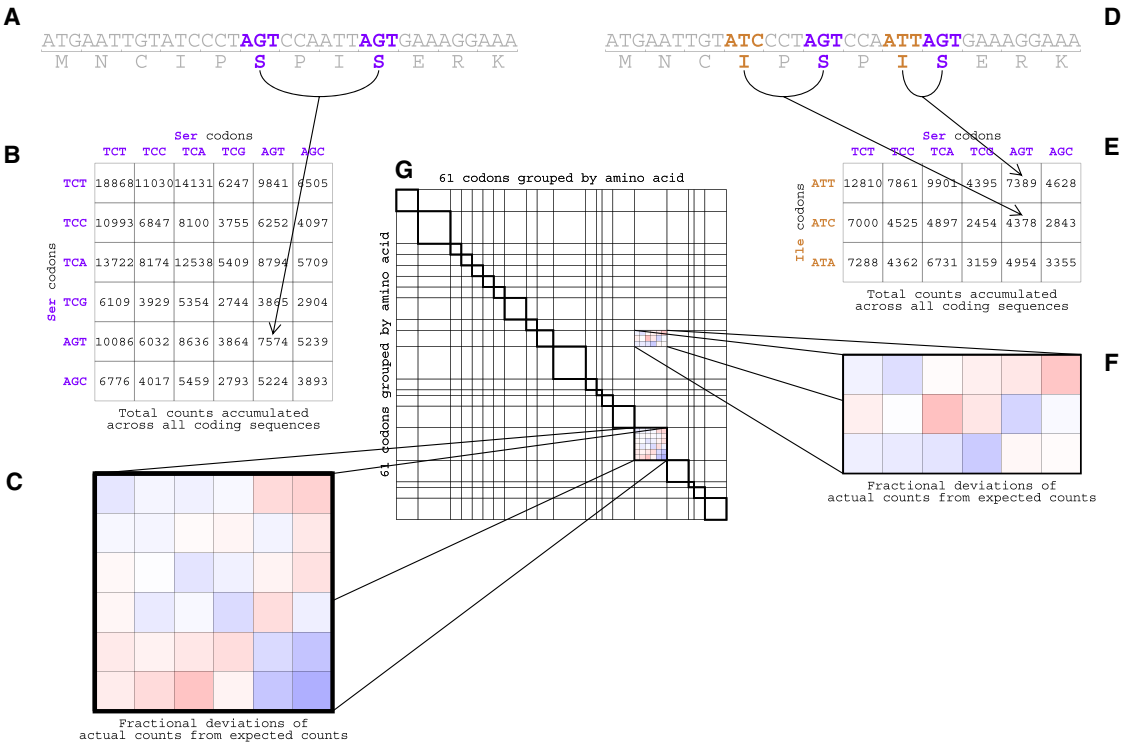


Figure 1. Arbitrary Codon Pairs Exhibit Comparable Local Covariance in Usage to Same-Amino-Acid Pairs

(A–C) Case 1: for codons encoding the same amino acids (the cases considered by Cannarozzi et al.), all sequential pairs of occurrences of an amino acid (in this case, serine) are identified (A) and the pairs of codons used to encode each pair of occurrences are counted (B). The total counts recorded over all coding sequences are then compared to the counts expected under a null model (see text for discussion of null model choice) to produce fractional deviations of actual counts from expected counts (C).

(D–G) Case 2: for codons encoding different amino acids, all ordered sequential pairs of occurrences of an ordered pair of amino acids (in this case, isoleucine and serine) are identified (D) and the pairs of codons used to encode each pair of occurrences are counted (E). Fractional deviations of actual counts from expected counts are produced as in the previous case (F). Collectively, the first case makes up the block diagonal of Figure 3, and the second case makes up the block off-diagonal portion of Figure 3 (G).

Deviations of comparable size are seen in (C) and (F).

in codon preference relative to genome-wide preferences. To compute these statistics, an ordered pair of codons translating the same amino acid is selected. The locations of all occurrences of the amino acid in all coding sequences of a genome are extracted, and the number of times that a sequential pair of occurrences is encoded by the pair of codons of interest is counted (Figures 1A and 1B). The count recorded is then compared to the number expected under a null model in which the codon used at each occurrence of the amino acid is an independent draw from a genome-wide codon preference distribution for the amino acid, estimated by the genome-wide frequencies with which each codon is used (Figure 1C). An amino acid encoded by d synonymous codons has d^2 possible ordered pairs of codons and therefore produces d^2 of these statistics, which can be naturally arranged in a $d \times d$ matrix. Terms on the diagonal of the matrix correspond to pairs consisting of repeated uses of the same codon, whereas terms off of the diagonal correspond to pairs consisting of two distinct codons. Cannarozzi et al. compute this set of statistics for several amino acids in *Saccharomyces cerevisiae* and find that diagonal terms are universally positive, corresponding to more occurrences of pairs of the same codon than expected under the null model. They interpret this observation as evidence that successive codon choices are not made independently but instead preferentially reuse the same codon.

The set of statistics considered do not provide specific support for this interpretation. The statistics are unable to distinguish between a model of codon usage in which the choices of codon used at consecutive occurrences of an amino acid are not independent and a model in which consecutive choices are independent but drawn from distributions whose parameters vary across the genome with any spatial structure at scales longer than the distance between amino acid occurrences but shorter than the entire genome.

To see this, consider an arbitrary amino acid translated by d synonymous codons and pick one of these codons. Let p_{local} be the location-specific probability with which the codon is used. Suppose that p_{local} varies as a function of location in the genome on scales longer than the typical distance between occurrences of the amino acid, so that the values of p_{local} at the two locations which make up a

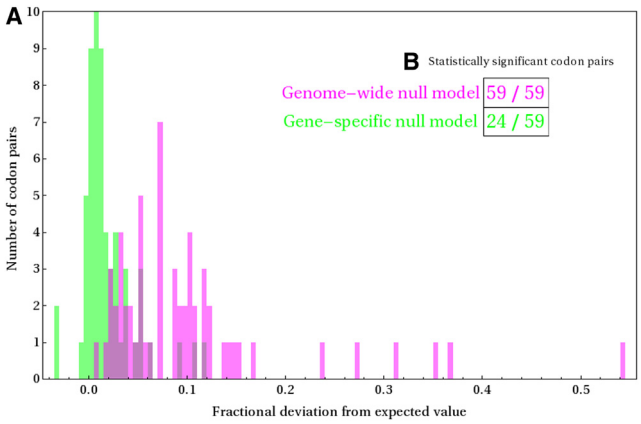


Figure 2. Most of the Apparent Signal in Codon Pair Usage Is due to Gene-Specific Codon Preferences

(A) Fractional deviations of actual counts of concordant codon pairs (diagonal entries in matrices) from expected counts under null models of (1) a single genome-wide codon preference distribution (magenta) or (2) gene-by-gene distributions (green). Most of the strength of the signal present relative to a genome-wide model disappears relative to a gene-by-gene model.

(B) Fractions of concordant codon pairs with statistically significant deviations from expected counts under null models at Benjamini-Hochberg false discovery rate of $\alpha = 0.05$.

p values for each codon pair to input to the Benjamini-Hochberg prescription were computed as (1) the fraction of 10,000 shuffles of codon assignments to amino acids within the entire genome (magenta) or (2) fraction of 10,000 shuffles of codon assignments to amino acids within each gene (green) for which the shuffled term was more extreme relative to the appropriate expected value than that of the actual data. Much of the statistical significance present relative to a genome-wide model disappears relative to a gene-by-gene model.

sequential pair of amino acid occurrences can be viewed as approximately equal. Let n_{genome} be the number of sequential pairs of occurrences of the amino acid in the genome, and let E_{genome} denote taking the expected value across all such pairs. Then (see the Supplemental Experimental Procedures), a null model of independent draws from a genome-wide codon preference distribution predicts that $n_{genome} E_{genome}[p_{local}]^2$ pairs of the codon will be observed, whereas the actual number expected is given by $n_{genome} E_{genome}[p_{local}^2]$. The statistic of interest, the deviation of observed counts from the genome-wide null model prediction, therefore has the expected value

$$n_{genome}\left[E_{genome}\left[p_{local}^2\right] - E_{genome}[p_{local}]^2\right]$$

and has a clear interpretation as a measure of the variance across the genome in the local independent probability with which the codon is used. In particular, the fact that this expression consists of the difference between the expected value of the square of a function and the square of the expected value of the function means that it is guaranteed (by Jensen's inequality) to be positive if p_{local} is not simply constant across the genome. Intuitively, the application of Jensen's inequality tells us that, whereas variation in p_{local} leads to the accumulation of excess consecutive pairs of the codon in regions where p_{local} is higher than its genome-wide average and the depletion of pairs in regions where p_{local} is lower than its average, the strict convexity of the function of p_{local} under consideration (namely, squaring) guarantees that the gains will always more than offset the losses. The fact that universally positive values of the statistics are observed on the diagonals of matrices is now seen to be unremarkable. It is expected under any model of codon usage in which codon preferences are not uniform across a genome. The two other types of statistics Cannarozzi et al. consider—the tRNA pairing index and tRNA correlation as a function of distance between amino acid occurrences—are variations on this theme and suffer from the same lack of specificity for essentially identical reasons (see the Supplemental Experimental Procedures).

Of course, codon preferences are not uniform across genomes. In particular, the existence of gene-specific codon preferences is a well-studied and well-accepted (if not completely well-understood) phenomenon (Sharp et al., 1995; Wright, 1990). Cannarozzi et al. correctly identify the need to control for gene-specific codon preferences and correctly identify that shuffling the assignments of codon choices to amino acid occurrences within each gene provides a way to do this. The striking feature of their controls, which compare their statistics computed on real data to statistics computed on a single shuffle of the data, however, is not that some signal survives the shuffle but that most of the signal does not. To directly quantify the extent to which the observed signals are explained by gene-specific codon preferences, we computed the expected numbers of pairs of sequential occurrences of each amino acid encoded by each pair of codons under such a gene-by-gene shuffle (see the Experimental Procedures). Fractional deviations of the data relative to this gene-specific null model are dramatically less extreme (Figure 2A) and less uniformly statistically significant (Figure 2B) than deviations over a genome-wide model. Having presented this control, it should be noted that Cannarozzi et al.'s argument that "if the correlation effect was simply due to the accumulation of frequent codons in genes with biased codon composition, this effect should also be highest for frequent codons and not observed for rare codon" misstates the effect that local bias in codon composition has on correlation effects. The effect will be highest for codons whose location-specific frequency exhibits the most variation around its average frequency in the genome, not those whose average frequency is highest.

Signal that Survives Gene-by-Gene Shuffling Is Also Nonspecific

The existence of statistically significant (but substantially reduced) residual positive diagonal values after replacing Cannarozzi et al.'s genome-wide null model with a gene-specific null model is no more specific evidence for tRNA recycling than the original signal was. By repeating the same argument as above with the phrase "gene-specific" substituted for "genome-wide," the expected value of the modified statistic is (see the Supplemental Experimental Procedures)

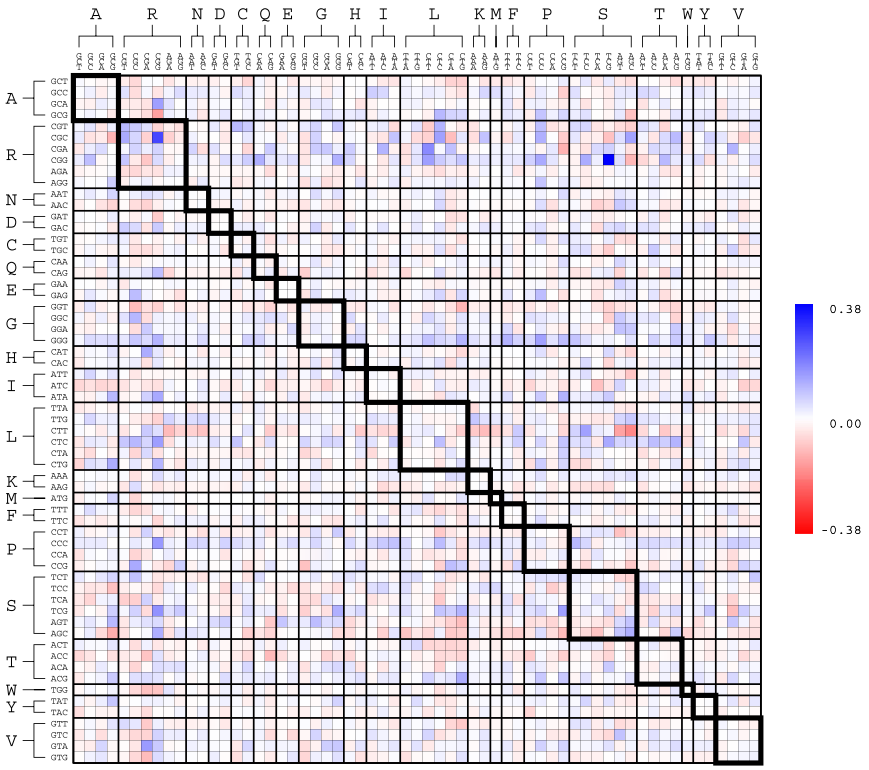


Figure 3. Complete Data for the Framework Shown in Figure 1G, Generated According to the Process Outlined in Figure 1

Fractional deviations of counts of actual usage of codon pairs in all coding sequences of *S. cerevisiae* with respect to counts expected under a shuffling of assignments of codons to amino acids within each gene. The thickly bordered diagonal blocks contain those pairs of codons that encode the same amino acid. These diagonal blocks are not a visually distinct subset of the full matrix. See Figure 4 for quantitative comparison.

$$\sum_{genes} n_{gene}\left(\mathrm{E}_{gene}\left[p_{local}^{2}\right]-\mathrm{E}_{gene}\left[p_{local}\right]^{2}\right),$$

where n_{gene} is the number of pairs of occurrences in a given gene, and positive values of the modified statistic are generic evidence for the existence of structure in codon preferences at scales larger than the distance between occurrences but smaller than genes. The existence of intragenic codon preference structure in many organisms is well established (Chen and Inouye, 1990; Eyre-Walker and Bulmer, 1993; Qin et al., 2004), and several models of sources for such structure have been proposed (Tuller et al., 2010; Gu et al., 2010).

A simple observation allows us to assess the amount of intragenic structure in codon preference that is due to sources that are not tRNA recycling. The set of statistics considered by Cannarozzi et al. can be extended in a natural way to consider pairs of codons encoding distinct amino acids. To construct this generalized set of statistics, label the 61 nonstop codons and select an arbitrary ordered pair (i,j). Let a_i be the amino acid translated by the first codon and a_j be the amino acid translated by the second codon. In each coding sequence in the genome, identify every sequential pair of occurrences of a_i and a_j (that is, a pair such that the occurrence of a_i is before that of a_j and there are no other occurrences of either amino acid in between the two; Figure 1D). Record the number of such pairs that are encoded by codons i and j (Figure 1E). The count produced can then be compared to the number expected under gene-specific shuffling of codon assignments to amino acid occurrences (see the Experimental Procedures; Figure 1F). The resulting statistics for all possible pairs of codons can be naturally arranged in a 61×61 matrix (Figure 1G). If codons are grouped according to the amino acid they translate, the original subset of codon pairs considered by Cannarozzi et al. (the special cases for which $a_i = a_j$) occupies blocks on the diagonal. Following a similar line of reasoning as above, if $p_{local}^{(i)}$ and $p_{local}^{(j)}$ are the local probabilities with which codons i and j, respectively, are used as a function of location in the genome, the (i,j) th entry in this matrix has an expected value approximately equal to

$$\sum_{genes} n_{gene}^{(a_i,a_j)}\left(\mathrm{E}_{gene}\left[p_{local}^{(i)}p_{local}^{(j)}\right]-\mathrm{E}_{gene}\left[p_{local}^{(i)}\right]\mathrm{E}_{gene}\left[p_{local}^{(j)}\right]\right),$$

where $n_{gene}^{(a_i,a_j)}$ is the number of sequential pairs of occurrences of a_i and a_j in a given gene. The motivation for calling this set of statistics the local covariance in codon preference is now apparent.

Significantly nonzero values for pairs of codons that do not encode the same amino acid cannot be caused by pressure for tRNA recycling because such pairs are neither translated by the same isoaccepting tRNA species nor forced to offset a potentially disproportionate share of expected counts taken up by a pair that is. Such values are, however, easily explained by models of location-specific variation in codon preferences. Intuitively, positive values of such an off-diagonal term indicate that regions in which codon i is used more often than its gene-wide frequency tend to overlap with regions in which codon j is used more often than its gene-wide frequency. Of course, this argument is unchanged if codons i and j are distinct but encode the same amino acid. As Cannarozzi et al. observe, in this case, positive values tend to be (i,j) pairs that are translated by the same tRNA species, an observation that survives the switch to a gene-specific null model. Whereas this signature could be caused by tRNA recycling, it could also simply indicate that local codon preferences are coupled, by selection, to the identities of tRNA species. For example, translation may be locally slowed down in portions of genes to prevent ribosomal "traffic jams" (Tuller et al., 2010) or to allow time for cotranslational folding of the

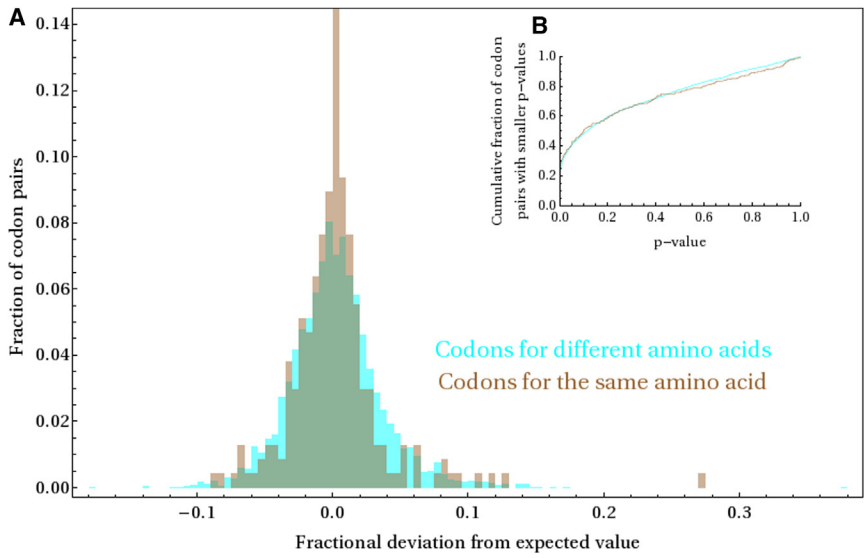


Figure 4. Comparison of Signals Observed for Codon Pairs Encoding the Same Amino Acid and Codon Pairs Encoding Different Amino Acids

(A) Distributions of fractional deviations shown in Figure 3 for (1) terms inside of the block diagonal, representing pairs of codons encoding the same amino acid (brown), and (2) terms outside of the block diagonal, representing pairs of codons encoding different amino acids (cyan). The distributions of signal strengths for these two classes of codon pairs are strikingly similar.

(B) Empirical cumulative distribution functions of p values for the fractional deviations in Figure 3 for (1) terms inside of the block diagonal (brown) and (2) terms outside of the block diagonal (cyan). p values for each term were computed as the fraction of 10,000 shuffles of codon assignments to amino acids within each gene for which the shuffled term was more extreme relative to the expected value than that of the actual data.

nascent polypeptide (Komar, 2009) through the preferential usage of codons translated by scarce tRNA species, or regions of genes prone to misfolding may experience pressure to use more accurate tRNAs (Drummond and Wilke, 2008). Such mechanisms would create positive covariances in location-specific preferences for codons translated by a given tRNA.

We now establish the plausibility of the second interpretation. Examining the strength and significance of local covariances between pairs of codons translating distinct amino acids, which can be caused by local independent codon preference variation but not by tRNA recycling, and comparing these to pairs translating the same amino acid allows us to determine if tRNA recycling is plausibly a major influence on codon usage. The 61×61 matrix of fractional deviations for all codon pairs in *S. cerevisiae* shows widespread structure (Figure 3). In particular, the block-diagonal segment corresponding to pairs encoding the same amino acid is not a visually or statistically distinct subset of the entire matrix. The distributions of fractional deviations and statistical significances corresponding to terms inside of the block-diagonal subset and those corresponding to terms outside of the block-diagonal subset are strikingly qualitatively similar (Figure 4A). Comparable fractions of terms from each class are indisputably statistically significant (Figures 4B and S1). The largest positive and negative values for pairs encoding distinct amino acids are as extreme as those for pairs of distinct codons encoding the same amino acid. Taken together, these observations suggest that values in the diagonal blocks can be explained entirely by local preference structure induced by non-tRNA-recycling mechanisms and therefore cannot be taken as specific evidence that tRNA recycling is a major force shaping codon choices.

EXPERIMENTAL PROCEDURES

Databases

Yeast coding sequences were retrieved from release 60 of the Ensembl databases using the Ensembl Perl API (Flicek et al., 2011).

Derivation of Expected Number of Pairs under Gene-Specific Shuffling

Pick an arbitrary amino acid that is translated by d codons. Let N be the total number of genes in a genome, n_g be the number of occurrences of the amino acid in gene g, and $c_{g,i}$ be the number of occurrences of codon i in gene g. What is the expected number of consecutive occurrences of the pair of choices (i, j) in a synthetic assignment of codon choices to occurrences produced by randomly shuffling the actual set of codon choices within each gene? First, note that, for any gene g and pair of amino acid occurrences k and $k+1$,

$$P_{shuffle}[\textit{occurrence } k \textit{ is codon } i \textit{ and occurrence } k+1 \textit{ is codon } j] = \begin{cases} \frac{c_{g,i}}{n_g}\,\frac{c_{g,j}-1}{n_g-1}, & i=j \\ \frac{c_{g,i}}{n_g}\,\frac{c_{g,j}}{n_g-1}, & i\neq j \end{cases}.$$

Let $1_{g,k}^{(i,j)}$ be 1 if the k th pair of consecutive occurrences of the amino acid in gene g (that is, occurrences k and $k+1$) consists of codon choices i and j and 0 otherwise.

Then,

$$\begin{aligned} \mathrm{E}_{shuffle}[\textit{number of pairs } i,j] &= \mathrm{E}_{shuffle}\left[\sum_{g=1}^{N}\sum_{k=1}^{n_g-1} 1_{g,k}^{(i,j)}\right] \\ &= \sum_{g=1}^{N}\sum_{k=1}^{n_g-1} \mathrm{E}_{shuffle}\left[1_{g,k}^{(i,j)}\right] \\ &= \begin{cases} \sum_{g=1}^{N}\sum_{k=1}^{n_g-1} \frac{c_{g,i}}{n_g}\,\frac{c_{g,j}-1}{n_g-1}, & i=j \\ \sum_{g=1}^{N}\sum_{k=1}^{n_g-1} \frac{c_{g,i}}{n_g}\,\frac{c_{g,j}}{n_g-1}, & i\neq j \end{cases} \\ &= \begin{cases} \sum_{g=1}^{N} \frac{c_{g,i}\left(c_{g,j}-1\right)}{n_g}, & i=j \\ \sum_{g=1}^{N} \frac{c_{g,i}c_{g,j}}{n_g}, & i\neq j \end{cases}. \end{aligned}$$

For the generalized set of statistics, consider codons i and j encoding amino acids a_i and a_j with $a_i \neq a_j$. Let $n_g^{(a_i)}$ be the number of occurrences of a_i, $n_g^{(a_j)}$ be the number of occurrences of a_j, $n_g^{(a_i,a_j)}$ be the number of pairs of occurrences of a_i followed at some distance by a_j such that there is no other occurrence

of a_i or a_j between the two in gene g, and $1_{g,k}^{(i,j)}$ be 1 if the k th such pair in gene g consists of codons i and j and 0 otherwise.

Then,

$$\mathrm{E}_{shuffle}[number\ of\ pairs\ i,j] = \mathrm{E}_{shuffle}\left[\sum_{g=1}^{N}\sum_{k=1}^{n_g^{(a_i,a_j)}-1} 1_{g,k}^{(i,j)}\right]$$

$$= \sum_{g=1}^{N}\sum_{k=1}^{n_g^{(a_i,a_j)}-1} \mathrm{E}_{shuffle}\left[1_{g,k}^{(i,j)}\right]$$

$$= \sum_{g=1}^{N} n_g^{(a_i,a_j)} \frac{c_{gj}}{n_g^{(a_i)}} \frac{c_{gj}}{n_g^{(a_j)}}.$$

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures and one figure and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2014.08.012.

ACKNOWLEDGMENTS

We thank A.D. Ellington, E.M. Marcotte, Joshua B. Plotkin, and Claus O. Wilke for helpful discussions as well as the anonymous referees for helpful suggestions.

Received: July 7, 2014
Revised: July 28, 2014
Accepted: August 6, 2014
Published: September 4, 2014

REFERENCES

Andersson, S.G.E., and Kurland, C.G. (1990). Codon preferences in free-living microorganisms. Microbiol. Rev. *54*, 198–210.

Bulmer, M. (1991). The selection-mutation-drift theory of synonymous codon usage. Genetics *129*, 897–907.

Cannarozzi, G., Schraudolph, N.N., Faty, M., von Rohr, P., Friberg, M.T., Roth, A.C., Gonnet, P., Gonnet, G., and Barral, Y. (2010). A role for codon order in translation dynamics. Cell *141*, 355–367.

Chamary, J.V., Parmley, J.L., and Hurst, L.D. (2006). Hearing silence: non-neutral evolution at synonymous sites in mammals. Nat. Rev. Genet. *7*, 98–108.

Chen, G.-F.T., and Inouye, M. (1990). Suppression of the negative effect of minor arginine codons on gene expression; preferential usage of minor codons within the first 25 codons of the Escherichia coli genes. Nucleic Acids Res. *18*, 1465–1473.

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

Duret, L. (2002). Evolution of synonymous codon usage in metazoans. Curr. Opin. Genet. Dev. *12*, 640–649.

Eyre-Walker, A., and Bulmer, M. (1993). Reduced synonymous substitution rate at the start of enterobacterial genes. Nucleic Acids Res. *21*, 4599–4603.

Flicek, P., Amode, M.R., Barrell, D., Beal, K., Brent, S., Chen, Y., Clapham, P., Coates, G., Fairley, S., Fitzgerald, S., et al. (2011). Ensembl 2011. Nucleic Acids Res. *39* (Database issue), D800–D806.

Gu, W., Zhou, T., and Wilke, C.O. (2010). A universal trend of reduced mRNA stability near the translation-initiation site in prokaryotes and eukaryotes. PLoS Comput. Biol. *6*, e1000664.

Hershberg, R., and Petrov, D.A. (2008). Selection on codon bias. Annu. Rev. Genet. *42*, 287–299.

Katz, L., and Burge, C.B. (2003). Widespread selection for local RNA secondary structure in coding regions of bacterial genes. Genome Res. *13*, 2042–2051.

Komar, A.A. (2009). A pause for thought along the co-translational folding pathway. Trends Biochem. Sci. *34*, 16–24.

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.

Lavner, Y., and Kotlar, D. (2005). Codon bias as a factor in regulating expression via translation rate in the human genome. Gene *345*, 127–138.

Pagani, F., Raponi, M., and Baralle, F.E. (2005). Synonymous mutations in CFTR exon 12 affect splicing and are not neutral in evolution. Proc. Natl. Acad. Sci. USA *102*, 6368–6372.

Parmley, J.L., and Hurst, L.D. (2007). How do synonymous mutations affect fitness? Bioessays *29*, 515–519.

Plotkin, J.B., and Kudla, G. (2011). Synonymous but not the same: the causes and consequences of codon bias. Nat. Rev. Genet. *12*, 32–42.

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.

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.

Sharp, P.M., Averof, M., Lloyd, A.T., Matassi, G., and Peden, J.F. (1995). DNA sequence evolution: the sounds of silence. Philos. Trans. R. Soc. Lond. B Biol. Sci. *349*, 241–247.

Tuller, T., Carmi, A., Vestsigian, K., Navon, S., Dorfan, Y., Zaborske, J., Pan, T., Dahan, O., Furman, I., and Pilpel, Y. (2010). An evolutionarily conserved mechanism for controlling the efficiency of protein translation. Cell *141*, 344–354.

Wright, F. (1990). The 'effective number of codons' used in a gene. Gene *87*, 23–29.

Zhang, F., Saha, S., Shabalina, S.A., and Kashina, A. (2010). Differential arginylation of actin isoforms is regulated by coding sequence-dependent degradation. Science *329*, 1534–1537.