Published in final edited form as:

Nat Ecol Evol. 2020 April ; 4(4): 589–600. doi:10.1038/s41559-020-1124-7.

Dissimilation of synonymous codon usage bias in virus-host coevolution due to translational selection

Feng Chen^{1,2,3}, Peng Wu², Shuyun Deng⁴, Heng Zhang⁵, Yutong Hou², Zheng Hu⁶, Jianzhi Zhang^{7,*}, Xiaoshu Chen^{4,8,*}, Jian-Rong Yang^{1,2,3,7,8,9,*}

¹Program in Cancer Research, Zhongshan School of Medicine, The Fifth Affiliated Hospital, Sun Yat-sen University, Guangzhou 510080, China

²Department of Biomedical Informatics, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou 510080, China

³Key Laboratory of Tropical Disease Control, Ministry of Education, Sun Yat-sen University, Guangzhou 510080, China

⁴Department of Medical Genetics, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou 510080, China

⁵Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou 510080, China

⁶Department of Obstetrics and Gynecology, Precision Medicine Institute, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou 510080, China

⁷Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109, USA

⁸RNA Biomedical Institute, Sun Yat-Sen Memorial Hospital, Sun Yat-sen University, Guangzhou 510120, China

⁹Guangdong Provincial Key Laboratory of Biomedical Imaging, The Fifth Affiliated Hospital, Sun Yat-sen University, Zhuhai, Guangdong Province 519000, China.

Abstract

*Corresponding authors. yangjianrong@mail.sysu.edu.cn (Yang JR), chenxshu3@mail.sysu.edu.cn (Chen X), jianzhi@umich.edu (Zhang J).

Author Contributions

J.-R.Y., X.C., J.Z. conceived the idea, designed and supervised the study. F.C., P.W., S.D., H.Z. and Y.H. conducted experiments and acquired data, S.D., Z.H., X.C. and J.-R.Y. contributed new reagents/analytic tools. F.C., P.W., S.D., H.Z., J.-R.Y. analyzed data. F.C., J.Z., X.C. and J.-R.Y. wrote the paper.

Data availability

For the yeast Ribo-Seq data underlying Fig. 1, all accession numbers for publicly available datasets were listed in Supplementary Table 1. For the human Ribo-Seq data underlying Fig. 2, the original dataset is available from NCBI SRA under accession number SRR3623932 and SRR3623937. The raw data underlying Fig. 3 is shown in Supplementary Figure 6 and Supplementary Table 3. The species identified as virus or its natural/symptomatic hosts were listed in Supplementary Table 5, with their genomic sequences obtained from NCBI GenBank.

Code availability

Custom R codes were used in data analysis, which are available on Github (<https://github.com/chenfengokha/CUB>).

Competing Interests statement

The authors declare no conflict of interest.

Eighteen of the twenty amino acids are each encoded by more than one synonymous codon. Due to the differential tRNA supply in the cell, synonymous codons are not used with equal frequencies, a phenomenon termed “codon usage bias” (CUB). Previous studies have demonstrated that CUB of endogenous genes trans-regulates the translational efficiency of other genes. We hypothesized similar effects for CUB of exogenous genes on host translation and test it in the case of viral infection, a common form of naturally occurring exogenous gene translation. We analyzed public Ribo-Seq datasets from virus-infected yeast and human cells and showed that virus CUB trans-regulated tRNA availability and therefore the relative decoding time of codons. Manipulative experiments in yeast using 37 synonymous fluorescent proteins confirmed that an exogenous gene with CUB more similar to that of the host would possess decreased translational load on host per unit of expression, whereas the expression of the exogenous gene was elevated. The combination of these two effects was that exogenous genes with CUB overly similar to that of the host would severely impede host translation. Finally, using a manually curated list of viruses, natural hosts and symptomatic hosts, we found that virus CUB tended to be more similar to that of symptomatic hosts than that of natural hosts, supporting a general deleterious effect of excessive CUB similarity between viruses and hosts. Our work revealed repulsion between virus and host CUBs when they are overly similar, a previously unrecognized complexity in the coevolution of virus and host.

Introduction

All amino acids except methionine and tryptophan are encoded by two or more synonymous codons. Some of the synonymous codons, called the “preferred codons”, are used more frequently than others, a phenomenon commonly termed “codon usage bias” (CUB). It is now clear that CUB is jointly determined by mutation, drift and selection^{1–6}. However, how exactly CUB affects the fitness of an organism is less clear. It has been shown that CUB cis-regulates translational efficiency^{7–10} and/or accuracy^{11–14} because the cognate tRNAs for preferred codons are more abundant than those for unpreferred codons^{15,16}. Additionally, CUB of highly expressed genes trans-regulates the overall translational efficiency in the cell because all translation in a cell shares resources, such as tRNAs and ribosomes^{17,18}. Indeed, theoretical and experimental studies have suggested that increased usage of unpreferred codons in a highly expressed endogenous gene would deplete the corresponding cognate tRNAs and subsequently stall translating ribosomes, decreasing the availability of free ribosomes in the cell and, thus, indirectly reducing the translational efficiency of other genes^{3,19–21}. Theoretically, the CUB of exogenous genes with sufficiently high expression levels will similarly influence host translation.

The most common type of naturally occurring translation of exogenous genes in host cells is perhaps the translation of viral proteins after viral infection. The majority of viruses have highly compact genomes that do not encode any tRNA; therefore the translation of viral proteins relies on the tRNA of their hosts^{22,23}. This circumstance creates a translational selection for the assimilation of virus CUB to host CUB. Indeed, previous studies have shown a strong resemblance of CUB between viruses and their corresponding hosts^{24,25}. However, how virus CUB impacts translation in host cells and how this type of virus-host interaction affects the evolution of viruses remain elusive.

While analyzing a published Ribo-Seq dataset from *Saccharomyces cerevisiae*, we serendipitously noticed a substantial fraction of ribosome-protected fragments that was unmappable to the reference genome. These ribosome-protected fragments were derived from viral protein-coding genes. Additional analyses suggested that the translation of viral proteins decelerated the decoding of codons that were scarcely supplied by the host and frequently used by the virus. Combined with similar results found in Ribo-Seq data from virus-infected human cells, our observations suggested that virus CUB trans-regulated host translation via differential depletion of tRNA. We used a fluorescence reporter system to further investigate how CUB of exogenous genes affected host translation and found that CUB similarity with host increased the expression of exogenous genes while reducing the translational load on the host per unit of exogenous gene expression. More importantly, the combined net effect was that exogenous genes with excessive CUB similarity to host imposed a higher translational load on the host. This result indicated that viruses with CUB too similar to that of the host would impede host translation. Finally, we examined the patterns of CUB in a manually curated list of species trios, each containing a virus, its natural host (a species with little symptoms during infection of the virus) and its symptomatic host (a species with obvious symptoms in one or more stage of the virus infection). Consistent with our theory, we found that virus CUB tended to be more similar to that of symptomatic hosts than that of natural hosts. Our work revealed repulsion between the virus and host CUB when they were overly similar, a previously unrecognized complexity in the coevolution of virus and host.

Results

Excessive viral translation differentially depletes host tRNA and decelerates translation

We analyzed four Ribo-Seq datasets from *Saccharomyces cerevisiae* generated by Gardin et al., named “Sc-Lys”, “Sc-His”, “YPD1”, and “YPD2” in their paper²⁶. While mapping the Ribo-Seq reads to the yeast genome, we noticed a substantial fraction of unmappable reads. To determine the origins of these reads, we randomly selected 10,000 unmappable reads and BLASTed them against the “nr” database at NCBI²⁷. Intriguingly, a high proportion of them hit one of two yeast double-stranded RNA viruses of the genus Totivirus, L-A and L-BC. L-A is a cytoplasmically transmitted virus associated with the yeast killer phenotype²⁸, while the phenotypic impact of L-BC on the host is unknown²⁹. In light of the numerous viral Ribo-Seq reads found in the pilot analysis, we remapped all the Ribo-Seq reads from Gardin et al.’s experiment to the yeast genome and the two viral genomes simultaneously. We then estimated the translational activity for each gene by averaging the number of Ribo-Seq reads per site for the gene (Supplementary Fig. 1). Note that this metric reflects the translational activity per gene, not per mRNA molecule, because it reflects the abundance of ribosome protected fragments without dividing by mRNA abundance, and is corrected for the gene length (see Methods). As a comparison, we analyzed five previously published yeast Ribo-Seq datasets (Supplementary Table 1). We found that the ratio between the average translational activity of the three viral genes and that of all yeast endogenous genes ranges from 1.4 to 7.9 in the four datasets of Gardin et al.²⁶ but is 0.275 to 0.930 in the other five yeast datasets (Fig. 1a and Supplementary Fig. 1). In each of Gardin et al.’s datasets, the translational activity of the most active viral gene is higher than that of 96% of the yeast

endogenous genes. In principle, such excessive translation of exogenous genes could deplete the host tRNA supply¹⁷ and hence deserves further scrutiny. In the following analysis, we concentrated on the Sc-Lys dataset, because the original research²⁶ also focused on this dataset.

To evaluate the impact of viral protein translation on the yeast tRNA supply we calculated its total consumption of each of the 61 sense codons (see Methods) based on the three viral genes and examined its relation to the corresponding tRNA supply by the host cell. Specifically for each codon, the codon consumption was the sum of codon consumption by each viral gene, which was calculated as the number of occurrence of the codon in the gene multiply by translational activity of the gene; and the tRNA supply by the host cell was approximated by tRNA adaptation index (tAI) calculated based on the genomic copy number of tRNA genes on the host genome (see Methods). The viral codon consumption appeared to be correlated with tAI (Spearman's rank correlation coefficient $\rho = 0.596$, $P < 10^{-6}$), yet this correlation was significantly weaker than that between tAI and the total codon consumption of yeast endogenous genes ($P = 0.001$, Fisher's r-to-z transformation; Fig. 1b). Furthermore, this correlation was also significantly weaker than that between tAI and the codon consumption of three randomly chosen highly expressed endogenous genes ($P = 0.035$, Fisher's r-to-z transformation; Fig. 1b), excluding the possibility that such reduction in correlation is caused by small size of the virus genome. Since the codon consumption of yeast endogenous genes should have evolved to match tRNA supply, the above result suggested a mismatch of tRNA demand by the virus and supply by the host, therefore differential tRNA shortages and corresponding changes in the decoding time of codons when the viral expression is sufficiently high (see results from our manipulative experiment below). Indeed, we found that the relative decoding time of a codon was positively correlated with its relative tRNA shortage due to viral translation (Spearman's rank correlation coefficient $\rho = 0.45$, $P < 4 \times 10^{-4}$; Fig. 1c), which was estimated for each codon by the ratio between viral consumption and tAI (see Methods). This pattern cannot be explained by the small viral genome size or interdependence between the parameters analyzed (Supplementary Fig. 2), and indicated that codons with severe tRNA shortage due to virus translation were translated more slowly than those with mild tRNA shortage. Since relative tRNA shortage due to viral translation is determined by virus CUB when virus expression level is given, the above results is thus compatible with a role of virus CUB in trans-regulating host translation through differential tRNA depletion when the virus expression level is high.

To further support the trans-regulatory role of virus CUB on host translation, we performed three additional analyses. First, we predicted that differences of codon decoding times between experiments should be correlated with the differences in relative tRNA shortage due to viral translation. This is indeed observed when we compared the Sc-Lys dataset with five other yeast Ribo-Seq experiments (Fig. 1d), which cannot be explained by the small size of virus genome (Supplementary Fig. 2). Second, we estimated for each codon the "typical decoding time", which was previously proposed as a better approximation for the effect of tRNA abundance as it explicitly excluded rare and extreme ribosomal pauses³⁰ (see Methods, Supplementary Text and Supplementary Fig. 3). We found a positive correlation between typical decoding time and relative tRNA shortage due to viral translation

(Spearman's rank correlation coefficient $\rho = 0.32$, $P = 0.013$; Fig. 1e), which again cannot be explained by the small viral genome size or interdependence between the parameters analyzed (Supplementary Fig. 4). Third, it was previously proposed that the supply of tRNA for translation cannot be approximated by the abundance of tRNA molecules since only a small fraction of all tRNA molecules are ready for translation¹⁷. We therefore recalculated relative tRNA shortage due to viral translation, with tRNA supply estimated by codon consumption of yeast endogenous genes (see Methods) instead of tAI. We found that the positive correlation between typical decoding time and relative tRNA shortage due to viral translation was retained (Spearman's rank correlation coefficient $\rho = 0.64$, $P < 10^{-8}$; Supplementary Fig. 4). Collectively, our observations of viral translation in yeast suggested a nonnegligible trans-regulatory effect of virus CUB on host translation if the expression of virus is sufficiently high.

The impact of viral translation in flu-infected human cells

To investigate whether the impact of virus CUB on host translation is generally applicable to other virus-host pairs, we downloaded another Ribo-Seq dataset generated from flu(Influenza A virus, IAV)-infected A549 cells³¹, a human lung adenocarcinoma cell line. Similar to that of the viral genes in yeast, the average translational activity of the IAV genes was ~300 folds higher than that of human endogenous genes, with the most active IAV gene higher than that of all human endogenous genes. Consistent with the yeast results, we found a positive correlation between typical decoding time and relative tRNA shortage due to viral translation (Spearman's rank correlation coefficient $\rho = 0.30$, $P = 0.020$; Fig. 2a. See also Supplementary Fig. 5). To further confirm our findings, we estimated the sensitivity of the typical decoding time to tRNA depletion due to viral translation by the ratio between the fold change in typical decoding time and the viral codon consumption. We found that the sensitivity of each codon was negatively correlated with its tRNA supply, suggesting that the typical decoding time of codons with adequate tRNA supply are less sensitive to viral consumption, or *vice versa*. (Spearman's rank correlation coefficient $\rho = -0.29$, $P = 0.023$; Fig. 2b. See also Supplementary Fig. 5). This result therefore suggested that the changes in typical decoding time of codons were indeed caused by the disruption to the share tRNA pool due to the excessive viral translation. The two observations made with flu-infected human cells cannot be explained by the small size of virus genome or interdependence between parameters (Supplementary Fig. 5). Overall, our observations in flu-infected human cells supported the general effect of virus CUB on host translation when the virus expression is high enough.

Experimental assessment for the general role of virus CUB on host translation

The above comparisons among codons suggested a model in which tRNA shortage due to virus translation decelerates decoding. When all codons used by the virus considered together, different codons could be decelerated to different levels. Since translation of an mRNA is a serial process, the codon with the slowest decoding time should become the rate-limiting step that dictates how rapidly ribosomes could be recycled and thus the overall translational efficiency of the host cell. This model predicts that viral genes with synonymous codon usage proportional to the tRNA supply of the host should impose minimum translational load on the host cell, because they give rise to similar levels of tRNA

shortage and, therefore, decoding deceleration for all codons. On the contrary, deviations from this proportionality (D_P . See Methods for how we quantified D_P) between virus codon usage and host tRNA supply would result in more severe tRNA shortage for some codons than the others, which would create even slower “rate-limiting” codons that sequester ribosomes and decrease overall translational efficiency of the host cell. In other words, when the expression level of the exogenous genes was fixed (but see below for when it was not fixed), exogenous genes with smaller D_P should impose lower translational load on the host, and *vice versa*. To test this prediction, we performed manipulative experiments in *S. cerevisiae*, using a highly expressed mCherry as the exogenous gene whose CUB trans-regulates host translational efficiency, and a lowly expressed YFP as a probe for the cellular translational efficiency (Fig. 3a. See Methods). A total of 37 synonymous versions of mCherry were designed with different D_P to the host (yeast) genome (Supplementary Table 2).

We first investigated the trans-regulatory effect of the CUB at a fixed expression level of the exogenous gene. To that end, we estimated the translational load imposed by various versions of mCherry, which is approximated by the reduction of YFP signal relative to the maximal YFP signal detected among all strains. Only cells whose mCherry expression level is at top 20% among all cells were considered to ensure observable impact on host translation by CUB and to limit the mCherry expression to a narrow range such that its variation is controlled. Under this scenario, we found that the translational load was higher for strains with mCherry of higher D_P (Fig. 3b, top of green-shaded area. See also Supplementary Figs. 6 and 7). Furthermore, we approximated the impact of each expression of exogenous gene by the ratio between the translational load and the mCherry expression level. This ratio was found as correlated with the D_P of mCherry (Fig. 3b, bottom of green-shaded area. See also Supplementary Figs. 6 and 7). Collectively, these results, in which the expression of exogenous gene was controlled, suggested that exogenous genes with CUB more similar to that of the host imposed smaller translational load on host cell. These observations were compatible with the results from our comparison among codons based on Ribo-Seq data, where mismatches between host tRNA supply and CUB of exogenous (viral) genes with sufficiently high expression had influenced the codon decoding time.

Our experiment also allowed us to find out the minimum expression level for the exogenous genes to exhibit significant impact on the host translational efficiency and therefore displaying a significant positive correlation between translational load and D_P of mCherry. To approximate this expression threshold, we stratified the > 50,000,000 individual cells from the 37 strains into 100 groups according to their mCherry expression, so that each group contained ~ 500,000 cells from 8 to 19 strains with different mCherry CUBs and very narrow range of (therefore effectively controlling) mCherry expression. We then calculated the within-group Spearman’s rank correlations between the translational load and D_P , which were found significantly positive in the top 43 groups (Supplementary Fig. 8). Further analyses based on qRT-PCR of mCherry *versus* actin and bulk RNA-seq data (see Methods) suggested that the minimum expression of mCherry in these 43 groups approximately ranked as the 208th highly expressed endogenous genes (Supplementary Fig. 8). Intriguingly, in the yeast Ribo-Seq data we analyzed, the total translational activity of virus genes in the Sc-Lys dataset was comparable to the top 201st highly expressed endogenous yeast gene.

This number is 48th for the Sc-His dataset, 201st for the YPD1 dataset, 220th for the YPD2 dataset, but are 306th to 722nd in other five datasets, suggesting that the killer virus indeed had sufficiently high expression level to affect translational efficiency in the majority of strains used by Gardin et al., but not in the strains used in the other five studies.

Nevertheless, besides trans-regulating the overall translational efficiency of the host cell via tRNA availability, CUB of exogenous genes was also known to cis-regulate the expression levels of the exogenous gene itself^{15,32,33}. Specifically, the assimilation of CUB of the exogenous gene to that of the host was found capable of increasing the expression level of exogenous gene¹⁵. This cis-regulatory effect of CUB on the expression level of the exogenous gene was confirmed by our experiment, as we found that mCherry sequences with lower D_P tended to have higher expression levels (Fig. 3b, red-shaded area. See also Supplementary Figs. 6 and 7). This result can be partially explained by the cis-regulatory effect of CUB on mRNA abundance^{32,33} (see Supplementary Table 3 and Supplementary Fig. 6), and is consistent with a known strategy of improving expression of exogenous gene by adjusting its CUB to mimic that of the host^{15,34,35}.

Another important question arises after the aforementioned experimental observations. Theoretically, the regulatory effects of virus CUB on host tRNA availability and virus gene expression level influence host translation in opposite directions. For example, when the virus CUB dissimilates the host CUB, virus expression should reduce and the translational load on the host per unit of viral expression should increase. What is the net effect on host translation when the cis-regulatory effect and trans-regulatory effect of the CUB of exogenous genes are combined, and what is its implication for the coevolution of virus and host? To answer these questions, we examined the net effect of CUB on host translation by directly comparing the CUB of mCherry with translational load as probed by YFP expression. We observed a negative correlation between D_P of mCherry and translational load (Fig. 3b, blue-shaded area. See also Supplementary Figs. 6 and 7), suggesting that exogenous genes with low D_P represented a heavier overall translational load to host cell, at least when the exogenous gene was driven by a strong promoter. Note that this result, in which the impact of changes in expression level of the exogenous genes was shown, was differed from the green-shaded results, in which the expression level of the exogenous genes was controlled. This result was not caused by a lowered YFP mRNA abundance or lowered cellular transcriptional efficiency, because the abundance of YFP mRNA as assessed by qRT-PCR is not correlated with D_P of mCherry (Spearman's rank correlation coefficient $\rho = 0.02$, $P = 0.9$. See Supplementary Table 3 and Supplementary Fig. 6). Also, the increased translational load by mCherry of lower D_P appeared unexplainable by altered translational accuracy caused by changes in tRNA supply, because the translational error rate measured in these strains were not correlated with D_P (Spearman's rank correlation coefficient $\rho = -0.020$, $P = 0.91$. See Methods, and Supplementary Table 4). On the other hand, the increased translational load by mCherry of lower D_P can at least be partially explained by the increased mCherry mRNA abundance (Supplementary Fig. 6), highlighting the role of CUB on translational load via cis-regulation of mRNA abundance of the exogenous genes.

To further confirm the fitness effect of virus CUB, we chose two strains with mCherry of high and low codon usage similarity with yeast ($D_P = 0.156$ and 0.815 , respectively) and

estimated their relative fitness by competitive co-cultures (see Methods). As a result, we found that the strain expressing mCherry with smaller D_p exhibited significant growth deficiency relative to the strain expressing mCherry with larger D_p (Fig. 3c). Collectively, our results based on translational load and fitness consistently suggested that, although the assimilation of CUB of the exogenous gene to that of the host could alleviate tRNA depletion per translation of the exogenous gene, the concomitant effect of increasing the expression of the exogenous gene dominated the way how CUB of the exogenous gene influenced host translation. Consequentially, cells expressing exogenous genes with excessive CUB similarity with the host display fitness disadvantage due to translational selection. Therefore, our results supported the repulsion between the virus and host CUB when they are overly similar, especially for virus-host pairs that coexist and coevolve for a long period, a situation in which both viral expression and minimized translational load on host cells are favored.

Virus CUB tends to be more similar to CUB of symptomatic host than CUB of asymptomatic natural host

Our manipulative experiment in yeast suggested that exogenous genes with CUBs too similar to that of their host might impose a strong translational load on the host cell, creating repulsion between the CUB of the virus and that of the host. If such repulsion indeed plays a role in the coevolution of viruses and hosts, we would predict that among all the hosts that a virus can infect (i.e., be expressed in the host cell), the hosts with CUBs more similar to that of the virus should be more likely to show symptoms compared to other hosts, since the translation in these hosts are more disrupted.

To test our prediction, we reviewed multiple databases of virus-host relationships, from which we extracted trios of species, each containing one virus, its natural host and its symptomatic host (VNS-trio. See Methods). A full list is available in Supplementary Table 5. For each VNS-trio, the natural host is a species that tolerates infection by the virus with little observable symptoms during any stages of the virus infection. The symptomatic host, on the other hand, is another species that, when infected by the virus, exhibits obvious symptoms during one or more stages of the virus life cycle. We collected a total of 52 VNS-trios, where the viruses covered DNA and RNA types, and the hosts ranged from insects to mammals. We then calculated the D_p between virus and natural host, or $D_p(V,N)$, and D_p between virus and symptomatic host, or $D_p(V,S)$. We found that, 43 out of 52 VNS trios showed $D_p(V,N) > D_p(V,S)$ (Binomial $P < 10^{-4}$. Supplementary Table 5), supporting our hypothesized deleterious effect of overly similar virus CUB to that of the host. We further reasoned that, viruses with small deviation from the proportionality rule of CUB are more likely subject to stronger selection for translational efficiency. Therefore, we calculated average ratio between $D_p(V,N)$ and $D_p(V,S)$ for VNS-trios whose $D_p(V,N) + D_p(V,S)$ is smaller than a certain threshold. As we lowered this threshold, the average ratio between $D_p(V,N)$ and $D_p(V,S)$ increases (Fig. 4a). To corroborate this result, we calculated an odds ratio (OR) to test whether each virus CUB was more similar to that of the natural host or that of the symptomatic host (See Methods), such that the greater the OR (relative to 1), the more similar the virus CUB is to the symptomatic host CUB. Consistent with previous result, we

found that for VNS-trios likely subject to stronger translational selection, the fraction of VNS-trios with $OR > 1$ becomes larger (Fig. 4b).

To further evaluate the CUB patterns among VNS-trios, we focused on Dengue and Zika virus, for which multiple sequenced variants allowed additional analyses. Intriguingly, all variants of Dengue and Zika virus downloaded from NCBI appeared to have CUBs more similar to that of their symptomatic host (human) than to that of their natural host (mosquito) (Fig. 4c, see also Supplementary Fig. 9). Notably, some of the virus strains were not isolated from the symptomatic host, meaning that the observed similarity between the virus and symptomatic host cannot be explained by rapid CUB assimilation to the symptomatic host. Collectively, our observations of VNS-trios supported the proposed role of CUB in virus-host coevolution.

Discussion

In the current study, we analyzed yeast and human Ribo-Seq data to show that virus CUB trans-regulates host translation via differential depletion of tRNA. Manipulative experiments in yeast revealed that CUB assimilation of the exogenous gene to the host alleviated the tRNA shortage caused by each translation of exogenous gene and elevated the expression level of the exogenous gene, such that the overall tRNA shortage became severed. These results suggest that viruses with CUBs that are too similar to those of their hosts might impede host translation. This role of virus CUB is further supported by the observation that virus CUB tends to be more similar to that of symptomatic hosts than that of natural hosts. Our results highlighted the impact of virus codon usage on host translation and point to a previously unrecognized complexity in virus-host coevolution.

Our research was inspired by an unexpected finding of a high fraction of reads with viral origin in a yeast Ribo-Seq dataset²⁶. This finding is in contrast with five other yeast Ribo-Seq datasets (Supplementary Table 1) in which viral reads are much rarer, likely explainable by inactive infections or contaminations. Consistent with previous experimental measures of translational efficiency¹⁷, our reanalysis indicated that high translational activities of one to several exogenous genes, whose codon usage did not match the tRNA supply by the host cell, will likely trigger differential tRNA shortage and, therefore, differential changes in decoding time among different codons. Thus, caution with respect to experimental conditions and genetic background should be taken when measuring decoding times.

Our yeast experiments using fluorescence proteins disentangled the two mechanistic components of the influence by CUB of exogenous genes on host cells, namely the cis-regulatory effect of CUB on expression of the exogenous gene, and the trans-regulatory effect on translational efficiency controlling for expression of the exogenous gene. However, the detail molecular mechanisms were likely more complex than this dichotomy. For example, the cis-regulatory effect of CUB on expression of the exogenous genes could have been via direct regulation of mRNA abundance³², which was supported by our data (Supplementary Fig. 6), or modulation of translational elongation rate^{10,36}. It will therefore be interesting to further investigate the independent contribution of these mechanisms in the future.

Previous studies of virus codon usage have focused on the assimilation of virus CUB to that of the host and how it facilitates successful viral expression in the host, whereas the deviation between virus and host CUB were assumed result of genetic drift or mutation bias³⁷. Our findings here suggested another scenario, in which viruses with codon usage too similar to that of the host are harmful to host cells due to their elevated expression and tRNA-depletion effects, thus creating repulsion between the CUBs of the virus and host. Theoretically, such repulsion is particularly important for virus-host pairs where the fitness of the host is important for the successful life cycle of the virus, e.g., pairs involving a natural host/reservoir. Indeed, the duration of virus-host coexistence is generally longer for natural hosts than for symptomatic hosts. Thus, more than enough time should have been available for the virus to evolve its CUB to be as similar to that of the natural host as to that of the symptomatic host. The observation that the virus CUB tended to be more similar to that of the symptomatic host, an organism with a shorter time of coevolution with the virus, than that of the natural host, strongly suggested that the optimal virus CUB, at least for its life stage in the natural host, is not an exact match to that of the host, but contains a slight deviation (Fig. 5). In conclusion, our results suggested an attraction-repulsion relationship between virus and host CUBs, a previously unrecognized complexity in virus-host coevolution.

Our findings also have potential practical implications. In viral epidemiology, synonymous mutations have been proven successful in creating virulence-reduced virus strains as vaccines. Considering the impact of virus codon usage on host translation, synonymous versions of the virus could be designed to control natural host populations, via translation disruption. Given that modifying the small virus genome is relatively easy, this strategy should be potentially feasible in multiple pairs of viruses and natural hosts.

Methods

Genome annotation, Ribo-Seq data, and mRNA-seq data

The *S. cerevisiae* genome sequence and gene annotations (strain S288c, version R64-1-1) were downloaded from SGD³⁸. The genomic sequences and annotations of the yeast viruses L-A (GenBank accession: [NC_003745](#)) and L-BC (GenBank accession: [U01060](#)) were downloaded from NCBI GenBank under accession numbers [NC_003745](#) and [U01060](#), respectively. The 4579-bp genome of L-A contains two open reading frames (ORFs) that are translated either separately or together as a fusion protein³⁹. Because only 6.2% of sites in the ORFs are discriminatory between one fused ORF and two separated ORFs, these two states are difficult to distinguish. In this paper, we arbitrarily considered the fused ORF, but considering the two separate ORFs did not alter our conclusions. The 4615-bp genome of L-BC contains two ORFs⁴⁰. The human genome sequence (GRCh38) and gene annotations were downloaded from Ensembl r85⁴¹. The sequences of IAV genes were downloaded under accession numbers NC_002016 to NC_002023 from NCBI GenBank. For yeast, we reanalyzed the Ribo-Seq data used by Gardin et al.²⁶ and found excessive viral translation. As a comparison, we analyzed five other published yeast Ribo-Seq datasets. The NCBI GEO or SRA accession numbers for these Ribo-Seq datasets are listed in Supplementary Table 1. Yeast mRNA expression levels were from previous mRNA-seq-based estimates⁴². For the

flu infection in human cells, both Ribo-Seq and mRNA-Seq datasets were from GSE82232. To concentrate on the tRNA-depletion effect of viral expression, only the uninfected (SRR3623932) and 2-hour-postinfection (SRR3623937) datasets were analyzed here, because decoding time in later post-infection time points appeared skewed by the progression of host shutoff³¹.

Ribo-Seq datasets from experimental repeats under the same conditions were combined. The four datasets from Gardin et al. had different experimental settings^{26,43} and thus were not combined. Attached adaptors as noted from the respective reports^{26,44} were removed by allowing minor mismatches due to sequencing errors. Any trimmed read shorter than 20 nucleotides or longer than 40 nucleotides was discarded. The remaining reads were first aligned to rRNA and tRNA transcripts in yeast, and the unmapped reads were aligned to the yeast genome by Bowtie2⁴⁵. Ten thousand unalignable reads were randomly selected and BLASTed against the NCBI nr database to identify their origins²⁷. After noticing a significant fraction of reads from two yeast viruses, we combined the viral genomes and the yeast genome for use as the reference in a second run of Bowtie2. To handle Ribo-Seq reads with multiple hits to the genome, we followed Dana and Tuller³⁰ to first record all uniquely mapped reads, and then randomly assign the multihit reads to one of the mapped locations, where the probability of choosing a particular location is proportional to the number of uniquely mapped reads in the region between -10 and +10 nucleotides from the location. Discarding reads with multiple hits did not change our results qualitatively. Also, ambiguous mappings on viral genomes is negligible because no commonly known repetitive elements were found by RepeatMasker⁴⁶ on the viral genome, and only < 3% of reads mapped to viral genome can be mapped to yeast genome with equal alignment score. Finally, each mapped Ribo-Seq reads was added to the ribosome count of one codon site in the ORFeome. Downstream estimation of decoding times was conducted following Gardin et al.'s procedure²⁶ or Dana and Tuller's EMG-based method³⁰ (see below).

Estimation of translational activity, codon consumption, tAI and relative/typical decoding times of codons

We first discarded the leading and trailing 20 codons of each ORF to avoid the influence of translational "ramp" effects⁴⁷ and termination on our estimation. The translational activity of a gene (abundance of ribosome protected fragments, not divided by mRNA abundance) was calculated as the mean ribosome count per site excluding sites with the top or bottom 25% of ribosome counts within each gene; this filter excludes the impact of extreme ribosomal pauses³⁰ and better approximates the translational initiation rate of the gene. Note that this metric reflects the translational activity per gene, not per mRNA molecule, and is corrected for the gene length. Codon consumption in a gene was estimated by the number of occurrences of the codon in the gene multiplied by the translational activity of the gene.

The relative tRNA shortage due to viral translation of a codon was then calculated as the ratio between its total consumption in the viral genes and tRNA supply by the host cell. Here tRNA supply was approximated by either tAI or total codon consumption of all host endogenous genes. For tAI, we first calculated the absolute adaptiveness value W_i for each codon i ,

$$W_i = \sum_{j=1}^{n_i} (1 - s_{ij}) \text{tGCN}_{ij},$$

where n_i is the number of tRNA isoacceptors that recognize the i th codon, tGCN_{ij} is the gene copy number of the j th tRNA that recognizes the i th codon, and s_{ij} is a selective constraint on the efficiency of the codon-anticodon coupling. Additionally, the W_i values can be easily produced by considering Crick's wobble rules⁴⁸. The s_{ij} values for eukaryotes were used as in a previous study³⁰. The copy numbers of the tRNA genes were retrieved from the Genomic tRNA Database⁴⁹. Then, we calculated the relative adaptiveness value $w_i = W_i / W_{\max}$ as the tAI of a codon, where $W_{\max}$ is the maximum W_i value. For total codon consumption of all host endogenous genes, if Ribo-Seq data was available for the estimation of translational activity, it was calculated as described in the last paragraph. If Ribo-Seq data was not available for the estimation of translational activity, we used transcriptome abundance of the gene to approximate translational activity (i.e., equal translational activity for each mRNA molecule was assumed).

We used the Perl scripts provided as supplementary files in Gardin et al.²⁶ to estimate relative decoding times for the yeast dataset. We were able to replicate the estimates in the original report with < 5% deviations, despite the stochasticity in the short read mapping procedure. For the EMG-based analysis, more details are given in Supplementary Text. Briefly, we first determined the distance between the 5' end of a ribosomal footprint and the decoding codon at the ribosome A site, referred to as the offset, as a function of the length of the footprint³⁰. This was achieved by finding the correct reading frame followed by identifying the most likely offset given the reading frame. Finally, the ribosome counts of all sites for the same codon were combined into a ribosome count profile, which was used to estimate typical decoding time by fitting an EMG-NB distribution. The main difference from the original EMG-based typical decoding time³⁰ was the incorporation of a negative binomial (NB) distribution, which avoided a major expression-dependent bias in the original EMG-based method (see Supplementary Text and Supplementary Fig. 3).

Fluorescence reporter gene assay in yeast

We designed 37 synonymous versions of mCherry (Supplementary Table 2) based on the previously used mCherry¹⁷. These coding sequences were designed to have CUBs with variable similarities to the host (yeast) CUB. The model of trans-regulation of host translational efficiency by CUB of exogenous genes predicted that exogenous genes with synonymous codon usage proportional to the tRNA supply of the host should impose minimum translational load on the host cell, because they give rise to similar levels of tRNA shortage and, therefore, decoding deceleration for all codons. On the contrary, deviations from this proportionality between virus CUB and host tRNA supply would result in more severe tRNA shortage for some codons than the others, which would create even slower "rate-limiting" codons that sequester ribosomes and decrease overall translational efficiency of the host cell. Therefore, to test our model, we need to quantify the deviation from proportionality (D_p) between codon usages of the exogenous gene and host tRNA supply. To this end, for each of the 18 amino acids encoded by at least two synonymous codons, we

first calculated the Euclidean distance in synonymous codon usage between the exogenous coding sequence and the tRNA supply by the host by the following equation.

$$D_i = \sqrt{\sum_{j=1}^{n_i} (Y_{ij} - X_{ij})^2},$$

Here n_i is the number of synonymous codons for amino acid i , Y_{ij} is the fraction of codon j among the synonymous codons of amino acid i for the exogenous coding sequence, X_{ij} is tRNA supply represented by either the fraction of codon j among the synonymous codons in the host transcriptome⁵⁰, or rescaled tAI values such that all tAIs within a group of synonymous codons sum up to 1. Then D_p of the gene is defined as the weighted geometric mean of all 18 D_i . Note that the approximation of tRNA supply by the transcriptomic codon usage was only valid when one assumed that codon usage is mostly shaped by selection for translational efficiency. It is also important to note that traditional measurements for CUB such as the codon adaptation index (CAI)⁵¹, I_{TE} ^{10,52}, or fraction of optimal codons (F_{op})⁵³ is not ideal for testing our model, because virus with excessive usage of preferred codons also imposes higher translational load than that with usage of preferred and unpreferred codons in proportion to the corresponding tRNA supply by the host, especially when the exogenous genes was highly expressed. For VNS-trios analyses (see below), in which the reliability of annotations in most species remains questionable, we resorted to genomic codon usage of top 100 highly expressed genes (expression determined by RNA-Seq datasets listed in Supplementary Table 5) for an approximation of X_{ij} unless otherwise noted. Although this practice is clearly not ideal¹⁷, such approximation of transcriptomic CUB by genomic CUB has been successful in multiple previous studies^{54–57}. Additionally, our test for CUB similarity in VNS-trios (MH-test based on 2×2 contingency tables. See below) should not be sensitive to the minor difference between genomic and transcriptomic codon usage.

We aimed to test the trans-regulatory role of CUB of exogenous genes on host translational efficiency using manipulative experiments in *S. cerevisiae*. To focus our experiment on tRNA depletion due to the translation of the exogenous/viral gene and exclude other translation-interfering mechanisms that are specific to any virus (such as host shutoff by IAV³¹), we decided to use mCherry, which encodes a fluorescent protein, as the exogenous gene to be excessively translated. A total of 37 synonymous versions of mCherry were designed with different levels of CUB similarity to the host (yeast) genome (Supplementary Table 2). These synonymous mCherry sequences have similar G+C content (41–45%), and identical sequences in the first 56 nucleotides of the coding region, where synonymous changes may affect the level of protein expression in a tRNA-independent manner^{7,19,58}. Furthermore, mCherry expression was driven by a constitutive strong promoter (pTDH3), ensuring that any impact on host translation would be observable. The overall translational efficiency of the host cell was detected using a reporter gene, the Venus yellow fluorescent protein (YFP). Here YFP was controlled by a relatively weak promoter (pDET1), minimizing its impact on host tRNA supply. The reporter cassette of mCherry and YFP were derived from previously used mCherry reporters¹⁷ and their modified versions. We designed four primers for each modified version of mCherry. The first and second primer or the third and fourth primers were respectively used to amplify two modified fragments, which were

then fused using the first and fourth primers in a fusion PCR. All the DNA fragments of mCherry and YFP were then cloned into the HO locus of BY4741 by homologous recombination using the lithium acetate protocol⁵⁹, with URA3 as the selection marker for successful transformation. All PCR and cloning primers are listed in Supplementary Table 6. This design allowed us to mimic the translational impact on the host of exogenous gene (mCherry) expression while tracking the overall translational efficiency of the host cell with a fluorescent reporter (YFP).

We measured the expression levels of mCherry and YFP of 300,000 cells for each of the 37 strains in the log phase in yeast extract/peptone/dextrose (YPD) media using flow cytometry (CytoFLEX S, Beckman). The fluorescence of mCherry was measured by a filter with a 20 nm bandpass centered on 610 nm, and the fluorescence of YFP was measured by a filter with a 40 nm bandpass centered on 525 nm. Yeast cells with mCherry and YFP fluorescence signals ten times greater than those of the BY4741 negative control cells were kept for later analyses. We retrieved the forward scatter (FSC, which is proportional to cell size) and mCherry and YFP fluorescence signals for all cells. The expression levels of the fluorescent proteins were defined as their fluorescence signals divided by FSC. All experiments were carried out with three biological and three technological replicates.

For qRT-PCR, total RNA samples were isolated from the log phase of all strains in yeast extract/peptone/dextrose (YPD) media using RNeasy Plus Mini Kit (QIAGEN, 74134) according to the manufacturer's instructions. The cDNAs were reversely transcribed from the RNAs using the Evo M-MLV RT Kit with gDNA Clean for qPCR (Takara, AG11705) according to the manufacturer's instructions. Then qRT-PCR were carried out by the cDNAs and primers (Supplementary Table 7) using QuantiNova SYBR Green PCR kit (QIAGEN, 208057–500T) according to the manufacturer's instructions.

To determine the minimum expression threshold for exogenous genes to have a significant impact on host translational efficiency, we stratified all single cells (~ 50,000,000) from the 37 strains into 100 groups according to their mCherry expression, so that each group has ~ 500,000 cells from 8 to 19 strains with mCherry of different CUB. We found that in the top 43 groups with high mCherry expression levels, the within-group Spearman's rank correlations between the translational load and D_p are significantly positive (Supplementary Fig. 8a), suggesting a significant impact of CUB of exogenous gene on the overall translational efficiency. We performed qRT-PCR in the strains with at least 1 million cells in these 43 groups, and determined that the abundance of mCherry mRNA in these strains were 0.7 to 3.4 fold of that of actin (Supplementary Table 3). According to RNA-seq-based transcriptome profile of bulk yeast cells⁴², 0.7 fold of the expression level of actin corresponds to the 208th highly expressed endogenous gene (Supplementary Fig. 8). This result suggested that CUB of exogenous genes whose total expression is not lower than the 208th highly expressed endogenous gene should have a significant impact on the overall translational efficiency of the yeast cell.

It is worth noting that our observations with exogenous genes translated in yeast were in agreement with a previous study, where codon CGU and CGC were synonymously converted to CGG in eight highly expressed endogenous genes, presumably creating severe

tRNA shortage, and the proteome-wide translation efficiency were subsequently reduced²¹. Moreover, rescuing the translation efficiency by increasing tRNA supply for CGG further supported the proposed mechanism of tRNA shortage²¹. Interestingly, unlike our observations that synonymous changes in mCherry cis-regulated its expression (Fig. 3b, red-shaded area), the codon conversion in the previous study only reduced one of the eight converted genes²¹. We speculated that the lack of significant expression change in the synonymously converted genes can be explained by the focused conversions in one group of synonymous codons (CGU/CGC/CGG, encoding Arginine), which should lead to relatively small changes in the CUB and therefore the expression of the converted genes, but severe shortage for one tRNA. On the contrary, synonymous substitutions of mCherry were conducted on all groups of synonymous codons in this present study, thus resulting in greater changes in the CUB and therefore the expression of mCherry.

We performed competitive co-culture assays to measure the relative fitness of two chosen strains ($D_p = 0.156$ and 0.815 , respectively). The two strains were first inoculated in 4ml YPD competition culture at 30°C with shaking 200 rpm for 16 hours. Then 2ml of each sample was left for DNA extraction and marked as day0 of the competition. At the same time we measured the total cell density of each competition using a Coulter counter (Beckman-Coulter) and diluted each competition into a fresh 100mL culture to reach concentration of $2-5 \times 10^5$ cells/ml. Then, cells were placed back at 30°C for 10–14h, such that the total cell density values were kept below 0.7 to maintain exponential growth throughout the experiment. We repeat this procedure for 6 days, resulting in one day0 sample and three for each of day2, day4, and day6 samples. We then extracted the DNA and amplified strain-specific region of mCherry sequences of the two strains with primers in Supplementary Table 8. The ten amplicon samples were mixed together and paired-end sequenced for 150 nt on either end on Illumina NovaSeq platform, for a total of 1Gbp raw data. Then, the data of ten samples were split using the 5bp index sequence at the 5' end of left primers. The paired reads were mapped to the two versions of mCherry using STAR⁶⁰. The numbers of reads uniquely mapped to either version of mCherry were used to approximate the relative population size of the two strains in the competition culture. The ratio between relative population size at day 2/4/6 and that of day 0 was further divided by such ratio of the fitter strain to yield a relative ratio of cells (Fig. 3c).

To assess the effect of codon usage of exogenous genes on the translational accuracy of the host, we followed previous studies^{36,61} to estimate the translational error rate in the BY4741 and the 36 yeast strains with different mCherry versions using a dual-luciferase system. This system contains two luciferases, Renilla and firefly, in a fusion protein. The measurement of concentration-independent firefly activity was therefore possible by the ratio between the observed firefly (F) and Renilla (R) activities. We used both wildtype and a mutant of the firefly, in which codon AAA (encoding Lys) at positive 529 was replaced with AGG (encoding Arg). Because no other side chain interacts with the luciferase substrate as the Lys side chain does⁶¹, the mutant can display firefly activity only when it is mistranslated as Lys at position 529. In other words, the concentration-independent firefly activity, relative to that of the wildtype firefly, measures the translational error rate^{36,61}. We successfully measured the translational error rate in the BY4741 and 35 out of the 37 yeast strains with mCherry (Supplementary Table 4).

Codon usage analysis of virus, natural host, and symptomatic host triples

To collect trios of virus, natural host and symptomatic host (VNS-trio), we manually collected host range information for various viruses from NCBI (<https://www.ncbi.nlm.nih.gov/taxonomy/>), ViralZone (<http://viralzone.expasy.org/>)⁶², Centers for Disease Control and Prevention (<https://www.cdc.gov/>) and Wikipedia (<https://www.wikipedia.org/>). To test our model, we needed viruses and hosts with known genomes or transcriptomes. It was also important to ensure that the virus genome encoded no tRNA gene, and the natural hosts were symptom-free or at least that their known symptoms were significantly milder than those of the symptomatic hosts. We thus used relatively stringent criteria to filter for highly reliable VNS-trios. The final list of 52 VNS-trios is given in Supplementary Table 5. The complete genome sequences of these organisms were obtained from GenBank.

We calculated an odds ratio (OR) to test whether each virus CUB was more similar to that of the natural host or that of the symptomatic host. Specifically, a 2×2 contingency table was constructed as

	$V_i > V_{mean}$	$V_i < V_{mean}$
$ V_i - N_i > V_i - S_i $	A	C
$ V_i - N_i < V_i - S_i $	B	D

where V , N and S are the frequencies of a codon in the virus, natural host and symptomatic host, respectively. Therefore, $V_i > V_{mean}$ means that a codon was a preferred codon in the virus, while $V_i < V_{mean}$ means it was unpreferred in the virus; $|V_i - N_i| > |V_i - S_i|$ means the virus codon frequency was more similar to that of the symptomatic host than to that of the natural host, while $|V_i - N_i| < |V_i - S_i|$ means the virus codon frequency was more similar to that of the natural host than to that of the symptomatic host. A 2×2 contingency table was constructed for each amino acid in each VNS-trio counting each codon toward A/B/C/D, such that biased usage of amino acids²⁴ was excluded. Finally, a common OR was calculated by combining the relevant contingency tables using the MH procedure implemented in R. An $OR > 1$ indicates that virus CUB tends to be more similar to that of the symptomatic host than that of the natural host, while an $OR < 1$ indicates that virus CUB tends to be more similar to that of the natural host than that of the symptomatic host.

For more thorough analyses of Dengue and Zika virus variants, all mutants recorded in NCBI (<https://www.ncbi.nlm.nih.gov/genome/viruses/variation/>) were downloaded. We calculated D_p for each mutant based on the transcriptomes of the natural and symptomatic hosts (accession numbers from Gene Expression Omnibus: GSE96605 for mosquito and GSE97949 for human).

There are a few caveats worth discussion in regard to in our VNS-trio analyses. First, during the viral infection process, the survival of the virus and the phenotypic outcome of the infection will be affected by the host immune system, a factor we did not address in our manipulative experiment in yeast or our VNS-trio analyses. One possibility is that all the

symptoms of infection are caused by a strong immune system reaction in the symptomatic hosts instead of high viral expression. This hypothesis, however, cannot explain zoonotic viruses such as the Ebola virus, whose natural (bat) and symptomatic (human) hosts both have adaptive immune systems. Indeed, among the 30 VNS-trios containing natural hosts with adaptive immune systems, 21 showed stronger CUB similarity ($D_p(V,S) < D_p(V,N)$) of the virus to the symptomatic host than to the natural host (binomial $P = 0.008$). Second, the stronger phenotypic consequence in symptomatic host relative to natural host might be a result of higher viral protein abundance in symptomatic host, but not translational disruption. However, this alternative hypothesis does not predict that the phenotypic consequences will intensify when translational selection is stronger (Figs. 4a, b). More importantly, the results from our manipulative experiments in yeast (Fig. 3c) suggest that even when the protein product of the exogenous gene has no potential virulence, the excessive CUB similarity between the exogenous gene and the host is still detrimental to host fitness. These evidences suggested that translational selection at least has a nonnegligible role in determining the phenotypic consequence of viral infection. Third, when we calculated D_p , we used transcriptomic codon usage of the host to approximate tRNA supply, which implicitly assumed that codon usage is mostly shaped by selection for translational efficiency. This approximation could be biased if codon usage is mostly shaped by other selective forces, such as selection for translational accuracy. Nevertheless, whenever we also approximated tRNA supply by tAI, another independent tRNA supply estimation based on genomic tRNA copy numbers, the result was always consistent with that found by transcriptomic codon usage. Therefore, it is highly unlikely that the aforementioned assumption heavily biased the calculation of D_p , or our conclusions based on D_p . Fourth, besides the repulsion caused by translational selection revealed in the current study, the increased CUB similarity between virus and natural host compared to that between virus and symptomatic host can also be explained by genetic drift (or weakened selection⁶³) and mutation bias. The relative contribution of these three factors might not be assessable until more quantitative data from specific virus-host pairs become available. Our study, while not rejecting the role of genetic drift and mutation bias, offered unequivocal support for the contribution of CUB repulsion between the virus and the host. For example, our study showed decelerated codon decoding time (Fig. 1 and 2) and reduced cellular translational efficiency (Fig. 3b) when the CUB of exogenous genes was overly similar to the host. The existence of such trans-regulatory effect dictates that viral translation should impose a translational load on host cell as long as the viral expression is high enough, therefore forming a mechanical basis for the CUB repulsion between the virus and the host. More importantly, the deleterious effect of excessive similarity between CUB of exogenous genes and that of the host was directly observed (Fig. 3c). Fifth, our model by no means rejected other clade-specific mechanisms causing differences in CUB deviation between viruses and hosts, such as virus-encoded tRNA gene^{64,65} and differential mutation bias in dsDNA versus ssDNA viruses^{65,66}. Last but not least, our model did not exclude other mechanism to adjust viral expression to avoid impediment of host translation, such as changes in mRNA levels due to mutations in promoters or other regulatory elements. Instead, our results highlighted that natural selection on viral expression is unlikely unidirectional, because various regulatory elements, including CUB and promoter sequences, collectively determined the viral expression.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgment

We thanked Xionglei He, Wenfeng Qian, Jianzhi Zhang, Zhenzhen Zhou, Zhuoxing Wu, Wenjun Shi, Zizhang Li for their comments on the manuscript. This work was supported by the National Special Research Program of China for Important Infectious Diseases (grant number 2018ZX10302103 to X. C.), the National Key R&D Program of China (grant number 2017YFA0103504 to X. C., and grant number 2018ZX10301402 to J.-R. Y. and Z. H.), and the National Natural Science Foundation of China (grant numbers 31671320, 31871320 and 81830103 to J.-R. Y., 31771406 to X. C.), and the start-up grant from “100 Top Talents Program” of Sun Yat-sen University (grant number 50000-18821112 to X. C. and grant number 50000-18821117 to J.-R. Y.).

References

1. Muto A & Osawa S The guanine and cytosine content of genomic DNA and bacterial evolution. *Proc. Natl. Acad. Sci. USA* 84, 166–169 (1987). [PubMed: 3467347]
2. Xia X Maximizing transcription efficiency causes codon usage bias. *Genetics* 144, 1309–1320 (1996). [PubMed: 8913770]
3. Bulmer M The selection-mutation-drift theory of synonymous codon usage. *Genetics* 129, 897–907 (1991). [PubMed: 1752426]
4. Hershberg R & Petrov DA Selection on codon bias. *Annu. Rev. Genet* 42, 287–299 (2008). [PubMed: 18983258]
5. Gouy M & Gautier C Codon usage in bacteria: correlation with gene expressivity. *Nucleic Acids Res.* 10, 7055–7074 (1982). [PubMed: 6760125]
6. Xia X How optimized is the translational machinery in *Escherichia coli*, *Salmonella typhimurium* and *Saccharomyces cerevisiae*? *Genetics* 149, 37–44 (1998). [PubMed: 9584084]
7. Tuller T, Waldman YY, Kupiec M & Ruppin E Translation efficiency is determined by both codon bias and folding energy. *Proc. Natl. Acad. Sci. USA* 107, 3645–3650 (2010). [PubMed: 20133581]
8. Robinson M et al. Codon usage can affect efficiency of translation of genes in *Escherichia coli*. *Nucleic Acids Res.* 12, 6663–6671 (1984). [PubMed: 6091031]
9. Sorensen MA, Kurland CG & Pedersen S Codon usage determines translation rate in *Escherichia coli*. *J. Mol. Biol* 207, 365–377 (1989). [PubMed: 2474074]
10. Xia X A major controversy in codon-anticodon adaptation resolved by a new codon usage index. *Genetics* 199, 573–579 (2015). [PubMed: 25480780]
11. Akashi H Synonymous codon usage in *Drosophila melanogaster*: natural selection and translational accuracy. *Genetics* 136, 927–935 (1994). [PubMed: 8005445]
12. Stoletzki N & Eyre-Walker A Synonymous codon usage in *Escherichia coli*: selection for translational accuracy. *Mol. Biol. Evol* 24, 374–381 (2006). [PubMed: 17101719]
13. Johnston TC, Borgia PT & Parker J Codon specificity of starvation induced misreading. *Mol. Gen. Genet* 195, 459–465 (1984). [PubMed: 6590943]
14. Johnston TC & Parker J Streptomycin-induced, third-position misreading of the genetic code. *J. Mol. Biol* 181, 313–315 (1985). [PubMed: 3920401]
15. Plotkin JB & Kudla G Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet* 12, 32 (2010). [PubMed: 21102527]
16. Ikemura T Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. *J. Mol. Biol* 146, 1–21 (1981). [PubMed: 6167728]
17. Qian W, Yang JR, Pearson NM, Maclean C & Zhang J Balanced codon usage optimizes eukaryotic translational efficiency. *PLOS Genet.* 8, e1002603 (2012). [PubMed: 22479199]
18. Akashi H & Eyre-Walker A Translational selection and molecular evolution. *Curr. Opin. Genet. Dev* 8, 688–693 (1998). [PubMed: 9914211]

19. Kudla G, Murray AW, Tollervey D & Plotkin JB Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* 324, 255–258 (2009). [PubMed: 19359587]
20. Shah P, Ding Y, Niemczyk M, Kudla G & Plotkin JB Rate-limiting steps in yeast protein translation. *Cell* 153, 1589–1601 (2013). [PubMed: 23791185]
21. Frumkin I et al. Codon usage of highly expressed genes affects proteome-wide translation efficiency. *Proc. Natl. Acad. Sci. USA* 115, E4940–E4949 (2018). [PubMed: 29735666]
22. Tian L, Shen X, Murphy RW & Shen Y The adaptation of codon usage of +ssRNA viruses to their hosts. *Infect. Genet. Evol* 63, 175–179 (2018). [PubMed: 29864509]
23. Albers S & Czech A Exploiting tRNAs to boost virulence. *Life (Basel)* 6 pii: E4 (2016). [PubMed: 26797637]
24. Bahir I, Fromer M, Prat Y & Linial M Viral adaptation to host: a proteome-based analysis of codon usage and amino acid preferences. *Mol. Syst. Biol* 5, 311 (2009). [PubMed: 19888206]
25. Lucks JB, Nelson DR, Kudla GR & Plotkin JB Genome landscapes and bacteriophage codon usage. *PLOS Comput. Biol* 4, e1000001 (2008). [PubMed: 18463708]
26. Gardin J et al. Measurement of average decoding rates of the 61 sense codons in vivo. *Elife* 3, e03735 (2014).
27. Camacho C et al. BLAST+: architecture and applications. *BMC Bioinformatics* 10, 421–421 (2009). [PubMed: 20003500]
28. Schmitt MJ & Breinig F The viral killer system in yeast: from molecular biology to application. *FEMS Microb. Rev* 26, 257–276 (2002).
29. Ribas JC & Wickner RB *Saccharomyces cerevisiae* L-BC double-stranded RNA virus replicase recognizes the L-A positive-strand RNA 3' end. *J. Virol* 70, 292–297 (1996). [PubMed: 8523538]
30. Dana A & Tuller T The effect of tRNA levels on decoding times of mRNA codons. *Nucleic Acids Res.* 42, 9171–9181 (2014). [PubMed: 25056313]
31. Bercovich-Kinori A et al. A systematic view on influenza induced host shutoff. *Elife* 5 (2016).
32. Chen S et al. Codon-resolution analysis reveals a direct and context-dependent impact of individual synonymous mutations on mRNA level. *Mol. Biol. Evol* 34, 2944–2958 (2017). [PubMed: 28961875]
33. Presnyak V et al. Codon optimality is a major determinant of mRNA stability. *Cell* 160, 1111–1124 (2015). [PubMed: 25768907]
34. Mignon C et al. Codon harmonization - going beyond the speed limit for protein expression. *FEBS Lett.* 592, 1554–1564 (2018). [PubMed: 29624661]
35. Angov E, Legler PM & Mease RM Adjustment of codon usage frequencies by codon harmonization improves protein expression and folding. *Methods Mol. Biol* 705, 1–13 (2011). [PubMed: 21125377]
36. Yang JR, Chen X & Zhang J Codon-by-codon modulation of translational speed and accuracy via mRNA folding. *PLOS Biol.* 12, e1001910 (2014). [PubMed: 25051069]
37. Jenkins GM & Holmes EC The extent of codon usage bias in human RNA viruses and its evolutionary origin. *Virus Res.* 92, 1–7 (2003). [PubMed: 12606071]
38. Cherry JM et al. *Saccharomyces* Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res.* 40, D700–705 (2012). [PubMed: 22110037]
39. Icho T & Wickner RB The double-stranded RNA genome of yeast virus L-A encodes its own putative RNA polymerase by fusing two open reading frames. *J. Biol. Chem* 264, 6716–6723 (1989). [PubMed: 2651431]
40. Bruenn JA A closely related group of RNA-dependent RNA polymerases from double-stranded RNA viruses. *Nucleic Acids Res.* 21, 5667–5669 (1993). [PubMed: 8284213]
41. Zerbino DR et al. Ensembl 2018. *Nucleic Acids Res.* 46, D754–D761 (2018). [PubMed: 29155950]
42. Nagalakshmi U et al. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 320, 1344–1349 (2008). [PubMed: 18451266]
43. Cai Y & Futcher B Effects of the yeast RNA-binding protein Whi3 on the half-life and abundance of CLN3 mRNA and Other targets. *PLOS ONE* 8, e84630 (2014).

44. Ingolia NT, Ghaemmaghami S, Newman JRS & Weissman JS Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 324, 218–223 (2009). [PubMed: 19213877]
45. Langmead B & Salzberg SL Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357 (2012). [PubMed: 22388286]
46. Tarailo-Graovac M & Chen N Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics* Chapter 4, Unit 4 10 (2009).
47. Tuller T et al. An Evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* 141, 344–354 (2010). [PubMed: 20403328]
48. Crick FHC Codon—anticodon pairing: The wobble hypothesis. *J. Mol. Biol* 19, 548–555 (1966). [PubMed: 5969078]
49. Chan PP & Lowe TM GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res.* 37, D93–D97 (2008). [PubMed: 18984615]
50. Holstege FC et al. Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* 95, 717–728 (1998). [PubMed: 9845373]
51. Sharp PM & Li WH The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* 15, 1281–1295 (1987). [PubMed: 3547335]
52. Bassik MC et al. Rapid creation and quantitative monitoring of high coverage shRNA libraries. *Nat. Methods* 6, 443–445 (2009). [PubMed: 19448642]
53. Drummond DA & Wilke CO Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* 134, 341–352 (2008). [PubMed: 18662548]
54. You E et al. Codon usage bias analysis for the spermidine synthase gene from *Camellia sinensis* (L.) O. Kuntze. *Genet Mol Res.* 14, 7368–7376 (2015). [PubMed: 26214415]
55. Liu Q, Hu H & Wang H Mutational bias is the driving force for shaping the synonymous codon usage pattern of alternatively spliced genes in rice (*Oryza sativa* L.). *Mol. Genet. Genomics* 290, 649–660 (2015). [PubMed: 25407289]
56. Jia X et al. Non-uniqueness of factors constraint on the codon usage in *Bombyx mori*. *BMC Genomics* 16, 356 (2015). [PubMed: 25943559]
57. Shen X, Wang H, Wang M & Liu B The complete mitochondrial genome sequence of *Euphausia pacifica* (Malacostraca: Euphausiacea) reveals a novel gene order and unusual tandem repeats. *Genome* 54, 911–922 (2011). [PubMed: 22017501]
58. Gu W, Zhou T & Wilke CO A universal trend of reduced mRNA stability near the translation-initiation site in prokaryotes and eukaryotes. *PLOS Comput. Biol* 6, e1000664 (2010). [PubMed: 20140241]
59. Gietz RD Yeast transformation by the LiAc/SS carrier DNA/PEG method *Yeast Genetics: Methods and Protocols*. Chapter 1, 1–12 (Springer New York, 2014).
60. Dobin A & Gingeras TR Mapping RNA-seq reads with STAR. *Curr. Protoc. Bioinformatics* 51, 11.14.11–11.14.19 (2015). [PubMed: 26334920]
61. Kramer EB, Vallabhaneni H, Mayer LM & Farabaugh PJ A comprehensive analysis of translational missense errors in the yeast *Saccharomyces cerevisiae*. *RNA* 16, 1797–1808 (2010). [PubMed: 20651030]
62. Hulo C et al. ViralZone: a knowledge resource to understand virus diversity. *Nucleic Acids Res.* 39, D576–582 (2011). [PubMed: 20947564]
63. Prabhakaran R, Chithambaram S & Xia X *Escherichia coli* and *Staphylococcus* phages: effect of translation initiation efficiency on differential codon adaptation mediated by virulent and temperate lifestyles. *J. Gen. Virol* 96, 1169–1179 (2015). [PubMed: 25614589]
64. Bailly-Bechet M, Vergassola M & Rocha E Causes for the intriguing presence of tRNAs in phages. *Genome Res.* 17, 1486–1495 (2007). [PubMed: 17785533]
65. Chithambaram S, Prabhakaran R & Xia X Differential codon adaptation between dsDNA and ssDNA phages in *Escherichia coli*. *Mol. Biol. Evol* 31, 1606–1617 (2014). [PubMed: 24586046]
66. Chithambaram S, Prabhakaran R & Xia X The effect of mutation and selection on codon adaptation in *Escherichia coli* bacteriophage. *Genetics* 197, 301–315 (2014). [PubMed: 24583580]

Fig. 1. Codon usage of viral genes with excessive translation impacted the decoding time of codons in yeast.

a, The average translational activity of viral genes is higher than that of yeast endogenous genes in the four datasets generated by Gardin et al²⁶, but not in other yeast Ribo-Seq datasets examined. **b**, Spearman's rank correlation between tAI and viral codon consumption is significantly weaker than that between tAI and the overall codon consumption of yeast endogenous genes ($P = 0.001$, Fisher's r-to-z transformation), and is significantly weaker than that between tAI and the codon consumption of three highly expressed yeast endogenous genes ($P = 0.035$, Fisher's r-to-z transformation). Error bar indicates standard errors estimated from 1,000 randomly sampled sets of three (among 200) highly expressed yeast endogenous genes. **c**, Relative tRNA shortages among different codons caused by viral translation are significantly correlated with relative decoding time. **d**, Comparison of the Sc-Lys dataset with five other yeast datasets. Nominal P values for Spearman's rank correlations are indicated by asterisks: **, $P < 0.01$. **e**, Typical decoding time is significantly correlated with relative tRNA shortages caused by viral translation. In **c** and **e**, each dot represents one codon, and the blue lines indicate fitted linear models and Spearman's rank correlations are shown. In **b-e**, statistical tests were all based on 61 codons.

Fig. 2. Impact of IAV codon usage on decoding time in infected human cells.

a, Spearman's rank correlation between typical decoding time and relative tRNA shortage due to viral translation is shown for human A549 cells infected by IAV. **b**, The sensitivity of typical decoding time was calculated as the fold change in typical decoding time between virus-infected and virus-free cells, divided by the viral codon consumption in infected cells. The sensitivity of the 61 sense synonymous codons is negatively correlated with their tRNA supply, which was approximated by tAI. In both panels, each dot represents one of the 61 codon, and the blue lines indicate fitted linear models and Spearman's rank correlations are shown.

Fig. 3. Manipulative experiments in yeast elucidate the regulatory effects of CUB on host translation.

a, The reporter cassette used for examining the impact of codon usage of exogenous genes on host translational efficiency. The exogenous gene, mCherry, which is driven by a strong promoter (pTDH3), imposes a translational load on the host cell, whereas the overall translational efficiency of the cell is probed by YFP, which is driven by a weak promoter (pDET1). The whole cassette was inserted into the HO locus of *S. cerevisiae* strain BY4741. Experiments for a total of 37 synonymous versions of mCherry were conducted. Promoter and terminators were indicated as dark gray and black boxes, whereas the ORF of YFP, URA3 and mCherry were indicated as orange, gray and red arrows. **b**, Regulatory effects by CUB of exogenous genes. The similarity of mCherry CUB to that of yeast was measured by D_p . The translational load ($YFP_{max} - YFP$) imposed by mCherry expression was approximated by the reduction of YFP signal relative to the maximal YFP signal detected among all strains. The Spearman's rank correlation coefficient between D_p and various metrics as indicated on the figure are shown. Nominal P values for Spearman's rank correlations are indicated by asterisks: **, $P < 0.01$. Sample sizes $n = 37$ strains, except that cells with top 20% mCherry expression only contained 15 strains. **c**, Two strains with mCherry of high and low codon usage similarity with yeast ($D_p = 0.156$ and 0.815 , respectively) were assayed for relative fitness in a competitive culture. The relative

population sizes of different strains were determined by high throughput sequencing of strain-specific region of mCherry. The ratio between relative population size at day x and that of day 0 was further divided by such ratio of the fitter strain. Error bars indicate standard error assessed by three biological replicates.

Fig. 4. CUB similarity among virus, natural host, and symptomatic host.

a. CUB similarity between virus and symptomatic or natural host was approximated by $D_P(V, S)$ or $D_P(V, N)$, respectively. VNS-trios with $D_P(V, S) + D_P(V, N)$ smaller than certain thresholds (x axis), which were more likely subject to stronger translational selection, were used to calculate the average $D_P(V, N)/D_P(V, S)$ (y axis). Error bars indicate the standard error. **b.** Similar to (A), except an odds ratio was calculated by the Mantel-Haenszel procedure (see Methods), and was used to identify the fraction of virus with CUB more similar to that of the symptomatic host than to that of the natural host, i.e., fraction of VNS-trios with OR > 1 (y axis). Error bars indicate the standard error, estimated by bootstrapping the VNS-trios 1000 times. **c.** The CUBs of all mutants of Dengue and Zika viruses recorded in the NCBI database were compared to those of the natural host (mosquito) and symptomatic host (human). The virus CUB was always more similar to the symptomatic host than the natural host, as the points are all above the dotted diagonal line of $x=y$. The origin of the virus strain that was used to determine the genomic sequence is indicated by the color of the point.

Fig. 5. A schematic diagram of the regulatory role of virus CUB and its evolutionary implication
 As the difference in codon usage between virus and host (x axis) decreased, expression of virus protein (solid curve) increased due to the cis-regulatory effect of virus CUB, meanwhile the translational efficiency (dashed curve) of host decreased due to the trans-regulatory effect of virus CUB. Combination of these effects created an attraction-repulsion relationship between CUB of virus and host.