nature ecology & evolution

ARTICLES

https://doi.org/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 Wu[2], Shuyun Deng[4], Heng Zhang[5], Yutong Hou[2], Zheng Hu[6], Jianzhi Zhang[7]✉, Xiaoshu Chen[4,7]✉ and Jian-Rong Yang[1,2,3,7,8,9]✉

Eighteen of the 20 amino acids are each encoded by more than one synonymous codon. Due to differential transfer RNA supply within the cell, synonymous codons are not used with equal frequency, 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 tested it in the case of viral infection, a common form of naturally occurring exogenous gene translation. We analysed 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 apply decreased translational load on the host per unit of expression, whereas 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 severely impeded host translation. Finally, using a manually curated list of viruses and natural 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 virus and host. 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.

All amino acids, except methionine and tryptophan, are encoded by two or more synonymous codons. Some synonymous codons, termed '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, the exact means by which 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 cognate transfer RNAs for preferred codons are more abundant than those for unpreferred codons[15,16]. Additionally, CUB of highly expressed genes *trans*-regulates overall translational efficiency in the cell because all translation within 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 following viral infection. Because the majority of viruses have highly compact genomes that do not encode any tRNA, the translation of viral proteins is reliant on host tRNA[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, the means by which virus CUB impacts translation in host cells, and how this type of virus–host interaction affects viral evolution, remain elusive.

While analysing 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 translation of viral proteins decelerated the decoding of codons whose cognate tRNAs 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*-regulates host translation via differential depletion of tRNA. We used a fluorescence reporter system to further investigate how CUB of exogenous genes would affect host translation, and found that CUB similarity to the host increased the expression of exogenous genes while reducing

[1]Program in Cancer Research, Zhongshan School of Medicine, The Fifth Affiliated Hospital, Sun Yat-sen University, Guangzhou, China. [2]Department of Biomedical Informatics, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, China. [3]Key Laboratory of Tropical Disease Control, Ministry of Education, Sun Yat-sen University, Guangzhou, China. [4]Department of Medical Genetics, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, China. [5]Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, China. [6]Department of Obstetrics and Gynecology, Precision Medicine Institute, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, China. [7]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA. [8]RNA Biomedical Institute, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, China. [9]Guangdong Provincial Key Laboratory of Biomedical Imaging, The Fifth Affiliated Hospital, Sun Yat-sen University, Zhuhai, Guangdong Province, China.
✉e-mail: jianzhi@umich.edu; chenxshu3@mail.sysu.edu.cn; yangjianrong@mail.sysu.edu.cn

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 the host imposed a higher translational load on the host. This result indicated that viruses with CUB overly 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 few symptoms during viral infection) and its symptomatic host (a species with obvious symptoms in one or more stages of 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 analysed four Ribo-Seq datasets from *S. cerevisiae* generated by Gardin et al., named 'Sc-Lys', 'Sc-His', 'YPD1' and 'YPD2' in their paper[26]. While mapping 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 the National Center for Biotechnology Information (NCBI)[27]. Intriguingly, a high proportion of these 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 phenotype[28], while the phenotypic impact of L-BC on the host is unknown[29]. In light of the numerous viral Ribo-Seq reads found in the pilot analysis, we simultaneously remapped all Ribo-Seq reads from the experiment of Gardin et al. to the yeast genome and two viral genomes. We then estimated translational activity for each gene by averaging the number of Ribo-Seq reads per site for that gene (Supplementary Fig. 1). Note that this metric approximates the translational activity per gene, not per messenger RNA molecule, because it reflects the abundance of ribosome-protected fragments without dividing by mRNA abundance, and is corrected for gene length (see Methods). As a comparison, we analysed five previously published yeast Ribo-Seq datasets[30–34] (Supplementary Table 1). We found that the ratio between 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 ranged 0.275–0.930 in the other five yeast datasets (Fig. 1a and Supplementary Fig. 1). In each of the datasets of Gardin et al., the translational activity of the most active viral gene was higher than that of 96% of yeast endogenous genes. In principle, such excessive translation of exogenous genes could deplete the host tRNA supply[17] and hence deserves further scrutiny. In the following analysis we concentrated on the Sc-Lys dataset, because the original research[26] also focused on this dataset.

To evaluate the impact of viral protein translation on yeast tRNA supply, we calculated total consumption for 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, codon consumption was the sum of codon consumption by each viral gene, which was calculated as the number of occurrences of the codon in the gene multiplied by translational activity of the gene; 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). 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 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 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 the 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, and thus differential tRNA shortages and corresponding changes in the decoding time of codons when viral expression is sufficiently high (see 'Experimental assessment of the general role of virus CUB in host translation' 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 analysed (Supplementary Fig. 2), and indicated that codons with marked 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 are thus compatible with a role of virus CUB in *trans*-regulating host translation through differential tRNA depletion when 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 in codon decoding times between experiments should be correlated with differences in relative tRNA shortage due to viral translation. This was 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 the 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 excludes rare and extreme ribosomal pauses[35] (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 small viral genome size or interdependence between the parameters analysed (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 is ready for translation[17]. We therefore recalculated relative tRNA shortage due to viral translation, with tRNA supply estimated by codon consumption of yeast endogenous genes (see Methods) rather than 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 non-negligible *trans*-regulatory effect of virus CUB on host translation when the expression of virus is sufficiently high.

Impact of viral translation on influenza-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 influenza A virus (IAV)-infected A549 cells[36], a human lung adenocarcinoma cell line. Similar to that of viral genes in yeast, the average translational activity of IAV genes was ~300-fold higher than that of human endogenous genes, with that of the most active IAV gene higher than that of all human endogenous genes. Consistent with the results for yeast, 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;

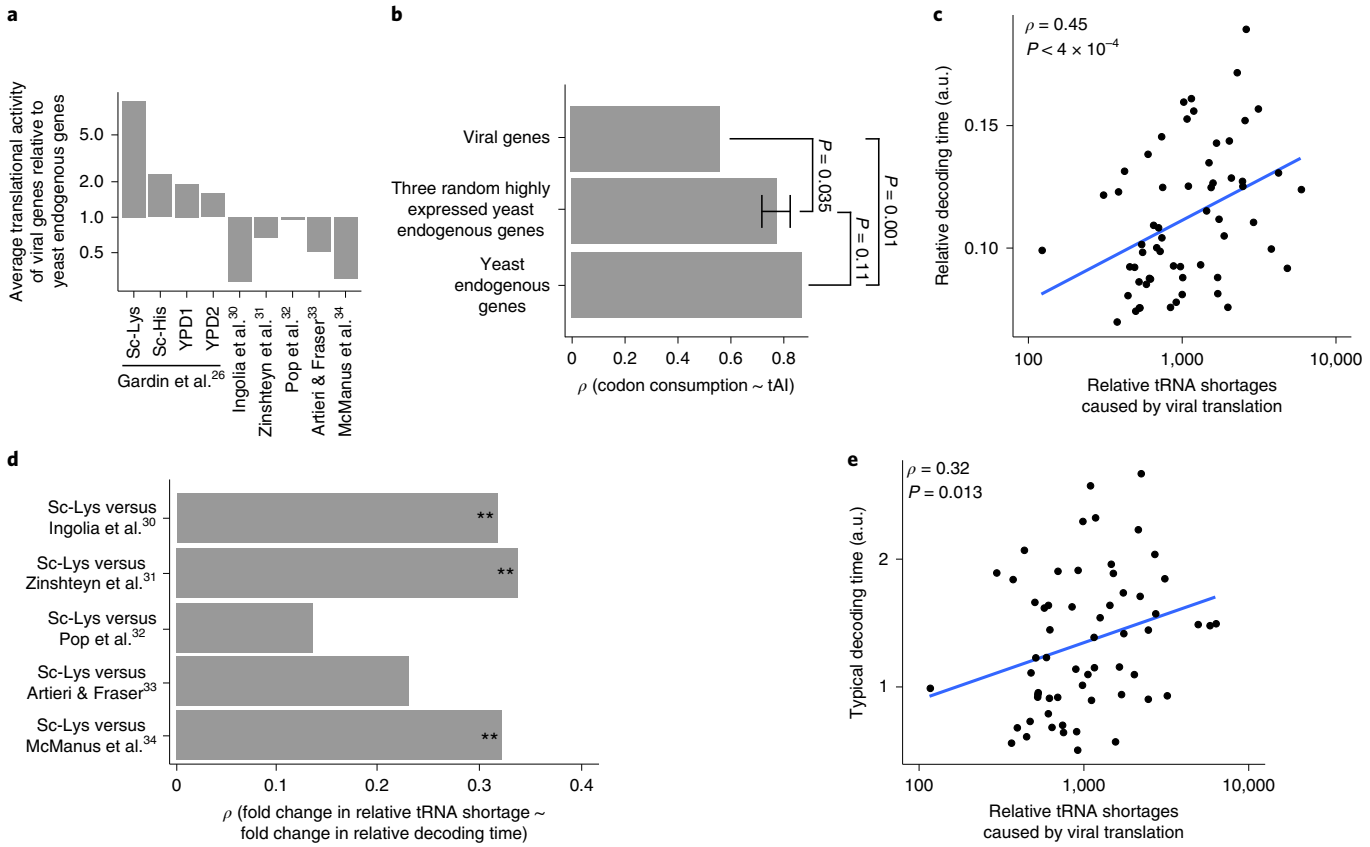

Fig. 1 | Codon usage of viral genes with excessive translation impacted the decoding time of codons in yeast. a, The average translational activity of viral genes was higher than that of yeast endogenous genes in the four datasets generated by Gardin et al.[26], but not in other yeast Ribo-Seq datasets examined. **b**, Spearman's rank correlation between tAI and viral codon consumption is significantly weaker than that between tAI and 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[30–34]. Nominal *P* value for Spearman's rank correlations, $^{**}P < 0.01$. **e**, Typical decoding time is significantly correlated with relative tRNA shortages caused by viral translation. **c**,**e**, Each dot represents one codon, blue lines indicate fitted linear models and Spearman's rank correlations are shown. **b**–**e**, All statistical tests based on 61 codons.

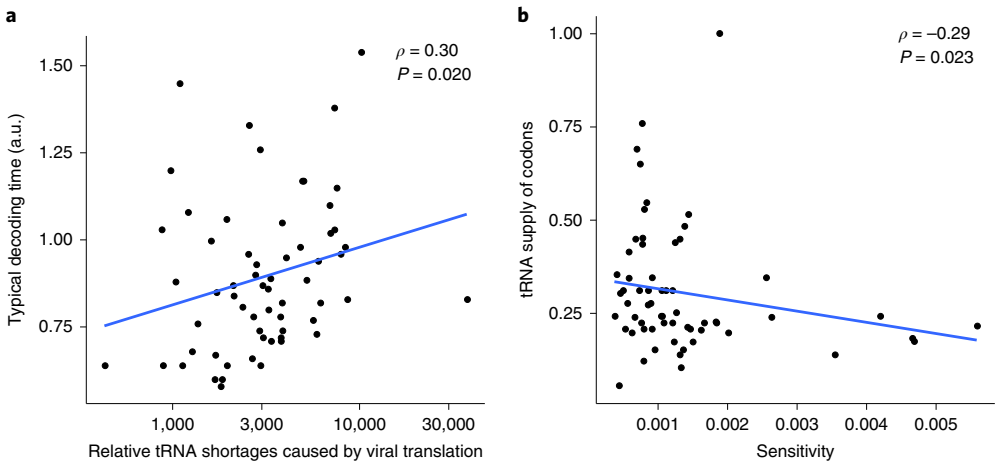

Fig. 2 | Impact of IAV codon usage on decoding time in infected human cells. a, Spearman's rank correlation between typical decoding time and relative tRNA shortage due to viral translation is shown for human A549 cells infected by IAV. **b**, The sensitivity of typical decoding time was calculated as fold change in typical decoding time between virus-infected and -free cells, divided by viral codon consumption in infected cells. Sensitivity of the 61 sense synonymous codons is negatively correlated with their tRNA supply, which was approximated by tAI. **a**,**b**, Each dot represents one of the 61 codons, blue lines indicate fitted linear models and Spearman's rank correlations are shown.

see also Supplementary Fig. 5). To further confirm our findings, we estimated the sensitivity of typical decoding time to tRNA depletion due to viral translation by the ratio between fold change in typical decoding time and 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 is 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 disruption to the shared tRNA pool due to excessive viral translation. The two observations made with influenza-infected human cells cannot be explained by the small size of the virus genome or interdependence between parameters (Supplementary Fig. 5). Overall, our observations in influenza-infected human cells support the general effect of virus CUB on host translation when virus expression is sufficiently high.

Experimental assessment of the general role of virus CUB in 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 are considered together, individual codons could be decelerated to different levels. Since translation of an mRNA is a serial process, the codon with the longest 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 exogenous genes was fixed (but see the paragraph after the next for when it was not fixed), exogenous genes with lower 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 cellular translational efficiency (Fig. 3a; Methods). A total of 37 synonymous versions of mCherry were designed with D_P different to the host (yeast) genome (Supplementary Table 2).

We first investigated the *trans*-regulatory effect of 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 maximal YFP signal detected among all strains. Only cells whose mCherry expression level is in the top 20% among all cells were considered, to ensure observable impact on host translation by CUB and to limit 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 translational load and mCherry expression level. This ratio was found to be 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 the host cell. These observations are 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 of sufficiently high expression influenced codon decoding time.

Our experiment also allowed us to determine the minimum expression level at which exogenous genes would exhibit significant impact on host translational efficiency and thus display a significant positive correlation between translational load and mCherry D_P. 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 8–19 strains with different mCherry CUBs and a very narrow range of mCherry expression, thereby effectively controlling it. We then calculated within-group Spearman's rank correlations between translational load and D_P, which were found significantly positive in the top 43 groups (Supplementary Fig. 8). Further analyses, based on quantitative PCR with reverse transcription (RT–qPCR) of mCherry versus actin and bulk RNA-sequencing (RNA-seq) data (see Methods), suggested that the minimum expression of mCherry in these 43 groups approximately ranked as the 208th most highly expressed endogenous genes (Supplementary Fig. 8). Intriguingly, in the yeast Ribo-Seq data we analysed, 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 306th to 722nd for the other five datasets, suggesting that the killer virus was indeed of sufficiently high expression level to affect translational efficiency in the majority of strains used by Gardin et al.[26], but not in strains used in the other five studies.

Nevertheless, besides *trans*-regulation of the overall translational efficiency of the host cell via tRNA availability, CUB of exogenous genes is also known to *cis*-regulate expression levels of the exogenous gene itself[15,37,38]. Specifically, assimilation of CUB of the exogenous gene to that of the host was found capable of increasing the expression level of the exogenous gene[15]. This *cis*-regulatory effect of CUB on the expression level of the exogenous gene was confirmed by our experiment, because 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[37,38] (see Supplementary Table 3 and Supplementary Fig. 6), and is consistent with the known strategy of improving expression of exogenous genes by adjusting their CUB to mimic that of the host[15,39,40].

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 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*- and *trans*-regulatory effects 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 mCherry CUB with translational load as probed by YFP expression. We observed a negative correlation between mCherry D_P 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 for the 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, differs from the green-shaded results in Fig. 3b, in which the expression level of the exogenous genes was controlled. This result was not caused by reduced YFP mRNA

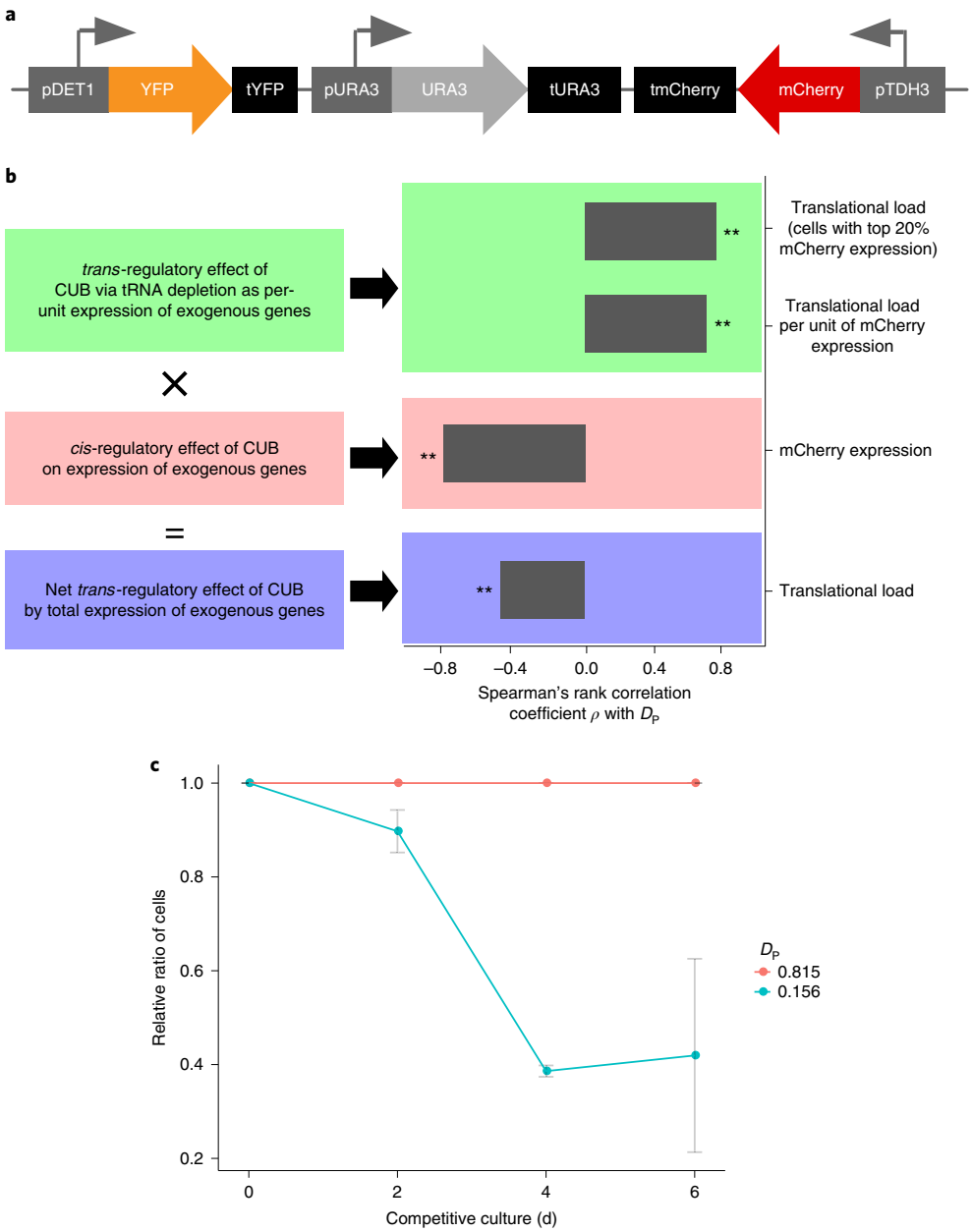

Fig. 3 | Manipulative experiments in yeast elucidate the regulatory effects of CUB on host translation. a, Reporter cassette used for examining the impact of codon usage of exogenous genes on host translational efficiency. The exogenous gene, mCherry, which is driven by a strong promoter (pTDH3), imposes 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. Promoters and terminators denoted by dark grey and black boxes, respectively; ORFs of YFP, URA3 and mCherry denoted by orange, grey and red arrows, respectively. **b**, Regulatory effects of CUB on 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 in YFP signal relative to maximal YFP signal detected among all strains. Spearman's rank correlation coefficient between D_P and various metrics as indicated on the figure is shown. Nominal *P* values for Spearman's rank correlations are indicated: $^{**}P < 0.01$. Sample, $n = 37$ strains except for cells with top 20% mCherry expression, which contained only 15 strains. **c**, Two strains with mCherry of high and low codon usage similarity to yeast ($D_P = 0.156$ and 0.815, respectively) were assayed for relative fitness in a competitive culture. Relative population sizes of different strains were determined by high-throughput sequencing of strain-specific regions of mCherry. The ratio between relative population size at day *x* and that at day 0 was further divided by this ratio of the fitter strain. Error bars indicate standard error as assessed by three biological replicates.

abundance or lowered cellular transcriptional efficiency, because the abundance of YFP mRNA as assessed by RT–qPCR is not correlated with mCherry D_P (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 resulting from changes in tRNA supply, because the translational error rate measured in these strains was 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

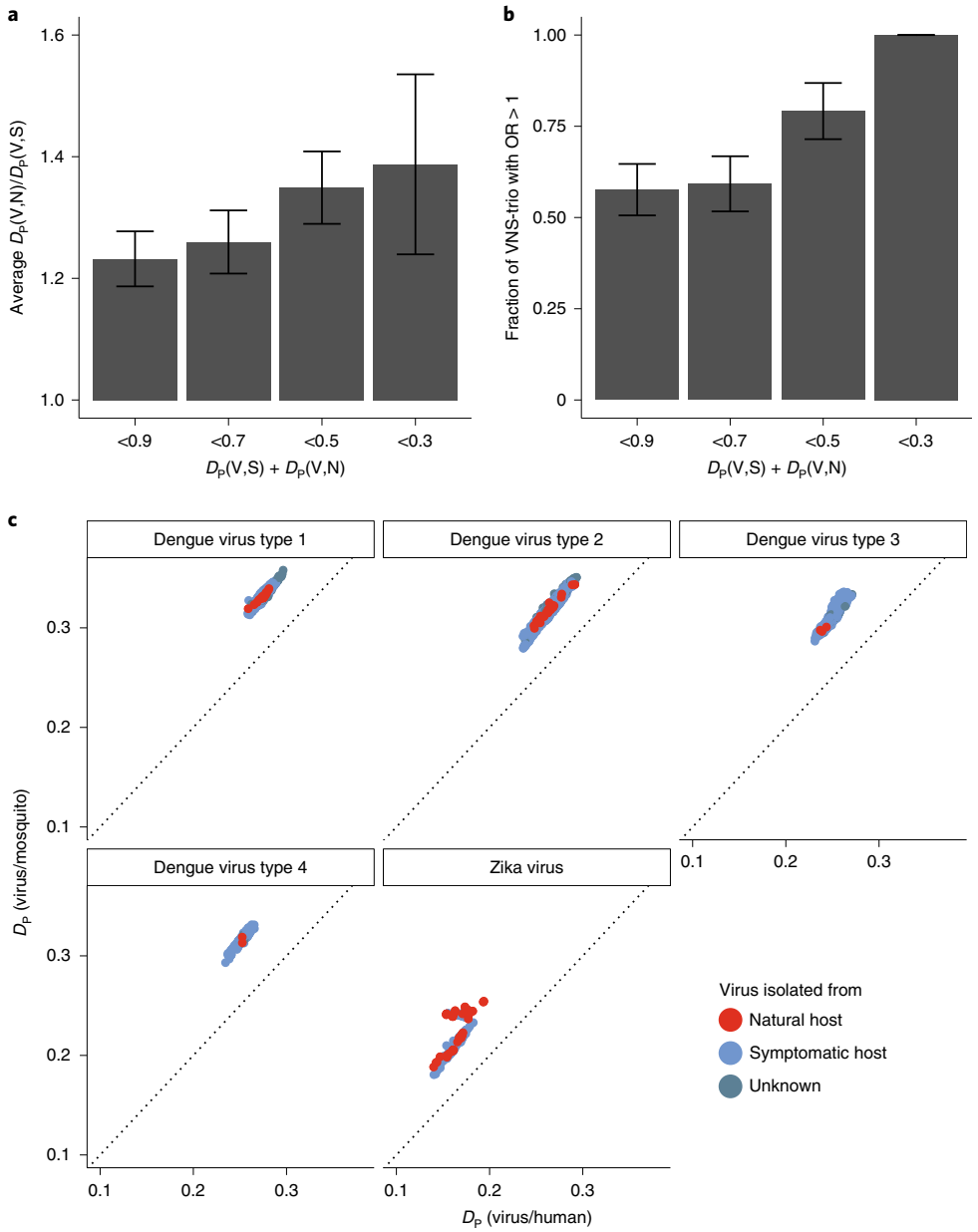

Fig. 4 | CUB similarity between virus, natural host and symptomatic host. a, CUB similarity between virus and symptomatic or natural host was approximated by D_P(V,S) or D_P(V,N), respectively. VNS-trios with D_P(V,S) + D_P(V,N) lower than certain thresholds (*x* axis), which were more likely to be subject to stronger translational selection, were used to calculate average D_P(V,N)/D_P(V,S) (*y* axis). Error bars indicate the standard error. **b**, Similar to **a**, except that odds ratio (OR) 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—that is, fraction of VNS-trios with OR > 1 (*y* axis). Error bars indicate the standard error, estimated by bootstrapping VNS-trios 1,000 times. **c**, CUB 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). Virus CUB was always more similar to the symptomatic host than to the natural host, as the points all lie 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 point colours.

mCherry of lower D_P can be explained, partially at least, by increased mCherry mRNA abundance (Supplementary Fig. 6), highlighting the role of CUB in translational load via *cis*-regulation of mRNA abundance of exogenous genes.

To further confirm the fitness effect of virus CUB, we chose two strains with mCherry of high and low codon usage similarity to yeast ($D_P = 0.156$ and 0.815, respectively) and estimated their relative fitness by competitive coculture (see Methods). As a result, we found that the strain expressing mCherry with lower D_P exhibited significant growth deficiency relative to that expressing mCherry with higher 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 by translation of the exogenous gene, the concomitant effect of increasing expression of the exogenous gene dominated the way in which CUB of the exogenous gene influenced host translation. Consequentially, cells expressing exogenous genes with excessive CUB similarity to the host display fitness disadvantage due to translational selection. Therefore, our results support the repulsion between virus and host CUB when they

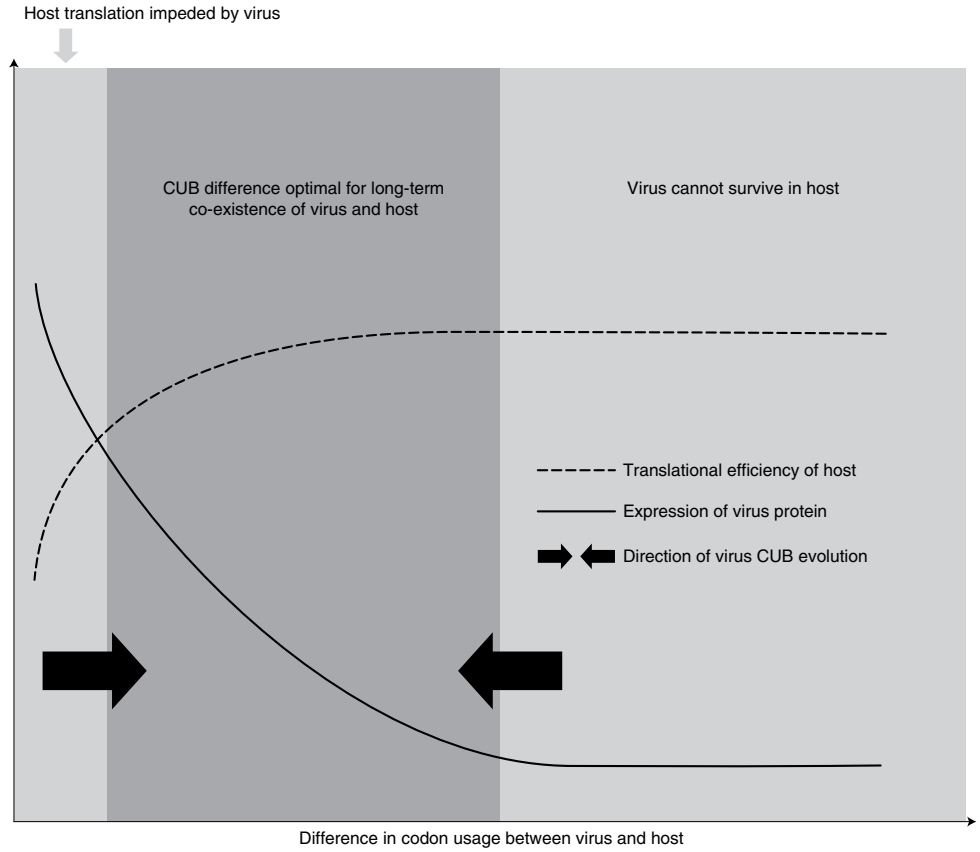

Fig. 5 | Schematic diagram of the regulatory role of virus CUB and its evolutionary implication. As the difference in codon usage between virus and host (*x* axis) decreased, expression of virus protein (solid curve) increased due to the *cis*-regulatory effect of virus CUB; meanwhile, translational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

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 favoured.

Virus CUB tends to be more similar to CUB of a symptomatic host than that of an asymptomatic natural host. Our manipulative experiment in yeast suggested that exogenous genes with CUB overly 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 (that is, be expressed in the host cell), hosts with CUB more similar to that of the virus would be more likely to show symptoms compared to other hosts since translation in these hosts is 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 few observable symptoms during any stage of 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 viruses covered DNA and RNA types, with hosts ranging from insects to mammals. We then calculated 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 virus CUB overly similar to that of the host. We further reasoned that viruses with minimal deviation from the proportionality rule of CUB are probably subject to stronger selection for translational efficiency. Therefore, we calculated the average ratio between D_P(V,N) and D_P(V,S) for VNS-trios whose $D_P(V,N) + D_P(V,S)$ was lower than a certain threshold. As we lowered this threshold, the average ratio between D_P(V,N) and D_P(V,S) increased (Fig. 4a). To corroborate this result, we calculated an odds ratio 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 odds ratio (relative to 1) the more similar the virus CUB is to the symptomatic host CUB. Consistent with previous results, we found that for VNS-trios probably subject to stronger translational selection, the fraction of VNS-trios with odds ratio > 1 increased (Fig. 4b).

To further evaluate CUB patterns among VNS-trios, we focused on Dengue and Zika viruses for which multiple sequenced variants allowed additional analyses. Intriguingly, all variants of Dengue and Zika viruses downloaded from NCBI appeared to have CUB 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 virus strains were not isolated from the symptomatic host, meaning that the observed similarity between virus and symptomatic host cannot be explained by rapid CUB assimilation to the latter. Collectively our observations of VNS-trios support the proposed role of CUB in virus–host coevolution.

Discussion

In the current study, we analysed 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 the exogenous gene and elevated the expression level of the exogenous gene, such that overall tRNA shortage became severed. These results suggest that viruses with CUB overly similar to that of their hosts might impede host translation. This role of virus CUB is further supported by the observation that it tends to be more similar to that of symptomatic hosts than that of natural hosts. Our results highlight 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 of viral origin in a yeast Ribo-Seq dataset[26]. This finding is in contrast to five other yeast Ribo-Seq datasets (Supplementary Table 1) in which viral reads are much rarer, probably explained by inactive infections or contaminations. Consistent with previous experimental measures of translational efficiency[17], 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 probably 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, detailed molecular mechanisms are probably more complex than this dichotomy. For example, the *cis*-regulatory effect of CUB on expression of exogenous genes could have occurred via direct regulation of mRNA abundance[37], which is supported by our data (Supplementary Fig. 6), or by modulation of translational elongation rate[10,41]. It will therefore be interesting to investigate further 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 was assumed to be the result of genetic drift or mutation bias[42]. Our findings here suggest 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 CUB of 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—for example, pairs involving a natural host/reservoir. Indeed, the duration of virus–host coexistence is generally longer for natural 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 that of the symptomatic host. The observation that 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 suggests that 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 suggest an attraction–repulsion relationship between virus and host CUB, a previously unrecognized complexity in virus–host coevolution.

Our findings also have potential practical implications. In viral epidemiology, synonymous mutations have been shown to be 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 modification of 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, v.R64-1-1) were downloaded from SGD[43]. Genomic sequences and annotations of the yeast viruses L-A (GenBank accession no. NC_003745) and L-BC (GenBank accession no. U01060) were downloaded from NCBI GenBank under accession nos. NC_003745 and U01060, respectively. The 4,579-base pair (bp) genome of L-A contains two open reading frames (ORFs) that are translated either separately or combined as a fusion protein[44]. Because only 6.2% of sites in the ORFs are discriminatory between one fused and two separated ORFs, these two states are difficult to distinguish. In this paper we arbitrarily focused on the fused ORF, but consideration of the two separate ORFs did not alter our conclusions. The 4,615-bp genome of L-BC contains two ORFs[45]. The human genome sequence (GRCh38) and gene annotations were downloaded from EnsEMBL r85 (ref. [46]). Sequences of IAV genes were downloaded under accession nos. NC_002016–NC_002023 from NCBI GenBank. For yeast, we reanalysed the Ribo-Seq data used by Gardin et al.[26] and found excessive viral translation. As a comparison, we analysed 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 derived from previous mRNA-seq-based estimates[47]. For influenza infection in human cells, both Ribo-Seq and mRNA-seq datasets were derived from GSE82232. To concentrate on the tRNA-depletion effect of viral expression, only the uninfected (SRR3623932) and 2-h post-infection (SRR3623937) datasets were analysed here, because decoding time at later post-infection time points appeared skewed by the progression of host shut-off[26].

Ribo-Seq datasets from experimental repeats under the same conditions were combined. The four datasets from Gardin et al. had different experimental settings[26,48] and thus were not combined. Attached adaptors as noted from the respective reports[26,30] were removed by allowing minor mismatches due to sequencing errors. Any trimmed reads shorter than 20 or longer than 40 nucleotides were discarded. The remaining reads were first aligned to rRNA and tRNA transcripts in yeast, and unmapped reads were aligned to the yeast genome by Bowtie2 (ref. [49]). Ten thousand unalignable reads were randomly selected and BLASTed against the NCBI nr database to identify their origins[27]. Having noticed a significant fraction of reads from two yeast viruses, we combined the viral and yeast genomes 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 Tuller[35] to first record all uniquely mapped reads and then randomly assigned multi-hit 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 that location. Discarding reads with multiple hits did not change our results qualitatively. Also, ambiguous mappings on viral genomes are negligible because no commonly known repetitive elements were found by RepeatMasker[50] on the viral genome, and only <3% of reads mapped to the viral genome can be mapped to yeast genome with equal alignment score. Finally, each mapped Ribo-Seq read was added to the ribosome count of one codon site in the ORFeome. Downstream estimation of decoding times was conducted following either the procedure of Gardin et al.[26] or Dana and Tuller's Exponentially Modified Gaussian (EMG)-based method[35] (see 'Estimation of translational activity, codon consumption, tAI and relative/typical decoding times of codons' 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' effects[51] 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 pauses[35] 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 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.

Relative tRNA shortage due to viral translation of a codon was then calculated as the ratio between its total consumption in 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,

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

where n_i is the number of tRNA iso-acceptors recognizing the ith codon, tGCN_{ij} is the gene copy number of the jth tRNA that recognizes the ith codon and s_{ij} is a selective constraint on the efficiency of the codon–anticodon coupling. Additionally, W_i values can readily be produced by considering Crick's wobble rules[52]. The s_{ij} values for eukaryotes were used as in a previous study[35]. Copy numbers of tRNA genes were retrieved from the Genomic tRNA Database[53].

We then 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, and if Ribo-Seq data were available for the estimation of translational activity, this was calculated as described in the foregoing paragraph. If Ribo-Seq data were not available for the estimation of translational activity, we used transcriptome abundance of the gene to approximate it (that is, 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 footprint[35]. This was achieved by determining the correct reading frame followed by identification of the most likely offset given by the reading frame. Finally, 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-negative binomial distribution. The main difference from the original EMG-based typical decoding time[35] was the incorporation of a negative binomial 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 mCherry[17]. These coding sequences were designed to have CUB with variable similarity 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 for 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 needed to quantify the D_p value between codon usage 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:

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

where 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 and X_{ij} is tRNA supply represented by either the fraction of codon j among the synonymous codons in the host transcriptome[54] or rescaled tAI values such that all tAIs within a group of synonymous codons sum 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 transcriptomic codon usage was valid only 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 codon adaptation index[55], index of translation elongation (I_{TE})[10] or fraction of optimal codons[56], are 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 exogenous genes are highly expressed. For VNS-trios analyses (see 'Codon usage analysis of virus, natural host and symptomatic host triples' below), in which the reliability of annotations in most species remains questionable, we resorted to genomic codon usage of the 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 ideal[17], such approximation of transcriptomic CUB by genomic CUB has been successful in many previous studies[57–60]. Additionally, our test for CUB similarity in VNS-trios (Mantel–Haenszel test based on 2 × 2 contingency tables; see 'Codon usage analysis of virus, natural host and symptomatic host triples' below) should not be sensitive to the minor difference between genomic and transcriptomic codon usage.

We aimed to test the *trans*-regulatory role of the CUB of exogenous genes in host translational efficiency using manipulative experiments in *S. cerevisiae*. To focus our experiment on tRNA depletion due to translation of the exogenous/viral gene and to exclude other translation-interfering mechanisms that are specific to any virus (such as host shut-off by IAV[36]), 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[7,19,61]. 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, 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 was derived from previously used mCherry reporters[17] and their modified versions. We designed four primers for each modified version of mCherry. The first and second 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 DNA fragments of mCherry and YFP were then cloned into the HO locus of BY4741 by homologous recombination using the lithium acetate protocol[62], 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 centred at 610 nm, and the that of YFP was measured by a filter with a 40-nm bandpass centred at 525 nm. Yeast cells with mCherry and YFP fluorescence signals tenfold greater than those of BY4741 negative control cells were kept for later analyses. We retrieved forward scatter (FSC, which is proportional to cell size) and mCherry and YFP fluorescence signals for all cells. Expression levels of fluorescent proteins were defined as their fluorescence signals divided by FSC. All experiments were carried out with three biological and three technological replicates.

For RT–qPCR, total RNA samples were isolated from the log phase of all strains in YPD media using the RNeasy Plus Mini Kit (Qiagen, no. 74134) according to the manufacturer's instructions. Complementary DNAs were reversely transcribed from RNAs using the Evo M-MLV RT Kit with gDNA Clean for qPCR (Takara, no. AG11705) according to the manufacturer's instructions. Next, RT–qPCR was carried out with cDNAs and primers (Supplementary Table 7) using the QuantiNova SYBR Green PCR kit (Qiagen, no. 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 had ~500,000 cells from 8–19 strains with mCherry of varying CUB. We found that in the top 43 groups with high mCherry expression levels, within-group Spearman's rank correlations between the translational load and D_p were significantly positive (Supplementary Fig. 8a), suggesting a significant impact of exogenous gene CUB on overall translational efficiency. We performed RT–qPCR in strains with at least 1 million cells in these 43 groups, and determined that the abundance of mCherry mRNA in these strains was 0.7–3.4-fold that of actin (Supplementary Table 3). According to the RNA-seq-based transcriptome profile of bulk yeast cells[47], 0.7-fold of the expression level of actin corresponds to the 208th most 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 are 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 proteome-wide translation efficiency was subsequently reduced[21]. Moreover, rescue of translation efficiency by increasing tRNA supply for CGG further supported the proposed mechanism of tRNA shortage[21]. Interestingly, unlike our observations that synonymous changes in mCherry *cis*-regulated its expression (Fig. 3b, red-shaded area), codon conversion in the previous study reduced only one of the eight converted genes[21]. We speculated that the lack of significant expression change in synonymously converted genes could 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 CUB and therefore 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 the present study, resulting in greater changes in CUB and therefore in the expression of mCherry.

We performed competitive coculture assays to measure the relative fitness of two chosen strains (D_p = 0.156 and 0.815, respectively). Both strains were first inoculated in 4 ml of YPD competition culture at 30 °C with shaking at 200 r.p.m. for 16 h. Next, 2 ml of each sample was left for DNA extraction and marked as day 0 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 100-ml culture to reach a concentration of $2–5 \times 10^5$ cells ml^{-1}. The cells were then replaced at 30 °C for 10–14 h, such that total cell density values were maintained >0.7 to maintain exponential growth throughout the experiment. We repeated this procedure for 6 d, resulting in one day 0 sample

and three for each sample of days 2, 4 and 6. We then extracted DNA and amplified strain-specific regions of mCherry sequences of both strains with the primers listed in Supplementary Table 8. The ten amplicon samples were mixed together and paired-end sequenced for 150 nt on either end on an Illumina NovaSeq platform, for a total of 1 giga base pairs raw data. Then, the data of ten samples were split using the 5-bp index sequence at the 5′ end of left primers. The paired reads were mapped to the two versions of mCherry using STAR[63]. 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 days 2/4/6 and that at day 0 was further divided by this 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[41,64] to estimate the translational error rate in 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 determining the ratio between the observed firefly and Renilla activities. We used both wild-type 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 does the Lys side chain[64], the mutant can display firefly activity only when it is mistranslated as Lys at position 529. In other words, concentration-independent firefly activity relative to that of the wild-type firefly measures the translational error rate[41,64]. We successfully measured the translational error rate in 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 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/)[65], 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 that the natural hosts were symptom-free or at least that their known symptoms were significantly milder than those of 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 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 shown in Table 1:

Table 1 | Contingency table

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

In Table 1, *V*, *N* and *S* are the frequencies of a codon in the virus, natural host and symptomatic host, respectively. Therefore, $V_i > V_{mean}$ denotes that a codon was a preferred codon in the virus while $V_i < V_{mean}$ denotes that it was unpreferred in the virus; $|V_i - N_i| > |V_i - S_i|$ means that virus codon frequency was more similar to that of the symptomatic host than that of the natural host, while $|V_i - N_i| < |V_i - S_i|$ means that virus codon frequency was more similar to that of the natural host than 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 acids[24] was excluded. Finally, a common odds ratio was calculated by combining the relevant contingency tables using the Mantel–Haenszel procedure implemented in R. An odds ratio > 1 indicates that virus CUB tends to be more similar to that of the symptomatic host than that of the natural host, while an odds ratio < 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 transcriptomes of 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 our VNS-trio analyses. First, during the viral infection process, survival of the virus and phenotypic outcome of the infection will be affected by the host immune system, a factor we addressed neither in our manipulative experiment in yeast nor in our VNS-trio analyses. One possibility is that all symptoms of infection are caused by a strong immune system reaction in symptomatic hosts rather than 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 the symptomatic host, but not translational disruption. However, this alternative hypothesis does not predict that phenotypic consequences will intensify when translational selection is stronger (Fig. 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, excessive CUB similarity between the exogenous gene and the host is still detrimental to host fitness. These findings suggested that translational selection at least has a non-negligible 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 determined 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 selection[66]) and mutation bias. The relative contribution of these three factors might not be assessable until further quantitative data from specific virus–host pairs become available. Our study, while not rejecting the role of genetic drift and mutation bias, offers unequivocal support for the contribution of CUB repulsion between virus and host. For example, our study showed decelerated codon decoding time (Figs. 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 the host cell providing that viral expression is sufficiently high, thereby forming a mechanical basis for CUB repulsion between virus and 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[67,68] and differential mutation bias in double- versus single-stranded DNA viruses[68,69]. Last but not least, our model did not exclude other mechanisms adjusting viral expression to avoid impediment of host translation, such as changes in mRNA levels due to mutations in promoters or other regulatory elements. Rather, our results highlight the fact that natural selection on viral expression is unlikely to be unidirectional because various regulatory elements, including CUB and promoter sequences, collectively determine viral expression.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

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

Code availability

Custom R codes were used in data analysis and are available at Github (https://github.com/chenfengokha/CUB).

Received: 6 September 2019; Accepted: 21 January 2020;
Published online: 02 March 2020

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).
2. Xia, X. Maximizing transcription efficiency causes codon usage bias. *Genetics* **144**, 1309–1320 (1996).
3. Bulmer, M. The selection-mutation-drift theory of synonymous codon usage. *Genetics* **129**, 897–907 (1991).
4. Hershberg, R. & Petrov, D. A. Selection on codon bias. *Annu. Rev. Genet.* **42**, 287–299 (2008).
5. Gouy, M. & Gautier, C. Codon usage in bacteria: correlation with gene expressivity. *Nucleic Acids Res.* **10**, 7055–7074 (1982).
6. Xia, X. How optimized is the translational machinery in *Escherichia coli*, *Salmonella typhimurium* and *Saccharomyces cerevisiae*? *Genetics* **149**, 37–44 (1998).

7. Tuller, T., Waldman, Y. Y., Kupiec, M. & Ruppin, E. Translation efficiency is determined by both codon bias and folding energy. *Proc. Natl Acad. Sci. USA* **107**, 3645–3650 (2010).
8. Robinson, M. et al. Codon usage can affect efficiency of translation of genes in *Escherichia coli*. *Nucleic Acids Res.* **12**, 6663–6671 (1984).
9. Sorensen, M. A., Kurland, C. G. & Pedersen, S. Codon usage determines translation rate in *Escherichia coli*. *J. Mol. Biol.* **207**, 365–377 (1989).
10. Xia, X. A major controversy in codon–anticodon adaptation resolved by a new codon usage index. *Genetics* **199**, 573–579 (2015).
11. Akashi, H. Synonymous codon usage in *Drosophila melanogaster*: natural selection and translational accuracy. *Genetics* **136**, 927–935 (1994).
12. Stoletzki, N. & Eyre-Walker, A. Synonymous codon usage in *Escherichia coli*: selection for translational accuracy. *Mol. Biol. Evol.* **24**, 374–381 (2006).
13. Johnston, T. C., Borgia, P. T. & Parker, J. Codon specificity of starvation induced misreading. *Mol. Gen. Genet.* **195**, 459–465 (1984).
14. Johnston, T. C. & Parker, J. Streptomycin-induced, third-position misreading of the genetic code. *J. Mol. Biol.* **181**, 313–315 (1985).
15. Plotkin, J. B. & Kudla, G. Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet.* **12**, 32–42 (2010).
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).
17. Qian, W., Yang, J. R., Pearson, N. M., Maclean, C. & Zhang, J. Balanced codon usage optimizes eukaryotic translational efficiency. *PLoS Genet.* **8**, e1002603 (2012).
18. Akashi, H. & Eyre-Walker, A. Translational selection and molecular evolution. *Curr. Opin. Genet. Dev.* **8**, 688–693 (1998).
19. Kudla, G., Murray, A. W., Tollervey, D. & Plotkin, J. B. Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* **324**, 255–258 (2009).
20. Shah, P., Ding, Y., Niemczyk, M., Kudla, G. & Plotkin, J. B. Rate-limiting steps in yeast protein translation. *Cell* **153**, 1589–1601 (2013).
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).
22. Tian, L., Shen, X., Murphy, R. W. & Shen, Y. The adaptation of codon usage of +ssRNA viruses to their hosts. *Infect. Genet. Evol.* **63**, 175–179 (2018).
23. Albers, S. & Czech, A. Exploiting tRNAs to boost virulence. *Life (Basel)* **6**, 4 (2016).
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).
25. Lucks, J. B., Nelson, D. R., Kudla, G. R. & Plotkin, J. B. Genome landscapes and bacteriophage codon usage. *PLoS Comput. Biol.* **4**, e1000001 (2008).
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 (2009).
28. Schmitt, M. J. & Breinig, F. The viral killer system in yeast: from molecular biology to application. *FEMS Microbiol. Rev.* **26**, 257–276 (2002).
29. Ribas, J. C. & Wickner, R. B. *Saccharomyces cerevisiae* L-BC double-stranded RNA virus replicase recognizes the L-A positive-strand RNA 3′ end. *J. Virol.* **70**, 292–297 (1996).
30. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S. & Weissman, J. S. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218–223 (2009).
31. Zinshteyn, B., Gilbert, W. V. & Copenhaver, G. P. Loss of a conserved tRNA anticodon modification perturbs cellular signaling. *PLoS Genet.* **9**, e1003675 (2013).
32. Pop, C. et al. Causal signals between codon bias, structure, and the efficiency of translation and elongation. *Mol. Syst. Biol.* **10**, 770 (2014).
33. Artieri, C. G. & Fraser, H. B. Accounting for biases in riboprofiling data indicates a major role for proline in stalling translation. *Genome Res.* **24**, 2011–2021 (2014).
34. McManus, C. J., May, G. E., Spealman, P. & Shteyman, A. Ribosome profiling reveals post-transcriptional buffering of divergent gene expression in yeast. *Genome Res.* **24**, 422–430 (2014).
35. Dana, A. & Tuller, T. The effect of tRNA levels on decoding times of mRNA codons. *Nucleic Acids Res.* **42**, 9171–9181 (2014).
36. Bercovich-Kinori, A. et al. A systematic view on influenza induced host shutoff. *eLife* **5**, e18311 (2016).
37. 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).
38. Presnyak, V. et al. Codon optimality is a major determinant of mRNA stability. *Cell* **160**, 1111–1124 (2015).
39. Mignon, C. et al. Codon harmonization—going beyond the speed limit for protein expression. *FEBS Lett.* **592**, 1554–1564 (2018).
40. Angov, E., Legler, P. M. & Mease, R. M. in *Heterologous Gene Expression in E. coli. Methods in Molecular Biology (Methods and Protocols)* vol. 705 (eds Evans Jr, T. & Xu, M. Q.) 1–13 (Humana Press, 2011).
41. Yang, J. R., Chen, X. & Zhang, J. Codon-by-codon modulation of translational speed and accuracy via mRNA folding. *PLoS Biol.* **12**, e1001910 (2014).
42. Jenkins, G. M. & Holmes, E. C. The extent of codon usage bias in human RNA viruses and its evolutionary origin. *Virus Res.* **92**, 1–7 (2003).
43. Cherry, J. M. et al. Saccharomyces genome database: the genomics resource of budding yeast. *Nucleic Acids Res.* **40**, D700–D705 (2012).
44. Icho, T. & Wickner, R. B. 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).
45. Bruenn, J. A. A closely related group of RNA-dependent RNA polymerases from double-stranded RNA viruses. *Nucleic Acids Res.* **21**, 5667–5669 (1993).
46. Zerbino, D. R. et al. Ensembl 2018. *Nucleic Acids Res.* **46**, D754–D761 (2018).
47. Nagalakshmi, U. et al. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320**, 1344–1349 (2008).
48. 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).
49. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357 (2012).
50. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics* **25**, 4.10.1–4.10.14 (2009).
51. Tuller, T. et al. An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* **141**, 344–354 (2010).
52. Crick, F. H. C. Codon–anticodon pairing: the wobble hypothesis. *J. Mol. Biol.* **19**, 548–555 (1966).
53. Chan, P. P. & Lowe, T. M. GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res.* **37**, D93–D97 (2008).
54. Holstege, F. C. et al. Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* **95**, 717–728 (1998).
55. Sharp, P. M. & Li, W. H. The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* **15**, 1281–1295 (1987).
56. Drummond, D. A. & Wilke, C. O. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* **134**, 341–352 (2008).
57. 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).
58. 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).
59. Jia, X. et al. Non-uniqueness of factors constraint on the codon usage in *Bombyx mori*. *BMC Genomics* **16**, 356 (2015).
60. 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).
61. Gu, W., Zhou, T. & Wilke, C. O. A universal trend of reduced mRNA stability near the translation-initiation site in prokaryotes and eukaryotes. *PLoS Comput. Biol.* **6**, e1000664 (2010).
62. Gietz, R. D. in *Yeast Genetics: Methods and Protocols* (eds Smith, J. S. & Burke, D. J.) 1–12 (Springer, 2014).
63. Dobin, A. & Gingeras, T. R. Mapping RNA-seq reads with STAR. *Curr. Protoc. Bioinformatics* **51**, 11.14.11–11.14.19 (2015).
64. Kramer, E. B., Vallabhaneni, H., Mayer, L. M. & Farabaugh, P. J. A comprehensive analysis of translational missense errors in the yeast *Saccharomyces cerevisiae*. *RNA* **16**, 1797–1808 (2010).
65. Hulo, C. et al. ViralZone: a knowledge resource to understand virus diversity. *Nucleic Acids Res.* **39**, D576–D582 (2011).
66. 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).
67. Bailly-Bechet, M., Vergassola, M. & Rocha, E. Causes for the intriguing presence of tRNAs in phages. *Genome Res.* **17**, 1486–1495 (2007).
68. Chithambaram, S., Prabhakaran, R. & Xia, X. Differential codon adaptation between dsDNA and ssDNA phages in *Escherichia coli*. *Mol. Biol. Evol.* **31**, 1606–1617 (2014).
69. Chithambaram, S., Prabhakaran, R. & Xia, X. The effect of mutation and selection on codon adaptation in *Escherichia coli* bacteriophage. *Genetics* **197**, 301–315 (2014).

Acknowledgements

We thank X. He, W. Qian, Z. Zhou, Z. Wu, W. Shi and Z. 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 nos. 2017YFA0103504 to X.C. and 2018ZX10301402 to J.-R.Y. and Z.H.), the National Natural Science Foundation of China (grant nos. 31671320, 31871320 and 81830103 to J.-R.Y. and 31771406 to X.C.), the start-up grant from '100 Top Talents Program' of Sun Yat-sen University (grant nos. 50000-18821112 to X.C. and 50000-18821117 to J.-R.Y.), and the US National Institutes of Health grant R01GM103232 to J.Z.

Author contributions

J.-R.Y., X.C. and 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. and J.-R.Y. analysed data. F.C., J.Z., X.C. and J.-R.Y. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41559-020-1124-7.

Correspondence and requests for materials should be addressed to J.Z., X.C. or J.-R.Y.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

nature research

Corresponding author(s): Xiaoshu Chen; Jian-Rong Yang

Last updated by author(s): Jan 18, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	CytExpert 2.0 for flowcytometry data
Data analysis	Custom R scripts were used in data analysis, which is publicly available on Github

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on 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.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For codon-based analysis, we analyzed all codons with synonymous counterparts (59 codons, excluding Met, Trp and 3 stop codons). For reporter gene assays, 37 mCherry sequences were a wide range of codon usage bias were designed such that effect of codon usage bias could be accessed. We manually curated a list of 52 species-trios, each containing one virus, its natural host and its symptomatic host. We strived to ensure reliable categorization of hosts by reviewing multiple data sources, and this resulting list of 52 trios are the most comprehensive list of this kind so far.
Data exclusions	We used all raw data in our analyses unless gene expression is too low for informative approximation of related parameters. The criteria for such exclusion are all explicitly described in the main text. For the reporter gene assay, we collected 300,000 cells for each strain, and retained at least 100,000 cells with mCherry and YFP signal > 10 times the wildtype BY4741 strain.
Replication	Effect of tRNA depletion by viral translation were replicated in human and yeast Ribo-Seq datasets. Fluorescence signal intensity were measured by flowcytometry for at least 100,000 cells for each strain. Luciferase reporter assays were conducted with three biological replicates, each with three technical replicates.
Randomization	In the reporter gene assay, the 30,000 cells for each strain measured by flowcytometry were treated equally. I.e., no grouping is applied.
Blinding	No grouping of raw samples is applied. Therefore blinding is not applicable here.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Flow Cytometry

Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	All samples were cultured overnight for 16 hours, and diluted to culture for 3-4 hours so that the total cell density values were kept below 0.7 to maintain log phase throughout the experiment. At last, all samples were diluted for 20 times so as to avoid the enormous granules which were mixed together by many yeast single cells.

Instrument	Fluorescence of mCherry was measured from FL13 with a filter having a 20 nm bandpass centered on 610 nm, and fluorescence of YFP was measure from FL1 with a filter having a 40 nm bandpass centered on 525 nm.
Software	We used the CytExpert 2.0 software to collect the data, and all Fcs data were analyzed using the "flowCore" library in R.
Cell population abundance	We collected 300,000 cells for each stain, and retained at least 100,000 cells of each stain after gating by wt BY4741
Gating strategy	Yeast cells with mCherry and YFP fluorescence signals ten times greater than those of the WT BY4741 were kept for later analyses.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.