Journal of Microbiology (2022) Vol. 60
DOI 10.1007/s12275-022-2371-x
eISSN 1976-3794
pISSN 1225-8873

Hepatitis B virus (HBV) codon adapts well to the gene expression profile of liver cancer: an evolutionary explanation for HBV's oncogenic role

Chunpeng Yu[1], Jian Li[1], Qun Li[1], Shuai Chang[1], Yufeng Cao[2], Hui Jiang[2], Lingling Xie[1], Gang Fan[3], and Song Wang[1]*

[1]Department of Interventional Radiology, The Affiliated Hospital of Qingdao University, Qingdao, Shandong 266005, P. R. China
[2]Department of Oncology, The Affiliated Qingdao Hiser Hospital of Qingdao University, Qingdao, Shandong 266005, P. R. China
[3]Department of Interventional Radiology, Jimo District Qingdao Hospital of Traditional Chinese Medicine, Qingdao, Shandong 266005, P. R. China

(Received Aug 16, 2022 / Revised Sep 15, 2022 / Accepted Sep 20, 2022)

Due to the evolutionary arms race between hosts and viruses, viruses must adapt to host translation systems to rapidly synthesize viral proteins. Highly expressed genes in hosts have a codon bias related to tRNA abundance, the primary RNA translation rate determinant. We calculated the relative synonymous codon usage (RSCU) of three hepatitis viruses (HAV, HBV, and HCV), SARS-CoV-2, 30 human tissues, and hepatocellular carcinoma (HCC). After comparing RSCU between viruses and human tissues, we calculated the codon adaptation index (CAI) of viral and human genes. HBV and HCV showed the highest correlations with HCC and the normal liver, while SARS-CoV-2 had the strongest association with lungs. In addition, based on HCC RSCU, the CAI of HBV and HCV genes was the highest. HBV and HCV preferentially adapt to the tRNA pool in HCC, facilitating viral RNA translation. After an initial trigger, rapid HBV/HCV translation and replication may change normal liver cells into HCC cells. Our findings reveal a novel perspective on virus-mediated oncogenesis.

Keywords: hepatitis viruses, evolutionary arms race, hepatocellular carcinoma (HCC), relative synonymous codon usage (RSCU), codon adaptation index (CAI)

Introduction

The host-virus evolutionary arms race is a long-term co-evolutionary process. The virus must optimize its genome sequence to adapt to host cells, while hosts must develop defense systems to prevent viral infection and proliferation. This host-virus arms race could be well understood from the current COVID-19 pandemic. SARS-CoV-2 must adapt to the host RNA translation system to rapidly produce viral proteins (Wang *et al.*, 2021). Therefore, the SARS-CoV-2 sequence has been optimized to achieve higher translation initiation efficiency (Zhang *et al.*, 2022; Zhu *et al.*, 2022) and elongation speed (Li *et al.*, 2020a; Yu *et al.*, 2021). This helps SARS-CoV-2 compete with the hosts' endogenous RNAs and proliferate faster. This sequence adaptation process occurred after SARS-CoV-2 split from RaTG13 (Li *et al.*, 2020c; Zhang *et al.*, 2021b). This suggests that virus evolution could be extremely fast (Cai *et al.*, 2022; Martignano *et al.*, 2022; Zong *et al.*, 2022) given the prevalence of mutations in the virus sequence (Li *et al.*, 2020b; Wei, 2022) and the power of natural selection (Zhang *et al.*, 2021a; Liu *et al.*, 2022b).

The translation efficiency of viral RNA is usually achieved by optimizing synonymous codon usage (Li *et al.*, 2020a; Zhang *et al.*, 2021b). Codon usage bias (CUB) occurs when organisms use synonymous codons unequally (Arella *et al.*, 2021). Preferred synonymous codons (usually G/C-ending codons) have higher matched tRNA concentrations than unpreferred synonymous codons (usually A/T-ending codons) (dos Reis *et al.*, 2004; Chu and Wei, 2019). As a result, despite encoding the same amino acid, preferred codons have higher translation rates than unpreferred codons (Yu *et al.*, 2015). Given this advantage of preferred codons, synonymous mutations are no longer evolutionarily neutral (Plotkin and Kudla, 2011). The fitness changes caused by synonymous mutations can reflect in several ways (Wei, 2020). Synonymous mutations (coupled with alterations in CUB and translation rate) partially contribute to oncogenesis in various cancer types (Supek *et al.*, 2014; Li *et al.*, 2021b). Therefore, optimizing synonymous codon usage is advantageous.

Viral codon usage deviates from their hosts. Under these conditions, viral RNA translation is less efficient because they cannot compete for tRNAs with endogenous mRNAs. However, throughout evolution, viruses have gradually optimized their coding sequences to adapt to host codon usage (Li *et al.*, 2020c; Zhang *et al.*, 2021b). This process elevates viral RNA translatability, allowing them to compete with endogenous RNAs for cellular resources such as tRNAs and translation machineries.

Moreover, tissue-specific genes also determine viral codon optimization. In hosts, such as humans, different tissues or cell types have distinct CUBs due to tissue-specific highly expressed genes (HEG). Widely used evolutionary indices of codon bias, such as relative synonymous codon usage (RSCU) and codon adaptation index (CAI), are derived from tissue-specific HEG rather than the whole genome (Sharp and Li, 1987). Thus, viruses would better mimic the codon usage of the target tissue rather than the host genome. For exam-

*For correspondence. E-mail: wangsongqyfy@163.com; Tel.: +86-0532-96166

ple, SARS-CoV-2 adapted its codon usage to human lungs rather than other tissues or the entire human genome (Li *et al.*, 2020c; Zhang *et al.*, 2021a). For viruses, "host" could refer to a specific tissue or cell type.

Given that viruses mutate and optimize their sequences to adapt to host translation systems, we speculated whether certain viruses promote oncogenesis following a similar mechanism. Our assumption has multiple theoretical bases: (1) Cancer has a high tissue specificity by definition, resembling the virus-host relationship. If the virus can adapt to specific tissues by codon optimization, it might trigger oncogenesis. (2) There are already known cases where the virus contributes to oncogeneses, such as the association between HBV (hepatitis B virus) and hepatocellular carcinoma (HCC) (Liu *et al.*, 2022a). If a codon optimization mechanism exists, then there might be a way to reduce viral transmission within host cells and between tumor and normal tissues; (3) Tumor and normal gene expression profiles must have distinct CUB. A positive feedback loop will be generated if oncogenesis-causing viruses are more adaptive in tumors than normal tissues. This assumption would naturally lead to widespread immortality and rapid cancer proliferation.

Our primary interest is to explore liver cancers such as HCC. Our previous studies have revealed the complex regulatory network of liver cancer oncogenesis, including the microRNA-circRNA pathway (Li *et al.*, 2022), A-to-I RNA editing (Li *et al.*, 2021a), and synonymous codon usage and translation rate (Li *et al.*, 2021b). We proposed that genomic mutations do not fully explain oncogenesis because some tumor-causing mutations exist at a low frequency in normal human populations (Chang *et al.*, 2021). However, the role of hepatitis viruses in developing liver cancer has never been investigated in the light of evolution and codon optimization (Liu *et al.*, 2022a). Based on the widespread virus-host arms race, our codon optimization hypothesis predicts that HBV may have adapted to codon usage in liver cancer.

This study aims to verify the codon optimization hypothesis of hepatitis viruses using the Genotype-Tissue Expression (GTEx) (The GTEx Consortium, 2013) database and HCC expression profile (Zou *et al.*, 2019; Wang and Wang, 2021). HCC had a higher correlation with HBV and HCV in CUB than all the available human normal tissues (GTEx). As a negative control, we discovered no strong relationship between SARS-CoV-2 and HCC, demonstrating that the difference was not driven by the highly divergent human and virus genome sequences. This observation clearly explains the adaptation of HBV/HCV in HCC. Once HBV/HCV triggers the initial transition from normal liver to HCC (under a particular molecular mechanism which is uninvestigated in this study) (Liu *et al.*, 2022a), then the HCC environment would be conducive to HBV/HCV proliferation and protein production. This positive feedback loop hastens the process of oncogenesis. We also discuss the limitations and some unexplainable issues in our theory. Our findings reveal a novel perspective on virus-mediated oncogenesis and deepen our understanding on the host-parasite arms race in the light of evolution.

Materials and Methods

Data acquisition

The reference genome of HBV was obtained using the NCBI ID NC_003977.2, Hepatitis B virus (strain *ayw*). The HAV and HCV reference IDs are NC_001489.1 and NC_004102.1, respectively. The human (*Homo sapiens*) reference genome was downloaded from Ensembl (https://asia.ensembl.org/) version hg38. The SARS-CoV-2 reference sequence was downloaded from NCBI with ID NC_045512.2.

The RNA-sequencing data of multiple tissues were downloaded from Genotype-Tissue Expression (GTEx) (The GTEx Consortium, 2013). The 17328 samples collected from different individuals comprised 30 unique tissues. Sample counts differed between tissues. For example, the brain had the highest number of samples (2,642 samples), while the fallopian tube had the lowest (nine samples). The median number of samples per tissue was 343, with a mean of 17328/30 = 578. Notably, all GTEx samples were normal tissues.

The liver cancer (HCC) RNA-sequencing data were retrieved from previous gene expression studies (Zou *et al.*, 2019; Wang and Wang, 2021). There were 10 samples (sequencing libraries) available from 10 HCC patients.

Relative synonymous codon usage (RSCU) and codon adaptation index (CAI)

RSCU was calculated by multiplying the codon frequency among all synonymous codons (from 0 to 1) with the number of synonymous codons for that amino acid (from 2 to 6) (Sharp and Li, 1987). Each codon had an RSCU value that ranged between zero and six. $RSCU_i$ is the RSCU of a codon divided by the maximum RSCU of an amino acid. The $RSCU_i$ values ranged from 0 to 1. The CAI is the geometric mean of all codons in a gene (Sharp and Li, 1987). Each gene had a CAI value that ranged from 0 to 1.

Due to an insufficient number of HAV and HCV genes for strong statistical power, we employed bootstrapping to calculate the CAI of HAV and HCV genes. The bootstrap approach was designed to increase statistical power with a small sample size. We randomly sampled N codons from the viral CDS (coding sequence), where N is the total number of codons in the coding region, and calculated the CAI using their RSCU. The sampling process was repeated 100 times to obtain 100 CAI values. This means that for each set of RSCU values, HAV and HCV had 100 simulated CAI values. As the N codons were sampled from the viral CDS, the distribution of 100 CAI values accurately reflected the codon usage of the virus.

Highly expressed genes (HEG) for each tissue

For all normal tissues in GTEx and the liver cancer sample HCC, gene expression levels were measured using RPKM (reads per kb per million reads). The top 5000 genes with the highest RPKM in each tissue were designated as the HEG of that tissue. Apart from a few housekeeping genes that are expressed in several tissues, most HEGs are tissue-specific. Moreover, we emphasize that tissue-specific CUB is achieved by all HEG in tissue rather than the "HEG minus housekeeping gene." Therefore, we used the top 5000 HEG in each

tissue to calculate the CUB parameters (such as RSCU and CAI) for that tissue.

Statistics

All statistical analyses (including the above-mentioned calculations), tests (Spearman's correlation tests), and graphic works (the figures in this article) were done using R studio.

Results

Hepatitis viruses, SARS-CoV-2, and human genomes had very different codon usage biases (CUB)

Our codon optimization hypothesis for hepatitis viruses stressed that the HBV CUB had adapted to the host CUB. Before considering the tissue specificity, we first compared viral and human CUBs. We calculated the RSCU of 61 sense codons (excluding three stop codons) of hepatitis viruses (HAV, HBV, and HCV) and the human genome (Fig. 1A). Notably, SARS-CoV-2 was used as a negative control to prove that the observed differences between hepatitis viruses and humans were unrelated to genomic divergence. *Codons* with RSCU > 1 were defined as preferred codons, while codons with RSCU < 1 were defined as unpreferred codons (Fig. 1A).

We found that hepatitis viruses and humans had distinct codon preferences. In humans, A/T-ending codons (the third codon position is A or T) were generally unpreferred codons, whereas G/C-ending codons (the third codon position is G or C) were the preferred codons (Fig. 1B). In HBV, codon preference did not appear to correlate with the third codon position (Fig. 1B); the percentage of preferred and unpreferred codons in the A/T-ending or G/C-ending codons were approximately 50–50%. However, the distinction between HBV and humans was not exclusive to their genetic divergence. SARS-CoV-2 did not function like HBV (Fig. 1B).

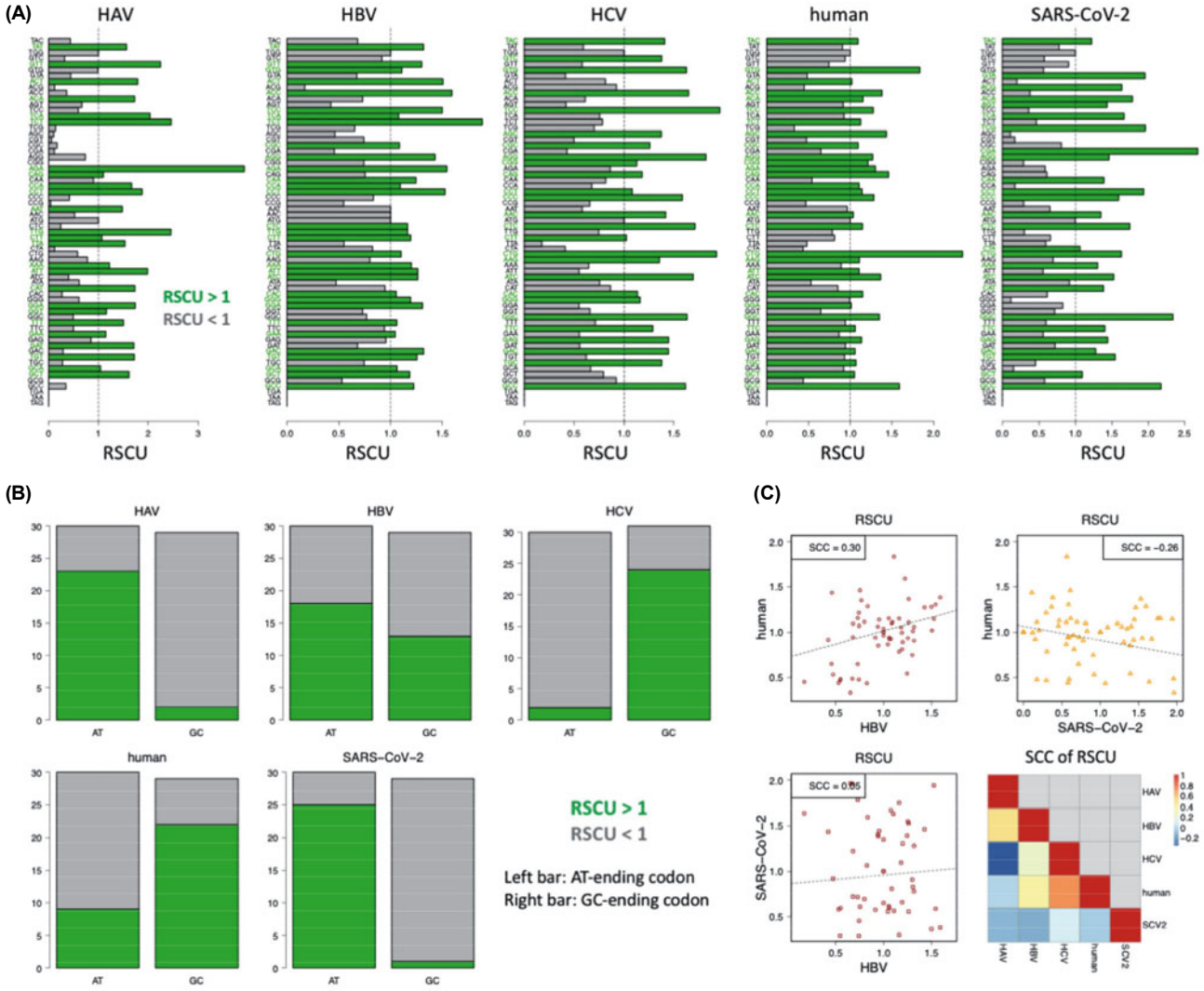

Fig. 1. The codon usage of HAV, HBV, HCV, human genome, and SARS-CoV-2. (A) The relative synonymous codon usage (RSCU) profile for the 61 sense codons. Codons with RSCU > 1 are green (preferred codons). Codons with RSCU < 1 are gray (unpreferred codons). (B) The number of codons with RSCU > 1 (green) and RSCU < 1 (gray). AT3 refers to codons ending with A/T. GC3 refers to codons ending with G/C. (C) Spearman correlation between the RSCU of different species. SCC, Spearman correlation coefficient. Heatmap shows the pairwise SCCs among these species.

Nearly all G/C-ending codons in SARS-CoV-2 were unpreferred codons, whereas A/T-ending codons were the most preferred codons. Unlike HBV, which had a "50%–50%" codon distribution, SARS-CoV-2 had a strong codon preference (on A/T-ending codons) (Fig. 1B). Furthermore, even within hepatitis viruses, HAV, HBV, and HCV had different codon preferences (Fig. 1B).

We performed Spearman's correlation tests to compare the RSCU values of humans, hepatitis viruses, and SARS-CoV-2 (Fig. 1C). We found that (1) the RSCU values of humans and HBV were positively correlated, although not very strongly; (2) the RSCU values of humans and SARS-CoV-2 were negatively correlated; (3) the RSCU values of hepatitis viruses and SARS-CoV-2 had no correlation; (4) HBV and HCV had the highest correlation with humans. (Fig. 1C). Interestingly, HBV and HCV are known to cause liver cancer, whereas HAV does not (Zaki *et al.*, 2022). This result prompted us to explore whether human and HBV/HCV codon usage and viral adaptability were related.

HBV and HCV were better adapted to liver cancer compared to other tissues

The codon optimization hypothesis of the host-parasite arms race emphasized tissue-specific viral adaptation. For example, studies found that SARS-CoV-2 codon usage is specifically adapted to the human lungs (Li *et al.*, 2020c; Zhang *et al.*, 2021b). Lung-specific codon usage and high correlation to lung tRNA pool explain the high translatability of SARS-CoV-2 RNA in the lungs. In contrast, HBV and HCV RSCU had a weak positive correlation with the human genome (Fig. 1), and it was uncertain whether HBV/HCV had a stronger association with specific human tissues. Given that HBV/HCV infects the liver, HBV/HCV having a better RSCU correlation with the liver than other human tissues is possible.

To address this issue, we obtained gene expression profiles of 30 unique human tissues from Genotype-Tissue Expression (GTEx) (The GTEx Consortium, 2013) (refer to the section "Materials and Methods" for the detailed data description). Each human tissue had multiple samples collected from different individuals (ranging from 9 to 2,642), totaling 17,328 samples. Moreover, because HBV is associated with liver cancer HCC (Liu *et al.*, 2022a), we collected HCC transcriptome data from previous studies (Zou *et al.*, 2019; Wang and Wang, 2021). Ten HCC samples were collected from 10 individuals. We compared HBV RSCU to that of each human tissue using Spearman's correlation (Fig. 2). The RSCU was calculated using the 5,000 most HEG in each sample, which were tissue-specific. Each human tissue had multiple samples, and each sample showed a Spearman correlation coefficient (SCC) with HBV/HCV, as shown in Fig. 1C. Consequently, we used box-and-whisker plots to compare the CUB between HBV/HCV and different human tissues (Fig. 2).

Notably, tissue-specific RSCU represents the tissue-specific tRNA pool required for mRNA translation (Sharp and Li, 1987; dos Reis *et al.*, 2004). The high viral correlation to tissue-specific RSCU indicates high adaptiveness to tissue-specific tRNA pools and, thus, high viral RNA translatability. Interestingly, regarding CUB (or tRNA pool) represented by tissue-specific RSCU, we observed that both HBV and HCV had the highest correlation with liver cancer HCC and the second highest correlation with normal liver tissue (Fig. 2), that is, HCC > normal liver > other tissues. This pattern verified the codon optimization hypothesis in the host-virus relationship: HBV/HCV first uses codons to adapt to the liver over other tissues (in normal humans) and then transforms the normal liver into HCC via an unknown mechanism (may be irrelevant

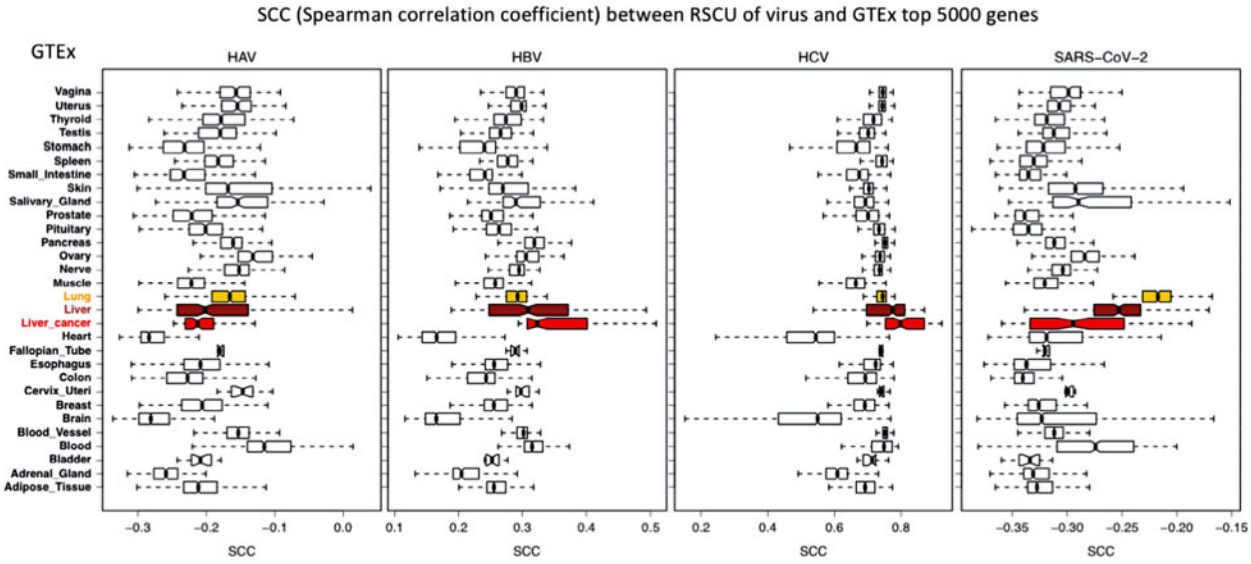

Fig. 2. The Spearman correlation coefficients (SCCs) of RSCU between viruses and human tissues. The RSCU of each human tissue was calculated using the 5,000 most highly expressed genes (HEG). As described in the section "Materials and Methods," the RNA-sequencing data of normal tissues and liver cancer (HCC) were obtained, respectively, from GTEx and the literature. We highlight three tissues: normal lung (orange), normal liver (brown), and liver cancer HCC (bright red). The results demonstrate that the codon usage of HBV and HCV (but not HAV) is most similar to that of liver cancer, whereas that of SARS-CoV-2 is most similar to that of the lung.

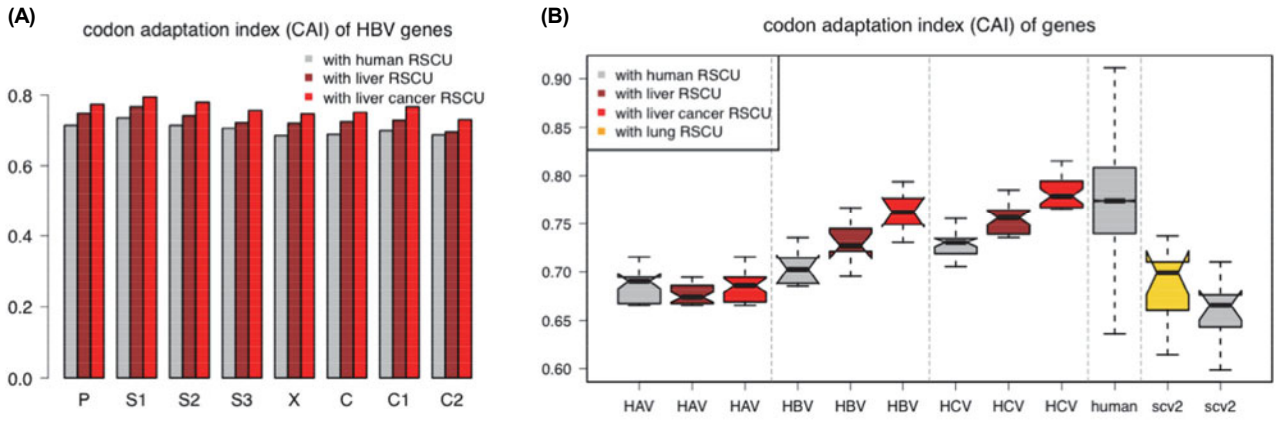

Fig. 3. The codon adaptation index (CAI) of genes. (A) The CAI of different HBV genes. Three sets of CAI values are shown. Gray bars are CAI calculated using human genome RSCU. Brown bars represent CAI calculated using normal liver RSCU. Bright red bars are CAI calculated using liver cancer RSCU. (B) Comparison of CAI of viral and human genes. CAI is calculated with different sets of RSCU. In general, the CAI of the virus is lower than that of humans. CAI of HBV and HCV is highest using the RSCU from liver cancer HCC (bright red), whereas CAI of SARS-CoV-2 is highest using the RSCU from lungs (orange). This also reflects that the HBV and HCV codon usage is most similar to that of liver cancer, whereas the SARS-CoV-2 codon usage is most similar to that of the lung.

to codon usage). Next, HCC codon usage (tRNA pool) promotes HBV/HCV RNA translation, facilitating viral production and transmission. The HBV/HCV replication cascade leads to more liver-to-HCC transitions. As a negative control, we employed SARS-CoV-2 to establish that the high RSCU correlation between HBV/HCV and HCC (or liver) was biological, not technical. SARS-CoV-2 had no strong correlation with the liver or HCC (Fig. 2). Instead, SARS-CoV-2 was most strongly associated with the lungs (Fig. 2), consistent with earlier findings (Li *et al*., 2020c; Zhang *et al*., 2021b). Moreover, HAV codon usage did not show a preference between normal liver and liver cancer, supporting the non-oncogenic role of HAV (Zaki *et al*., 2022).

The apparent high variability of the HBV-liver cancer correlation, as depicted by the box-and-whisker distribution (Fig. 2), was possibly due to two aspects: (1) only 10 HCC patients and 10 samples were available (compared to tens to hundreds samples per normal tissue in GTEx); and (2) cancers are typically more heterogeneous than normal tissues. Similarly, the RSCU correlation between SARS-CoV-2 and liver cancer was variable, although the correlation coefficients were negative (Fig. 2).

Codon adaptation index (CAI) of HBV/HCV genes was highest in patients with HCC

While RSCU evaluates the optimality of a single codon, CAI evaluates the optimality of a gene (Sharp and Li, 1987) (refer to the section "Materials and Methods"). Because CAI is determined using the geometric mean of the RSCU, the RSCU chosen significantly impacts the CAI value. We calculated RSCU for the human genome, viruses, and tissues in the previous sections. Here, we calculated the CAI of the genes using several RSCU sets. Notably, CAI for viral genes should be calculated using the host RSCU, which resembles the tRNA pool. However, since viruses can only employ their hosts' translation machineries and resources, calculating CAI with viral RSCU is pointless.

HBV has several genes, but the exact number depends on how "one gene" is defined. HBV contains the *P* gene (reverse transcriptase; DNA polymerase), *S* gene (surface protein), *X* gene (undetermined), and *C* gene (external core antigen). The *S* gene is further divided into *S1*, *S2*, and *S3*, encoding long, middle, and small surface proteins, respectively. The *C* gene is further divided into *C1* and *C2* encoding signal peptide and core antigen, respectively.

We first calculated the CAI of HBV genes using RSCU of (1) human genome, (2) liver, and (3) liver cancer HCC. Intriguingly, the CAI of HBV genes was the highest in HCC-related malignancies (Fig. 3A). This result was supported by the data shown in Fig. 2, which suggested that HBV is best adapted to liver cancer environments. Then, we expanded our analysis to SARS-CoV-2 genes and found the following patterns (Fig. 3B): (1) Regardless of the RSCU set employed, human gene CAI > HBV gene CAI > SARS-CoV-2 gene CAI; (2) for SARS-CoV-2, lung RSCU provided the highest CAI; (3) for HBV, HCC RSCU provided the highest CAI. Next, we evaluated the CAI of the HAV and HCV genes (refer to the section "Materials and Methods"). HCV exhibited a similar pattern to HBV, while HAV showed similar CAI between different sets of RSCU values (Fig. 3B). These findings supported that HBV and HCV (but not HAV) have an oncogenic role (Zaki *et al*., 2022). Our results mostly corroborated the codon optimization hypothesis of the host-virus arms race. The viral tRNA pool (RSCU) of the preferred tissue was optimal for viral RNA translation.

Discussion

Viruses replicate, mutate, transcribe, and translate in host cells, using the hosts' energies, resources and editing enzymes (Chen *et al*., 2000; Palladino *et al*., 2000; Liang and Landweber, 2007). RNA translation is the rate-limiting step in the central dogma (Stadler and Fire, 2011; Zhang *et al*., 2022). Therefore, increasing the translation rate is favorable for viruses. The cel-

lular tRNA pool, intimately associated with the CUB of a species, regulates most RNA translation rates. Therefore, the RSCU and CAI parameters (measuring CUB) reflect the "translatability" of a codon or gene. Moreover, CUB and tRNA pool are tissue-specific because unique gene sets are significantly expressed in different tissues. The same viral RNA may have different translatabilities in different tissues.

In this study, we found that HBV is translationally adapted to liver cancer HCC by observing that (1) HBV CUB was best correlated with RSCU of HCC, but not other human tissues; (2) CAI of HBV genes was highest in HCC (i.e., CAI calculated with RSCU of HCC). Because the HBV-HCC correlation was higher than the HBV-normal liver correlation, we hypothesized a model to consider how HBV leads to HCC. Possibly, the codon usage of HBV first helps it adapt to the human liver and then, via an as-yet-unknown mechanism, transforms the normal liver to HCC. Next, the HCC codon usage (and thus the tRNA pool) is better suited for HBV RNA translation, facilitating viral production and transmission. Then, a replication cascade (or positive feedback loop) activates in HCC when HBV reproduces uncontrollably, increasing the transformation of normal liver cells into HCC tumor cells.

The validity of the HBV-HCC axis resembles the "Mathematical Induction (MI)" methodology. The MI methodology states that (1) if a formula is valid when N = 1, then (2) it is presumed to be valid when N = k, and (3) proved to be valid when N = k + 1. The formula is valid when these three steps are met.

Similarly, to prove the validity of the HBV-HCC axis, we considered each normal liver cell as a separate unit. HBV is known to be associated with hepatocarcinogenesis via a specific molecular pathway (Liu *et al*., 2022a) (the axis is valid when N = 1), and we assumed that HBV had converted multiple normal liver cells to HCC cells (here, N = k); we would expect HBV to fully exploit the tRNA pools of HCC cells and rapidly translate the HBV RNAs, thereby producing more HBV "individuals" and infecting more liver cells (we thus proved that the axis is valid when N = k + 1). Thus, the oncogenic role of HBV was explained. Indeed, faster viral replication usually does not ensure higher oncogenicity of tumor viruses. Therefore, these two features might not be directly correlated. However, as we have proposed, HBV should be associated with hepatocarcinogenesis via a specific molecular pathway/mechanism (which is not the codon usage issue mentioned in this study). HBV codon optimization induced a cascade of HBV proliferation; however, the HBV oncogenic role might not be relevant to codon optimization.

On the other hand, it is believed that tumor viruses including HBV generally proliferate more efficiently in normal differentiated cells than in less differentiated cancerous cells. This seems to contradict with our theory. We have the following explanations:

(1) The observed virus proliferation rate is connected to not only RNA translation rate but also many other processes like RNA processing, RNA degradation rate, and even protein degradation rate.

(2) Our codon optimization hypothesis is solely based on the *cis* feature (sequence) but does not consider potential *trans* factors in liver cancer that may inhibit HBV proliferation. We can envision that if the codon usage of HBV does not adapt to liver cancer, then the virus may proliferate even more slowly than what is currently observed. Therefore, the observed low proliferation rate of HBV in liver cancer does not necessarily reflect the failure of codon optimization of the virus. It may be due to *trans* effects which are not investigated in this study.

We found that the positive correlation between HBV codon usage and the human genome, liver, and HCC was not perfect. In other words, HBV had not completely adapted to the tRNA pool in human cells. Consequently, human genes had higher global CAI values than the virus genes when evaluated using human RSCU (Fig. 3). This is expected because the cellular tRNA pool is designed to meet the requirements of highly expressed, tissue-specific genes. In addition, there is a "home court advantage" that the endogenous mRNAs must have higher translatability than exogenous viral RNAs because viruses can rapidly mutate and adapt. The facts that (1) HBV codon usage best adapts to liver or liver cancers, (2) SARS-CoV-2 codon usage best adapts to the lungs (Li *et al*., 2020c; Zhang *et al*., 2021b) proves that the viral sequences have already been optimized during evolution.

Notably, the measurement of CUB, like the RSCU, closely resembles the tRNA pool; however, they are not identical (dos Reis *et al*., 2004). The ultimate purpose of codon optimization (for viruses) is to adapt to the tRNA pool of the host. Therefore, using tissue-specific tRNA-sequencing data to determine tRNA abundance should be more accurate than using the RSCU calculated with highly expressed mRNAs. However, tRNA-sequencing data for all tissues are not always available. This aspect could be improved in future studies on the evolutionary arms race between viruses and hosts.

Acknowledgements

We thank the Wiley Editing Service for language editing. No funding has supported this study.

Conflict of Interest

The authors declare that they have no conflict of interest.

Author Contributions

Conceptualization: S.W.
Supervision: S.W.
Methodology: C.Y., J.L., Q.L., S.C., Y.C., H.J., L.X., and G.F.
Writing – original draft preparation: S.W.
Writing – review and editing: S.W.
All authors have read and agreed to the published version of the manuscript.

Ethical Statement

This article does not contain any studies with human participants or animals performed by any of the authors.

Data Availability

All data needed to evaluate the conclusions in the paper are present in the paper.

References

Arella, D., Dilucca, M., and Giansanti, A. 2021. Codon usage bias and environmental adaptation in microbial organisms. *Mol. Genet. Genomics* **296**, 751–762.

Cai, H., Liu, X., and Zheng, X. 2022. RNA editing detection in SARS-CoV-2 transcriptome should be different from traditional SNV identification. *J. Appl. Genet.* **63**, 587–594.

Chang, S., Li, J., Li, Q., Yu, C.P., Xie, L.L., and Wang, S. 2021. Retrieving the deleterious mutations before extinction: genome-wide comparison of shared derived mutations in liver cancer and normal population. *Postgrad. Med. J.* **98**, 584–590.

Chen, C.X., Cho, D.S., Wang, Q., Lai, F., Carter, K.C., and Nishikura, K. 2000. A third member of the RNA-specific adenosine deaminase gene family, ADAR3, contains both single- and double-stranded RNA binding domains. *RNA* **6**, 755–767.

Chu, D. and Wei, L. 2019. Characterizing the heat response of *Arabidopsis thaliana* from the perspective of codon usage bias and translational regulation. *J. Plant Physiol.* **240**, 153012.

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

Li, J., Li, Q., Yu, C.P., Chang, S., Xie, L.L., and Wang, S. 2021a. Genome-wide expression changes mediated by A-to-I RNA editing correlate with hepatic oncogenesis. *Transl. Cancer Res.* **10**, 2725–2737.

Li, Q., Li, J., Yu, C.P., Chang, S., Xie, L.L., and Wang, S. 2021b. Synonymous mutations that regulate translation speed might play a non-negligible role in liver cancer development. *BMC Cancer* **21**, 388.

Li, Y., Yang, X., Wang, N., Wang, H., Yin, B., Yang, X., and Jiang, W. 2020a. GC usage of SARS-CoV-2 genes might adapt to the environment of human lung expressed genes. *Mol. Genet. Genomics* **295**, 1537–1546.

Li, Y., Yang, X.N., Wang, N., Wang, H.Y., Yin, B., Yang, X.P., and Jiang, W.Q. 2020b. Pros and cons of the application of evolutionary theories to the evolution of SARS-CoV-2. *Future Virol.* **15**, 369–372.

Li, Y., Yang, X.N., Wang, N., Wang, H., Yin, B., Yang, X., and Jiang, W.Q. 2020c. The divergence between SARS-CoV-2 and RaTG13 might be overestimated due to the extensive RNA modification. *Future Virol.* **15**, 341–347.

Li, J., Yu, C.P., Li, Q., Chang, S., Xie, L.L., and Wang, S. 2022. Large-scale omics data reveal the cooperation of mutation-circRNA-miRNA-target gene network in liver cancer oncogenesis. *Future Oncol.* **18**, 163–178.

Liang, H. and Landweber, L.F. 2007. Hypothesis: RNA editing of microRNA target sites in humans? *RNA* **13**, 463–467.

Liu, W., Cai, S., Pu, R., Li, Z., Liu, D., Zhou, X., Yin, J., Chen, X., Chen, L., Wu, J., *et al.* 2022a. HBV preS mutations promote hepatocarcinogenesis by inducing endoplasmic reticulum stress and up-regulating inflammatory signaling. *Cancers* **14**, 3274.

Liu, X., Liu, X., Zhou, J., Dong, Y., Jiang, W., and Jiang, W. 2022b. Rampant C-to-U deamination accounts for the intrinsically high mutation rate in SARS-CoV-2 spike gene. *RNA* **28**, 917–926.

Martignano, F., Di Giorgio, S., Mattiuz, G., and Conticello, S.G. 2022. Commentary on "Poor evidence for host-dependent regular RNA editing in the transcriptome of SARS-CoV-2". *J. Appl. Genet.* **63**, 423–428.

Palladino, M.J., Keegan, L.P., O'Connell, M.A., and Reenan, R.A. 2000. *dADAR*, a Drosophila double-stranded RNA-specific adenosine deaminase is highly developmentally regulated and is itself a target for RNA editing. *RNA* **6**, 1004–1018.

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

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

Stadler, M. and Fire, A. 2011. Wobble base-pairing slows *in vivo* translation elongation in metazoans. *RNA* **17**, 2063–2073.

Supek, F., Miñana, B., Valcárcel, J., Gabaldón, T., and Lehner, B. 2014. Synonymous mutations frequently act as driver mutations in human cancers. *Cell* **156**, 1324–1335.

The GTEx Consortium. 2013. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585.

Wang, Y., Gai, Y., Li, Y., Li, C., Li, Z., and Wang, X. 2021. SARS-CoV-2 has the advantage of competing the iMet-tRNAs with human hosts to allow efficient translation. *Mol. Genet. Genomics* **296**, 113–118.

Wang, N. and Wang, D. 2021. Genome-wide transcriptome and translatome analyses reveal the role of protein extension and domestication in liver cancer oncogenesis. *Mol. Genet. Genomics* **296**, 561–569.

Wei, L. 2020. Selection on synonymous mutations revealed by 1135 genomes of *Arabidopsis thaliana*. *Evol. Bioinform. Online* **16**, 1176934320916794.

Wei, L. 2022. Reconciling the debate on deamination on viral RNA. *J. Appl. Genet.* **63**, 583–585.

Yu, C.H., Dang, Y., Zhou, Z., Wu, C., Zhao, F., Sachs, M.S., and Liu, Y. 2015. Codon usage influences the local rate of translation elongation to regulate co-translational protein folding. *Mol. Cell* **59**, 744–754.

Yu, Y., Li, Y., Dong, Y., Wang, X., Li, C., and Jiang, W. 2021. Natural selection on synonymous mutations in SARS-CoV-2 and the impact on estimating divergence time. *Future Virol.* **16**, 447–450.

Zaki, M.Y.W., Fathi, A.M., Samir, S., Eldafashi, N., William, K.Y., Nazmy, M.H., Fathy, M., Gill, U.S., and Shetty, S. 2022. Innate and adaptive immunopathogeneses in viral hepatitis; crucial determinants of hepatocellular carcinoma. *Cancers* **14**, 1255.

Zhang, Y., Jiang, W., Li, Y., Jin, X.J., Yang, X.P., Zhang, P.R., Jiang, W., and Yin, B. 2021a. Fast evolution of SARS-CoV-2 driven by deamination systems in hosts. *Future Virol.* **16**, 587–590.

Zhang, Y., Jin, X., Wang, H., Miao, Y., Yang, X., Jiang, W., and Yin, B. 2021b. Compelling evidence suggesting the codon usage of SARS-CoV-2 adapts to human after the split from RaTG13. *Evol. Bioinform. Online* **17**, 11769343211052013.

Zhang, Y., Jin, X., Wang, H., Miao, Y., Yang, X., Jiang, W., and Yin, B. 2022. SARS-CoV-2 competes with host mRNAs for efficient translation by maintaining the mutations favorable for translation initiation. *J. Appl. Genet.* **63**, 159–167.

Zhu, L., Wang, Q., Zhang, W., Hu, H., amd Xu, K. 2022. Evidence for selection on SARS-CoV-2 RNA translation revealed by the evolutionary dynamics of mutations in UTRs and CDSs. *RNA Biol.* **19**, 866–876.

Zong, J., Zhang, Y., Guo, F., Wang, C., Li, H., Lin, G., Jiang, W., Song, X., Zhang, X., Huang, F., *et al.* 2022. Poor evidence for host-dependent regular RNA editing in the transcriptome of SARS-CoV-2. *J. Appl. Genet.* **63**, 413–421.

Zou, Q., Xiao, Z., Huang, R., Wang, X., Wang, X., Zhao, H., and Yang, X. 2019. Survey of the translation shifts in hepatocellular carcinoma with ribosome profiling. *Theranostics* **9**, 4141–4155.