

Article

Codon Usage Bias Analysis of Citrus Leaf Blotch Virus

Xin Ren ¹, Lifang Xu ¹, Yuqian Yan ¹, Ying Wang ¹ and Aijun Huang ^{1,2,*}
¹ College of Life Science, Gannan Normal University, Ganzhou 341000, China; 1251016006@gnnu.edu.cn (X.R.); 1241016015@gnnu.edu.cn (L.X.); 221602038@gnnu.edu.cn (Y.Y.); wangying@gnnu.edu.cn (Y.W.)

² National Navel Orange Engineering Research Center, Ganzhou 341000, China

* Correspondence: huangaijun@gnnu.edu.cn

Abstract

Citrus leaf blotch virus (CLBV) is a positive-sense single-stranded RNA virus belonging to the genus *Citivirus* within the family *Betaflexiviridae*. It infects a broad range of economically significant fruit crops, including citrus, kiwifruit, and apple. Surveys conducted in the field have documented appreciable incidence rates in several hosts, thereby emphasizing its emerging threat to global pomiculture. Comprehensive surveillance of CLBV genetic diversity is indispensable for predicting strain-specific epidemics and designing durable, broadly protective control strategies. Current surveys of CLBV diversity are still gene-fragment-centric, with whole-genome resolution remaining largely untapped. In this study, an analysis of codon usage bias analysis was performed using all available CLBV full-length genomes. The findings revealed that CLBV exhibits low codon usage bias, with natural selection, rather than mutational drift, being the primary driver. Phylogenetic analysis has been demonstrated to categorize isolates according to their host of origin rather than their geographical location. This observation suggests that host adaptation may supersede spatial structure in CLBV evolution and reinforce natural selection as the dominant force shaping its codon usage landscape. From the perspective of the codon adaptation index, *Prunus avium* is the host that exerts the greatest influence on the formation of its codon usage bias. The present study provides the first genome-wide portrait of CLBV codon usage bias, offering a robust framework for future investigations into its origin and evolutionary dynamics.

Keywords: citrus leaf blotch virus; codon usage; host adaptation


Academic Editor: Gian Paolo Accotto

Received: 14 October 2025

Revised: 7 November 2025

Accepted: 11 November 2025

Published: 12 November 2025

Citation: Ren, X.; Xu, L.; Yan, Y.; Wang, Y.; Huang, A. Codon Usage Bias Analysis of Citrus Leaf Blotch Virus. *Viruses* **2025**, *17*, 1497. <https://doi.org/10.3390/v17111497>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Citrus leaf blotch virus (CLBV), assigned to the genus *Citivirus* of the family *Betaflexiviridae*, has a broad natural host range encompassing citrus, kiwifruits, apples, peonies, and mulberries [1–6]. Infected plants may remain asymptomatic or display faint leaf mottling and vein clearing, depending on host genotype and environment [7,8]. Virions are flexuous filaments, about 960 × 14 nm that package an 8747 nt, positive-sense, single-stranded RNA genome containing a 3′ poly(A) tail [9,10]. Three ORFs are expressed: ORF1 yields a 227 kDa replication polyprotein harboring methyltransferase, AlkB, OTU-like peptidase, papain-like protease, helicase, and RdRp domains; ORF2 encodes a 40 kDa movement protein (MP) of the 30 K superfamily; and ORF3 specifies a 41 kDa coat protein (CP) [11]. MP and CP are translated from dedicated sub-genomic RNAs, and the MP additionally functions as a suppressor of RNA silencing [12].

CLBV remains an economically overlooked pathogen, and has not yet caused large-scale economic losses in its known hosts, so the virus has attracted only limited research

interest thus far. Published investigations are largely confined to field incidence surveys and fragment-based descriptions of genetic diversity. Recent surveys of major kiwifruit and citrus production areas in China revealed CLBVB detection rates of between 11.3% and 28.5% in kiwifruit and approximately 7.7% in citrus. These results indicate that the virus has already established a measurable prevalence in these economically important crops [13,14]. Phylogenetic analyses based on the movement protein (MP) and coat protein (CP) genes show minimal divergence among isolates from different geographical origins, but substantial divergence among isolates from different host species, implying that host adaptation has driven the observed variability. Because existing studies have relied almost exclusively on partial genomic sequences, comprehensive analyses using complete genome sequences are now required to provide a more robust understanding of CLBVB evolution and epidemiology.

Codon usage bias (CUB) describes the non-random choice among synonymous triplets that specify the same amino acid [15]. Across all kingdoms of life, 61 sense codons are deployed with markedly unequal frequencies, generating species-specific yet genome-wide patterns that are more conserved than protein sequences themselves [16]. The observed bias can be attributed to a tripartite balance, comprising directional mutation pressure, natural selection for translational efficiency and accuracy, and genetic drift, which fixes minor alleles in small populations [17]. Additional layers, including gene length, mRNA secondary structure, protein-fold hydrophobicity, replication timing, and environmental stress, fine-tune local codon choice [18], whereas the tRNA gene copy number and wobble-base modification provide the ultimate selective arena [19–22]. In plant viruses these forces are amplified by an intracellular arms race: RNA viruses exhibit unusually low CUB, and adopt a “generalist” strategy to evade competition with highly expressed host photosynthetic genes. Non-recombinant isolates of the Potato virus Y display uniformly low codon preference, and every gene of the Turnip mosaic virus exhibits ENC values > 45, reducing competition with the divergent tRNA pools of their cruciferous hosts [23,24].

Despite the increasing number of CLBVB genome sequences, a systematic assessment of codon usage bias as an evolutionary driver is still lacking. Here, we employed all available CLBVB full-length genomes to dissect, for the first time, the codon usage landscape of citrus leaf blotch virus and to reconstruct its phylogenetic history. This combined approach offers new insight into the virus’s genetic divergence and the determinants that shape its synonymous codon choice.

2. Materials and Methods

2.1. Virus Isolates

A total of 56 full genomic sequences of CLBVB isolates were downloaded from the National Center for Biotechnology Information (NCBI) (<https://www.ncbi.nlm.nih.gov/>, accessed on 20 June 2025). Metadata for each isolate, including date of collection, geographical location, and host plant, are provided in Table S1.

2.2. Recombination and Phylogenetic Analysis

The alignment of all CLBVB sequences was conducted utilizing the MUSCLE method and employing the MEGA 12 software. The potential occurrence of recombination events was screened via seven algorithms (RDP, BOOTSCAN, MAXCHI, SISCAN, CHIMAERA, 3SEQ, GENECONV) within the RDP4 software suite. The assumptions regarding parent/donor assignments were confirmed through the implementation of a phylogenetic approach in the RDP4 suite. Each candidate recombinant event was corroborated by a minimum of four independent algorithms within the RDP4 suite, with p -values below 1.0×10^{-6} .

The phylogenetic relationships of the three protein-coding sequences of CLBV were assessed using the maximum-likelihood (ML) method with MEGA12 software. The general time-reversible substitution model with gamma distribution (G) and invariant sites (I) (GTR + G + I) is suitable for the coding sequences of the three regions, namely polyprotein, MP, and CP. Branch support was evaluated by a bootstrap analysis using 1000 pseudoreplicates for the ML analyses. The resulting trees were then subjected to visualization in TreeView.

2.3. Nucleotide Composition Analysis

Following the exclusion of five non-biased codons (UAA, UGA, UAG, termination codons) and the single-codon amino acids AUG (Met) and UGG (Trp), compositional parameters were calculated for the three CLBV protein-coding sequences. Overall nucleotide composition (A, C, U, and G%) and AU and GC content were determined in BioEdit. The nucleotide composition at the third codon position of the three CLBV protein-coding sequences (A3, C3, U3, and G3%) were calculated using the CodonW 1.4.2. The GC content at the 1st, 2nd, and 3rd codon positions (GC1, GC2, GC3), and the GC12 (the mean of GC1 and GC2) were computed by Python v3.13. The dinucleotide analysis was computed using the online platform EMBOSS (<https://www.bioinformatics.nl/cgi-bin/emboss/compseq>, accessed on 25 June 2025). Codon pair bias (CPB) refers to the non-random usage of synonymous codon pairs within protein-coding regions of genomes [25]. The codon pair bias (CPB) analysis was performed with Python v3.13.

2.4. Relative Synonymous Codon Usage (RSCU) Analysis

The RSCU is defined as the ratio of the observed frequency of a codon to the frequency expected under uniform synonymous codon usage [26]. It used to measure the degree of codon preference. RSCU values > 1.6 are considered over-represented for the more abundant codons, and <0.6 are regarded under-represented for the less abundant [27]. The calculation formula is as follows:

$$\text{RSCU}_{ij} = \frac{g_{ij}}{\sum_j^{ni} g_{ij}} \times ni$$

RSCU_{ij} represents the value of the i th codon for the j th amino acid; whereas g_{ij} is the cumulative count of all synonymous codons that encode the j th amino acid, and ni is the total number of synonymous codons that encode the j th amino acid. Python v3.13 was used to calculate the RSCU values of the CLBV three protein-coding sequences.

2.5. Effective Number of Codons (ENC) Analysis

The ENC is a quantitative metric that quantifies the degree of sequence codon preference. It ranges from 20, indicating an extreme preference for one synonymous codon per amino acid, to 61, indicating no preference and an equal use of synonymous codons per amino acid [28]. The ENC values were calculated via the EMBOSS online platform (<https://emboss.bioinformatics.nl/cgi-bin/emboss/>, accessed on 27 June 2025).

2.6. ENC-Plot Analysis and Selection Pressure Analysis

The ENC-plot was generated using R 4.5.0 studio, with the X-axis designated as GC3 and the Y-axis assigned to the ENC value. In the absence of selection, codon usage is dictated solely by mutational bias, and data dots fall on, or close to, the standard curve. Deviations below the curve indicate that natural selection, together with mutational pressure and other evolutionary forces, shapes codon bias patterns [29]. We performed a selection pressure analysis using the Branch-site Unrestricted Statistical Test for Episodic

Diversification with Multiple Hits (BUSTED-MH) method on the Datamonkey platform (<https://www.datamonkey.org/>, accessed on 6 July 2025).

2.7. Parity Rule 2 Analysis

PR2-bias plots were employed to investigate the relative contributions of natural selection and mutation pressure on the codon usage of the CLBv, with G3/(G3 + C3) and A3/(A3 + T3) plotted on the abscissa and ordinate, respectively. When codon usage bias is solely driven by mutation pressure, A=T and G=C, with data dots locating at the midpoint (X = 0.5, Y = 0.5) [30]. The contents of A3, T3, G3, and C3 for each CLBv sequence were calculated and plotted using R 4.5.0 studio.

2.8. Neutrality Analysis

In the neutrality plot graph, GC12 and GC3 are shown as the ordinate and abscissa, respectively, to quantify the relative influences of mutation pressure, natural selection, and other factors on codon usage bias. In the absence of selection pressure, or if the selection pressure is weak, the slope of the regression line is near 1.0 [31]. Conversely, if the regression line slope deviates from 1.0, this indicates that natural selection is pivotal in codon bias [32]. The neutrality plots were generated in R 4.5.0 studio.

2.9. Codon Adaptation Index (CAI) Analysis

The CAI reflect the adaptability of individual viral genes to their hosts. The range of CAI is from 0 to 1, with higher values indicating closer codon usage matching and, consequently, stronger host adaptation [33]. The CAI values were calculated on the CAIcal online server (<http://genomes.urv.cat/CAIcal>, accessed on 9 July 2025) with host codon table data retrieved from the Codon Usage Database (<http://www.kazusa.or.jp/codon>, accessed on 12 July 2025). Following the grouping of the CAI values by host, the generation of visualizations was undertaken using R 4.5.0 studio.

3. Results

3.1. Recombination Analysis

Prior to downstream analyses, a rigorous screening for recombination events was conducted on the 56 CLBv genomes. This is an essential step because recombination has the capacity to distort both phylogenetic topology and genome-wide codon usage signatures. Following the identification and exclusion of ten isolates exhibiting clear recombinant signals, 46 non-recombinant genomes were selected for further analysis. The 46 CLBv remaining genome sequences included 18 from citrus, 26 from *Actinidia*, and 2 from *malus*. In the recombination events detected, breakpoints were predominantly localized within the CP region of CLBv, with sporadic events detected in the MP and the polyprotein regions. Notably, *Citrus* spp. constituted the major or minor parental lineage for the majority of these recombinant isolates (Table 1).

Table 1. Recombination events detected in CLBv full-length genomes.

Recombinant Isolates	Major Parent	Minor Parent	Breaking Point Position (bp)	Coding Protein
<i>Nandina</i> (MT078932)	<i>Malus</i> (MW713062)	<i>Citrus</i> (OL871235)	7008–8786	MP, CP
<i>Morus</i> (MT767171)	<i>Actinidia</i> (OK513339)	<i>Citrus</i> (OL871235)	7274–8161	MP, CP
<i>Actinidia</i> (MG604237)	<i>Actinidia</i> (OK513347)	<i>Actinidia</i> (OK513324)	1–2229	Polyprotein
<i>Prunus</i> (KR023647)	<i>Citrus</i> (LC758583)	<i>Malus</i> (MW713062)	7567–8762	CP
<i>Morus</i> (OP971103)	<i>Malus</i> (MW713062)	<i>Citrus</i> (AJ318061)	7280–8617	CP
<i>Viburnum</i> (OP751940)	<i>Citrus</i> (OR838782)	<i>Actinidia</i> (OK513336)	8246–8817	CP
<i>Malus</i> (MW713062)	<i>Actinidia</i> (OK513339)	<i>Citrus</i> (OL871235)	7259–8491	CP
<i>Actinidia</i> (MK135436)	<i>Nandina</i> (MT078932)	<i>Actinidia</i> (JN900477)	6889–8678	MP, CP

Table 1. Cont.

Recombinant Isolates	Major Parent	Minor Parent	Breaking Point Position (bp)	Coding Protein
<i>Citrus</i> (MH558590)	<i>Actinidia</i> (OK513326)	<i>Citrus</i> (LC758583)	6279–6577	MP, CP
<i>Malus</i> (OR855671)	<i>Citrus</i> (OR838782)	<i>Viburnum</i> (OP751940)	8205–8787	CP

Note: The parentheses contain the GenBank accession numbers of the sequences.

3.2. Phylogenetic Analysis

Maximum-likelihood phylogenies were inferred separately for each of the three CLB- encoded proteins. The results showed that the sequences exhibited pronounced host origin clustering. In particular, the polyprotein and CP-based trees revealed that all *Actinidia*- derived isolates formed a single, well-supported clade, the two *malus*-derived isolates constituted a second distinct clade, while the *citrus*-derived isolates were resolved into three separate lineages: one exclusive *citrus* clade, one nested within the *Actinidia* clade, and the remaining forming a second *citrus*-specific clade (Figure 1). These findings suggest that host adaptation has played a significant role in shaping viral diversification during the process of spread, and that host origin exerts a pronounced influence on sequence variation. Consequently, host origin was incorporated as a pivotal factor in all subsequent codon usage analyses.

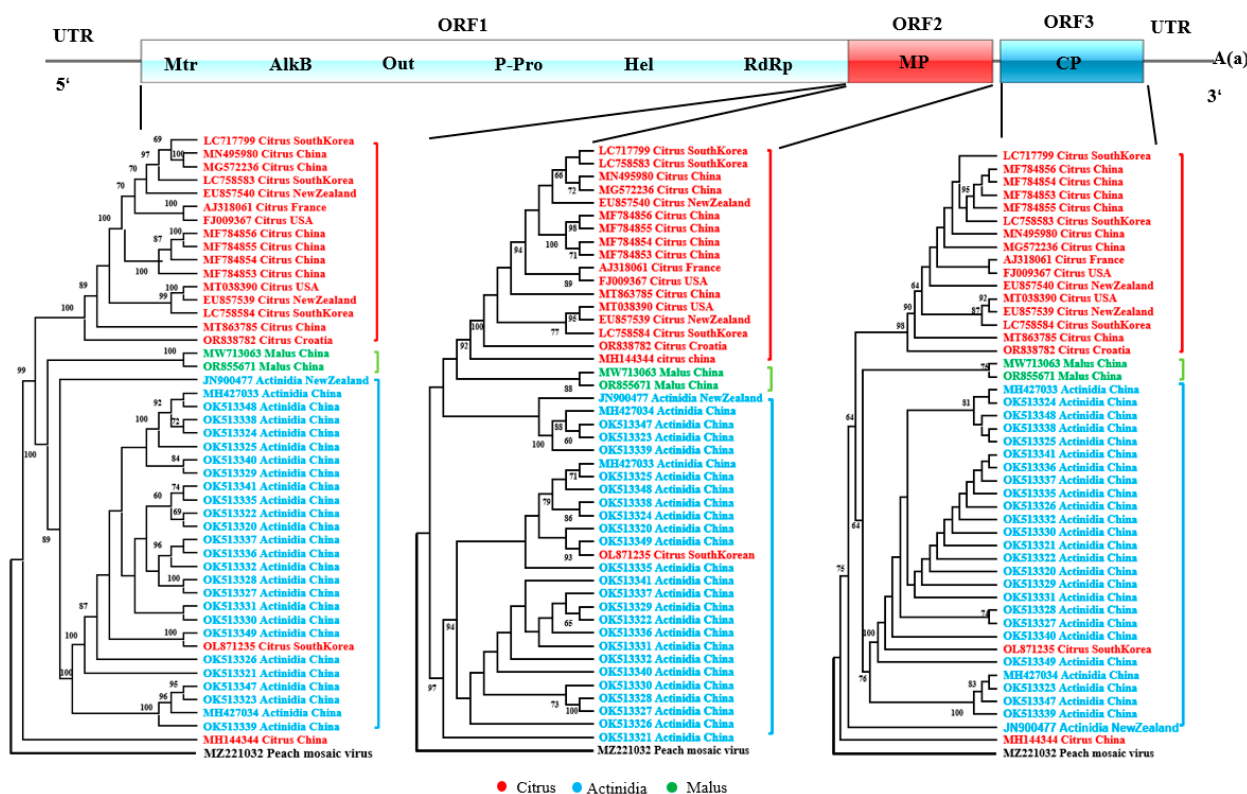


Figure 1. The maximum-likelihood (ML) tree calculated from the individual protein-coding sequences of CLB.

3.3. Nucleotide Composition Analysis

The nucleotide compositions of the three protein-coding sequences of CLB were assessed to explore the effect of compositional constraints on codon usage. Across the CLB polyprotein region, adenine ($30.8 \pm 0.38\%$) and uracil ($29.4 \pm 0.33\%$) were the dominant nucleotides; the CP region displayed the same A/U bias, whereas the MP region showed A ($32.79 \pm 0.32\%$) and G ($25.08 \pm 0.34\%$) as the two most abundant bases. At the third codon position, the polyprotein and MP regions exhibited the hierarchy $U_3 > A_3 > G_3 > C_3$

(polyprotein: $33.03 \pm 0.63\%$, $25.96 \pm 0.96\%$, $20.99 \pm 0.80\%$, $20.02 \pm 0.67\%$, and MP), while the CP region followed $A_3 \approx U_3 > G_3 > C_3$ ($31.02 \pm 1.78\%$, $30.57 \pm 1.37\%$, $20.77 \pm 1.16\%$, $17.64 \pm 0.72\%$) (Table S2). In conclusion, the predominance of A/U was observed at all three codon positions. The composition of A/U is superior to the GC of three protein-coding sequences, suggesting that A/U is more prevalent in the CLB V coding sequences. The results of the dinucleotide analysis and codon pair bias (CPB) are provided in Supplementary Table S4, respectively. Our dinucleotide analysis revealed a genome-wide depletion of CG dinucleotides and an enrichment of AU/UA dinucleotides in the CLB V genome. Subsequent codon pair analyses further corroborated this trend.

3.4. RSCU Analysis of CLB V

In order to estimate the codon usage pattern of the protein-coding sequences of CLB V, the RSCU values were first calculated. The CLB V protein-coding sequence exhibited marked codon usage biases for multiple amino acids, including arginine (Arg), asparagine (Asn), histidine (His), leucine (Leu), threonine (Thr), and serine (Ser). For instance, Arg is preferentially encoded by AGA and AGG, with the RSCU value of AGA ranging from 2.65 to 3.66 across different hosts and coding regions, while the codons CGA, CGT, CGC, and CGG are strongly avoided. Furthermore, it was observed that most high-frequency codons favored termination with A/T, while a majority of low-frequency codons favored to end with G/C. Simultaneously, we identified pronounced host-dependent differences in codon usage within specific coding regions. Specifically, cysteine (Cys) is preferentially encoded by TGC in the polyprotein-coding region in citrus hosts (with TGT disfavored), whereas the opposite pattern—strong preference for TGT and avoidance of TGC—is observed in *Actinidia* hosts (Figure 2, Table S3).

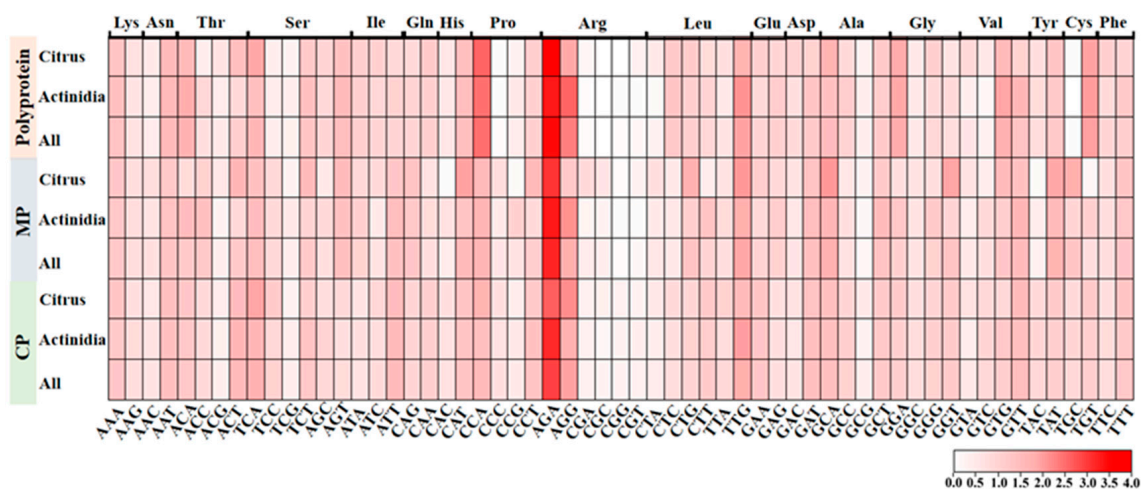


Figure 2. RSCU heatmap of 59 codons encoding 18 amino acids across 3 CLB V proteins.

3.5. Codon Usage Bias of CLB V

The effective number of codons (ENC) is a metric of codon usage bias that is both length and composition independent. The scale ranges from 20 (extreme bias) to 61 (no bias); values < 35 denote strong bias, whereas those > 55 indicate weak bias [28]. Across the three CLB V protein-coding regions, ENC varied narrowly (47.06–51.63), reflecting a globally conserved nucleotide composition and low overall bias. Nevertheless, host-specific divergence was evident: the MP region exhibited the most pronounced shift, with ENC values of 48.55 in *Citrus* and 52.94 in *Actinidia*, implying adaptive modulation of codon usage during host specialization (Figure 3).

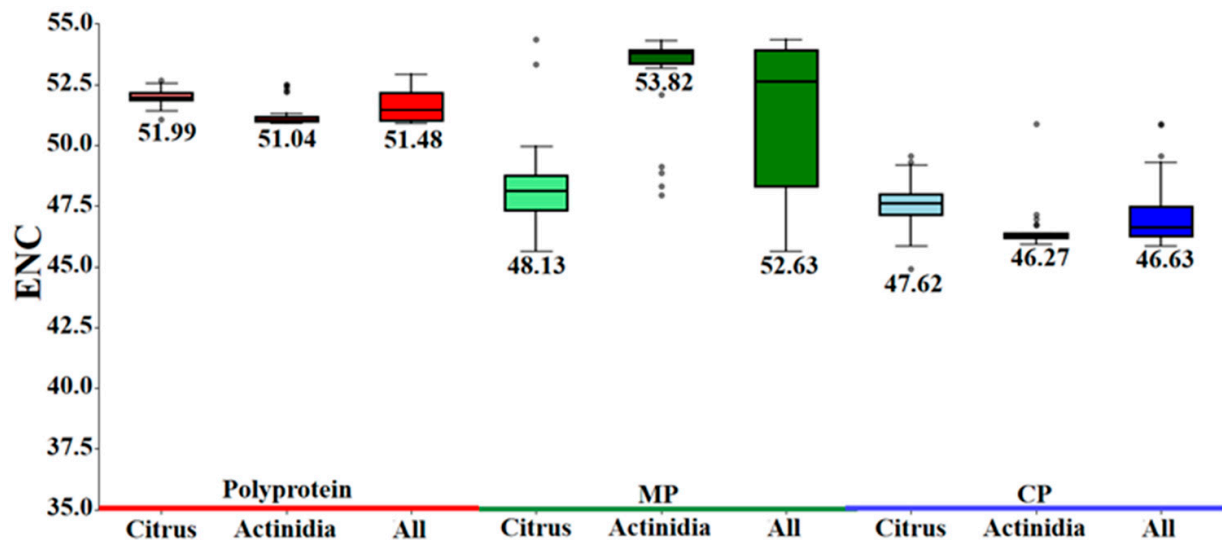


Figure 3. ENC values for the three coding sequences of CLB.

3.6. ENC-Plot and Neutrality Analysis

ENC-plot and neutrality analysis jointly dissect the evolutionary forces shaping codon usage by contrasting observed ENC or GC₁₂–GC₃ relationships with their mutational expectations. This approach enables the quantification of the relative contributions of natural selection and mutational bias. It is hypothesized that under conditions of pure mutational pressure, ENC-plot points will be expected to lie on the theoretical curve, and that the neutrality plot will yield a regression slope of 1. In the ENC-plot generated for the three protein-coding sequences of CLB, all data points representing CLB isolates from different hosts cluster below the expected curve. This finding suggests that natural selection dominated over mutation pressure, while the influence of mutation was not completely absent. Consistently, neutrality analysis showed that no significant correlation was observed between the GC₁₂ and GC₃ for the three protein-coding sequences of CLB. The slope of the linear regression for protein-coding sequences was found to be close to zero, indicating that codon usage is predominantly governed by natural selection. However, the slopes differed among the coding regions. The regression slope for the MP coding region was found to be -0.136 , indicating a mutation pressure of 14.6% and a natural selection effect of 85.4%, representing the weakest natural selection among the three coding regions (Figure 4). The results obtained collectively indicated that natural selection is the principal driving force for the formation of codon usage bias in CLB. The detailed data from the selection pressure analysis are provided in Supplementary Table S4. The selection pressure analysis results revealed that significant episodic positive selection was detected in both the CP and polyprotein genes: In the CP gene, 99.754% of sites were under strong negative selection ($dN/dS = 0.044/0.048$), while 0.246% of sites experienced extremely strong positive selection ($dN/dS = 65.717$). In the polyprotein gene, 99.947% of sites were under negative selection ($dN/dS = 0.001/0.207$), whereas 0.053% of sites were under strong positive selection ($dN/dS = 50.510$). These findings provide supporting evidence for the view that natural selection influences the formation of codon usage bias in CLB.

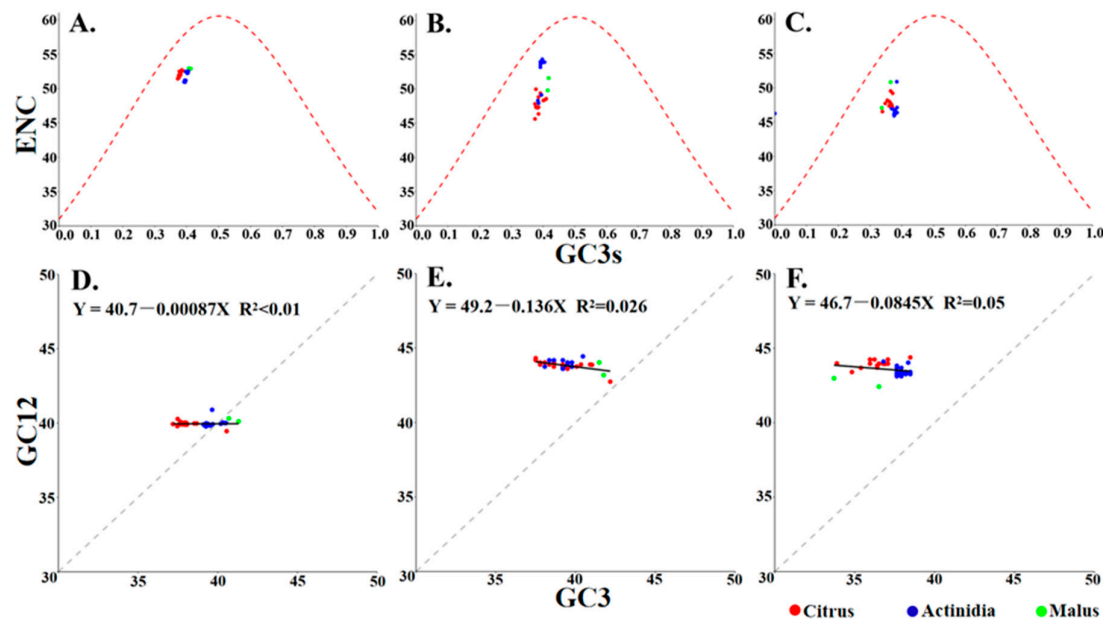


Figure 4. ENC-plot and neutrality analysis of the individual protein-coding sequences of CLBV. (A–C) ENC-plots for polyprotein, MP, and CP; (D–F) Neutrality plots for polyprotein, MP, and CP.

3.7. PR2 Analysis

In the PR2-bias plot, the absence of both directional mutation pressure and natural selection is represented by the intersection of the two reference lines ($X = 0.5$ and $Y = 0.5$), where the frequencies of G and C, as well as A and T, are equal. In the CP region, all points exhibited Y values > 0.5 , indicating a consistent bias toward A over T. Most citrus samples exhibited a clustering in the quadrant where X values exceeded 0.5, thereby revealing a pronounced preference for G over C, whereas Actinidia-derived samples were positioned close to $X = 0.5$, denoting nearly equal usage of G and C. Within the MP region, the majority of points fell below $Y = 0.5$, thus signifying a clear bias towards T relative to A. Citrus samples once again displayed a G-biased pattern ($X > 0.5$), whereas Actinidia samples showed a C-biased pattern ($X < 0.5$). In the polyprotein region, most points were located in the quadrant where both X and Y values were less than 0.5, indicating a simultaneous preference for T and C over their complementary nucleotides (Figure 5). Collectively, these findings revealed markedly heterogeneous nucleotide preferences among coding regions, with the same region displaying host-specific GC/AT skews in citrus versus Actinidia-infecting isolates. This implies that viral genome evolution is jointly shaped by the functional constraints and host environment.

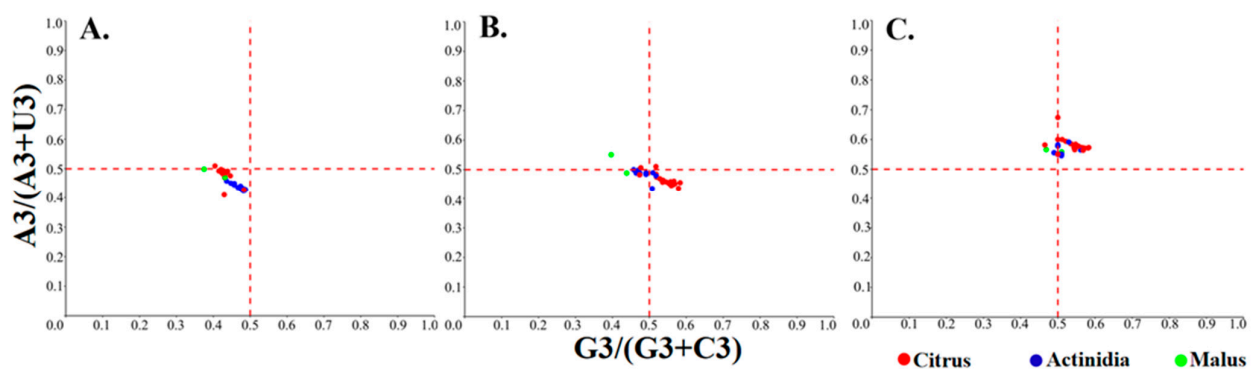


Figure 5. Parity Rule 2 analysis of the individual protein-coding sequences of CLBV. (A–C) PR2 plots for polyprotein, MP, and CP.

3.8. CAI Analysis of Sequences to Different Hosts

To quantify the adaptation and codon usage optimization of CLBV to its hosts, codon adaptation index (CAI) values were calculated. It is generally accepted that higher CAI values are indicative of stronger translational adaptation to the host tRNA pool, and that this, in turn, leads to superior host compatibility. The highest average CAI values were identified in *Prunus avium* (0.828–0.848) and the lowest were identified in *Nandina domestica* (0.654–0.666), indicating that among the documented hosts, *Prunus avium* represents the most suitable host, whereas *Nandina* as the least favorable (Figure 6). Consistently, among citrus cultivars, *Citrus sinensis* exhibited the highest CAI values, thereby further suggesting that *Citrus sinensis* is the most suitable citrus host for CLBV. Nevertheless, the current number of full genome sequences for certain hosts remains limited, which may introduce bias into the CAI estimation. Consequently, the incorporation of additional genome sequences will be essential to obtain a more robust and representative assessment.

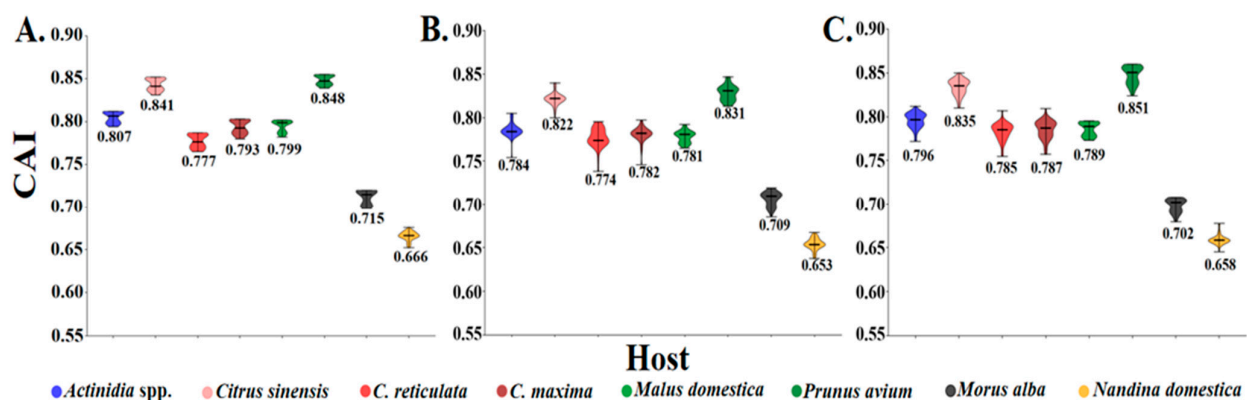


Figure 6. The codon adaptation index (CAI) analysis of the individual protein-coding sequences of CLBV. (A–C) The plots for polyprotein, MP, and CP.

4. Discussion

The present study utilized a genome-wide analysis of 46 non-recombinant CLBV to uncover that the virus evolution is predominantly shaped by host-specific translational selection. Despite the globally low codon usage bias, synonymous codon choice is stratified by host taxonomy rather than geography, phylogenies across all three ORFs recapitulate host lineage, and recombination breakpoints are disproportionately enriched in the coat protein gene, the very locus where host-determined codon signatures peak. Collectively, these data imply natural selection for tRNA compatibility as a major driver of CLBV divergence, with RNA recombination serving as an expedient route to import pre-adapted, host-optimized modules.

4.1. Recombinant Event Analysis

We detected ten unique recombination events, 80% of which map to the CP cistron and involve citrus-derived parental lineages. This breakpoint preference is unlikely to be stochastic: the CP is the interface between virion and host chaperones during cell-to-cell movement, and its surface-exposed residues are subject to strong selective sweeps. The AU-richness and predicted stem-loop structures within this region may promote RdRp template switching via the copy-choice mechanism [34,35], as documented in other flexiviruses. By shuffling CP modules that have already been fine-tuned for citrus translation machinery, CLBV can instantaneously increase fitness in new hosts, a strategy reminiscent of TYLCV recombinants that expanded their tropism to common genes [36,37]. Conse-

quently, recombination appears to act as an “evolutionary shortcut”, allowing the virus to sample pre-adapted sequences without the slow accumulation of point mutations.

4.2. Host-Driven Divergence Predominates over Geographic Structuring in CLBV Phylogeny

CLBV phylogenies exhibit pronounced host fidelity: *Actinidia* spp. isolates form a robust monophyletic clade, whereas citrus-derived sequences are dispersed across several lineages, mirroring the host-associated topology reported for CABYV [38]. This host-first, geography-second structure implicates adaptation to divergent host translational environments as the primary driver of CLBV divergence, a scenario reinforced by the parallel host-specific codon usage split most strikingly seen in the movement protein gene. Natural selection, not spatial isolation, is therefore the dominant force sculpting the genome of this generalist virus.

4.3. Genome-Wide and Gene-Specific Nucleotide Signatures Reflect Host-Driven Selection in CLBV

The genome of CLBV exhibits a pronounced nucleotide bias, with significant enrichment of adenine (A) and uracil (U) at multiple sites and a near-complete exclusion of CpG-containing CGN codons. This tactic is not unique: a substantial fraction of plant viruses also favor A-rich sequences [39,40], a preference driven by convergent evolutionary pressures including immune evasion, replication efficiency, translational optimization, and long-term adaptive selection [41–44]. Strikingly, the A-rich, C-depleted signature of CLBV mirrors the CpG-suppressed profiles documented in animal RNA viruses, implying that both plant and animal pathogens exploit analogous ‘linguistic camouflage’ to evade conserved antiviral defenses [45]. Discovered in animals, Zinc Finger Antiviral Protein (ZAP) recognizes CpG dinucleotides in viral RNA through its N-terminal RNA-binding domain containing four CCH-type zinc fingers and a specific binding pocket, and distinguishes non-self RNA based on the CpG suppression feature of the host genome. After binding to CpG, ZAP forms a complex with TRIM25, recruits nucleases to degrade viral RNA, and may also indirectly activate the interferon pathway, thereby triggering the host’s immune response. In order to circumvent the triggering of the host’s immune response, animal RNA viruses have evolved to reduce CpG content [46,47].

Beyond genome-wide compositional bias, CLBV also displays gene-specific codon usage patterns. Within the polyprotein region of citrus-adapted isolates, cysteine is almost exclusively encoded by TGT (RSCU = 1.94), whereas TGC is strongly avoided (RSCU = 0.06). Conversely, the MP region exhibits the opposite preference: TGC is favored (RSCU = 1.82) and TGT is disfavored (RSCU = 0.18). These reciprocal biases corroborate earlier reports that distinct genes within a single viral genome can experience different translational or selective environments [48,49].

There is a large difference in ENC values between the CP and MP genes of CLBV in *Actinidia* hosts. A similar pattern has been reported for the Usutu virus, where the ENC of the NS4A gene differs markedly from those of NS2B and NS4B. This discrepancy arises because the different genes of the Usutu virus experience distinct equilibria between selective and mutational pressures during evolution; for this flavivirus, codon usage is governed primarily by natural selection rather than by mutation pressure [50]. Analogously, codon usage in CLBV is also dominated by natural selection with only minor influence from mutation pressure, giving rise to the same inter-gene heterogeneity in ENC values.

4.4. Codon Adaptation Index

From the perspective of the CAI, *Prunus avium* appears to be the most adapted host, while *Nandina domestica* is the least adapted. A higher CAI is indicative of tighter adaptation to a host and, by extension, a potentially longer co-evolutionary history. Consequently, the

hypothesis can be formulated that *P. avium* represents an ancestral rather than a recently acquired host of CLB. However, given that the virus was initially reported in citrus and subsequently described in *Actinidia*, the majority of existing studies, and the corresponding genomic datasets have focused exclusively on these two hosts, with few studies focusing on cherry trees. Therefore, comprehensive investigations into the prevalence, geographic distribution, and genomic characteristics of CLB infecting cherry trees are urgently warranted. Meanwhile, the question of whether superior host adaptation translates into higher viral accumulation and, consequently, more pronounced symptoms remains an open question that requires systematic experimental validation in future studies. A parallel question has been raised in the context of citrus: CAI analyses indicate that CLB is more adapted to *Citrus sinensis* than to other citrus cultivars. Whether this enhanced codon adaptation corresponds to higher viral titres or more severe symptoms in *Citrus sinensis* remains to be empirically tested.

5. Conclusions

By integrating full-genome phylogenetics with multilayer codon usage analyses, we show that host-driven translational selection, rather than geographic distance, play a major role in the molecular evolution of CLB. Despite an overall low codon usage bias, CLB exhibits host-specific nucleotide and codon signatures that are strongest in the coat protein gene, a region that has been identified as a recombination hotspot. The virus achieves optimal adaptation in *Prunus avium* (highest CAI) and undergoes accelerated evolution in *Nandina domestica* (lowest CAI), indicating that synonymous site evolution is a reliable proxy for host compatibility. Natural selection, reinforced by modular RNA recombination, thus provides CLB with a rapid pathway to fine-tune its translational landscape across divergent hosts. These findings establish a genome-wide baseline for the monitoring of emerging CLB lineages and inform the design of host-adapted, codon-optimized viral vectors for sustainable pomiculture.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/v17111497/s1>, Table S1: Metadata for the CLB isolates used in this study; Table S2: Details of the nucleotide composition at synonymous third codon positions of CLB coding sequences; Table S3: RSCU values of CLB coding regions across host-associated sequence sets; Table S4: Dinucleotide analysis of CLB coding sequences.

Author Contributions: Conceptualization, A.H. and Y.W.; methodology, X.R. and L.X.; software, X.R. and L.X.; validation, Y.Y. and L.X.; formal analysis, X.R.; data curation, X.R. and A.H.; writing—original draft preparation, X.R.; writing—review and editing, A.H.; visualization, X.R. and A.H.; supervision, A.H. and Y.W.; project administration, A.H.; funding acquisition, A.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Science and Technology Project of Jiangxi Province (20225BCJ22005) and National Natural Science Foundation of China (grant numbers 32460646).

Data Availability Statement: All data are presented in the manuscript and the Supplementary Materials.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Guardo, M.; Sorrentino, G.; Marletta, T.; Caruso, A. First report of *Citrus leaf blotch virus* on Kumquat in Italy. *Plant Dis.* **2007**, *91*, 1054. [\[CrossRef\]](#)
2. Xuan, Z.Y.; Xie, J.X.; Yu, H.D.; Zhang, S.; Li, R.H.; Cao, M.J. Mulberry (*Morus alba*) is a new natural host of *Citrus leaf blotch virus* in China. *Plant Dis.* **2021**, *105*, 716. [\[CrossRef\]](#)
3. Wang, J.; Zhu, D.; Tan, Y.; Zong, X.; Wei, H.; Liu, Q. First report of *Citrus leaf blotch virus* in Sweet Cherry. *Plant Dis.* **2016**, *100*, 1027. [\[CrossRef\]](#)

4. Gress, J.C.; Smith, S.; Tzanetakis, I.E. First report of *Citrus leaf blotch virus* in Peony in the USA. *Plant Dis.* **2017**, *101*, 637. [\[CrossRef\]](#)
5. Li, H.F.; Zhao, Q.; Gray, S.M.; Xu, Y. Viral Small RNA-Based Screening of *Malus* spp. reveals *Citrus leaf blotch virus* infection of kaido crab apple in China. *Plant Dis.* **2020**, *104*, 3272. [\[CrossRef\]](#)
6. Liu, H.; Song, S.; Wu, W.; Mi, W.; Shen, C.; Bai, B.; Wu, Y. Distribution and molecular characterization of *Citrus leaf blotch virus* from actinidia in Shaanxi province, China. *Eur. J. Plant Pathol.* **2019**, *154*, 855–862. [\[CrossRef\]](#)
7. Velázquez, K.; Agüero, J.; Vives, M.C.; Aleza, P.; Pina, J.A.; Moreno, P.; Navarro, L.; Guerri, J. Precocious flowering of juvenile Citrus induced by a viral vector based on *Citrus leaf blotch virus*: A new tool for genetics and breeding. *Plant Biotechnol. J.* **2016**, *14*, 1976–1985. [\[CrossRef\]](#)
8. Niu, E.; Liu, H.; Zhou, H.; Luo, L.; Wu, Y.; Andika, I.B.; Sun, L. Autophagy inhibits intercellular transport of *Citrus leaf blotch virus* by targeting viral movement protein. *Viruses* **2021**, *13*, 2189. [\[CrossRef\]](#)
9. Umer, M.; Liu, J.; You, H.; Xu, C.; Dong, K.; Luo, N.; Kong, L.; Li, X.; Hong, N.; Wang, G.; et al. Genomic, morphological and biological traits of the viruses infecting major fruit trees. *Viruses* **2019**, *11*, 515. [\[CrossRef\]](#)
10. Chavan, R.R.; Blouin, A.G.; Cohen, D.; Pearson, M.N. Characterization of the complete genome of a novel citrivirus infecting *Actinidia chinensis*. *Arch. Virol.* **2013**, *158*, 1679–1686. [\[CrossRef\]](#)
11. Vives, M.C.; Martín, S.; Ambrós, S.; Renovell, A.; Navarro, L.; Pina, J.A.; Moreno, P.; Guerri, J. Development of a full-genome cDNA clone of *Citrus leaf blotch virus* and infection of citrus plants. *Mol. Plant Pathol.* **2008**, *9*, 787–797. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Renovell, A.; Vives, M.C.; Ruiz-Ruiz, S.; Navarro, L.; Moreno, P.; Guerri, J. The *Citrus leaf blotch virus* movement orotein acts as silencing suppressor. *Virus Genes* **2012**, *44*, 131–140. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Peng, Q.; Ning, J.; Xu, Q.; Yang, T.; Wang, Y.; Zheng, T.; Zhuang, Q.; Xi, D. Development and application of a reverse transcription loop-mediated isothermal amplification combined with lateral flow dipstick for rapid and visual detection of *Citrus leaf blotch virus* in kiwifruit. *Crop Prot.* **2021**, *143*, 105555. [\[CrossRef\]](#)
14. Yi, L.; Chen, Y.Q.; Chen, B.; Zhou, J. Occurrence and molecular characteristics of *Citrus leaf blotch virus* from citrus in China based on coat protein genes. *Trop. Plant Pathol.* **2021**, *46*, 714–718. [\[CrossRef\]](#)
15. Hershberg, R.; Petrov, D.A. Selection on codon bias. *Annu. Rev. Genet.* **2008**, *42*, 287–299. [\[CrossRef\]](#)
16. Grantham, R.; Gautier, C.; Gouy, M.; Mercier, R.; Pavé, A. Codon catalog usage and the genome hypothesis. *Nucleic Acids Res.* **1980**, *8*, r49–r62. [\[CrossRef\]](#)
17. Bulmer, M. The selection-mutation-drift theory of synonymous codon usage. *Genetics* **1991**, *128*, 897–907. [\[CrossRef\]](#)
18. Reis, M.; Savva, R.; Wernisch, L. Solving the riddle of codon usage preferences: A test for translational selection. *Nucleic Acids Res.* **2004**, *32*, 5036–5044. [\[CrossRef\]](#)
19. Sharp, P.M.; Emery, L.R.; Zeng, K. Forces that influence the evolution of codon bias. *Philos. Trans. R. Soc. B Biol. Sci.* **2010**, *365*, 1203–1212. [\[CrossRef\]](#)
20. Bennetzen, J.L.; Hall, B.D. Codon selection in yeast. *J. Biol. Chem.* **1982**, *257*, 3026–3031. [\[CrossRef\]](#)
21. van Hemert, F.; van der Kuyl, A.C.; Berkhout, B. Impact of the biased nucleotide composition of viral RNA genomes on RNA structure and codon usage. *J. Gen. Virol.* **2016**, *97*, 2608–2619. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Hussain, S.; Rasool, S.T. Analysis of synonymous codon usage in Zika virus. *Acta Trop.* **2017**, *173*, 136–146. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Gómez, M.M.; Volotão, E.M.; Assandri, I.R.; Peyrou, M.; Cristina, J. Analysis of codon usage bias in potato virus Y non-recombinant strains. *Virus Res.* **2020**, *286*, 198077. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Qin, L.; Ding, S.; Wang, Z.; Jiang, R.; He, Z. Host plants shape the codon usage pattern of turnip mosaic virus. *Viruses* **2022**, *14*, 2267. [\[CrossRef\]](#)
25. Wang, Y.; Chi, C.; Zhang, J.; Zhang, K.; Deng, D.; Zheng, W.; Chen, N.; Meurens, F.; Zhu, J. Systematic analysis of the codon usage patterns of African swine fever virus genome coding sequences reveals its host adaptation phenotype. *Microb. Genom.* **2024**, *10*, 001186. [\[CrossRef\]](#)
26. Sharp, P.M.; Li, W.-H. An evolutionary perspective on synonymous codon usage in unicellular organisms. *J. Mol. Evol.* **1986**, *24*, 28–38. [\[CrossRef\]](#)
27. Bansal, S.; Mallikarjuna, M.G.; Balamurugan, A.; Nayaka, S.C.; Prakash, G. Composition and codon usage pattern results in divergence of the zinc binuclear cluster (Zn(II)2Cys6) sequences among ascomycetes plant pathogenic fungi. *J. Fungi* **2022**, *8*, 1134. [\[CrossRef\]](#)
28. Buitrago, S.P.; Garzón-Ospina, D. Genetic diversity of SARS-CoV-2 in South America: Demographic history and structuration signals. *Arch. Virol.* **2021**, *166*, 3357–3371. [\[CrossRef\]](#)
29. Zhang, K.; Wang, Y.; Zhang, Y.; Shan, X.F. Codon usage characterization and phylogenetic analysis of the mitochondrial genome in *Hemerocallis citrina*. *BMC Genom. Data* **2024**, *25*, 6. [\[CrossRef\]](#)
30. Patil, S.S.; Indrabalan, U.B.; Suresh, K.P.; Shome, B.R. Analysis of codon usage bias of classical swine fever virus. *Vet. World* **2021**, *14*, 1450–1458. [\[CrossRef\]](#)
31. Li, J.; Guo, Y.; Roellig, D.M.; Li, N.; Feng, Y.; Xiao, L. *Cryptosporidium felis* differs from other *Cryptosporidium* spp. in codon usage. *Microb. Genom.* **2021**, *7*, 000711. [\[CrossRef\]](#)

32. Fang, J.; Zheng, L.; Liu, G.; Zhu, H. Comparative analysis of chloroplast genomes in *Cephaleuros* and its related genus (*Trentepohlia*): Insights into adaptive evolution. *Genes* **2024**, *15*, 839. [\[CrossRef\]](#) [\[PubMed\]](#)
33. 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.* **1987**, *15*, 1281–1295. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Wang, H.; Cui, X.; Cai, X.; An, T. Recombination in positive-strand RNA viruses. *Front. Microbiol.* **2022**, *13*, 870759. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Bull, R.A.; Hansman, G.S.; Clancy, L.E.; Tanaka, M.M.; Rawlinson, W.D.; White, P.A. Norovirus recombination in ORF1/ORF2 overlap. *Emerg. Infect. Dis.* **2005**, *11*, 1079–1085. [\[CrossRef\]](#)
36. Lefeuvre, P.; Moriones, E. Recombination as a motor of host switches and virus emergence: Geminiviruses as case studies. *Curr. Opin. Virol.* **2015**, *10*, 14–19. [\[CrossRef\]](#)
37. Cui, J.; Li, F.; Shi, Z.-L. Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol.* **2019**, *17*, 181–192. [\[CrossRef\]](#)
38. Maachi, A.; Donaire, L.; Hernando, Y.; Aranda, M.A. Genetic differentiation and migration fluxes of viruses from melon crops and crop edge weeds. *J. Virol.* **2022**, *96*, e00421–22. [\[CrossRef\]](#)
39. He, Z.; Qin, L.; Xu, X.; Ding, S. Evolution and host adaptability of plant RNA viruses: Research insights on compositional biases. *Comput. Struct. Biotechnol. J.* **2022**, *20*, 2600–2610. [\[CrossRef\]](#)
40. Qin, L.; Ding, S.; He, Z. Compositional biases and evolution of the largest plant RNA virus order *Patatavirales*. *Int. J. Biol. Macromol.* **2023**, *240*, 124403. [\[CrossRef\]](#)
41. Anderson, B.R.; Muramatsu, H.; Jha, B.K.; Silverman, R.H.; Weissman, D.; Karikó, K. Nucleoside modifications in RNA limit activation of 2′-5′-oligoadenylate synthetase and increase resistance to cleavage by RNase L. *Nucleic Acids Res.* **2011**, *39*, 9329–9338. [\[CrossRef\]](#)
42. Wang, X.; Zhu, B. SARS-CoV-2 nsp15 preferentially degrades AU-rich dsRNA via its dsRNA nickase activity. *Nucleic Acids Res.* **2024**, *52*, 5257–5272. [\[CrossRef\]](#) [\[PubMed\]](#)
43. Bera, B.C.; Virmani, N.; Kumar, N.; Anand, T.; Pavulraj, S.; Rash, A.; Elton, D.; Rash, N.; Bhatia, S.; Sood, R.; et al. Genetic and codon usage bias analyses of polymerase genes of equine influenza virus and its relation to evolution. *BMC Genom.* **2017**, *18*, 652. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Tyagi, N.; Sardar, R.; Gupta, D. Natural selection plays a significant role in governing the codon usage bias in the novel SARS-CoV-2 variants of concern (VOC). *PeerJ* **2022**, *10*, e13562. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Kustin, T.; Stern, A. Biased mutation and selection in RNA viruses. *Mol. Biol. Evol.* **2021**, *38*, 575–588. [\[CrossRef\]](#)
46. Lo, R.; Gonçalves-Carneiro, D. Sensing nucleotide composition in virus RNA. *Biosci. Rep.* **2023**, *43*, BSR20230372. [\[CrossRef\]](#)
47. Gonçalves-Carneiro, D.; Takata, M.A.; Ong, H.; Shilton, A.; Bieniasz, P.D. Origin and evolution of the zinc finger antiviral protein. *PLoS Pathog.* **2021**, *17*, e1009545. [\[CrossRef\]](#)
48. Xiong, B.; Wang, T.; Huang, S.; Liao, L.; Wang, X.; Deng, H.; Zhang, M.; He, J.; Sun, G.; He, S.; et al. Analysis of codon usage bias in xyloglucan endotransglycosylase (XET) genes. *Int. J. Mol. Sci.* **2023**, *24*, 6108. [\[CrossRef\]](#)
49. Hockenberry, A.J.; Sirer, M.I.; Amaral, L.A.N.; Jewett, M.C. Quantifying position-dependent codon usage bias. *Mol. Biol. Evol.* **2014**, *31*, 1880–1892. [\[CrossRef\]](#)
50. Zhou, J.; Xing, Y.; Zhou, Z.; Wang, S. A comprehensive analysis of Usutu virus (USUV) genomes revealed lineage-specific codon usage patterns and host adaptations. *Front. Microbiol.* **2023**, *13*, 967999. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.