

## Article

# Host Plants Shape the Codon Usage Pattern of Turnip Mosaic Virus

Lang Qin <sup>1</sup>, Shiwen Ding <sup>1</sup>, Zhilei Wang <sup>1</sup>, Runzhou Jiang <sup>1</sup> and Zhen He <sup>1,2,\*</sup><sup>1</sup> College of Plant Protection, Yangzhou University, Wenhui East Road No.48, Yangzhou 225009, China<sup>2</sup> Joint International Research Laboratory of Agriculture and Agri-Product Safety of Ministry of Education of China, Yangzhou University, Yangzhou 225009, China

\* Correspondence: hezhen@yzu.edu.cn

**Abstract:** Turnip mosaic virus (TuMV), an important pathogen that causes mosaic diseases in vegetable crops worldwide, belongs to the genus *Potyvirus* of the family *Potyviridae*. Previously, the areas of genetic variation, population structure, timescale, and migration of TuMV have been well studied. However, the codon usage pattern and host adaptation analysis of TuMV is unclear. Here, compositional bias and codon usage of TuMV were performed using 184 non-recombinant sequences. We found a relatively stable change existed in genomic composition and a slightly lower codon usage choice displayed in TuMV protein-coding sequences. Statistical analysis presented that the codon usage patterns of TuMV protein-coding sequences were mainly affected by natural selection and mutation pressure, and natural selection was the key influencing factor. The codon adaptation index (CAI) and relative codon deoptimization index (RCDI) revealed that TuMV genes were strongly adapted to *Brassica oleracea* from the present data. Similarity index (SiD) analysis also indicated that *B. oleracea* is potentially the preferred host of TuMV. Our study provides the first insights for assessing the codon usage bias of TuMV based on complete genomes and will provide better advice for future research on TuMV origins and evolution patterns.

**Keywords:** turnip mosaic virus; codon usage; composition bias; natural selection; host adaptation



**Citation:** 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. <https://doi.org/10.3390/v14102267>

Academic Editors: Tao Zhou and Zhiyou Du

Received: 14 September 2022

Accepted: 14 October 2022

Published: 15 October 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 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

Turnip mosaic virus (TuMV) belongs to one of the largest genera of plant RNA viruses, namely, *Potyvirus*, which is in the family *Potyviridae* [1]. TuMV is known to infect a wide range of plant species that mostly belong to the family *Brassicaceae* [2]. In nature, TuMV can be transmitted by aphids in a non-persistent manner. The size of TuMV virions is approximately 720 nm × 15~20 nm, and they are composed of 95% coat protein and 5% RNA. TuMV virions contain a positive single-stranded RNA molecule, which is approximately 9830 nucleotides (nts) in length. The five ends of the TuMV genome are covalently linked to a viral coding protein (VPg). The genome has a main open reading frame (ORF) encoding a large polyprotein and non-translated regions on each end of the molecule. Under the action of viral encoded proteases, a total of 10 functional proteins are obtained after the proteolytic process of the polyprotein, such as the first protein (P1; the molecular weight is 40 kDa), helper component protease (HC-Pro; 52 kDa), protein 3 (P3; 40 kDa), the first 6 kDa protein (6K1; 6 kDa), cylindrical inclusion body protein (CI; 72 kDa), the second 6 kDa protein (6K2; 6 kDa), encoding viral genome-related protein (VPg; 22 kDa), small nuclear inclusion body a (NIa; 27 kDa), small nuclear inclusion body b (NIb; 60 kDa), and coat protein (CP; 33 kDa) [1]. A small, overlapping ORF encodes a truncated frameshift product, namely, PIPO protein [3].

Normally, the genetic code allows 61 triplet codons to encode 20 amino acids, and codons encoding the same amino acid are termed synonymous codons [4,5]. Intriguingly, synonymous codons are not randomly used, the synonymous codons are also not used equally in various organisms or even in different gene groups of the same genome, creating

a bias in codon usage, which is known as codon usage bias (CUB) [6–9]. Codon usage patterns are influenced by many factors, such as mutation pressure, compositional constraints, natural selection, gene length, replication, hydrophobicity, selective transcription, gene function, secondary protein structure, and the external environment [4,6,8–14]. Evolution, adaptation, overall viral fitness, and evasion of host cell responses and survival are affected by the codon usage bias of viruses and their hosts. Due to their low codon usage bias, most RNA viruses can reduce their level of competition with host genes and thereby effectively replicate in host cells.

As one of the most-studied plant-infecting RNA viruses in the area of evolution, the genetic variation, population structure, evolutionary rate, timescale, and migration of TuMV isolated from *Raphanus sativus*, *Brassica oleracea*, *Brassica juncea*, *Brassica rapa*, *Sisymbrium loeseli*, and *Rapistrum rugosum* in Belgium, China, Greece, Germany, South Korea, Japan, Iran, Turkey, UK, Poland, Russia, USA, Slovakia, Mexico, Italy, and Brazil have been stated based on analyses of partial or complete genome sequences [15–21]. To date (August 2022), one hundred and eighty-four genomic sequences of TuMV non-recombinant isolates from South Korea, China, Iran, Turkey, the UK, Poland, Russia, Greece, Belgium, Germany, Italy, The Czech Republic, Australia, and Brazil have been reported [22,23]. However, the synonymous codon usage pattern of TuMV is still not fully reported.

In the present study, we conducted detailed codon usage analyses of TuMV non-recombinant isolates based on 184 genomic sequences to assess the evolutionary adaptation of this virus to its hosts. We explored the factors that shape the codon usage patterns of TuMV and provided a new perspective on the genetic divergence of TuMV. To the best of our knowledge, this study provides the first insights into the codon usage patterns of TuMV.

## 2. Materials and Methods

### 2.1. Virus Isolates

One hundred and eighty-four genomic sequences of TuMV non-recombinant isolates were retrieved from GenBank [22,23]. The details of those isolates, such as geographical location, date of collection, and host, are shown in Table S1.

### 2.2. Recombination and Phylogenetic Analysis

All of the TuMV sequences that are described in Table S1 were aligned using CLUSTAL X2 [24]. Putative recombination incidents of aligned TuMV were identified by several methods within the RDP4 software package [25], such as the GENECONV, RDP, BOOTSCAN, 3SEQ, CHIMAERA, MAXCHI, and SISCAN programs [26–32]. Through a phylogenetic approach in the RDP4 package, parent/donor assignments were proved. There were at least four different methods ( $p$ -value of  $<1.0 \times 10^{-6}$ ) in the RDP4 package that supported the putative recombinants. These analyses were performed by the different detection programs using the default settings.

We used the neighbor-joining (NJ) method in MEGA v11 [33] to assess the phylogenetic relationships of the polyprotein-coding sequences of TuMV. The NJ analyses used were evaluated by Kimura's two-parameter with 1000 bootstrap replicates [34]. The inferred trees were displayed using TreeView [35].

### 2.3. Nucleotide Composition Analysis

The nucleotide composition of TuMV polyprotein and the 11 protein-coding sequences were calculated after removing five non-bias codons, such as UGA, UAG, and UAA (termination codons) and UGG and AUG (the only codons encoding Trp and Met, respectively). The component parameters of the TuMV polyprotein and the 11 protein-coding sequences were then calculated. The entire nucleotide composition (e.g., A, C, U, and G%) and the total AU and GC contents were calculated using BioEdit version 5.0.9 [16]. CodonW 1.4.2 package was used for the analysis of the nucleotide composition at the third codon position

of the TuMV coding sequences (e.g., A3, C3, U3, and G3%). The GC contents of the first base of codon (GC1), the second base, and the third base of codon (GC2, GC3) were employed for analysis in an online program (<http://www.bioinformatics.nl/emboss-explorer/> (accessed on 25 August 2022)), where GC12 is the mean of GC1 and GC2.

#### 2.4. Relative Synonymous Codon Usage (RSCU) Analysis

The RSCU value of a codon is the ratio between the observed and expected usage frequencies [36]. The RSCU values were calculated using the following formula:

$$RSCU_{ij} = \frac{g_{ij}}{\sum_j^{ni} g_{ij}} \times ni \quad (1)$$

In this formula,  $RSCU_{ij}$  is the value of the  $i$ -th codon for the  $j$ -th amino acid, the  $g_{ij}$  is the observed number of the  $i$ -th codon for the  $j$ -th amino acid, and “ $ni$ ” kind of represents the degenerate numbers of synonymous codons which encode the  $j$ -th amino acid. An RSCU value of 1 suggests no bias for the codon. While codons with RSCU values  $<0.6$  and  $>1.6$  are defined as low and high-frequency codons, respectively. MEGA v11 software was used to calculate the RSCU values of the TuMV polyprotein and the 11 protein-coding sequences [33]. The available coding sequences of *R. sativus*, *B. rapa*, *B. oleracea*, and *R. juncea* were downloaded from the GenBank database. The host RSCU values were calculated using MEGA v11 software [33].

#### 2.5. Principal Component Analysis (PCA)

A multivariate statistical method called PCA was used to identify the correlations between variables and samples. After removing the three termination codons and UGG and AUG codons, a 59-dimensional vector was used to represent each strain of the 12 data sets where different dimensions corresponded to each sense codon’s RSCU value. PCA analysis was used by Origin 8.0.

#### 2.6. Effective Number of Codons Analysis (ENC)

The ENC values, which were calculated using CodonW v1.4.2 software and indicate the degree of codon usage bias, ranged from 20 (an extreme codon usage bias for which only one synonymous codon was used) to 61 (no bias, the synonymous codons were equally used) [37]. The ENC values were calculated as:

$$ENC = 2 + \frac{9}{\bar{F}_2} + \frac{1}{\bar{F}_3} + \frac{5}{\bar{F}_4} + \frac{3}{\bar{F}_6} \quad (2)$$

where  $\bar{F}_k$  ( $k = 2, 3, 4, 6$ ) is on behalf of the average of  $F_k$ , and  $k$  indicates the  $k$ -fold degenerate amino acids.  $F_k$  is estimated as follows:

$$F_k = \frac{nS - 1}{n - 1} \quad (3)$$

where  $n$  is the total number of the observed values of the codon for the corresponding amino acid and

$$S = \sum_{i=1}^k \left( \frac{n_i}{n} \right)^2 \quad (4)$$

where  $n_i$  stands for the total number of the  $i$ -th codon for that amino acid.

The ENC analysis is used to measure the absolute codon usage bias of the TuMV genes. Typically, a gene with ENC values  $\leq 35$  indicates significant CUB. It is considered that smaller ENC values show stronger CUB.

### 2.7. ENC-Plot Analysis

ENC-plot analysis (with the GC3s value on the horizontal ordinate and the ENC value on the longitudinal coordinate) was used to provide the role of mutation pressure in codon usage bias. When mutation pressure is the only factor, the dot lies on or around the standard curve. Otherwise, it is influenced by selection and other factors. The expected ENC was conducted as:

$$ENC\ expected = 2 + s + \left( \frac{29}{s^2 + (1 - s)^2} \right) \quad (5)$$

where  $s$  represents the value of GC3s.

### 2.8. Parity Rule 2 Analysis (PR2)

Applying PR2 bias plots to investigate the influence of natural selection and mutation pressure on the codon usage of the TuMV. The value of AU-bias ( $A3/(A3 + U3)$ ) as the ordinate against GC-bias ( $G3/(G3 + C3)$ ) as the abscissa, respectively. The center of the plot is 0.5, which indicates a balance between mutation pressure and natural selection.

### 2.9. Neutrality Analysis

In the neutrality plot graph, GC12 and GC3 are shown as the ordinate and abscissa, respectively. The mutational force is represented by the slope of the regression line between GC12 and GC3 contents. If there is no selection pressure or the selection pressure is weak, the slope of the regression line is near 1.0. Conversely, if the regression line slope deviates from 1.0, indicating that natural selection has a key role in codon bias.

### 2.10. Codon Adaptation Index (CAI) Analysis

The CAI analysis was computed by a web server (<http://genomes.urv.cat/CAIcal/RCDI/> (accessed on 25 August 2022)) and was used to predict the adaptation of individual TuMV genes to their potential host. Normally, higher CAI values (e.g., from 0 to 1) indicate stronger adaptability to the host. Due to the lack of relevant codon usage data of the hosts *R. rugosum* and *S. loeselii*, the CAI analysis was performed by the remaining four hosts.

### 2.11. Relative Codon Deoptimization Index (RCDI) Analysis

The RCDI values are calculated using an online program (<http://genomes.urv.cat/CAIcal/RCDI/> (accessed on 25 August 2022)) for the TuMV polyprotein, and the 11 protein-coding sequences were used to identify trends in codon deoptimization. If the RCDI values were equal to 1, this indicated that the virus displayed a host-adapted codon usage pattern. Conversely, RCDI values higher than 1 indicate lower adaptability due to the lack of relevant codon usage data of the hosts *R. rugosum* and *S. loeselii*. The RCDI analysis was performed by the remaining four hosts.

### 2.12. Similarity Index (SiD) Analysis

SiD analysis is a widely used method for determining the effect of the codon usage bias of hosts. The SiD value was calculated as:

$$R(A, B) = \frac{\sum_{i=1}^{59} a_i b_i}{\sqrt{\sum_{i=1}^{59} b_i^2 \sum_{i=1}^{59} a_i^2}} \quad (6)$$

$$D(A, B) = \frac{1 - R(A, B)}{2} \quad (7)$$

where  $a_i$  represents the RSCU values of 59 synonymous codons of the TuMV coding sequences, and  $b_i$  represents the RSCU values of identical codons of the host. The potential impact of the host's entire codon usage on the different clades of the TuMV gene is repre-

sented by the SiD [ $D(A, B)$ ] (from 0 to 1.0). Higher values generally indicate that the host plays a significant role in codon usage.

### 3. Results

#### 3.1. Recombination and Phylogenetic Analysis

Generally, recombination can influence the topology of phylogenetic trees and overall codon usage patterns regardless of gene or genome levels [38,39]. A total of 184 TuMV non-recombinant coding sequences [22,23] from *B. juncea*, *B. oleracea*, *B. rapa*, *R. sativus*, *R. rugosum*, and *S. loeselii* were used in the following phylogenetic and codon usage analyses.

Phylogenetic analyses were conducted using the NJ methods based on the complete polyprotein of TuMV. The NJ trees that are based on complete polyprotein sequences are shown (Figure S1). Six lineages with certain degrees of host origins were formed based on the complete polyprotein-coding sequences (Figure S1). As Yasaka et al. (2017), Kawakubo et al. (2021), and Kawakubo et al. (2022) reported, these six major genetic groups that are based on polyprotein-coding sequences were clustered into the Orchis, Asian-BR, basal-B, basal-BR, Iranian, and world-B groups.

#### 3.2. Nucleotide Bias Analysis

The nucleotide compositions of the complete polyprotein and 11 protein-coding sequences of TuMV were assessed to explore the effect of compositional constraints on codon usage. For the polyprotein, nucleotides A and G were most abundant, with mean compositions of  $32.09 \pm 0.55\%$  and  $24.21 \pm 0.51\%$  (Table S2), respectively, and were followed by U ( $22.55 \pm 0.38\%$ ) and C ( $21.15 \pm 0.45\%$ ). Similarly, for the individual protein-coding sequences, we also found that the nucleotides A and G were most abundant in the P1, HC-Pro, 6K2, VPg, NIb, CP, and PIPO coding regions (Table S2), while nucleotides A and U were rich in P3, CI, and NIa coding regions. The nucleotides A and C were most abundant in the 6K1 (Table S2). However, the third position's nucleotide composition of synonymous codons (e.g., A3S, U3S, G3S, and C3S) was inconsistent with the nucleotide composition at the complete polyprotein level. The most frequent nucleotide was A3S ( $39.55 \pm 1.78\%$ ), which was followed by C3S ( $30.82 \pm 1.35\%$ ), U3S ( $28.96 \pm 1.21\%$ ), and G3S ( $29.03 \pm 1.70\%$ ) (Table S2). For the protein-coding sequences, the A3S and G3S were only found to be most abundant in the coding sequences of the CP, NIb, and PIPO coding regions (Table S2); compared with the nucleotides A3S and U3S, which were most abundant in the P3 and CI coding region sequences, the nucleotides A3S and C3S were most abundant in the 6K1, 6K2, HC-Pro, NIa, P1, and VPg coding sequences (Table S2). The composition of AG is better than the UC of complete polyproteins and 11 protein-coding sequences (Table S2), indicating that there is an AG-rich composition for the TuMV coding sequences.

#### 3.3. Relative Synonymous Codon Usage Analysis of TuMV and Its Hosts

RSCU analysis was conducted to estimate the codon usage patterns of TuMV polyprotein and 11 protein-coding sequences. Thirteen of the 18 preferred codons were A/C (A-ended 7, C-ended 6) in the complete polyprotein-coding region (Table 1). A/C-ended codons were also preferred in individual protein-coding sequences, P1 (A-ended 5, C-ended 6), HC-Pro (A-ended 8, C-ended 5), 6K2 (A-ended 5, C-ended 7), Vpg (A-ended 6, C-ended 6), NIa (A-ended 7, C-ended 5), and PIPO (A-ended 8, C-ended 5), except P3 (A-ended 6, U-ended 5), 6K1 (A-ended 6, U-ended 6), CI (A-ended 8, U-ended 4) and CP (C-ended 5, G-ended 5) (Table 1). The results show that the A/C-terminal codon is slightly popular in the TuMV coding sequence. Among these preferred codons in the complete polyprotein-coding region, the RSCU values of four codons were  $>1.6$ , and the highest value was for CCA (2.36), indicating extreme overrepresentation, and the remaining preferred codons had RSCU values  $>0.6$  and  $<1.6$ . Additionally, to determine the potential influences of hosts on the codon usage patterns of the TuMV isolates, the RSCU patterns of the TuMV polyprotein-coding sequences were correlated with those of *B. juncea*, *B. oleracea*, *B. rapa*, *R. sativus*, *R. rugosum*, and *S. loeselii*. Seventeen of the 18 preferred codons were A/U

-ended (A-ended: 5; U-ended: 12) for *B. juncea*, and *R. rugosum* had a similar pattern of use, with A-ended: 6, U-ended: 12. Whereas the four hosts *B. oleracea*, *B. rapa*, *R. sativus*, and *S. loeselii* had almost the same codon usage pattern, 13 of the 18 preferred codons were C/U-ended (C-ended: 4; U-ended: 9) (Table S3). Overall, a mixture of antagonism and coincidence was discovered in the codon usage patterns of TuMV and its six hosts based on polyprotein-coding sequences (Table S3).

**Table 1.** The relative synonymous codon usage (RSCU) value of 59 codons encoding 18 amino acids according to 11 protein and polyprotein of turnip mosaic virus.

Codon	aa	TuMV											
		P1	HC	P3	6K1	CI	6K2	Vpg	N1a	N1b	CP	PIPO	Pol
TTT	F	0.98	0.66	<b>1.02</b>	<b>1.41</b>	0.95	0.37	0.79	0.90	0.81	0.91	<b>1.75</b>	0.87
TTC	F	<b>1.02</b>	<b>1.34</b>	0.98	0.59	<b>1.05</b>	<b>1.63</b>	<b>1.21</b>	<b>1.10</b>	<b>1.19</b>	<b>1.09</b>	0.24	<b>1.13</b>
TTA	L	0.65	0.72	1.35	0.81	0.62	0.49	0.49	1.04	0.50	<b>1.32</b>	2.35	0.80
TTG	L	<b>1.76</b>	<b>1.31</b>	<b>1.44</b>	0.78	0.90	1.16	1.08	<b>1.31</b>	<b>1.41</b>	1.02	<b>2.40</b>	<b>1.25</b>
CTT	L	0.66	0.82	0.68	1.00	<b>1.73</b>	<b>1.69</b>	1.13	0.89	0.89	0.93	0.07	1.03
CTC	L	1.29	0.97	0.68	<b>2.53</b>	1.31	0.75	<b>1.68</b>	1.08	1.12	1.15	0.01	1.14
CTA	L	0.79	1.23	1.11	0.70	0.98	1.07	0.57	1.26	0.87	0.83	0.44	0.98
CTG	L	0.86	0.95	0.75	0.17	0.46	0.85	1.05	0.42	1.19	0.75	0.74	0.80
ATT	I	<b>1.22</b>	0.65	0.87	0.81	0.81	<b>1.45</b>	0.85	1.11	0.74	<b>1.20</b>	0.26	0.90
ATC	I	1.13	1.12	<b>1.36</b>	0.62	1.06	1.11	0.89	0.76	<b>1.15</b>	0.84	<b>1.73</b>	<b>1.08</b>
ATA	I	0.65	<b>1.23</b>	0.77	<b>1.57</b>	<b>1.12</b>	0.44	<b>1.26</b>	<b>1.13</b>	1.10	0.96	1.01	1.02
GTT	V	1.05	1.11	0.90	0.94	<b>1.14</b>	1.31	0.67	0.68	<b>1.03</b>	0.90	0.09	1.02
GTC	V	0.94	<b>1.18</b>	<b>1.08</b>	<b>1.57</b>	0.86	0.22	0.47	1.10	1.02	0.66	<b>2.81</b>	0.95
GTA	V	0.79	0.70	0.97	0.39	0.90	0.82	0.52	0.67	0.98	1.20	0.32	0.84
GTG	V	<b>1.23</b>	1.01	1.05	1.11	1.11	<b>1.65</b>	<b>2.34</b>	<b>1.56</b>	0.97	<b>1.25</b>	0.78	<b>1.19</b>
TCT	S	0.42	0.56	0.67	1.56	0.78	0.88	1.61	0.49	0.21	1.18	0.47	0.62
TCC	S	0.69	0.51	0.39	1.55	0.49	0.83	0.46	0.71	0.32	1.16	0.37	0.55
TCA	S	1.38	<b>1.73</b>	1.06	0.09	<b>1.53</b>	1.24	1.07	1.18	<b>1.78</b>	0.60	<b>1.64</b>	1.38
TCG	S	0.38	0.71	0.50	0.00	0.53	0.23	0.06	0.47	1.27	0.09	1.00	0.57
AGT	S	<b>1.61</b>	1.10	<b>1.88</b>	1.15	1.20	0.34	0.61	1.34	1.11	<b>1.61</b>	1.02	1.35
AGC	S	1.53	1.40	1.50	<b>1.65</b>	1.47	<b>2.48</b>	<b>2.19</b>	<b>1.81</b>	1.32	1.36	1.50	<b>1.52</b>
CCT	P	0.67	0.44	0.62	2.30	0.62	1.63	0.81	0.42	0.87	0.16	0.00	0.62
CCC	P	0.60	0.32	0.33	1.54	0.34	<b>2.28</b>	0.42	0.24	0.38	0.65	0.39	0.43
CCA	P	<b>2.15</b>	<b>2.70</b>	<b>2.64</b>	0.15	<b>2.45</b>	0.09	<b>2.26</b>	<b>2.54</b>	<b>2.41</b>	<b>1.79</b>	<b>2.61</b>	<b>2.36</b>
CCG	P	0.58	0.53	0.41	0.00	0.59	0.00	0.51	0.81	0.34	1.40	0.02	0.59
ACT	T	0.66	0.84	1.03	1.33	0.87	0.42	0.54	0.89	0.89	0.65	0.00	0.81
ACC	T	1.33	0.57	0.66	0.25	0.57	1.38	0.74	0.81	0.72	0.76	0.40	0.76
ACA	T	<b>1.54</b>	<b>1.65</b>	<b>1.49</b>	<b>2.13</b>	<b>2.01</b>	0.76	<b>2.16</b>	<b>1.21</b>	<b>1.80</b>	1.11	<b>2.22</b>	<b>1.67</b>
ACG	T	0.47	0.93	0.82	0.30	0.55	<b>1.45</b>	0.56	1.08	0.59	<b>1.48</b>	0.83	0.76
GCT	A	0.76	0.78	0.83	0.29	0.89	0.00	<b>1.50</b>	0.77	0.89	0.83	1.21	0.85
GCC	A	0.88	0.45	0.81	0.82	0.61	<b>0.07</b>	1.15	0.69	0.30	0.52	0.19	0.64
GCA	A	<b>1.81</b>	<b>2.15</b>	<b>1.65</b>	<b>2.20</b>	<b>1.81</b>	0.03	0.92	<b>2.09</b>	<b>2.13</b>	<b>2.04</b>	<b>2.35</b>	<b>1.88</b>
GCG	A	0.55	0.62	0.71	0.69	0.69	0.03	0.43	0.45	0.69	0.61	0.24	0.63
TAT	Y	0.73	0.75	0.63	0.63	<b>1.07</b>	0.00	0.75	0.90	<b>1.13</b>	0.61	0.58	0.88
TAC	Y	<b>1.27</b>	<b>1.25</b>	<b>1.37</b>	<b>1.37</b>	0.93	0.00	<b>1.25</b>	<b>1.10</b>	0.87	<b>1.39</b>	<b>1.42</b>	<b>1.12</b>
CAT	H	0.68	0.94	<b>1.17</b>	<b>1.31</b>	0.79	0.44	0.67	0.00	0.60	<b>1.24</b>	0.00	0.87
CAC	H	<b>1.32</b>	<b>1.06</b>	0.83	0.69	<b>1.21</b>	<b>1.56</b>	<b>1.33</b>	<b>0.97</b>	<b>1.40</b>	0.76	0.00	<b>1.13</b>
CAA	Q	0.95	<b>1.02</b>	<b>1.24</b>	<b>1.53</b>	0.85	<b>1.02</b>	0.88	<b>1.33</b>	<b>1.03</b>	0.80	<b>1.33</b>	<b>1.02</b>
CAG	Q	<b>1.05</b>	0.98	0.76	0.47	<b>1.15</b>	0.97	<b>1.12</b>	0.00	0.97	<b>1.20</b>	0.67	0.98
AAT	N	0.53	0.79	<b>1.11</b>	0.20	0.91	0.82	0.90	0.69	0.76	0.83	0.03	0.83
AAC	N	<b>1.47</b>	<b>1.21</b>	0.89	<b>1.76</b>	<b>1.09</b>	<b>1.18</b>	<b>1.10</b>	<b>1.31</b>	<b>1.24</b>	<b>1.17</b>	<b>1.69</b>	<b>1.17</b>
AAA	K	0.73	0.97	0.97	0.60	<b>1.02</b>	<b>1.09</b>	0.83	<b>1.11</b>	0.91	0.89	<b>1.31</b>	0.92
AAG	K	<b>1.27</b>	<b>1.03</b>	<b>1.03</b>	<b>1.40</b>	0.98	0.91	<b>1.17</b>	0.89	<b>1.09</b>	<b>1.11</b>	0.69	1.08
GAT	D	0.55	<b>1.10</b>	<b>1.23</b>	<b>1.14</b>	<b>1.02</b>	0.83	<b>1.08</b>	<b>1.02</b>	<b>1.19</b>	0.92	<b>1.21</b>	<b>1.05</b>
GAC	D	<b>1.45</b>	0.90	0.77	0.86	0.98	<b>1.17</b>	0.92	0.98	0.81	<b>1.08</b>	0.79	0.95
GAA	E	0.94	0.92	<b>1.08</b>	<b>1.23</b>	0.94	<b>1.80</b>	<b>1.20</b>	0.70	<b>1.18</b>	0.90	<b>1.15</b>	<b>1.01</b>
GAG	E	<b>1.06</b>	<b>1.08</b>	0.92	0.77	<b>1.06</b>	0.20	0.80	<b>1.30</b>	0.82	<b>1.10</b>	0.85	0.99

Table 1. Cont.

Codon	aa	TuMV											
		P1	HC	P3	6K1	CI	6K2	Vpg	NIa	NIb	CP	PIPO	PoI
TGT	C	0.33	<b>1.08</b>	<b>1.36</b>	<b>1.64</b>	<b>1.24</b>	<b>1.77</b>	<b>1.79</b>	<b>1.16</b>	<b>1.01</b>	0.00	<b>1.96</b>	<b>1.12</b>
TGC	C	<b>1.66</b>	0.92	0.64	0.36	0.76	0.23	0.21	0.84	0.99	<b>2.00</b>	0.03	0.88
CGT	R	0.59	0.53	0.41	1.08	0.53	1.07	0.85	0.56	0.36	1.19	0.34	0.63
CGC	R	0.86	1.33	1.10	0.70	0.37	0.49	0.50	0.21	1.03	0.75	1.48	0.77
CGA	R	0.75	0.71	0.53	<b>1.52</b>	1.17	<b>2.39</b>	1.12	1.56	<b>1.90</b>	1.15	0.70	1.09
CGG	R	0.21	0.32	0.10	0.62	0.44	1.21	0.13	1.14	0.72	0.43	0.12	0.41
AGA	R	<b>2.07</b>	<b>1.81</b>	<b>2.30</b>	1.43	<b>2.44</b>	0.48	1.66	<b>1.85</b>	1.23	<b>1.43</b>	<b>3.24</b>	<b>1.90</b>
AGG	R	1.52	1.29	1.56	0.65	1.05	0.36	<b>1.74</b>	0.69	0.77	1.06	0.13	1.20
GGT	G	0.66	0.95	1.11	0.11	1.07	0.74	1.03	<b>1.25</b>	0.95	<b>1.31</b>	0.48	1.01
GGC	G	0.63	1.15	<b>1.33</b>	0.02	0.67	0.83	0.80	0.74	0.72	1.23	<b>3.52</b>	0.86
GGA	G	<b>1.89</b>	<b>1.40</b>	1.23	<b>3.41</b>	<b>1.53</b>	<b>2.34</b>	<b>1.72</b>	1.24	<b>1.67</b>	1.18	0.01	<b>1.54</b>
GGG	G	0.82	0.50	0.32	0.46	0.73	0.08	0.45	0.77	0.66	0.27	1.75	0.59

The most frequently used codons are shown in bold.

### 3.4. Trends in Codon Usage Variations

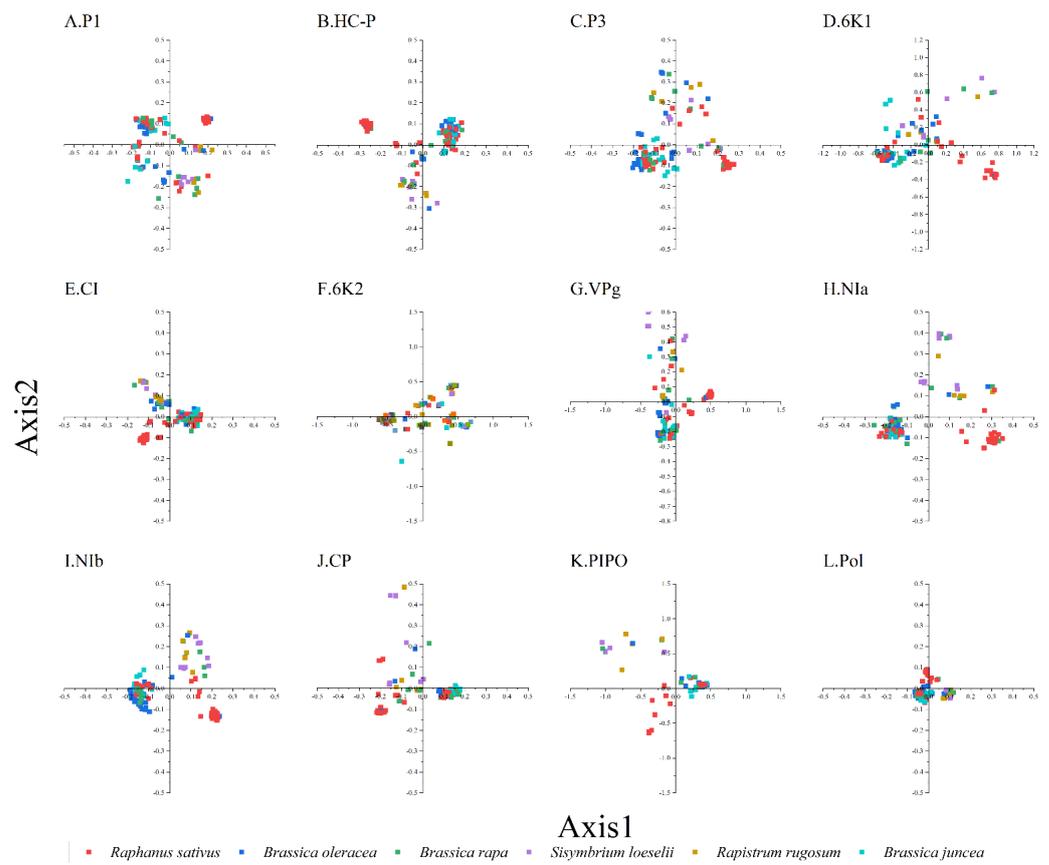
To study the synonymous codon usage variations in the coding sequences of TuMV, principal component analysis was used. The first four axes (axes 1–4) of the complete polyprotein and individual protein-coding sequences were recorded for more than 60% of the variation (Figure S2). In addition, it can be seen from the figure that axis 1 is the key factor that affects the codon usage for the TuMV coding regions (Figure 1). Moreover, based on the RSCU values on the first two axes, we discovered the distribution of the complete polyprotein and 11 protein-coding sequences in different hosts (Figure S2). We found obvious an overlap between the different hosts in the PCA analysis of TuMV complete polyprotein and individual protein-coding regions, which suggests distinct codon usage trends (Figure 1).

### 3.5. Codon Usage Bias of TuMV

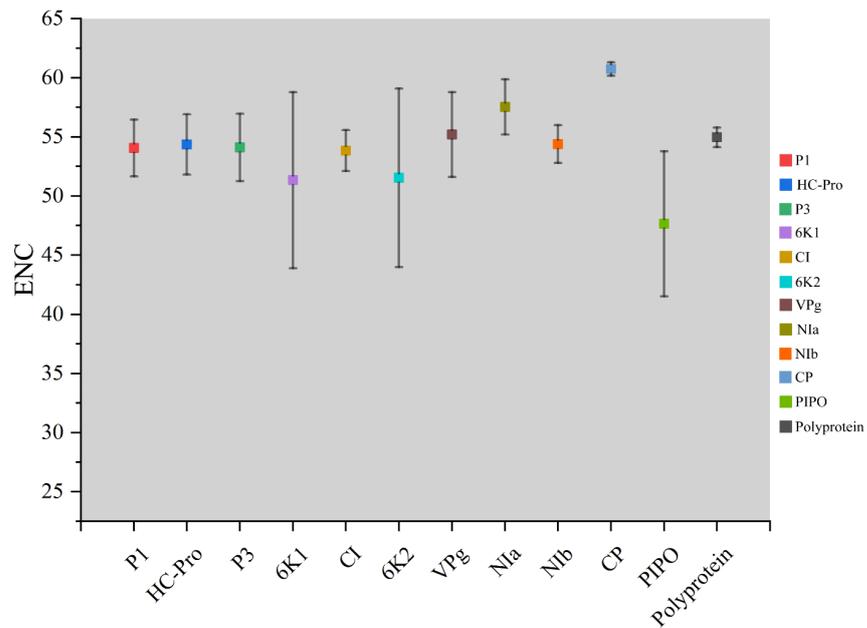
The ENC values were calculated to show the magnitude of the choice of TuMV genome codon usage. Individually, maximum ENC values were observed for the CP coding sequences, while minimum values were found in the PIPO coding sequences (Figure 2). Polyprotein and 10 coding sequences of TuMV, the average ENC values were all more than 45 (Table S2) (Figure 2). These results suggested that there was a relatively conserved nucleotide composition with slightly lower codon usage choice in the TuMV coding sequences.

### 3.6. ENC-Plot Analysis

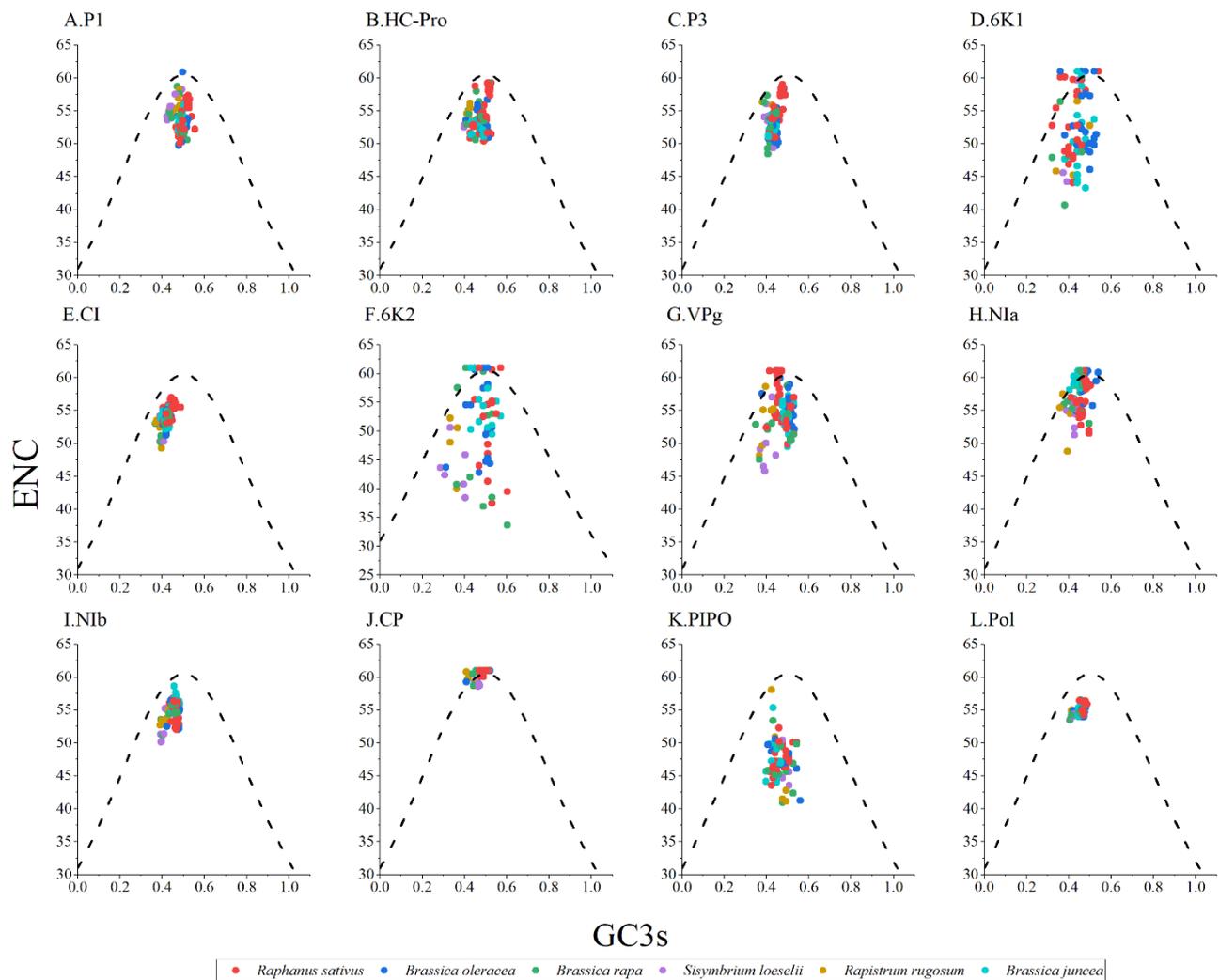
ENC-GC3s plot analysis was performed to study the forces that influenced the codon usage bias of the TuMV protein-coding regions. Generally, if the points fall below the expected curve, it means that the codon usage is more strongly affected by natural selection pressure. However, mutation pressure is indicated when the data points fall on the expected curve. As shown in Figure 3, the TuMV isolates from different hosts typically cluster below the expected curve; it implies that natural selection dominated over mutation pressure, while the influence of mutation was not completely absent (Figure 3).



**Figure 1.** Principal component analysis (PCA) is based on the relative synonymous codon usage (RSCU) values of the 59 synonymous codons for the individual protein (A–K), complete polyprotein (L), and coding sequences of turnip mosaic virus. The *Raphanus sativus*, *Brassica oleracea*, *Brassica rapa*, *Sisymbrium loeseli*, *Rapistrum rugosum* and *Brassica juncea* hosts are represented in red, blue, green, purple, orange, wathet blue and brown, respectively.



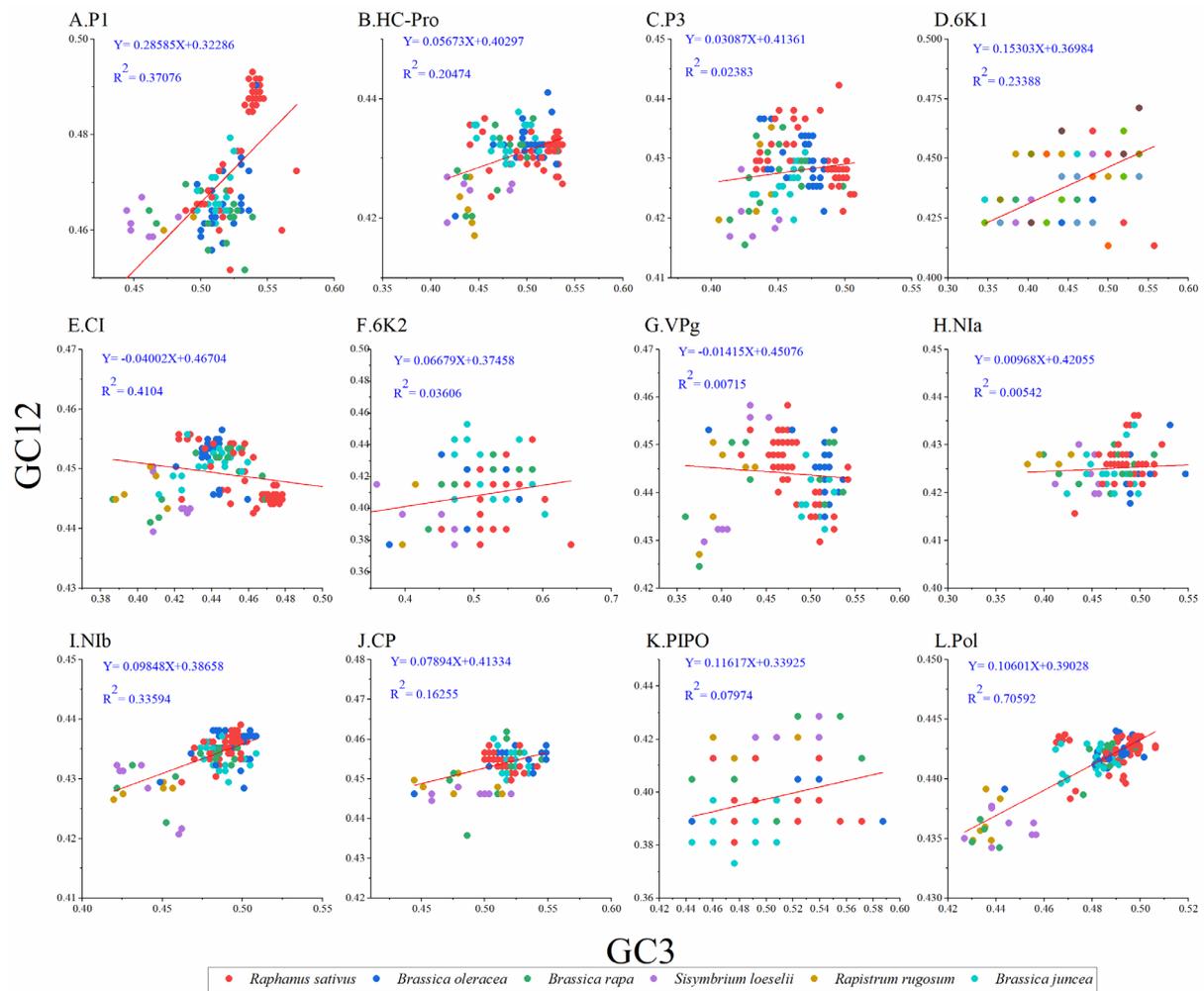
**Figure 2.** Effective number of codons analysis (ENC) values for the eleven protein and complete polyprotein-coding sequences of turnip mosaic virus.



**Figure 3.** Effective number of codons analysis (ENC)-plot analysis of the individual protein (A–K) and complete polyprotein (L) coding sequences of turnip mosaic virus, with ENC against the GC3s of different hosts. The black dotted line represents the standard curve when the codon usage bias is determined by only the GC3s composition. The *Raphanus sativus*, *Brassica oleracea*, *Brassica rapa*, *Sisymbrium loeseli*, *Rapistrum rugosum* and *Brassica juncea* hosts are represented in red, blue, green, purple, orange, wathet blue and brown, respectively.

### 3.7. Neutrality Plot

To unravel the extent of influence between mutation pressure and natural selection on codon usage in TuMV, we performed a neutrality analysis between GC12 and GC3. Normally, nucleotide changes at the third position of the codon do not result in amino acid changes, which are considered to reflect only a mutational force. Whereas, if a nucleotide change produces a change in the amino acid, it is considered a mutation pressure. Among the protein-coding sequences of TuMV, significant positive correlations were observed between the GC12 and GC3 values for the TuMV polyprotein (Figure 4) and the P1, HC-Pro, P3, 6K1, 6K2, NIa, NIb, PIPO, and CP coding sequences (Figure 4A, B, C, D, F, H, I, J, and K respectively); in contrast, the GC12 and GC3 values for the TuMV CI and VPg coding sequences (Figure 4G) showed no significant correlations. The slope of the linear regression for the polyprotein-coding sequences was 0.106 (Figure 4), indicating that mutation pressure accounted for 10.6% of the pressure on codon usage, while natural selection accounted for 89.4% of the pressure. All of these results showed that natural selection was the principal force driving the TuMV coding sequences' codon usage bias.



**Figure 4.** Neutrality plot analysis of GC3s against GC12s for the individual protein (A–K) and complete polyprotein (L) coding sequences of turnip mosaic virus. The *Raphanus sativus*, *Brassica oleracea*, *Brassica rapa*, *Sisymbrium loeseli*, *Rapistrum rugosum* and *Brassica juncea* hosts are represented in red, blue, green, purple, orange, wathet blue and brown, respectively.

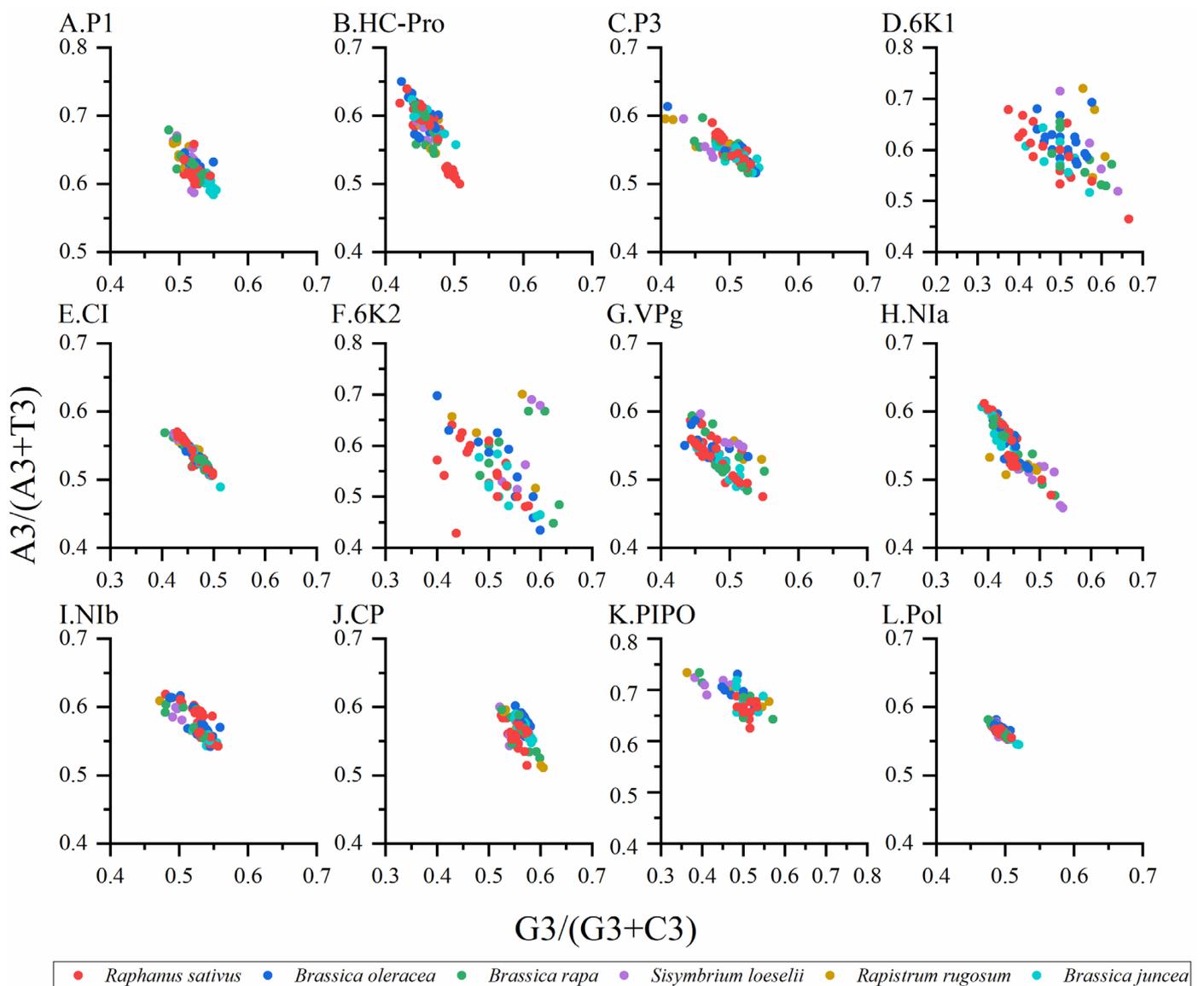
### 3.8. Parity Analysis

Normally, when PR2 biases at the third codon position are plotted in four-codon sequences of individual genes, it is considered that the PR2 plots are especially useful. Therefore, we constructed PR2 plots to confirm the influence of mutation pressure and natural selection on the CUB. When the plot lies in the center (e.g., A = U and G = C), both coordinates are 0.5, and no bias is present in the selection or mutation pressure [6]. The results showed that nucleotides A was more frequently used than U, while nucleotides G and C were used at similar frequencies in the TuMV coding sequences (Figure 5A–L), which indicated that the codon usage bias of TuMV was also shaped by natural selection and other factors.

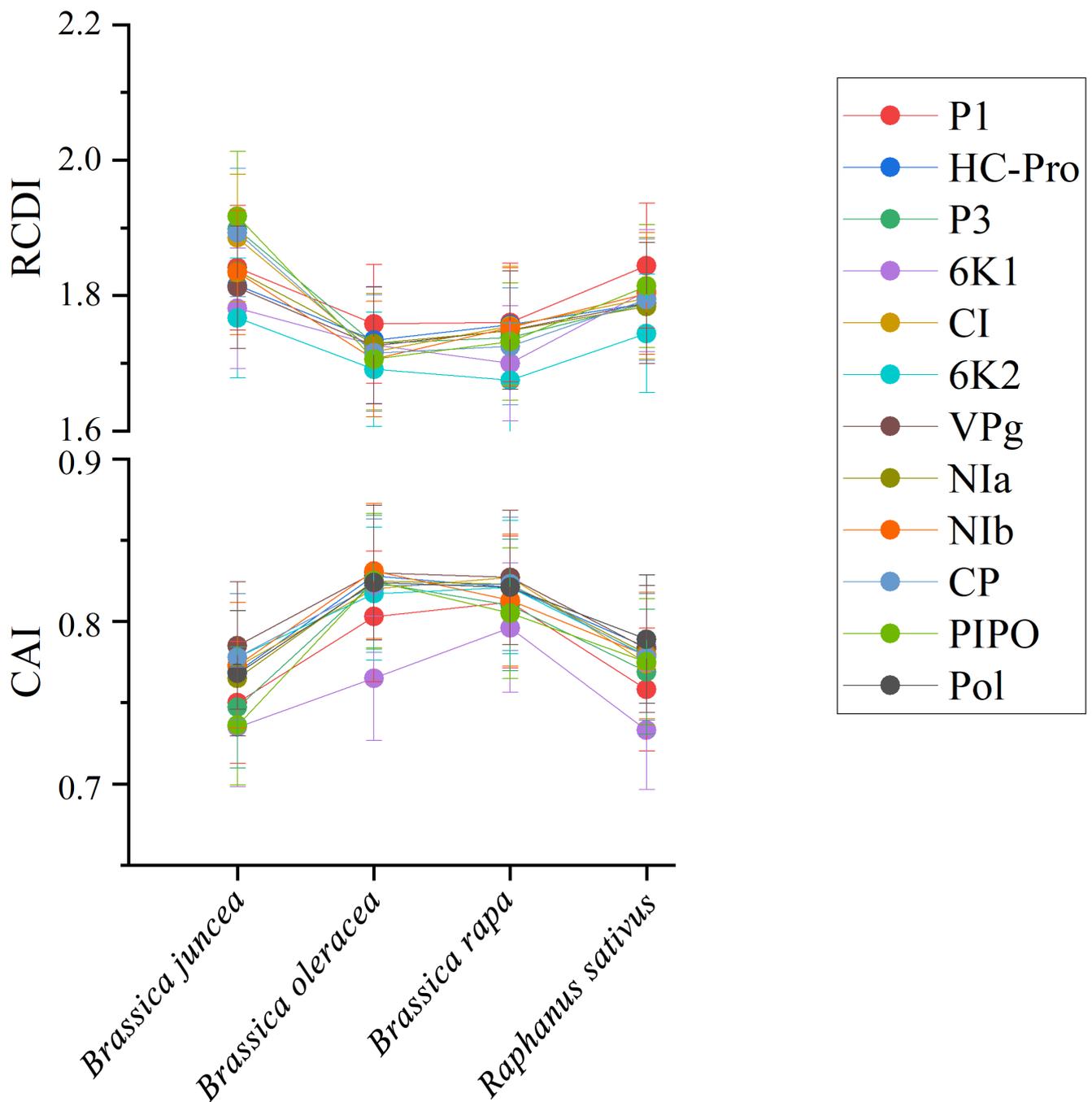
### 3.9. Codon Usage Adaptation in TuMV

To quantify the adaptation and codon usage optimization of TuMV to its hosts, codon adaptation index values were calculated. Normally, genes with higher CAI values are more suitable for the host than those with lower CAI values. The average CAI values of polyprotein sequences were 0.824, 0.821, 0.768, and 0.789 for *B. oleracea*, *B. rapa*, *B. juncea*, and *R. sativus* respectively, whereas the highest values for the eleven coding sequences were identified in *B. oleracea* (Figure 6). These results suggest that *B. oleracea* was the most suitable host of TuMV. Additionally, RCDI analysis was conducted to show the cumulative effects

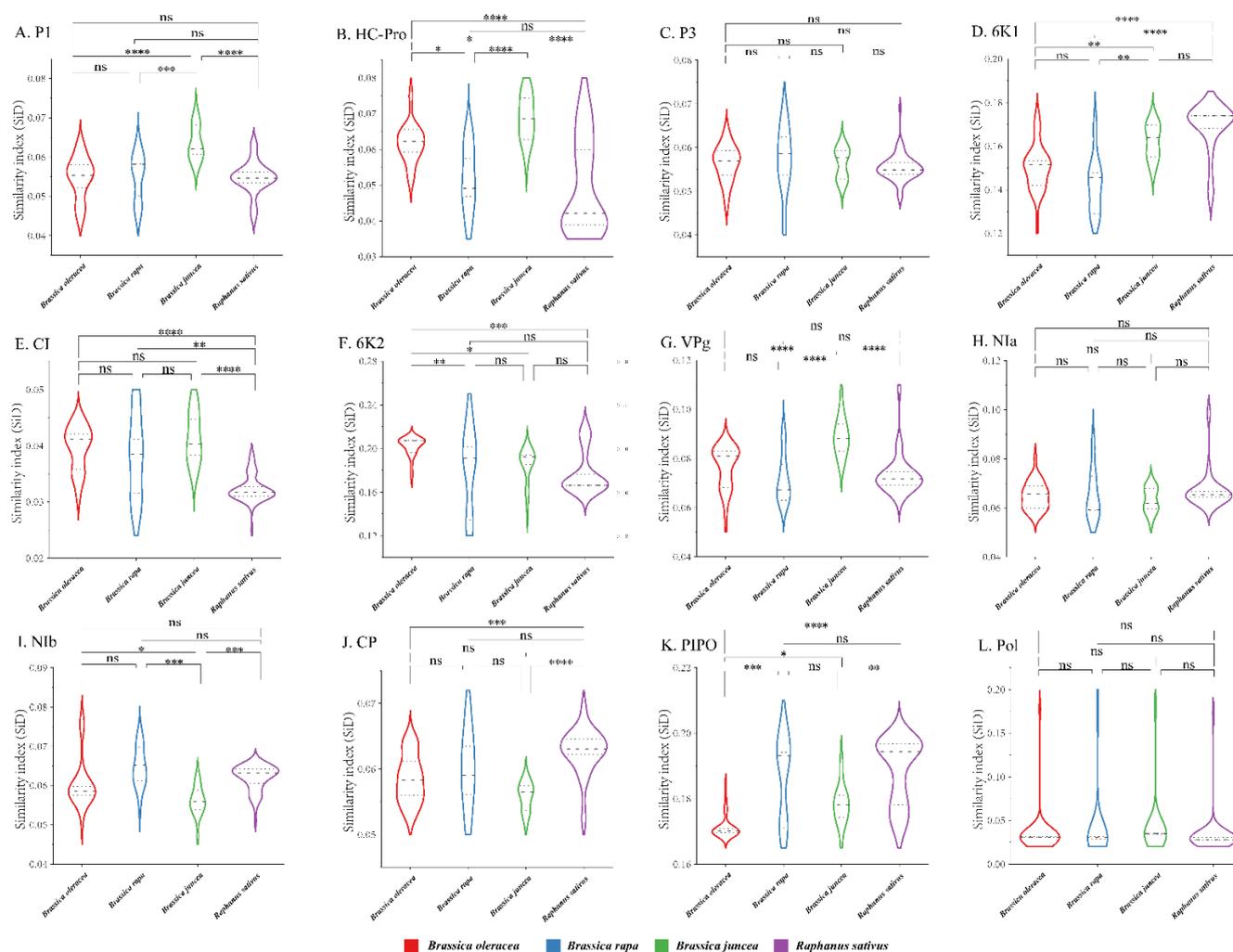
of codon bias on a single gene expression. The means of the RCDI values were highest for *B. juncea*, and the lowest RCDI values were observed for *B. oleracea* (Figure 6), indicating that codon usage deoptimization was highest for *B. juncea* and lowest for *B. oleracea*. Then, a SiD analysis was performed to understand how the codon usage patterns of *B. oleracea*, *B. rapa*, *B. juncea*, and *R. sativus* affected the TuMV codon usage pattern (Figure 7). The SiD value among the complete polyproteins is similar for *B. juncea* and *B. oleracea* (Figure 7). In the 11 protein-coding sequences of TuMV, the highest SiD values were observed in *B. oleracea*, *B. juncea*, and *R. sativus* (except P3). Combining the above CAI and RCDI analysis shows that through TuMV evolution, *B. oleracea* perhaps had a greater impact on the virus than the other hosts from the present data.



**Figure 5.** Parity plot showing the presence of AT bias [ $A3/(A3 + T3)$ ] and GC bias [ $G3/(G3 + C3)$ ] for the individual protein (A–K) and complete polyprotein (L) coding sequences of turnip mosaic virus. The center of the plot, where the value of both the coordinates is 0.5, indicates the place where there is no bias in the mutation or selection rates. The *Raphanus sativus*, *Brassica oleracea*, *Brassica rapa*, *Sisymbrium loeselii*, *Rapistrum rugosum* and *Brassica juncea* hosts are represented in red, blue, green, purple, orange, wathet blue and brown, respectively.



**Figure 6.** The codon adaptation index (CAI) analysis and relative codon deoptimization index (RCDI) analysis of the eleven protein and complete polyprotein-coding sequences of turnip mosaic virus in relation to the natural hosts. The x-axis indicates the sequences isolated from different hosts.



**Figure 7.** The similarity index (SiD) analysis of the individual protein (A–K) and complete polyprotein (L) coding sequences of turnip mosaic virus in relation to the natural hosts. One-way ANOVA and Tukey’s test were employed to compare the mean of the SiD values pertaining to the different hosts. Asterisk indicated the differential SiD value of turnip mosaic virus between four hosts is statistically significant or very significant ( $p < 0.001$  or  $p < 0.0001$ ), “ns”, not significant,  $p > 0.05$ . \*  $p < 0.05$ ; \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ ; \*\*\*\*  $p < 0.0001$ .

#### 4. Discussion

TuMV is an important viral disease in vegetable crops, and predecessors also published many reports about vegetable crop virus disease [40–42]. Previously, the genetic evolution of TuMV in terms of phylogenetics, dynamics, and migration was effectively performed and based on analyses of complete or partial genome sequences in Europe, the Middle East, East Asia, and Oceania [15–22]. Previous studies have reported that approximately 75% of the isolates from TuMV populations are recombinants [17,43]. Phylogenetic analyses performed by Ohshima et al. (2002), Nguyen et al. (2013), and Kawakubo et al. (2022), which were based on the complete polyprotein sequence, found six divergent evolutionary lineages [15,17]. Recently, Yasaka et al. (2017) and Kawakubo et al. (2021) reported that six lineages were gathered based on TuMV non-recombinant sequences from Europe, the Middle East, East Asia, and Oceania [21].

Codon usage patterns of viruses reflect evolutionary changes, such as adaption, evolution, evasion from host immune systems, and survival [44–52]. Now, only limited reports show the codon usage patterns of plant viruses, such as citrus tristeza virus (CTV) [50], rice stripe virus (RSV) [51], papaya ringspot virus (PRSV) [52], potato virus M (PVM) [53], sug-

arcane mosaic virus (SCMV) [54], broad bean wilt virus 2 (BBWV2) [55], rice black-streaked dwarf virus (RBSDV) [56], narcissus degeneration virus (NDV) [57], narcissus late season yellows virus (NLSYV) [57], and narcissus yellow stripe virus (NYSV) [57]. Here, the composition and codon usage patterns of TuMV based on the complete genome were estimated. In general, genomes with AU-rich virus compositions tend to contain codons ending with A and U as opposed to viral genomes with GC-rich compositions, which tend to contain codons that end with G and C [43,46,47,53]. In this study, in comparison, A/C-terminal codons are more strongly preferred in TuMV gene sequences. The nucleotide bias analysis in previous studies showed that the composition constraint was mainly affected by the preferred codon [43,46,47,53]. In the present study, A/G was found to be most abundant in the TuMV protein-coding regions, which supports the presence of mutation pressure.

Low codon usage biases were also observed for RSV, CTV, PRSV, PVM, SCMV, BBWV2, RBSDV, NDV, NLSYV, and NYSV [50–57]. For TuMV, similar lower codon usage patterns were also found with ENC values higher than 35, which indicated a low degree of preference. Additionally, the neutrality plot, ENC-plot, and PR2 analyses showed that the evolution of the TuMV genome has been shaped by mutation and natural selection to varying degrees. Moreover, the neutrality plot and ENC-plot analyses indicated that natural selection is the major factor that induces the codon usage bias of TuMV, which is consistent with PVM and SCMV [53,54].

Previously, several studies have shown that codon usage patterns could affect virus host-specific adaptations [43,46,48,53,54]. In the present study, TuMV and its host adaptations were assessed from the viewpoint of codon usage bias. CAI analysis demonstrated that TuMV genes were more strongly adapted to *B. oleracea* than to *B. juncea*, *B. rapa*, and *R. sativus*. Furthermore, RCDI analysis showed that strong codon usage deoptimization occurred in *B. oleracea*. Generally, low RCDI values mean strong adaptations to hosts [58]. Thus, both the CAI and RCDI results were consistent. Our SiD analysis indicated that the selection pressure of host plants on TuMV was similar because almost consistent SiD values for isolates were observed from *B. oleracea*, *B. juncea*, *B. rapa*, and *R. sativus* based on polyprotein. However, *B. oleracea*, *B. juncea*, and *R. sativus* showed differences in impacts on the evolution of TuMV 11 protein-coding sequences. It is worth noting that only reference genome sequences of four host species were used in this study. The host population panel is not adequately represented in this study due to only the reference genomes of some species are available in the database. Similarly, during the evolution of the Zika virus (ZIKV), both the CAI and RCDI results showed that the Zika virus (ZIKV) was most strongly adapted to *Aedes aegypti* or *Homo sapiens* while in the SiD analysis, *Ae. albopictus* is potentially the new, preferred vector of ZIKV because the selection pressure exerted by *Ae. albopictus* on codon usage patterns was greater than the selection pressure imposed by *Ae. aegypti* or *H. sapiens* [43].

In conclusion, the detailed codon usage patterns of TuMV were studied for the first time according to complete genome sequences to gain knowledge into the genetic evolution and host adaptability of TuMV. Our study also provides a better understanding of the evolutionary changes of TuMV, which should be considered for the prevention and control of this virus.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/v14102267/s1>. Figure S1: Neighbor-Joining (NJ) trees calculated from the polyprotein sequences of non-recombinant turnip mosaic virus. Numbers at each node indicate the percentage of bootstrap samples in the NJ trees. Horizontal branch lengths are drawn to scale, with the bar indicating 0.1 nt replacements per site; Figure S2: Relative and cumulative inertia of the 35 axes from a COA of the RSCU values based on the individual protein (A–K) and complete polyprotein (L) coding sequences of turnip mosaic virus; Table S1: The turnip mosaic virus isolates used in this study; Table S2: The detail of codons in the individual protein encoding regions of turnip mosaic virus; Table S3: The relative synonymous codon usage (RSCU) value of 59 codons of 18 amino acids of turnip mosaic virus polyprotein and the RSCU value of different hosts.

**Author Contributions:** Conceptualization, Z.H.; methodology, Z.W., R.J., S.D. and L.Q.; software, Z.H. and L.Q.; validation, Z.H., S.D. and L.Q.; formal analysis, Z.H.; writing—original draft preparation, Z.H. and S.D.; writing—review and editing, Z.H., S.D. and L.Q.; supervision, Z.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by grants from the National Natural Science Foundation of China (No. 32272485), Natural Science Foundation of Jiangsu Province of China (BK20211323), High-Level Talent Support Program of Yangzhou University, the Qing Lan Project of Yangzhou University.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The Tables S1–S3 in supplementary file.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Wylie, S.J.; Adams, M.; Chalam, C.; Kreuze, J.; López-Moya, J.J.; Ohshima, K.; Praveen, S.; Rabenstein, F.; Stenger, D.; Wang, A.; et al. ICTV Virus Taxonomy Profile: *Potyviridae*. *J. Gen. Virol.* **2017**, *98*, 352–354. [[CrossRef](#)] [[PubMed](#)]
2. Walsh, J.A.; Jenner, C.E. Turnip Mosaic Virus and the Quest for Durable Resistance. *J. Mol. Plant. Pathol.* **2002**, *3*, 289–300. [[CrossRef](#)] [[PubMed](#)]
3. Chung, B.Y.W.; Miller, W.A.; Atkins, J.F.; Firth, A.E. An Overlapping Essential Gene in the *Potyviridae*. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 5897–5902. [[CrossRef](#)] [[PubMed](#)]
4. Hasegawa, M.; Yasunaga, T.; Miyata, T. Secondary Structure of MS2 Phage RNA and Bias in Code Word Usage. *Nucleic. Acids. Res.* **1979**, *7*, 2073–2079. [[CrossRef](#)] [[PubMed](#)]
5. Sharp, P.M.; Tuohy, T.M.F.; Mosurski, K.R. Codon Usage in Yeast: Cluster Analysis Clearly Differentiates Highly and Lowly Expressed Genes. *Nucleic. Acids. Res.* **1986**, *14*, 5125–5143. [[CrossRef](#)] [[PubMed](#)]
6. Sueoka, N. Directional Mutation Pressure and Neutral Molecular Evolution. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 2653–2657. [[CrossRef](#)] [[PubMed](#)]
7. Sharp, P.M.; Cowe, E. Synonymous Codon Usage in *Saccharomyces Cerevisiae*. *Yeast* **1991**, *7*, 657–678. [[CrossRef](#)] [[PubMed](#)]
8. Comeron, J.M.; Aguadé, M. An Evaluation of Measures of Synonymous Codon Usage Bias. *J. Mol. Evol.* **1998**, *47*, 268–274. [[CrossRef](#)]
9. Hershberg, R.; Petrov, D.A. Selection on Codon Bias. *Annu. Rev. Genet.* **2008**, *42*, 287–299. [[CrossRef](#)]
10. Kyte, J.; Doolittle, R.F. A Simple Method for Displaying the Hydropathic Character of a Protein. *J. Mol. Biol.* **1982**, *157*, 105–132. [[CrossRef](#)]
11. Sueoka, N. Translation-Coupled Violation of Parity Rule 2 in Human Genes Is Not the Cause of Heterogeneity of the DNA G+C Content of Third Codon Position. *Gene* **1999**, *238*, 53–58. [[CrossRef](#)]
12. Duret, L.; Mouchiroud, D. Expression Pattern and, Surprisingly, Gene Length Shape Codon Usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 4482–4487. [[CrossRef](#)]
13. Fuglsang, A. Accounting for Background Nucleotide Composition When Measuring Codon Usage Bias: Brilliant Idea, Difficult in Practice. *Mol. Biol. Evol.* **2006**, *23*, 1345–1347. [[CrossRef](#)]
14. Coleman, J.R.; Papamichail, D.; Skiena, S.; Futcher, B.; Wimmer, E.; Mueller, S. Virus Attenuation by Genome-Scale Changes in Codon Pair Bias. *Science* **2008**, *320*, 1784–1787. [[CrossRef](#)]
15. Ohshima, K.; Yamaguchi, Y.; Hirota, R.; Hamamoto, T.; Tomimura, K.; Tan, Z.; Sano, T.; Azuhata, F.; Walsh, J.A.; Fletcher, J.; et al. Molecular Evolution of Turnip Mosaic Virus: Evidence of Host Adaptation, Genetic Recombination and Geographical Spread. *J. Gen. Virol.* **2002**, *83*, 1511–1521. [[CrossRef](#)]
16. Tomimura, K.; Gibbs, A.J.; Jenner, C.E.; Walsh, J.A.; Ohshima, K. The Phylogeny of Turnip Mosaic Virus; Comparisons of 38 Genomic Sequences Reveal a Eurasian Origin and a Recent “emergence” in East Asia. *Mol. Ecol.* **2003**, *12*, 2099–2111. [[CrossRef](#)]
17. Nguyen, H.D.; Tomitaka, Y.; Ho, S.Y.W.; Duchêne, S.; Vetten, H.J.; Lesemann, D.; Walsh, J.A.; Gibbs, A.J.; Ohshima, K. Turnip Mosaic Potyvirus Probably First Spread to Eurasian Brassica Crops from Wild Orchids about 1000 Years Ago. *PLoS ONE* **2013**, *8*, e55336. [[CrossRef](#)]
18. Nguyen, H.D.; Tran, H.T.N.; Ohshima, K. Genetic Variation of the Turnip Mosaic Virus Population of Vietnam: A Case Study of Founder, Regional and Local Influences. *Virus. Res.* **2013**, *171*, 138–149. [[CrossRef](#)]
19. Gibbs, A.J.; Nguyen, H.D.; Ohshima, K. The “emergence” of Turnip Mosaic Virus Was Probably a “Gene-for-Quasi-Gene” Event. *Curr. Opin. Virol.* **2015**, *10*, 20–26. [[CrossRef](#)]
20. Yasaka, R.; Ohba, K.; Schwinghamer, M.W.; Fletcher, J.; Ochoa-Corona, F.M.; Thomas, J.E.; Ho, S.Y.W.; Gibbs, A.J.; Ohshima, K. Phylodynamic Evidence of the Migration of Turnip Mosaic Potyvirus from Europe to Australia and New Zealand. *J. Gen. Virol.* **2015**, *96*, 701–713. [[CrossRef](#)]
21. Yasaka, R.; Fukagawa, H.; Ikematsu, M.; Soda, H.; Korkmaz, S.; Golnaraghi, A.; Katis, N.; Ho, S.Y.W.; Gibbs, A.J.; Ohshima, K. The Timescale of Emergence and Spread of Turnip Mosaic Potyvirus. *Sci. Rep.* **2017**, *7*, 4240. [[CrossRef](#)] [[PubMed](#)]

22. Kawakubo, S.; Gao, F.; Li, S.; Tan, Z.; Huang, Y.-K.; Adkar-Purushothama, C.R.; Gurikar, C.; Maneechoat, P.; Chiemsoombat, P.; Aye, S.S.; et al. Genomic Analysis of the Brassica Pathogen Turnip Mosaic Potyvirus Reveals Its Spread along the Former Trade Routes of the Silk Road. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2021221118. [[CrossRef](#)] [[PubMed](#)]
23. Kawakubo, S.; Tomitaka, Y.; Tomimura, K.; Koga, R.; Matsuoka, H.; Uematsu, S.; Yamashita, K.; Ho, S.Y.W.; Ohshima, K. The Recombinogenic History of Turnip Mosaic Potyvirus Reveals Its Introduction to Japan in the 19th Century. *Virus Evol.* **2022**, *8*, veac060. [[CrossRef](#)] [[PubMed](#)]
24. Larkin, M.A.; Blackshields, G.; Brown, N.P.; Chenna, R.; McGettigan, P.A.; McWilliam, H.; Valentin, F.; Wallace, I.M.; Wilm, A.; Lopez, R.; et al. Clustal W and Clustal X Version 2.0. *Bioinformatics* **2007**, *23*, 2947–2948. [[CrossRef](#)] [[PubMed](#)]
25. Martin, D.P.; Murrell, B.; Golden, M.; Khoosal, A.; Muhire, B. RDP4: Detection and Analysis of Recombination Patterns in Virus Genomes. *Virus Evol.* **2015**, *1*, vev003. [[CrossRef](#)] [[PubMed](#)]
26. Martin, D.; Rybicki, E. RDP: Detection of Recombination amongst Aligned Sequences. *Bioinformatics* **2000**, *16*, 562–563. [[CrossRef](#)] [[PubMed](#)]
27. Sawyer, S.A. *GENECONV: A Computer Package for the Statistical Detection of Gene Conversion*; Department of Mathematics, Washington University in Louis: Washington, WA, USA, 1999.
28. Salminen, M.O.; Carr, J.K.; Burke, D.S.; Mccutchan, F.E. Identification of Breakpoints in Intergenotypic Recombinants of HIV Type 1 by Bootscanning. *AIDS Res. Hum. Retroviruses*. **1995**, *11*, 1423–1425. [[CrossRef](#)] [[PubMed](#)]
29. Smith, J.M. Analyzing the Mosaic Structure of Genes. *J. Mol. Evol.* **1992**, *34*, 126–129. [[CrossRef](#)]
30. Posada, D.; Crandall, K.A. Evaluation of Methods for Detecting Recombination from DNA Sequences: Computer Simulations. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 13757–13762. [[CrossRef](#)]
31. Boni, M.F.; Posada, D.; Feldman, M.W. An Exact Nonparametric Method for Inferring Mosaic Structure in Sequence Triplets. *Genetics* **2007**, *176*, 1035–1047. [[CrossRef](#)]
32. Gibbs, M.J.; Armstrong, J.S.; Gibbs, A.J. Sister-Scanning: A Monte Carlo Procedure for Assessing Signals in Recombinant Sequences. *Bioinformatics* **2000**, *16*, 573–582. [[CrossRef](#)]
33. Tamura, K.; Stecher, G.; Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol. Biol. Evol.* **2021**, *38*, 3022–3027. [[CrossRef](#)]
34. Kimura, M. A Simple Method for Estimating Evolutionary Rates of Base Substitutions through Comparative Studies of Nucleotide Sequences. *J. Mol. Evol.* **1980**, *16*, 111–120. [[CrossRef](#)]
35. Page, R.D.M. Treeview: An Application to Display Phylogenetic Trees on Personal Computers. *Bioinformatics* **1996**, *12*, 357–358. [[CrossRef](#)]
36. Sharp, P.M.; Li, W.H. An Evolutionary Perspective on Synonymous Codon Usage in Unicellular Organisms. *J. Mol. Evol.* **1986**, *24*, 28–38. [[CrossRef](#)]
37. Wright, F. The “effective Number of Codons” Used in a Gene. *Gene* **1990**, *87*, 23–29. [[CrossRef](#)]
38. Gerton, J.L.; DeRisi, J.; Shroff, R.; Lichten, M.; Brown, P.O.; Petes, T.D. Global Mapping of Meiotic Recombination Hotspots and Coldspots in the Yeast *Saccharomyces Cerevisiae*. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 11383–11390. [[CrossRef](#)]
39. Steel, M. *The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing*, 2nd ed.; Lemey, P., Salemi, M., Vandamme, A.M., Eds.; Cambridge University Press: Cambridge, UK, 2010; Volume 66, pp. 324–325.
40. Fajinmi, A.A. Interactive Effect of Blackeye Cowpea Mosaic Virus and Cucumber Mosaic Virus on *Vigna Unguiculata*. *Hortic. Plant. J.* **2019**, *5*, 88–92. [[CrossRef](#)]
41. Yasaka, R.; Nguyen, H.D.; Ho, S.Y.W.; Duchêne, S.; Korkmaz, S.; Katis, N.; Takahashi, H.; Gibbs, A.J.; Ohshima, K. The Temporal Evolution and Global Spread of Cauliflower Mosaic Virus, a Plant Pararetrovirus. *PLoS ONE* **2014**, *9*, e85641. [[CrossRef](#)]
42. He, Z.; Dong, T.; Wang, T.; Chen, W.; Liu, X.; Li, L. Genetic Variation of the Novel Badnaviruses Infecting *Nelumbo Nucifera* Based on the RT/RNase H Coding Region Sequences. *Hortic. Plant. J.* **2020**, *6*, 335–342. [[CrossRef](#)]
43. Butt, A.M.; Nasrullah, I.; Qamar, R.; Tong, Y. Evolution of Codon Usage in Zika Virus Genomes Is Host and Vector Specific. *Emerg. Microbes. Infect.* **2016**, *5*, 107. [[CrossRef](#)]
44. Li, G.; Wang, H.; Wang, S.; Xing, G.; Zhang, C.; Zhang, W.; Liu, J.; Zhang, J.; Su, S.; Zhou, J. Insights into the Genetic and Host Adaptability of Emerging Porcine Circovirus 3. *Virulence* **2018**, *9*, 1301–1313. [[CrossRef](#)]
45. He, W.; Zhao, J.; Xing, G.; Li, G.; Wang, R.; Wang, Z.; Zhang, C.; Franzo, G.; Su, S.; Zhou, J. Genetic Analysis and Evolutionary Changes of Porcine Circovirus 2. *Mol. Phylogenet. Evol.* **2019**, *139*, 106520. [[CrossRef](#)]
46. Yan, Z.; Wang, R.; Zhang, L.; Shen, B.; Wang, N.; Xu, Q.; He, W.; He, W.; Li, G.; Su, S. Evolutionary Changes of the Novel Influenza D Virus Hemagglutinin-Esterase Fusion Gene Revealed by the Codon Usage Pattern. *Virulence* **2019**, *10*, 1–9. [[CrossRef](#)]
47. Zhang, W.; Zhang, L.; He, W.; Zhang, X.; Wen, B.; Wang, C.; Xu, Q.; Li, G.; Zhou, J.; Veit, M.; et al. Genetic Evolution and Molecular Selection of the HE Gene of Influenza C Virus. *Viruses* **2019**, *11*, 167. [[CrossRef](#)]
48. Xu, X.Z.; Liu, Q.P.; Fan, L.J.; Cui, X.F.; Zhou, X.P. Analysis of Synonymous Codon Usage and Evolution of Begomoviruses. *Zhejiang Univ. Sci. B.* **2008**, *9*, 667–674. [[CrossRef](#)]
49. 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](#)]
50. Biswas, K.K.; Palchoudhury, S.; Chakraborty, P.; Bhattacharyya, U.K.; Ghosh, D.K.; Debnath, P.; Ramadugu, C.; Keremane, M.L.; Khetarpal, R.K.; Lee, R.F. Codon Usage Bias Analysis of Citrus Tristeza Virus: Higher Codon Adaptation to Citrus Reticulata Host. *Viruses* **2019**, *11*, 331. [[CrossRef](#)]

51. He, M.; Guan, S.Y.; He, C.Q. Evolution of Rice Stripe Virus. *Mol. Phylogenet. Evol.* **2017**, *109*, 343–350. [[CrossRef](#)]
52. Chakraborty, P.; Das, S.; Saha, B.; Sarkar, P.; Karmakar, A.; Saha, A.; Saha, D.; Saha, A. Phylogeny and Synonymous Codon Usage Pattern of Papaya Ringspot Virus Coat Protein Gene in the Sub-Himalayan Region of North-East India. *Can. J. Microbiol.* **2015**, *61*, 555–564. [[CrossRef](#)]
53. He, Z.; Gan, H.; Liang, X. Analysis of Synonymous Codon Usage Bias in Potato Virus M and Its Adaption to Hosts. *Viruses* **2019**, *11*, 752. [[CrossRef](#)] [[PubMed](#)]
54. He, Z.; Dong, Z.; Gan, H. Genetic Changes and Host Adaptability in Sugarcane Mosaic Virus Based on Complete Genome Sequences. *Mol. Phylogenet. Evol.* **2020**, *149*, 106848. [[CrossRef](#)] [[PubMed](#)]
55. He, Z.; Dong, Z.; Qin, L.; Gan, H. Phylodynamics and Codon Usage Pattern Analysis of Broad Bean Wilt Virus 2. *Viruses* **2021**, *13*, 198. [[CrossRef](#)] [[PubMed](#)]
56. He, Z.; Dong, Z.; Gan, H. Comprehensive Codon Usage Analysis of Rice Black-Streaked Dwarf Virus Based on P8 and P10 Protein Coding Sequences. *Infect. Genet. Evol.* **2020**, *86*, 104601. [[CrossRef](#)]
57. He, Z.; Ding, S.; Guo, J.; Qin, L.; Xu, X. Synonymous Codon Usage Analysis of Three Narcissus Potyviruses. *Viruses* **2022**, *14*, 846. [[CrossRef](#)]
58. Puigbò, P.; Aragonès, L.; Garcia-Vallvé, S. RCDI/ERCDI: A Web-Server to Estimate Codon Usage Deoptimization. *BMC. Res. Notes* **2010**, *3*, 87. [[CrossRef](#)]