Iranian Journal of Virology 2024;18(2):8-18.

Original Article

Decoding of Codon Usage Patterns of Two Betaflexiviruses Cherry Virus A and Prunus Virus T

Reza Pourrrahim[1]*, Shirin Farzadfar[1], Mohammadreza Rezapanah[2,3], Tayyebeh Keshavarz [1]

1. Plant Virus Research Department, Iranian Research Institute of Plant Protection (IRIPP), Agricultural Research, Education and Extension Organization (AREEO), Tehran, Iran.
2. Insect Molecular Virology laboratory, BioControl Department, Iranian Research Institute of Plant Protection (IRIPP), Agricultural Research, Education and Extension Organization (AREEO), Tehran, Iran.
3. Iranian Network for Research in Viral Diseases, Tehran, Iran.

Abstract

Background and Aims: Among the more than fifty viruses that infect stone fruit trees and belong to the *Betaflexiviridae* family, cherry virus A (*Capillovirus*) and prunus virus T (*Tepovirus*) are common members. In this study, we provide codon usage analyses and composition of CVA and PrVT based on their full genome nucleotide sequences. Codon usage patterns studies provide great information about host-pathogen co-evolution, phylogenetic relations among specific species, and molecular evolution of genes.

Material and methods: Genome sequences of CVA and PrVT isolates available at GenBank including two Iranian isolates from sweet cherry were used for nucleotide composition analysis using MEGAX soft-ware. The nucleotide mixtures at the 3rd codon position (A3s, C3s, T3s, G3s%, and GC3s), were determined using CodonW 1.4.2 package (http://codonw.sourceforge.net). The RSCU values for the ORFs coding sequences were determined to explore aspects of synonymous codon usage without the complex influence of amino acid composition and coding sequence size of samples. The ENC value was determined using Codon W v1.4.2. ENC analysis was used to quantify the absolute codon usage bias by evaluating the degree of codon usage bias exhibited by the coding sequences, regardless of gene length and the number of amino acids.

Results: Our analysis indicates synonymous codon usage in CVA and PrVT possibly mirroring a dynamic process of mutations from the viruses along with selection pressure to adjust their codon usage to various hosts, and conditions. Furthermore, relative synonymous codon usage-RSCU, nucleotide composition, and an effective number of codon-ENC-plots analysis results suggest that the selection pressure was greater than mutation pressure.

Conclusions: This bioinformatics analysis expands our knowledge of the differences and common features in codon usage patterns of betaflexiviruses with their various hosts.

Keywords: *Betaflexiviridae*, CVA, PrVT, Codon Usage Patterns

Introduction

Over fifty viruses and virus-like agents infecting stone fruit trees have been determined and this list is continually being expanding using high throughput sequencing (HTS) (1-4). The *Betaflexiviridae* family (the order *Tymovirales*) is one of the most diverse positive-strand RNA virus ensembles. Currently, this family is composed of 15 genera divided between two subfamilies; members of the subfamily *Quinvirinae* have a triple gene block (TGB) module that facilitates cell-to-cell movement whereas viruses in the subfamily *Trivirinae* encode a 30K-like movement protein. The plant host range of the Betaflexiviruses members is very large and most of them infect fruit trees such as apples, stone fruits, and grapevines (5-9).

Cherry virus A (CVA) (*Capillovirus* genus), is one of the most widespread betaflexiviruses

*Corresponding author:
Reza Pourrrahim, Ph.D,
Plant Virus Research Department, Iranian Research Institute of Plant Protection (IRIPP), Agricultural Research, Education and Extension Organization (AREEO), Tehran, Iran.
Email: pourrahim@yahoo.com

affecting sweet cherry (*P. avium*) and sour cherry (*P. ceracus*) worldwide (10). This virus has also been reported to a lesser extent from non-cherry hosts including, apricot (*P. armeniaca*), peach (*P. persica*) (11), cherry plum (*P. cerasifera*), flowering cherry (*P. serrulata*), Japanese apricot (*P. mume*), and Japanese (Chinese) plum (*P. salicina*) (1, 4, 12, 13). Nevertheless of the high prevalence of CVA infection, no natural vector has been reported, but it can be readily transmitted from plant to plant through vegetative propagation and grafting (1, 4, 12, 13).

While no clear specific symptoms have been associated with CVA, studies have shown that mixed infections with other viruses can lead to more severe symptomss and affect rootstock/ scion compatibility (14, 15). The genome of CVA is comprised of a positive single-stranded RNA with 7,383 nucleotides (nt), except the 3'-terminal poly (A) tail. The genomic structure of CVA consisting of two overlapping open reading frames (ORFs), is similar to *Apple stem grooving virus* (ASGV), the type species of the *Capillovirus* genus. A large polyprotein (266 kDa) including the RNA-dependent RNA polymerase (RdRp) fused in-frame to the coat protein (CP) and a 52 kDa putative movement protein (MP) are encoded by ORF1 and ORF2, respectively (9). A high genomic diversity has been reported in the CVA population, and the nucleotide identity of the entire genome varies between 79 to 99% (15). CVA isolates have been classified into five molecular groups based on RdRp, however, six different molecular groups have been reported according to the full genome sequences (15). At present CVA is grouped into seven phylogenetic clusters (16).

In 2015, Marais et al. (7) proposed prunus virus T (PrVT) species in the *Tepovirus* genus. However, the biological information concerning PrVT is still scarce, but molecular screening of prunus samples allowed the identification of three hosts (*P. domestica*, *P. avium*, and *P. cerasifera*) in Italy and Azerbaijan, with an overall prevalence of 1%, and from (*Pyrus communis*) in Kyrgyzstan (14). Since all reported host plants of PrVT were simultaneously infected by *apricot pseudo-chlorotic leaf spot virus*-APCLSV (*Trichovirus*), apple chlorotic leaf spot virus-ACLSV (*Trichovirus*), CVA (*Capillovirus*), prune dwarf virus-PDV (*Ilarvirus*), or little cherry virus-LChV1 (*Closteroviridae*) as well-known prunus-infecting viruses, therefore, symptoms clearly associated with PrVT infection have not yet been described and it is difficult to associate specific symptoms with PrVT infection (7).

Using high-throughput sequencing, and 5'/3' RACE, the full genome sequences of two betaflexiviruses, *Cheery virus A*-CVA (*Capillovirus* genus) and *Prunus virus T*-PrVT (*Tepovirus* genus) isolates were determined in Iran. RT-PCR tests indicates a high prevalence of infections with these viruses in the surveyed cherry orchards and nurseries (17). Furthermore, in this study, the diversity and population structure of CVA and PrVT, and the evolutionary forces that shape these populations have been considered. In this respect, we provide codon usage analyses and composition of CVA and PrVT based on their full genome nucleotide sequences. Codon usage patterns studies provide valuable information about the host-pathogen co-evolution, phylogenetic relationships, and molecular evolution of genes (18-21).

Methods and Materials

Virus Sources

In this study, the nucleotide sequences of CVA (n=114) and PrVT (n=5) isolates available at the GenBank database including two previously sequenced Iranian isolates CVA-Qz5 (Acc. No. PP736769) and PrVT-Qz5 (Acc. No. PP736770) were used for nucleotide analyses. Additionally, three aphid species were tested for their ability to transmit CVA and PrVT. CVA-Qz5 and PrVT-Qz5 isolates were maintained on young sweet cherry seedlings in greenhouse conditions and used as virus sources in aphid transmission tests. Clonal populations of the aphids *Aphis faba*, *Myzus persicae*, and *M. cerasi*, originally from Tehran, Iran, were kept on turnip plants. These plants were housed in separate growth chambers at teperatures ranging from 15 to 20 degrees Celsius with 16 hours of light and 8 hours of darkness cycles. Aphids were starved for 3 hours before being

allowed to acquire the viruses during periods of 15 to 30 min. Aphids were then manually transferred from source plants to young sweet cherry seedlings using a paintbrush. Approximately 15 aphids were placed on each test plant to inoculate the plants. All aphids were killed with Actara insecticide (Thiamethoxam 24% SC, Syngenta), and the plants were monitored for 3 months to determine transmission rates using the RT-PCR diagnostic method as described previously (17).

Nucleotide Composition Analysis

Nucleotide composition analysis was conducted using MEGAX software (22). The overall nucleotides occurrence frequency (A%, C%, T/U%, and G%), the overall occurrence of nucleotide frequency at the third position of codons (A3%, C3%, U3%, and G3%), and the overall occurrence of nucleotides frequencies of G+C at different codon positions were determined. Codons AUG and UGG were excluded from the analysis due to their lack of codon usage bias. Termination codons (UAG, UGA, UAA) were also excluded from the analysis, as they do not encode any amino acid.

Codon Usage Bias and Relative Synonymous Codon Usage

The nucleotide mixtures at the 3rd codon position (A3s, C3s, T3s, G3s%, and GC3s), were determined using CodonW 1.4.2 package (http://codonw.sourceforge.net). The component parameters of the coding sequences were calculated after the deletion of five non-bias codons (AUG, UGG, UAA, UGA, and UAG). In general, codon adaptation is estimated using metrics derived from the relative synonymous codon usage (RSCU), the ratios between the observed usage frequency of a specific codon in a gene and its expected one in the correspondent synonymous family (23). The RSCU values for the ORFs coding sequences were determined to explore aspects of synonymous codon usage without the complex influence of amino acid composition and coding sequence size of samples. The RSCU values >1.6 and <0.6, indicate synonymous codon usage is overrepresented or underrepresented, respectively. Furthermore, those groups of codons with RSCU=1, RSCU > 1, and RSCU < 1 are considered as having no bias, being more-abundant and less-abundant, respectively (19, 24).

Effective Number of Codons (ENC), and ENC-Plot

The ENC value was determined using Codon W v1.4.2. ENC analysis was used to quantify the absolute codon usage bias by evaluating the degree of codon usage bias exhibited by the coding sequences, regardless of gene length and the number of amino acids. ENC values range from 20 (a high codon usage bias using only one of the possible synonymous codons for the matching amino acid) to 61 (no bias using all available synonymous codons equally for the matching amino acid) (25). In addition, the ENC-plot (ENC values versus GC3s values), which refers to the effect of mutational pressure on codon usage bias, was analyzed for each ORF of CVA and PrVT. Selection pressure is the main factor driving codon usage bias, if points are below the standard curve or else the influence of other factors such as mutation can be considered.

Results

Nucleotide Composition

The compositional constraints' effect on codon usage of CVA and PrVT was analyzed using nucleotide composition. For the CVA polyprotein, according to different hosts, the nucleotide composition bias was in order A > U > G > C, with an average of 31.0 to 31.5%, 29.1 to 29.8%, 20.5 to 20.8%, and 18.3 to 18.8%, respectively. However synonymous codons at the third position followed the bias T3s > A3s > G3s > C3s indicating an AU-biased composition (Figure 1a). According to different hosts the nucleotide composition bias for the PrVT CP gene was in order G > A > U > C, with an average of 30.8%, 26.9, 23.5 and 18.8, respectively. Synonymous codons at the third position followed the bias G3s > T3s > A3s > C3s (Figure 1b). The overall GC content was higher than AU, suggesting a GC bias in the CP gene. For the MP gene of PrVT, the nucleotide composition bias was G > A > U > C

similar to the CP gene, however, the synonymous codons at the third position followed the bias T3s > G3s > A3s > C3s, with overall AU bias (Figure 1c).

Patterns of Codon Usage in Polyprotein of CVA and its Hosts

In order to realize the influence of selection pressure on codon usage patterns, the preferred codons with higher frequency were determined. Among the CVA coding sequence, 16 out of 18 preferred codons were A/U-ending (A-ending: 9; U-ending: 7), while exhibiting an equal distribution of G/C-ending (C-ending: 1; G-ending: 1) (Table 1), therefore A- and U-ending codons are preferred in the CVA coding sequence.

The extreme over-presentation (RSCU > 1.6) was found for three preferred codons with the highest being AGA (2.80), whereas the remaining preferred codons had 1.6 < RSCU values > 1.02. Twelve optional synonymous codons were underrepresented (RSCU < 0.6) (Table 1).

To understand the influence of selection pressure on codon usage patterns, the preferred codons with higher frequency were compared between CVA, and all the hosts, as shown in Table 2 and Figure 2. Twelve amino acids Leu, Iie, Val, Ser, Pro, His, Gln, Asn, Lys, Asp, Glu, and Arg showed identical usage of preferred codons (UUG, AUU, GUU, UCA, CCA, CAU, CAA, AAU, AAA, GAU, GAA, and

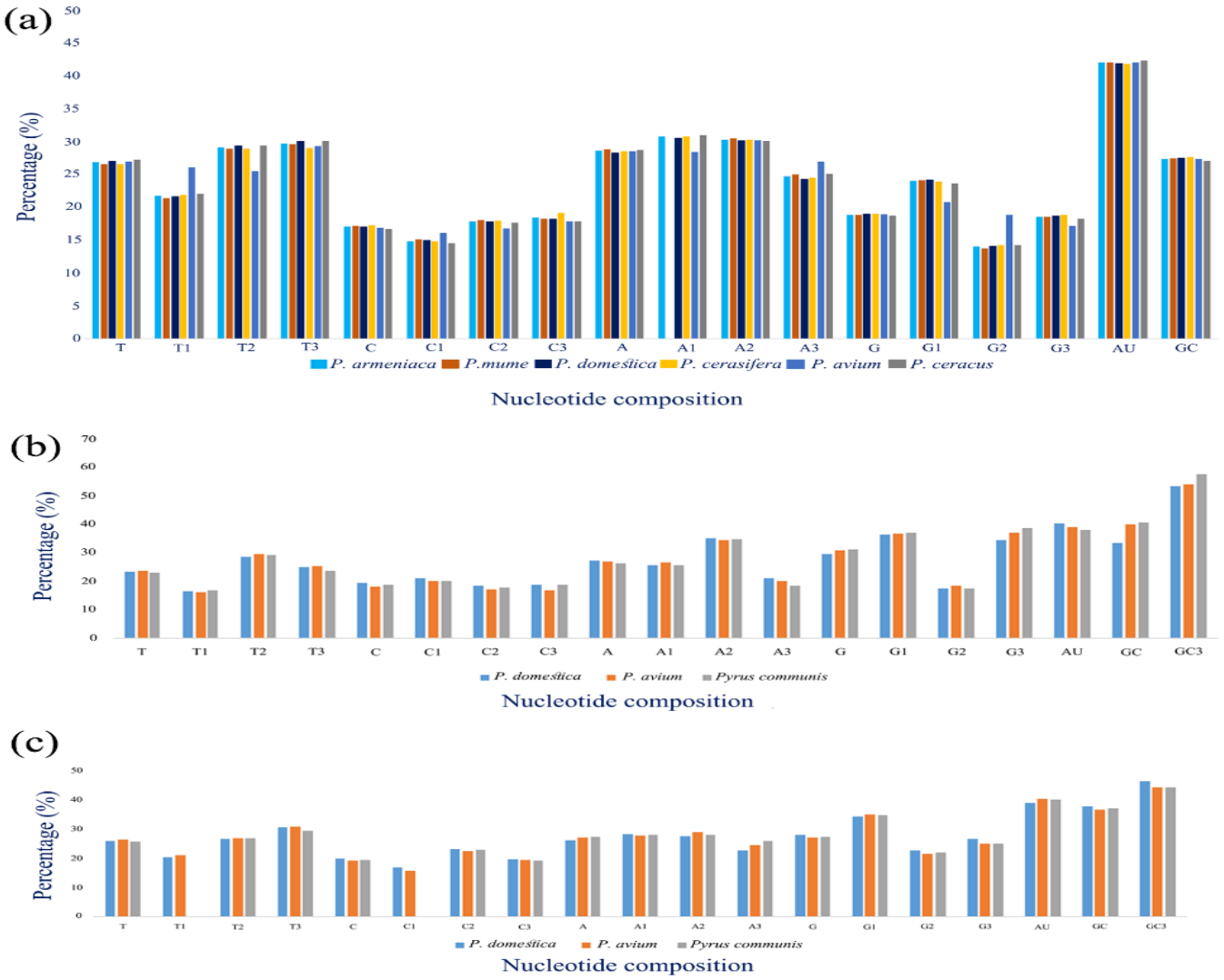

Fig. 1. Comparative analysis of nucleotide composition patterns between CVA (a), PrVT-CP (b), PrVT-MP (c) and its hosts.

Table 1. The RSCU value of 59 codons encoding 18 amino acids according to CVA, and PrVT coding sequences

Codon	aa	CVA	PrVT		
		Full genome	Replicase	MP	CP
UUU	Phe	**1.20***	**1.02**	**1.06**	0.82
UUC		0.80	0.98	0.94	**1.18**
UUA	Leu	0.96	0.45	0.4	0.50
UUG		**1.41**	**1.50**	**2.17**	1.00
CUU		1.36	1.1	0.91	0.89
CUC		0.91	0.99	0.76	0.89
CUA		0.56	0.61	0.40	0.39
CUG		0.81	1.35	1.36	**2.33**
AUU	Iie	**1.53**	**1.18**	**1.31**	0.50
AUC		1.01	1.07	0.93	0.83
AUA		0.46	0.75	0.77	**1.67**
GUU	Val	**1.53**	**1.39**	**1.70**	1.18
GUC		0.89	0.86	0.91	0.41
GUA		0.46	0.39	0.26	0.18
GUG		1.12	1.37	1.13	**2.23**
UCU	Ser	1.40	1.20	**1.91**	0.70
UCC		0.78	0.85	1.00	1.13
UCA		**1.73**	**1.51**	1.13	**1.48**
UCG		0.17	0.47	0.59	0.09
AGU		1.09	1.11	0.67	1.13
AGC		0.83	0.86	0.70	1.47
CCU	Pro	1.26	1.11	0.94	1.16
CCC		0.67	0.63	0.94	0,00
CCA		**1.81**	**1.68**	**1.72**	1.29
CCG		0.25	0.58	0.39	**1.55**
ACU	Thr	1.35	1.05	**1.26**	0.86
ACC		0.80	1.23	0.93	0.38
ACA		**1.56**	**1.24**	1.21	**2.10**
ACG		0.29	0.48	0.60	0.67
GCU	Ala	1.45	**1.49**	**2.07**	**1.72**
GCC		0.87	0.96	0.59	0.74
GCA		**1.47**	1.26	0.93	0.93
GCG		0.21	0.29	0.41	0.6
UAU	Tyr	**1.11**	0.96	**1.25**	**1.33**
UAC		0.89	**1.04**	0.75	0.67
CAU	His	**1.43**	**1.09**	**1.38**	**1.25**
CAC		0.57	0.91	0.62	0.75
CAA	Gln	**1.30**	0.79	0.88	0.74
CAG		0.70	**1.21**	**1.12**	**1.26**
AAU	Asn	**1.40**	**1.13**	**1.11**	**1.22**
AAC		0.60	0.87	0.89	0.78
AAA	Lys	**1.13**	**1.12**	**1.03**	0.73
AAG		0.87	0.88	0.97	**1.27**
GAU	Asp	**1.35**	**1.22**	**1.23**	0.96
GAC		0.65	0.78	0.77	**1.04**
GAA	Glu	**1.25**	**1.03**	**1.19**	0.77
GAG		0.75	0.97	0.81	**1.23**
UGU	Cys	0.98	0.96	**1.08**	**1.33**
UGC		**1.02**	**1.04**	0.92	0.67
CGU	Arg	0.22	0.20	0.50	1.75
CGC		0.30	0.11	0.29	0.33
CGA		0.51	0.28	1.00	0.17
CGG		0.27	0.28	0.79	0.50
AGA		**2.80**	**2.62**	1.54	0.83
AGG		1.90	2.51	**1.89**	**2.42**
GGU	Gly	1.09	1.08	0.84	1.13
GGC		0.73	0.69	0.41	0.67
GGA		**1.46**	**1.12**	**1.68**	0.56
GGG		0.73	1.11	1.07	**1.64**

Table 2. Preferred codons for each amino acid in the Polyprotein of CVA and its hosts

Codon	aa	Hosts						CVA
		P. armeniaca	*P.mume*	*P. domestica*	*P. cerasifera*	*P. avium*	*P. ceracus*	**Poly protein**
UUU	Phe	**1.07**	**1.09**	**1.10**	0.94	**1.2**	**1.24**	**1.20**
UUC		0.93	0.91	0.90	**1.06**	0.8	0.76	0.80
UUA	Leu	1.01	0.98	1.00	0.95	0.97	1.09	0.96
UUG		**1.61**	**1.51**	**1.59**	**1.78**	**1.41**	**1.72**	**1.41**
CUU		1.06	1.08	0.98	0.90	1.38	1.09	1.36
CUC		0.82	0.79	0.78	0.98	0.89	0.71	0.91
CUA		0.60	0.65	0.56	0.67	0.54	0.52	0.56
CUG		0.86	0.98	1.08	0.72	0.8	0.87	0.81
AUU	Iie	**1.39**	**1.35**	**1.44**	**1.44**	**1.53**	**1.36**	**1.53**
AUC		0.89	0.91	0.82	0.88	1	0.91	1.01
AUA		0.72	0.74	0.75	0.68	0.47	0.74	0.46
GUU	Val	**1.50**	**1.59**	**1.53**	**1.40**	**1.53**	**1.29**	**1.53**
GUC		0.94	0.92	0.94	1.03	0.88	1	0.89
GUA		0.52	0.41	0.33	0.49	0.46	0.39	0.46
GUG		1.05	1.08	1.20	1.08	1.13	1.32	1.12
UCU	Ser	1.09	1.05	0.96	1.13	1.4	1.28	1.4
UCC		0.76	0.71	0.60	0.68	0.79	0.72	0.78
UCA		**2.02**	**2.15**	**2.25**	**2.00**	**1.73**	**1.81**	**1.73**
UCG		0.22	0.23	0.22	0.21	0.17	0.19	0.17
AGU		1.16	1.16	1.21	0.96	1.09	1.15	1.09
AGC		0.76	0.71	0.77	1.02	0.82	0.85	0.83
CCU	Pro	1.07	1.03	1.38	1.11	1.27	1.20	1.26
CCC		0.44	0.43	0.39	0.38	0.66	0.4	0.67
CCA		**2.30**	**2.41**	**2.06**	**2.28**	**1.8**	**2.22**	**1.81**
CCG		0.19	0.13	0.17	0.23	0.26	0.18	0.25
ACU	Thr	**1.53**	**1.36**	**1.65**	**1.73**	1.34	1.43	1.35
ACC		0.84	0.99	0.78	0.69	0.81	0.83	0.80
ACA		1.32	1.32	1.24	1.22	**1.56**	**1.57**	**1.56**
ACG		0.31	0.33	0.33	0.37	0.28	0.17	0.29
GCU	Ala	**1.69**	**1.66**	**1.52**	**1.74**	1.45	**1.65**	1.45
GCC		0.84	0.88	1.09	0.87	0.85	0.66	0.87
GCA		1.28	1.28	1.09	1.16	**1.50**	1.49	**1.47**
GCG		0.20	0.19	0.30	0.23	0.2	0.20	0.21
UAU	Tyr	**1.09**	**1.25**	**1.05**	0.99	**1.11**	**1.11**	**1.11**
UAC		0.91	0.75	0.95	**1.01**	0.89	0.89	0.89
CAU	His	**1.42**	**1.41**	**1.47**	**1.44**	**1.43**	**1.41**	**1.43**
CAC		0.58	0.59	0.53	0.56	0.57	0.59	0.57
CAA	Gln	**1.20**	**1.23**	**1.15**	**1.19**	**1.29**	**1.32**	**1.30**
CAG		0.80	0.78	0.85	0.81	0.71	0.68	0.70
AAU	Asn	**1.38**	**1.35**	**1.31**	**1.43**	**1.39**	**1.41**	**1.40**
AAC		0.62	0.65	0.69	0.57	0.61	0.59	0.60
AAA	Lys	**1.06**	**1.07**	**1.09**	**1.08**	**1.14**	**1.22**	**1.13**
AAG		0.94	0.93	0.91	0.92	0.86	0.78	0.87
GAU	Asp	**1.33**	**1.28**	**1.41**	**1.31**	**1.35**	**1.28**	**1.35**
GAC		0.67	0.72	0.59	0.69	0.65	0.72	0.65
GAA	Glu	**1.10**	**1.06**	**1.11**	**1.11**	**1.25**	**1.11**	**1.25**
GAG		0.90	0.94	0.89	0.89	0.75	0.89	0.75
UGU	Cys	**1.14**	**1.13**	**1.14**	**1.06**	0.98	**1.02**	0.98
UGC		0.86	0.87	0.86	0.94	**1.02**	0.98	**1.02**
CGU	Arg	0.17	0.17	0.25	0.18	0.21	0.18	0.22
CGC		0.24	0.29	0.19	0.18	0.31	0.30	0.3
CGA		0.4	0.52	0.25	0.30	0.51	0.42	0.51
CGG		0.17	0.06	0.19	0.18	0.27	0.18	0.27
AGA		**2.99**	**2.97**	**3.31**	**2.68**	**2.83**	**2.79**	**2.80**
AGG		2.03	1.98	1.81	2.47	1.87	2.14	1.90
GGU	Gly	**1.37**	**1.40**	**1.32**	**1.44**	1.07	**1.37**	1.09
GGC		0.79	0.72	0.83	0.69	0.74	0.85	0.73
GGA		1.05	1.09	1.12	1.19	**1.45**	1.16	**1.46**
GGG		0.79	0.79	0.73	0.69	0.74	0.62	0.73

* The most frequently used codons are shown in bold.

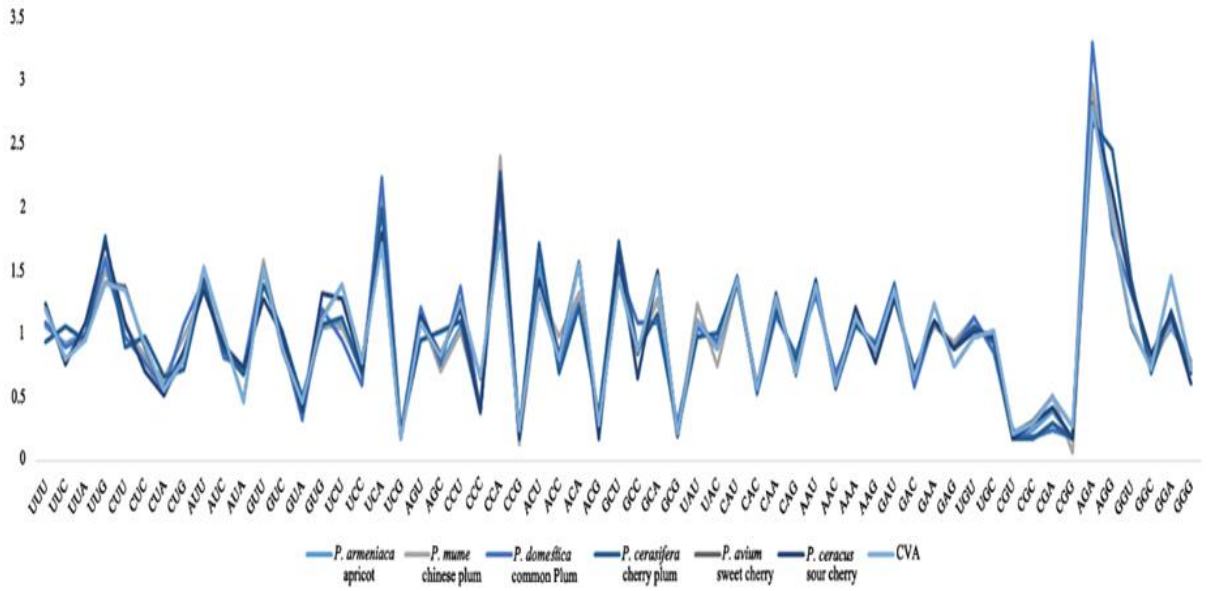

Fig. 2. Comparative analysis of relative synonymous codon usage (RSCU) patterns between CVA and its hosts

AGA) among hosts and CVA, indicating mutual codon preference. However, differences in the choice of preferred codons were found among the remaining six amino acids (Table 2, Figure 2). Similarities and differences in preferred codons were also found among the CVA and various hosts. For example, CVA and *P. avium* for encoding the Ala, Cys, and Gly amino acids are common in three preferred codons CGA, UGC, and GGA, respectively. However, other hosts used different preferred codons GCU, UGC, and GGU for encoding these three amino acids (Ala, Cys, and Gly), respectively. This indicates that the synonymous codon usage of CVA is a mixture of the two types of codon usage "coincidence and antagonism".

Patterns of Codon Usage in MP and CP of PrVT and its Hosts

The RSCU results of PrVT MP, and CP coding sequences are shown in Table 1. Among the Pr VT-CP gene, 10 out of 18 preferred codons were G/C-ending (G-ending: 8; C-ending:2), and the remaining 8 were A/U-ending (A-ending: 3; U-ending: 5) (Table 1). As indicated, G- and U-ending codons were preferred in the PrVT-CP coding sequence. Within these preferred codons, 7 had an RSCU value > 1.6, i.e. (CUG, AUA, GUG, ACA, GCU, AGG, GGG) indicating overrepresented codons, with the highest being AGG (2.42), while the remaining 11 had RSCU values >1.04 and <1.6, indicating abundant codons. For the PrVT CP gene out of these 18 preferred codons, 31 codons had RSCU values <1, i.e., less-abundant codons, while 9 had RSCU ≥1, i.e., abundant codons. One optional synonymous codon CCC (Pro amino acid) was underrepresented (RSCU = 0.0) (Table 1). Among PrVT-MP gene, 15 out of 18 preferred codons were A/U-ending (A-ending: 4; U-ending: 11), and 3 were G–ending (G-ending: 3). Extreme over-representation (RSCU > 1.6) was found for seven preferred codons, whereas the remaining had RSCU values >1.03 and < 1.6. No optional synonymous codon was under-represented (Table 2). In addition, preferred codons encoding amino acids with higher frequency for both PrVT MP and CP genes, and its hosts have been listed. For PrVT-MP 10 amino acid (Leu, Lie, Val, Ser, Pro, Ala, Tyr, His, Asp, and Gly), showed similar preferred codons (UUG, AUU, GUU, UCU, CAA, GCU, UAU, CAU, GAU, and GGA) among three hosts, indicating evidence of mutual codon preference, however, they used different preferred codons for the other eight amino acids (Phe, Thr, Gln, Asn, Lys, Glu, Cys, and Arg) (Table 3, Figure 3). A high number of similar preferred codons were detected for

Table 3. Preferred codons for each amino acid in the MP and CP genes of PrVT and its hosts

Codon	aa	CP			MP		
		P. avium	*P. domestica*	*P. communis*	*P. avium*	*P. domesticus*	*P. communis*
UUU	Phe (F)	0.88	0.75	0.8	**1.13**	0.95	**1.06**
UUC		**1.13**	**1.25**	**1.20**	0.87	**1.05**	0.94
UUA	Leu (L)	0.44	0.29	0.65	0.26	0.5	0.5
UUG		0.88	1.14	1.04	**2.04**	**1.75**	**2.50**
CUU		0.88	1.14	0.78	0.89	1.25	0.75
CUC		0.73	0.86	1.04	1.28	0.25	0.50
CUA		0.29	0.57	0.39	0.38	0.00	0.63
CUG		**2.78**	**2.00**	**2.09**	1.15	2.25	1.13
AUU	Iie (I)	0.86	0.60	0.00	**1.38**	**1.58**	**1.11**
AUC		0.86	0.60	1.00	1.14	0.63	0.87
AUA		**1.29**	**1.80**	**2.00**	0.49	0.79	1.03
GUU	Val (V)	1.37	1.00	1.06	**1.67**	**1.63**	**1.76**
GUC		0.32	0.50	0.47	0.78	0.88	1.06
GUA		0.11	0.50	0.12	0.24	0.38	0.24
GUG		**2.21**	**2.00**	**2.35**	1.31	1.13	0.94
UCU	Ser (S)	0.41	1.38	0.67	**1.89**	**2.00**	**1.89**
UCC		1.03	0.46	1.56	1.08	0.93	0.94
UCA		**1.66**	**1.38**	**1.33**	1.15	1.07	1.15
UCG		0.00	0.46	0.00	0.74	0.4	0.54
AGU		1.24	0.92	1.11	0.67	0.67	0.67
AGC		1.65	1.37	1.32	0.47	0.93	0.81
CCU	Pro	1.00	1.14	1.33	1.48	0.75	0.55
CCC		0.00	0.00	0.00	0.89	0.75	1.10
CCA		**1.67**	0.57	1.33	**1.48**	**1.50**	**2.07**
CCG		1.33	**2.29**	**1.34**	0.15	1.00	0.28
ACU	Thr	0.94	0.57	0.89	1.09	**1.56**	**1.26**
ACC		0.71	0.57	0.00	0.97	0.89	0.91
ACA		**1.88**	**2.29**	**2.22**	**1.45**	1.33	0.91
ACG		0.47	0.57	0.89	0.48	0.22	0.91
GCU	Ala	**1.58**	**1.89**	**1.76**	**2.00**	**2.00**	**2.17**
GCC		0.61	0.84	0.82	0.70	0.73	0.43
GCA		1.33	1.05	0.47	0.90	0.73	1.04
GCG		0.48	0.21	0.94	0.40	0.55	0.35
UAU	Tyr	**1.33**	**1.33**	**1.33**	**1.20**	**1.20**	**1.33**
UAC		0.67	0.67	0.67	0.80	0.80	0.67
CAU	His	**1.71**	0.67	**1.01**	**1.60**	**1.33**	**1.20**
CAC		0.29	**1.33**	1.00	0.40	0.67	0.80
CAA	Gln	0.77	0.62	0.77	0.74	0.80	**1.03**
CAG		**1.23**	**1.38**	**1.23**	**1.26**	**1.20**	0.97
AAU	Asn	**1.23**	**1.33**	**1.14**	**1.20**	0.55	**1.28**
AAC		0.77	0.67	0.86	0.80	**1.45**	0.72
AAA	Lys	0.83	0.67	0.65	**1.02**	**1.08**	0.90
AAG		**1.17**	**1.33**	**1.35**	0.98	0.92	**1.00**
GAU	Asp	**1.05**	0.91	0.91	**1.23**	**1.24**	**1.22**
GAC		0.95	**1.09**	**1.09**	0.77	0.76	0.78
GAA	Glu	0.67	**1.10**	0.71	**1.25**	0.87	**1.28**
GAG		**1.33**	0.90	**1.29**	0.75	**1.13**	0.72
UGU	Cys	**1.33**	**2.00**	1.00	**1.29**	**1.11**	0.86
UGC		0.67	0.00	0.90	0.71	0.89	**1.14**
CGU	Arg	1.80	1.60	1.78	0.45	0.60	0.51
CGC		0.20	0.40	0.44	0.36	0.40	0.17
CGA		0.20	0.40	0.00	1.07	0.80	1.01
CGG		0.40	0.80	0.44	0.63	0.80	0.93
AGA		1.20	0.80	0.44	**1.79**	1.60	1.27
AGG		**2.20**	**2.00**	**2.89**	1.70	**1.80**	**2.11**
GGU	Gly	1.07	0.80	1.33	0.86	1.06	0.69
GGC		0.67	1.07	0.48	0.43	0.35	0.41
GGA		0.40	0.53	0.73	**1.66**	**1.41**	**1.86**
GGG		**1.87**	**1.60**	**1.45**	1.05	1.18	1.03

***** The most frequently used codons are shown in bold.

fifteen amino acids among the PrVT-CP gene and its hosts, except for Pro, and His (Table 3, Figure 4). This analysis also indicates the phenomena of "coincidence and antagonism" for the PrVT genome.

The Effective Number of Codons (ENC), and ENC-Plot

Low codon usage bias in coding sequences was found for CVA-polyprotein, and both CP and MP genes of PrVT, with average ENC values of (49.71 ± 0.414), (58.187± 0.829), and (56.102 ± 0.414), respectively. This represents an approximately constant and conserved genomic composition. ENC-plot analysis was done to find the main factor influencing the codon usage bias for CVA-polyprotein (Figure 5a), PrVT-CP (Fig. 5b), and PrVT-MP (Figure 5c). The dots from different phylogroups CVA (Figure 5a), PrVT-CP (Fig. 5b), and PrVT-MP (Figure 5c) are below the standard curve, which indicates that the selection pressure is the main factor driving codon usage bias. SiD values (ranging from 0 to 1) were calculated to measure the effect of the host codon usage bias on the CVA-polyprotein and PrVT-CP gene.

Discussion

The CVA polyprotein revealed an over-representation of A, followed by U, G, and C with an overall A/U codon bias pattern in the nucleotide composition. For CVA polyprotein, according to different hosts, the nucleotide composition bias was as follows: A > U > G > C, however synonymous codons at the third position followed the bias T3s > A3s > G3s > C3s, indicating an AU-biased composition. According to different hosts the nucleotide composition bias for PrVT suggested AU and GC biases in MP and CP genes, respectively. This indicates the role of mutation in codon usage bias for both CVA and PrVT. The adaptation of the common ancestor of new strains in terms of proximity to the nucleotide composition of the host(s) is a necessary evolutionary process, and thus this nucleotide composition bias in CVA and PrVT genomes can be interpreted (26).

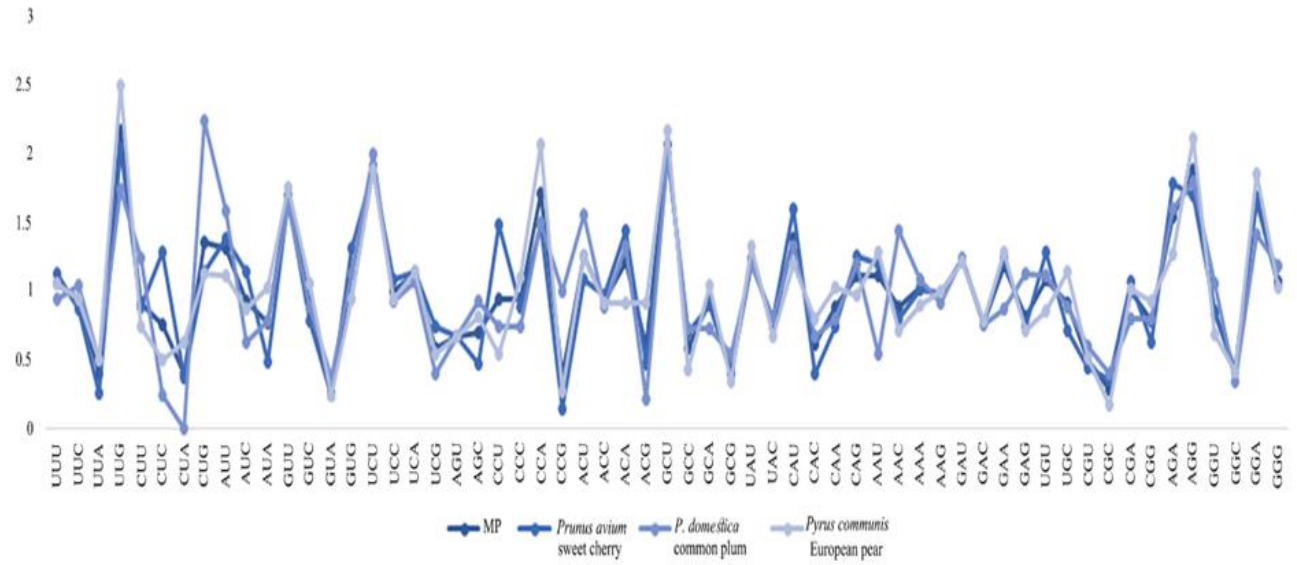

Fig. 3. Comparative analysis of relative synonymous codon usage (RSCU) patterns between PrVT-MP and its hosts

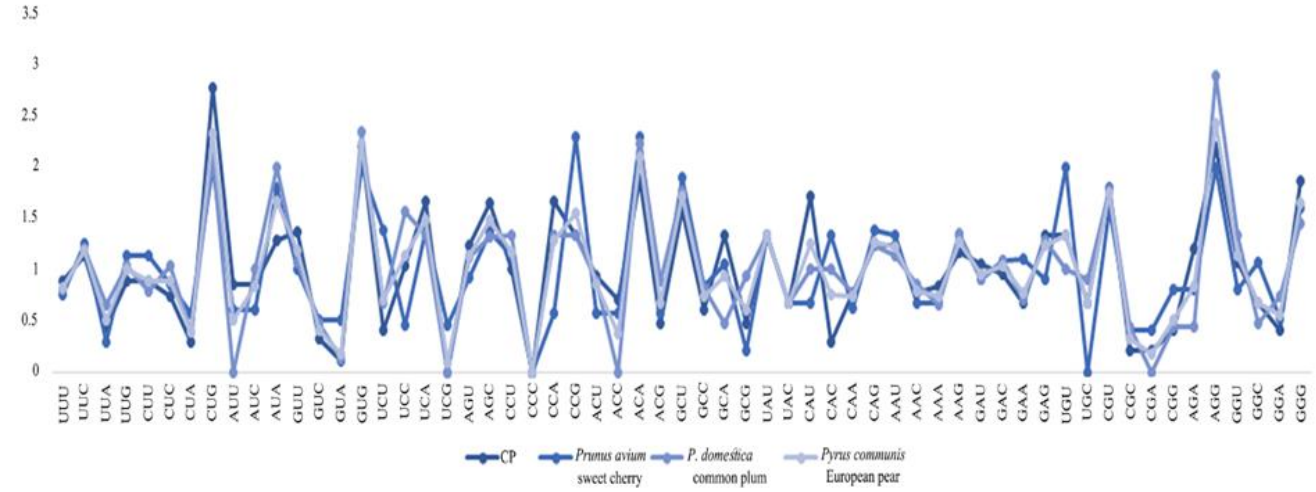

Fig. 4. Comparative analysis of relative synonymous codon usage (RSCU) patterns between PrVT-CP and its hosts

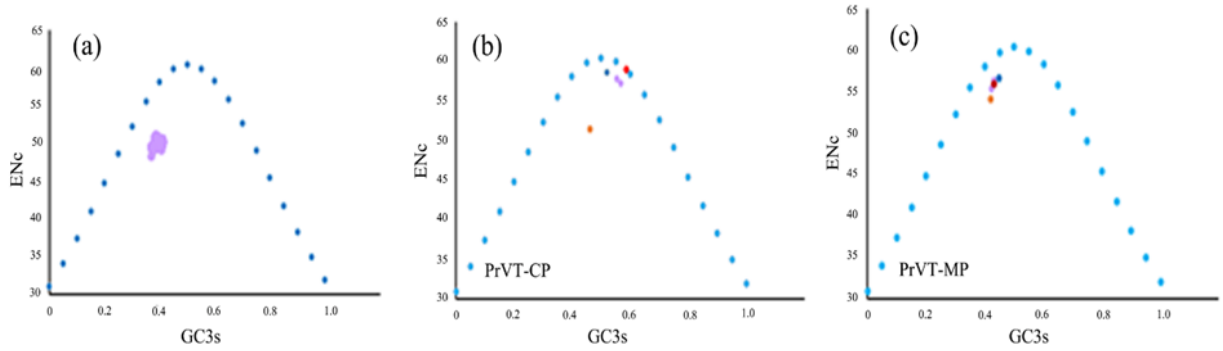

Fig. 5. ENC-plot analysis of the CVA and PrVT ORFs with ENC curve drawn against GC3s. The standard curve was plotted while using the codon usage bias (calculated by the GC3s composition only) shown by blue points. (a): CVA-polyprotein; (b) PrVT-CP; (c) PrVT-MP.

In general, codon adaptation is estimated using metrics derived from the RSCU, the ratios between the observed usage frequency of a specific codon in a gene and its expected one in the correspondent synonymous family. AU or GC-rich genomes are more likely to correlate with the RSCU patterns, particularly in viruses. For example, an AU or GC rich virus genome prefers codons ending with A/U or G/C, which supports the impact of mutation pressure (23, 27-28).

The probability of synonymous codons of different species being used is not always equal. Some synonymous codons are used more frequently, while some codons are used less frequently (29). Analyzing the codon bias features and variations of dissimilar species is fruitful for understanding the genetic structure and evolution bias of species (30, 31). Optimal codon selection in obligate parasites, including viruses, significantly depends on the translation machinery of their host cells, thus, the interplay between the codon usage of the virus and their hosts is expected to influence the overall viral survival, fitness, eluding of the host immune system and evolution (32). It can also be concluded from this study that the synonymous codon usage of the aforementioned viruses was a mixture of coincidence and antagonism. This phenomenon was previously reported for human infecting viruses, and to our knowledge, there are very few studies of this kind in plant viruses (33).

Also, for some of the amino acids, the preferred codon in our studied viruses showed similarity with their hosts. This scientific fact which is called "mutual codon preference of host-pathogens" indicates that a similar pattern of codon usage among a virus and its host could result in more effective protein synthesis by the virus. Meanwhile, dissimilarity of codon preference between any of the studied viruses and hosts causes weak translation and protein synthesis of the corresponding virus.

The low ENC values represent an approximately constant and conserved genome composition. The lower codon usage bias has already been shown for some DNA and RNA plant viruses (19, 21, 34-36). This is in favor of the viruses because it reduces the competition between the virus and their hosts for the use of the protein synthesis machine, and thus the replication efficiency in the host cell increases (37).

Furthermore, the RSCU, nucleotide composition, and ENC-plot analysis results indicate that the selection pressure was greater than the mutation pressure. According to the viral features, the interplay between the codon usage of the virus and that of its host is expected to affect the overall viral survival, fitness, and evolution. Our analysis indicates synonymous codon usage in CVA and PrVT possibly mirroring a dynamic process of mutation from the viruses along with selection pressure to adjust thier codon usage to various hosts, and conditions. Such bioinformatics analysis expands our knowledge of differences and common features in the codon usage patterns of betaflexiviruses with their various hosts.

Acknowledgment

This work was carried out in the laboratory of the Plant Virus Research Department, Iranian Research Institute of Plant Protection (IRIPP), Agricultural Research, Education and Extension Organization (AREEO). Financial support from the IRIPP project (Projects No. 961463 and 83063) is appreciatively acknowledged.

Disclosure

None

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Code Availability

No custom or special code or mathematical algorithm was used in this study.

Availability of Data and Material

The data sets analyzed during the current study

are available in the GenBank repository (https: //www.ncbi.nlm.nih.gov/)

Funding

The Iranian Research Institute of Plant Protection-IRIPP has supported this work in par, under project numbers (010549, 93134, and 92196).

Ethics Approval

The authors confirm that the ethical policy of the journal, as mentioned on the journal's authors guideline page, was ensured and no ethical approval was required as the study did not collect any sample or questionnaires from animals and humans.

Data Availability Statement

All data generated or analyzed during this study are included in this article.

Authors' contributions

Authors' contributions are equal.

Consent to Participate & Publication

Written informed consent to participate in the study was obtained from all individual participants.

References

1. Rubio M, Martínez-Gómez P, Marais A, Sánchez-Navarro JA, Pallás V, Candresse T. Recent advances and prospects in Prunus virology. Ann Appl Biol. 2017;171 (2):125-138.
2. Umer M, Liu J, You H, Xu C, Dong K, Luo N, et al. Genomic, morphological and biological traits of the viruses infecting major fruit trees. Viruses 2019;11(6): 515.
3. Çağlayan K, Roumi V, Gazel M, Elçi E, Acioğlu M, Plesko IM, et al. Identifcation and characterization of a novel robigovirus species from sweet cherry in Turkey. Pathogens 2019;8(2):57.
4. Kinoti WM, Nancarrow N, Dann A, Rodoni BC, Constable FE. Updating the quarantine status of prunus infecting viruses in Australia. Viruses 2020;12(2):246.
5. Da Silva LA, Oliveira A S, Melo FL, Ardisson-Araujo DMP, Resende FV, Resende RO, et al. A new virus found in garlic virus complex is a member of possible novel genus of the family Betaflexiviridae (order Tymovirales). PeerJ. 2019;7:e6285.
6. Liu H, Wu L, Zheng L, Cao M, Li R. Characterization of three new viruses of the family Betafexiviridae associated with camellia ringspot disease. Virus Res. 2019; 272:197668.
7. Marais A, Faure C, Mustafayev E, Barone M, Alioto D, Candresse T. Characterization by deep sequencing of prunus virus T, a novel tepovirus infecting Prunus species. Phytopathology 2015;105(1):135-40.
8. Marais A, Faure C, Mustafayev E, Candresse T. Characterization of new isolates of apricot vein clearing associated virus and of a new prunus-infecting virus: Evidence for recombination as a driving force in *Betaflexiviridae* evolution. PLoS One. 2015;10(6):e0129469.
9. Rwahnih MA, Alabi OJ, Hwang MS, Stevens K, Golino D. Identification and genomic characterization of grapevine Kizil Sapak virus, a novel grapevine infecting member of the family Betaflexiviridae. Arch Virol. 2019;164(12):3145-49.
10. Marais A, Candresse T, Svanella-Dumas L, Jelkmann W. "Cherry virus A", In: Virus and Virus-Like Diseases of Pome and Stone Fruits, eds. Hadidi A, Barba M, Candresse T, Jelkmann W (USA). 2011;147-150.
11. James D and Jelkmann W. Detection of cherry virus A in Canada and Germany. 17th International Symposium on Virus and Virus-Like Diseases of Temperate Fruit Crops: Fruit Tree Dis. 1998;472:299-303.
12. Ben Mansour K, Komínek P, Komínková M, Brožová J. Characterization of Prunus necrotic ringspot virus and Cherry virus A infecting myrobalan rootstock. Viruses 2023;15(8):1723.
13. Marais A, Faure C, Svanella-Dumas L, Candresse T. First report of Cherry virus A in *Prunus mume* in China. Plant Dis. 2008;92(11):1589.
14. Costa LC, Hu Xi, Malapi-Wight M, Foster J, McFarland C, Hurtado-Gonzales OP. Identification of a novel robigovirus and a Prunus-infecting tepovirus in Pyrus communis and their transmissibility on Malus spp. Eur J Plant Pathol. 2021;162:275-88.
15. Noorani MS, Khan AJ, Khursheed S. Molecular characterization of cherry virus A and prunus necrotic ringspot virus and their variability based on nucleotide polymorphism. Arch Phytopathol Plant Protect. 2022;55(4): 387-404.
16. Gao R, Xu Y, Candresse T, He Z, Li S, Ma Y, et al. Further insight into genetic variation and haplotype diversity of Cherry virus A from China. PLoS One. 2017;12(10):e0186273.
17. Pourrahim R, Farzadfar Sh. The incidence and genetic analysis of two betaflexiviruses capillovirus alphavii and Tepovirus tafpruni in Iran. Plant Pathol J. 2025;41 (1):38-50.

18. Gu H, Chu DKW, Peiris M, Poon LLM. Multivariate analyses of codon usage of SARS-CoV-2 and other beta-coronaviruses. Virus Evol. 2020;6(1);veaa032.
19. He Z, Gan H, Liang X. Analysis of synonymous codon usage bias in *Potato virus M* and its adaption to hosts. Viruses. 2019;11(8):752.
20. Liu XS, Zhang YG, Fang YZ, Wang YL. Patterns and influencing factor of synonymous codon usage in porcine circovirus. Virol J. 2012;9:68.
21. Xu XZ, Liu QP, Fan LJ, Cui XF, Zhou XP. Analysis of synonymous codon usage and evolution of begomoviruses. J Zhejiang Uni Sci B. 2008;9(9):667-74.
22. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGAX: Molecular evolutionary genetics analysis across Computing Platforms Sudhir. Mol Biol Evol. 2018;35(6):1547-49.
23. Sharp PM, Li WH. An evolutionary perspective on synonymous codon usage in unicellular organisms. J Mol Evol. 1986;24(1-2):28-38.
24. Wong EHM, Smith DK, Rabadan R, Peiris M, Poon LLM. Codon usage bias and the evolution of infuenza A viruses. Codon usage biases of infuenza virus. BMC Evol Biol. 2010;10:253.
25. Sharp PM, Tuohy TMF, Mosurski KR. Codon usage in yeast: Cluster analysis clearly differentiates highly and lowly expressed genes. Nuc Acids Res. 1986;14(3): 5125-5143.
26. Bouquet J, Cherel P, Pavio N. Genetic characterization and codon usage bias of full-length hepatitis E virus sequences shed new lights on genotypic distribution, host restriction and genome evolution. Infect Genet Evol. 2012;12(8):1842-53.
27. Gu W, Zhou T, Ma J, Sun X, Lu Z. Analysis of synonymous codon usage in SARS coronavirus and other viruses in the Nidovirales. Virus Res. 2004;101(2): 155-61.
28. Jenkins GM, Holmes EC. The extent of codon usage bias in human RNA viruses and its evolutionary origin. Virus Res. 2003;92(1):1-7.
29. Liu H, Lu Y, Lan B, Xu J. Codon usage by chloroplast gene is bias in Hemiptelea davidii. J Genet. 2020;99:8.
30. Li Q, Luo Y, Sha A, Xiao W, Xiong Z, Chen X, et al. Analysis of synonymous codon usage patterns in mitochondrial genomes of nine Amanita species. Front Microbiol. 2023;14:1134228.
31. Yang C, Zhao Q, Wang Y, Zhao J, Qiao L, Wu B, et al. Comparative analysis of genomic and transcriptome sequences reveals divergent patterns of codon bias in wheat and its ancestor species. Front Genet. 2021;12: 732432.
32. Moratorio G, Iriarte A, Moreno P, Musto H, Cristina J. A detailed comparative analysis on the overall codon usage patterns in West Nile virus. Infect Genet Evol. 2013;14:396-400.
33. Shafat Z, Ahmed A, Parvez MK, Parveen S. Decoding the codon usage patterns in Y-domain region of hepatitis E viruses. J Genet Eng Biotechnol. 2022;20 (1):56.
34. Biswas KK, Palchoudhury S, Chakraborty P, Bhattacharyya UK, Ghosh DK, Debnath P. et al. Codon usage bias analysis of *Citrus tristeza virus*: Higher codon adaptation to citrus reticulata host. Viruses 2019;11(4):331.
35. Chakraborty P, Das S, Saha B, Sarkar P, Karmakar A, Saha A, et al. 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(8):555-64.
36. Pourrahim R, Farzadfar Sh. Insights into the factors affecting synonymous codon usage in apple mosaic virus and its host adaptability. J Plant Biotech. 2022;49 (1):46-60.
37. Zhang Z, Dai W, Wang Y, Lu C, Fan H. Analysis of synonymous codon usage patterns in torque teno sus virus 1 (TTSuV1). Arch Virol. 2012;158(1):145-54.