agronomy

MDPI

Article

Study of Codon Usage Patterns and Influencing Factors in Rice Yellow Mottle Virus Based on Coding Sequence Data

Siddiq Ur Rahman [1,*], Sajid Nawaz [1], Sifat Ullah [1], Inayat Ur Rahman [2,3,*], Muhammad Inam Ul Haq [1], Muazzam Ali Khan [4], Abdullah Ahmed Al-Ghamdi [5], Fahad M. Al-Hemaid [5], Mohamed S. Elshikh [5], Reem M. Aljowaie [5] and Wafa Ali Eltayb [6]

1 Department of Computer Science and Bioinformatics, Khushal Khan Khattak University, Karak 27200, Khyber Pakhtunkhwa, Pakistan
2 Department of Botany, Khushal Khan Khattak University, Karak 27200, Khyber Pakhtunkhwa, Pakistan
3 Missouri Botanical Garden, P.O. Box 299, St. Louis, MO 63166-0299, USA
4 Department of Botany, Bacha Khan University, Charsadda 24631, Khyber Pakhtunkhwa, Pakistan
5 Department of Botany and Microbiology, College of Science, King Saud University, P.O. Box 2455, Riyadh 11451, Saudi Arabia
6 Biotechnology Department, Faculty of Science and Technology, Shendi University, Shendi 11111, Sudan
* Correspondence: siddiqur.rahman@kkkuk.edu.pk (S.U.R.); iurahman@kkkuk.edu.pk (I.U.R.)

Abstract: Rice yellow mottle virus (RYMV), transmitted by chrysomelids, is one of the major viral pathogens that has devastated rice production in Sub-Saharan Africa. RYMV is a member of the genus *Sobemoviruses* in the family *Solemoviridae* and harbors a positive-sense single-stranded RNA (+ssRNA). Here, we used 50 RYMV strains, applying the codon usage bias (CUB). Both base content and relative synonymous codon usage (RSCU) analysis revealed that GC-ended codons were more frequently used in the genome of RYMV. Further low codon usage bias was observed from the effective number of codons (ENC) value. The neutrality plot analysis suggested the dominant factor of natural selection was involved in the structuring of CUB in RYMV. Based on RSCU values, the RYMV and its host relationship indicate that the RYMV develops codon usage patterns similar to its host. Generally, both natural selection and mutational pressure impact the codon usage pattern of the protein-coding genes in RYMV. This study is important because it characterized the codon usage pattern in the RYMV genomes and provided the necessary data for a basic evolutionary study on them. Additionally, we recommend that experiments such as whole genome sequencing (WGS) or dual RNA sequencing (DRS) should be considered in order to correlate these in-silico findings with viral diseases in the future.

Keywords: rice yellow; codon usage; mutation pressure; natural selection

check for updates

Citation: Rahman, S.U.; Nawaz, S.; Ullah, S.; Rahman, I.U.; Haq, M.I.U.; Khan, M.A.; Al-Ghamdi, A.A.; Al-Hemaid, F.M.; Elshikh, M.S.; Aljowaie, R.M.; et al. Study of Codon Usage Patterns and Influencing Factors in Rice Yellow Mottle Virus Based on Coding Sequence Data. *Agronomy* **2022**, *12*, 1990. https://doi.org/10.3390/agronomy12091990

Academic Editor: Caterina Morcia

Received: 23 July 2022
Accepted: 19 August 2022
Published: 23 August 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

The rice yellow mottle virus (RYMV) is a member of the *Solemoviridae* family, which belongs to the genus *Sobemovirus*. It has four ORFs for protein-coding and a positive-sense single-stranded RNA (ssRNA) molecule [1]. The ORF1, which encodes the small protein P1, is implicated in both the suppression of gene silencing and viral mobility [2]. The coat protein is fixed by ORF3 while ORFs 2a and 2b translate the central polyproteins P2a and P2ab [3]. The genome also has three non-coding regions at the 5′ and 3′ ends and between the first two ORFs [4]. The virus particles are highly infectious and mechanically transmissible [3,5].

Rice yellow mottle virus (RYMV) causes a vector-borne viral infection (Rice yellow mottle disease (RYMD)) that has a significant impact on rice production, particularly in Africa [6]. The major insect vectors are chrysomelids, which play an essential role in primary infection, while secondary infection is due mainly to plant-to-plant contact. The chrysomelids are also accountable for increasing the same virus in grass species

such as *Panicum repens* L., *Echinochloa colona* L. Link, and *Ischaemum rugosum* Salisb [7,8]. Despite the fact that RYMV is not seed transmitted [9,10]. The symptoms of RYMD include mottling, yellow to orange discoloration of the affected leaves, stunted growth, sterile panicles, and likely crop loss (depending on the cultivar). Usually, these signs appear at different developmental stages, including leaf initiation, early heading, tillering, and panicle initiation [6,11].

RYMD is a disease that could be lethal with an extensive past in rice production. RYMV infection was first identified in 1970 in the western part of Kenya, Nyanza province [5], and later spread to West, Central, and East Africa. From 1976 until late 1980, several incidents of RYMV infection were detected in Sierra Leone, Nigeria, Tanzania, and Liberia [12,13], and infections spread to several locations in Côte d'Ivoire (Ivory Coast) [14], Koba in Guinea and Ghana [15], Mali, and Burkina Faso, along with in Malawi [16] and Rwanda, producing a global economic emergency [6]. The severity of the infection has been increasing since 1989, when rice tilling was uncontrolled in Lake Alaotra and at Marovoay [6,17]. At the beginning of the 21st century, the virus infection had been described in Chad and Cameroon [18]. Recently, the virus's occurrence has been restricted to Africa. The prevalence and detrimental effects of the disease change depending on rice variety, location, and vegetation zones [19].

Besides the transmission through insects, other ways of RYMV transmission include plant-to-plant, contaminated soil and water, and infected rice plants. The RYMV is also transmitted through donkeys, grass rats, and cows in watered rice fields [8]. Restrictions on the spread of the RYMV include limited vector movement, a narrow host variety, and a lack of seed spread. Highly and moderately challenging varieties have been discovered. Nonetheless, the infection is currently not adequately operated, and its prevalence in Africa is rising significantly [20].

To explore the molecular and evolutionary research of the virus, we may need to implement the phenomenon called codon usage bias (CUB) [21]. Apart from tryptophan and methionine, distinct synonymous codons encode all the other 18 amino acids (aa). This is codon degeneracy, in which the synonymous codons occurring in the genome are not in an equal number and follow a proper codon usage pattern [22,23]. There are several factors which impact the configuration of CUB. Natural selection and mutational pressure, on the other hand, were two major factors that shaped CUB [24,25]. Despite notable advancements made in scientific research, the CUB's evolutionary progression is still an arguable subject [22,26,27].

Several previous studies on codon usage using viruses have revealed that mutational pressure is the most important factor influencing the shape of codon usage associated with selection pressure [24,26]. However, as our understanding of codon usage increases, it becomes clear that, while mutational pressure is a major force, it is not the only one when various types of DNA and RNA viruses are studied [27]. Comparison of the genomes of prokaryotes and eukaryotes with viruses and highlights such as dependance on the host environment for major procedures comprising the synthesis of protein, replication, and spreading, based on codon usage, the interaction between hosts and viruses is supposed to influence general viral suitability, existence, or evasion of host cell retorts and evolution [28–30]. So, study on viruses related to codon usage not only supports information on evolution at the molecular level but also enhances knowledge of viral gene expression regulation and helps in designing a vaccine, wherever the effective appearance of viral proteins might be required to cause resistance.

In this study, we analyzed 50 RYMV strains associated with African and European families. Presently, the virus infection poses a serious threat to a rice health concern. The goal of this study was to determine the effects of viral relationships with their hosts on viral capability, as well as to investigate factors that influence the patterns of codon usage in the RYMV genome.

2. Materials and Methods

2.1. Sequences

A total of 50 coding sequences (CDS) of the RYMV were retrieved from the National Centre for Biotechnology Information (NCBI). The number of nucleotides in the coding sequence was an exact multiple of three [31]. Complete information about the overall 50 RYMV isolates associated with African and European families is listed in Table S1. As the CDS for the Asian isolates is not available in GeneBank, they were excluded from the study. According to recent studies, the rice yellow mottle virus (RYMV) is only found in Africa [32,33].

2.2. Analysis of Nucleotide Composition

Here we employed CodonW software to determine the total base composition (G, C, A, and U%) and the contents of the nucleotide at the 3rd codon location (C3, U3, G3, and A3%) for all synonymous codons in this virus. The GC% contents of all three codon locations (GC1, GC2, and GC3%) were measured. Additionally, the average frequency of G/C at GC12 locations and the overall GC/AU compositions were also determined. Further, only 59 synonymous codons encoding 18 amino acids were considered for the present study, not including the 1st ATG codon, the codon (TGG) encoding tryptophan, and the three ended codons (TAG, TAA, and TGA), respectively [34].

2.3. Analysis of Relative Synonymous Codon Usage (RSCU)

RSCU values compare the observed codon occurrence to its random occurrence, indicating that the virus uses all the identical codons equally. The CDS contains both frequently occurring codons with RSCU values greater than one and rarely occurring codons with RSCU values lower than one [35]. Higher CUB or more frequently used codons were determined through high RSCU. In the coding sequence, the overrepresented codon represents the codon RSCU value greater than 1.6 and the underrepresented codon represents the codon RSCU value less than 0.6 in the coding sequence [36].

The RSCU value was determined for each codon by following the formula [37]:

$$RSCU = \frac{X_{ij}}{\sum_{j=1}^{n_i} X_{ij}} n_i$$

Here, n_i is the number of codons for the ith amino acid and X_{ij} denotes the frequency of the jth codon for the ith amino acid.

2.4. Analysis of Codon Adaptation Index (CAI)

The CAI is applied to calculate the gene expression level depending on its codon-based sequence. The value of the CAI value varies from zero to one; a value near one indicates higher levels of codon usage bias [38]. The CAI was determined through an online tool used for CAI calculation, "CAIcal" [39], where the *O. sativa* genome was used as a reference source. Furthermore, the e-CAI (expected CAI) was analysed using the online tool "CAIcal". The values of RSCU for the *O. sativa* genome were retrieved from the database of codon usage.

2.5. Indices of Codon Usage

To determine the proper measurement of codon bias, ENC (effective number of codons) was performed, which measures the total usage of the codon in a certain gene [27,40]. It clarifies the ratio of codon variation in a gene from the total even usage of codons which are synonymous. The ENC value varies from 20 (where one amino acid encodes one codon only) to 61 (where each amino acid is used randomly for all codons). ENC value <35 implies significant CUB [40,41]. A plot between ENC-GC3s is drawn out to determine the impact

of GC3s composition on codon usage [40]. For each GC3, the expected ENC values were considered by using the following formula:

$$ENC = 2 + s + 29/(s^2 + (1 - s)^2)$$

wherever 's' is the GC3s content of each codon.

2.6. Correspondence Analysis (COA)

COA is a multidimensional critical method to resolve the important developments in the codon usage patterns of CDS through codon RSCU values [36,42,43]. To create the COA plot, RSCU values of 59 codons were considered. To study the tendencies in the deviation of the use of codon, relative inertia was used to hold a specific position in the graph.

2.7. Analysis of Neutrality Plot

GC12 and GC3 were studied for attaining a neutrality plot to identify the role of selection-mutation symmetry in the usage of codon discrepancy. In the graph, each point signifies a discrete gene. The line of regression slop between GC3 and GC12 indicates mutation pressure is the major factor in affecting CUB, i.e., for values coming close to 1, although if the value comes close to 0, it indicates selection pressure has been the main factor in defining CUB [44,45].

2.8. Correlation Analysis

To illustrate the relations between nucleotide content and codon usage patterns, RYMV correlation analysis was performed. These analyses were conducted by using Spearman's rank correlation method [46]. All processes were executed using the R corrplot package. For codon usage index analysis, CodonW (1.4.4) software (http://codonw.sourceforge.net/) was applied [27,47].

3. Results

3.1. Base Compositional Analysis in RYMV Coding Sequences

CUB can be considerably predisposed by the general base composition of genomes. The nucleotide contents of 50 RYMV genes were studied, which are presented in Table 1. Here, our outcomes disclosed that the mean G (28.73%) and C (26.03%) were maximum, tailed by A (21.86%) and U (23.39%) across all genomes. The mean G3 (29.22%) and C3 (27.22%) occurred at a maximum level higher than the A3 (20.27%) and U3 (23.29%) (Figure 1, Table 1). The total GC and AU compositions were found to be 54.75% and 45.27%, respectively, suggesting RYMV strains have strong GC.

Table 1. A nucleotide compositional analysis of RYMV coding sequences (%).

No	A	C	U	G	GC	AU	A3	C3	U3	G3	GC1	GC2	GC3	AU3	GC3s	ENC	Gravy	Aromo
1	21.99	25.66	23.72	28.63	54.29	45.71	19.82	23.32	22.22	34.63	47.45	57.46	57.96	42.04	56.41	57.45	−0.3005	0.092184
2	21.57	26.13	23.20	29.10	55.23	44.77	16.50	30.90	21.90	30.70	57.60	46.50	61.60	38.40	60.23	56.82	−0.2919	0.092092
3	21.67	26.18	23.76	28.38	54.57	45.43	15.68	27.56	24.75	32.01	55.12	49.01	59.57	40.43	58.85	57.58	−0.3021	0.07438
4	21.73	26.29	23.71	28.27	54.57	45.43	15.84	27.72	24.59	31.85	54.95	49.17	59.57	40.43	58.78	57.32	−0.2988	0.071074
5	22.03	25.87	23.40	28.70	54.57	45.43	17.50	30.40	22.50	29.60	57.20	46.50	60.00	40.00	58.52	57.33	−0.2933	0.092092
6	22.17	26.46	23.21	28.16	54.62	45.38	17.33	27.72	23.27	31.68	54.95	49.50	59.41	40.59	58.75	57.61	−0.2925	0.07438
7	21.57	26.43	22.93	29.07	55.50	45.50	17.20	31.60	21.30	29.90	57.70	47.30	61.50	38.50	60.12	56.63	−0.2891	0.089089
8	21.53	26.13	23.33	29.00	55.13	44.87	16.90	30.80	22.00	30.30	57.40	46.90	61.10	38.90	59.67	57.83	−0.2876	0.091091
9	21.67	25.90	23.40	29.03	54.93	45.07	16.90	30.30	22.40	30.40	57.10	47.00	60.70	39.30	59.21	57.80	−0.3038	0.093093
10	21.47	26.27	23.13	29.13	55.40	44.60	17.00	31.50	21.60	29.90	57.80	47.00	61.40	38.60	59.98	57.24	−0.2798	0.091091
11	21.51	26.18	23.76	28.55	54.73	45.27	15.02	27.72	24.26	33.00	54.46	49.01	60.73	39.27	60.03	56.10	−0.3044	0.069421
12	21.62	26.57	23.43	28.38	54.95	45.05	15.84	27.72	24.09	32.34	55.12	49.67	60.07	39.93	59.36	55.92	−0.3173	0.069421
13	21.84	26.40	23.54	28.22	54.62	45.38	15.84	27.72	24.42	32.01	54.79	49.34	59.74	40.26	58.95	57.36	−0.3292	0.069421
14	21.83	26.30	23.22	28.66	54.96	45.04	27.58	25.42	24.14	22.86	60.28	56.30	48.28	51.72	48.47	59.60	−0.4022	0.074859
15	21.92	25.93	23.49	28.66	54.59	45.41	17.02	29.73	23.02	30.23	56.86	46.95	59.96	40.04	58.52	57.33	−0.2975	0.089178
16	21.66	25.57	23.73	29.04	54.61	45.39	18.15	25.98	24.22	31.65	51.08	55.13	57.62	42.38	56.61	58.32	−0.2354	0.081211
17	21.68	25.78	23.55	28.99	54.77	45.23	18.42	25.84	24.22	31.51	51.55	55.40	57.35	42.65	56.35	58.58	−0.2448	0.079725
18	21.66	25.73	23.55	29.06	54.79	45.21	18.42	25.78	24.29	31.51	51.62	55.47	57.29	42.71	56.28	58.51	−0.2439	0.079092
19	22.22	25.66	23.66	28.45	54.12	45.88	22.60	28.34	23.14	25.91	55.60	52.50	54.25	45.75	53.61	59.27	−0.3880	0.088687
20	21.95	26.07	23.23	28.74	54.81	45.19	19.84	27.33	23.48	29.35	53.10	54.66	56.68	43.32	55.79	59.73	−0.3074	0.074534
21	21.93	25.91	23.44	28.72	54.63	45.37	20.04	26.79	24.02	29.15	53.24	54.72	55.94	44.06	55.01	59.64	−0.2971	0.071923
22	21.61	25.62	23.82	28.95	54.57	45.43	20.78	29.42	24.29	25.51	56.95	51.82	54.93	45.07	54.13	58.86	−0.4036	0.089013
23	21.79	25.78	23.55	28.88	54.66	45.34	18.96	25.84	24.16	31.04	51.42	55.67	56.88	43.12	55.96	58.64	−0.2448	0.07777

Table 1. *Cont.*

No	A	C	U	G	GC	AU	A3	C3	U3	G3	GC1	GC2	GC3	AU3	GC3s	ENC	Gravy	Aromo
24	21.82	25.71	23.57	28.90	54.61	45.39	18.89	25.71	24.29	31.11	51.42	55.60	56.82	43.18	55.89	58.69	−0.2486	0.07777
25	22.29	25.96	23.17	28.59	54.54	45.46	22.20	27.73	23.55	26.52	56.55	52.83	54.25	45.75	53.59	59.21	−0.4560	0.084203
26	21.69	26.39	23.17	28.75	55.14	44.86	21.85	28.86	22.39	26.90	57.11	53.53	55.77	44.23	55.22	58.25	−0.3994	0.086171
27	22.14	26.34	22.95	28.57	54.91	45.09	22.52	29.06	22.72	25.69	57.18	52.80	54.75	45.25	54.43	58.93	−0.4301	0.082696
28	22.13	26.09	23.19	28.59	54.68	45.32	20.11	27.19	23.41	29.28	53.04	54.52	56.48	43.52	55.58	59.61	−0.3165	0.074586
29	21.94	25.87	23.62	28.57	54.44	45.56	25.02	21.38	23.47	30.14	53.61	58.19	51.52	48.48	50.40	59.71	−0.2559	0.079132
30	21.66	26.23	23.17	28.95	55.17	44.83	22.13	28.61	22.47	26.79	56.82	53.31	55.40	44.60	54.94	58.74	−0.3946	0.083043
31	22.29	25.91	23.23	28.57	54.48	45.52	22.20	27.73	23.62	26.45	56.68	52.56	54.18	45.82	53.52	59.23	−0.4475	0.084899
32	21.81	26.27	23.09	28.83	55.10	44.90	24.58	21.74	22.89	30.79	54.42	58.34	52.53	47.47	51.51	59.24	−0.2734	0.068966
33	22.12	25.85	23.32	28.71	54.56	45.44	25.47	21.83	22.10	30.59	53.37	57.88	52.43	47.57	51.65	59.07	−0.2822	0.078762
34	22.27	25.69	23.60	28.43	54.12	45.88	25.56	21.04	22.59	30.82	53.14	57.38	51.85	48.15	50.77	58.95	−0.2796	0.082691
35	22.25	25.62	23.65	28.48	54.10	45.90	22.59	27.51	23.20	26.70	56.17	51.92	54.21	45.79	53.67	59.14	−0.3769	0.085834
36	21.86	25.86	23.50	28.78	54.64	45.36	21.48	29.97	23.30	25.25	57.17	51.52	55.22	44.78	54.35	59.00	−0.4505	0.089708
37	21.97	26.07	23.30	28.65	54.72	45.28	19.97	27.26	23.68	29.08	53.04	54.79	56.34	43.66	55.55	59.93	−0.2964	0.070491
38	21.79	26.19	23.18	28.84	55.03	44.97	23.52	29.45	23.05	23.99	59.03	52.63	53.44	46.56	53.17	59.06	−0.4666	0.080724
39	21.83	26.05	23.40	28.73	54.78	45.22	22.39	28.12	22.99	26.50	57.05	52.66	54.62	45.38	53.84	59.20	−0.3835	0.087561
40	21.69	26.21	23.24	28.86	55.07	44.93	22.05	28.66	22.52	26.77	57.18	52.60	55.43	44.57	54.92	58.50	−0.3786	0.086866
41	21.66	25.78	23.59	28.97	54.75	45.25	18.69	25.64	24.43	31.24	51.42	55.94	56.88	43.12	56.02	58.83	−0.2484	0.077029
42	21.75	25.71	23.62	28.92	54.63	45.37	18.69	25.64	24.36	31.31	51.28	55.67	56.95	43.05	56.06	58.69	−0.2461	0.07777
43	21.71	26.30	23.06	28.93	55.23	44.77	22.39	28.32	22.93	26.37	57.99	53.00	54.69	45.31	54.30	59.30	−0.4041	0.082754
44	21.92	26.12	23.29	28.68	54.80	45.20	22.25	28.46	23.20	26.10	57.05	52.80	54.55	45.45	53.90	58.95	−0.4244	0.085595
45	22.12	26.41	22.95	28.52	54.93	45.07	22.52	29.20	22.72	25.56	56.91	53.14	54.75	45.25	54.48	58.84	−0.4297	0.081533
46	21.92	26.25	22.97	28.86	55.11	44.89	24.81	21.71	22.52	30.95	54.75	57.92	52.66	47.34	51.90	59.57	−0.2966	0.070827
47	21.58	26.39	23.13	28.91	55.29	44.71	21.51	24.07	22.52	31.89	48.21	61.70	55.97	44.03	54.85	59.80	−0.2944	0.069376
48	22.09	25.84	23.55	28.52	54.36	45.64	22.13	28.21	23.35	26.32	55.67	52.90	54.52	45.48	54.12	59.26	−0.4056	0.085434
49	22.18	25.92	23.42	28.48	54.39	45.61	19.76	26.57	23.74	29.94	52.73	53.94	56.51	43.49	55.28	59.64	−0.3229	0.07687
50	21.73	25.62	23.73	28.92	54.54	45.46	18.89	25.71	24.29	31.11	51.42	55.40	56.82	43.18	55.96	58.74	−0.2293	0.07989
Mean	**21.86**	**26.03**	**23.39**	**28.73**	**54.75**	**45.27**	**20.27**	**27.22**	**23.29**	**29.22**	**54.88**	**52.97**	**56.44**	**43.56**	**55.59**	**58.55**	**−0.32791**	**0.08092**
SD	**0.23**	**0.27**	**0.25**	**0.24**	**0.32**	**0.31**	**3.04**	**2.62**	**0.87**	**2.67**	**2.76**	**3.63**	**2.97**	**2.97**	**2.76**	**1.00**	**0.067628**	**0.007244**

ENC represents the effective number of codons. GC1 represents the G + C content at the first position of codons. GC2 represents the G + C content at the second position of codons. GC3 represents the G + C content at the third positions of codons. AU3 represents the A + U content at the third positions of codons. Gravy represents the hydrophobicity of protein. ARO represents the aromaticity of protein.

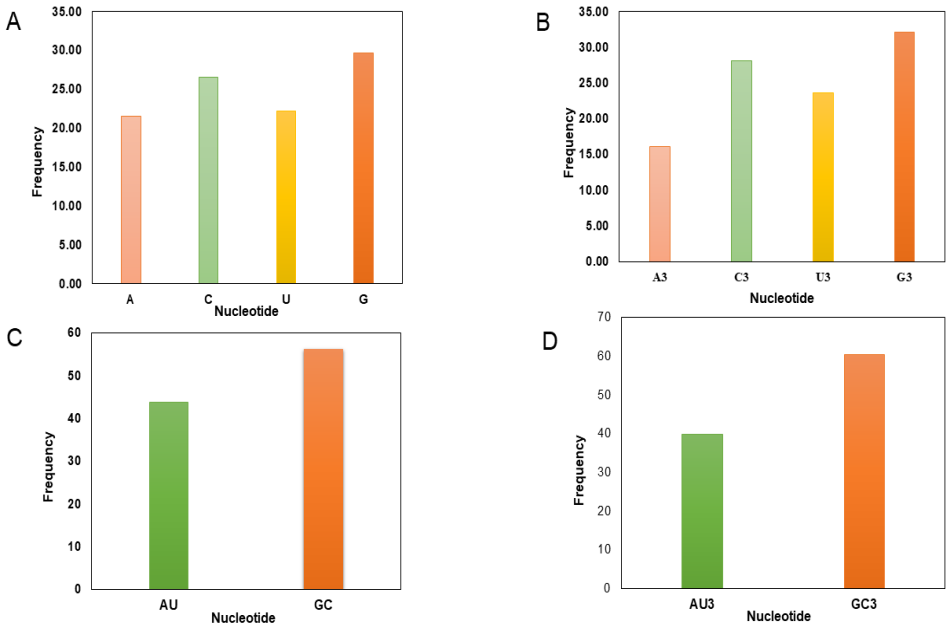


Figure 1. A nucleotide composition analysis: (**A**) The average A, U, G, and C nucleotide composition of the entire viral genome. (**B**) The average values of the nucleotide composition at the third codon position, indicating G/C richness followed A/U richness. (**C**) The mean frequency for GC and AU composition (**D**) The mean frequency of AU and GC at the codon's third position, indicating that GC3 is more common than AU3.

Nucleotide content analysis at the 1st, 2nd, and 3rd synonymous codon positions disclosed that the values of GC1 ranged from 47.45 to 60.28% (mean: 54.88%; SD: 2.76), while GC2 ranged from 46.50 to 61.70% (mean: 52.97%; SD: 3.63). Whereas the GC3 values ranged from 51.52 to 61.70% (mean: 56.44%; SD: 2.97), in contrast, the values of AU3 ranged from 38.40 to 51.72% (mean: 43.56%; SD: 2.97). These data further supported the notion that an extensive area of RYMV is self-possessed of C/G content (Table 1).

3.2. Defining Codon Usage Patterns

An RSCU analysis was used to regulate the identical pattern of codon usage in the RYMV CDS. Fifteen of the eighteen most abundantly used codons in RYMVs [UUC (Phe), CUC (Leu), AUC (Ile), GUG (Val), UCC (Ser), ACG (Thr), CAC (His), UAC (Tyr), CAG (Gln), AAC (Asn), AAG (Lys), GAC (Asp), GAG (Glu), CGG (Arg), and GGG (Gly)] ended with C or G (C: 8; G: 7). Despite this, the remaining three codons (CCA, GCU and UGU) had ended with A or U. Thus, the C or G-end codon bases are more shared in the genome of RYMV. Furthermore, from the RSCU analysis, we found that the over-represented (>1.6) codons are rarely seen in the genome of RYMV. Nearly all the ideal and non-ideal codons are in the range of 0.6 to 1.6. We observed that most codons ending in A were under-represented (<0.6), while codons ended in C were over-represented (>1.6) (Figure 2, Table 2), which was similar in *O. sativa*, where the C/G ending codons are preferred over the A/U ending codons.

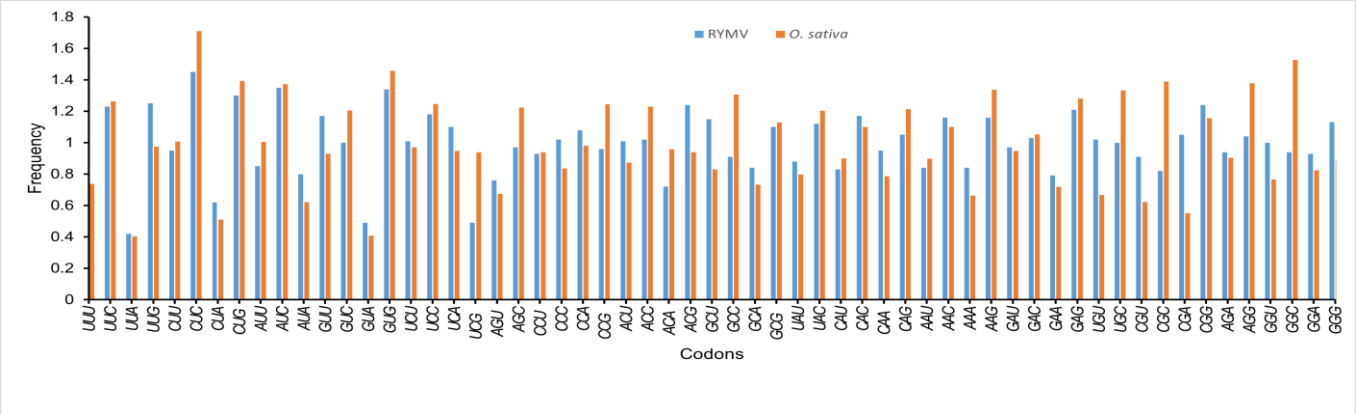


Figure 2. A comparative analysis of RSCU patterns between Rice yellow mottle virus (RYMV), and its host *O. Sativa*. *X*-axis represents Codons. *Y*-axis represents Frequency.

Table 2. The relative synonymous codon usage frequency of RYMV and its natural host.

	RSCU				RSCU		
AA	**Codon**	**RYMV**	***O. sativa***	**AA**	**Codon**	**RYMV**	***O. sativa***
Phe	UUU	**0.77**	**0.738**	Ala	GCU	**1.15**	**0.831**
	UUC	**1.23**	**1.262**		GCC	**0.91**	**1.306**
Leu	UUA	**0.42**	**0.404**		GCA	**0.84**	**0.734**
	UUG	**1.25**	**0.975**		GCG	**1.10**	**1.128**
	CUU	**0.95**	**1.008**	Tyr	UAU	**0.88**	**0.797**
	CUC	**1.45**	**1.71**		UAC	**1.12**	**1.203**
	CUA	**0.62**	**0.51**	His	CAU	**0.83**	**0.900**
	CUG	**1.30**	**1.392**		CAC	**1.17**	**1.100**
Ile	AUU	**0.85**	**1.005**	Gln	CAA	**0.95**	**0.787**
	AUC	**1.35**	**1.373**		CAG	**1.05**	**1.213**
	AUA	**0.80**	**0.623**	Asn	AAU	**0.84**	**0.899**
Val	GUU	**1.17**	**0.93**		AAC	**1.16**	**1.101**
	GUC	**1.00**	**1.205**	Lys	AAA	**0.84**	**0.663**
	GUA	**0.49**	**0.408**		AAG	**1.16**	**1.337**
	GUG	**1.34**	**1.457**	Asp	GAU	**0.97**	**0.948**
Ser	UCU	**1.01**	**0.971**		GAC	**1.03**	**1.052**
	UCC	**1.18**	**1.246**	Glu	GAA	**0.79**	**0.718**
	UCA	**1.10**	**0.948**		GAG	**1.21**	**1.282**
	UCG	**0.49**	**0.94**	Cys	UGU	**1.02**	**0.667**
	AGU	**0.76**	**0.673**		UGC	**1.00**	**1.333**

Table 2. *Cont.*

		RSCU				RSCU	
AA	**Codon**	**RYMV**	***O. sativa***	**AA**	**Codon**	**RYMV**	***O. sativa***
	AGC	**0.97**	**1.223**	Arg	CGU	**0.91**	**0.621**
					CGC	**0.82**	**1.388**
Pro	CCU	**0.93**	**0.94**		CGA	**1.05**	**0.552**
	CCC	**1.02**	**0.836**		CGG	**1.24**	**1.155**
	CCA	**1.08**	**0.981**		AGA	**0.94**	**0.905**
	CCG	**0.96**	**1.244**		AGG	**1.04**	**1.379**
Thr	ACU	**1.01**	**0.874**	Gly	GGU	**1.00**	**0.766**
	ACC	**1.02**	**1.229**		GGC	**0.94**	**1.527**
	ACA	**0.72**	**0.957**		GGA	**0.93**	**0.823**
	ACG	**1.24**	**0.94**		GGG	**1.13**	**0.885**

AA represents amino acid; the "RSCU" value represents the pattern of relative synonymous codon usage; over-represented (RSCU > 1.6), and under-represented (RSCU < 0.6) codons are marked as bold with Blue and green, respectively; the codons favored by RYMV, and host (RSCU > 1) are marked as bold with red colors, and the optimal codons for RYMV are underlined.

From both the nucleotide content and RSCU analysis, we assumed that the selection of the preferred codons has been generally inclined by compositional restraints, which also determine the existence of mutational pressure. We are unsure that the compositional pressure could not be the single aspect related to RYMV patterns of codon usage, as although the total values of RSCU could disclose the pattern of codon usage for the genomes, it may conceal the codon usage variation amongst distinct genes in a genome [48]. Here, the patterns of codon usage and selection of the preferred codons in RYMV genomes showed complete synergism with those of the host, *O. sativa* (Figure 2).

3.3. Adaptation of RYMV to the Host Genome

The Codon adaptation index (CAI) analysis was executed to regulate the optimization of codon usage and RYMV adaptation to its host [49]. The values of CAI range from zero to one. For all codons, the CAI values were measured through the reference of *O. sativa* codon usage. We determined that, concerning *O. sativa*, the mean CAI value of RYMV coding regions was 0.78 (>0.5), which revealed that RYMV has a good adaptation to its host and a minimal translation pressure (Figure 3).

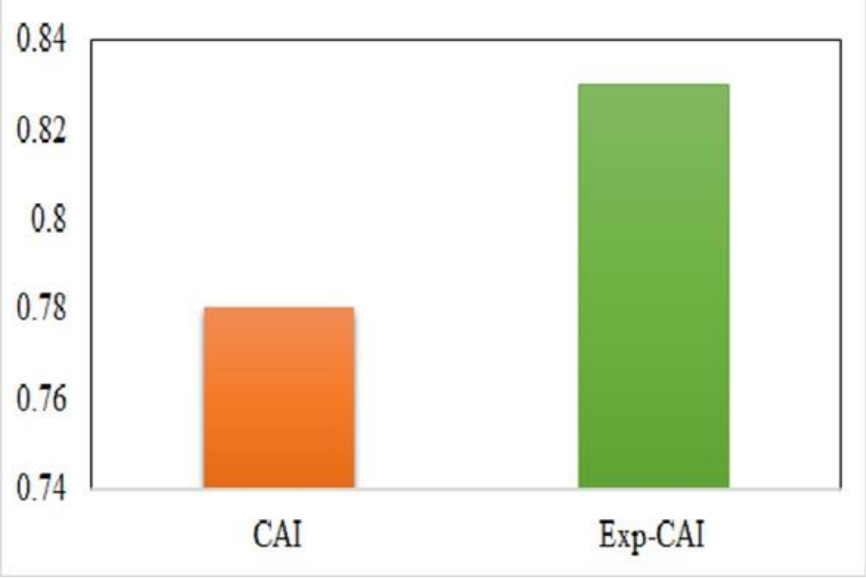


Figure 3. A codon usage adaptation index of RYMV to its host. In the plot, *x*-axis represents CAI and Exp-CAI. *Y* axis represents CAI value.

To check the observed significant statistical differences arises in the values of CAI [40,50], the values of expected CAI (e-CAI) were considered for RYMV CDS with *O. Sativa* codon usage sets. The result of the e-CAI value was 0.83 ($p < 0.05$) in relation to *O. sativa* revealing that the generated sequences keep to a normal distribution (Figure 3).

3.4. Use of Codons Biases in RYMV

To know the strength of CUB within RYMV coding sequences, the genes' ENC value was assessed and mapped next to the GC content at the 3rd codon position (GC3) (Table 1). To determine the codon usage of the genes, a plot of distribution was employed that deviated from the same usage of indistinguishable codons. The values of ENC were observed to vary from 55.92 to 59.93, indicating a high level of genetic differences in the codon's usage. Nevertheless, the average value of ENC was 58.55 > 35, implying that the whole CUB was low (Figure 4).

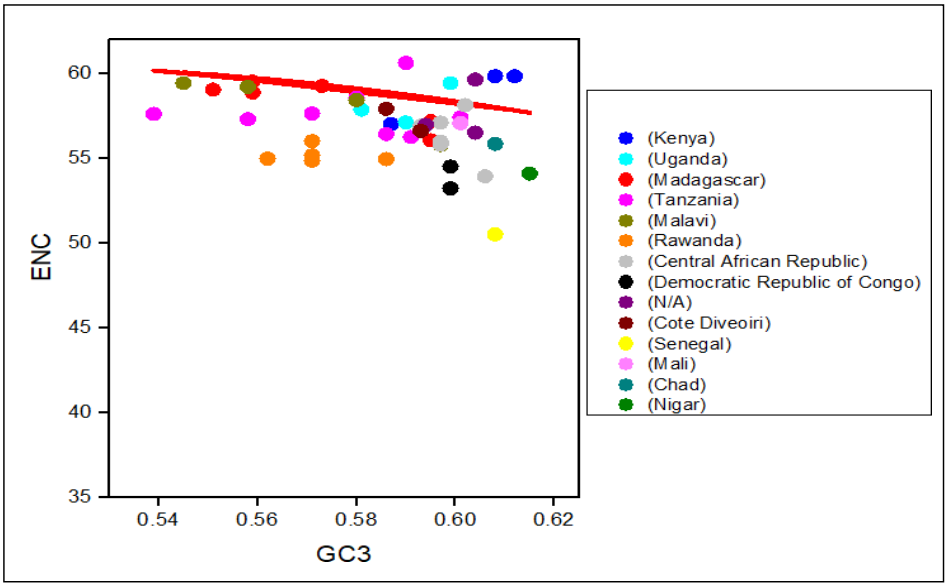


Figure 4. ENC-GC3 plots of 50 RYMVs strains: The effective number of codons (ENC-values, *Y*-axis) was plotted against the GC-content at the third synonymous codon positions (GC3-values, *X*-axis). The ENC values ranged from 30 to 61. Interestingly, all the RYMV isolates were between the 55.92 and 59.93 range and slightly below the 61, suggesting that codon usage bias is caused by selection pressure. The continuous red line represents theoretical ENC values for random codon usage as a function of GC3.

3.5. Discrepancy in the Usage of Codon among RYMVs

Scattered data on the principle axis represents diverse geographical ancestries and their relationships. All the RYMV strains were found in groups using COA (Figure 5). RYMV isolates from Uganda and Madagascar were assembled into one clade, while isolates from Niger, Sierra Leone, Tanzania, and Mali were found in another. The Kenya (first epidemic site of RYMV) and Uganda isolates were assembled into one clade. Though isolates from France were assembled in another clade, they gathered with Togo. Moreover, Nigeria and Kenya separately were dispersed into distinct groups. Interestingly, the same genetic lineage was found in Kenya-Uganda, where both relatives were isolated from the same origin (Figure 5).

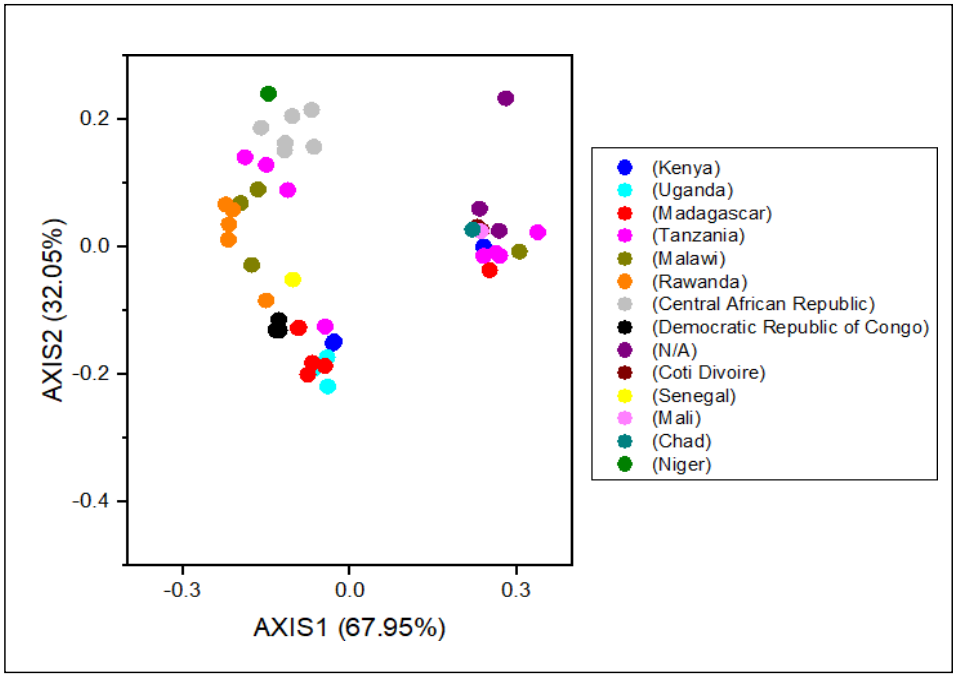


Figure 5. A correspondence analysis (COA) of the genes in RYMV genomes. Each point represents a gene corresponding to the coordinates of the first and second axes of variation generated from the correspondence analysis.

The impact of evolutionary processes on the RYMV codon usage pattern was assessed using a phylogenetic analysis. All the RYMV strains were clustered into distinct groups, similar to the pattern detected in COA (Figures 5 and 6).

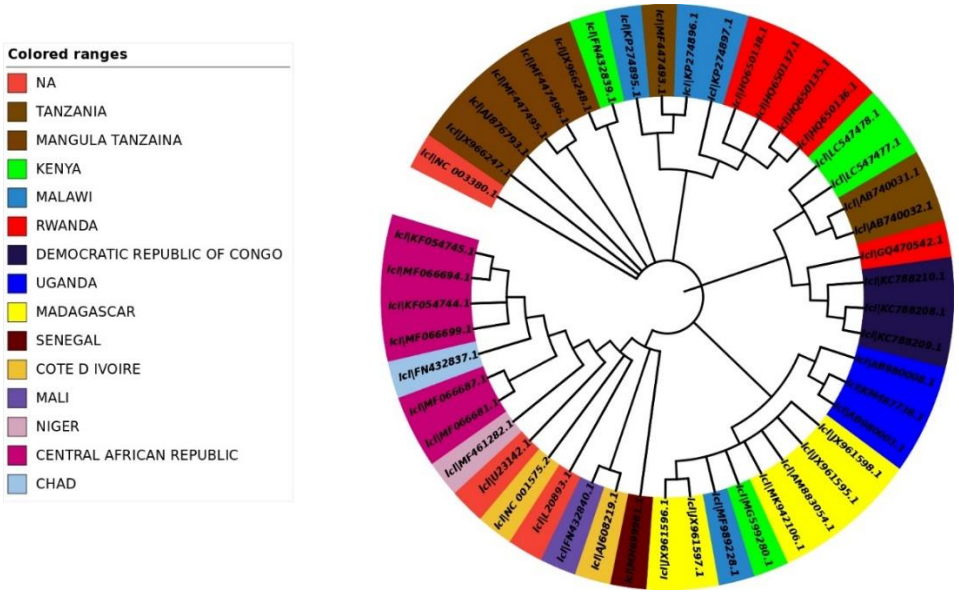


Figure 6. Phylogenetic trees are based on the polyprotein-coding regions of 50 RYMV strains. The tree was generated by the maximum likelihood (ML) method using Clustal X2. The tree was designed by using the online tool "iTOL".

3.6. Neutrality Plot Analysis

The plot of neutrality was performed, which implies the bond between GC1/2 and GC3 composition to determine the position of mutation and selection pressure that impact the codon usage bias form. To observe the association, we paradigm the plot of neutrality

between GC3 and GC1/2 for the RYMV genome. Here the plot shows that no significant association was found between GC3 and GC1/2 contents for the reason that the regression value and link were $p > 0.05$ and r = −0.06 (Figure 7). Finally, we suggested that natural selection and mutational pressure both impact the codon usage shaping of RYMV.

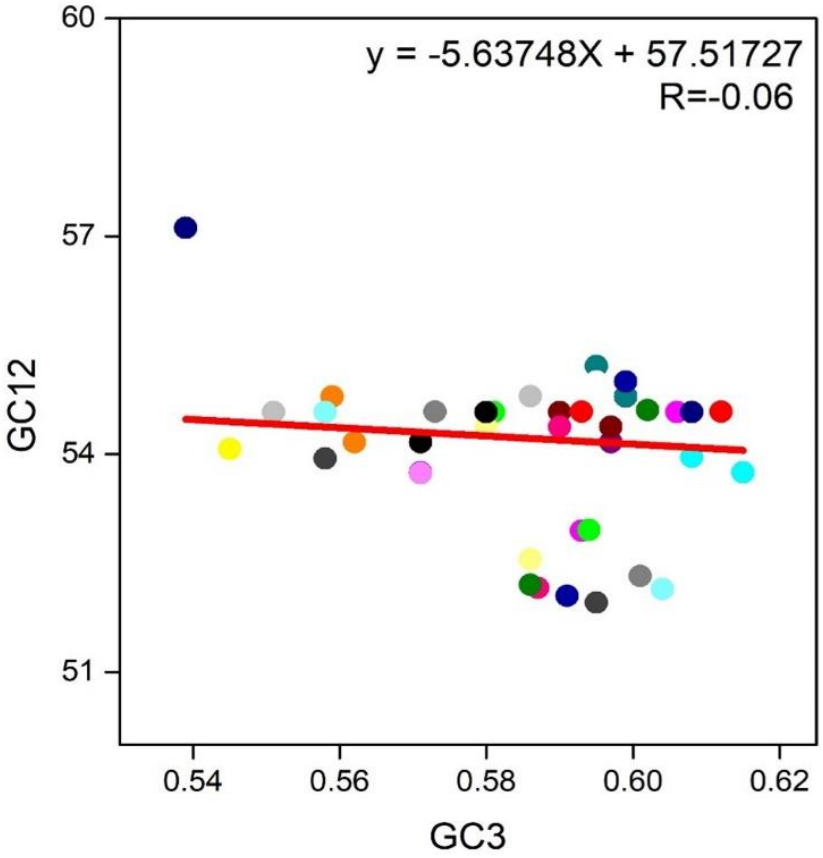


Figure 7. A Neutrality Plot between (GC3 vs. GC1,2) for the entire coding sequence of RYMV. GC1,2 represent GC at 1st and 2nd of the codon. While GC3 represents GC at 3rd codon position. And the Red solid represents the Regression analysis of GC1,2 against GC3.

3.7. Codon Usage Pattern Dominating Effects on RYMV

Here we take two factors into consideration: natural selection and mutational pressure, in order to determine codon usage bias in RYMV. Accordingly, we performed a correlation analysis between ENC and CAI values. The CAI values of the RYMV sequences seemed to have a negative relationship between the RYMV sequences' CAI values in relation to O. sativa with the ENC values that expected the effect of the RYMV codon usage pattern. Such impact was also observed among GC3s (−0.79), GC content (0.68), hydrophobicity (GRAVY) and aromaticity (AROMO) with ENC. This outcome imitates the impact of mutational pressure and natural selection on the RYMV codon usage pattern.

GRAVY and AROMO are also associated with the pattern of codon usage in different viruses [51]. It is clear from the assessment that the value of AROMO has a substantial negative relative to ENC. Likewise, GRAVY values had a significant negative association with ENC (r = −0.25, $p < 0.01$). The findings suggest that the level of protein AROMO and GRAVY are concerned with the discrepancy of codon usage in RYMV, implying the significance of selection pressure on getting the pattern of codon usage.

The study of correlation was also determined amidst the compositions of nucleotides and the first two-principles axis in RYMV genomes. Here the study revealed various significant relationships among the 2nd axes and the nucleotide compositions. The 1st axis exhibited a statistically significant association with GC3, T3, and G3, whereas the 2nd axis had a significant association with C, U, A, A3, G, G3 (r = 0.17, $p < 0.001$), T3 (r = 0.25,

$p < 0.001$), AU (r = 0.53, $p < 0.001$), GC2 (r = 0.97, $p < 0.001$), AU3 (r = 0.7, $p < 0.001$), and GC (r = 0.59, $p < 0.001$) (Figure 8). These consequences suggest that the contents of a nucleotide impact the codon usage and synonymous pattern of RYMVs.

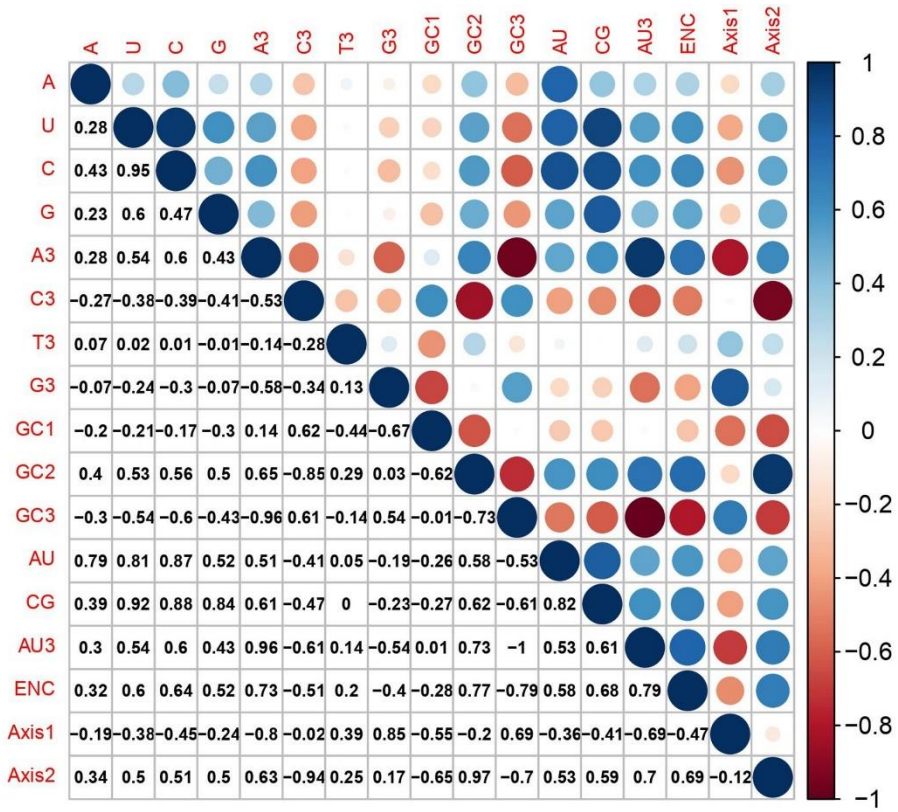


Figure 8. Correlation analysis among different nucleotides contents of RYMVs, and the first two principal axes of COA. The dark blue means positive correlation, and the dark red means negative correlation; the value larger means more significant correlation.

4. Discussion

In this study, we examined the use of synonymous codons in coding sequences from 50 RYMV genomes to comprehend how the virus evolved molecularly under the influence of host, viral, and environmental factors. It has been demonstrated in the past that the overall genomic composition can have a significant impact on codon usage bias or the preference for one type of codon over another [52]. An analysis of the nucleotide composition of RYMV genomes revealed that G and C nucleotides constitute the majority of the total nucleotide composition. This result was inconsistent with previous studies on the genomes of dicot species, i.e., cotton [53], tomato [54], *Arabidopsis thaliana* [55], peas, tobacco [56], and rice tungro bacilliform virus (RTBV) [57], wherein A and U frequencies were higher than G and C frequencies [58]. However, this finding supports the previous reports [59–62]. Although the biotic causes for increased G and decreased A are unresolved, it is essential to verify the causes of these drifts in virus-related RNA genomes [57]. The RSCU study also revealed that the genome of RYMV demonstrates higher codon usage bias toward G and C-ended codons (Table 2), agreeing with the previous reports [63,64]. Therefore, once it is proved that there is codon bias toward G and C-ended codons in RYMV genomes, we next determined the patterns of codon usage and selection of the preferred codons in RYMV genomes showed complete synergism to those of the host, *O. sativa* (Figure 2), which is inconsistent with our recent studies on RTBV, which indicated complete antagonism with the host [57,65].

CAI is considered a measure of gene expression and is useful for determining how well viral genes have adapted to their hosts. It has been suggested that genes that are

highly expressed show a larger preference for specific codons than genes that are less expressed. The CAI calculates the difference between a protein-coding sequence and a reference set of genes [63]. If the CAI score is high, then the influence of natural selection is dominant, and the codon usage bias is quite high [66]. Here, the high CAI value tendency of *O. sativa* suggests that selection pressure from *O. sativa* can impact the RYMV codon usage, and that the codon usage evolution in RYMV permits it to use the translation machinery of *O. sativa* more capably [29,67]. The outcomes of this study regarding the preferences of codon usage are comparable with the previous research that has indicated disparate *Pseudomonas fuscovaginae*, *Xanthomonas oryzae*, against *O. sativa* [68]. Here, our result was consistent with the published work [69]. The mean ENC value in RYMV coding sequences was found to be 58.55, indicating slightly biased, comparatively stable, and conserved genomic composition among various RYMV genomes. Studies have demonstrated an inverse relationship between ENC and gene expression, i.e., a lower ENC value indicates a higher codon use preference and higher gene expression [27]. Overall, it was observed that the gene expression and codon usage bias among various RYMV genomes were comparable and slightly biased. Our results were aided by the low codon usage bias that has been seen in prior studies relating to TEs genes, which are less biased (ENC: 57) than rice nuclear genes (ENC: 41) [70]. To estimate the variations in the usage of codons of AP2 genes in various species of plant, the CDS value of ENC was revealed with a mean value of 59.37, which suggested low CUB in various species of plant genome [71]. Previous studies focused on *Herbaceous Peony* and *Z. mays*, the value of ENC was also observed to be higher than 35. The analysis disclosed that low codon bias was seen along with the position of natural selection on the genes [46,72]. As a result, low codon bias within RYMV coding sequences has allowed RYMV presence in the host, despite the fact that the host's codon usage preferences differ from those of RYMV. The low codon bias of RNA viruses may be favorable for effective replication in host cells by minimizing the competition between the virus and host, each of which may have different codon preferences for synthesis machinery. Following that, an ENC plot is commonly used to determine variation in codon usage among genes in different organisms [27,73]. When the RYMVs' ENC and GC3 values were plotted, it was discovered that all the isolates clustered tightly below the curve, demonstrating the significant influence of mutation pressure on RYMV codon usage patterns as well as the influence of natural selection to some extent. Additionally, it has been previously reported that both natural selection and mutation pressure can affect the overall ENC and that this index may not be reliable for illustrating the respective contributions of selection and mutation to structuring codon usage patterns [73].

CO analysis describes the discrepancy in the usage of codons. The changes occur in the patterns of codon usage revealed through RSCU values. In the plot of CO analysis, axis 1 and 2 are the two main factors of general discrepancy [27,43,74]. Our results propose that the first axis signifies the virus strains, and the second axis signifies the countries where the virus arises. Using CO analysis, the RYMV isolates were found in clustered form (Figure 3). The mobility of RYMV's natural host, *O. sativa*, has a significant influence on the grouping of distinct RYMV lineages that are thousands of miles apart inside one cluster, as well as in the clustering of closely related lineages into different clusters. In addition, it seems that geographic diversity and related factors, including the presence of a natural host in the infection zone, climatic features, and host susceptibility, play a major part in the evolution of RYMV and in a synonymous codon usage pattern, where in the future such investigations may assist in discovering the essence of rising RYMV strains. Furthermore, present outcomes also show that more than one widespread genetic lineage was found in every infected country. A phylogenetic study was used to determine how evolutionary processes have affected the codon usage pattern of RYMV, which shows similarity to the pattern observed in COA [62]. This supports the main outcome of evolutionary processes and topographical dispersal on codon usage patterns, which further exposed the signs of recombination and genome re-assortment during single host co-infection, signifying the potential for the upcoming arrival of the novel alternates [34,35,75–77].

Natural selection and mutation pressure are two of the main influencing factors that are thought to have a significant impact on codon usage patterns [39]. The majority of the codon usage among some RNA viruses would undoubtedly be explained by a general mutation pressure that affects the entire genome [52]. Previous studies suggested that the base compositions at the 3rd positions of the codon, mutational bias is mostly explained, whilst base compositions at the 1st and 2nd positions, selective pressure were mostly validated [61,62,78]. The results of the ENC and CAI analyses made clear how natural selection and mutation pressure both affected the codon use patterns of the RYMV genes. A neutrality plot analysis was carried out to assess the relative contributions of each component to the evolution of RYMV codon usage patterns, and the results revealed that the influence of mutation pressure predominates over natural selection [29]. Furthermore, using correlation and linear regression analyses between various nucleotide compositional constraints, ENC, and principal axes, we also investigated the impact of mutation pressure on RYMV codon usage. Strong and significant connections were found, which suggests that mutation pressure has a prevailing influence. This was further confirmed by the fact that significant, strong correlations were seen when these indices were plotted using PCA against the first two principal axes. However, in the case of RYMV genomes, the involvement of factors other than mutation pressure, such as natural selection, cannot be ignored. When viruses attempt to adapt to the host cell, natural selection may result in a weak codon usage bias [48,49,79]. Moreover, a correlation study between primary axes and GRAVY and ARO also demonstrated that, although natural selection has somewhat influenced RYMV codon usage patterns, it is far weaker than mutation pressure [61,62,80].

Virus genomes in general contain a considerable proportion of repetitive sequences and microsatellites. Recombination, nucleotide content, and coding region are all influenced by the presence of a repetitive sequence and microsatellite. Whole genome sequencing, which also explores microsatellites, is warranted for further comparative genomics and evolutionary biology studies in RYMV species [81,82].

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/agronomy12091990/s1, Table S1: Detailed information of RYMV isolates studied in this study.

Author Contributions: S.U.R. and I.U.R.: Conceptualization, Methodology, Software, Data curation, Writing—Original draft preparation. M.A.K. and S.N.: Helped in write-up and editing, Validation. S.N. and S.U.: Methodology, Visualization, Validation. M.I.U.H. and M.A.K.: Reviewing-editing, Validation. A.A.A.-G., F.M.A.-H., R.M.A., W.A.E. and M.S.E.: Reviewing and Editing, Validation. All authors have read and agreed to the published version of the manuscript.

Funding: The authors would like to extend their sincere appreciation to the Researchers Supporting Project Number (RSP2022R483), King Saud University, Riyadh, Saudi Arabia.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to extend their sincere appreciation to the Researchers Supporting Project Number (RSP2022R483), King Saud University, Riyadh, Saudi Arabia.

Conflicts of Interest: All the authors declare that they have no conflict of interest.

Ethical Approval: This article does not contain any studies with human participants performed by any of the authors.

References

1. Rakotomalala, M.; Abera, B.; Rakotoarisoa, J.; Alemu, D.; Hébrard, E. Complete Genome Sequences of Rice Yellow Mottle Virus Isolates from the Federal Democratic Republic of Ethiopia. *Microbiol. Resour. Announc.* **2019**, *8*, e00589-19. [CrossRef] [PubMed]
2. Voinnet, O.; Pinto, Y.M.; Baulcombe, D.C. Suppression of Gene Silencing: A General Strategy Used by Diverse DNA and RNA Viruses of Plants. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 14147–14152. [CrossRef] [PubMed]

3. Koudamiloro, A.; Nwilene, F.E.; Togola, A.; Akogbeto, M. Insect Vectors of Rice Yellow Mottle Virus. *J. Insects* **2015**, *2015*, 721751. [CrossRef]
4. Fargette, D.; Abubakar, Z.; Traore, O.; Brugidou, C.; Choisy, M.; Se, Y.; Fauquet, C.; Konate, G. Inferring the Evolutionary History of Rice Yellow Mottle Virus from Genomic, Phylogenetic, and Phylogeographic Studies. *J. Virol.* **2004**, *78*, 3252–3261. [CrossRef]
5. Bakker, W. *Characterization and Ecological Aspects of Rice Yellow Mottle Virus in Kenya*; Centre for Agricultural Publishing and Documentatio: Wageningen, The Netherlands, 1974.
6. Kouassi, N.K.; N'Guessan, P.; Albar, L.; Fauquet, C.M.; Brugidou, C. Distribution and Characterization of Rice Yellow Mottle Virus: A Threat to African Farmers. *Plant Dis.* **2005**, *89*, 124–133. [CrossRef]
7. Konate, G.; Traore, O.; Coulibaly, M.M. Characterization of Rice Yellow Mottle Virus Isolates in Sudano-Sahelian Areas. *Arch. Virol.* **1997**, *142*, 1117–1124. [CrossRef]
8. Sarra, S.; Peters, D. Rice Yellow Mottle Virus Is Transmitted by Cows, Donkeys, and Grass Rats in Irrigated Rice Crops. *Plant Dis.* **2003**, *87*, 804–808. [CrossRef]
9. Konate, G.; Sarra, S.; Traore, O. Rice Yellow Mottle Virus Is Seed-Borne but Not Seed Transmitted in Rice Seeds. *Eur. J. Plant Pathol.* **2001**, *107*, 361–364. [CrossRef]
10. Hubert, J.; Lyimo, H.J.F.; Luzi-kihupi, A. Geographical Variation, Distribution and Diversity of Rice Yellow Mottle Virus Phylotypes in Tanzania. *Am. J. Plant Sci.* **2017**, *8*, 1264–1284. [CrossRef]
11. Ayaka, U.; Naswiru, T.; Keiko, T.N.; Nobuhito, S.; Nobuaki, O. Characterization of Rice Yellow Mottle Virus in North-Eastern Tanzania. *J. Agric. Sci.* **2015**, *60*, 116–126.
12. N'Guessan, P.; Pinel, A.; Caruana, M.L.; Frutos, R.; Sy, A.; Ghesquière, A.; Fargette, D. Evidence of the Presence of Two Serotypes of Rice Yellow Mottle Sobemovirus in Côte d'Ivoire. *Eur. J. Plant Pathol.* **2000**, *106*, 167–178. [CrossRef]
13. Rossel, H.W.; Ayotade, K.A.; Thottapilly, G.; Adeoti, A.A.; Alluri, K.; Alam, M.S.; Zan, K. A new record of rice yellow mottle virus disease in Badeggi, Nigeria. *Int. Rice Comm. Newsl.* **1982**, *31*, 23–24.
14. Fauquet, C.; Thouvenel, J.C. Isolation of the rice yellow mottle virus in Ivory Coast. *Plant Dis. Report.* **1977**, *61*, 4–7.
15. Raymundo, S.A.; Konteh, I. Distribution, importance, screening methods and varietal reaction to rice pale yellow mottle disease [in West Africa]. *Int. Rice Comm. Newsl.* **1980**, *29*, 51–53.
16. John, V.T.; Thottappilly Ibadan, G.; Awoderu, V.A. Occurrence of Rice Yellow Mottle Virus in Some Sahelian Countries in West Africa. *FAO Plant Prot. Bull.* **1984**, *32*, 86–87.
17. Reckhaus, P.M.; Andriamasintseheno, H.F. Rice Yellow Mottle Virus in Madagascar and Its Epidemiology in the Northwest of the Island/Rice Yellow Mottle Virus in Madagaskar Und Seine Epidemiologie Im Nordwesten Der Insel. *Z. Pflanzenkrankh. Pflanzenschutz/J. Plant Dis. Prot.* **1997**, *104*, 289–295.
18. Traoré, O.; Pinel, A.; Fargette, D.; Konaté, G. First Report and Characterization of Rice Yellow Mottle Virus in Central Africa. *Plant Dis.* **2001**, *85*, 920. [CrossRef] [PubMed]
19. Awoderu, V.A. The Rice Yellow Mottle Virus Situation in West Africa. *J. Basic Microbiol.* **1991**, *31*, 91–99. [CrossRef]
20. Asante, M.D.; Amadu, B.; Traore, V.S.E.; Oppong, A.; Adebayo, M.A.; Aculey, P.; Marfo, E.A.; Kang, K.-H. Assessment of Korean Rice Lines for Their Reaction to Rice Yellow Mottle Virus in Ghana. *Heliyon* **2020**, *6*, e05551. [CrossRef]
21. Butt, A.M.; Nasrullah, I.; Qamar, R.; Tong, Y. Evolution of Codon Usage in Zika Virus Genomes Is Host and Vector Speci Fi C. *Emerg. Microbes Infect.* **2016**, *5*, 1–14. [CrossRef]
22. Hershberg, R.; Petrov, D.A. Selection on Codon Bias. TL-42. *Annu. Rev. Genet.* **2008**, *42*, 287–299. [CrossRef] [PubMed]
23. Andersson, S.G.; Kurland, C.G. Codon Preferences in Free-Living Microorganisms. *Microbiol. Rev.* **1990**, *54*, 198–210. [CrossRef] [PubMed]
24. Cristina, J.; Moreno, P.; Moratorio, G.; Musto, H. Genome-Wide Analysis of Codon Usage Bias in Ebolavirus. *Virus Res.* **2015**, *196*, 87–93. [CrossRef] [PubMed]
25. Rahman, S.U.; Mao, Y.; Tao, S. Codon Usage Bias and Evolutionary Analyses of Zika Virus Genomes. *Genes Genom.* **2017**, *39*, 855–866. [CrossRef]
26. Sharp, P.M.; Li, W.H. Codon Usage in Regulatory Genes in Escherichia Coli Does Not Reflect Selection for "rare" Codons. *Nucleic Acids Res.* **1986**, *14*, 7737–7749. [CrossRef]
27. Butt, A.M.; Nasrullah, I.; Tong, Y. Genome-Wide Analysis of Codon Usage and Influencing Factors in Chikungunya Viruses. *PLoS ONE* **2014**, *9*, e90905. [CrossRef]
28. Sharp, P.M.; Emery, L.R.; Zeng, K. Forces That Influence the Evolution of Codon Bias. *Philos. Trans. R. Soc. B* **2010**, *365*, 1203–1212. [CrossRef]
29. Rahman, S.U.; Abdullah, M.; Khan, A.W.; Haq, M.I.U.; Haq, N.U.; Aziz, A.; Tao, S. A Detailed Comparative Analysis of Codon Usage Bias in Alongshan Virus. *Virus Res.* **2022**, *308*, 198646. [CrossRef]
30. Mueller, S.; Papamichail, D.; Coleman, J.R.; Skiena, S.; Wimmer, E. Reduction of the Rate of Poliovirus Protein Synthesis through Large-Scale Codon Deoptimization Causes Attenuation of Viral Virulence by Lowering Specific Infectivity. *J. Virol.* **2006**, *80*, 9687–9696. [CrossRef]
31. Costafreda, M.I.; Pérez-Rodriguez, F.J.; D'Andrea, L.; Guix, S.; Ribes, E.; Bosch, A.; Pintó, R.M. Hepatitis A Virus Adaptation to Cellular Shutoff Is Driven by Dynamic Adjustments of Codon Usage and Results in the Selection of Populations with Altered Capsids. *J. Virol.* **2014**, *88*, 5029–5041. [CrossRef]

32. Burns, C.C.; Shaw, J.; Campagnoli, R.; Jorba, J.; Vincent, A.; Quay, J.; Kew, O. Modulation of Poliovirus Replicative Fitness in HeLa Cells by Deoptimization of Synonymous Codon Usage in the Capsid Region. *J. Virol.* **2006**, *80*, 3259–3272. [CrossRef] [PubMed]
33. Karumathil, S.; Raveendran, N.T.; Ganesh, D.; Kumar Ns, S.; Nair, R.R.; Dirisala, V.R. Evolution of Synonymous Codon Usage Bias in West African and Central African Strains of Monkeypox Virus. *Evol. Bioinform. Online* **2018**, *14*, 1176934318761368. [CrossRef] [PubMed]
34. Odongo, P.J.; Onaga, G.; Ricardo, O.; Natsuaki, K.T.; Alicai, T.; Geuten, K. Insights Into Natural Genetic Resistance to Rice Yellow Mottle Virus and Implications on Breeding for Durable Resistance. *Front. Plant Sci.* **2021**, *12*, 671355. [CrossRef] [PubMed]
35. Pidon, H.; Chéron, S.; Ghesquière, A.; Albar, L. Allele Mining Unlocks the Identification of RYMV Resistance Genes and Alleles in African Cultivated Rice. *BMC Plant Biol.* **2020**, *20*, 222. [CrossRef]
36. Yao, X.; Fan, Q.; Yao, B.; Lu, P.; Rahman, S.U.; Chen, D.; Tao, S. Codon Usage Bias Analysis of Bluetongue Virus Causing Livestock Infection. *Front. Microbiol.* **2020**, *11*, 655. [CrossRef]
37. Wong, E.H.M.; Smith, D.K.; Rabadan, R.; Peiris, M.; Poon, L.L.M. Codon Usage Bias and the Evolution of Influenza A Viruses. Codon Usage Biases of Influenza Virus. *BMC Evol. Biol.* **2010**, *10*, 253. [CrossRef]
38. Liu, J.; Zhu, D.; Ma, G.; Liu, M.; Wang, M.; Jia, R.; Chen, S.; Sun, K.; Yang, Q.; Wu, Y.; et al. Genome-Wide Analysis of the Synonymous Codon Usage Patterns in Riemerella Anatipestifer. *Int. J. Mol. Sci.* **2016**, *17*, 1304. [CrossRef]
39. Chakraborty, A.; Uechi, T.; Higa, S.; Torihara, H.; Kenmochi, N. Loss of Ribosomal Protein L11 Affects Zebrafish Embryonic Development through a P53-Dependent Apoptotic Response. *PLoS ONE* **2009**, *4*, e4152. [CrossRef]
40. Puigbò, P.; Bravo, I.G.; Garcia-Vallve, S. CAIcal: A Combined Set of Tools to Assess Codon Usage Adaptation. *Biol. Direct* **2008**, *3*, 38. [CrossRef]
41. Wright, F. The "effective Number of Codons" Used in a Gene. *Gene* **1990**, *87*, 23–29. [CrossRef]
42. Cristina, J.; Fajardo, A.; Sonora, M.; Moratorio, G.; Musto, H. A Detailed Comparative Analysis of Codon Usage Bias in Zika Virus. *Virus Res.* **2016**, *223*, 147–152. [CrossRef] [PubMed]
43. Greenacre, M. *Theory and Applications of Correspondence Analysis*; Academic Press: London, UK, 1984; ISBN 0122990501.
44. Guan, D.L.; Ma, L.B.; Khan, M.S.; Zhang, X.X.; Xu, S.Q.; Xie, J.Y. Analysis of Codon Usage Patterns in Hirudinaria Manillensis Reveals a Preference for GC-Ending Codons Caused by Dominant Selection Constraints. *BMC Genom.* **2018**, *19*, 542. [CrossRef] [PubMed]
45. Sueoka, N. Directional Mutation Pressure and Neutral Molecular Evolution. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 2653–2657. [CrossRef] [PubMed]
46. Wu, Y.; Zhao, D.; Tao, J. Analysis of Codon Usage Patterns in Herbaceous Peony (*Paeonia Lactiflora* Pall.) Based on Transcriptome Data. *Genes* **2015**, *6*, 1125–1139. [CrossRef]
47. Singh, N.K.; Tyagi, A. A Detailed Analysis of Codon Usage Patterns and Influencing Factors in Zika Virus. *Arch. Virol.* **2017**, *162*, 1963–1973. [CrossRef]
48. Hassan, S.; Mahalingam, V.; Kumar, V. Synonymous Codon Usage Analysis of Thirty Two Mycobacteriophage Genomes. *Adv. Bioinform.* **2009**, *2009*, 316936. [CrossRef]
49. Sharp, P.M.; Li, W.H. An Evolutionary Perspective on Synonymous Codon Usage in Unicellular Organisms. *J. Mol. Evol.* **1986**, *24*, 28–38. [CrossRef]
50. Rahman, S.U.; Yao, X.; Li, X.; Chen, D.; Tao, S. Analysis of Codon Usage Bias of Crimean-Congo Hemorrhagic Fever Virus and Its Adaptation to Hosts. *Infect. Genet. Evol.* **2018**, *58*, 1–16. [CrossRef]
51. Wang, Z.; Cao, R.; Taylor, K.; Briley, A.; Caldwell, C.; Cheng, J. The Properties of Genome Conformation and Spatial Gene Interaction and Regulation Networks of Normal and Malignant Human Cell Types. *PLoS ONE* **2013**, *8*, e58793. [CrossRef]
52. Jenkins, G.M.; Holmes, E.C. The Extent of Codon Usage Bias in Human RNA Viruses and Its Evolutionary Origin. *Virus Res.* **2003**, *92*, 1–7. [CrossRef]
53. Wang, L.; Xing, H.; Yuan, Y.; Wang, X.; Saeed, M.; Tao, J.; Feng, W.; Zhang, G.; Song, X.; Sun, X. Genome-Wide Analysis of Codon Usage Bias in Four Sequenced Cotton Species. *PLoS ONE* **2018**, *13*, e0194372. [CrossRef] [PubMed]
54. Sablok, G.; Wu, X.; Kuo, J.; Nayak, K.C.; Baev, V.; Varotto, C.; Zhou, F. Combinational Effect of Mutational Bias and Translational Selection for Translation Efficiency in Tomato (Solanum Lycopersicum) Cv. Micro-Tom. *Genomics* **2013**, *101*, 290–295. [CrossRef] [PubMed]
55. Ossowski, S.; Schneeberger, K.; Lucas-Lledó, J.I.; Warthmann, N.; Clark, R.M.; Shaw, R.G.; Weigel, D.; Lynch, M. The Rate and Molecular Spectrum of Spontaneous Mutations in Arabidopsis Thaliana. *Science* **2010**, *327*, 92–94. [CrossRef] [PubMed]
56. Kawabe, A.; Miyashita, N.T. Patterns of Codon Usage Bias in Three Dicot and Four Monocot Plant Species. *Genes Genet. Syst.* **2003**, *78*, 343–352. [CrossRef] [PubMed]
57. Nguyen, T.H.; Wang, D.; Rahman, S.U.; Bai, H.; Yao, X.; Chen, D.; Tao, S. Analysis of Codon Usage Patterns and Influencing Factors in Rice Tungro Bacilliform Virus. *Infect. Genet. Evol.* **2021**, *90*, 104750. [CrossRef]
58. Anwar, A.M.; Aljabri, M.; El-soda, M. Patterns of Genome-Wide Codon Usage Bias in Tobacco, Tomato and Potato. *Biotechnol. Biotechnol. Equip.* **2021**, *35*, 657–664. [CrossRef]
59. Wei, L.; He, J.; Jia, X.; Qi, Q.; Liang, Z.; Zheng, H.; Ping, Y.; Liu, S.; Sun, J. Analysis of Codon Usage Bias of Mitochondrial Genome in Bombyx Moriand Its Relation to Evolution. *BMC Evol. Biol.* **2014**, *14*, 262. [CrossRef]
60. Li, N.; Li, Y.; Zheng, C.; Huang, J.; Zhang, S. Genome-Wide Comparative Analysis of the Codon Usage Patterns in Plants. *Genes Genom.* **2016**, *38*, 723–731. [CrossRef]

61. Cardinale, D.J.; Derosa, K.; Duffy, S. Base Composition and Translational Selection Are Insufficient to Explain Codon Usage Bias in Plant Viruses. *Viruses* **2013**, *5*, 162–181. [CrossRef]
62. 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]
63. Song, H.; Liu, J.; Song, Q.; Zhang, Q.; Tian, P.; Nan, Z. Comprehensive Analysis of Codon Usage Bias in Seven Epichloë Species and Their Peramine-Coding Genes. *Front. Microbiol.* **2017**, *8*, 1419. [CrossRef] [PubMed]
64. van Hemert, F.; Berkhout, B. Nucleotide Composition of the Zika Virus RNA Genome and Its Codon Usage. *Virol. J.* **2016**, *13*, 95. [CrossRef]
65. Yang, C.; Zhao, Q.; Wang, Y.; Zhao, J.; Qiao, L.; Wu, B.; Yan, S.; Zheng, J.; Zheng, X. Comparative Analysis of Genomic and Transcriptome Sequences Reveals Divergent Patterns of Codon Bias in Wheat and Its Ancestor Species. *Front. Genet.* **2021**, *12*, 732432. [CrossRef] [PubMed]
66. Liu, H.; He, R.; Zhang, H.; Huang, Y.; Tian, M.; Zhang, J. Analysis of Synonymous Codon Usage in Zea Mays. *Mol. Biol. Rep.* **2010**, *37*, 677–684. [CrossRef] [PubMed]
67. Subramanian, A.; Rup Sarkar, R. Data in Support of Large Scale Comparative Codon Usage Analysis in Leishmania and Trypanosomatids. *Data Br.* **2015**, *4*, 269–272. [CrossRef]
68. Vasanthi, S.; Dass, J.F.P. Comparative Genome-Wide Analysis of Codon Usage of Different Bacterial Species Infecting Oryza Sativa. *J. Cell. Biochem.* **2018**, *119*, 9346–9356. [CrossRef] [PubMed]
69. Pan, X.; Welti, R.; Wang, X. Quantitative Analysis of Major Plant Hormones in Crude Plant Extracts by High-Performance Liquid Chromatography–Mass Spectrometry. *Nat. Protoc.* **2010**, *5*, 986–992. [CrossRef]
70. Jia, J.; Xue, Q. Codon Usage Biases of Transposable Elements and Host Nuclear Genes in Arabidopsis Thaliana and Oryza Sativa. *Genom. Proteom. Bioinform.* **2009**, *7*, 175–184. [CrossRef]
71. Qing, Y.; Cheng, L.; Li, R.; Liu, G.; Zhang, Y.; Tang, X.; Wang, J.; Liu, H.; Qin, Y. Potential Antibacterial Mechanism of Silver Nanoparticles and the Optimization of Orthopedic Implants by Advanced Modification Technologies. *Int. J. Nanomed.* **2018**, *13*, 3311–3327. [CrossRef]
72. Ma, Y.P.; Liu, Z.X.; Hao, L.; Ma, J.Y.; Liang, Z.L.; Li, Y.G.; Ke, H. Analysing Codon Usage Bias of Cyprinid Herpesvirus 3 and Adaptation of This Virus to the Hosts. *J. Fish Dis.* **2015**, *38*, 665–673. [CrossRef]
73. Nasrullah, I.; Butt, A.M.; Tahir, S.; Idrees, M.; Tong, Y. Genomic Analysis of Codon Usage Shows Influence of Mutation Pressure, Natural Selection, and Host Features on Marburg Virus Evolution. *BMC Evol. Biol.* **2015**, *15*, 174. [CrossRef] [PubMed]
74. Kumar, N.; Bera, B.C.; Greenbaum, B.D.; Bhatia, S.; Sood, R.; Selvaraj, P.; Anand, T.; Tripathi, B.N.; Virmani, N. Revelation of Influencing Factors in Overall Codon Usage Bias of Equine Influenza Viruses. *PLoS ONE* **2016**, *11*, e0154376. [CrossRef] [PubMed]
75. Chamberlain, J.; Cook, N.; Lloyd, G.; Mioulet, V.; Tolley, H.; Hewson, R. Co-Evolutionary Patterns of Variation in Small and Large RNA Segments of Crimean-Congo Hemorrhagic Fever Virus. *J. Gen. Virol.* **2005**, *86*, 3337–3341. [CrossRef] [PubMed]
76. Zhang, Y.; Shen, Z.; Meng, X.; Zhang, L.; Liu, Z.; Liu, M.; Zhang, F.; Zhao, J. Codon Usage Patterns across Seven Rosales Species. *BMC Plant Biol.* **2022**, *22*, 65. [CrossRef]
77. Hewson, R.; Gmyl, A.; Gmyl, L.; Smirnova, S.E.; Karganova, G.; Jamil, B.; Hasan, R.; Chamberlain, J.; Clegg, C. Evidence of Segment Reassortment in Crimean-Congo Haemorrhagic Fever Virus. *J. Gen. Virol.* **2004**, *85*, 3059–3070. [CrossRef]
78. Cheng, X.; Wu, X.; Wang, H.; Sun, Y.; Qian, Y.; Luo, L. High Codon Adaptation in Citrus Tristeza Virus to Its Citrus Host. *Virol. J.* **2012**, *9*, 15–19. [CrossRef]
79. Gupta, S.K.; Bhattacharyya, T.K.; Ghosh, T.C. Synonymous Codon Usage in Lactococcus Lactis: Mutational Bias versus Translational Selection. *J. Biomol. Struct. Dyn.* **2004**, *21*, 527–536. [CrossRef]
80. 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]
81. George, B.; Alam, C.M.; Kumar, R.V.; Gnanasekaran, P.; Chakraborty, S. Potential Linkage between Compound Microsatellites and Recombination in Geminiviruses: Evidence from Comparative Analysis. *Virology* **2015**, *482*, 41–50. [CrossRef]
82. George, B.; Gnanasekaran, P.; Jain, S.K.; Chakraborty, S. Genome Wide Survey and Analysis of Small Repetitive Sequences in Caulimoviruses. *Infect. Genet. Evol.* **2014**, *27*, 15–24. [CrossRef]