



## Article

# Adaptation and Codon-Usage Preference of Apple and Pear-Infecting Apple Stem Grooving Viruses

Jaedeok Kim <sup>1,2</sup>, Aamir Lal <sup>1</sup>, Eui-Joon Kil <sup>1,3</sup> , Hae-Ryun Kwak <sup>4</sup>, Hwan-Su Yoon <sup>5</sup> , Hong-Soo Choi <sup>4</sup>, Mikyeong Kim <sup>4,6,\*</sup> , Muhammad Ali <sup>7,\*</sup> and Sukchan Lee <sup>1,\*</sup>

<sup>1</sup> Department of Integrative Biotechnology, Sungkyunkwan University, Suwon 16419, Korea; elhanan06@korea.kr (J.K.); aamirchaudhary43@gmail.com (A.L.); viruskil@anu.ac.kr (E.-J.K.)

<sup>2</sup> Incheon International Airport Regional Office, Animal and Plant Quarantine Agency, Seoul 22382, Korea

<sup>3</sup> Department of Plant Medicals, Andong National University, Andong 36729, Korea

<sup>4</sup> Crop Protection Division, National Institute of Agricultural Sciences, Rural Development Administration, Wanju 55365, Korea; hrkwakhahn@korea.kr (H.-R.K.); hschoi@korea.kr (H.-S.C.)

<sup>5</sup> Department of Biological Science, Sungkyunkwan University, Suwon 16419, Korea; hsyoon2011@skku.edu

<sup>6</sup> College of Agriculture, Life and Environment Sciences, Chungbuk National University, Cheongju 28644, Korea

<sup>7</sup> Department of Life Sciences, School of Science, University of Management and Technology (UMT), Johar Town, Lahore 54770, Pakistan

\* Correspondence: mkim00@chungbuk.ac.kr (M.K.); muhammad\_ali@umt.edu.pk (M.A.); cell4u@skku.edu (S.L.); Tel.: +82-43-261-2509 (M.K.); +92-312-9959558 (M.A.); +82-31-290-7866 (S.L.); Fax: +82-43-271-4414 (M.K.); +82-31-290-7892 (S.L.)

**Abstract:** Apple stem grooving virus (ASGV; genus *Capillovirus*) is an economically important virus. It has an approx. 6.5 kb, monopartite, linear, positive-sense, single-stranded RNA genome. The present study includes identification of 24 isolates—13 isolates from apple (*Pyrus malus* L.) and 11 isolates from pear (*Pyrus communis* L.)—from different agricultural fields in South Korea. The coat protein (CP) gene of the corresponding 23 isolates were amplified, sequenced, and analyzed. The CP sequences showed phylogenetic separation based on their host species, and not on the geography, indicating host adaptation. Further analysis showed that the ASGV isolated in this study followed host adaptation influenced and preferred by the host codon-usage.

**Keywords:** ASGV; coat protein; host adaptation; RNA virus; virus evolution



**Citation:** Kim, J.; Lal, A.; Kil, E.-J.; Kwak, H.-R.; Yoon, H.-S.; Choi, H.-S.; Kim, M.; Ali, M.; Lee, S. Adaptation and Codon-Usage Preference of Apple and Pear-Infecting Apple Stem Grooving Viruses. *Microorganisms* **2021**, *9*, 1111. <https://doi.org/10.3390/microorganisms9061111>

Academic Editor: Elvira Fiallo-Olivé

Received: 27 February 2021

Accepted: 5 May 2021

Published: 21 May 2021

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



**Copyright:** © 2021 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

An understanding of the molecular evolution of viruses is essential to design knowledge-based control strategies. Relatively fewer studies are available on plant viruses covering sequence origin, mutation rates, selection pressure, reassortment, and recombination [1]. Previous studies suggest point-mutations, recombination, and reassortments as major driving forces in evolution for the natural selection. The mechanisms of adaptation and selections contributing a virus fitness are yet to be explored exactly [2].

The adaptation of viruses to novel hosts is a general evolutionary phenomenon of invasion and adaptation to a new niche. The new host may present challenges at the level of viral entry into cells, replication, and transmission/exit. Similarly, the new host may encode different/stringent defense mechanisms, which the newly invading virus has to overcome. Only a small minority of the initial pool of viral genotypes may survive these hurdles and that subsequent adaptation will likely lead to improved adaptation [3].

Parallel evolution, i.e., independent evolution of similar or identical features in distinct lineages subject to similar selection pressures [4], were reported extensively in both natural and experimental populations of microbes—most often viruses [4–6], bacteria [7], yeast [8], and protozoa [9]. Differences in the founding genotypes may result in divergent evolutionary trajectories [10,11]. It is, therefore, important to know the dynamics of evolution that could be predicted through parallel evolution.

*Apple stem grooving virus* (type species, genus *Capillovirus*) is an economically important virus of pome fruits and citrus, particularly plant species of *Malus* and *Pyrus* [12]. A virion of apple stem grooving virus (ASGV) contains monopartite, linear single-strand positive-sense 6.7–7.4 kb RNA genome. As host plants are often asymptomatic, the virus survives long enough, unnoticed, to adapt to woody plant hosts.

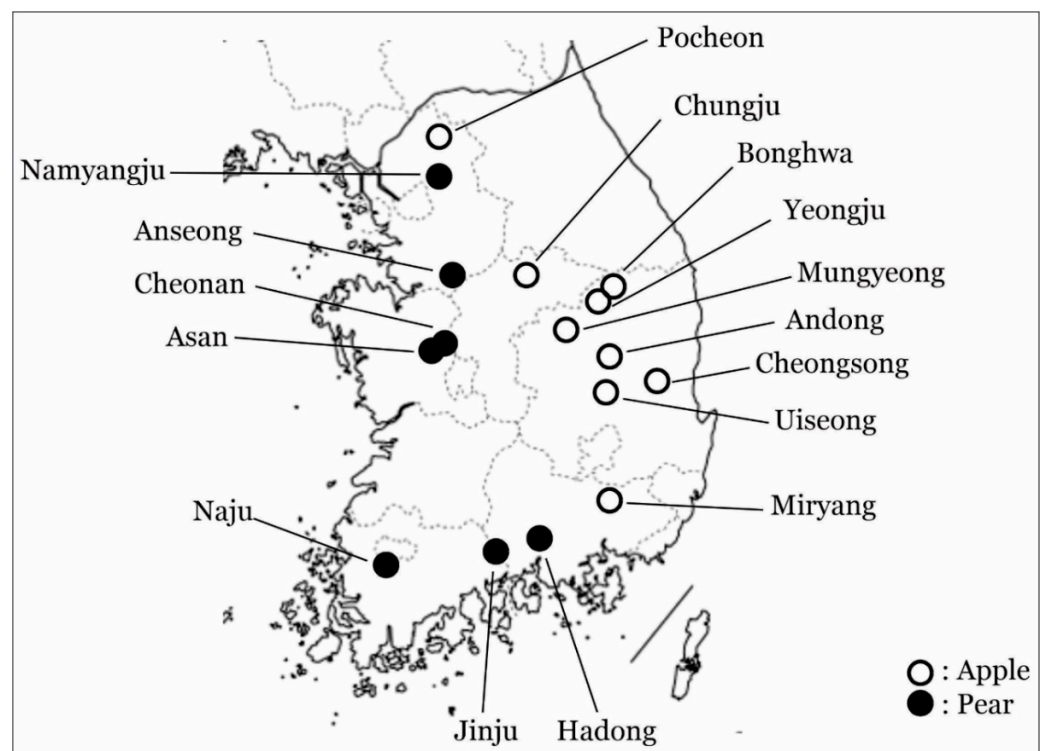
Although ASGV is latent in most commercial rootstock–scion combinations, serious disease burden is evident in apples grown on *Malus sieboldii* rootstocks in Japan [13]. In pear, ASGV claimed black necrotic leaf spot disease where a decline up to 50% has been estimated [14].

Looking into the importance of the virus, we assessed the dynamics of evolution to examine the molecular basis of adaptation in a viral population under host-derived natural selection in apple and pear plants.

## 2. Materials and Methods

### 2.1. Sample Collection

Apple and pear plant samples were collected from major agricultural areas in South Korea (Figure 1) in the years 2008 and 2009. Petal and leaf tissues were collected from randomly chosen trees in each field. All collected tissues were stored at  $-80\text{ }^{\circ}\text{C}$  until plant total RNA was extracted using the Easy-Spin™ (Intron Co., Seongnam, Korea) RNA extraction kit following manufacturer’s instructions.



**Figure 1.** Geography of plant samples collected in this study for apple stem grooving virus amplification.

### 2.2. cDNA Amplification

Primers for the ASGV coat protein (CP) gene were designed as: *forward primer* 5'-AGG CWA TCA GCG GCT GTG-3' and *reverse primer* 5'-AGA GTG GAC AAA CTC TAG ACT CTA GAA A-3'. A reference sequence (GenBank AB004063) was used to design the primers. The cDNA synthesis was carried out in a 5  $\mu\text{L}$  reaction mixture. Briefly, 0.1 ng to 5  $\mu\text{g}$  plant total RNA, 20 pmol reverse primer was taken. The volume was made up to 3  $\mu\text{L}$  with DEPC-treated water. The mixture was incubated for 5 min at  $65\text{ }^{\circ}\text{C}$ , followed by chilling on ice. Before incubating reaction mixture for 60 min at  $42\text{ }^{\circ}\text{C}$ ,  $1\times$  reaction buffer

(Promega Co., Madison, WI, USA), 1 mM final concentration of dNTP mix and 1 unit of AMV Reverse Transcriptase was added. The product of reverse transcription was directly used in downstream PCR or stored at  $-20^{\circ}\text{C}$ .

A final volume of 25  $\mu\text{L}$  reaction mixture—comprised of  $1\times$  Taq buffer, 200  $\mu\text{M}$  dNTPs mix, 1.5 mM  $\text{MgCl}_2$ , 10 pM of each forward and reverse primer, 1 unit of Taq DNA Polymerase (Cat# M7660, Promega Co.), and 5  $\mu\text{L}$  of reverse transcription product—were incubated in a thermal cycler. The incubation program was executed as—denaturation for 3 min at  $94^{\circ}\text{C}$ , followed by amplification for 35 cycles ( $94^{\circ}\text{C}$  for 30 s,  $55^{\circ}\text{C}$  for 30 s, and  $72^{\circ}\text{C}$  accordingly as 1 min per kb amplicon). The amplification was followed with a final extension of 10 min at  $72^{\circ}\text{C}$ . The PCR product was visualized on 1% agarose gel.

### 2.3. Phylogenetic Analysis

Amplified PCR amplicons were sequenced in both orientations, assembled and analyzed. The sequences were then submitted to the GenBank (Table S1). The ASGV sequences identified in this study were aligned with the previously available sequences in the databases. The evolutionary history was inferred from the nucleotide sequences using the MUSCLE method based on K2+G model (Figure 4). The bootstrap consensus tree was inferred using 100 replicates representing the evolutionary history. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site.

Nucleotide sequences were translated into amino acid sequences. The evolutionary history of the amino acid sequences was inferred using the MUSCLE method, based on the JTT+G best-fit-model (Figure 5). The bootstrap consensus tree was inferred from 100 replicates representing the evolutionary history. All other options followed as described above.

### 2.4. Protein Structure Prediction

Multi-aligned amino acids sequences were analyzed using Geneious Pro 8 (Biomatters Co., Auckland, New Zealand). Secondary structures were predicted using the original Garnier Osguthorpe Robson algorithm (GOR I) provided by the EMBOSS suite embedded in Geneious Pro 8. Tertiary structures were predicted and modeled using the i-TASSER server [15].

### 2.5. Codon-by-Codon Behavior Analysis for Synonymous and Non-Synonymous Substitutions

Among the sequences identified in this study, only fully characterized ASGV CP sequences were selected. These sequences were analyzed to determine the cumulative behavior of sequence variation in a codon-by-codon manner with the synonymous and non-synonymous analysis program [16].

### 2.6. Codon-by-Codon Maximum Likelihood Analysis of Natural Selection

For each codon, estimates of inferred synonymous (s) and nonsynonymous (n) substitutions were presented along with the numbers of sites that were estimated to be synonymous (S) and nonsynonymous (N). These estimates were produced using joint maximum likelihood reconstructions of ancestral states under a Muse-Gaut model of codon substitution and Tamura-Nei model of nucleotide substitution. For estimating ML values, a tree topology was automatically computed. The test statistic  $dN-dS$  was used to detect codons that had undergone positive selection; where,  $dS$  is the number of synonymous substitutions per site ( $s/S$ ) and  $dN$  is the number of nonsynonymous substitutions per site ( $n/N$ ). The analysis involved 25 Korean ASGV isolates. The codon positions included were 1st + 2nd + 3rd + Noncoding. All positions containing gaps and missing data were eliminated. There were 675 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 [17].

### 2.7. Preferred Codon Usage of ASGV CP Gene

Codon usage of ASGV CP genes was analyzed using Geneious Pro 8 with multi-aligned nucleotide and the corresponding amino acid sequences. Codon usage of ASGV

was analyzed in two ways. First, a correspondence analysis for codon usage was performed for all Chinese and Korean isolates. Second, codon usage frequencies among apple and pear isolates were processed on the codon usage database and codon usage analyzer [18].

### 3. Results

A total of 24 amplicons were sequenced in both orientations. Sequencing results confirmed ASGV sequences. The sequences were 710–711 nucleotides long covering complete coat protein gene coding sequence (CDS) except the stop codon (i.e., TAG and, in some cases, a T before the stop codon). The sequences were then submitted to the GenBank (JN792471 to JN792494).

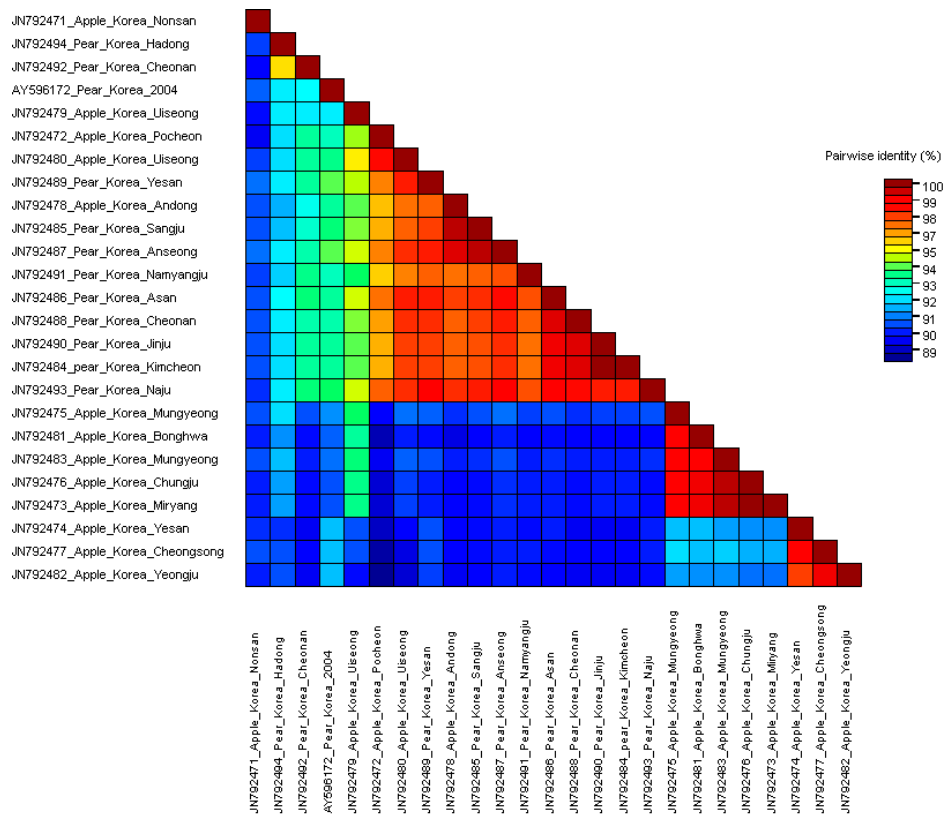
#### 3.1. Phylogenetic Analysis of ASGV Showed Host-Specificity

To determine ASGV host-specificity, 56 CP sequences, including 24 isolates identified in this study (i.e., 13 isolates from apple, 11 isolates from pear) and a former Korean isolate (GenBank AY596172) from pear, identified prior to this study, were used. Phylogenetic analysis based on nucleotide sequences revealed clustering of clear distinct patterns corresponding to their isolation hosts. Additionally, intermediate groups were also observed (Figure 4).

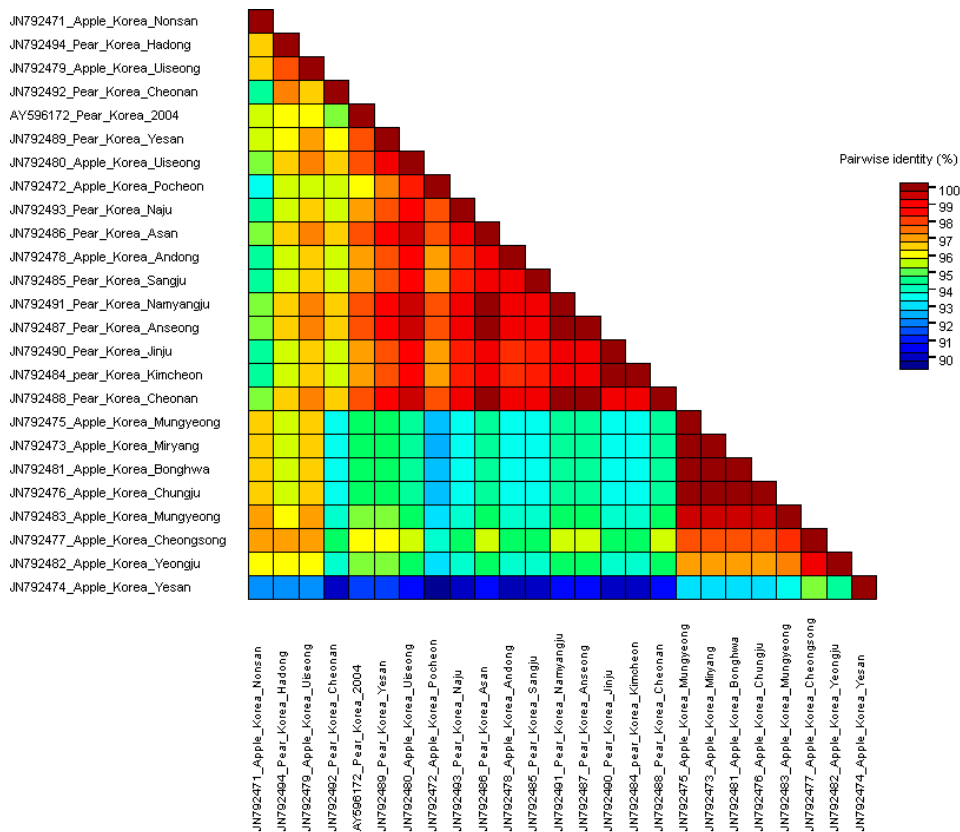
Most ASGV isolates worldwide were categorized in clade 4. Countries from which the isolates were reported are the Latvia, Serbia, and Turkey in Europe; China, India, and Korea in Asia; and Brazil in the Americas. Regardless of the location, ASGV were easily distinguished by their isolation hosts. Similar to the clade 4, ASGV isolates from China in intermediate<sup>2</sup> was also separated to their geography.

Interestingly, clade 3 contains Korean isolates identified on apple trees. Along with this, the clade 3 contains a Japanese isolate identified on apple trees. Most importantly, all isolates collected from pear trees from the distant regions were categorized in intermediate clade 1, referred to as the Korean pear group, and the clade 2. These distant regions included Anseong (GenBank JN792487), Asan (GenBank JN792486), Cheonan (GenBank JN792488), Kimcheon (GenBank JN792484), Naju (GenBank JN792493), Namyangju (GenBank JN792491), and Sangju (GenBank JN792485). To conclude, the distinctly located isolates showed high identities based on their isolation host—e.g., the two pear isolates Cheonan (GenBank JN792492) and Hadong (GenBank JN792494) showed 95.6% nucleotide identities and clustered in the clade 2 (Figures 2, 4 and S1).

The isolates from the pear group exhibited significant differences in nucleotide sequence identity up to 10% against the apple isolates (Figure 6). Four apple isolates cluster into clade 1, GenBank JN792472 (Pochen), JN792479 (Uiseong), and JN792480 (Uiseong) (Figure 4). Although the pear isolates were collected from far flung distant areas, the pear isolates showed close relationship with each other and were distinguished from the apple isolates (Figures 1 and 4). Isolates from apple and pear identified from the same or nearby regions did not show much relevance. For example, the accession with GenBank JN792492 from Yeongju in pear and JN792478 from Andong in apple were collected from nearby regions were clustered in distinct clades 2 and intermediate<sup>1</sup> clade, respectively (Figures 4 and 5). Similarly, the two accessions from Yesan (GenBank JN792474 (apple) and JN792489 (pear)) shared only 90.0% nucleotide and 91.5% amino acid identities with each other (Figures S1 and S2). The distribution of the Yesan isolates into distinct clades, i.e., clade 4 and intermediate<sup>1</sup> clade (Figures 4 and 5) indicates the isolation host specificity (also evident from Figures 2 and 3).



**Figure 2.** Apple stem grooving virus (ASGV) coat protein gene nucleotide sequence identities. MUSCLE alignment was used implemented in Sequence demarcation tool (SDTV1.2).



**Figure 3.** Apple stem grooving virus (ASGV) coat protein amino acid sequence identities. MUSCLE alignment was used implemented in Sequence demarcation tool (SDTV1.2).

The CP sequences were translated to amino acid sequences and aligned using the MUSCLE algorithm. The phylogenetic analysis showed that the amino acids sequences were similar to the nucleotide sequences in context with host relevance (Figure 5). Contrary to the nucleotide sequence tree, the amino acid sequence tree showed low bootstrap support, which may mean that the amino acid sequence variation was lower than the nucleotide sequences.

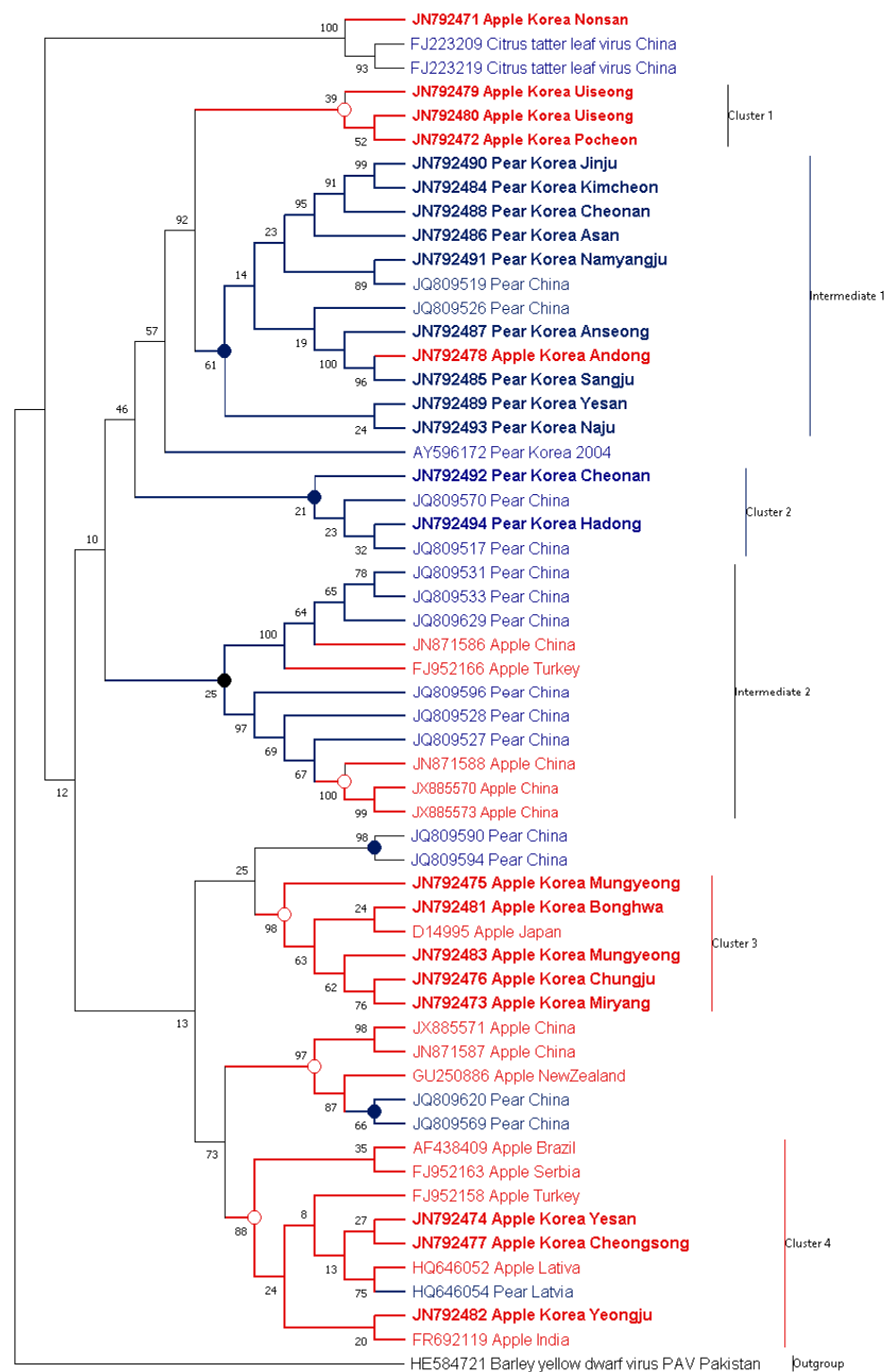
Both phylogenies followed the same clustering trend (Figures 4 and 5). The Korean apple isolates had Gln<sup>9</sup>, Ala<sup>27</sup>, Leu<sup>30</sup>, Gly<sup>38</sup>, Lys<sup>48</sup>, Lys<sup>96</sup>, Glu<sup>103</sup>, Arg<sup>104</sup>, Ala<sup>108</sup>, Ser<sup>110</sup>, Ile<sup>117</sup>, and Arg<sup>199</sup> residues, whereas the Korean pear isolates had Leu<sup>9</sup>, Gly<sup>27</sup>, Ser<sup>38</sup>, Arg<sup>48</sup>, Arg<sup>96</sup>, Glu<sup>103</sup>, Lys<sup>104</sup>, Glu<sup>108</sup>, Met<sup>110</sup>, Val<sup>117</sup>, and Lys<sup>199</sup>. The results of the amino acid sequence analysis showed differences between the apple and pear isolates in terms of amino acids residues. These changes suggest that the differences could be expressed in their structure. To explore these differences, secondary structures of the peptide chains were predicted.

The secondary structure predictions showed categorization corresponding to the phylogenetic groupings of nucleotide and amino acid sequences. Representative schematic diagrams of the predicted secondary structures are shown in Figure S4. Amino acid residue variations are reflected in the secondary structure prediction. Residue variations in the apple isolate group included Gln<sup>9</sup>, Ala<sup>27</sup>, Leu<sup>30</sup>, Gly<sup>38</sup>, Lys<sup>48</sup>, Lys<sup>96</sup>, Glu<sup>103</sup>, Arg<sup>104</sup>, Ala<sup>108</sup>, Ser<sup>110</sup>, Ile<sup>117</sup>, and Arg<sup>199</sup>, compared to the pear isolates and reflected helix length, strand and coil structure, and appearance of the beta-turn.

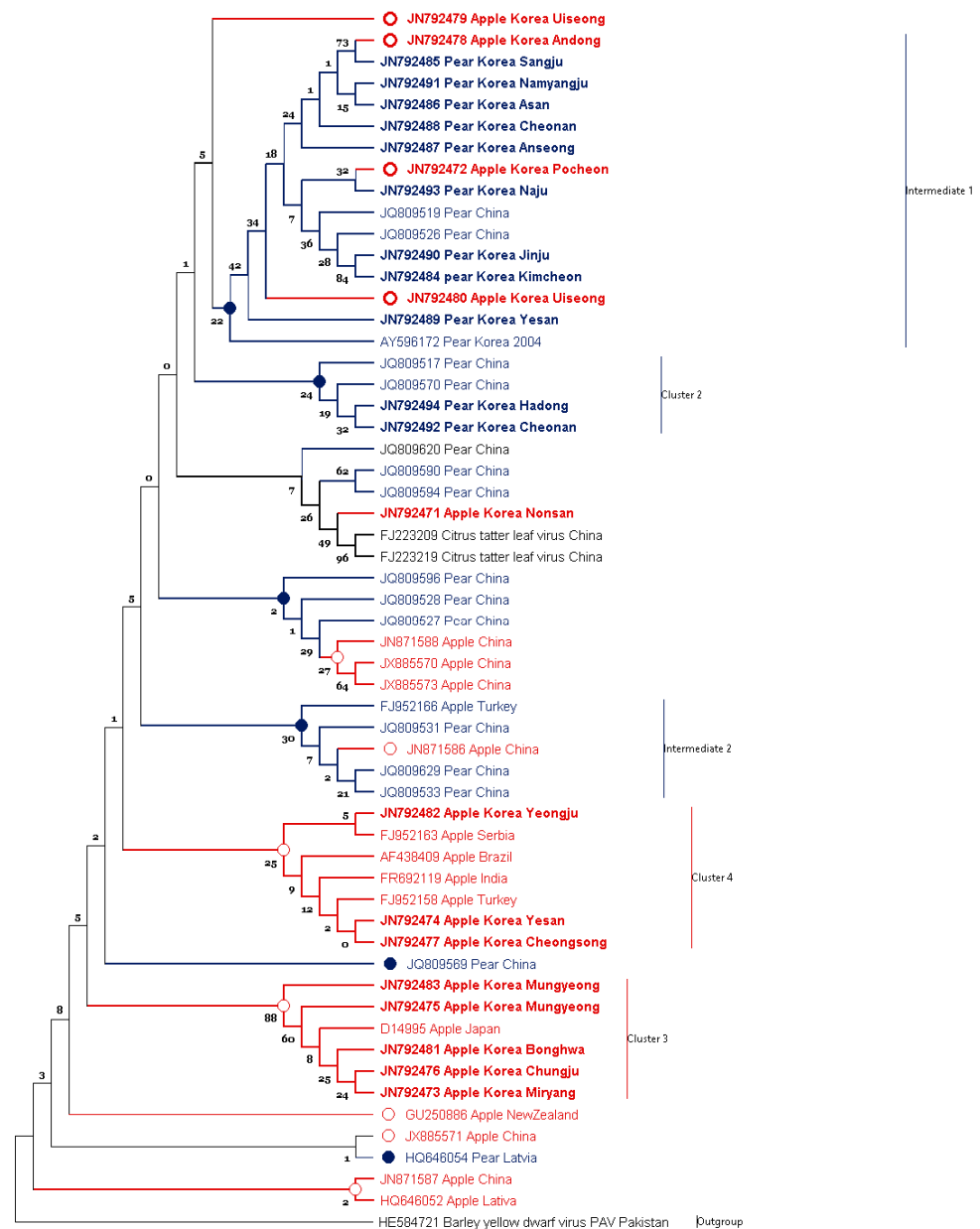
Among these variations, the most distinguishable locus was located between the 100th and 130th amino acids. At this locus, the apple isolates had shorter helix structure than the pear isolates, due to a change in residues at the 110th position; the pear isolates had a methionine residue instead of serine as in the apple isolates. Similarly, Lys<sup>198</sup> residue substitution in the pear isolate group for Arg<sup>198</sup> in the apple isolate group resulted in clear differences in the secondary structure. This residue difference changed the length of the helix structure and appearance of the beta-turn structure (Figure S4).

Furthermore, all these variations in the secondary structure are reflected in the tertiary structure prediction model. From the phylogenetic analyses based on the amino acid sequences, a total of eight isolates were selected and the tertiary structures were determined for comparisons between the apple and pear groups. From the pear group, the isolates GenBank JN792486, JN792484, JN79247, JN792485, and JN792472 were selected and from the apple group, the isolates GenBank JN792477, JN 792483, and JN792475 were selected (Figure S6). These were representative of the subgroups observed in the phylogenetic tree of the amino acid sequences.

The predicted structures differed significantly from each other. In particular, GenBank JN792486 and JN792484 were the most divergent among the eight isolates. Their structure predictions were quite different from the structure of the apple isolate group; the N-terminal region was folded in a different direction compared with that in the apple isolate group. In contrast, three other sequences had similar structures to the apple isolate group. In addition, all isolates showed structural differences corresponding to their clustering in the amino acid phylogenetic tree (Figure S2).



**Figure 4.** Phylogenetic analysis of apple stem grooving virus (ASGV) coat protein gene nucleotide. The tree is divided into distinct and intermediate clades. The Korean isolates identified in this study are shown with bold. Taxon names are color coded by isolation host (red for apple and blue for pear). The bootstrap support value is displayed on each branch node. The tree was generated using Maximum Likelihood with K2+G best-fit-model implemented in MEGAX.

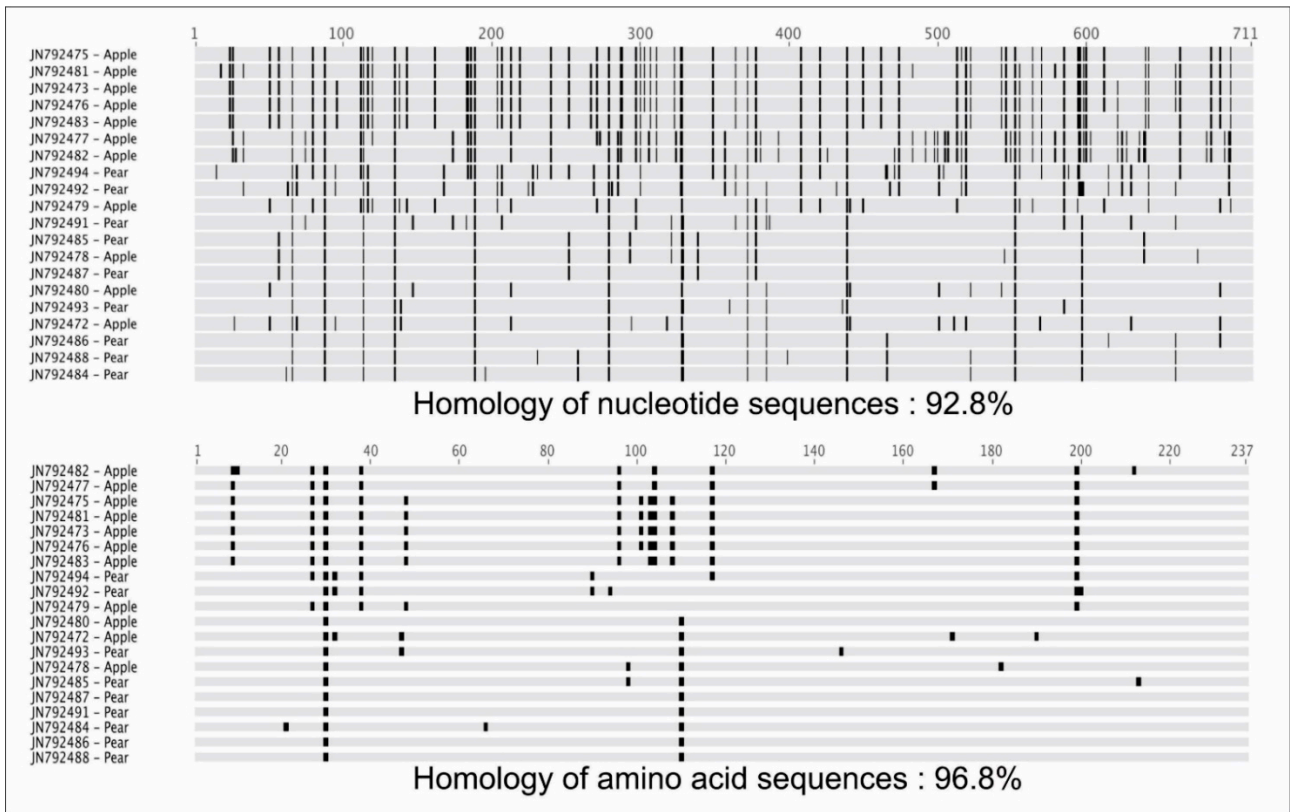


**Figure 5.** Phylogenetic analysis of apple stem grooving virus (ASGV) coat protein amino acid sequences. The tree was divided into distinct and intermediate clades. The Korean isolates identified in this study are shown with bold. Taxon names are color coded by isolation host (red for apple and blue for pear). The bootstrap support value is displayed on each branch node. The tree was generated using Maximum Likelihood with JTT+G best-fit-model implemented in MEGAX.

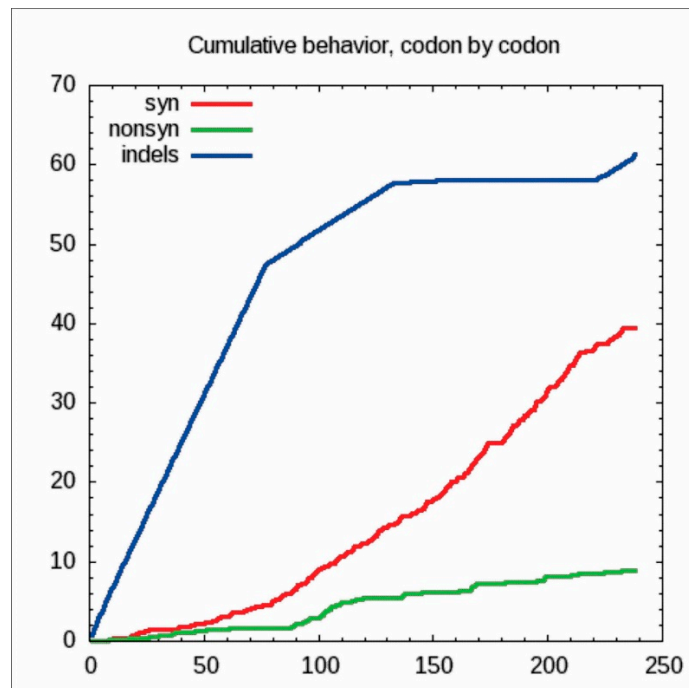
### 3.2. Comparisons of Nucleotide and Protein Sequences

Few differences were observed when phylogenetic trees based on amino acid and nucleotide sequences were compared. For instance, the maximum nucleotide and amino acid sequence identities of Korean ASGV isolates were 92.8% and 96.4%, respectively (Figure 6). As predicted, the amino acid sequence alignment showed fewer differences compared to the nucleotide sequences. The cumulative codon-by-codon behavior analysis for synonymous and non-synonymous substitutions showed a higher rate of synonymous (dS) than non-synonymous (dN) substitutions indicating purifying selection (Figure 7).





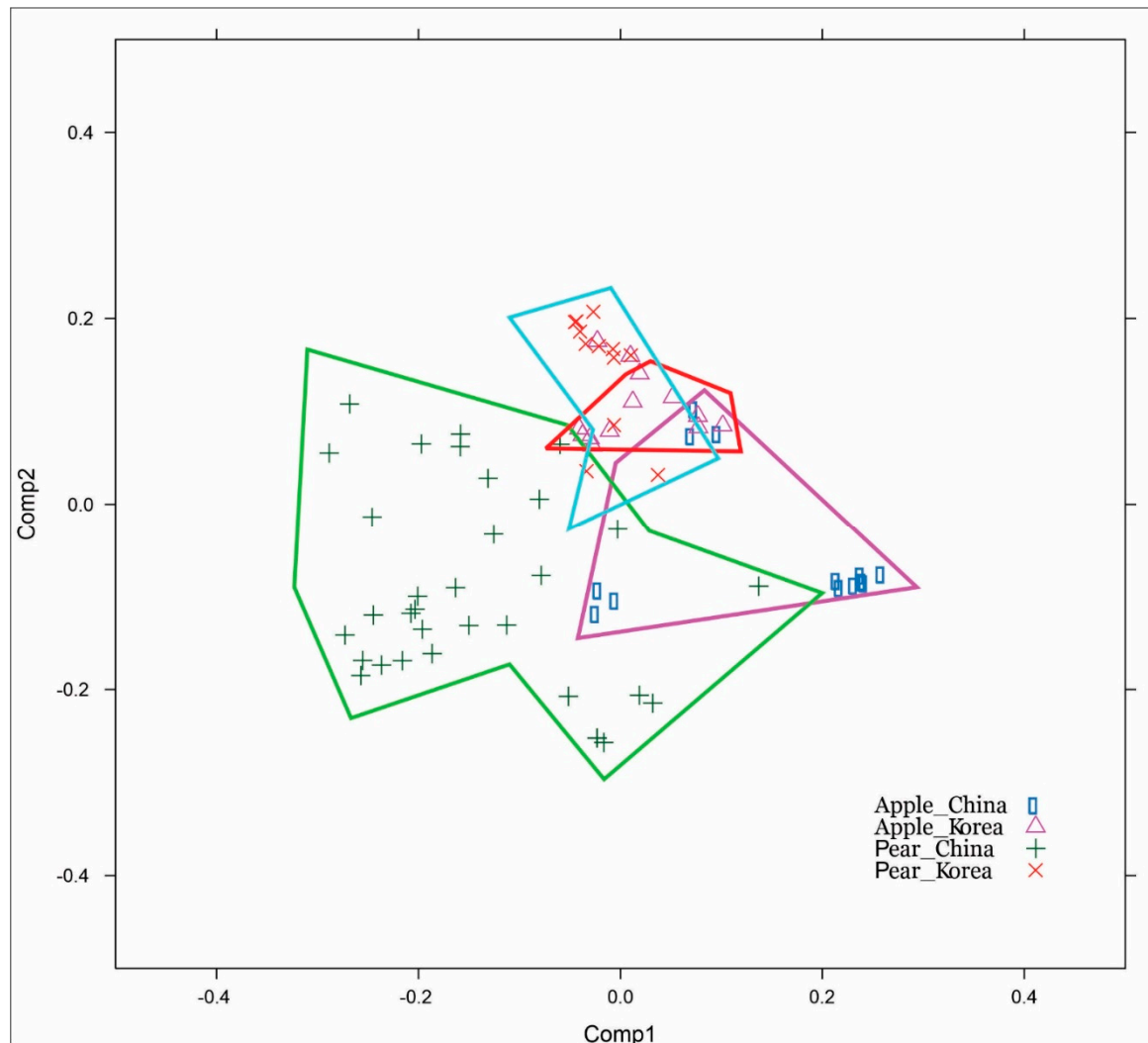
**Figure 6.** Multi-aligned sequence comparisons between nucleotides and amino acids. Identities of nucleotide and amino acid sequences were different from each other. Black highlighted positions represent differences between residues and the consensus sequence. Nucleotide sequences showed various residue substitutions coding for the same amino acids.



**Figure 7.** Cumulative behavior (codon-by-codon) of Korean isolates of *Apple stem grooving virus*.

### 3.3. The CP Gene Corresponds with the Host Codon Usage Frequency

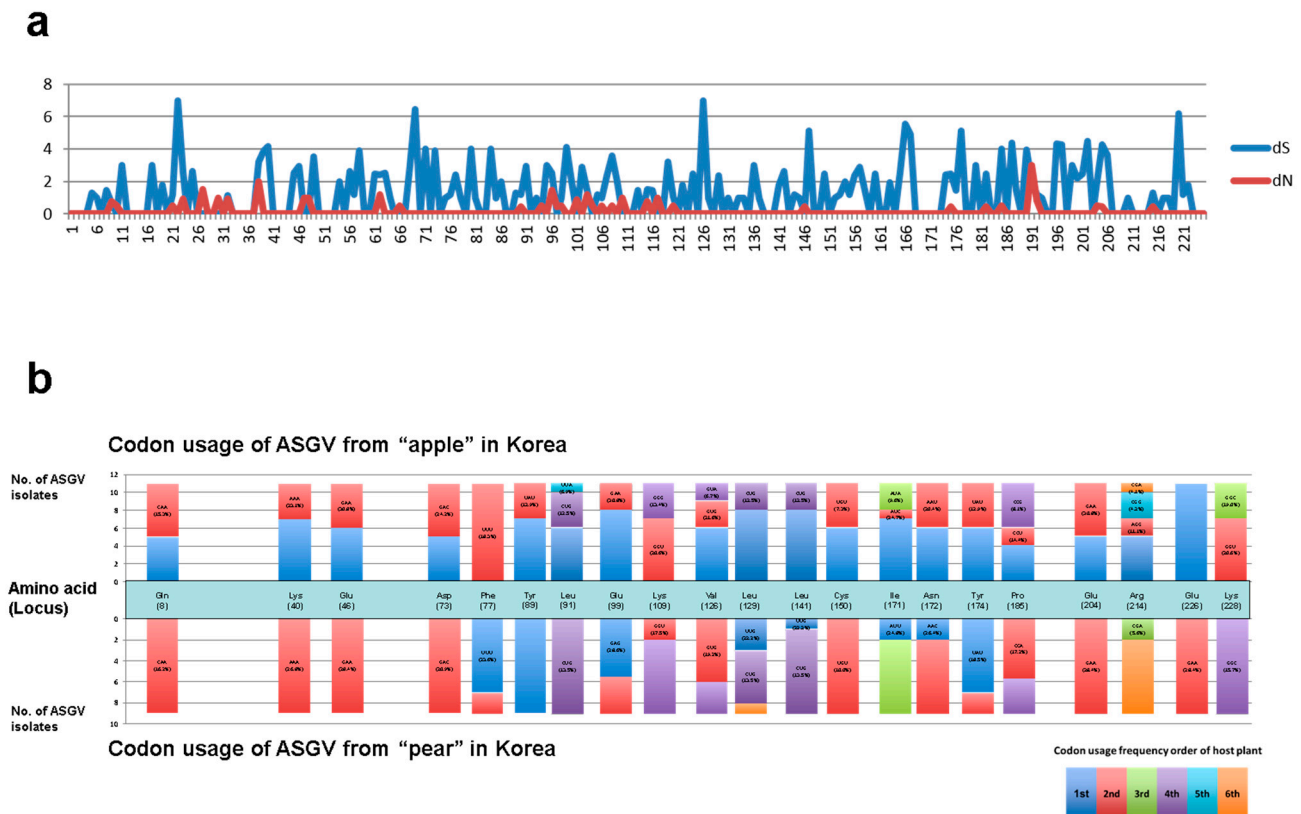
The differences in the nucleotide sequences might better be explained by variation in codon usage. Specifically, Chinese and Korean isolates were clearly separated according to their hosts. More precisely, codon usage of the ASGV CP gene from Korea and China illustrated that isolates could be distinguished by the isolation country as well (Figure 8).



**Figure 8.** Correspondence analysis with codon usage of apple stem grooving virus (ASGV) coat protein gene from Korea and China. Codon usage of ASGV corresponded with the host and country of collection. Region with red boundary line indicates in majority the Korean ASGV collected from apple. The area with a blue boundary shows Korean ASGV collected from pear. Region with a pink boundary indicates Chinese ASGV collected from apple. Region with green circumference indicates Chinese ASGV collected from pear.

For the Korean ASGV isolates, the codon usage ratio was examined codon-by-codon via a maximum likelihood analysis of natural selection. Codon usage variations appeared throughout the entire sequence of the CP gene. Specifically, these variations were not relevant to their genomic distance on the CP gene. The dS rate was higher than the dN rate for all codons among the 237 residues (Figure 9a). To characterize codon usage variation tendencies, the usage of every codon between isolates obtained from apple and pear were compared on the multiple sequence alignment. We analyzed 208 dN among a total of 237 residues for the Korean ASGV CP gene. The codon usage analysis showed variations between the CP gene of apple and pear isolates at 95 loci among 18 amino acids. Moreover, apple isolates showed more variation (showing more diversity) in codon usage than the

pear isolates. Apple isolates showed variation at 79 loci as compared to 56 loci in pear isolates. Both the variation and variety of the codon usage was high in the apple isolates when compared with pear isolates. For example, there was more codon usage variety for glutamine residues among apple isolates, whereas the codon usage was conserved in pear isolates. Similarly, codon usage variations for other amino acids were more highly conserved in pear isolates than apple isolates. Except for those variants, most codon use was similar for the entire CP gene (Figure 9b).



**Figure 9.** Codon usage analysis of Korean apple stem grooving virus (ASGV) coat protein gene. (a) Maximum likelihood analysis of natural selection codon-by-codon. The analysis involved 25 nucleotide sequences with 675 positions of ASGV codons. Most nucleotide substitutions were synonymous over the entire coat protein gene of Korean ASGV. (b) Analysis of ASGV codon usage at loci of synonymous substitutions compared with host preferred codon usage frequency. Substitutions of virus gene codons tended to follow the preferred codon usage of the isolated host.

Among these variants, codon usages of apple isolates seem to follow the codon usage frequency of the apple host at 21 amino acid residues loci (Figure 9b). In pear isolates, the codon for Cys<sup>150</sup> was UGU only, whereas among the 13 apple isolates, six were UGC and five were UGU. In apple, the UGC codon was 3.4% more frequent than UGU. In pear isolates, for Asp<sup>73</sup>, the codon was GAC, whereas among the 13 apple isolates five were GAU and six were GAC. The frequency of GAG was 6% higher than that of GAC in apple isolates. Codon usages for Glu were tended to follow the apple gene codon plant usage. Compared to pear isolates, apple isolates were more likely to follow host codon usage frequency. Codon usage frequency for Glu was similar across pear isolates, with only a 0.2% difference. However, GAG was 6.1% more frequent than GAA in apple isolates. In apple isolates, Glu<sup>46</sup> was represented by the GAG and GAA codons six and five times, respectively, whereas in pear isolates, only the GAA codon was observed. Similarly, in apple isolates, for Glu<sup>99</sup>, GAG and GAA appeared eight and three times, respectively, whereas, in pear isolates, they appeared one and eight times, respectively. At the 205th locus, in apple isolates, GAG and GAA were observed five and six times, respectively,

whereas among pear isolates, only the GAA codon was observed. At the 226th locus, GAG was observed only in apple isolates, whereas GAA was exclusively observed in pear isolates.

Codon usage variations in apple isolates for Gly<sup>109</sup>, Gly<sup>228</sup>, Ile<sup>171</sup>, Lys<sup>40</sup>, Leu<sup>91</sup>, Leu<sup>129</sup>, Leu<sup>141</sup>, Pro<sup>185</sup>, Arg<sup>214</sup>, Val<sup>126</sup>, Tyr<sup>89</sup>, and Tyr<sup>174</sup> were similar to the apple host codon usage. The codon usage frequency was similar for the similar amino acid, for instance, the codon frequency of apple for glutamine represented a difference of 1 percent only. Furthermore, analysis of Chinese ASGV isolates showed similar trend. These results indicate that the codon usage follows host codon usage frequency (Figure S3).

#### 4. Discussion

Mutations, recombinations, and reassortments are the major driving forces explored extensively in plant virus evolution [2]. Even single amino acid mutations result in dramatic changes in the biological properties of viruses, such as symptom severity and systemic infectivity. Increased symptom severity in tolerant zucchini cultivars attributed to the point mutations in Zucchini yellow mosaic virus sequence [19]. Similarly, the loss and recovery of the systemic infectivity to cucumber mosaic virus in squash linked to a single amino acid mutation in the CP (Thompson et al., 2006).

In a similar way, recombination of viral genomes could also change the biological properties of viruses. RNA-RNA homologous recombination and virus evolution studies have a long history [20]. Most studies have examined phylogenetic relationships and viral genome recombination [1,21,22]. Furthermore, the effects of viral genome recombination have been studied using artificial recombinant genomes to demonstrate the loss and recovery of viral infectivity [23]. Genome reassortment is yet another driving force of evolution in multi-partite viruses. Studies on plant virus genome reassortment have mostly been conducted with the cucumber mosaic virus mainly due its split genome organization, i.e., three-segments [24,25].

In this study, 24 ASGV CP isolates (i.e., 13 isolates from apple and 11 isolates from pear) were amplified and analyzed along with 01 former Korean isolates from pear. The amplicons were sequenced in both orientations. Mix infection of a sample with different isolates remains a major possibility; however, was not studied. The sequences were then analyzed to identify the forces driving plant virus evolution. Phylogenetic analyses showed interesting categorizations of the ASGV CP gene. All the analyzed isolates could be categorized in two groups through their corresponding hosts, pear or apple. First, recombinations among virus CP gene were analyzed. All apple and pear isolates were infected with different viruses, Apple chlorotic leaf spot virus (ACLSV) and Apple stem pitting virus, in different combinations. These viruses were classified in the same virus family, *Flexiviridae* [26,27]. To determine the relationship between recombination and phylogenetic categorization, we analyzed recombination in ASGV CP genes with co-infected virus genes. According to the viral infection status, most apple isolates were co-infected with ACLSV, whereas pear isolates were not. However, we did not detect any recombination event reliable/acceptable among viral species (data not shown).

Second, to explain the categories detected, we took clues from studies regarding host adaptation of viruses. Previous host adaptation studies have examined different viral species with various experimental hosts. Ohshima et al. (2002) analyzed the molecular variability of 66 isolates of turnip mosaic virus (TuMV) isolated from several hosts. In TuMV, phylogenetic categorization showed host and geographical differences. Sacristán et al. (2005) assayed the infectivity of cucumber mosaic virus during the 10th passage of virus inoculation to examine host adaptation. After 10 passages of inoculation, viral accumulation significantly increased on newly introduced hosts. Similarly, viruses on woody plants showed similar phenomena. Symptom severity changed when the viruses were introduced to different hosts [28,29].

Molecular analysis has been conducted to characterize the biological properties of viruses that have adapted to the introduced hosts. RNA genome sequence variations were

analyzed to determine the evolutionary history of viruses through the host passages and examine changes in amino acid sequence substitutions [30,31].

All sequences were categorized into two major groups accordingly as per their hosts. The bifurcation is evident both in nucleotides and amino acids sequence analysis. These groups showed significant differences in predicted CP structures. Torres et al. (2005) and Pinel-Galzi et al. (2007) previously suggested selective forces and pathways of virus evolution. The present genome variation analysis elucidated the direction of the virus evolution pathway. Together, these studies suggested that genomic evolution is still occurring, and that the direction can be determined. Similar to that in previous studies, we found evidence of evolution in the ASGV CP, with phylogenetic differences corresponding to the hosts in both nucleotide and amino acid levels (Figures 4 and 5). The synonymous microevolution of viral and host genomes suggests that viral genomic evolution is related to adaptation to newly introduced hosts. As Taq DNA polymerase (Cat# M7660, Promega Co.) was used for amplification, there might be some un-intended errors/mutations introduced in the amplicons due to lack of the polymerase proof reading activity. However, a trend of host-related adaption while amplifying multiple samples indicate a minimal impact of the PCR induced mutations.

The phylogenies based on nucleotide and amino acid sequences showed interesting variations. The variations were mostly caused by silent mutations in the genome as an index of wobble hypothesis. All codon usage differences followed the phylogenetic categorization corresponding to their isolated hosts. To explain the relationship between phylogenetic categorization and hosts, viral gene codon usages were compared with host codon usage frequencies. We found that the codon usage variations followed the host codon usage frequency. Most variations that agreed with host codon usage appeared in isolates from apple. For 20 loci of the ASGV CP, more than 5 isolates from apple had higher frequencies of codon similarity to the hosts, whereas pear isolates had lower frequencies of codon similarity. Pear isolates also followed the higher frequency host codons at least at two loci.

In plant evolution studies, codon usage has not been considered an important selective force [32]. However, the dicots- and monocots-infecting sobemoviruses were categorized in the same groups [10]. Nevertheless, studies of plant virus evolution that examine codon usage were not considered useful to better understand host specificity. Unlike studies of plant viruses, in animal viruses, codon usage is considered an important factor in phylogenetic relationship along with base composition [33]. Additional studies of animal viruses suggested that relationships in codon usage between the host and the virus are driving forces of virus evolution [34]. Moreover, viral protein codon adaptation to the host, as observed in viral proteins, does not require host-specific recognition [35].

We detected host adaptation with synonymous changes in the CP gene among apple isolates of ASGV. In addition, codon usage for the ASGV CP gene was conserved in isolates from pear compared with those from apple. A study on the evolutionary trajectory of *Turnip mosaic virus* reported higher nucleotide diversity among viral genomes that had been introduced to new hosts [3]. Codon usage among pear isolates was conserved, unlike the high frequency of codon variations of pear. Conserved codon usage followed tendencies for low frequency codon use of hosts. Our results are similar to the analyses of codon usage bias of Antoniwi and Adams (2003). Codon usage variation among apple isolates may, therefore, constitute evidence of ASGV adaptation to hosts. As far as protein structure analysis is concerned, the structural variability was evident in the 3D conformations of the two groups of CP (Supplementary Figure S3).

Based on the ASGV CP gene, host preference/adaptation is evident from the nucleotide and protein sequence analysis (Figures 4 and 5). More importantly, non-synonymous nucleotide substitutions were common showing significant separation in accordance with their hosts. Synonymous substitutions of the CP gene reflect parallel evolution. Variations among synonymous nucleotide substitutions follow host codon usage frequency. This indicated that viral genomes evolve alongside host molecular factors.

Mutational selection, therefore, may drive microevolution within translational selection forces, especially host codon usage frequency.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/microorganisms9061111/s1>. Figure S1. Percentage nucleotide identities of Korean Apple stem grooving virus (ASGV) coat protein gene using MUSCLE alignment implemented in SDTv1.2. Figure S2. Percentage amino acid identities of Korean apple stem grooving virus (ASGV) coat protein gene using MUSCLE alignment implemented in SDTv1.2. Figure S3. Synonymous (dS) and Non-synonymous (dN) substitution of the apple stem grooving virus (ASGV) coat protein gene (a). Codon usage of ASGV from apple and pear reported sequences from China (b). Figure S4. Secondary structure prediction of Korean apple stem grooving virus (ASGV) coat protein. Amino acid sequences were rearranged corresponding to the phylogenetic analysis categorization of nucleotide sequences. Boxes represent the significant differences between the two clades from the phylogenetic analysis. Predicted secondary structure results showed similar categorization of nucleotide phylogenetic analysis. Secondary structures were predicted using the original Garnier Osguthorpe Robson algorithm (GOR I) provided by the EMBOSS suite embedded in Geneious Pro 8. Figure S5. Protein model prediction of selected Korean apple stem grooving virus (ASGV) coat protein sequences using i-TASSER. Phylogenetic relationship of the isolates is also shown. The accessions used for model prediction were under-lined. Table S1. List of coat protein gene sequences used in the nucleotide sequence analyses.

**Author Contributions:** J.K., M.K. and S.L. conceived the research. J.K., E.-J.K. and H.-R.K. generated the data. J.K., E.-J.K., A.L., H.-R.K., H.-S.Y., H.-S.C. and M.K. analyzed the data. J.K., M.A. and S.L. wrote the article. All authors have edited and contributed to the writing and reviewed the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by a grant from the Agenda Program (PJ01426101) funded by the Rural Development Administration of Korea and a fund (Project Code No. Z-1543086-2017-21-01) by Research of Animal and Plant Quarantine Agency, South Korea.

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

**Informed Consent Statement:** Written informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The data are not available due to the nature of this research. Participants of this study did not agree for their data to be shared publicly.

**Conflicts 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.

## References

- Ohshima, K.; Yamaguchi, Y.; Hirota, R.; Hamamoto, T.; Tomimura, K.; Tan, Z.; Sano, T.; Azuhata, F.; Walsh, J.A.; Fletcher, J.; et al. Molecular evolution of Turnip mosaic virus: Evidence of host adaptation, genetic recombination and geographical spread. *J. Gen. Virol.* **2002**, *83*, 1511–1521. [[CrossRef](#)] [[PubMed](#)]
- Roossinck, M.J. Mechanisms of plant virus evolution. *Annu. Rev. Phytopathol.* **1997**, *35*, 191–209. [[CrossRef](#)] [[PubMed](#)]
- Ohshima, K.; Akaishi, S.; Kajiyama, H.; Koga, R.; Gibbs, A.J. Evolutionary trajectory of turnip mosaic virus populations adapting to a new host. *J. Gen. Virol.* **2010**, *91*, 788–801. [[CrossRef](#)] [[PubMed](#)]
- Wichman, H.A. Different Trajectories of Parallel Evolution during Viral Adaptation. *Science* **1999**, *285*, 422–424. [[CrossRef](#)] [[PubMed](#)]
- Novella, I.S.; Zárate, S.; Metzgar, D.; Ebandick-Corpus, B.E. Positive selection of synonymous mutations in vesicular stomatitis virus. *J. Mol. Biol.* **2004**, *342*, 1415–1421. [[CrossRef](#)]
- Wolf, Y.I.; Viboud, C.; Holmes, E.C.; Koonin, E.V.; Lipman, D.J. Long intervals of stasis punctuated by bursts of positive selection in the seasonal evolution of influenza A virus. *Biol. Direct.* **2006**, *1*, 34. [[CrossRef](#)]
- Woods, R.; Schneider, D.; Winkworth, C.L.; Riley, M.A.; Lenski, R.E. Tests of parallel molecular evolution in a long-term experiment with *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 9107–9112. [[CrossRef](#)]
- Bazykin, G.A.; Kondrashov, F.A.; Brudno, M.; Poliakov, A.; Dubchak, I.; Kondrashov, A.S. Extensive parallelism in protein evolution. *Biol. Direct.* **2007**, *2*, 20. [[CrossRef](#)] [[PubMed](#)]
- Musset, L.; Le Bras, J.; Clain, J. Parallel Evolution of Adaptive Mutations in Plasmodium falciparum Mitochondrial DNA during Atovaquone-Proguanil Treatment. *Mol. Biol. Evol.* **2007**, *24*, 1582–1585. [[CrossRef](#)]

10. Zhou, H.; Wang, H.; Huang, L.F.; Naylor, M.; Clifford, P. Heterogeneity in codon usages of sobemovirus genes. *Arch. Virol.* **2005**, *150*, 1591–1605. [[CrossRef](#)] [[PubMed](#)]
11. Pinel-Galzi, A.; Rakotomalala, M.; Sangu, E.; Sorho, F.; Kanyeka, Z.; Traoré, O.; Sérémé, D.; Poulicard, N.; Rabenantoandro, Y.; Séré, Y.; et al. Theme and variations in the evolutionary pathways to virulence of an RNA plant virus species. *PLoS Pathog.* **2007**, *3*, e180. [[CrossRef](#)]
12. Magome, H.; Yoshikawa, N.; Takahashi, T.; Ito, T.; Miyakawa, T. Molecular variability of the genomes of capilloviruses from apple, Japanese pear, European pear, and citrus trees. *Phytopathology* **1997**, *87*, 389–396. [[CrossRef](#)] [[PubMed](#)]
13. Yanese, H.; Mink, G.I.; Sawamura, K.; Yamaguchi, A. Apple topworking disease. In *Compendium of Apple and Pear Diseases*; Jones, A.L., Aldwinckle, H.S., Eds.; APS Press: St. Paul, MN, USA, 1990; pp. 74–75.
14. Shim, H.; Min, Y.; Hong, S.; Kwon, M.; Kim, D.; Kim, H.; Choi, Y.; Lee, S.; Yang, J. Nucleotide sequences of a Korean isolate of apple stem grooving virus associated with black necrotic leaf spot disease on pear (*Pyrus pyrifolia*). *Mol. Cells* **2004**, *18*, 192–199. [[PubMed](#)]
15. Roy, A.; Kucukural, A.; Zhang, Y. I-TASSER: A unified platform for automated protein structure and function prediction. *Nat. Protoc.* **2010**, *5*, 725–738. [[CrossRef](#)] [[PubMed](#)]
16. Foley, B.; Leitner, T.; Apetrei, C.; Hahn, B.; Mizrachi, I.; Mullins, J.; Rambaut, A.; Wolinsky, S.; Korber, B. (Eds.) *HIV Sequence Compendium 2018*; LA-UR 18-25673; Theoretical Biology and Biophysics Group, Los Alamos National Laboratory: Santa Fe, NM, USA, 2018.
17. Tamura, K.; Stecher, G.; Peterson, D.; Filipiński, A.; Kumar, S. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **2013**, *30*, 2725–2729. [[CrossRef](#)] [[PubMed](#)]
18. Nakamura, Y.; Gojobori, T.; Ikemura, T. Codon usage tabulated from international DNA sequence databases: Status for the year 2000. *Nucleic Acids Res.* **2000**, *28*, 292. [[CrossRef](#)]
19. Desbiez, C.; Gal-On, A.; Girard, M.; Wipf-Scheibel, C.; Lecoq, H. Increase in Zucchini yellow mosaic virus Symptom Severity in Tolerant Zucchini Cultivars Is Related to a Point Mutation in P3 Protein and Is Associated with a Loss of Relative Fitness on Susceptible Plants. *Phytopathology* **2003**, *93*, 1478–1484. [[CrossRef](#)]
20. Simon, A.E.; Bujarski, J.J. RNA-RNA Recombination and Evolution in Virus-Infected Plants. *Annu. Rev. Phytopathol.* **1994**, *32*, 337–362. [[CrossRef](#)]
21. Revers, F.; Le Gall, O.; Candresse, T.; Le Romancer, M.; Dunez, J. Frequent occurrence of recombinant potyvirus isolates. *J. Gen. Virol.* **1996**, *77 Pt 8*, 1953–1965. [[CrossRef](#)]
22. Ali, M.; Tahir, M.; Hameed, S. Phylogenetic and genome-wide pairwise distance analysis of the genus *Luteovirus*. *Pak. J. Agric. Sci.* **2017**, *54*, 363–371. [[CrossRef](#)]
23. Hajimorad, M.R.; Eggenberger, A.L.; Hill, J.H. Adaptation of Soybean mosaic virus Avirulent Chimeras Containing P3 Sequences from Virulent Strains to Rsv1-Genotype Soybeans Is Mediated by Mutations in HC-Pro. *Mol. Plant-Microbe Interact.* **2008**, *21*, 937–946. [[CrossRef](#)] [[PubMed](#)]
24. Lin, H.X.; Rubio, L.; Smythe, A.B.; Falk, B.W. Molecular Population Genetics of Cucumber Mosaic Virus in California: Evidence for Founder Effects and Reassortment. *J. Virol.* **2004**, *78*, 6666–6675. [[CrossRef](#)] [[PubMed](#)]
25. Chen, Y.; Chen, J.; Zhang, H.; Tang, X.; Du, Z. Molecular evidence and sequence analysis of a natural reassortant between Cucumber mosaic virus subgroup IA and II strains. *Virus Genes* **2007**, *35*, 405–413. [[CrossRef](#)] [[PubMed](#)]
26. Adams, M.J.; Antoniw, J.F.; Bar-Joseph, M.; Brunt, A.A.; Candresse, T.; Foster, G.D.; Martelli, G.P.; Milne, R.G.; Fauquet, C.M. Virology Division News: The new plant virus family Flexiviridae and assessment of molecular criteria for species demarcation. *Arch. Virol.* **2004**, *149*, 1045–1060. [[CrossRef](#)] [[PubMed](#)]
27. Martelli, G.P.; Adams, M.J.; Kreuze, J.F.; Dolja, V.V. Family Flexiviridae: A Case Study in Virion and Genome Plasticity. *Annu. Rev. Phytopathol.* **2007**, *45*, 73–100. [[CrossRef](#)] [[PubMed](#)]
28. Moury, B.; Cardin, L.; Onesto, J.P.; Candresse, T.; Poupet, A. Survey of Prunus necrotic ringspot virus in Rose and Its Variability in Rose and *Prunus* spp. *Phytopathology* **2001**, *91*, 84–91. [[CrossRef](#)] [[PubMed](#)]
29. Sentandreu, V.; Castro, J.A.; Ayllón, M.A.; Rubio, L.; Guerri, J.; González-Candelas, F.; Moreno, P.; Moya, A. Evolutionary analysis of genetic variation observed in citrus tristeza virus (CTV) after host passage. *Arch. Virol.* **2005**, *151*, 875–894. [[CrossRef](#)]
30. Rico, P.; Ivars, P.; Elena, S.F.; Hernández, C. Insights into the selective pressures restricting Pelargonium flower break virus genome variability: Evidence for host adaptation. *J. Virol.* **2006**, *80*, 8124–8132. [[CrossRef](#)] [[PubMed](#)]
31. Moury, B. A new lineage sheds light on the evolutionary history of Potato virus Y. *Mol. Plant Pathol.* **2010**, *11*, 161–168. [[CrossRef](#)]
32. Antoniw, J.F.; Adams, M.J. Codon usage bias amongst plant viruses. *Arch. Virol.* **2003**, *149*, 113–135. [[CrossRef](#)]
33. Jenkins, G.M.; Pagel, M.; Gould, E.A.; de A Zannotto, P.M.; Holmes, E.C. Evolution of base composition and codon usage bias in the genus *Flavivirus*. *J. Mol. Evol.* **2001**, *52*, 383–390. [[CrossRef](#)] [[PubMed](#)]
34. van Hemert, F.J.; Berkhout, B.; Lukashov, V.V. Host-related nucleotide composition and codon usage as driving forces in the recent evolution of the Astroviridae. *Virology* **2007**, *361*, 447–454. [[CrossRef](#)] [[PubMed](#)]
35. Bahir, I.; Fromer, M.; Prat, Y.; Linial, M. Viral adaptation to host: A proteome-based analysis of codon usage and amino acid preferences. *Mol. Syst. Biol.* **2009**, *5*, 1–14. [[CrossRef](#)] [[PubMed](#)]