On the Evolution and Biogeography of Apple stem grooving capillovirus

Kamal Shujaei[1], Adrian John Gibbs[2], Mohammad Hajizadeh[*1]

[1]Plant Protection Department, Faculty of Agriculture, University of Kurdistan, Sanandaj, Iran;

[2]Emeritus Faculty, Australian National University, Canberra, Australia

*****Corresponding author: M. Hajizadeh

E-mail: hajizadeh2003@yahoo.com

ORCIDs: https://orcid.org/0000-0002-0774-3523

Abstract

Apple stem grooving capillovirus (ASGV) has a wide host range, notably apples, pears and citrus. It is found worldwide. In this study, the published nucleotide sequences of the genomes of 60 isolates and of over 200 coat protein (CP) genes, including those of seven newly determned Iranian isolates, were analyzed. The non-recombinant genomes showed that sequences concatenated from the replicase genes (minus a short variable region), together with the movement and CP genes gave the best supported ML phylogenies, and these also correlated best with CP phylogenies, so that information from phylogenies of the genomic sequences could be augmented, with caution, by that from CP phylogenies. The basal ASGV population was only isolated from CP samples from Himalchan Pradesh of northern India. Its progeny lineages diverged into populations in south and east Asia, especially China. The ASVG population in the USA probably came from a Chinese lineage, and the Eurasian one, including seven Iranian isolates, from Indian lineages. The analyses suggests that trade, negative selection and founder effects were the most important factors affecting the genetic evolution of ASGV.

Key words: Apple stem grooving virus, Genome and Gene phylogeny, Biogeography Population Genetics, Overlapping Gene

Analysis of gene sequences provides essential information for planning effective control measures for a virus as it may indicate its 'centre of emergence' and the mode and timing of its subsequent spread and evolution (Gibbs et al. 2008). As most viruses with RNA genomes replicate

fast using RNA polymerases that lack proof-reading activity, they often generate large and variable populations containing a significant numbers of recombinants. Therefore, they have a great potential for rapid evolution and adaptation to new conditions (Yasaka et al. 2017; Hajizadeh et al. 2019). Their genome stability is mostly maintained by pervasive negative selection (Domingo and Holland 1997), which ensures deleterious mutants are removed (Gao et al. 2017a), and this is critical for their survival.

Apple stem grooving capillovirus (ASGV) is the type species of the genus *Capillovirus*, subfamily *Trivirinae*, family *Betaflexiviridae*, and order *Tymovirales* (Adams et al. 2016). Nowadays it is widely distributed in rosaceous fruit trees such as apple, apricot, cherry and pear, causing a latent infection in most commercial cultivars but some apple cultivars may develop severe symptoms such as xylem pitting and grooving, phloem necrosis, decreased trunk diameter and the complete decay of the tree thus greatly decreasing its productivity (Nickel et al. 2001; Maxim et al. 2004). ASGV is also an important pathogen of many varieties of citrus, causing the tatter leaf disease (Magome et al. 1997).

Although several papers have reported analyses of gene sequences of regional populations of ASGV (Nickel et al. 2001; Han et al. 2016; Souza et al. 2017; Wang et al. 2018), the global population structure of this virus has not been reported. Currently (March, 2020) 60 full length ASGV genomic sequences have been reported, together with the sequences of coat protein (CP) genes of more than 140 more isolates (https://www.ncbi.nlm.nih.gov/nuccore/). The aims of our study was to investigate whether the available gene sequences of not only full length genomes of ASGV, but also the more numerous of the CP region, could provide useful insights to the phylogeny, population evolution and biogeography of ASGV. Furthermore, we report and include the CP gene sequences of seven Iranian ASVG isolates as they are potentially of particular interest; Iran is at the western end of the legendary "Silk Road", which from around 200 BCE to 1800 CE connected Eurasia to China and east Asia, and which is bounded in central Asia by the Tien Shan mountains, which are considered the site of domestication of apples (Zohary et al. 2011; Spengler 2019), if therefore ASGV had first infected apples while it was being domesticated then isolates from Iran might be more basal than others in an ASGV phylogeny.

ASGV is a capillovirus. It has filamentous virions, each composed of a single positive-sense, single-stranded RNA genome of about 6.5 x 10^3 nucleotides excluding the non-stranslated 5' and the 3' polyadenylated tail (Yoshikawa et al. 1992; Yoshikawa and Takahashi 1998) enclosed in a helically constructed tube of c. 1.3 x 103 subunits of a single species of coat protein (CP). The genome of ASGV has two ORFs (Ohira et al. 1995; Liebenberg et al. 2012). ORF1 covers most of the genome, except

for the termini, and has two very variable regions (Tatineni et al. 2009a; Liebenberg et al. 2012); a small one around nts 1590-1710 and a muchlarger one around nts 4749-5604. ORF1 encodes a replicase polyprotein (241 kDa), and the 3' terminal CP (27 kDa), which is expressed by a sub-genomic RNA (Hirata et al. 2003). A search of the Pfam database with the replicase polyprotein sequence shows that it encodes, from N- to C-terminus, a viral methyltransferase (v-Mtase), papain-like protease, viral super family 1 helicase, a RNA-depemdent RNA-polymerase (RdRp). The amino acid sequences encoded by the two variable regions only match ASGV sequences in BLASTn or BLASTp searches of the Genbank database, and their functions are unknown. The ORF2 is in the -2 reading frame of the second variable region, and encodes a movement protein (MP) of 36-kDa (Magome et al. 1997; Tatineni et al. 2009b) of the '30 kDa superfamily'. Thus, in summary, the genome has five regions; region 1 encodes a v-Mtase, the short region 2 is variable and of unknown function, region 3 encodes a papain-like endopeptidase, a helicase and a RdRp, region 4a is the second variable region and encodes a "polyprotein linker domain" (Pavesi 2019) that only matches ASGV sequences in BLASTp searches, however it completely overlaps region 4b, which encodes a MP that matches the MPs of many other triniviruses in BLASTp searches of the Genbank database, and ORF 1 is completed by region 5, which is the CP gene.

Sixty full length ASVG genomic sequences (SDFile 1) were found in, and downloaded from, Genbank (March 2020). The 5' and 3' terminal non-coding regions were removed, and the aligned ORFs checked for recombination using RDP 4.95 (Martin et al. 2015) and the 19 found to be recombinant were removed from further analyses. The relationships, genetic diversity and phylogenetic information of the remaining 41 genomic sequences, and each of their five genomic regions, were compared in various ways. First their nucleotide sequences and the amino acid sequences they encoded were aligned using the TranslatorX server (Abascal et al. 2010), and used to generate ML phylogenetic trees using PhyML (Guindon and Gascuel 2003) and the patristic distances within each of these trees were compared pairwise using PATRISTIC (Fourment and Gibbs 2006). The resulting correlation coefficients (Table 1) show (row 2) that, with the exception of the 4b region, comparisons of ASGV nucleotide trees correlate better than those calculated from the amino acid sequences they encode. Also the phylogenies calculated from regions 1 and 3 (i.e. the major regions of the replicase gene) are most similar (correlation coefficient 0.925; $p < 0.001$), whereas the variable regions (2 and 4a) are least similar (correlation coefficient 0.632; $p < 0.001$), furthermore the MP and CP phylogenies correlate most closely with those of regions 1 and 3 of the replicase gene (0.727 and 0.771; $p < 0.001$), which together represent 85% of the genome. These results confirm those of Pavesi

(2019) who showed that dominance of the ASGV overlap was particularly assymetric, and indicates that region 4b, not 4a, is the 'parental' reading frame of that region of the genome. This conclusion was confirmed by a pairwise comparison of codon usage in the different regions (Keese and Gibbs 1992), which showed (Table S1) that the usages in regions 1, 3, 4b and 5 were significantly similar, and differed from those in regions 2 and 4a, which differed from one another.

The relationships between the five genomic regions was further confirm by population genetic analysis Table 2, which showed that in all the population genetic features, the 1, 3, 4b and 5 regions were closely similar, and significantly different from those of regions 2 and 4a. In particular the dN/dS ratio, which is a measure of selection, indicates that regions 1, 3, 4b and 5 are all under strong negative selection (range 0.027 - 0.072), whereas those of regions 2 and 4a (0.525 and 2.813) respectively, may be under positive selection. The Tajima's D metrics for most regions are similar, but that of region 5 (the CP gene) differs from the others and is negative (i.e. indicative of a population expansion after a bottleneck), although this value is not statistically significant.

The five gene regions were also checked for the possibility of positive selection using the Fast Unconstrained Baysian AppRoximation (FUBAR) and Fixed effect Likelihood (FEL) in the Datamonkey programs (Weaver et al. 2018). The FUBAR method detected codons 3, 187, and 5, in regions 2, 4a and 5 respectively, to be under strong positive selection, whereas there were no codons identified as being under positive selection for regions 1, 3, and 4b. Of the regions that may have been positively selected, amino acids 159 and 4 of regions 4a and 5 were confirmed by the FEL method ($P > 0.05$) to be positively selected. Unexpectedly, none of the region 2 sites found to be under positive selection were confirmed by this method (Table 3). These results indicated that the regions of 1, 3 and 4b were under strong evolutionary constraints and region 4a under strong positive selection pressure.

In summary, it is clear that the phylogenetic relationships of regions 1 and 3 (RP), 4b (MP) and 5 (CP) of the ASGV genome (94% of it) are closely similar, whereas the two variable regions, 2 and 4a, are not. Fig. 1 shows a ML tree calculated from the concatenated sequences of the phylogenetic regions. This phylogeny was rooted at its midpoint, close to the position of the root, at the base of branch "2", obtained using Yacon virus Y as an outgroup; Fig. S1 shows a ML tree of the amino acid sequences of the RNA dependent RNA polymerases of one ASVG isolate (NC_001749) and those of all the viruses in the Genbank database that are phylogenetically close to it, and it shows that Yacon virus A (YVA; NC_030657) is the closest to ASGV. This tree also shows that ASGV and YVA share a common ancestor related to the Hobart betaflexivirus 1 metagenome isolated from honey bees, *Apis*

mellifera, collected in Hobart, Tasmania, Australia. This might indicate that ASGV, which is not known to have natural vectors, may be spread naturally via pollen or nectar in addition to grafting.

The phylogeny in Fig. 1 shows that the concat sequences fall into clusters, that are mostly congruent with the clusters found in a phylogeny of CP sequences (see below), and have been given the same numbers. The clusters are well-supported (SH values > 0.9), but the mid and basal nodes within the tree are less certain. The phylogeny indicates clearly that the population of ASGV isolates from which complete genomic sequences were obtained, originated in China and spread in several lineages to other parts of south and east Asia, but samples from other parts of the world are only found in three recently diverged clusters; "6", "8" and "9". The distribution of hosts in the phylogeny show a clear grouping within clusters, irrespective of the country of origin, as expected for a virus spread by trade and grafting, and suggests that there was little or no spread by other means.

The patristic distances within the phylogeny in Fig. 1 tree gave a correlation coefficient of 0.817 ($p < 0.001$) with a tree calculated from region 5 (CP) sequences alone and therefore we conclude that CP sequences could, with caution, be used as surrogates for complete genomes in phylogenetic/biogeographic studies of ASGV. We downloaded from Genbank (March 2020) the CP gene sequences of more than 250 isolates from 11 countries: Brazil, China, Czech Republic, Germany, India, Japan, Serbia, South Korea, Taiwan, Turkey, the USA. To these we added the newly determined CP sequences of seven ASGV isolates from Iran. These came from 174 leaf samples collected during summer of 2017 and 2018 from apple, apricot, and pear of symptomatic and non-symptomatic plants from several locations of west and northwest of Iran. Total RNA was extracted from each leaf sample using the method described by Foissac et al. (2000) with minor modifications. Reverse transcription (RT) reactions were done using a cDNA solution synthesis kit (HyperScriptTM Reverse Transcriptase, GeneAll, South Korea) and random hexamer primers. PCR reactions done using specific primers, ASGV-U: 5’CCCGCTGTTGGATTTGATACACCTC3’ and ASGV-2: 5’GGAATTTCACACAGACTCCTAACCCTCC3’ (Kundu et al. 2003) in the PCR master mix (GeneAll, South Korea). PCR products were electrophoresed in a 1.2% (w/v) agarose gel (containing 0.5 µg/µl EtBr) in 1X TAE buffer. ASGV was detected in 18 (10.3%) of samples and no RNAs were obtained from healthy or from water controls. Sporadic symptoms observed in orchard trees were not consistently associated with the presence of this virus. Seven isolates (Table S2), obtained from five apple, one apricot and one pear trees infected by ASGV were used to amplify the full-CP gene by the specific primer pairs, ASGV-MF: 5’ATGAGTTTGGAAGACGTGCTTC3’ and ASGV-MR:

5’AACCCCTTTTTGTCCTTCAGTACGAA3’. The PCR products were ligated into the pTG19 cloning vector (SinaClon, Iran), and the ligation mix was used to transform *Escherichia coli* strain DH5α as described by Chung et al. (1989). The identities of the recombinant plasmids were confirmed by restriction analysis and a purified clone from each isolate was sequenced by Macrogen Inc. (Seoul, South Korea). All seven sequences contained inserts 714 nt in length with 237 deduced amino acid residues. ASGV had a TAG at the end as stop codon in the all sequences. The greatest nucleotide diversity was at position 590, and least diversity was in their 5′-terminal regions. The encoded amino acid sequences, showed only two differences, at positions 52 and 122, and the C and N-terminal regions were completely conserved. Sequence identity among isolates from different hosts and location in our study ranged from 98.9 to 100% and 99.2 to 100% in the nucleotide and amino acid sequences (aa), respectively. No differences were found between the encoded CP sequences of the Iranian isolates and those of isolates from other countries except for one aa (Proline replaced Leucine at position 122) in isolate NG. The sequence of the CP gene of KS3 (isolated from apple), KZ (appricot), D3 (apple), NG (pear), SS (apple), SS4 (apple), X4 (apple) were deposited in GenBank with Accession Codes MK354030- MK354036, respectively.

The ASGV CP sequences were aligned using Clustal W (Thompson et al. 1997) in MEGA X program (Kumar et al. 2018) or using Translator X with default parameters. No putative recombinant events were found in the Iranian ASGV isolates or the other isolates using RDP. For further analysis, duplicate sequences were deleted leaving 212 sequences (Table S3). Nucleotide diversity (π) was estimated as the mean nucleotide distance between sequence pairs using the Mega X Program. The global nucleotide diversity of ASGV was 0.81 ± 0.006, very close to the nucleotide diversity of the CP gene of cherry virus A (0.084) (Gao et al. 2017b), another Capillovirus, but with a smaller genome than other viruses in the family *Betaflexiviridae* (Alabi et al. 2014; Yoon et al. 2014).

Phylogenies were calculated by PhyML (Guindon and Gascuel 2003) using the Shimodaira-Hasegawa test (Shimodaira and Hasegawa 1999) for node support. Fig. 2 shows the ML tree calculated from the non-recombinant and unique CP gene sequences. They fall into 11 clusters, mostly collapsed (details in File S1), and numbered to indicate which include CPs from the concats used for Fig. 1. The major difference is the fully supported (SH=1) extra cluster ("1"), which is a basal sister lineage to all other isolates. It contains 10 isolates from four hosts, all known only from their CP sequences, all from the Himachal Pradash region of North India, and with relationships within the cluster suggesting that apple (*Malus domestica*) was probably the original host. Cluster 2, known from one complete sequence

and 13 CPs, is mostly from Chinese isolates and is also strongly supported. The other nine clusters are found in both trees with > 0.8 SH support, but the relationships between the clusters is not resolved in either tree, except clusters "6" and "11". The ASVG isolates that have been sequenced most frequently are from samples collected in south and east Asian countries; 72 from India, 54 from China, 42 from South Korea, 11 from Japan and six from Taiwan, compared with nine from the Americas (clusters "2", "6", "8" and "9"), and 18 from Eurasia (cluster "6" and "7"). All seven Iranian isolates were members of cluster "6", together with isolates from Czech Republic, Germany, Serbia and Turkey, and also regions of east Asia and the Americas. The most likely interpretation of these patterns is that the earliest ASGV isolates currently known were from the Himalchan Pradesh region of India, that the virus diverged forming several lineages in south and east Asia before invading the Americas and Eurasia. Two thirds of the samples were from apple and pear trees, and only 15% from citrus.

As the pattern of divergence of the ASGV population was poorly resolved, perhaps because this is a recently diverged population, the genetic structure of the ASGV populations in the different world regions, were compared by grouping the isolates into the four main subpopulations: India and East Asia (China, Japan, South Korea, Taiwan), Eurasia (Iran, Czech Republic, Germany, Serbia and Turkey) and the Americas (Brazil and the USA). All geographic subpopulations showed small nucleotide diversity values (0.068, 0.086, 0.019, and 0.065, respectively; Table 4) that were in the same range as those found between the phylogenetic subpopulations (0.051 to 0.093). The extent of genetic differentiation and gene flow (Fst) were checked using the DnaSP 5.10 program (Rozas 2009); values of $F_{ST} > 0.33$ indicate that there are large genetic differences between the subpopulations, and they have probably not been genetically connected, whereas if < 0.33 then gene flow may have occurred (Balloux and Lugon-Moulin 2002). The pairwise Fst comparisons of the geographic populations (Table 4) gave clear evidence of gene flow between India and East Asia, also between India and Eurasia. Links between East Asia, probably China, and the Americas were supported more clearly than between India and the Americas. These analyses provide clear support for the phylogenetic evidence linking the ASGV population of India with those of East Asia, but also show that the Eurasian ASGV population came from the Indian population.

Evolutionary events such as population expansion, bottlenecks and selection, are tested by Tajima's D (Tajima 1989) and by Fu & Li's D and F static tests (Fu and Li 1993), therefore these were tested using the DnaSP 5.10 program. The significantly negative values of Tajima's D and Fu and Li's F* metrics for the Eurasian population (Table 5) is evidence of recent population expansion of that

population, however the statistics of similar past population expansions of the Indian, East Asian and USA were indicative but not statistically significant.

In conclusion, our use of both genomic and CP genes of ASGV, together with a closer outgroup, YVA, than was available before, has enabled us to establish a likely biogeographic framework for this virus. The earliest population of this virus currently known is in north India, it diverged from there into East Asia and from there into the Americas, whereas the Eurasian population was established directly from the Indian population. The relationship of ASGV with YVA is interesting because yacon (*Smallanthus sonchifolius* – *Asteraceae*) is a species of perennial daisy that was domesticated and grown as a root vegetable in the northern and central Andes of South America, however the YVA from which the sequence was obtained came from a yacon plant grown in Poland, and YVA has not been reported from South America, even though yacon crops have been tested for viruses (R.A.C. Jones; personal communication). As stated above apples were initially domesticated at least five thousand years ago from *Malus sieversii* in the Tien Shan mountains of central Asia, and then spread westwards along the 'Silk Road' progressively introgressing with other Eurasian *Malus* species [13, 14], however our results indicate that it is unlikely that they carried ASGV with them to Iran, but, more likely, the Iranian population came in recent trade as part of a worldwide lineage from the Indian population.

Figure legends

Fig. 1 Maximum likelihood phylogram of the concatenated replicase, MP and CP genes (minus short variable region of replicase) of 41 ASGV genomes. Midpoint rooted. Twelve clusters or branches are identified and numbered to identify the clusters with the same isolates in the CP phylogram (Fig. 2). Blue disks on nodes with > 0.9 SH support. Branch length scale in nucleotide substitutions per site.

Fig. 2. Maximum likelihood phylogram of CP gene sequences of 212 ASGV isolates, using that of Yacon virus A as the outgroup and with its branch shortened by 95%. Blue disks on nodes with > 0.9 SH support. The hosts are: A, *Actindia deliciosa*; C, *Citrus jambhiri*; F, *Ficus palmata*; M, *Malus domestica*; P, *Pyrus communis*. Branch length scale in nucleotide substitutions per site. Eleven clusters with > 0.8 SH support are identified and numbered to identify the clusters in Fig. 1 that include the same sequences.

Fig. S1. Maximum Likelihood phylogram of the amino acid sequences of the RNA dependent RNA polymerase from members of the *Betaflexiviridae*.

Table 1 Pairwise comparisons of the patristic distances of ML trees inferred from the five regions of 41 ASGV genomic sequences - correlation coefficients

Region	1	2	3	4a	4b	5
nts v aas	0.847	0.883	0.904	0.940	0.500	0.757
nts v nts (aas v aas)	1	2	3	4a	4b	5
1		0.760 (0.688)	0.925 (0.815)	0.797 (0.699)	0.792 (0.473)	0.727 (0.528)
2			0.812 (0.702)	0.636 (0.698)	0.636 (0.445)	0.714 (0.559)
3				0.872 (0.655)	0.869 (0.482)	0.771 (0.427)
4a					0.999 (0.534)	0.662 (0.354)
4b						0.656 (0.259)

Table 2 . Genetic diversity of ASGV in different regions of genome

Regions[a]	**Lenght (nt)**	**Θ**[b]	**π**	**d_N**	**d_S**	**d_N/d_S**	**Tajima's D**
1	1545	0.459	0.173	0.042	0.638	0.066	0.225ns
2	171	0.678	0.455	0.380	0.724	0.525	0.101ns
3	3033	0.456	0.180	0.037	0.707	0.052	0.271ns
4a	858	0.411	0.144	0.169	0.060	2.813	0.407ns
4b	853	0.410	0.144	0.015	0.572	0.027	0.402ns
5	711	0.313	0.080	0.021	0.286	0.072	-0.426ns

[a] ASGV genomic regions analyzed: 1, the methyltransferase domain of the RNA-dependent RNA polymerase (RdRp); 2, located within the RdRp; 3, the P-pro, Hel and RdRp domains of ORF1; 4, located within the movement protein (MP) and 5; the coat protein coding region.

[b] Proportion of segregating sites

Table 3 . Estimates of selection pressures acting on the five genome regions of global isolates of ASGV using FABUR and FEL methods in Datamonkey

Regions	**sites**	**PS**[a] **(FABUR)**	**NS**[b] **(FABUR)**	**PS (FEL)**	**NS (FEL)**
1	515	0	460	0	436
2	57	3 (34, 38, 46)	18	0	20
3	1011	0	958	0	905
4a	286	187	5	159	6
4b	284	0	251	0	246
5	237	5 (27, 94, 98, 101, 103)	116	4 (27, 94, 98, 101)	111

[a]PS: Positive selection, [b]NS: Negative selection

Table 4 Genetic distance of ASGV isolates in and between different geographic areas

Subpopulation	N[a]	DWP[b] ± SE	DBP[c] ± SE / Fst[d]		
India	72	0.068 ± 0.005	India		
East Asia	113	0.086 ± 0.006	0.093 ± 0.006 / **0.168**	East Asia	
Eurasia	18	0.019 ± 0.002	0.051 ± 0.004 / **0.152**	0.088 ± 0.007 / 0.394	Eurasia
Americas	9	0.065 ± 0.005	0.093 ± 0.007 / 0.298	0.092 ± 0.007 / **0.189**	0.087 ± 0.008 / 0.526

[a] N = number of ASGV isolates, [b] DWP= Nucleotide diversities (average nucleotide substitutions per site between sequence pairs) within each cluster ± standard error, [c] BDP= Nucleotide diversities between each cluster ± standard error, [d] Fst= values between subpopulations to estimate the extent of gene flow are in parentheses. An absolute value of Fst [< 0.25 suggests infrequent gene flow]

Table 5. Genetic diversity of coat protein gene of ASGV in different geographic area

Populations	N	S	π	d_N	d_S	d_N/d_S	Tajima's D	Fu & Li's D*	Fu & Li's F*
India	72	196	0.067	0.017	0.239	0.069	0.065 ns	-0.274 ns	-0.163 ns
Eurasia	18	89	0.019	0.003	0.073	0.046	-2.042*	-2.987**	-3.149**
East Asia	113	303	0.083	0.021	0.299	0.070	-0.750 ns	-0.839 ns	-0.958 ns
Americas	9	129	0.063	0.017	0.220	0.076	-0.895 ns	-0.896 ns	-1.007 ns
Total	212	351	0.081	0.018	0.293	0.063	-0.972ns	-1.949 ns	-1.733 ns

S: number of polymorphic (segregating) sites, π: nucleotide diversity, d_N: non-synonymous nucleotide diversity, d_S: synonymous nucleotide diversity, ns: non significant, Tajima's D: comparing the estimated number of segregating sites with the mean pairwise difference among sequences

Table S1 Codon usage in different regions of the ASGV genome

Codon usage	Correlation coefficients				
Regions	2	3	4a	4b	5
1	0.227	0.830	0.262	0.745	0.641
2		0.267	0.299	0.364	0.119
3			0.457	0.761	0.737
4a				0.260	0.296
4b					0.537

The codons used in each gene region were counted using BioEdit (Hall, 1999), and the correlations between the codon usage of different gene regions calculated using Excel. Correlation 0.408 $p < 0.001$; degrees of freedom = 63

Table S2 Data on full coat protein gene sequences of Iranian isolates of ASGV characterized in this study

Collection site	Isolate name	Host	Symptoms[c]	Accession numbers
Kamyaran[a]	Ks3	Apple	LD	MK354030
Kamyaran[a]	KZ	Apricot	Y	MK354031
Divandareh[a]	D3	Apple	M	MK354032
Negel[a]	NG	Pear	LD, BLM	MK354033
Sanandaj[a]	SS	Apple	LD	MK354034
Sanandaj[a]	SS4	Apple	LD, VC	MK354035
Khoye[b]	X4	Apple	LC, LD	MK354036

[a] Kurdistan province, Iran; [b]West-Azarbayjan, Iran; [c]LD; leaf deformation, Y; yellowing, M; mosaic, mo; mottle, NS; non-symptoms, BLM; burn leaf margin, VC; vein clearing, LC; leaf curl These symptoms may not referred to ACLSV and ASGV and caused by other agents.

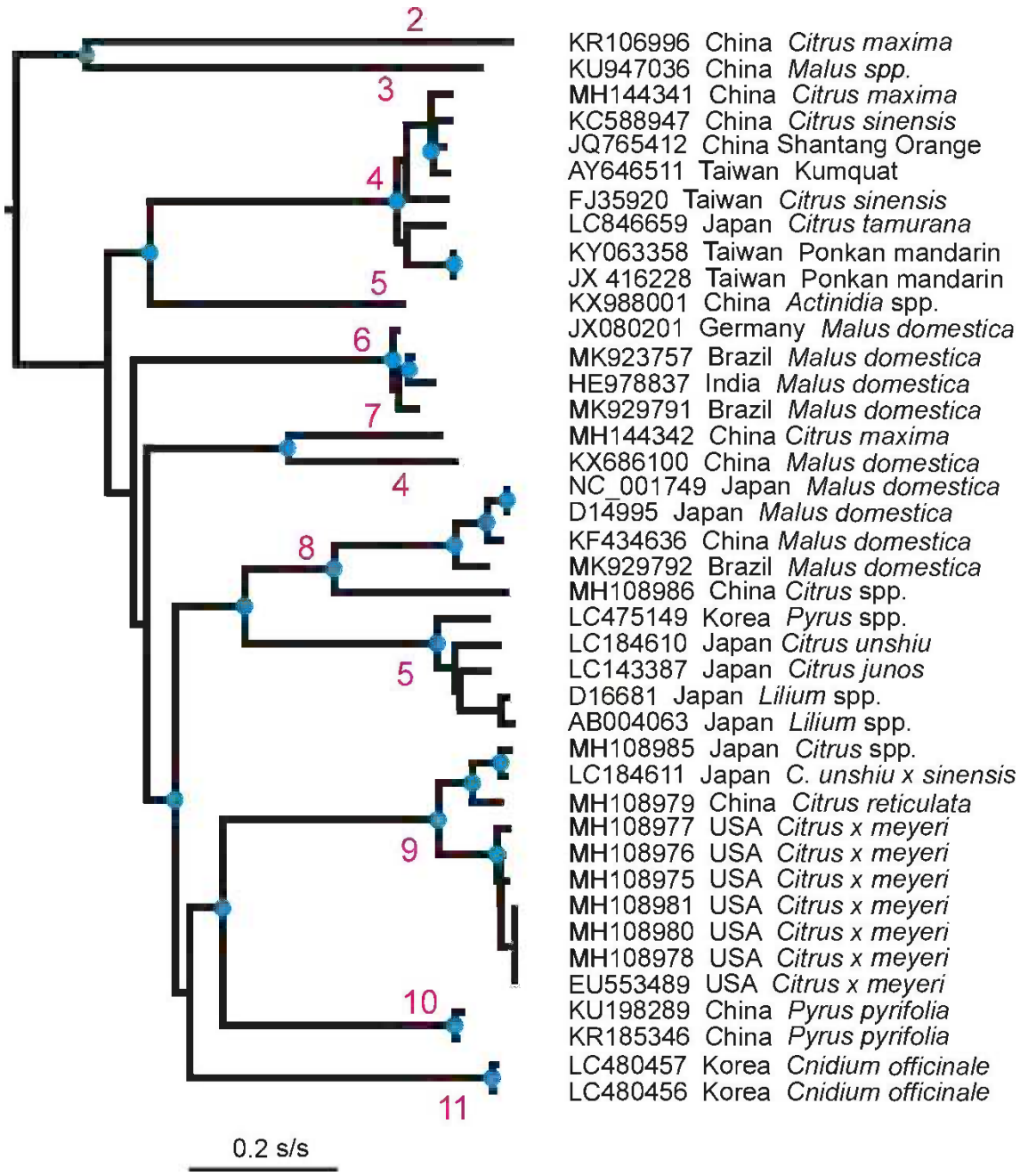
2
3
4
5
6
7
4
8
5
9
10
11
0.2 s/s
KR106996 China Citrus maxima
KU947036 China Malus spp.
MH144341 China Citrus maxima
KC588947 China Citrus sinensis
JQ765412 China Shantang Orange
AY646511 Taiwan Kumquat
FJ35920 Taiwan Citrus sinensis
LC846659 Japan Citrus tamurana
KY063358 Taiwan Ponkan mandarin
JX 416228 Taiwan Ponkan mandarin
KX988001 China Actinidia spp.
JX080201 Germany Malus domestica
MK923757 Brazil Malus domestica
HE978837 India Malus domestica
MK929791 Brazil Malus domestica
MH144342 China Citrus maxima
KX686100 China Malus domestica
NC_001749 Japan Malus domestica
D14995 Japan Malus domestica
KF434636 China Malus domestica
MK929792 Brazil Malus domestica
MH108986 China Citrus spp.
LC475149 Korea Pyrus spp.
LC184610 Japan Citrus unshiu
LC143387 Japan Citrus junos
D16681 Japan Lilium spp.
AB004063 Japan Lilium spp.
MH108985 Japan Citrus spp.
LC184611 Japan C. unshiu x sinensis
MH108979 China Citrus reticulata
MH108977 USA Citrus x meyeri
MH108976 USA Citrus x meyeri
MH108975 USA Citrus x meyeri
MH108981 USA Citrus x meyeri
MH108980 USA Citrus x meyeri
MH108978 USA Citrus x meyeri
EU553489 USA Citrus x meyeri
KU198289 China Pyrus pyrifolia
KR185346 China Pyrus pyrifolia
LC480457 Korea Cnidium officinale
LC480456 Korea Cnidium officinale

Fig. 1

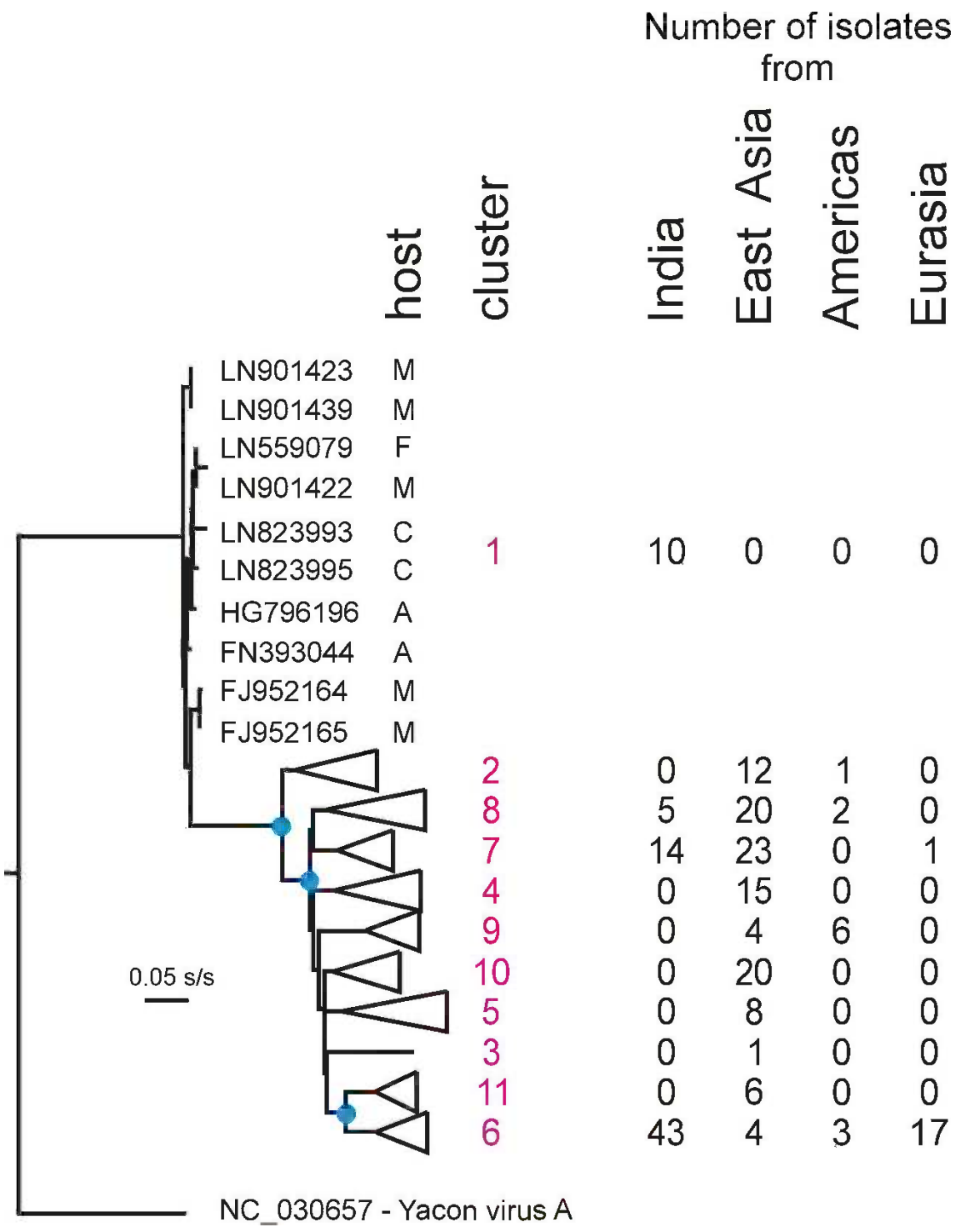
Number of isolates from
host
cluster
India
East Asia
Americas
Eurasia
LN901423 M
LN901439 M
LN559079 F
LN901422 M
LN823993 C
LN823995 C
HG796196 A
FN393044 A
FJ952164 M
FJ952165 M
1 10 0 0 0
2 0 12 1 0
8 5 20 2 0
7 14 23 0 1
4 0 15 0 0
9 0 4 6 0
10 0 20 0 0
5 0 8 0 0
3 0 1 0 0
11 0 6 0 0
6 43 4 3 17
0.05 s/s
NC_030657 - Yacon virus A

Fig. 2

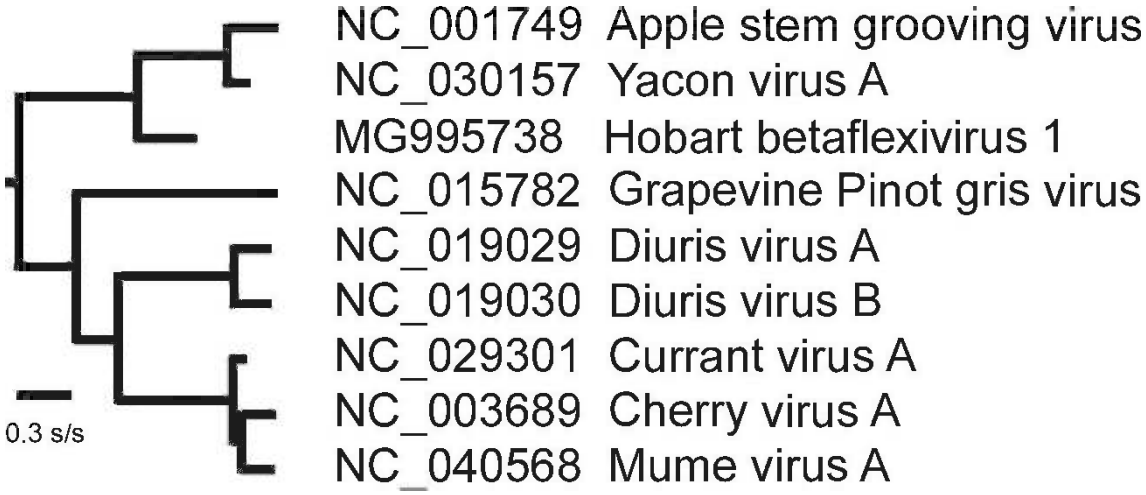

Fig. S1

References

Abascal, F., Zardoya, R., & Telford, M. J. (2010). TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Research, 38*, W7–W13.

Adams, M. J., Lefkowitz, E. J., & King, A. M. Q. (2016). Ratification vote on taxonomic proposals to the International Committee on Taxonomy of Viruses. *Archives of Virology, 161*, 2921–2949.

Alabi, O. J., Al Rwahnih, M., Mekuria, T. A., & Naidu, R. A. (2014). Genetic diversity of grapevine virus a in washington and california vineyards. *Virology, 104*, 548–560.

Balloux, F., & Lugon-Moulin, N. (2002). The estimation of population differentiation with microsatellite markers. *Molecular Ecology, 11*, 55–165.

Chung, C. T., Niemela, S. N., & Muller, R. H. (1989). One step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution. *Proceedings of the National Academy of Sciences, 86*, 2172–2175.

Domingo, E., & Holland, J. J. (1997). RNA virus mutations and ftness for survival. *Annual Review of Microbiology, 51*, 151–178.

Foissac, X., Savalle-Dumas, L., Gentit, P., Dulucq, M. J., & Candresse, T. (2000). Polyvalent detection of fruit tree Tricho, Capillo and Faveaviruses by nested RT-PCR using degenerated and inosine containing primers (PDO RT-PCR). *Acta Horticulturae, 357*, 52–59.

Fourment, M., & Gibbs, M. J. (2006). PATRISTIC: a program for calculating patristic distances and graphically comparing the components of genetic change. *BMC Evolutionary Biology, 6*, 1.

Fu, Y. X., & Li, W. H. (1993). Statistical tests of neutrality of mutations. *Genetics, 133*, 693–709.

Gao, F., Zou, W., Xie, L., & Zhan, J. (2017a). Adaptive evolution and demographic history contribute to the divergent population genetic structure of *Potato virus Y* between China and Japan. *Evolutionary Applications, 10*, 379–390.

Gao, R., Xu, Y., Candresse, T., He, Z., Li, S., Ma, Y., & Lu, M. (2017b). Further insight into genetic variation and haplotype diversity of *Cherry virus A* from China. *PloS ONE, 12*, e0186273.

Gibbs, A. J., Gibbs, M., Ohshima, K., & García-Arenal, F. (2008). More about plant virus evolution; past, present and future. In E. Domingo, C. R. Parrish, J. J. Holland, San Diego (Eds.) *Origin and Evolution of Viruses* (pp. 229–250). Elsevier Academic Press.

Guindon, S., & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology, 52*, 696–704.

Hajizadeh, M., Gibbs, A. J., Amirnia, F., & Ohshima, K. (2019). The global phylogeny of *Plum pox virus* is emerging. *Journal of General Virology, 100*, 1457–1468.

Han, J-Y., Park, C-H., Seo, E-Y., Kim, J. K., Hammond, J., & Lim, H. S. (2016). Occurrence of *Apple stem grooving virus* in commercial apple seedlings and analysis of its coat protein sequence. *Korean Journal of Agricultural Science, 43*, 21–27.

Hirata, H., Lu, X., Yamaji, Y., Kagiwada, S., Ugaki, M., & Namba, S. (2003). A single silent substitution in the genome of Apple stem grooving virus causes symptom attenuation. *Journal of General Virology, 84*, 2579–2583.

Keese, P., & Gibbs, A. (1992). Origins of genes: big bang or continuous creation? *Proceedings of the National Academy of Sciences, 89*, 9489–9493.

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution*, 35, 1547–1549.

Kundu, J. K. (2003). The occurrence of *Apple stem pitting virus* and *Apple stem grooving virus* within field-grown apple cultivars evaluated by RT-PCR. *Plant Protection Science*, 39, 88–92.

Liebenberg, A., Moury, B., Sabath, N., Hell, R., Kappis, A., & Wetrzel, T. (2012). Molecular Evolution of the Genomic RNA of Apple Stem Grooving Capillovirus. *Journal of Molecular Evolution, 75*, 92–101.

Magome, H., Yoshikawa, N., Takahashi, T., Ito, T., & Miyakawa, T. (1997). Molecular variability of the genomes of capilloviruses from apple, Japanese pear, European pear, and citrus trees. *Phytopathology, 87*, 389–396.

Martin, D. P., Murrell, B., Golden, M., Khoosal, A., & Muhire, B. RDP4: (2015). Detection and analysis of recombination patterns in virus genomes. *Virus Evolution, 1*, vev003.

Maxim, A., Zagrai, L., Zagrai, I., & Isac, M. (2004). Studies on the influence of *Apple stem grooving virus* on tree growth of various apple cultivars in the nursery. *Acta Horticulturae, 657*, 41–44.

Nickel, O., Fajardo, T. V., Jelkmann, W., & Kuhn, G. B. (2001). Sequence analysis of the capsid protein gene of an isolate of *Apple stem grooving virus*, and its survey in Southern Brazil. *Fitopatologia Brasileira, 26*, 655–659.

Ohira, K., Namba, S., Rozanov, M., Kusumi, T., & Tsuchizaki, T. (1995). Complete sequence of an infectious full-length cDNA clone of citrus tatter leaf capillovirus: comparative sequence analysis of capillovirus genomes. *Journal of General Virology, 76*, 2305-2309.

Pavesi A. (2019). Asymmetric evolution in viral overlapping genes is a source of selective protein adaptation. *Virology, 532*, 39–47.

Rozas, J. (2009). DNA Sequence Polymorphism Analysis using DnaSP. In D. Posada (Ed.), *Bioinformatics for DNA sequence analysis; methods in molecular biology series* (pp. 337–350). USA: Humana Press.

Shimodaira, H., & Hasegawa, M. (1999). Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution, 16*, 1114–1116.

Souza, E. B., Nickle, O., Fajardo, T. V. M, Silva, J. M. F., & Barros, D. R. (2017). Biological and molecular characterization of two Brazilian isolates of *Apple stem grooving virus*. *Tropical Plant Pathology, 42*, 391–396.

Spengler, R. N. (2019). '*Fruits from the Sands: the Silk Road origins of the foods we eat*'. Oakland, California: University of California Press.

Tajima F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics, 123*, 585–595.

Tatineni, S., Afunian, M. R., Gowda, S., Hilf, M. E., Bar–Joseph, M., & Dawson, W. O. (2009a). Characterization of the 5'–and 3'–terminal subgenomic RNAs produced by a *capillovirus*: Evidence for a CP subgenomic RNA. *Virology, 385*, 521–528.

Tatineni, S., Afunian, M. R., Hilf, M. E., Gowda, S., Dawson, W. O., & Garnsey, S. M. (2009b). Molecular characterization of *Citrus tatter leaf virus* historically associated with Meyer Lemon trees: complete genome sequence and development of biologically active in vitro transcripts. *Phytopathology, 99*, 423–431.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., & Higgins, D. G. (1997). The Clustal_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research, 25*, 4876–4882.

Wang, Y., Zhuang, H., Yang, Z., Wen, L., Wang, G., & Hong, N. (2018). Molecular characterization of an *Apple stem grooving virus* isolate from kiwifruit (*Actinidia chinensis*) in China. *Canadian Journal of Plant Pathology, 40*, 76–83.

Weaver, S., Shank, S. D., Spielman, S. J., Li, M., Muse, S. V., & Pond, S. L. K. (2018). Datamonkey 2.0: A Modern Web Application for Characterizing Selective and Other Evolutionary Processes. *Molecular Biology and Evolution, 35*, 773–777.

Yasaka, R., Fukagawa, H., Ikematsu, M., Soda, H., Korkmaz, S., & Golnaraghi, A. (2017). The time of emergence and spread of turnip mosaic potyvirus. *Scientific Reports, 7*, 4240.

Yoon, J. Y., Joa, J. H., Choi, K. S., Do, K. S., Lim, H. C., & Chung, B. N. (2014). Genetic diversity of a natural population of apple stem pitting virus ısolated from apple in korea. *The Plant Pathology Journal*, *30*, 195–199.

Yoshikawa, N., Sasaki, E., Kato, M., & Takahashi, T. (1992). The nucleotide sequence of apple stem grooving capillovirus genome. *Virology*, *191*, 98–105.

Yoshikawa, N., & Takahashi, T. (1998). Properties of RNAs and proteins of Apple stem grooving and Apple chlorotic leaf spot viruses. *Journal of General Virology, 69*, 241–245.

Zohary, D., Hopf, M., & Weiss, E. (2011). *Domestication of Plants in the Old World: The origin and spread of domesticated plants in south-west Asia, Europe, and the Mediterranean Basin*. UK: Oxford Unversity Press.