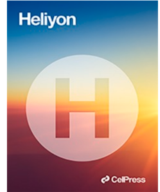
Heliyon 9 (2023) e17006

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Genetic diversity analysis of lychnis mottle virus and first identification of *Angelica sinensis* infection

Weijie Jin [a,b,c], Yubao Zhang [a,c,*], Xuesi Su [a,b,c], Zhongkui Xie [a,c], Ruoyu Wang [a,c], Zhongpei Du [d], Yajun Wang [a,c], Yang Qiu [a,c]

[a] *Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou, 730000, China*
[b] *University of Chinese Academy of Sciences, Beijing, 100049, China*
[c] *Gansu Gaolan Field Scientific Observation and Research Station for Agricultural Ecosystem, Lanzhou, 730000, China*
[d] *Jiuquan Vocational and Technical College, Jiuquan, 735000, China*

ARTICLE INFO

Keywords:
Angelica sinensis (Oliv.) Diels
Lychnis mottle virus
Coat protein
Virus population
Genetic diversity

ABSTRACT

Gansu Province is a district renowned for the cultivation of *Angelica sinensis* (Oliv.) Diels, accounting for greater than 90% of China's total annual production. However, virus infection has caused a reduction in *A. sinensis* yield. Here, we collected suspected virus-infected *A. sinensis* leaf samples from Gansu Province's *A. sinensis* cultivation area. For the first time, using small RNA deep sequencing and RT-PCR, lychnis mottle virus (LycMoV) was found to naturally infect *A. sinensis.* The coat protein (*cp*) gene of the Gansu *A. sinensis* LycMoV isolate was obtained through cloning, where its nucleotide and amino acid identity was highest while having the closest affinity to the China Pearl (i.e., *Prunus persica*) isolate. Recombination analysis indicated that genetic recombination had only a limited influencing effect on the molecular evolution of LycMoV. Moreover, results from genetic diversity analysis indicated that the host, geographic isolation, and genetic drift may be the main factors that contributed to the formation of genetic diversity and differentiation in LycMoV. Furthermore, the LycMoV population trend was expansionary. Selection pressure may also be the main driver for the evolution of the entire LycMoV population, while the driving effect of genetic recombination is limited. This study marks a new LycMoV host (i.e., *A. sinensis*) for the first time and provides scientific support for the identification, prevention, and control of LycMoV.

1. Introduction

Angelica sinensis (Oliv.) Diels is a perennial herb of the *Angelica* L. genus in the family *Umbelliferae* [1]. Along with the rapid increase in the demand for Chinese herbs, the annual demand for *A. sinensis* has exceeded 30,000 t [2]. *A. sinensis* is cultivated in Minxian, Weiyuan, Tanchang, and Zhangxian counties and surrounding areas in Gansu Province, China, which is commonly considered the "district road" of *A. sinensis* production, whose planting area exceeds >90% of that of the entire country [3]. Plant virus infections can affect the growth of *A. sinensis* stem and leaf components and ultimately lead to plant death. There are three known plant viruses that infect *A. sinensis*: tomato mosaic virus (ToMV) [4], konjak mosaic virus (KoMV) [5], and Japanese hornwort mosaic virus (JHMV) [2]. Plant viruses are known as plant "cancers" [6]. Although no effective control measures are available, strengthening quarantine

* Corresponding author. Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou, 730000, China.
E-mail address: zyubao@lzb.ac.cn (Y. Zhang).

https://doi.org/10.1016/j.heliyon.2023.e17006
Received 27 August 2022; Received in revised form 13 April 2023; Accepted 3 June 2023
Available online 5 June 2023
2405-8440/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

measures and eliminating vectors are two typical approaches often considered [7]. Therefore, the accurate and rapid identification of all viruses and virus-like components in samples is an important prerequisite for disease control and basic research. The advantage of small RNA sequencing is that it does not solely rely on known virus information. Moreover, this technique is gradually becoming routine in plant virus detection [8].

Lychnis mottle virus (LycMoV) belongs to an unclassified genus of the *Secoviridae* family. According to the International Committee on Taxonomy of Viruses (ICTV), the cnidium vein yellowing virus (CnVYV) is an isolate of LycMoV [9]. This virus possesses two RNA strands, namely, RNA1 and RNA2, each containing an open reading frame. RNA1 encodes protease cofactors, RNA helicases, viral genome-linked proteins (VPgs), and helicases, while RNA2 encodes RNA-dependent RNA polymerase (RdRP) and movement proteins, large coat proteins (CPs), and small CPs. Since the initial discovery of LycMoV in *Lychnis cognata* Maxim. in Korea [10], the virus was found to be capable of infecting ornamental plants (e.g., peonies and lilies), agricultural plants (e.g., alfalfa) [11], and medicinal plants (e.g., *Ligusticum chuanxiong* Hort) [12,13]. The host range suggests that the virus has the potential to cause significant economic loss.

The mutation rate of RNA viruses is high, subsequently facilitating the adaptation and survival of the virus in different hosts and under different environments. For epidemic monitoring and the ecological control of plant virus diseases, it is necessary to analyze both host and regional differences in viral populations to reveal the genetic structure of plant viral populations as well as their dynamic evolutionary patterns. Studies have shown that the population structure of LycMoV in the United States of America (which is currently expanding) is subject to negative selection [12]. However, this finding has yet to be reported elsewhere. It is critical that we investigate LycMoV genetics, variability, and causes for variation and epidemiological patterns during field occurrences. This will provide us a theoretical basis for the development of long-term sustainable prevention and control strategies, such as the establishment of specific assays to prevent false negatives, the determination of necessary control requirements based on changes in viral infectivity and the selection and breeding of resistant varieties.

2. Materials and methods

2.1. Sample collection and small RNA sequencing

In July 2019, *A. sinensis* leaf samples were collected from fields in Zhangxian, Tanchang, and Zhangxian counties, Gansu Province, China, suspected of being infected with viral disease. Seven *A. sinensis* samples were mixed into a composite sample and entrusted to Sangon Biotech (Shanghai) Co., Ltd. for small RNA sequencing (sRNA-Seq) using the Illumina sequencing platform. RNA was extracted from the mixed samples using the Total RNA Extraction (Trizol) Kit (Sangon Biotech (Shanghai) Co., Ltd.). Small RNA-Seq libraries were prepared using the TruSeq™ Small RNA Sample Prep Kit (Illumina, San Diego, USA). The constructed libraries were then sequenced using Illumina's HiSeq 2000/2500 sequencing system, applying a (single-end) x 50 bp read length. The Cutadapt tool was used to preprocess downstream sequences and to quality score preprocessing methods to raw sequencing reads, including the removal of pre-spliced sequences and low-quality sequences. Qualified reads were spliced using the SPAdes algorithm [14] with a *k*-mer value of 17. Assembly results were compared with the National Center for Biotechnology Information (NCBI) NT database to screen the sequences used to compare viruses.

2.2. Primer design

The primer design was based on the LycMoV *cp* gene sequence. Primer Premier 5 software was used to design the specific primers: LycMoV-F 5′-ACTGAGTGGGGGCGGTTTTAT-3′ and LycMoV-R 5′-GGGGACGGATCTGGGATGGTA-3′, which were used to verify LycMoV infection. Primers LycMoV-CP-F5′-GGTAATCATCATGAAGAGCTGGTTC-3′ and LycMoV-CP-R 5′-CCACCAACAAAAGTGTTCTCA-3′ were used to amplify the LycMoV *cp* gene. The primers were synthesized at Sangon Biotech (Shanghai) Co. Ltd.

2.3. Total RNA extraction and RT-PCR

RNA was extracted from *A. sinensis* leaf samples that were suspected of being infected with the virus using the total RNA extraction kit (Tiangen Biochemical Technology (Beijing) Co. Ltd.), while the 1st strand cDNA was synthesized through reverse transcription using the PrimeScript II 1st strand cDNA Synthesis Kit (Takara Biomedical Technology Co., Ltd.) Reverse Transcription (RT)- polymerase chain reaction (PCR) was used to amplify *cp* genes. Amplification conditions were as follows: pre-denaturation at 94 °C for 4 min, denaturation at 94 °C for 30 s, annealing at 58 °C (LycMoV-F/LycMoV-R) and 62 °C (LycMoV-CP-F/LycMoV-CP-R) for 45 s, extension at 72 °C for 90 s, 35 amplification cycles, and extension at 72 °C for 10 min. PCR products were detected using 1.5% agarose gel electrophoresis. Target fragments were recovered using a DNA recovery kit (i.e., the Agarose Gel DNA Fragment Recovery Kit), the purified fragments were ligated into pMD19-T vectors (Takara Biomedical Technology Co., Ltd.) and transformed into *Escherichia coli* DH5α cells (Takara Biomedical Technology Co., Ltd.). Positive colonies were screened on an ampicillin screening medium, and the recombinant plasmids were then extracted. Positive plasmids were verified using PCR and Sanger sequencing (Tsingke Biotechnology Co. Ltd). The BLAST algorithm was used to compare sequencing results.

2.4. Recombinant analysis

Recombinant analysis is the basis of all evolutionary analyses. Phylogenetic and natural selection analyses can only be performed after first removing recombinant isolates. Accordingly, SplitsTree (v. 4.13.1) was first used to construct a phylogenetic network and to

calculate pairwise homoplasy index (PHI) [15]. Then, the seven tools provided in RDP5 software [16] (i.e., RDP, GENECONV, BOOTSCAN, MAXCHI, Chimera, SISCAN, and 3SEQ) were used to detect recombination events based on the viral *cp*gene dataset. Default values were used for all parameters, and a Bonferroni-corrected *P*-value of 0.05 was used to evaluate recombination events detected in greater than half of four out of at least seven of the tools as valid.

2.5. Construction of evolutionary tree

All LycMoV *cp* sequences, including that of the LycMoV *A. sinensis* isolate which was amplified in this study, were retrieved from the NCBI. A total of 61 sequences (i.e., totaling 60 after removing the recombinant isolates) were obtained. These sequences were used to analyze LycMoV phylogenetic relationships. Multiple sequence comparisons were performed using MEGA7 software. The most accurate nucleotide and amino acid substitution model was selected using ModelFinder in IQ-TREE software. The tree was constructed using the maximum likelihood estimation method [17,18]with a spanning value of 1000.

2.6. Population genetic analysis

Nucleotide diversity (π), number of segregation sites (S), total number of mutations (η), average number of nucleotide differences between sequences (K), haplotype diversity (Hd), number of non-synonymous mutations (Ka), synonymous mutations (Ks), and the ratio of non-synonymous to synonymous mutations (Ka/Ks) were calculated using DnaSP software [19]. Tajima's D, Fu & Li's D, and Fu & Li's F were calculated to test whether they fit the neutral evolutionary model. When all three parameter values are positive, the population is stably selected and the number is in decline. A negative value means that the population is negatively selected and the population is expanding following a recent bottleneck [20,21].

2.7. Population difference test

Four independent population differences were verified for the Ks*, Kst*, Z*, and Snn *cp* gene sequences. Genetic differences between different populations were tested for significance by means of repeated sampling (1000 times) using a permutation test. Fst is the genetic differentiation coefficient, used to measure the degree of diversity among populations, ranging from zero (0) (indicating no differentiation among populations) to a theoretical maximum of one (1) (where populations show significant differentiation) [15,22,23].

2.8. Gene flow analysis

The Fst parameter can also be used to estimate the degree of genetic exchange among populations. An absolute Fst value greater than 0.33 indicates infrequent genetic exchanges between two populations, while the opposite indicates frequent genetic exchanges [24]. The Nm parameter can be used to assess gene flow and genetic drift. An Nm value of less than one (1) indicates that the population is prone to genetic drift. Genetic drift is the main cause of genetic differentiation in a population. An Nm value greater than one (1) indicates no genetic drift between two populations, where gene flow can occur through an unknown channel.

3. Results

3.1. Small RNA sequencing results and validation

Samples were sequenced to obtain 1199561625 raw_bases and 281093971 clean_bases by removing unqualified reads. After assembly using the SPAdes algorithm, 56 were obtained and compared with the NCBI's NT database, wherein seven counts were homologous to LycMoV (Table 1). These specific primers were designed according to matched counting, and PCR amplification was performed to verify the target band of 1654 bp (Fig. 1S). Target bands were verified to be the expected size. Sequencing results were compared using BLAST on the NCBI data, showing 85.06% nucleotide similarity to LycMoV (LC579913.1), which infects *Prunus persica* (China Pearl). LycMoV infection was then confirmed in *A. sinensis* leaf samples and was determined to be natural infection. LycMoV infected *A. sinensis* displayed yellow blotches and mottled symptoms (Fig. 1B and C).

3.2. Cloning and sequencing of cp genes

The cDNA obtained by reverse transcription was used as a template for PCR to amplify the full length of the LycMoV *cp* genes, Additionally, electrophoresis showed that the amplified fragment was between 1000 and 2000 bp, which was consistent with the

Table 1
Statistical sample Quality Control results.

Raw_num	Raw_bases	Clean_bases	Mean_len	Contig	Contig belonging to LycMoV
15994155	1199561625	281093971	21.81	56	7

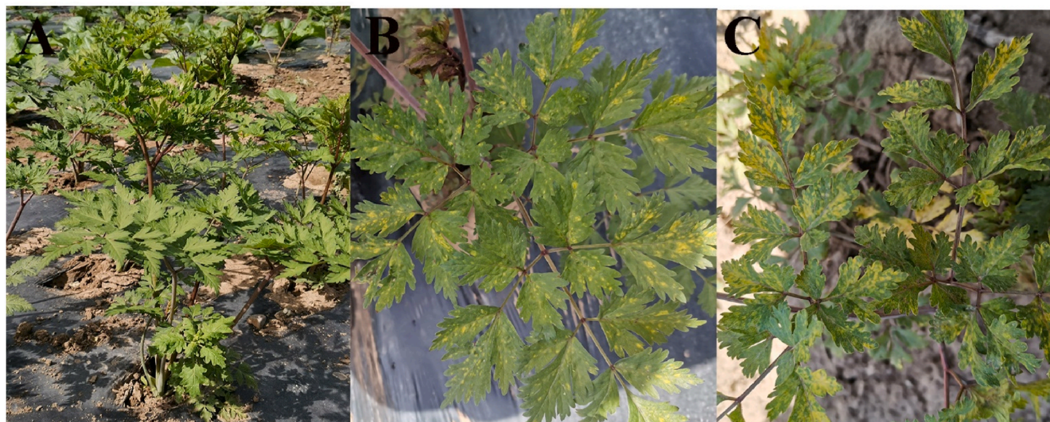


Fig. 1. (A) Healthy *Angelica sinensis* plants; (B and C) *A. sinensis* plants exhibiting virus-like symptoms of leaf mosaic, mottling, and chlorosis, which have been found positive for lychnis mottle virus infection.

expected size of 1887 bp (Fig. 2S). Sequences obtained by sequencing were submitted to the NCBI database (GenBank accession number: OQ185382).

3.3. Genetic recombination

The split network showed evidence of a reticular structure; however, PHI = 0.083 (PHI below a 0.05 threshold ($\Phi w < 0.05$) indicates that there is significant recombination present in the dataset), indicating that no verifiable significant recombination took place (Fig. 3S). On the other hand, a potential recombinant MW035178.1 (Arkansas (USA)/*P. lactiflora* cv. 'Coral N′ Gold) was identified among the 61 isolates using recombinant RDP5 software analysis. Secondary and primary parents were MW035151.1 (Michigan (USA)/*P. lactiflora* cv. La Perle) and MW035153.1 (Michigan (USA)/*P. lactiflora* cv. Illini Belle), respectively. The recombination loci were 218–942 (95% confidence interval) (Table 2).

3.4. Genetic diversity analysis of LycMoV

A phylogenetic tree was constructed using IQ-TREE and referred to the BIC criteria during GTR + F + I + G4 model selection. A total of five branches were clustered, and the LycMoV Gansu *A. sinensis* isolate (GSDG) was clustered with Chinese, Korean, and Japanese isolate branches. The isolates from Michigan (USA), Alaska (USA), and New York (USA) were combined, and the isolate from Arkansas (USA) was split into two. MW035190.1 (USA: Alaska) was clustered into one (Fig. 2). Nucleotide and amino acid identity analyses of *cp* genes (detected using BioAider) showed that the nucleotide identity of the 60 LycMoV isolates ranged from 74.68% to 99.52%, and amino acids ranged from 89.49% to 100% (Fig. 3A). The isolates obtained in this study showed the highest nucleotide and amino acid identity to LC579920.1 (China/*P. persica*), with 84.77% and 96.5% sequence identity, respectively. Nucleotide identity with BK011047.1 (China/*Medicago sativa*) was the lowest at 77.65%, but amino acid identity was higher at 93.63%. The lowest amino acid identity (92.99%) (Fig. 3B) was found with MW035176.1 (USA/Arkansas *Paeonia lactiflora* cv. 'Coral N′ Gold) and MW035144.1 (USA: New York/*P. lactiflora*). Although these findings indicated that CPs exhibited extensive mutations in nucleotides during speciation, most were synonymous mutations, and that CP proteins were highly conserved in their evolution.

3.5. Characteristics and diversity of LycMoV populations

To explore LycMoV population diversity, we divided LycMoV *cp*gene sequences based on the geographic location (China, Japan/Korea, Arkansas, Alaska, Michigan, and New York) of six populations. Because one isolate from Japan could not be grouped with certainty, it was grouped with Korea. Haplotype diversity (Hd) and nucleotide diversity (π) were greater than 0.500 and 0.005, respectively, in all six populations investigated. These results indicated high genetic diversity within populations. The highest mean

Table 2
Detection of recombination event within *cp* gene.

Recombinant sequence(s)	Minor parental sequence(s)	Major parental sequence(s)	*P*-value of the six detection methods in RDP						
			3seq	RDP	GENECONV	Bootscan	Maxchi	Chimaera	SiSscan
MW035178.1	MW035151.1	MW035153.1	6.620×10^{-03}	NS	3.634×10^{-03}	4.964×10^{-04}	2.856×10^{-04}	NS	5.130×10^{-04}

Note: NS: Non-significant.

Fig. 2. The phylogenetic tree of lychnis mottle virus (LycMoV) isolates based on nucleotide sequences of the *cp* genes (where the GenBank accession number of the LycMoV sequence of infected A. sinensis is OQ185382.1).

number of *cp* nucleotide differences (K = 363.000) and nucleotide diversity (π = 0.19268) were found in the Chinese population, having the highest number of segregating loci (S = 687) and the highest number of mutations (η = 884). These results indicated that the genetic diversity of the Chinese population was comparatively the highest. The Alaska population had the least number of segregating loci (S = 60) and the least number of mutations (η = 60). The Michigan population exhibited the lowest mean nucleotide differences (K = 36.3810) and nucleotide diversity (π = 0.01982). Haplotype diversity (Hd = 0.905) was lowest in ROK and JPN, while Hd = 1 for all remaining populations. Moreover, Ka/Ks values of the six populations ranged from 0.02356 to 0.08844, indicating that predominant synonymous mutations were subjected to negative (or purifying) selection. The Alaska population had the lowest Ka/Ks (0.02356) value and was under the least selection pressure. The New York population had the highest selection pressure with a Ka/Ks value of 0.08844 (Table 3).

Statistical analysis showed that all three parameters of the Korean population had values greater than zero (0) indicating that the Korean and Japan population is more selectively stable while being in decline. For the USA populations, Tajima's D, Fu & Li's D, and Fu & Li's F were all less than zero (0). Additionally, Tajima's D and Fu & Li's F were less than zero (0) for the Chinese populations, indicating that these populations could potentially be expanding (Table 4).

3.6. Genetic differences and gene flow analysis of LycMoV populations

In this study, Snn *P*-values were equal to 1 corresponding Ks*, Kst*, Z*, and Snn parameters were all less than 0.01 among the Chinese and various other populations, indicating significant differences. These results demonstrated considerable genetic differences

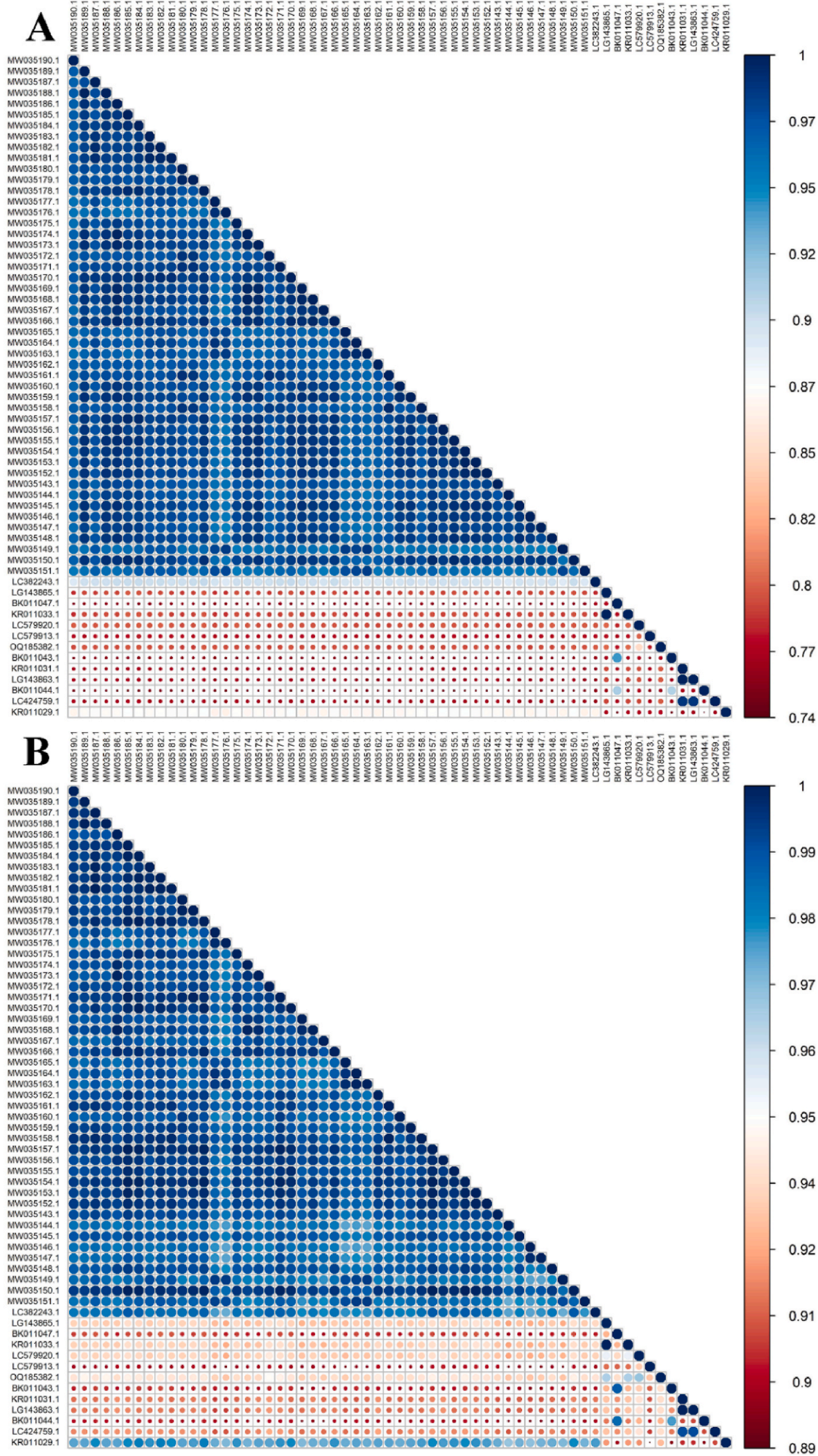


Fig. 3. Two-dimensional graphic representation of the percentage of the pairwise identity plot of (A) the nucleotide sequence and (B) the amino acid sequence of lychnis mottle virus (LycMoV) *cp* genes (where the GenBank accession number of the LycMoV sequence of infected *A. sinensis* is OQ185382.1).

Table 3
Genetic diversity parameters estimated for CP genes of the lychnis mottle virus.

population	M	η	K	π	Hd	SS	NS	Ks	ka	Ka/Ks
ALL	60	1279	179.318	0.09584	0.999	465.89	1403.11	0.31517	0.02294	0.07279
China	6	884	363.000	0.19268	1.000	471.14	1412.86	0.64067	0.04323	0.06748
Japanese and Korean	7	800	335.000	0.17781	0.905	469.88	1414.12	0.58190	0.04353	0.07481
Alaska	3	60	40.000	0.02123	1.000	467.78	1416.22	0.07979	0.00188	0.02356
Arkansas	30	192	41.669	0.02222	1.000	467.31	1407.69	0.07691	0.00407	0.05292
Michigan	7	108	36.381	0.01982	1.000	469.05	1414.95	0.06709	0.00346	0.05157
New York	7	137	46.667	0.02489	1.000	465.95	1406.05	0.07903	0.00699	0.08844

M: number of sequences used; η: total number of mutations; K: average number of nucleotide differences; π: nucleotide diversity; Hd: haplotype (gene) diversity; SS: synonymous sites. Number of sites; NS: nonSynonymous sites; Ka: number of non-synonymous mutations; Ks: synonymous mutations; Ka/Ks: the ratio of non-synonymous to synonymous mutations.

among the Chinese population and the other populations. However, Snn *P*-value parameters between the Japanese/Korean and the Alaska populations were not significant, but Ks*, Kst*, and Z* *P*-values were all less than 0.01. The *P*-values of Ks*, Kst*, Z*, and Snn among the Japanese/Korean population and the Alaska, Michigan, and New York populations were all less than 0.01. These results also demonstrated large genetic differences between the Japanese/Korean and the USA populations. The *P*-values of all four Arkansas, Michigan, and New York population parameters were significant, indicating a large genetic gap. Only the Snn parameter *P*-value was significant between the Alaska population and the New York population. Furthermore, Kst* was greater than zero (0) and Snn was greater than 0.5 among the New York population and the Alaska and Michigan populations, indicating genetic differentiation; however, the *P*-value was greater than 0.01 and was therefore not significant. The genetic differentiation coefficient (i.e., Fst) between the New York and Michigan populations was close to zero (0), and none of the four parameters significantly correlated (Table 5).

The absolute Fst values between the Chinese and Korean/Japan populations and among the USA populations were all less than 0.33, while the Nm value was greater than one (1). This was indicative of the frequent gene exchanges that have occurred among these populations. However, there was no evidence of genetic drift, and gene flow may occur through other channels. The absolute Fst values among the USA, Japanese/Korean, and Chinese populations were all greater than 0.33, which was indicative of infrequent gene exchange. However, the Nm value was less than one (1), which was potentially indicative of genetic drift (Table 6).

4. Discussion and conclusions

High-throughput sequencing methods are uniquely suited to study the distribution patterns and genetic evolution of viruses [25]. In this study, LycMoV was detected in *A. sinensis* for the first time by the small RNA sequencing, where in the corresponding *cp* gene sequence was cloned and upon which *A. sinensis* was determined to be a new LycMoV host. The isolates obtained for this study had the highest nucleotide and amino acid identity with *P. persica* (China Pearl) isolates. In the phylogenetic tree, *A. sinensis* and P. persica populations were also the most closely related.

Results from haplotype and nucleotide diversity in *cp* gene analysis indicated that the entire LycMoV population is rich in genetic resources and is high in genetic diversity, facilitating the virus to adapt to different hosts and environments. The LycMoV population may have evolved from a large and stable population over an extended period or from secondary contact between the populations of two different lineage groups (Grant & Bowen, 1998). Moreover, genetic diversity parameters (i.e., S, η, K, and π) were significantly higher in the Chinese and Japanese/Korean populations than in the USA populations. This could may be due to either regional geographic differences or to host differences, given that *P. lactiflora* was the host of all USA isolates. Chinese and Japanese/Korean population hosts included a variety of plants, such as lilies, alfalfa, and *A. sinensis*, suggesting that host differences could potentially be one of the reasons for the formation of genetic diversity in LycMoV populations. In this study, only one recombinant isolate was observed among 61 *cp* genes, and the level of significance of two recombinant isolates was negligible. This result could potentially be related to the fact that *cp* gene sequences are evolutionarily conserved and shorter compared to the whole genome, limiting the production of certain recombinant isolates. However, to some extent, this result could also indicate that the driving effect of recombination on the molecular evolution of LycMoV is limited. Selection pressure analysis suggested that negative selection pressure is the main driving factor for LycMoV evolution. From neutral assay analysis results, the Japanese/Korean population was found to be more stable in selection than the Chinese population while its size is on the decline. Conversely, other populations, including those in Gansu Province, China, are increasing. Since LycMoV was first reported in Korea, it has widely been speculated that the first LycMoV

Table 4
Neutrality tests of the LycMoV population from a pairwise comparison of lychnis mottle virus sequences.

Population	Tajima's D	Fu & Li's D	Fu & Li's F
ALL	−1.24091	0.28941	−0.38531
China	−0.40640	0.00222	−0.09368
Japanese and Korean	0.15286	0.66416	0.60950
Arkansas	−0.54234	−0.65797	−0.73293
Michigan	−1.02105	−1.13948	−1.23362
New York	−0.96899	−0.93120	−1.04198

Table 5
Genetic differentiation measurements between subpopulations of lychnis mottle virus sequences and the summary of test statistics examined for permutation probability.

Population	Ks*	Kst*	Ks*, Kst*P-value	Z*	Z* p-value	Snn	Snn *P*-value
China/Japanese and Korean	5.46232	0.05315	0.0120 *	3.14249	0.0130 *	1.00000	0.0020 **
China/Alaska	5.30817	0.07360	0.0000 ***	2.21976	0.0220 *	1.00000	0.0350 *
China/Arkansas	3.92104	0.10699	0.0000 ***	5.22617	0.0000 ***	1.00000	0.0000 ***
China/Michigan	4.45094	0.15866	0.0000 ***	2.83773	0.0010 **	1.00000	0.0010 **
China/New York	4.65804	0.13578	0.0010 **	2.83531	0.0010 **	1.00000	0.0010 **
Japanese and Korean/Alaska	4.85492	0.09592	0.0220 *	2.66236	0.0220 *	0.80000	0.0930 ns
Japanese and Korean/Arkansas	3.88025	0.11701	0.0000 ***	5.18916	0.0000 ***	0.94595	0.0000 ***
Japanese and Korean/Michigan	4.26471	0.16757	0.0000 ***	3.07001	0.0000 ***	0.85714	0.0140 *
Japanese and Korean/New York	4.45069	0.14628	0.0000 ***	3.04314	0.0000 ***	0.85714	0.0060 **
Alaska/Arkansas	3.63597	0.00716	0.0920 ns	5.21460	0.0620 ns	0.96970	0.0080 **
Alaska/Michigan	3.33785	0.02820	0.1210 ns	2.74677	0.1920 ns	0.85000	0.0290 *
Alaska/New York	3.65073	0.00399	0.3630 ns	2.85995	0.3710 ns	0.70000	0.2110 ns
Arkansas/Michigan	3.60210	0.01284	0.0140 *	5.38820	0.0080 **	0.92793	0.0010 **
Arkansas/New York	3.66029	0.01102	0.0180 *	5.41872	0.0230 *	0.91892	0.0000 ***
Michigan/New York	3.54220	0.00939	0.1690 ns	3.42835	0.0690 ns	0.71429	0.0760 ns

Note: Ks*, Kst*, Ks*, Z* and Snn are test statistics of genetic differentiation; Significance thresholds: *, $0.01 < P\text{-value} < 0.05$; **, $0.001 < P\text{-value} < 0.01$; ***, $P\text{-value} < 0$.

Table 6
Estimate of gene flow among lychnis mottle virus populations.

Population	Nm	Fst
China/Japanese and Korean	1.18	0.17434
China/Alaska	0.22	0.53260
China/Arkansas	0.22	0.53016
China/Michigan	0.21	0.53791
China/New York	0.22	0.52671
Japanese and Korean/Alaska	0.26	0.48864
Japanese and Korean/Arkansas	0.26	0.48658
Japanese and Korean/Michigan	0.26	0.49012
Japanese and Korean/New York	0.27	0.48139
Alaska/Arkansas	2.31	0.09775
Alaska/Michigan	4.77	0.04976
Alaska/New York	56.87	0.00438
Arkansas/Michigan	2.39	0.09469
Arkansas/New York	3.00	0.07688
Michigan/New York	−27.25	−0.00926

Note: Nm: gene flow; Fst: genetic differentiation coefficient.

outbreak must have occurred in that country.

Population genetic differentiation and gene flow analysis showed significant genetic differences among the Chinese, Japanese/Korean, and USA populations. Results showed a significant genetic difference in populations subject to infrequent gene exchange, which were susceptible to genetic drift. However, the genetic differentiation between the Chinese and Japanese/Korean populations was less than that of the USA populations, which indicated frequent gene exchange and lower genetic drift susceptibility. Genetic differences were significant between the Arkansas and Michigan populations and between the Arkansas and New York populations. However, no genetic differences were observed between the New York and Michigan populations. Collectively, geographic analysis suggested that geographic isolation and genetic drift could be responsible for the differences observed among these populations. Although gene exchange between the virus has frequently occurred among the various USA populations, no genetic drift has been was detected, which could potentially be host related. Peonies, as an ornamental plant, are widely distributed throughout the USA and are mostly propagated through grafting. Virus could therefore be carried from parent to progeny.

For the first time, this study has confirmed that LycMoV is able to infect *A. sinensis*, a popular medicinal plant. Results from genetic diversity analysis indicated that the host, geographic isolation, and genetic drift could be the main factors that have contributed to the formation of genetic diversity and differentiation of LycMoV, and that selection could may be the main driver for LycMoV population evolutionary processes as a whole. Results from this study provide a basis for LycMoV identification and control measures, which can be used in designing long-term, sustainable LycMoV management strategies.

Author contribution statement

Weijie Jin: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Yubao Zhang: Conceived and designed the experiments; Contributed reagents, materials, analysis tools or data.
Xuesi Su, Yang Qiu: Performed the experiments.
Zhongkui Xie, Ruoyu Wang, Zhongpei Du, Yajun Wang：Contributed reagents, materials, analysis tools or data.

Data availability statement

Data associated with this study has been deposited at the NCBI database under the GenBank accession number: OQ185382.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (Grant No. 32072394), the Chinese Academy of Sciences Regional Development Projects for Young Scholars (Grant No. 2022-01), the Crop Varietal Improvement and Insect Pests Control by Nuclear Radiation (Grant No. E190050401), the Gansu Province Key Research and Development Program (Grant No. 21YF5NA136), the Gansu Province Science and Technology Plan project (Grant No. 20JR10RA071) and the Jiuquan Key Technology R&D Program (Grant No. 2022CA1009).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2023.e17006.

References

[1] X. Zhang, Y. Peng, Y. Wang, Z. Zhang, D. Li, J. Yu, C. Han, Simultaneous detection and differentiation of three genotypes of Brassica yellows virus by multiplex reverse transcription-polymerase chain reaction, Virol. J. 13 (2016).
[2] Y. Zhang, R. Wang, Z. Xie, Y. Wang, X. Zhao, Y. Liu, Q. Zhou, Rapid visual detection of Japanese hornwort mosaic virus infectingAngelica sinensisby reverse transcription loop-mediated isothermal amplification, Ann. Appl. Biol. 178 (3) (2021) 489–497.
[3] W.-L. Wei, R. Zeng, C.-M. Gu, Y. Qu, L.-F. Huang, Angelica sinensis in China-A review of botanical profile, ethnopharmacology, phytochemistry and chemical analysis, J. Ethnopharmacol. 190 (2016) 116–141.
[4] W. Liu, L. Yan, Z. Wen, X. Chen, Survey of virus diseases on 18 species of medical herbs in Gansu Province and identification of two virus disease, Plant Prot. 40 (5) (2014) 133–137, 163.
[5] Y. Zhang, R. Wang, J. Wang, J. Chang, X. Zhang, T. Chen, L. An, S. Xu, A new potyvirus first isolated and identified from Angelica sinensis, Virus Gene. 39 (1) (2009) 120–125.
[6] J. Chen, X. Luo, Y. Chen, Y. Wang, J. Peng, Z. Xing, Recent research progress: discovery of anti-plant virus agents based on natural scaffold, Front. Chem. (2022) 10.
[7] R.N. Strange, P.R. Scott, Plant disease: a threat to global food security, Annu. Rev. Phytopathol. 43 (2005) 83–116.
[8] G.T. Mehetre, V.V. Leo, G. Singh, A. Sorokan, I. Maksimov, M.K. Yadav, K. Upadhyaya, A. Hashem, A.N. Alsaleh, T.M. Dawoud, K.S. Almaary, B.P. Singh, Current developments and challenges in plant viral diagnostics: a systematic review, Viruses-Basel 13 (3) (2021).
[9] J.R. Thompson, I. Dasgupta, M. Fuchs, T. Iwanami, A.V. Karasev, K. Petrzik, H. Sanfacon, I. Tzanetakis, R. van der Vlugt, T. Wetzel, N. Yoshikawa, I. R. Consortium, ICTV virus Taxonomy profile: Secoviridae, J. Gen. Virol. 98 (4) (2017) 529–531.
[10] R.H. Yoo, F. Zhao, S. Lim, D. Igori, S.-H. Lee, J.S. Moon, The complete nucleotide sequence and genome organization of lychnis mottle virus, Arch. Virol. 160 (11) (2015) 2891–2894.
[11] I.P. Adams, R.H. Glover, W.A. Monger, R. Mumford, E. Jackeviciene, M. Navalinskiene, M. Samuitiene, N. Boonham, Next-generation sequencing and metagenomic analysis: a universal diagnostic tool in plant virology, Mol. Plant Pathol. 10 (4) (2009) 537–545.
[12] C.M. Shaffer, D.C. Michener, N.B. Vlasava, H. Chotkowski, K. Lamour, D. Stainton, I.E. Tzanetakis, The population structure of the secovirid lychnis mottle virus based on the RNA2 coding sequences, Virus Res. (2021) 303.
[13] P. Jiang, J. Shao, L.G. Nemchinov, Identification of emerging viral genomes in transcriptomic datasets of alfalfa (Medicago sativa L.), Virol. J. 16 (1) (2019).
[14] A. Bankevich, S. Nurk, D. Antipov, A.A. Gurevich, M. Dvorkin, A.S. Kulikov, V.M. Lesin, S.I. Nikolenko, P. Son, A.D. Prjibelski, A.V. Pyshkin, A.V. Sirotkin, N. Vyahhi, G. Tesler, M.A. Alekseyev, P.A. Pevzner, SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing, J. Comput. Biol. 19 (5) (2012) 455–477.
[15] R.R. Hudson, D.D. Boos, N.L. Kaplan, A statistical test for detecting geographic subdivision, Mol. Biol. Evol. 9 (1) (1992) 138–151.
[16] D. Martin, E. Rybicki, RDP: detection of recombination amongst aligned sequences, Bioinformatics 16 (6) (2000) 562–563.
[17] L.-T. Nguyen, H.A. Schmidt, A. von Haeseler, B.Q. Minh, IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies, Mol. Biol. Evol. 32 (1) (2015) 268–274.
[18] H. Diep Thi, O. Chernomor, A. von Haeseler, B.Q. Minh, V. Le Sy, UFBoot2: improving the ultrafast bootstrap approximation, Mol. Biol. Evol. 35 (2) (2018) 518–522.
[19] P. Librado, J. Rozas, DnaSP v5: a software for comprehensive analysis of DNA polymorphism data, Bioinformatics 25 (11) (2009) 1451–1452.
[20] F. Tajima, STATISTICAL-METHOD for testing the neutral mutation hypothesis by DNA polymorphism, Genetics 123 (3) (1989) 585–595.
[21] Y.X. Fu, W.H. Li, Statistical tests of neutrality of mutations, Genetics 133 (3) (1993) 693–709.
[22] R.R. Hudson, A new statistic for detecting genetic differentiation, Genetics 155 (4) (2000) 2011–2014.

[23] F. Gao, W. Lin, J. Shen, F. Liao, Genetic diversity and molecular evolution of arabis mosaic virus based on the CP gene sequence, Arch. Virol. 161 (4) (2016) 1047–1051.
[24] F. Rousset, Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance, Genetics 145 (4) (1997) 1219–1228.
[25] J.S. Okonya, H. Gamarra, A. Nduwayezu, A. Bararyenya, J. Kroschel, J. Kreuze, Serological survey and metagenomic discovery of potato viruses in Rwanda and Burundi reveals absence of PVY in Burundi and first report of TRV in potatoes in sub-Saharan Africa, Virus Res. (2021) 302.