Infection, Genetics and Evolution 73 (2019) 167–174

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Research paper

Temporal analysis and adaptive evolution of the global population of potato virus M

Zhen He[a,*], Wen Chen[a], Ryosuke Yasaka[b], Chunfeng Chen[a], Xijun Chen[a]

[a] School of Horticulture and Plant Protection, Yangzhou University, Wenhui East Road No.48, Yangzhou 225009, Jiangsu Province, PR China
[b] The United Graduate School of Agricultural Sciences, Kagoshima University, 1-21-24, Kagoshima 890-0065, Japan

ARTICLE INFO

Keywords:
Potato virus M
Evolutionary rate
Migration
Adaption evolution
Global

ABSTRACT

Potato virus M (PVM), which is a member of the genus *Carlavirus* in the family *Betaflexviridae*, causes critical economic losses of nightshade crops. PVM is transmitted by aphids in a non-persistent manner, by sap inoculation and also transmitted in tubers. Previously, several reports described the genetic structure of PVM. However, the evolutionary rate, timescale, spread and adaptation evolution of the virus have not been examined. In this study, we investigated the phylodynamics of PVM using 145 nucleotide sequences of the coat protein gene and 117 sequences of the cysteine-rich nucleic acid-binding protein (NABP) gene, which were sampled between 1985 and 2013. We found that at least three lineages with isolates that were defined geographically but not by the original host were clustered. The evolutionary rate of the NABP (1.06×10^{-2}) was faster than that of the CP (4.12×10^{-3}). The time to the most recent common ancestors (TMRCAs) is similar between CP (CIs 31–110) and NABP (CIs 28–33) genes. Based on CP and NABP genes, PVM migrated from China to Canada, Iran, India and European countries, and it circulated within China. Our study is the first attempt to evaluate the evolutionary rates, timescales and migration dynamics of PVM.

1. Introduction

The evolution of the plant virus population is shaped by selection, founder effects and recombination (Gibbs and Ohshima, 2010; Gibbs et al., 2015; Ohshima et al., 2002), including not only RNA viruses, for example the turnip mosaic virus (TuMV), potato virus Y (PVY) and sugarcane streak mosaic virus (SCSMV) in the family *Potyviridae* (Cuevas et al., 2012a,b; He et al., 2014, 2016; Ohshima et al., 2002), but also single- or double-strand DNA viruses, such as tomato yellow leaf curl virus (TYLCV) and maize streak virus (MSV) in the family *Geminiviridae* (Duffy and Holmes, 2008; Harkins et al., 2009; Rocha et al., 2013) and cauliflower mosaic virus (CaMV) in the family *Caulimoviridae* (Yasaka et al., 2014). Under various ecological or physiological factors, those plant virus populations may lead to adaptive subdivision. Therefore, studies on the molecular evolution of populations of plant viruses are important for understanding virus emergence, geographical spread and adaptation to new hosts, and for planning control strategies (García-Arenal et al., 2001; Gibbs et al., 2010, 2017; Gibbs et al., 2008; Jones, 2009).

Potato virus M (PVM) is a species of the genus *Carlavirus* in the family *Betaflexviridae*. This virus contains a single-stranded positive-sense RNA molecule, which is appropriately 8.5 kb in length, encapsidated in flexuous filamentous virions (Zavriev et al., 1991). The genome of PVM possesses a cap structure at the 5′ end and a poly(A) tail at the 3′ end and encodes six open reading frames (ORFs) (Flatken et al., 2008; Ge et al., 2012; Zavriev et al., 1991). ORF1 encodes a polyprotein including methyltransferase, helicase, and polymerase domains and is involved in RNA replication (Koonin et al., 1993; Rozanov et al., 1992). The overlapping ORF2, ORF3, and ORF4 encode three putative proteins named triple gene blocks (TGBs), which are 25, 12 and 7 kDa in size, respectively, and are involved in cell-to-cell movement, suppression of RNA silencing and membrane binding (Ju et al., 2005; Morozov et al., 1999; Senshu et al., 2011). ORF5 and ORF6 encode the coat protein (CP) and a cysteine-rich nucleic acid-binding protein (NABP), respectively (Gramstat et al., 1990; Zavriev et al., 1991). In 1923, PVM was first isolated from potato (*Solanum tuberosum*) in the United States (Schultz and Folsom, 1923). Subsequently, PVM was identified from pepino (*Solanum muricatum*), tomato and tobacco (Ge et al., 2012, 2014). At present, PVM has become an economically important pathogen of potato and pepino worldwide (Ge et al., 2014; Su et al., 2017; Tabasinejad et al., 2015, 2014). PVM is transmitted by aphids in a non-persistent manner and also by mechanical inoculation with sap from young infected leaves (Ge et al., 2012; Xu et al., 2010).

Ge et al. (2014) and Tabasinejad et al. (2014) reported the genetic

* Corresponding author.
E-mail addresses: hezhen@yzu.edu.cn (Z. He), st9522@cc.saga-u.ac.jp (R. Yasaka), xjchen@yzu.edu.cn (X. Chen).

https://doi.org/10.1016/j.meegid.2019.04.034
Received 16 March 2019; Received in revised form 28 April 2019; Accepted 30 April 2019
Available online 02 May 2019

structure of PVM populations in China and Iran based on analyses of single gene sequences. Tabasinejad et al. (2015) described the population genetics of PVM on a global scale according to the cysteine-rich NABP gene with limited sequences. To date (March 2018), only eight full genomic sequences of PVM isolates from China, Russia, Iran, Germany, Poland, and the Czech Republic have been reported (Flatken et al., 2008; Ge et al., 2012; Zavriev et al., 1991). However, these studies reported uncertainty about the evolutionary rate, timescale, host and geographical adaptation of the virus around the world.

In this study, we analyzed the genetic structure of PVM by using 145 nucleotide sequences of the CP gene and 117 sequences of the cysteine-rich NABP gene to assess the accurate phylogeny, the evolutionary timescale, and the degree of divergence between populations in different countries and hosts. To our knowledge, our analyses provide the first insights into the spatial and temporal evolution of PVM.

2. Materials and methods

2.1. Virus isolates

One hundred and forty-five CP gene sequences and 117 NABP gene sequences were retrieved from GenBank. The details of these PVM isolates, such as geographical locations, collection time, and host origins, are given in Table S1.

2.2. Alignment and recombination analysis

All above PVM sequences were used for evolutionary analyses. As Ge et al. (2014) reported, the sequences of *Narcissus common latent virus* (NCLV) [accession number: AM158439] and *Phlox virus M* (PhlVM) [accession number: FJ159381] were used as an outgroup of the CP and NABP genes of PVM because BLAST searches showed that the CP gene of PVM is most closely related to that from NCLV, while NABP is most similar to that from PhlVM. The predicted PVM amino acid (aa) sequences were aligned with each nearest species as an outgroup by using CLUSTAL X2 (Larkin et al., 2007) with TRANSALIGN (kindly supplied by Georg Weiller) to maintain the degapped alignment of the encoded amino acids. This gave the lengths of sequences for the CP (900 nt) and NABP (324 nt) genes.

Putative recombination sites of the above aligned sequences were identified using RDP (Martin and Rybicki, 2000), GENECONV (Sawyer, 1999), BOOTSCAN (Salminen et al., 1995), MAXCHI (Smith, 1992), CHIMAERA (Posada and Crandall, 2001), 3SEQ (Boni et al., 2007) and SISCAN (Gibbs et al., 2000) programs in the RDP4 software package (Martin et al., 2015) and the original SISCAN Version 2 (Gibbs et al., 2000) software. The phylogenetic approach in the RDP4 package was used to verify the parent/donor assignments, and the putative recombination isolates were supported by at least three different methods in the RDP4 package with an associated *P*-value of $< 1.0 \times 10^{-6}$. These analyses were performed using the default settings for the different detection programs and a Bonferroni corrected *P*-value cut-off of 0.05 or 0.01. These analyses also determined which non-recombinant sequences had regions that were closest to those of the recombinant sequences and, hence, indicated the likely lineages that provided those regions of the recombinant genomes. For convenience, we called these the 'parental isolates' of the recombinants. Finally, all of the above PVM sequences were also aligned without outgroup sequences and were checked using the programs described above.

2.3. Phylogenetic analysis

The phylogenetic analysis of all above PVM sequences were assessed using the Neighbor-Net method implemented in SPLITSTREE v4.11.3 (Huson and Bryant, 2006), the maximum-likelihood (ML) method in PhyML v3 (Guindon et al., 2010), and the neighbor-joining (NJ) method in MEGA v7 (Kumar et al., 2016). For the ML tree, the best fit model of each dataset was selected using jModeltest v0.1.1 (Posada, 2008), the general time-reversible (GTR) substitution model with site rate variations and a gamma distribution (r4) (GTR + r4) supported the best model for the CP gene sequences, while GTR with invariant sites, a gamma distribution and a proportion of invariable sites (GTR + I + r4) provided the best fit for the NABP gene sequences. Branch support was evaluated by bootstrap analysis according to 1000 pseudoreplicates for ML analyses, and Kimura's two-parameter option (Kimura, 1980) was used to calculate the 1000 bootstrap replications for NJ analysis. The ML trees were compared using PATRISTIC (Fourment and Gibbs, 2006). The inferred trees were displayed with TreeView (Page, 1996).

2.4. Estimation of the substitution rates and divergence times

The evolutionary rate and timescale of PVM was estimated by BEAST v1.8.2 software (Drummond et al., 2012). Recombinant sequences were discarded from the two data sets. The molecular clock was calibrated based on the sampling times (Table S1) of the PVM isolates. The best-fitting molecular-clock models of each data set were supported by Bayes factors. In addition the strict, uncorrelated exponential, and uncorrelated lognormal molecular clocks (Drummond et al., 2006) with constant population size, logistic growth, exponential growth, expansion growth, and the Bayesian skyline plot demographic models of each data sets were also compared. Markov chain Monte Carlo (MCMC) sampling was used to estimate the posterior distributions of the parameters. During these analyses, a total of 6×10^8-step MCMC chains were explored every 10^4 steps, and the first 10% samples were discarded as burn-in. The estimation of the relevant evolutionary parameters was checked using Tracer v1.6 (http://tree.bio.ed.ac.uk/software/tracer/). To check the temporal signal of the seven PVM data sets, ten data-randomized replicates of each data set were produced. The acceptable criterion of temporal structure that is used to verify the mean estimate from the original data set should not be contained in 95% CIs of the date-randomized replicates (Duchêne et al., 2015; Ramsden et al., 2008).

The spatial diffusion patterns of PVM were inferred in BEAST using an asymmetric discrete-time Markov chain phylogeographic mode (Lemey et al., 2009). Five geographical locations (Jilin, Beijing, Qinghai, Gansu and Yunnan) in China, along with Canada, Iran, India, and European countries (Russia, Germany, Poland, Latvia and the Czech Republic), were selected and coded as discrete states. The migration pathways between pairs of locations can be identified by using Bayes factors > 10 and a mean indicator of > 0.5 (Suchard et al., 2001). The well-supported spatiotemporal movements of PVM were identified using SPREAD v 1.0.7 (Bielejec et al., 2011) and Google Earth (http://www.google.com/earth). Ten randomized location states data sets were produced and compared to assess the reliability of the most plausible location at the root node.

2.5. Genetic differentiation and measurement of gene flow

Three values of *Ks** and *Z* based on permutation statistical tests were analyzed to examine the genetic differentiation between populations (Hudson, 2000). DnaSP v6.0 (Rozas et al., 2017) was used to test the level of gene flow between populations by estimating the absolute value of the standardized variance in allele frequencies across populations (*Fst*). The absolute value of *Fst* ranges from 0 to 1 for undifferentiated to fully differentiated populations, respectively. Based on the criterion, an absolute value of $Fst < 0.33$ suggests frequent gene flow, while an absolute value of $Fst > 0.33$ suggests infrequent gene flow (He et al., 2016; Nguyen et al., 2013).

2.6. Demography analysis

Tajima's *D* (Tajima, 1989), Fu and Li's *D** and *F** (Fu and Li, 1993) statistical tests and haplotype diversity for each of the data set

sequences was estimated by DnaSP v6 (Rozas et al., 2017). The haplotype diversity was determined based on the frequency and number of haplotypes in the population. Nonsynonymous (dN) and synonymous (dS) differences that correlated with phylogenetic relationships were estimated separately using the Pamilo-Bianchi-Li (PBL) method in MEGA v7.0 (Kumar et al., 2016).

3. Results

3.1. Genome sequences

In total, 137 nucleotide sequences of the CP gene and 109 sequences of the cysteine-rich NABP gene without ambiguous nucleotides and eight full genomic sequences of PVM isolates, sampled from China, Iran, Canada, India, and European countries (Russia, Germany, Poland, Latvia and the Czech Republic) between 1990 and 2013, were retrieved from GenBank (Table S1).

3.2. Recombination analysis

To analyze the evidence of recombination, we investigated both CP and NABP genes sequences using SPLITSTREE v4.11.3 (Huson and Bryant, 2006) with the NeighborNet method. The conflicted phylogenetic signal of the reticulated phylogenetic trees (Fig. S1) showed that the recombination sites were probably present in the CP gene. In addition, these sequences were checked for recombination using the detection programs that were assembled in the RDP v4.0 package (Martin et al., 2015). Only one novel recombinant (501 isolated from Iran, KC129095) was supported by clear *P* values ($< 1 \times 10^{-6}$) in the RDP (4.996×10^{-20}), GENECONV (1.578×10^{-18}), BOOTSCAN (1387×10^{-17}), MAXCHI (4.409×10^{-6}), CHIMAERA (4.287×10^{-6}), SISCAN (4.512×10^{-10}) and 3Seq (1.105×10^{-25}) programs in the RDP 4.0 software for the CP coding region in the present study. There were no obvious recombination events detected in the NABP gene of PVM.

3.3. Phylogenetic analysis

We discarded the recombinant and conducted phylogenetic analyses using the ML and NJ methods based on above data sets. The ML trees are shown (Fig. 1) because similar topologies appeared in both ML and NJ trees (data not shown). One novel lineage was formed based on the CP and NABP gene sequences, which were compared with Ge et al.'s results that two lineages were clustered in trees derived from the CP and NABP gene sequences.

The CP ML tree partitioned all of the PVM isolates into three groups supported by high bootstrap: GP1 consists of Canadian and Iranian subgroups, GP2 consists of Indian and Iranian subgroups, and GP3 consists of Chinese and Iranian/Indian subgroups (Fig. 1A). The Iranian isolates belonged to all of the three groups in the CP trees, the Indian isolates belonged to two of the three groups (GP2 and GP3), and the Canadian and Chinese isolates only belonged to one group each (GP1 and GP3, respectively). These suggest that the Iranian and Indian populations probably were more diverse than the Canadian and Chinese populations. Nucleotide diversities of the four PVM populations also showed that the PVM isolates from Iran and India were more diverse than the Canadian and Chinese populations based on the CP gene (Table S2). The PVM isolates from potato were widely distributed in the CP ML tree while the pepino, tomato and tobacco isolates were clustered into GP3 only. Furthermore, the nucleotide diversities of the PVM potato isolates were higher than the pepino and tomato isolates (Table S2). Based on the host, both the phylogenetic analysis and nucleotide diversities supported that the PVM potato isolates were more diverse than the pepino and tomato populations.

For the NABP gene, similar topologies with the CP ML tree were observed (Fig. 1B). All of the PVM isolates were also clustered into three groups with clear geographical and host specificity. The ML trees of the CP and NABP genes were compared by PATRISTIC software. The distance plots of the CP distances against the ML trees of the NABP genes showed three distinct lineages (Fig. 2).

3.4. Genetic population structure and selection

The demography, haplotype and nucleotide diversities of the Canadian, Indian, Iranian and Chinese populations of PVM were calculated (Table S2) based on the CP and NABP gene sequences. High haplotype diversity and low genetic diversity were observed in most of the PVM geographical populations. Based on the CP gene, the Iranian population showed the highest genetic diversity, while the Chinese population had the lowest genetic diversity. Negative demographic statistical test values of most Chinese populations were found in the CP and NABP coding regions of PVM. Both demography and haplotype and nucleotide diversity analyses supported that a recent expansion had occurred in China. The dN/dS ratios were used to estimate the selective constraints on the CP and NABP genes using the SLAC, FEL, and REL methods in DataMonkey (http://www.datamonkey.org) and MEGA v7. A strong purifying selection was found in both the CP and NABP coding regions of PVM, with varied values. The CP gene was found under the strongest selection pressure with the lowest dN/dS value (0.059), while the weakest selection was in the NABP gene with the highest value (0.304) (Table S3). However, no positive selection sites were found in either of the two coding regions.

3.5. Evolutionary rates and timescales

To estimate the evolutionary rates and timescales of the individual PVM genes or genomic regions, a Bayesian phylogenetic method in BEAST v1.8.4 (Drummond et al., 2012) was selected and used in this study. The Bayesian skyline plot was supported as the best demographic model for the CP gene, while the expansion growth demographic model was best supported for the NABP gene (Table 1), based on a comparison of marginal likelihoods that were calculated using the harmonic-mean estimator in Tracer v1.6 (http://tree.bio.ed.ac.uk/software/tracer/). For the two data sets, a relaxed-clock model provided a better fit than the strict-clock model (Table 1), which indicated the presence of rate variation among groups. Both CP and NABP datasets passed the date-randomization tests (Duchêne et al., 2015; Ramsden et al., 2008) and met the more conservative criterion proposed by Duchêne et al. (2015), which indicated the presence of an adequate temporal signal in these data sets (Fig. S3).

The mean evolutionary rates of the CP and NABP genes were different; 4.12×10^{-3} for CP, and 1.06×10^{-2} for NABP (Table 1). The time to the most recent common ancestors (TMRCAs) was 62 years (31 – 110) for CP (Fig. 3, Table 1). and 29 years (28–33) for NABP (Fig. S2, Table 1).

3.6. Geographical spread of PVM

We assessed the likely migration routes of PVM into Canada, China, India, European countries (including Germany, Poland, Russia, the Czech Republic and Latvia) and Iran according to the non-recombinant sequences of the CP and NABP genes using a Bayesian phylogeographical analysis (Lemey et al., 2009). The Chinese PVM isolates that were used in the present study were tagged with their provinces of provenance, while other isolates were tagged with their countries because of limited detailed information in the public data bases. Our results showed that PVM migrated from China to Canada, Iran, India and European countries, and it circulated within China (Fig. 4, Table S4), based on the CP and NABP data sets. These results were confirmed by the gene flows and ML trees of the CP and NABP regions. The absolute values of *Fst* between Chinese and European populations and between Chinese and Indian populations were < 0.33 based on the CP gene

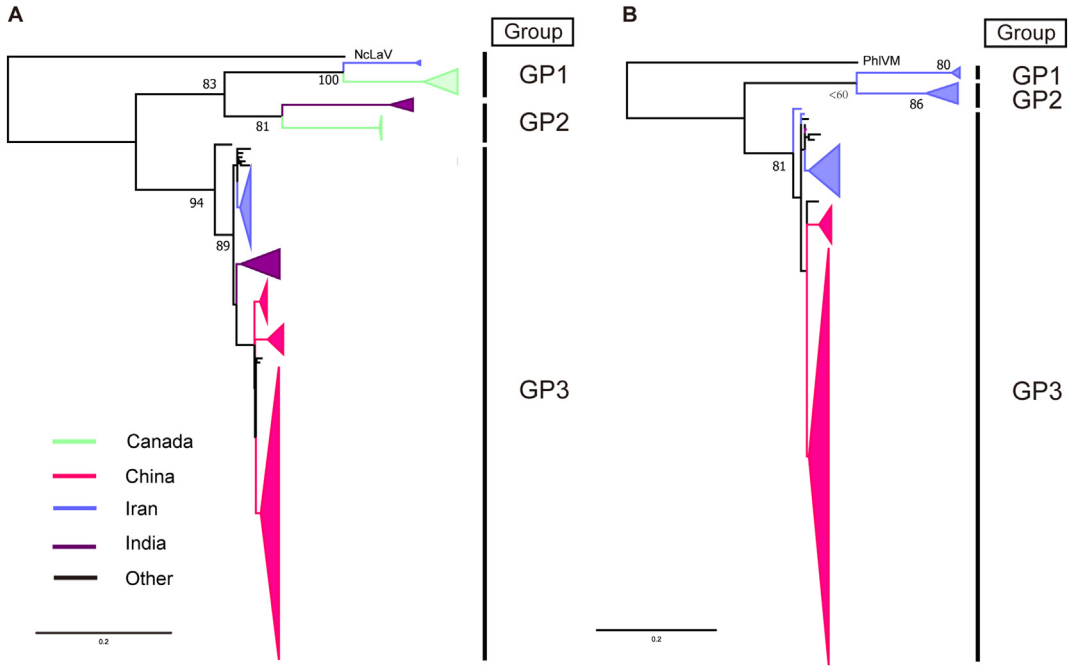
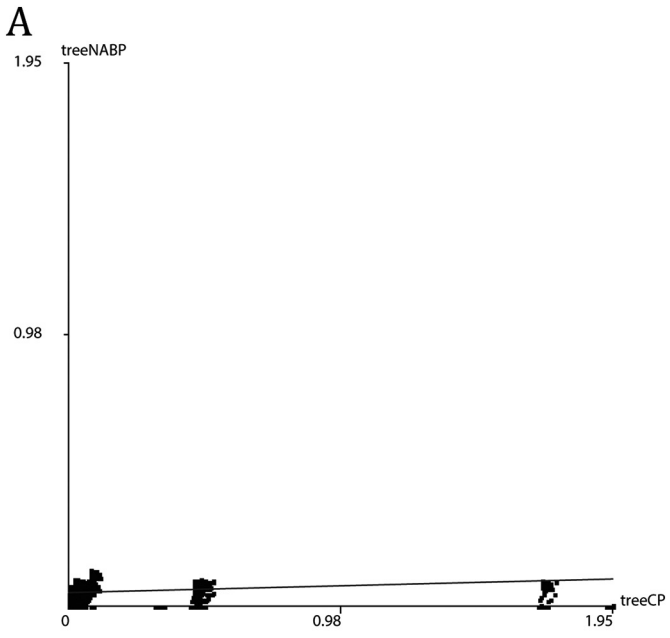


Fig. 1. Maximum-likelihood trees that were calculated from the CP (A) and NABP (B) gene sequences of potato virus M obtained in this study. Numbers at each node indicate the percentage of supporting puzzling steps (or bootstrap samples) in the maximum-likelihood trees. The horizontal branch length is drawn to scale with the bar indicating 0.2 nt replacements per site. The sequences of narcissus common latent virus (NCLV) [accession number: AM158439] was used as an outgroup of the CP gene tree. While the sequences of phlox virus M (PhlVM) [accession number: FJ159381] was used as an outgroup of the NABP gene tree.



B

Open reading frame	Correlation coefficient	Slope
CP vs. NABP	0.275	0.026

Fig. 2. Graphs comparing patristic distances in pairs of maximum-likelihood trees. Isolates from groups I, II and III were used.

(Table S5), and the absolute values between Chinese and European populations and between Chinese and Iranian populations were also smaller than 0.33 based on the NABP gene (Table S5), which supported the frequent gene flows between Chinese and European populations and between Iranian and Indian populations. In the CP and NABP ML trees, the Chinese isolates were closest to the European, Iranian and Indian clusters of the GP3 group (Fig. 1), which also supported the probable migration routes.

4. Discussion

Previously, the genetic structure of PVM populations was investigated in China and Iran based on only single gene's sequences (Ge et al., 2014; Tabasinejad et al., 2014, 2015). In the present study, the migration dynamics and adaptive evolution of PVM populations were estimated. We found that (1) one novel recombinant was identified; (2) one novel geographically confined phylogenetic lineage was found from each of the CP and ML trees; (3) PVM in potato possesses more diversity than that in pepino; (4) the evolutionary rate of the NABP gene (1.06×10^{-2}) was faster than that of the CP gene (4.12×10^{-3}); (5) TMRCAs is similar between CP (CIs 31–110) and NABP (CIs 28–33) genes; and (6) there is evidence that PVM circulated within China and migrated from China to Canada, European countries, Iran and India.

Recombination is considered to be one of the main factors shaping the evolution of plant viruses (Chare and Holmes, 2006; Gibbs and Ohshima, 2010), especially for potyviruses (Green et al., 2017; Ogawa et al., 2008; Ohshima et al., 2002, 2007) and begomoviruses (Lefeuvre et al., 2009, 2007; Lefeuvre and Moriones, 2015; Martin et al., 2011). However, our previous report found a low recombination frequency in PVM (Ge et al., 2014). Here, only one novel recombinant was found in the CP gene, which further supported our result. This is possibly due to that PVM only natural infect several solanaceae crops, which are similar to SCSMV and the narcissus viruses (He et al., 2016; Ohshima et al., 2016), which also have infrequent recombination events.

Tabasinejad et al. (2014, 2015) have found two divergent evolutionary lineages that did not reflect the geographical regions of origin in

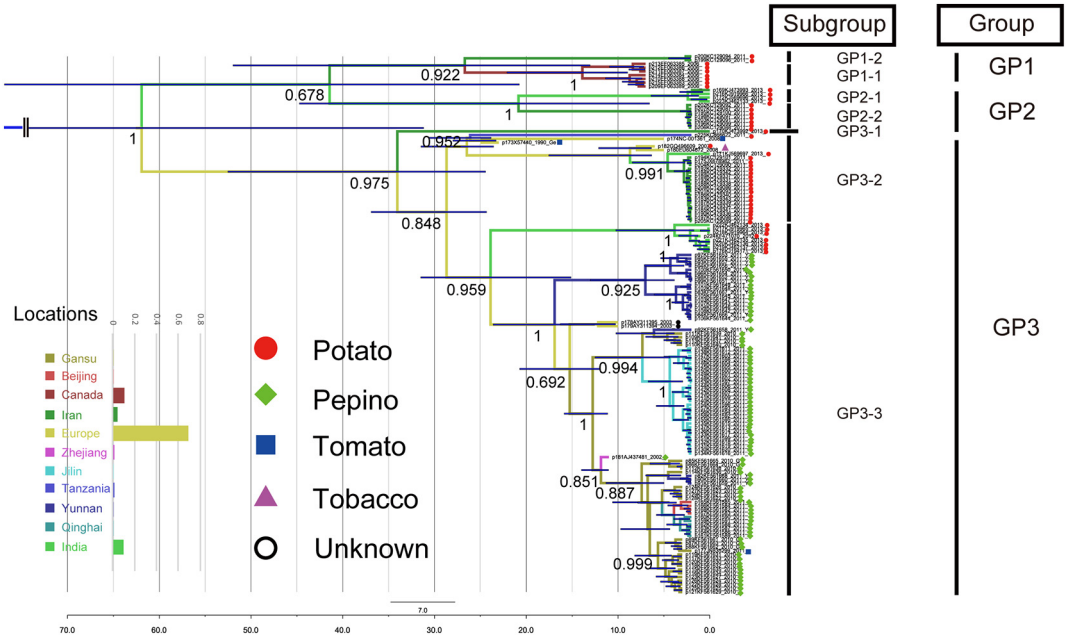


Fig. 3. Bayesian maximum-clade-credibility tree inferred from 144 sequences of the CP gene of potato virus M. Horizontal blue bars represent the 95% credibility intervals of the estimates of node ages. The tree topology has been chosen to maximize the product of node posterior probabilities. Only posterior probability values above 0.95 are shown. The bar graph shows the root state posterior probabilities for each location. Year before present; 2013. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the phylogenetic analysis of PVM based on the CP gene and cysteine-rich protein (CRPs) motifs. At the same time, our previous studies showed that two lineages were also found in PVM ML trees based on the TGB 2, CP, and NABP genes, but these lineages seemed to reflect the geographic origin (Ge et al., 2014). In the present study, we found at least three geographically confined phylogenetic lineages supported by high bootstrap based on the CP and NABP gene sequences (Fig. 1). The third lineage (GP2) consists of isolates from India and Iran in the CP ML tree, and consists of isolates from Iran in the NABP ML tree.

PVM was first isolated from potato in the early 1920s (Schultz and Folsom, 1923) and was isolated from other nightshade crops, such as pepino, tomato and tobacco, early in the decade (Ge et al., 2012, 2014). These suggested that most, if not all, PVM were prevalent in potato much earlier than in pepino and tomato, although the dates that were found for the virus emergences may be not consistent with the actual emergence. Host–parasite interactions could influence the dynamics, emergence, and evolution of infectious diseases (Gibbs et al., 2010; Irwin et al., 2012; Rodelo-Urrego et al., 2013; Torres-Pérez et al., 2011). Cuevas et al. reported that the genetic structure of PVY was shaped by host-driven adaptation (Cuevas et al., 2012b). In this study, the PVM isolates in potato are more diverse than in pepino and tomato, possibly because of the long-term coevolution between PVM and potato. However, the effect of the biased location sampling cannot be excluded here, because only Chinese PVM isolates are available from pepino.

The most recent common ancestor (TMRCAs) of PVM that was estimated here corresponds very closely to the timing of the virus emergence dates that were found in Ireland (Schultz and Folsom, 1923). The potato (*Solanum tuberosum* L.) was first domesticated in Peru between 8000 and 5000 BCE, and its landrace cultivars had descended from a monophyletic origin (Spooner et al., 2005). In the latter half of the 16th century, potato was introduced to Europe by the Spanish conquistadors and then diffused widely in Europe as a major food resource. In China, potato was introduced from Europe at the end of the Ming dynasty (1600–1700) and became popular throughout China in the Qianlong era (1735–1796) of the Qing dynasty (https://en.wikipedia.org/wiki/History_of_the_potato). Similarly, potato was introduced into the western coast of India in the early seventeenth century by the Portuguese, and it circulated across northern hill areas in the end of the 18th century (Nunn et al., 2008). In North America, potato was widely planted as early as 1838, although the first record of potatoes in North America were introduced from Europe in approximately 1719 (https://en.wikipedia.org/wiki/History_of_the_ potato). TMRCAs of the basal node in potato virus Y phylogeny are most likely represents an event that occurred in South America (Gibbs et al., 2017). Similar, Santillan et al. speculated that potato virus S emerged in South America at least 5000 years ago (Santillan et al., 2018). Our Bayesian analysis places the root of the CP and NABP trees in Europea rather than South America,

Table 1
Details of the CP and NABP datasets used for estimation of nucleotide substitution rate and time to the most recent common ancestor for PVM.

Parameter	CP	NABP
Best-fit substitution model	GTR + Γ_4	GTR + I + Γ_4
Best-fit population growth model	Relaxed uncorrelated exponential	Relaxed uncorrelated exponential
Best-fit molecular-clock model	Coalescent Extended Bayesian Skyline	Coalescent expansion
Sequence length (nt)	900	324
No. of sequences	144	117
Sampling date range	1990–2013	1985–2013
Chain length (in millions)	600	600
TMRCA (years)	62 (31–110)	29 (28–33)
Substitution rate (nt per site per year)	4.12×10^{-3}(2.51×10^{-3}-5.87×10^{-3})	1.06×10^{-2}(6.91×10^{-3}-1.45×10^{-2})

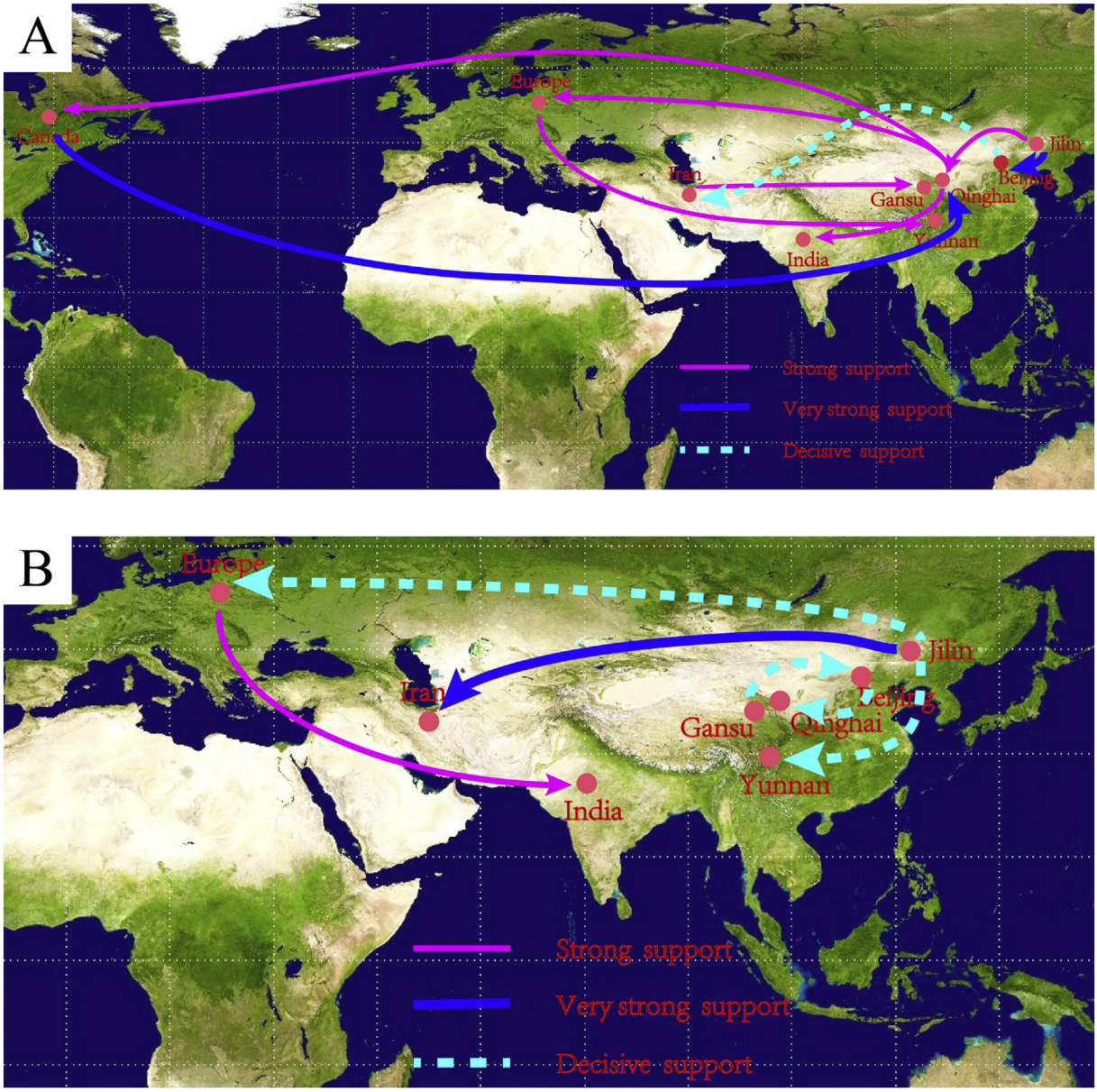


Fig. 4. Spatial diffusion pathways of potato virus M that were inferred using the CP (A) and NABP (B) gene sequences using non-recombinant sequences. Dissemination routes are only shown when supported by a Bayes factor > 10 and a mean indicator > 0.5. Bayes factor values can be interpreted as follows: $10 \leq BF < 30$, strong support; $30 \leq B < 100$, very strong support; and $BF \geq 100$, decisive support.

mostly due to the limitations that there is no sequences of the virus isolates available from the South America, or because of the earlier South America, Europe, China and India PVM populations were either not sampled or did not survive to the present day.

Our Bayesian phylogeographic analysis supports the presence of several migration links in the diffusion processes of PVM, while China might have played a key role in seeding the PVM epidemics. Europe was the largest producer of potatoes in a long time of the 20th century (NPCS Board of Consultants and Engineers, 2007). However, in the past few decades, the most rapid expansion of potato has occurred in southern and eastern Asia (https://en.wikipedia.org/wiki/Potato), and China has become the largest producer of potatoes in the late 20th centry (Gao et al., 2017; Duan et al., 2018). Thus, the results of the spatial analysis of PVM is consistent with the global trade in potatoes over the past few decades, suggesting that movements of PVM have been associated with human-mediated activities.

The pepino (*Solanum muricatum*) originated from the tropical and subtropical Andes in Peru as an important, esteemed crop (Prohens et al., 1996). In the last two decades of the 20th century, it was grown as a commercial fruit with high prices in New Zealand, European countries, the United States and China (Luo, 1994; Prohens et al., 1996; Sweet, 1986). Presently, pepino is widely grown in Gansu, Qinghai, Jilin, and Yunnan provinces of China (Ge et al., 2014). The PVM pepino isolates are clustered with potato isolates as sister sublineages in the GP3 group (Fig. 3), which possibly suggests that the PVM pepino isolate spread from the potato, and its diversity center is China. However, further pepino sampling of the PVM sublineage is needed to investigate the history of its emergence, particularly from Andean countries.

In conclusion, our study has firstly identified the evolutionary rate of the NABP (1.06×10^{-2}) and CP (4.12×10^{-3}) genes, and found that PVM migrated from China to Canada, Iran, India and European countries, and it circulated within China. Our study is the first attempt to evaluate the evolutionary rates, timescales and migration dynamics of PVM.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.meegid.2019.04.034.

Acknowledgments

This work was supported by grants from the National Natural Science Foundation of China (No. 31601604, 31772210), the Yunnan Key Laboratory of Sugarcane Genetic Improvement (2018DG018-1), the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (No. 16KJB210015), and Science and Technology Project of Jiangsu Province (BE2018359).

Disclosure statement

The authors declare no conflict of interest.

References

Bielejec, F., Rambaut, A., Suchard, M.A., Lemey, P., 2011. SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. Bioinformatics 27, 2910–2912.
Boni, M.F., Posada, D., Feldman, M.W., 2007. An exact nonparametric method for inferring mosaic structure in sequence triplets. Genetics 176, 1035–1047.
Chare, E.R., Holmes, E.C., 2006. A phylogenetic survey of recombination frequency in plant RNA viruses. Arch. Virol. 151, 933–946.
Cuevas, J.M., Delaunay, A., Rupar, M., Jacquot, E., Elena, S.F., 2012a. Molecular evolution and phylogeography of potato virus Y based on the CP gene. J. Gen. Virol. 93, 2496–2501.
Cuevas, J.M., Delaunay, A., Visser, J.C., Bellstedt, D.U., Jacquot, E., Elena, S.F., 2012b. Phylogeography and molecular evolution of potato virus Y. PLoS One 7, e37853.
Drummond, A.J., Ho, S.Y.W., Phillips, M.J., Rambaut, A., 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4, e88.
Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29, 1969–1973.
Duan, G., Zhan, F., Du, Z., Ho, S.Y.W., Gao, F., 2018. Europe was a hub for the global spread of potato virus S in the 19th century. Virology 525, 200–204.
Duchêne, S., Duchêne, D., Holmes, E.C., Ho, S.Y.W., 2015. The performance of the date-randomization test in phylogenetic analyses of time-structured virus data. Mol. Biol. Evol. 32, 1895–1906.
Duffy, S., Holmes, E.C., 2008. Phylogenetic evidence for rapid rates of molecular evolution in the single-stranded DNA begomovirus tomato yellow leaf curl virus. J. Virol. 82, 957–965.
Flatken, S., Ungewickell, V., Menzel, W., Maiss, E., 2008. Construction of an infectious full-length cDNA clone of potato virus M. Arch. Virol. 153, 1385–1389.
Fourment, M., Gibbs, M.J., 2006. PATRISTIC: a program for calculating patristic distances and graphically comparing the components of genetic change. BMC Evol. Biol. 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., 2017. Adaptive evolution and demographic history contribute to the divergent population genetic structure of potato virus Y between China and Japan. Evol. Appl. 10, 379–390.
García-Arenal, F., Fraile, A., Malpica, J.M., 2001. Variability and genetic structure of plant virus populations. Annu. Rev. Phytopathol. 39, 157–186.
Ge, B.B., He, Z., Jiang, D.M., Zhang, Z.X., Liu, G.J., Wang, H.Q., 2012. Characterization and complete nucleotide sequence of potato virus M isolated from tomato in China. Acta Virol. 56, 261–263.
Ge, B., He, Z., Zhang, Z., Wang, H., Li, S., 2014. Genetic variation in potato virus M isolates infecting Pepino (*Solanum muricatum*) in China. Arch. Virol. 159, 3197–3210.
Gibbs, A., Ohshima, K., 2010. Potyviruses and the digital revolution. Annu. Rev. Phytopathol. 48, 205–223.
Gibbs, M.J., Armstrong, J.S., Gibbs, A.J., 2000. Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. Bioinformatics 16, 573–582.
Gibbs, A.J., Ohshima, K., Phillips, M.J., Gibbs, M.J., 2008. The prehistory of potyviruses: their initial radiation was during the dawn of agriculture. PLoS One 3, e2523.
Gibbs, A.J., Fargette, D., García-Arenal, F., Gibbs, M.J., 2010. Time-the emerging dimension of plant virus studies. J. Gen. Virol. 91, 13–22.
Gibbs, A.J., Nguyen, H.D., Ohshima, K., 2015. The 'emergence' of turnip mosaic virus was probably a 'gene-for-quasi-gene' event. Curr. Opin. Virol. 10, 20–26.
Gibbs, A.J., Ohshima, K., Yasaka, R., Mohammadi, M., Gibbs, M.J., Jones, R.A.C., 2017. The phylogenetics of the global population of potato virus Y and its necrogenic recombinants. Virus Evol. 3, 1–12.
Gramstat, A., Courtpozanis, A., Rohde, W., 1990. The 12 kDa protein of potato virus M displays properties of a nucleic acid-binding regulatory protein. FEBS Lett. 276, 34–38.
Green, K.J., Brown, C.J., Gray, S.M., Karasev, A.V., 2017. Phylogenetic study of recombinant strains of *Potato virus Y*. Virology 507, 40–52.
Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59, 307–321.
Harkins, G.W., Martin, D.P., Duffy, S., Monjane, A.L., Shepherd, D.N., Windram, O.P., Owor, B.E., Donaldson, L., van Antwerpen, T., Sayed, R.A., Flett, B., Ramusi, M., Rybicki, E.P., Peterschmitt, M., Varsani, A., 2009. Dating the origins of the maize-adapted strain of maize streak virus, MSV-A. J. Gen. Virol. 90, 3066–3074.
He, Z., Li, W., Yasaka, R., Huang, Y., Zhang, Z., Ohshima, K., Li, S., 2014. Molecular variability of sugarcane streak mosaic virus in China based on an analysis of the P1 and CP protein coding regions. Arch. Virol. 159, 1149–1154.
He, Z., Yasaka, R., Li, W., Li, S., Ohshima, K., 2016. Genetic structure of populations of sugarcane streak mosaic virus in China: comparison with the populations in India. Virus Res. 211, 103–116.
Hudson, R.R., 2000. A new statistic for detecting genetic differentiation. Genetics 155, 2011–2014.
Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23, 254–267.
Irwin, N.R., Bayerlová, M., Missa, O., Martínková, N., 2012. Complex patterns of host switching in New World arenaviruses. Mol. Ecol. 21, 4137–4150. https://doi.org/10.1111/j.1365-294X.2012.05663.x.
Jones, R.A.C., 2009. Plant virus emergence and evolution: origins, new encounter scenarios, factors driving emergence, effects of changing world conditions, and prospects for control. Virus Res. 141, 113–130.
Ju, H.J., Samuels, T.D., Wang, Y.S., Blancaflor, E., Payton, M., Mitra, R., Krishnamurthy, K., Nelson, R.S., Verchot-Lubicz, J., 2005. The potato virus X TGBp2 movement protein associates with endoplasmic reticulum-derived vesicles during virus infection. Plant Physiol. 138, 1877–1895.
Kimura, M., 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16, 111–120.
Koonin, E.V., Dolja, V.V., Morris, T.J., 1993. Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences. Crit. Rev. Biochem. Mol. Biol. 28, 375–430.
Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874.
Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and clustal X version 2.0. Bioinformatics 23, 2947–2948.
Lefeuvre, P., Moriones, E., 2015. Recombination as a motor of host switches and virus emergence: Geminiviruses as case studies. Curr. Opin. Virol. 10, 14–19.
Lefeuvre, P., Martin, D.P., Hoareau, M., Naze, F., Delatte, H., Thierry, M., Varsani, A., Becker, N., Reynaud, B., Lett, J.M., 2007. Begomovirus "melting pot" in the south-West Indian Ocean islands: molecular diversity and evolution through recombination. J. Gen. Virol. 88, 3458–3468.
Lefeuvre, P., Lett, J.M., Varsani, A., Martin, D.P., 2009. Widely conserved recombination patterns among single-stranded DNA viruses. J. Virol. 83, 2697–2707.
Lemey, P., Rambaut, A., Drummond, A.J., Suchard, M.A., 2009. Bayesian phylogeography finds its roots. PLoS Comput. Biol. 5, e1000520.
Luo, W., 1994. New sort of vegetable *Solanum muricatum* and techniques for its cultivation. Chinese Veg. 1, 53–54.
Martin, D.P., Rybicki, E.P., 2000. RDP: detection of recombination amongst aligned sequences. Bioinformatics 16, 562–563.
Martin, D.P., Biagini, P., Lefeuvre, P., Golden, M., Roumagnac, P., Varsani, A., 2011. Recombination in eukaryotic single stranded DNA viruses. Viruses 3, 1699–1738.
Martin, D.P., Murrell, B., Golden, M., Khoosal, A., Muhire, B., 2015. RDP4: detection and analysis of recombination patterns in virus genomes. Virus Evol. 1.
Morozov, S.Y., Solovyev, A.G., Kalinina, N.O., Fedorkin, O.N., Samuilova, O.V., Schiemann, J., Atabekov, J.G., 1999. Evidence for two nonoverlapping functional domains in the *Potato virus X* 25K movement protein. Virology 260, 55–63.
Nguyen, H.D., Tran, H.T.N., Ohshima, K., 2013. Genetic variation of the *Turnip mosaic virus* population of Vietnam: a case study of founder, regional and local influences. Virus Res. 171, 138–149.
NPCS Board of Consultants and Engineers, N.B.O.C, 2007. Potato and Potato Products Cultivation, Seed Production, Manuring, Harvesting, Organic Farming, Storage and Processing. NIIR Project Consultancy Services.
Nunn, N., Nunn, N., Qian, N., 2008. Columbus's Contribution to World Population and Urbanization: A Natural Experiment Examining the Introduction of Potatoes. Working paper. Harvard University.
Ogawa, T., Tomitaka, Y., Nakagawa, A., Ohshima, K., 2008. Genetic structure of a population of *Potato virus Y* inducing potato tuber necrotic ringspot disease in Japan; comparison with north American and European populations. Virus Res. 131, 199–212.
Ohshima, K., Yamaguchi, Y., Hirota, R., Hamamoto, T., Tomimura, K., Tan, Z., Sano, T., Azuhata, F., Walsh, J.A., Fletcher, J., Chen, J., Gera, A., Gibbs, A., 2002. Molecular evolution of *Turnip mosaic virus*: evidence of host adaptation, genetic recombination and geographical spread. J. Gen. Virol. 83, 1511–1521.
Ohshima, K., Tomitaka, Y., Wood, J.T., Minematsu, Y., Kajiyama, H., Tomimura, K., Gibbs, A.J., 2007. Patterns of recombination in turnip mosaic virus genomic sequences indicate hotspots of recombination. J. Gen. Virol. 88, 298–315.
Ohshima, K., Nomiyama, R., Mitoma, S., Honda, Y., Yasaka, R., Tomimura, K., 2016. Evolutionary rates and genetic diversities of mixed potyviruses in Narcissus. Infect. Genet. Evol. 45, 213–223.
Page, R.D.M., 1996. Tree view: an application to display phylogenetic trees on personal computers. Bioinformatics 12, 357–358.
Posada, D., 2008. jModelTest: phylogenetic model averaging. Mol. Biol. Evol. 25, 1253–1256.
Posada, D., Crandall, K.A., 2001. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. Proc. Natl. Acad. Sci. U. S. A. 98, 13757–13762.
Prohens, J., Ruiz, J.J., Nuez, F., 1996. The Pepino (*Solanum muricatum*, Solanaceae): a "new" crop with a history. Econ. Bot. 50, 355–368.
Ramsden, C., Holmes, E.C., Charleston, M.A., 2008. Hantavirus evolution in relation to its rodent and insectivore hosts: no evidence for codivergence. Mol. Biol. Evol. 26, 143–153.
Rocha, C.S., Castillo-Urquiza, G.P., Lima, A.T.M., Silva, F.N., Xavier, C.A.D., Hora-Júnior,

B.T., Beserra-Júnior, J.E.A., Malta, A.W.O., Martin, D.P., Varsani, A., Alfenas-Zerbini, P., Mizubuti, E.S.G., Zerbini, F.M., 2013. Brazilian begomovirus populations are highly recombinant, rapidly evolving, and segregated based on geographical location. J. Virol. 87, 5784–5799.

Rodelo-Urrego, M., Pagán, I., González-Jara, P., Betancourt, M., Moreno-Letelier, A., Ayllón, M.A., Fraile, A., Piñero, D., García-Arenal, F., 2013. Landscape heterogeneity shapes host-parasite interactions and results in apparent plant-virus codivergence. Mol. Ecol. 22, 2325–2340.

Rozanov, M.N., Koonin, E.V., Gorbalenya, A.E., 1992. Conservation of the putative methyltransferase domain: a hallmark of the "Sindbis-like" supergroup of positive-strand RNA viruses. J. Gen. Virol. 73, 2129–2134.

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J.C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S.E., Sánchez-Gracia, A., 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol. Biol. Evol. 34, 3299–3302.

Salminen, M.O., Carr, J.K., Burke, D.S., McCutchan, F.E., 1995. Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. AIDS Res. Hum. Retrovir. 11, 1423–1425.

Santillan, F.W., Fribourg, C.E., Adams, I.P., Gibbs, A.J., Boonham, N., Kehoe, M.A., Maina, S., Jones, R.A.C., 2018. The biology and phylogenetics of *Potato virus S* isolates from the Andean region of South America. Plant Dis. 102, 869–885.

Sawyer, S.A., 1999. GENECONV: A Computer Package for the Statistical Detection of Gene Conversion. Distributed by the author. Department of Mathematics, Washington University, Louis. http://www.math.wustl.edu/sawyer.

Schultz, E., Folsom, D., 1923. Transmission, variation, and control of certain degeneration diseases of Irish potatoes. J. Agric. Res. 25, 43–147.

Senshu, H., Yamaji, Y., Minato, N., Shiraishi, T., Maejima, K., Hashimoto, M., Miura, C., Neriya, Y., Namba, S., 2011. A dual strategy for the suppression of host antiviral silencing: two distinct suppressors for viral replication and viral movement encoded by potato virus M. J. Virol. 85, 10269–10278.

Smith, J., 1992. Analyzing the mosaic structure of genes. J. Mol. Evol. 34, 126–129.

Spooner, D.M., McLean, K., Ramsay, G., Waugh, R., Bryan, G.J., 2005. A single domestication for potato based on multilocus amplified fragment length polymorphism genotyping. Proc. Natl. Acad. Sci. 102, 14694–14699.

Su, X., Wu, K., Zhang, L.Z., Rahman, M.S., Zheng, K., Li, T., Zhang, Z., Dong, J., 2017. Complete genome sequence of a new isolate of *Potato virus M* in Yunnan, China. Arch. Virol. 162, 2485–2488.

Suchard, M.A., Weiss, R.E., Sinsheimer, J.S., 2001. Bayesian selection of continuous-time Markov chain evolutionary models. Mol. Biol. Evol. 18, 1001–1013.

Sweet, C., 1986. Pepino: Can it Grow Here? Calif Grow. vol. 10. pp. 19–34.

Tabasinejad, F., Jafarpour, B., Zakiaghl, M., Siampour, M., Rouhani, H., Mehrvar, M., 2014. Genetic structure and molecular variability of potato virus M populations. Arch. Virol. 159, 2081–2090.

Tabasinejad, F., Jafarpour, B., Zakiaghl, M., Siampour, M., Rowhani, H., Mehrvar, M., 2015. Molecular variability in the cysteine rich protein of potato virus M. Virus Dis. 26, 117–122.

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

Torres-Pérez, F., Palma, R.E., Hjelle, B., Holmes, E.C., Cook, J.A., 2011. Spatial but not temporal co-divergence of a virus and its mammalian host. Mol. Ecol. 20, 4109–4122.

Xu, H., D'Aubin, J., Nie, J., 2010. Genomic variability in potato virus M and the development of RT-PCR and RFLP procedures for the detection of this virus in seed potatoes. Virol. J. 7, 25.

Yasaka, R., Nguyen, H.D., Ho, S.Y.W., Duchêne, S., Korkmaz, S., Katis, N., Takahashi, H., Gibbs, A.J., Ohshima, K., 2014. The temporal evolution and global spread of *Cauliflower mosaic virus*, a plant pararetrovirus. PLoS One 9, e85641.

Zavriev, S.K., Kanyuka, K.V., Levay, K.E., 1991. The genome organization of *Potato virus M* RNA. J. Gen. Virol. 72, 9–14.