Virology 528 (2019) 110–117

Contents lists available at ScienceDirect

Virology

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

Bayesian phylodynamic analysis reveals the dispersal patterns of tobacco mosaic virus in China

Fangluan Gao[a,b,1], Xiaowei Liu[c,1], Zhenguo Du[a], Han Hou[c], Xiaoyan Wang[d], Fenglong Wang[c,*], Jinguang Yang[c,*]

[a] Fujian Key Laboratory of Plant Virology, Institute of Plant Virology, Fujian Agriculture and Forestry University, Fuzhou 350002, Fujian, PR China
[b] School of Life and Environmental Sciences, University of Sydney, Sydney, NSW 2006, Australia
[c] Tobacco Research Institute of Chinese Academy of Agricultural Sciences, Key Laboratory of Tobacco Pest Monitoring, Controlling & Integrated Management, Qingdao 266101, Shandong, PR China
[d] Foxcroft School, Middleburg, VA 20118, United States

ARTICLE INFO

Keywords:
Tobacco mosaic virus
Bayesian phylogenetics
Viral phylodynamics
Spatial transmission

ABSTRACT

Tobacco mosaic virus (TMV) is widespread in China and causes considerable economic losses to tobacco production. The molecular epidemiology of this virus is, however, poorly understood. In this study, we sequenced the genomes of 51 TMV isolates from five tobacco-producing regions in China and investigated the dispersal patterns of this virus. Our phylogenetic analysis showed that TMV might have been introduced to China in the early 1900s, probably first to southwest China. However, TMV then moved to the north of the country, where it expanded. The north became the main seeding region for the subsequent movements of the virus within China. The north-to-south movement of TMV coincides with a shift of major tobacco-producing areas from north to south in this century, suggesting a link between human activities and the dispersal of TMV in China.

1. Introduction

Tobacco (*Nicotiana tabacum*), native to South America (Holmes, 1951), is one of the most important economic crops worldwide. Tobacco was introduced to China in the 16th and 17th centuries (Benedict, 2011). From about the year 1914, flue-cured tobacco became widely cultivated in China. By the year 1948, the cultivation area of flue-cured tobacco in China had reached 0.16 million ha and this number further increased to about 0.41 million by the year 1967. From about 1966–1973, the cultivation area of flue-cured tobacco decreased slightly. However, this was followed by a sharp increase from about 0.61 million ha in 1978 to about 1.71 million in 1997. After 1997, the cultivation area of flue-cured tobacco decreased (Lu et al., 2011; Zhu, 2008). However, China remains the largest producer of flue-cured tobacco in the world, with a cultivation area greater than 1 million ha (Zhu, 2008). Flue-cured tobacco planting in China can be roughly divided into five producing regions: the Southwest producing region (SWR), Southeast producing region (SER), the producing region of the upper and middle reaches of the Yangtze River (YRR), Huanghuai producing region (HHR), and the Northern region (NR) (Li et al., 2016). The contribution of these regions to the total cultivation area of flue-cured tobacco in China has changed over time. In the past, NR and HHR were major producers. However, the cultivation area of flue-cured tobacco in these two regions decreased greatly in this century. In contrast, the other three regions, especially SWR and YRR, have become much more important in flue-cured tobacco production in China.

As in other parts of the world, tobacco mosaic virus (TMV) has long been a significant pathogen of flue-cured tobacco in China (Scholthof, 2004). Despite intensive efforts (Gooding, 1986; Li et al., 2011; Shen et al., 2014), attempts to control TMV in tobacco fields have been unsuccessful. One reason for this is that TMV has unusually stable virions (Alonso et al., 2013). The infectivity of TMV virions can be remained for years, or even decades under ideal conditions, and these virions can be easily transmitted to tobacco plants through mechanical inoculation (Creager et al., 1999; Harrison and Wilson, 1999).

TMV has a non-segmented, positive-sense, single stranded RNA genome of about 6.4 kb, which contains four open reading frames encoding at least four proteins: two replication proteins (126 K and 183 K), a movement protein (MP), and a coat protein (CP) (King et al., 2011). Among these proteins, 126 K has been shown to suppress gene silencing and is thought to play important roles in viral replication and pathogenicity (Ding et al., 2004). The 183 K protein is a read-through

* Corresponding authors.
E-mail addresses: wangfenglong@caas.cn (F. Wang), jinguangyang@126.com (J. Yang).
[1] Contributed equally to the work.

https://doi.org/10.1016/j.virol.2018.12.001
Received 22 June 2018; Received in revised form 19 November 2018; Accepted 1 December 2018
Available online 27 December 2018
0042-6822/

product of the amber stop codon of 126 K (Valle et al., 1992). Both 126 K and 183 K are RNA-specific, RNA-dependent RNA polymerases that are possibly associated with biological membrane structures (Hagiwara et al., 2003; Liu and Nelson, 2013). MP is involved in the cell-to-cell spread of viral progeny during infection, whereas CP is involved in symptom induction and is essential for long-distance movement of the virion (Heinlein, 2002).

As the first plant virus to be described, TMV has been a model for understanding the molecular biology of viruses and the mechanisms underlying virus-plant interactions (Scholthof, 2004). However, only a few studies have focused on the evolutionary dynamics of this virus. Based on an assumption of virus-host codivergence, it was estimated that most tobamoviruses have a lower evolutionary rate of 2.20×10^{-8} substitutions/site/year (Gibbs et al., 2010). In contrast, a long-term passaging experiment showed that TMV evolves at a mean rate of 3.10×10^{-4} substitutions/site/year (Kearney et al., 1999). The substitution rate in the 126 K protein gene has been estimated at 7.90×10^{-4} substitutions/site/year (95% credibility interval 3.90×10^{-4}–1.30×10^{-3}), based on an analysis of time-structured sequence data (Pagan et al., 2010). This gene seems to be evolving more rapidly than others genes in TMV genome.

The evolutionary and demographic history of TMV, especially the dynamics of its transmission and spread in China, remain poorly understood, although such information is potentially valuable for developing efficient and sustainable management strategies. In this study, we infer the phylodynamics of TMV in China using complete TMV genomes collected over a 10-year period (2001–2011). The data obtained here may be relevant for TMV management in China and may also serve as a case study for understanding the epidemiology of TMV in the world.

2. Materials and methods

2.1. Viral sequences

Total RNAs were extracted using an RNAsimple Total RNA Kit and reverse-transcribed. The TMV genome was obtained by sequencing three overlapping fragments (nucleotides 1–2777, 2593–4592, 4090–6394) using three pairs of degenerate primers (Table S1), which were designed from highly conserved regions of published TMV genomes (accession numbers AB369275-AB369276, AF165190, AF395127-AF395129, AJ011933, NC_001367, and X68110). PCR amplifications of cDNA were performed in a total volume of 50.00 μL, containing 0.50 μL of PrimerSTAR HS DNA Polymerase (2.50 U/μL; TaKaRa), 10.00 μL of 5 × PrimeSTAR Buffer, 4.00 μL dNTP mixture (2.50 mM each), 1.00 μL of forward primer (10.00 μmol/L), 1.00 μL of reverse primer (10.00 μmol/L), 32.50 μL of ddH_2O, and 1.00 μL of template cDNA. The PCR program comprised 30 cycles of 98 °C for 10 s, 50 °Cfor 15 s, and 72 °C for 2 min. PCR products were separated on 1% agarose gels by electrophoresis, visualized by UV transilluminator, and cleaned using an EasyPure Quick Gel Extraction Kit (TransGen, Beijing).

PCR products were ligated to pGEM18T Vector (Takara, China) and subsequently propagated in cells of *E. coli* strain DH5α. The recombinant plasmids were extracted and sequenced in both directions by Sangon Biological Co., Ltd. (Shanghai). At least three cDNA clones from each transformation were sequenced to obtain a consensus sequence. Sequences were assembled using DNAMAN 6.0 (Lynnon, Quebec, Canada). In addition to the novel sequence data, five complete genome sequences of TMV isolates from Fujian and Shanxi provinces in China and 20 complete sequences from other parts of the world (including France, South Korea, United Kingdom, and United States of America) were obtained from GenBank (Table S2). The combined sequence data were used for the recombination and phylogenetic analyses described below.

2.2. Recombination and phylogenetic analyses

We used two different approaches to identify potential recombination events among the TMV genomes. First, we constructed a split network using split decomposition and calculated the pairwise homoplasy index using the neighbor-net method in SplitsTree 4.13.1 (Huson, 1998). We then tested for potential recombinants in the sequence alignment using seven methods (RDP, GENECONV, BOOTSCAN, MAXCHI, CHIMAERA, SISCAN, and 3SEQ) available in the RDP 4.95 package (Martin et al., 2015). The probability of a putative recombination event was corrected by a Bonferroni procedure, with a cut-off of $p = 0.01$. To minimize false positives, recombination events were only considered to be significant if they were supported by at least four of the seven methods.

Phylogenetic analysis was performed based on the 76 genome sequences using maximum likelihood in IQ-TREE 1.5.5 (Nguyen et al., 2014). The data set was partitioned into gene groups, with the GTR $+\Gamma_4$, GTR + I, HKY $+\Gamma_4$, K2P + I substitution models applied to four protein-coding genes, respectively. The substitution models were selected using the Bayesian information criterion in ModelFinder (Kalyaanamoorthy et al., 2017). Node support in the phylogenetic tree was estimated using an ultrafast bootstrap with 5000 replicates (Minh et al., 2013), as well as the Shimodaira–Hasegawa approximate likelihood-ratio test (SH-aLRT) with 1000 replicates (Guindon et al., 2010).

2.3. Temporal dynamics of TMV

To infer the evolutionary rate and timescale of TMV, we analysed the genome sequences using a molecular clock. After we removed 20 isolates that did not have attached sampling dates, our analyses were based on 56 genome sequences from Chinese isolates of TMV. To test for temporal structure in the data set, we performed a regression of root-to-tip genetic distances against year of sampling in TempEst (Rambaut et al., 2016). For this analysis, we inferred the tree topology and branch lengths using maximum likelihood in IQ-TREE (Nguyen et al., 2014), as described above. We then performed a date-randomization test (Ramsden et al., 2008), based on 20 permutations of the sampling dates produced using the TipDatingBeast package (Rieux and Khatchikian, 2017). The date-randomized replicates of the data set were analysed using a Bayesian phylogenetic approach in BEAST 1.8.4 (Drummond et al., 2012). In this test, the data set is considered to have sufficient temporal signal for molecular dating if the 95% credibility interval of the rate estimate does not fall within the 95% credibility intervals of the rate estimates from the date-randomized replicates (Duchêne et al., 2015).

We found no evidence of temporal structure in the data set using either of the two methods, so we applied a previous estimate of the substitution rate in order to estimate the evolutionary timescale of TMV. We performed a Bayesian phylogenetic analysis in BEAST, with a separate substitution model applied to each of the four genes. A uniform prior of 3.90×10^{-4}–1.30×10^{-3} substitutions/site/year was specified for the absolute substitution rate of the 126 K protein gene, based on a previous estimate (Pagan et al., 2010). The other three genes were allowed to have distinct relative rates. The GTR $+\Gamma_4$, GTR + I, HKY $+\Gamma_4$, K2P + I substitution models were used for the four protein-coding genes, determined by ModelFinder with Bayesian information criterion. We confirmed the adequacy of the selected substitution models (Fig. S1) using PhyloMAd (Duchêne et al., 2018).

To compare the fit of the strict clock and the uncorrelated lognormal relaxed clock (Drummond et al., 2006), we computed marginal likelihoods using stepping-stone sampling (Baele et al., 2012). An extended Bayesian skyline plot was used as a flexible tree prior, allowing us to infer changes in effective population size through time (Heled and Drummond, 2008). Posterior distributions of parameters were estimated by Markov chain Monte Carlo sampling, with samples drawn every 2000 steps over 4×10^8 steps. The MCMC analysis was run for

more than twice as much, and two of them were selected to check for convergence. After discarding the first 25% of samples as burn-in, we checked for sufficient sampling by ensuring that the effective sample size of each parameter was greater than 200.

2.4. Genetic differentiation among TMV populations in China

We defined five populations of TMV, corresponding to the tobacco-producing regions of China: Southwest producing region (SWR), Southeast producing region (SER), the producing region of the upper and middle reaches of the Yangtze River (YRR), Huanghuai producing region (HHR), and Northern region (NR). We computed pairwise F_{ST}, a measure of genetic differentiation among populations, using Arlequin 3.5 (Excoffier and Lischer, 2010). We interpreted the F_{ST} values using the following thresholds: moderate degree of differentiation, 0.05–0.15; large degree of differentiation, 0.15–0.25; and great degree of differentiation, > 0.25 (Balloux and Lugon-Moulin, 2002). F_{ST} values were also used to measure levels of gene flow. F_{ST} values below 0.33 suggest frequent gene flow, whereas F_{ST} values greater than 0.33 indicate infrequent gene flow.

2.5. Phylogeographic and demographic history

To gain insight into the circulation of TMV across the tobacco-producing regions of China, we reconstructed spatial transmission patterns using a phylogeographic analysis in BEAST. Five tobacco-producing regions, as described above, were coded as discrete states. Spatial diffusions between these regions were allowed to be asymmetric and model averaging was performed using Bayesian stochastic search variable selection (Lemey et al., 2009). We used the same substitution models as described above. Posterior distributions of parameters were estimated by Markov chain Monte Carlo sampling, as described above.

The best-supported pairwise diffusions were identified using Bayes factors in SPREAD3 0.9.7 (Bielejec et al., 2011). Migration pathways were considered to be important when they yielded a Bayes factor greater than 3 and when the mean posterior value of the corresponding was greater than 0.50. Bayes factors were interpreted according to the guidelines of Kass and Raftery (1995). We also estimated the number of expected location-state transitions (Markov jump counts) along the branches of the phylogeny (Minin and Suchard, 2008). This provides a measure of asymmetric migration between the tobacco-producing regions.

To investigate the impact of imbalanced sample sizes in our data set, we used a bootstrapping approach to standardize sample sizes and performed analyses of 10 replicate subsamples. For each bootstrap replicate, we randomly sampled 7 sequences with replacement from each tobacco-producing region. Each of these datasets was analysed using BEAST, as described above. We run MCMC simulations for 100 million steps across three independent Markov chains and collected samples every 10,000 steps.

3. Results

3.1. Sampling and sequencing of TMV in China

In July 2009 and August 2010, we collected 51 isolates of TMV from tobacco plants across 12 provinces of China (Fig. 1A). The identities of the virus samples were confirmed by double-antibody sandwich enzyme-linked immunosorbent assay (DAS-ELISA) (Adgen, UK). The TMV isolates were sap-inoculated to *Nicotiana tabacum* L. cv. Samsun NN and serially cloned through single lesions.

Fifty-one complete genomes were obtained in this study and deposited in the GenBank databases under accession numbers HE818410–HE818460 (Table S2). The splits network did not show evidence of reticulations, suggesting that TMV has not experienced substantial recombination (Fig. 1B). The pairwise homoplasy index also consistently failed to identify a signal of recombination in the data set ($p > 0.05$). Likewise, no significant signals of recombination were identified by any of the seven algorithms in the RDP package. Therefore, we used the complete data set for all our phylogenetic analyses.

3.2. Phylogenetic inference

The maximum-likelihood phylogenetic analysis revealed that TMV isolates form four distinct groups, each with high degree of support (SH-aLRT ≥75% and bootstrap support values ≥ 95%; Fig. 1C). Group I contains 55 isolates from China, together with one isolate from South Korea. Group II includes the ten isolates from Europe (France and United Kingdom), four from China, and two from South Korea. Two Chinese isolates (Chuxiong-1 and Xiongfan-1) form a distinct lineage (Group III), whereas the highly divergent Group IV contains one isolate from Taiwan and one from USA.

3.3. Time-scaled Bayesian phylodynamic analyses

The uncorrelated lognormal relaxed clock yielded a higher log marginal likelihood (-19,706) than the strict clock (-19,915), indicating that the former provided the best fit to our sequence data. The results of our Bayesian phylogenetic analysis showed that TMV in China probably emerged in the Southwest producing region (root posterior probability = 0.99), with a most recent common ancestor in 1935 (95% credibility interval 1915–1954; Fig. 2A, Table 1). We conducted subsampling analyses to minimize the influence of potential sampling biases on reconstruction of the evolutionary history and dispersal patterns of TMV. Our results provide further evidence that the probability of the Southwest producing region at the root is higher than that of other producing regions either in the original data set or among the 10 replicate subsamples (Fig. S2). The most recent common ancestors of TMV isolates in Group I, Group II, and Group II were placed in 1985 (95% credibility interval 1981–1989), 2001 (95% credibility interval 1997–2004), and 2008 (95% credibility interval 2007–2009), respectively (Fig. 2A, Table 1). The four genes in the TMV genome were estimated to have evolved at very similar rates (Table 1).

3.4. Migration and demographic history of TMV in China

Our investigation of F_{ST} values did not show significant genetic differentiation between any pairs of populations, except for the NR/SER, YRR/SER, NR/SWR, and YRR/SWR comparisons (Table S3). This suggests low levels of spatial genetic structure of TMV in China. All F_{ST} values between isolates from different planting regions were lower than 0.33 (Table S3), indicating high levels of gene flow between these TMV populations.

Bayesian phylogeographic analysis supports the presence of seven migration links, with mean rates (that is, migration events per lineage per year) ranging from 0.53 to 1.49 (Fig. 3A, Table S4). The highest mean rates were observed for migration from SWR to YRR, whereas the lowest mean rates were observed for migration from NR to YRR (Table S4). Three introductions of TMV into YRR were supported by the MCC tree with high posterior probability (Fig. 2A), and their directionality was confirmed by our phylogeographic analysis (Fig. 3A, Table S4). Two of the introductions of TMV into YRR originated from SWR (in the periods 1986–2002 and 1989–2009), whereas the third introduction occurred from NR between 2002 and 2009 (Fig. 2A). This is further supported by the number of observed state changes, with migration from NR being much greater than from any other regions (Fig. 3B). The total mean rate per Markov jump for all tobacco-producing regions supports the role of NR as a seeding population. Similar results were obtained for the 10 replicate subsamples (Fig. S3). SWR and YRR have strong epidemiological links with multiple regions in this transmission network. Markov rewards for YRR (687, 95% credibility interval 491–950) are higher than those for SWR (75, 95% credibility interval

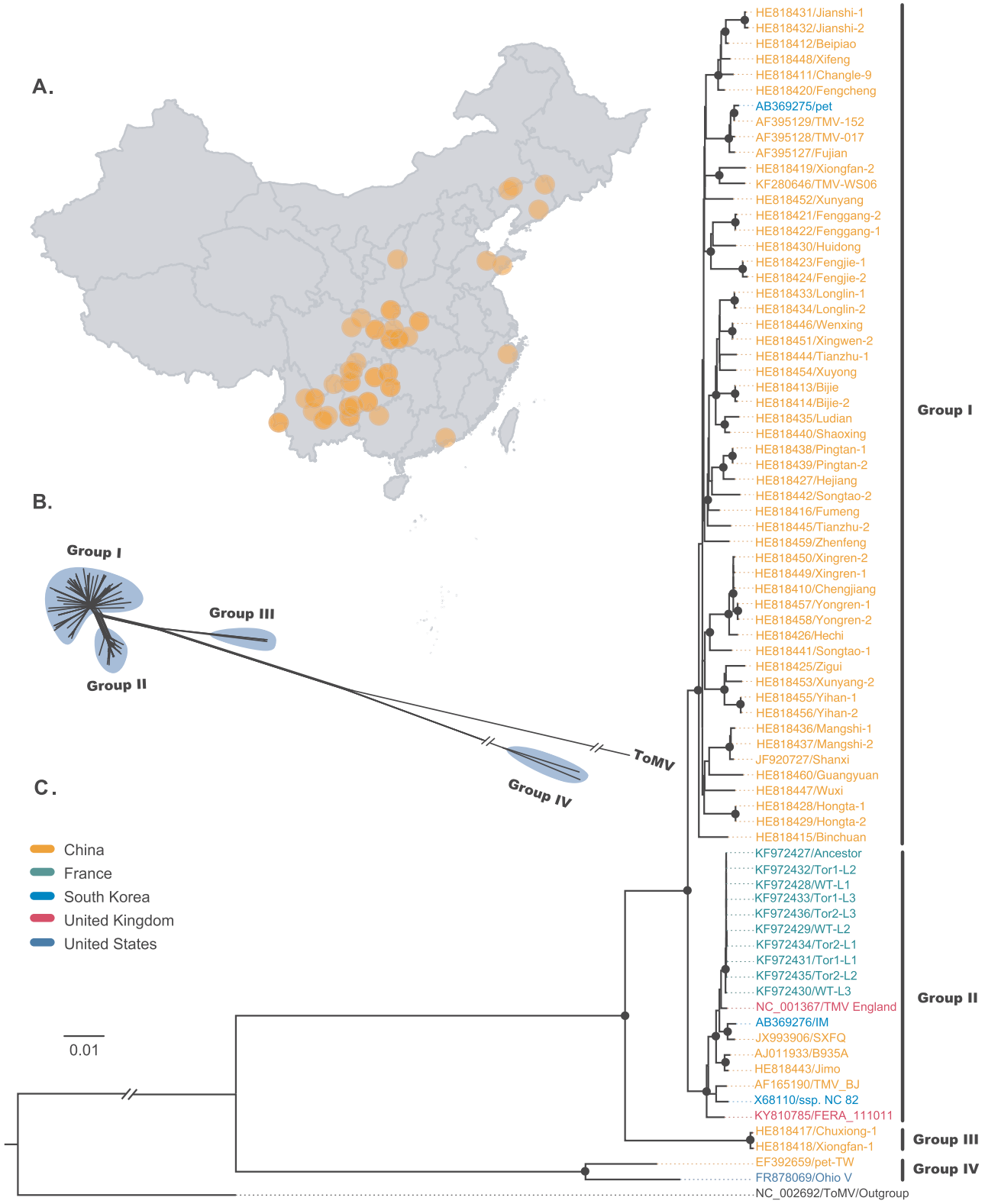
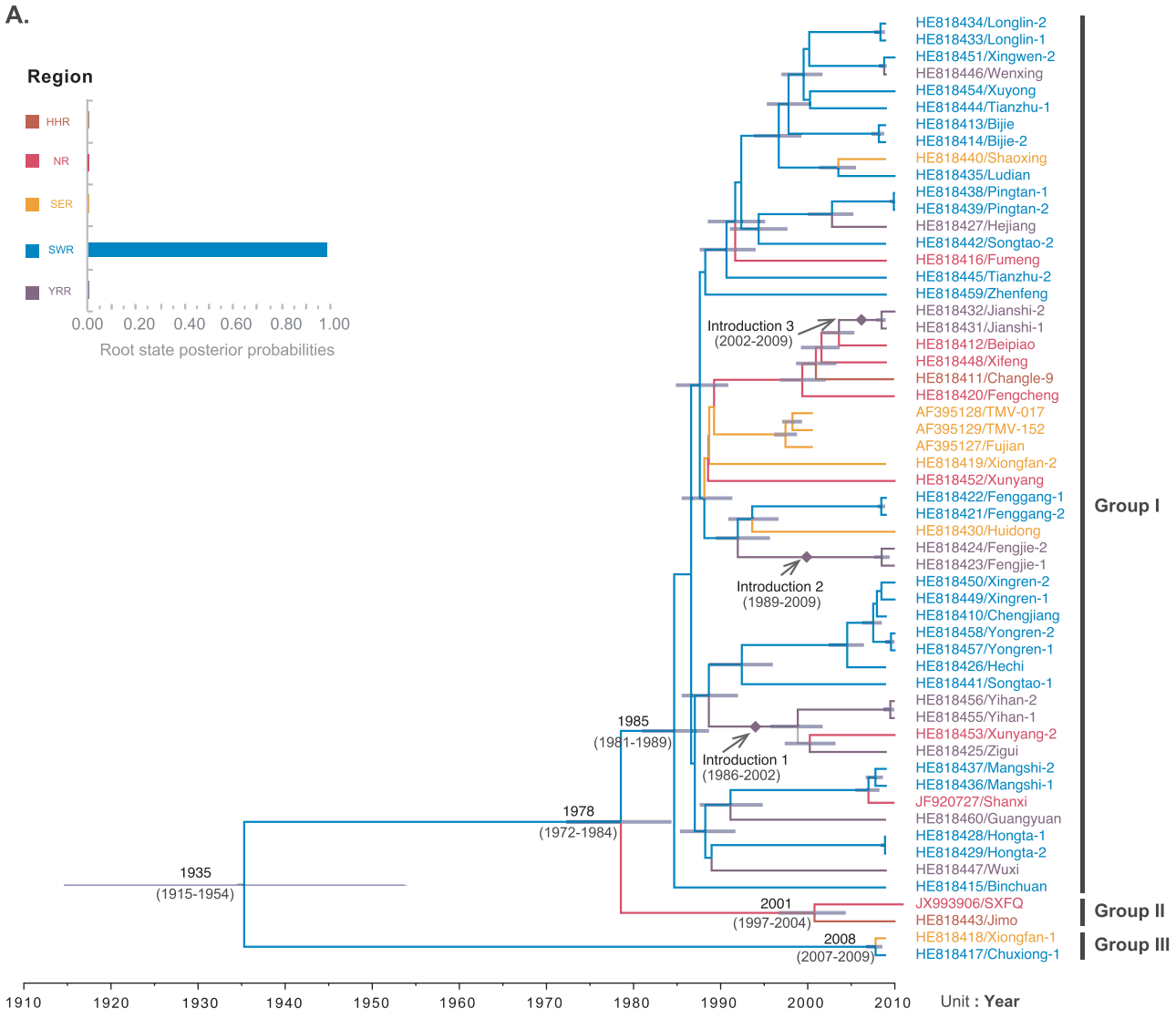
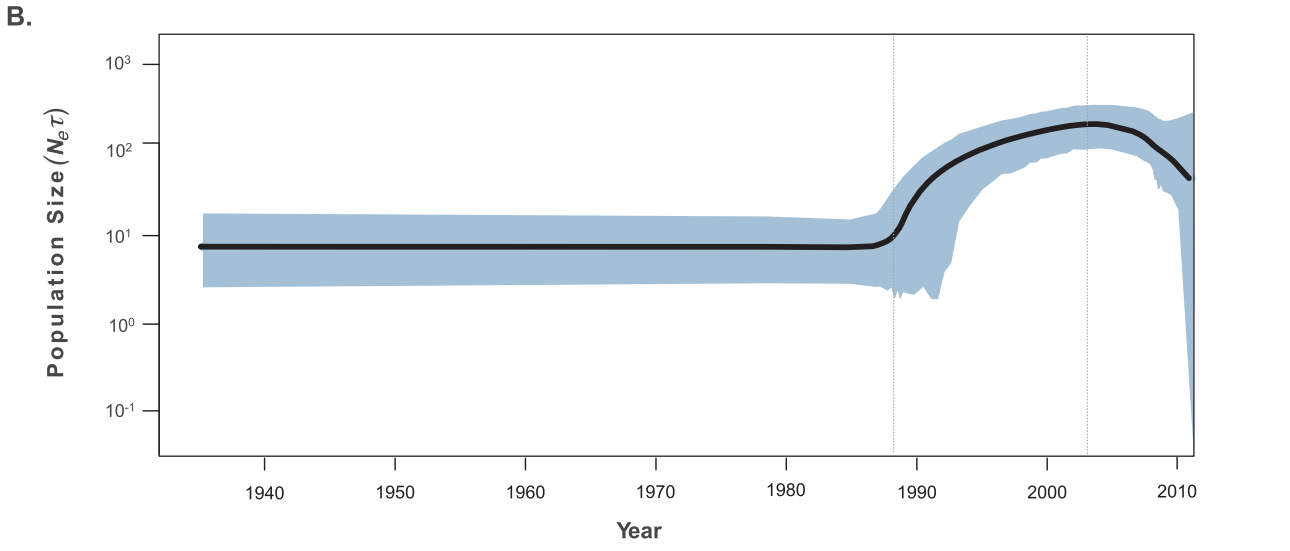


Fig. 1. Phylogenetic relationships of tobacco mosaic virus isolates. (A) Map showing the localities of viral isolates from China that were sequenced in this study. (B) Splits network analysis of the complete genome sequences of 76 tobacco mosaic virus isolates. An isolate of tomato mosaic virus (accession number NC_002692) was used as an outgroup. (C) Maximum-likelihood tree of 76 tobacco mosaic virus isolates based on complete genome sequences. Black circles indicate strong node support (SH-aLRT ≥ 75% and bootstrap support ≥ 95%). Colours indicate sampling locations, as shown in the key. The scale bar is given in substitutions/site.

34–84), suggesting that YRR has played a major role in the evolution and persistence of TMV over the time period investigated.

Reconstruction of the demographic history, using the extended Bayesian skyline plot, revealed that the TMV population in China experienced multiple shifts in size (Fig. 2B). The mean estimated number of population-size changes was 2.45 (95% credibility interval 1–4). The population size was relatively constant from 1935 to 1987, then expanded between 1988 and 2002. From 2003 to 2011, the last





(caption on next page)

Fig. 2. (A) Time-tree showing the evolutionary relationships and timescale of tobacco mosaic virus isolates from China. The tree topology has been chosen to maximize the product of node posterior probabilities. Branch lengths are scaled according to time, as indicated by the horizontal axis. Branch colours denote inferred location states, as shown in the colour key. Grey node bars show the 95% credibility intervals of node-age estimates and are shown only for nodes with posterior probabilities of at least 0.90. The root state posterior probabilities for the five producing regions are shown in the inset panel: SWR, Southwest producing region; SER, Southeast producing region; YRR, the producing region of the upper and middle reaches of Yangtze River; HHR, Huanghuai producing region; and NR, Northern region. Three introductions of tomato mosaic virus into YRR are indicated by purple solid diamonds. (B) Extended Bayesian skyline plot showing population size through time for tobacco mosaic virus in China. The *y* axis represents a measure of genetic diversity, given as the product of effective population size (N_e) and virus generation time (τ). The *x* axis is measured in calendar years. The black line shows the median estimate of the population size and the light blue shading shows the 95% credibility interval. The dotted vertical lines indicate two population-size changes for tobacco mosaic virus in China.

Table 1
Estimates of evolutionary parameters of the genome of tobacco mosaic virus.

Parameter estimate	Mean (95% Highest probability density, HPD)
Date range (year)	10 (2001–2011)
Sample size	56
tMRCA (year)	
All isolates	1935 (1915–1954)
Group I	1985 (1981–1989)
Group II	2001 (1997–2004)
Group III	2008 (2007–2009)
Substitution rate (subs/site/year)	
126 K	4.16×10^{-4} (3.90×10^{-4}–4.70×10^{-4})
183 K[a]	3.92×10^{-4} (3.16×10^{-4}–4.77×10^{-4})
MP	4.03×10^{-4} (3.16×10^{-4}–4.98×10^{-4})
CP	4.38×10^{-4} (3.27×10^{-4}–5.61×10^{-4})

[a] Other parts of the 183 K protein-coding region that excluded 126 K protein-coding one.

sampling year, the TMV population experienced a decline in size.

4. Discussion

In all, we obtained new sequence data for 51 TMV isolates from China. Using these data, we investigated the molecular epidemiology of TMV in this country using Bayesian phylogeographic inference. Our analyses show that the TMV in the world can be divided into four groups (Group I, Group II, Group III and Group IV, Fig. 1C). All the four groups can be found in China. However, more than 94.12% the TMV isolates sequenced in this study fell into Group I, indicating the predominance of this group. The most recent common ancestor of Group I TMV was dated to about 1985, a time when China was beginning to experience a sharp increase in the cultivation area of flue-cured tobacco (Lu et al., 2011; Zhu, 2008).

Our phylogenetic analysis placed the root of the tree for Chinese TMV in SWR with strong support (Fig. 2A and Fig. S2). However, Markov jump estimates between different locations pinpointed NR as an important source of TMV epidemics in China (Fig. 2B and Fig. S3). Similar patterns, in which the locations of viral origin do not match those responsible for dispersal, have been reported recently for two other plant viruses (Olarte Castillo et al., 2011; Stainton et al., 2015). This can be explained by a scenario in which the virus was first introduced into SWR, then it moved north and expanded there. The north then became the main seeding region for the subsequent movements of the virus within China.

Our phylogeographic analysis revealed a general north-to-south dispersal of TMV in China (Fig. 3A, Table S4). This coincides with a north-to-south transition of the production of flue-cured tobacco in China. This suggests that the dispersal pattern of TMV might be associated with multiple human-mediated factors (i.e., market demand and government policy). In particular, farmers' planting behaviour caused by the comparative advantage in the production of the flue-cured tobacco between the Northern and Southern regions plays an important role in the north-to-south dispersal. For example, the enthusiasm of farmers for planting flue-cured tobacco in Huanghuai region and the Northern region was reduced by higher disease incidences and frequent natural disasters after continuous cropping over multiple years. The Southwest region is more suitable for tobacco planting because of its

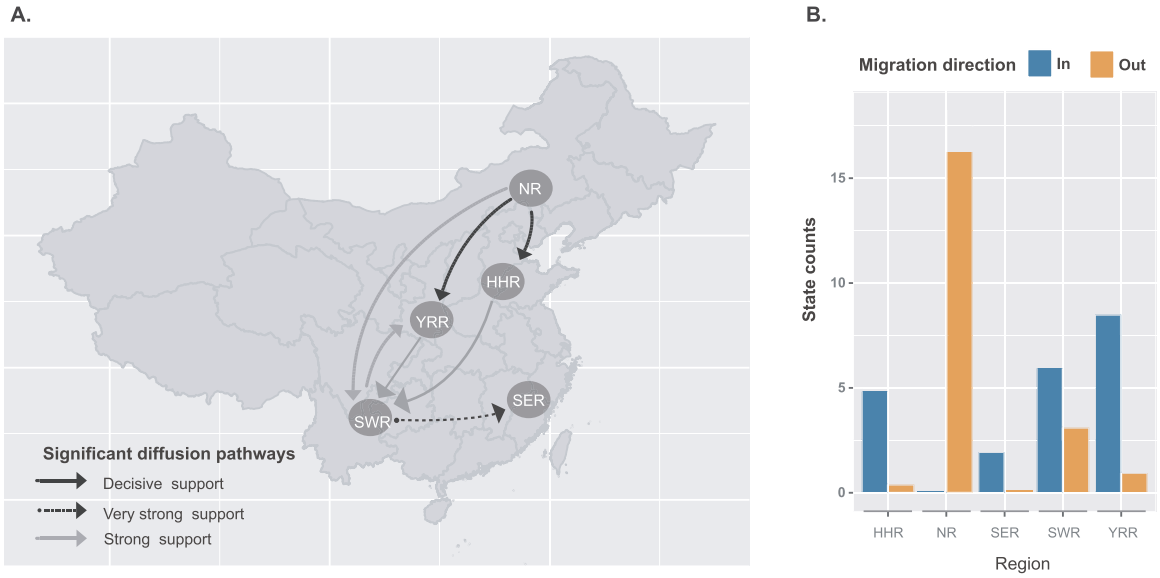


Fig. 3. Spatial diffusion of tobacco mosaic virus in China. (A) Spatial diffusion pathway and (B) histogram of the total number of location state transitions inferred from 56 viral isolates collected from five tobacco-producing regions during 2001–2011. Thickness of lines represents supported migration rates with a mean indicator of > 0.5: solid black arrows, decisive support with BF > 1000; dashed black arrows, very strong support with 150 < BF < 1000; solid grey arrows, strong support with 20 < BF < 150; and dashed grey arrows, supported rates with 3 < BF < 20. SWR, Southwest producing region; SER, Southeast producing region; YRR, the producing region of the upper and middle reaches of Yangtze River; HHR, Huanghuai producing region; and NR, Northern region.

favourable growth conditions, including more suitable temperatures and adequate sunlight. In the past few years, more than half of the tobacco production in China has been concentrated in the Southwest producing region. In particular, Yunnan is the biggest tobacco-planting province, with a cultivation area of more than 400,000 ha (Li et al., 2009), and accounts for about 45% of the total national yield. The majority of tobacco manufacturers in China are supplied from Yunnan.

In contrast with the observed north-to-south patterns, we found one migration pathway with very strong support from SWR to SER from our phylogeographic analyses (Fig. 3A). However, this movement was restricted due to the differences in tobacco cultivars and varieties across the tobacco-producing regions in China. 'Yunyan 87', a tobacco variety with high sensitivity to TMV, was grown mainly in SWR whereas 'Bicui No.1', a variety with moderate sensitivity to TMV, was planted in SER.

Sudden changes in population size can affect the generation, maintenance, and distribution of genetic variation. These effects can occur through shifts in the balance between genetic drift and mutation, but also through impacts on the efficiency of natural selection (Wang and Whitlock, 2003). Indeed, our demographic analyses reveal that TMV populations in China have been small but have undergone recent expansion, possibly associated with an overproduction of tobacco in China in the mid-1990s. Historical accounts indicate an oversupply of tobacco leaf on the market (Hu et al., 2006), reaching a peak of 4.26 million tonnes in China in 1997 (http://faostat.fao.org). In 2002, however, stocks of tobacco held in China steadily declined to 1.33 million tonnes after the State Tobacco Monopolization Administration closed down 11 small and inefficient factories with old equipment (Hu et al., 2010). This also is in accordance with our estimate that the TMV population declined in size from 2002.

The results of our study suggest a link between human activities and the spatio-temporal transmission of TMV across tobacco-producing regions in China. This information provides a platform for developing effective management strategies to control this virus. Although the sampling window of the sequence data analysed in this study might be too short (~10 years) to allow a comprehensive reconstruction of the evolutionary history of TMV in China, our study provided new insights into the geographical dispersal of TMV in China. Further analyses of larger data sets from wider sampling windows and with broader geographic representation will lead to a more comprehensive picture of TMV evolution.

CRediT authorship contribution statement

Fangluan Gao: Formal analysis, Methodology, Writing - original draft, Writing - review & editing. **Xiaowei Liu:** Data curation, Formal analysis, Visualization. **Zhenguo Du:** Formal analysis, Writing - review & editing. **Han Hou:** Data curation, Software, Visualization. **Xiaoyan Wang:** Data curation, Visualization. **Fenglong Wang:** Conceptualization, Project administration, Writing - review & editing. **Jinguang Yang:** Conceptualization, Funding acquisition, Project administration, Writing - review & editing.

Acknowledgments

This work was supported by grants from the Natural Science Foundation of China (Grant no. 31570146, 31772103), the State Tobacco Monopoly Bureau (Grant no. 110201601024(LS-04)), Hongyunhonghe Tobacco (Group) Co., Ltd. (Grant no. HYHH2016YL02), and Yunnan Tobacco Company of China National Tobacco Corporation (Grant no. 2018530000241014), P.R. China. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We thank Prof. Simon Ho (The University of Sydney) for his constructive comments and suggestions on the manuscript. We also thank Prof. Manfred Heinlein (Institute of Molecular Biology of Plants, CNRS, France) for providing sampling information for TMV isolates from France.

Conflicts of interest

The authors declare that they have no conflicts of interest.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.virol.2018.12.001.

References

Alonso, J.M., Gorzny, M.L., Bittner, A.M., 2013. The physics of tobacco mosaic virus and virus-based devices in biotechnology. Trends Biotechnol. 31, 530–538.
Baele, G., Lemey, P., Bedford, T., Rambaut, A., Suchard, M.A., Alekseyenko, A.V., 2012. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. Mol. Biol. Evol. 29, 2157–2167.
Balloux, F., Lugon-Moulin, N., 2002. The estimation of population differentiation with microsatellite markers. Mol. Ecol. 11, 155–165.
Benedict, C., 2011. Golden-Silk Smoke: A History of Tobacco in China. University of California Press, California.
Bielejec, F., Rambaut, A., Suchard, M.A., Lemey, P., 2011. SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. Bioinformatics 27, 2910–2912.
Creager, A.N.H., Scholthof, K.-B.G., Citovsky, V., Scholthof, H.B., 1999. Tobacco mosaic virus: pioneering research for a century. Plant Cell 11, 301–308.
Ding, X.S., Liu, J., Cheng, N.H., Folimonov, A., Hou, Y.M., Bao, Y., Katagi, C., Carter, S.A., Nelson, R.S., 2004. The tobacco mosaic virus 126-kDa protein associated with virus replication and movement suppresses RNA silencing. Mol. Plant Microbe Interact. 17, 583–592.
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.
Duchêne, D.A., Duchêne, S., Ho, S.Y.W., 2018. PhyloMAd: efficient assessment of phylogenomic model adequacy. Bioinformatics 34, 2300–2301.
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.
Excoffier, L., Lischer, H.E., 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. 10, 564–567.
Gibbs, A., Fargette, D., Garcia-Arenal, F., Gibbs, M., 2010. Time – the emerging dimension of plant virus studies. J. Gen. Virol. 91, 13–22.
Gooding, G.V., 1986. Tobacco mosaic virus epidemiology and control. In: Van Regenmortel, M.H.V., Fraenkel-Conrat, H. (Eds.), The Plant Viruses: The Rod-Shaped Plant Viruses. Springer US, Boston.
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.
Hagiwara, Y., Komoda, K., Yamanaka, T., Tamai, A., Meshi, T., Funada, R., Tsuchiya, T., Naito, S., Ishikawa, M., 2003. Subcellular localization of host and viral proteins associated with tobamovirus RNA replication. EMBO J. 22, 344–353.
Harrison, B.D., Wilson, T.M., 1999. Milestones in the research on tobacco mosaic virus. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 354, 521–529.
Heinlein, M., 2002. The spread of tobacco mosaic virus infection: insights into the cellular mechanism of RNA transport. Cell. Mol. Life Sci. 59, 58–82.
Heled, J., Drummond, A., 2008. Bayesian inference of population size history from multiple loci. BMC Evol. Biol. 8, 1–15.
Holmes, F.O., 1951. Indications of a new-world origin of tobacco-mosaic virus. Phytopathology 41, 341–349.
Hu, T.-W., Mao, Z., Shi, J., Chen, W., 2010. The role of taxation in tobacco control and its potential economic impact in China. Tob. Control 19, 58–64.
Hu, T.W., Mao, Z., Ong, M., Tong, E., Tao, M., Jiang, H., Hammond, K., Smith, K.R., de Beyer, J., Yurekli, A., 2006. China at the crossroads: the economics of tobacco and health. Tob. Control 15, i37–i41.
Huson, D.H., 1998. SplitsTree: analyzing and visualizing evolutionary data. Bioinformatics 14, 68–73.
Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermiin, L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14, 587–589.
Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Am. Stat. Assoc. 90, 773–795.
King, A.M., Lefkowitz, E., Adams, M.J., Carstens, E.B., 2011. Virus Taxonomy, Ninth Report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, Amsterdam.
Lemey, P., Rambaut, A., Drummond, A.J., Suchard, M.A., 2009. Bayesian phylogeography finds its roots. PLOS Comput. Biol. 5, e1000520.
Li, C., He, X., Zhu, S., Zhou, H., Wang, Y., Li, Y., Yang, J., Fan, J., Yang, J., Wang, G., Long, Y., Xu, J., Tang, Y., Zhao, G., Yang, J., Liu, L., Sun, Y., Xie, Y., Wang, H., Zhu, Y., 2009. Crop diversity for yield increase. PLOS ONE 4, e8049.
Li, J.G., Cao, J., Sun, F.F., Niu, D.D., Yan, F., Liu, H.X., Guo, J.H., 2011. Control of tobacco mosaic virus by PopW as a result of induced resistance in tobacco under greenhouse and field conditions. Phytopathology 101, 1202–1208.
Li, Y., Feng, J., Liu, H., Wang, L., Hsiang, T., Li, X., Huang, J., 2016. Genetic diversity and

pathogenicity of *Ralstonia solanacearum* causing tobacco bacterial wilt in China. Plant Dis. 100, 1288–1296.
Liu, C., Nelson, R., 2013. The cell biology of tobacco mosaic virus replication and movement. Front. Plant Sci. 4, 1–10.
Lu, K., Xiong, Z.G., Zhu, K., Yang, D.H., Zheng, W., Zi, W.H., Jiang, M.H., 2011. Review and discussion on history of flue-cured tobacco production in China. J. Kunming Univ. 33, 48–52.
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, vev003.
Minh, B.Q., Nguyen, M.A., von Haeseler, A., 2013. Ultrafast approximation for phylogenetic bootstrap. Mol. Biol. Evol. 30, 1188–1195.
Minin, V.N., Suchard, M.A., 2008. Counting labeled transitions in continuous-time Markov models of evolution. J. Math. Biol. 56, 391–412.
Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2014. IQ-tree: a fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. Mol. Biol. Evol. 32, 268–274.
Olarte Castillo, X.A., Fermin, G., Tabima, J., Rojas, Y., Tennant, P.F., Fuchs, M., Sierra, R., Bernal, A.J., Restrepo, S., 2011. Phylogeography and molecular epidemiology of papaya ringspot virus. Virus Res. 159, 132–140.
Pagan, I., Firth, C., Holmes, E.C., 2010. Phylogenetic analysis reveals rapid evolutionary dynamics in the plant RNA virus genus tobamovirus. J. Mol. Evol. 71, 298–307.
Rambaut, A., Lam, T.T., Max Carvalho, L., Pybus, O.G., 2016. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol. 2, vew007.
Ramsden, C., Melo, F.L., Figueiredo, L.M., Holmes, E.C., Zanotto, P.M., 2008. High rates of molecular evolution in hantaviruses. Mol. Biol. Evol. 25, 1488–1492.
Rieux, A., Khatchikian, C.E., 2017. TipDatingBeast: an R package to assist the implementation of phylogenetic tip-dating tests using BEAST. Mol. Ecol. Resour. 17, 608–613.
Scholthof, K.B., 2004. Tobacco mosaic virus: a model system for plant biology. Annu. Rev. Phytopathol. 42, 13–34.
Shen, L., Wang, F., Yang, J., Qian, Y., Dong, X., Zhan, H., 2014. Control of tobacco mosaic virus by *Pseudomonas fluorescens* CZ powder in greenhouses and the field. Crop Prot. 56, 87–90.
Stainton, D., Martin, D.P., Muhire, B.M., Lolohea, S., Halafihi, M., Lepoint, P., Blomme, G., Crew, K.S., Sharman, M., Kraberger, S., Dayaram, A., Walters, M., Collings, D.A., Mabvakure, B., Lemey, P., Harkins, G.W., Thomas, J.E., Varsani, A., 2015. The global distribution of banana bunchy top virus reveals little evidence for frequent recent, human-mediated long distance dispersal events. Virus Evol. 1, vev009.
Valle, R.P., Drugeon, G., Devignes-Morch, M.D., Legocki, A.B., Haenni, A.L., 1992. Codon context effect in virus translational readthrough. A study in vitro of the determinants of TMV and Mo-MuLV amber suppression. FEBS Lett. 306, 133–139.
Wang, J., Whitlock, M.C., 2003. Estimating effective population size and migration rates from genetic samples over space and time. Genetics 163, 429–446.
Zhu, J., 2008. Analyses of the tobacco leaf production trend in world and the northern-to-southward movement of tobacco production in China. World Agric. 351, 22–24.