Virology 525 (2018) 200–204

Contents lists available at ScienceDirect

Virology

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

Brief communications

Europe was a hub for the global spread of potato virus S in the 19th century

Guohua Duan[a,1], Fangfang Zhan[a,1], Zhenguo Du[a], Simon Y.W. Ho[b], Fangluan Gao[a,b,*]

[a] Fujian Key Laboratory of Plant Virology, Institute of Plant Virology, Fujian Agriculture and Forestry University, Fuzhou 350002, PR China
[b] School of Life and Environmental Sciences, University of Sydney, Sydney, NSW 2006, Australia

ARTICLE INFO

Keywords:
Potato virus S
Bayesian phylogeography
Molecular clock
Model selection

ABSTRACT

Potato virus S (PVS) is a major plant pathogen that causes considerable losses in global potato production. Knowledge of the evolutionary history and spatio-temporal dynamics of PVS is vital for developing sustainable management schemes. In this study, we investigated the phylodynamics of the virus by analysing 103 nucleotide sequences of the coat protein gene, sampled between 1985 and 2014. Our Bayesian phylogenetic analyses showed that PVS has been evolving at a rate of 3.32×10^{-4} substitutions/site/year (95% credibility interval 1.33×10^{-4}–5.58×10^{-4}). We dated the crown group to the year 1325 CE (95% credibility interval 762–1743 CE). Our phylogeographic analyses pointed to viral origins in South America and identified multiple migration pathways between Europe and other regions, suggesting that Europe has been a major hub for PVS transmission. The results of our study have potential implications for developing effective strategies for the control of this pathogen.

1. Introduction

Potato virus S (PVS) is a member of the genus *Carlavirus* in the family *Betaflexiviridae* (King et al., 2011) and has become one of the most prevalent viruses in potato crops around the world. Depending on whether or not they are able to cause systemic infection in *Chenopodium* spp., PVS isolates from potato are biologically classified into the two strains PVS^A (Andean) and PVS^O (Ordinary), respectively. PVS^A can infect *Chenopodium* spp. systemically, whereas PVS^O induces local lesions but cannot infect the plant systemically (Cox and Jones, 2010). Moreover, PVS^A is transmitted by aphids in a non-persistent manner and is reported to induce more severe leaf symptoms on potato plants than does PVS^O (Slack, 1983). In addition to transmission by aphids, PVS is transmitted mechanically and by vegetative propagation (Franc and Banttari, 1996; Khalil and Shalla, 1982).

More than 40 plant viruses infect potato crops, which are vegetatively propagated. PVS ranks behind potato leafroll virus and potato virus Y as the third most important pathogen affecting potato production worldwide (Lacomme et al., 2017). It is highly prevalent in potato seed but produces mild symptoms in most potato varieties. However, it can cause serious reductions in tuber yield when occurring in mixed infections with other viruses, such as potato virus A, potato virus X, and potato virus Y (Loebenstein et al., 2001). Since it was first reported in Europe in 1952 (De Bruyn Ouboter, 1952), PVS has spread throughout the major potato-planting areas across the world. Several studies have found serological and pathotype diversity among PVS isolates (Gutiérrez et al., 2013; Santillan et al., 2018). Knowledge of the evolutionary history and spatio-temporal dynamics of PVS is vital for developing sustainable management schemes.

PVS has a single-stranded, positive-sense RNA genome of ~ 8.5 kb. The viral genome is encapsidated in a 34-kDa coat protein (CP) and contains a 5′ cap structure, six open reading frames, and a poly-A tail at the 3′ terminus. The six open reading frames encode RdRp (RNA-dependent RNA-polymerase, 223 kDa), TGBp1–3 (triple gene-block proteins of 25 kDa, 12 kDa, and 7 kDa), CP (32.3 kDa), and NABP (cysteine-rich nucleic-acid-binding protein, 11 kDa) (Martelli et al., 2007).

Many potato viruses are globally distributed, but their important migration pathways have not yet been identified. Similarly, few studies have focused on the evolutionary dynamics and timescale of the carlaviruses (Benitez-Galeano et al., 2017). As with most carlaviruses, the biology of PVS has been intensively studied for several decades. The most recent common ancestor of PVS lineages has been dated at 1067 CE (95% credibility interval 68–1369 CE) and the substitution rate of the PVS genome has been estimated at 4.8×10^{-4} subs/site/year (Santillan et al., 2018). However, the evolutionary dynamics of this pathogen, particularly its global spatial spread, is poorly understood. In this study, we use a Bayesian phylogenetic approach to estimate the evolutionary rate of the CP gene and to investigate the spatio-temporal

* Correspondence to: Fujian Agriculture and Forestry University, Fuzhou 350002, Fujian, PR China.
E-mail address: raindy@fafu.edu.cn (F. Gao).
[1] G. Duan and F. Zhan contributed equally to this work.

https://doi.org/10.1016/j.virol.2018.09.022
Received 27 August 2018; Received in revised form 26 September 2018; Accepted 26 September 2018

dynamics of PVS. This gene is commonly used in evolutionary studies of plant viruses (Gao et al., 2018, 2017; Moury and Simon, 2011; Xu et al., 2017). In particular, the N-terminal variable region of the CP gene is believed to be involved in the differences between the biological properties of PVS^A and PVS^O.

2. Materials and methods

2.1. Data set

We obtained the CP gene sequences of 103 PVS isolates from GenBank (Table S1). These sequences had been collected from 13 countries between 1985 and 2014, and had known sampling dates, geographic locations, and host origins. The isolates came from six geographic regions: Asia ($n = 19$), Europe ($n = 27$), Middle East ($n = 27$), North America ($n = 8$), Oceania ($n = 14$), and South America ($n = 8$). Codon-based sequence alignments were performed using the MAFFT algorithm (Katoh and Standley, 2013) in TranslatorX (Abascal et al., 2010). To test for the presence of recombinants, we used seven different algorithms, including RDP, GENECONV, BOOTSCAN, MAXCHI, CHIMAERA, SISCAN, and 3SEQ, implemented in RDP 4.95 (Martin et al., 2015). To reduce the presence of false positives, we only considered any recombination events supported by at least four of the seven methods with $P < 10^{-6}$. No significant signals of recombination were discovered and so we used the complete data set for all of our subsequent analyses. Excluding the stop codons, the sequence alignment had a length of 882 nucleotides.

2.2. Temporal dynamics of potato virus S

To infer the evolutionary rate and timescale of PVS, we performed a Bayesian phylogenetic analysis using BEAST 1.8.4 (Drummond et al., 2012). The sequences were analysed using the GTR+Γ_4 substitution model, which was selected using the Bayesian information criterion by ModelFinder (Kalyaanamoorthy et al., 2017) in IQ-TREE 1.5.5 (Nguyen et al., 2014). We used PhyloMAd (Duchêne et al., 2018) to confirm the absolute fit ('adequacy') of the selected substitution model (Fig. S1). Using marginal likelihoods estimated by path sampling (Baele et al., 2012), we compared the constant-size, exponential-growth, and Bayesian skyline coalescent tree priors, as well as the strict molecular clock and uncorrelated lognormal relaxed clock (Drummond et al., 2006). A constant-size coalescent tree prior and uncorrelated lognormal relaxed clock provided the best fit to our sequence data (Table S2). The sampling times of the viral isolates were used to calibrate the molecular clock. Posterior distributions of parameters were estimated by Markov chain Monte Carlo (MCMC) sampling, with samples drawn every 10^4 steps over 10^8 steps. Sufficient effective sample sizes (> 200) and convergence to the stationary distribution were checked using Tracer 1.7 (Rambaut et al., 2018) after we discarded the first 10% of samples as burn-in.

Bayesian analyses of time-structured data sets can be misled when population structure is not adequately modelled (Möller et al., 2018). To exclude this potential effect on molecular dating, we first conducted a Mantel test of the correlation between pairwise genetic distances and differences in sampling dates and calculated the *P*-value of this test with 1000 permutations (Murray et al., 2016). We found no evidence of confounding of temporal and genetic structure in the data ($P = 0.935$, Fig. S2A, Fig. S2B). We then confirmed the presence of temporal signal in the data set (Fig. S2C) using a date-randomization test (Duchêne et al., 2015; Ramsden et al., 2008), based on 20 permutations of the sampling dates produced using the TipDatingBEAST package (Rieux and Khatchikian, 2017). As an additional assessment on the effect of population structure on evolutionary dynamics, we performed a Bayesian phylogenetic analysis using a structured coalescent tree prior and strict clock, as described by Vaughan et al. (2014). We also estimated the evolutionary rate using a regression of root-to-tip genetic distances against year of sampling in TempEst 1.5 (Rambaut et al., 2016). For this analysis, we inferred the tree topology and branch lengths using maximum likelihood under the GTR+Γ_4 substitution model in IQ-TREE.

2.3. Bayesian phylogeographic analysis

To explore the spatial diffusion patterns of PVS through time, we employed an asymmetric continuous-time Markov chain phylogeographic model in BEAST, coupled with model averaging using Bayesian stochastic search variable selection (Lemey et al., 2009). Six geographical locations, as described above, were selected and coded as discrete states. Well-supported pairwise diffusions were identified using Bayes factors in SPREAD3 0.9.7 (Bielejec et al., 2016). These migration pathways were based on a combination of a Bayes factor > 3 and a mean indicator of > 0.5. Our interpretation of Bayes factors followed 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) using the asymmetric migration model described above, and plotted the total number of state counts for migration into and out of each location. To assess the reliability of the most plausible location at the root node, we compared the results with those from 20 replicate data sets in which the location states were randomized among the sequences.

3. Results and discussion

Our estimated mean substitution rate for CP was 3.32×10^{-4} subs/site/year (95% credibility interval 1.33×10^{-4}–5.58×10^{-4}). This is similar to the estimate of 2.71×10^{-4} subs/site/year (95% credibility interval 1.37×10^{-4}–4.10×10^{-4}) using the structured coalescent tree prior. The estimate from our Bayesian analysis is also consistent with our estimate of 3.45×10^{-4} subs/site/year from a regression of root-to-tip genetic distances against the sampling dates. Our estimates are similar to a recent rate estimate of 4.8×10^{-4} subs/site/year for th whole genome of PVS (Santillan et al., 2018). Given the results of the date-randomization test (Fig. S2C) and the consistency among our estimates, we consider the substitution rate inferred from our Bayesian analysis to be reliable. Further analysis of sequence data from a wider sampling window will provide insight into the applicability of our rate estimate across broader evolutionary timescales.

Our time-scaled maximum-clade-credibility tree showed that PVS isolates could be separated into three lineages (Fig. 1). Eighty-five isolates were placed into the PVS^O lineage and 17 isolates were placed into the PVS^A lineage. PVS^O forms the sister lineage to a highly divergent isolate from Colombia (PVS-RVC, GenBank accession number JX419379) with a posterior probability of 0.92. Our phylogenetic results support a previous study based on the CP gene that suggested that RVC might represent a new PVS lineage (Fernando et al., 2013), which was confirmed by Gutiérrez et al. (2013) based on a complete genome sequence. One plausible explanation for the divergent nature of RVC is that it was isolated from *Solanum phureja*, a host species native to the Andes, and that it might be the product of genetic exchange with other PVS strains (Gutiérrez et al., 2013).

Our Bayesian analysis places the root of the tree in South America, with a posterior probability of 0.4 (Fig. 1). This is outside the range of probabilities (0.06–0.32) obtained in our analyses of 20 location-randomized data sets (Fig. S3). The potato, *Solanum tuberosum*, originated in the Andean region (modern-day southern Peru). The first likely cultivation of potatoes has been dated to 400 CE (Grun, 1990). However, we have dated the crown group of PVS to the year 1325 CE (95% credibility interval 762–1743 CE), similar to a recent date estimate for the most recent common ancestor of 1067 CE (95% credibility interval 68–1369 CE; Santillan et al., 2018). These estimates suggest that the earlier PVS isolates from South America were either not sampled or have not survived to the present day.

The date estimates from our analyses place the most recent common