Virology 487 (2016) 188–197

Contents lists available at ScienceDirect

Virology

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

Temporal analysis of reassortment and molecular evolution of *Cucumber mosaic virus*: Extra clues from its segmented genome

Kazusato Ohshima [a,b,*], Kosuke Matsumoto [a], Ryosuke Yasaka [a,b], Mai Nishiyama [a], Kenta Soejima [a], Savas Korkmaz [c], Simon Y.W. Ho [d], Adrian J. Gibbs [e], Minoru Takeshita [f]

[a] Laboratory of Plant Virology, Faculty of Agriculture, Saga University, Saga, Japan
[b] The United Graduate School of Agricultural Sciences, Kagoshima University, Kagoshima, Japan
[c] Department of Plant Protection, Faculty of Agriculture, University of Canakkale Onsekiz Mart, Canakkale, Turkey
[d] School of Biological Sciences, University of Sydney, Sydney, New South Wales, Australia
[e] Emeritus Faculty, Australian National University, Canberra, Australia
[f] Laboratory of Plant Pathology, Faculty of Agriculture, University of Miyazaki, Miyazaki, Japan

ARTICLE INFO

Article history:
Received 8 August 2015
Returned to author for revisions
29 August 2015
Accepted 28 September 2015
Available online 3 November 2015

Key words:
Cucumber mosaic virus
Cucumovirus
Evolution
Timescale
Reassortment
Selection
Synonymous sites

ABSTRACT

Cucumber mosaic virus (CMV) is a damaging pathogen of over 200 mono- and dicotyledonous crop species worldwide. It has the broadest known host range of any virus, but the timescale of its evolution is unknown. To investigate the evolutionary history of this virus, we obtained the genomic sequences of 40 CMV isolates from brassicas sampled in Iran, Turkey and Japan, and combined them with published sequences. Our synonymous ('silent') site analyses revealed that the present CMV population is the progeny of a single ancestor existing 1550–2600 years ago, but that the population mostly radiated 295–545 years ago. We found that the major CMV lineages are not phylogeographically confined, but that recombination and reassortment is restricted to local populations and that no reassortant lineage is more than 251 years old. Our results highlight the different evolutionary patterns seen among viral pathogens of brassica crops across the world.

Introduction

Most brassicas now grown as crops were first domesticated in South-West Eurasia (e.g. Crisp, 1995; Hodgkin, 1995). Viruses damage these crops worldwide. Some of the most widespread and well studied of these pathogens are *Turnip mosaic virus* (TuMV), *Cauliflower mosaic virus* (CaMV) and *Cucumber mosaic virus* (CMV). Planning their control is improved if we know when they first became established in their host populations, and the locations of their 'centres of emergence' (Ohshima et al., 2002; Nguyen et al., 2013).

CMV is the type species of the genus *Cucumovirus* in the family *Bromoviridae* (King et al., 2012). CMV is distributed worldwide and has the broadest known host range of any virus, infecting more than 1000 species of plants, including monocots and dicots, herbaceous plants, shrubs, and trees. It is transmitted by seed and by more than 60 aphid species in a non-persistent manner (Palukaitis and García-Arenal, 2003; Jacquemond, 2012). However, we cannot be sure whether this great diversity of hosts and vectors reflects a recent rapid radiation of the virus into man-made ecosystems, or results from a more ancient origin. This uncertainty is best resolved by a molecular dating analysis.

CMV has isometric virions with a tripartite genome of positive-sense, single-stranded RNAs of about 8700 nucleotides (nt) with five open reading frames (ORFs). The 1a and 2a ORFs, encoded on RNAs 1 and 2, respectively, are translated into the viral components of the replicase. The 2b ORF, which overlaps the 3′-end of the 2a ORF, is expressed from a subgenomic RNA, RNA 4A (Ding et al., 1994). It encodes a suppressor of posttranscriptional gene silencing (Brigneti et al., 1998), and probably arose by overprinting (Ding et al., 1994; Keese and Gibbs, 1992). RNA 3 encodes the 3a protein, which is the viral movement protein, and the coat protein (CP), which is expressed from subgenomic RNA 4 (for reviews see Palukaitis and García-Arenal, 2003; Jacquemond, 2012).

Roossinck et al. (1999) and Roossinck (2002) reported that the CMV population had two clear lineages, which they designated subgroups I and II. Subgroup I was further divided, especially by the 2b ORF, into subgroups IA and IB. However none of these subgroupings, defined by maximum-parsimony analysis (Swofford

* Corresponding author at: Laboratory of Plant Virology, Faculty of Agriculture, Saga University, Saga, Japan.
E-mail address: ohshimak@cc.saga-u.ac.jp (K. Ohshima).

http://dx.doi.org/10.1016/j.virol.2015.09.024
0042-6822/

and Wilgenbusch, 2003) of gene sequences, was correlated with other features of the isolates, such as host range or provenance. There were also differences between the gene trees estimated from different RNA segments using maximum parsimony, but whether this resulted from differences in evolutionary rates, reassortment, or lineage sampling, was not identified (White et al., 1995). There have been attempts to estimate the evolutionary timescales of other viruses with segmented genomes, including *Bluetongue virus* (Carpi et al., 2010), and *Influenza A virus* (Dudas et al., 2014; Westgeest et al., 2014), and these reported differences in the mean substitution rate of different ORFs. Studies of segmented genomes are valuable because they provide opportunities to test the performance of methods of estimating evolutionary rates and timescales using real data.

Several studies have reported the genetic structure of CMV populations in Spain, Italy, South Korea, and USA based on analyses of entire genomes or partial genomic sequences (Fraile et al., 1997; García-Arenal et al., 2000; Lin et al., 2004; Kim et al., 2014; Nouri et al., 2014). A recent review described the population genetics of CMV on a global scale and noted the importance of analyzing sequences of all three of its genome segments (Jacquemond, 2012). The complete nucleotide sequences of 106 CMV genomes have been reported so far (August, 2014). However, most are from countries of East Asia (e.g., 54, 13, and 11 isolates from South Korea, China, and Japan, respectively) and from solanaceous hosts. Data from a larger sample of the global CMV population are clearly required to characterize the full population structure of the virus across time and space.

Here, we report the genomic sequences of 40 more CMV isolates, mostly from brassicas in the Eurasian region (Iran and Turkey) and from Japan, collected during 2000–2013. We analyzed these sequences, in combination with published sequences, to estimate the timescale and rate of evolution of the individual genes of CMV, and the dates of reassortment events in the CMV population. We compared these estimates from complete sequences with those from which non-synonymous and invariate codons had been removed.

Results and discussion

Genome sequences

We sequenced the complete genomes of 13 isolates from Iran, 20 from Japan, and seven from Turkey (Table S1). The lengths of the genomic segments of these isolates were 3319–3337 nt for RNA 1, 3012–3039 nt for RNA 2, and 2160–2194 nt for RNA 3, excluding the 5′ non-coding region (NCR) primer sequences (26 nt, 25 nt, and 22 nt for RNA 1, RNA 2, and RNA 3, respectively) and 3′ NCR primer sequence of 10 nt (5′-AAGGAGACCA-3′). These gave ORFs of 2979 nt (1a protein), 2571–2583 nt (2a protein), 330–339 nt (2b protein), 837–840 nt (3a protein), and 654–657 nt (CP). Furthermore, the overlapping regions located between 2a and 2b proteins of most isolates were 242 nt in length, whereas that of TRT611J was 251 nt. The new genomic sequences determined in this study are available in international nucleotide sequence databases with Accession numbers LC066399 – LC066518.

Published data comprising a further 144, 152, and 203 sequences with complete ORFs and no ambiguous nucleotides for CMV RNA 1, RNA 2, and RNA 3, respectively, were obtained from public databases. One hundred and six of these represented complete genomes. The concatenated 5′ NCR plus main ORF(s) and 3′ NCR sequences of individual genomic segments were aligned with those of the ER isolate of *Peanut stunt virus* (PSV) (Accession numbers U15728-30) as an outgroup taxon.

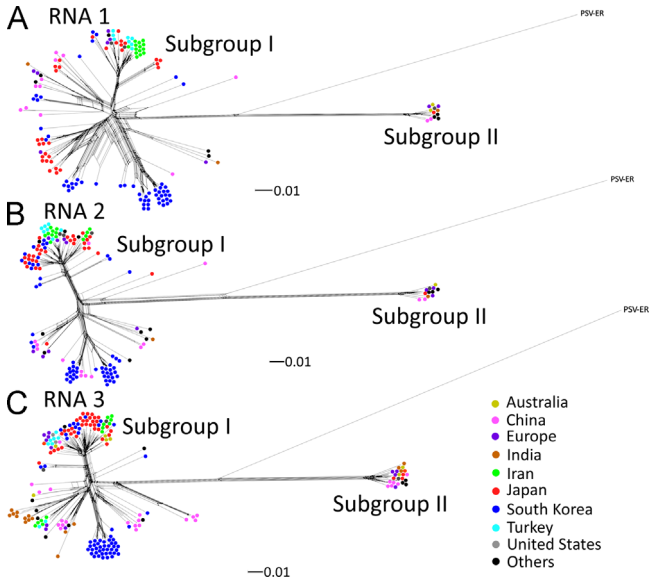

Fig. 1. Phylogenetic networks of RNA sequences of *Cucumber mosaic virus*. The network inferred from segments (A) RNA 1, (B) RNA 2 and (C) RNA 3 of Iran, Japan, Turkey and worldwide isolates. The three homologous RNA sequences of *Peanut stunt virus* ER isolate (PSV-ER) (Accession Numbers, U15728-30) were used as the outgroup. The isolates obtained in this study are listed in Table S1.

Phylogenetic and recombination analyses

We used SPLITSTREE v4.11.3 (Huson and Bryant, 2006) to analyze all the complete CMV sequences, including the concatenated 5′ NCR plus main ORFs and 3′ NCR regions. Incomplete segments and those that contained ambiguous nucleotides were discarded. The analysis yielded reticulated phylogenetic networks (Fig. 1), reflecting conflicts in the phylogenetic signal that are possibly due to the presence of recombinant sequences (see below). Furthermore, although there were small-scale geographical groupings of isolates in the networks, there was no major congruence between the relationships among the isolates and their provenance. This is despite the data having probably captured a representative sample of the global genetic diversity of CMV, including the geographical regions in which various brassicas were first domesticated. We confirmed this result using a Bayesian phylogeographic analysis (Lemey et al., 2009) of the CMV data (results not shown). The lack of a phylogeographic signal in the CMV data stands in contrast with those obtained from our earlier studies of TuMV (Ohshima et al., 2002; Nguyen et al., 2013; Yasaka et al., 2015) and CaMV (Yasaka et al., 2014). It is possible that there are significant parts of the CMV population in non-brassica hosts and these have not yet been sampled. Although this might have biased our results, there might be genuine biogeographic differences among CMV, TuMV, and CaMV.

The sequences were then checked for recombination using the RDP4 package (Martin et al., 2015). The genomic segments of 100 isolates appeared to be free of recombination. Each of the identified recombination sites was examined individually, and the phylogenetic relationships were used to verify the parent/donor assignments. We found many recombinant segments with an associated *P*-value of around 10^{-6} or lower (Table S2). Recombination sites were mostly found in the genomes of isolates from Asian countries of India, Japan, and South Korea.

Phylogenetic relationships were estimated using maximum likelihood (ML) (PhyML; Guindon and Gascuel, 2003) analysis of each of the five ORFs, after all recombinant sequences had been removed (Table S2). Trees obtained for all non-recombinant sequences (i.e., 112, 143, 399, 198, and 198 sequences for the 1a, 2a, 2b, 3a, and CP ORFs; Fig. S1) had very similar topologies, showing a major primary