Population structure of blueberry mosaic associated virus: Evidence of reassortment in geographically distinct isolates

Thanuja Thekke-Veetil a, James J. Polashock b, Mojca V. Marn c, Irena M. Plesko c, Annemiek C. Schilder d, Karen E. Keller e, Robert R. Martin e, Ioannis E. Tzanetakis a,∗

a Department of Plant Pathology, Division of Agriculture, University of Arkansas System, Fayetteville, AR 72701, United States
b USDA-ARS, CIFRL, Chatsworth, NJ 08019, United States
c Agricultural Institute of Slovenia, Hacquetova 17, Ljubljana, Slovenia
d Department of Plant, Soil and Microbial Sciences, Michigan State University, East Lansing, MI 48824, United States
e USDA-ARS, Corvallis, OR 97330, United States

A R T I C L E   I N F O

Article history:
Received 15 October 2014
Received in revised form 19 February 2015
Accepted 22 February 2015
Available online 28 February 2015

Keywords:
Diversity
Genetic variation
Selection
Ophiovirus
Reassortment

A B S T R A C T

The population structure of blueberry mosaic associated virus (BIMaV), a putative member of the family Ophioviridae, was examined using 61 isolates collected from North America and Slovenia. The studied isolates displayed low diversity in the movement and nucleocapsid proteins and low ratios of non-synonymous to synonymous nucleotide substitutions, indicative of strong purifying selection. Phylogenetic analyses revealed grouping primarily based on geography with some isolates deviating from this rule. Phylogenetic incongruence in the two regions, coupled with detection of reassortment events, indicated the possible role of genetic exchange in the evolution of BIMaV.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Blueberry mosaic has been reported from several regions around the world (Martin et al., 2012). Diseased plants display mild to brilliant mosaic and mottling on the foliage with colors of yellow, yellow–green and sometimes pink (Varney, 1957) as well as delayed fruit ripening, reduced yield and fruit quality (Ramsdell and Stretch, 1987). Recently, a new member of the Ophioviridae, blueberry mosaic associated virus (BIMaV), was detected in all symptomatic material tested (Thekke-Veetil et al., 2014). Ophioviruses are a group of negative-strand RNA viruses with segmented genomes comprised of three or four RNAs. The genome of BIMaV is comprised of three RNAs (RNAs 1–3) encoding four proteins in the viral complementary strand (Fig. 1). RNA 1 encodes a 272 kDa replicase and a 23 kDa protein of unknown function. RNAs 2 and 3 code for a 58 kDa movement protein (MP) and 40 kDa nucleocapsid protein (NP) respectively (Thekke-Veetil et al., 2014).

Processes that increase genetic variability are key drivers of evolution as they provide viruses the tools to mediate the ever evolving arms race between pathogen and host. RNA viruses generally display elevated genetic variation because of their error-prone replication and genetic exchange through recombination or segment reassortment (Domingo and Holland, 1994; Nagy, 2008). Distinct genetic variants in a virus population may differ in severity of symptom expression, in their host range or in the transmission efficiency by their biological vectors. Examples are the mild and severe strains of citrus tristeza virus, cucumber mosaic virus and tomato spotted wilt virus which cause symptoms ranging from mild mosaic to plant death (Aramburu et al., 2010; Lin et al., 2003; Roistacher and Moreno, 1991). Analysis of virus population structure allows the grouping of isolates with unique molecular and possibly biological characteristics and provides information that could be used to predict the durability of genetic or transgenic resistance. Therefore, knowledge of the virus population structure is important in understanding disease development and implementing effective management strategies including development of reliable diagnostic assays for screening propagation material. In addition, information on the virus population structure provides insights on the virus dynamics and factors influencing virus evolution (Garcia-Arenal and McDonald, 2003).

There have been several reports on the population structure of positive-strand RNA viruses (Alabi et al., 2011; Fargette et al., 2004;...
Lopez et al., 1998; Poudel et al., 2012; Poudel and Tzanetakis, 2013; Rubio et al., 2001; Thekke-Veetil et al., 2013; Tomimura et al., 2004; Turturo et al., 2005; Vigne et al., 2004; Wang et al., 2011). On the other hand, research on negative-strand plant RNA viruses is limited (Callaghan and Dietzgen, 2005; Miranda et al., 2000; Tentchev et al., 2011; Wei et al., 2009) and in the case of oophoviruses they are limited to mirafiori lettuce big-vein virus (MiLBV) and citrus psorosis virus (CPSV) (Alioto et al., 2003; Martin et al., 2006; Navarro et al., 2005). This communication reports on the population structure of 61 BlaMV isolates obtained from Canada, Slovenia and the United States and analyzes the potential role of evolutionary forces shaping the genetic structure of the population.

2. Methods

2.1. Virus isolates, amplification and sequencing

Isolates were collected from symptomatic plants between 2009 and 2013 (Supplementary Table 1) from British Columbia in Canada (BC1 to 2), Slovenia (Slo1 to 9) and the United States (Arkansas, AR1 to 5; Kentucky, KY1 to 4; Michigan, MI1 to 6; New Jersey, NJ1 to 11; Oregon, OR1 to 22). Presence of BlaMV in the samples was confirmed by RT-PCR as described (Poudel et al., 2013) using the virus specific primer set [(+ 5'-CCCCGGATTTCCCGAACCTT-3'; (- 5'-ATTCTGACACCCCCGTCTTG-3')] designed from the MP region, with the PCR program consisting of initial denaturation at 94 °C for 3 min followed by 40 cycles of 94 °C for 20 s, 54 °C for 20 s and 72 °C for 40 s.

The amplification of the two regions used in the study was done as follows: cDNAs were diluted 1:2 in water to minimize the effect of potential inhibitors in downstream reactions. One thousand two hundred (1200) nucleotide (nt) regions in the MP and NP genes were amplified (Supplementary Table 2) using either Taq DNA polymerase (Genscript Corporation, NJ) or Phire Hot Start II DNA Polymerase enzymes (Thermo Fisher Scientific Inc., MA) according to manufacturers’ instructions. More than one primer pairs were required to amplify both regions in all isolates due to the sequence variations at the priming sites. Amplicons were cloned and at least three recombinant clones were sequenced in both directions and assembled essentially as described earlier (Thekke-Veetil et al., 2013). The sequences were deposited in GenBank under accession numbers KJ849092-KJ849150, KP793911-2 (MP region) and KJ849151-KJ849209, KP793913-4 (NP region).

2.2. Phylogenetic analysis, selection pressure and detection of recombination events

Nucleotide sequences were aligned using Clustal W (Thompson et al., 1994). Variations in nt and predicted amino acid (aa) sequences were determined using BioEdit (Hall, 1999). The phylogenetic trees were constructed using the maximum likelihood method with 1000 bootstrap replicates in MEGA v.6 (Tamura et al., 2013). The trees were redrawn after collapsing all the branches with <70% bootstrap support as those were considered unreliable.

SNAP (Korber, 2000) was used to estimate $d_{N/d_{S}}$ where $d_{N}$ represents the average number of non-synonymous substitutions per non-synonymous site and $d_{S}$ represents the average number of synonymous substitutions per synonymous site (Nei and Gojobori, 1986). Depending on the $d_{N}/d_{S}$ ratio, the selection pressure on the genes was considered negative or purifying ($d_{N}/d_{S} < 1$), neutral ($d_{N}/d_{S} = 1$), or positive or diversifying ($d_{N}/d_{S} > 1$). Cumulative behavior of codon mutations and distribution of mutations in the regions of analysis were also obtained from SNAP. Screening for positive selection on codons was conducted using three different codon-based maximum likelihood methods (SLAC, FEL and REL) (http://www.datamonkey.org; Kosakovsky Pond and Frost, 2005). Synonymous codon usage bias was measured by calculating the effective number of codons (ENC, Wright, 1990) used in a gene by applying DnaSP v5 (Librado and Rozas, 2009). ENC values range from 20 (the number of amino acids), indicative of maximum bias as only one codon is used from each synonymous–codon group, to 61 (the number of sense codons), which indicates no codon-usage bias where all synonymous codons for each amino acid are equally used.

Putative recombination/reassortment events in the MP/RNA2 and NP/RNA3 were investigated using nine algorithms employed in RDP4 (Martin et al., 2010). Analyses were conducted on the two regions separately and also after concatenating the sequences of both regions into one for each isolate. Only events supported by at least five RDP4-implemented algorithms were considered plausible recombination/reassortment events. The recombinants/reassortants were compared with the parental sequences using SimPlot software (Lole et al., 1999). All potential reassortants were reanalyzed for the presence of mixed infection after sequencing at least 10 additional plasmids for the respective regions where reassortments were reported. When sequence diversity indicated infection by multiple isolates, an additional 10 clones were sequenced as confirmation.

3. Results

3.1. Population structure

Analysis of over 20% of the virus genome (2400 nt per isolate; regions indicated in Fig. 1) revealed significant sequence conservation. For the MP region, overall sequence variation among isolates peaked at 9% and 4% in the nt and aa sequences respectively (Supplementary Tables 3 and 4). Variation reached 13% and 5%, respectively, in the nt and aa sequences of the NP (Supplementary Tables 5 and 6). Two isolates from New Jersey (NJ10 and NJ11) were unique in the NP region with a Gly95 insertion and exhibited 87–90% nt identity to the rest of the isolates. Phylogenetic trees were generated from both the nt and aa sequences to assess the evolutionary relationships of the isolates. Trees derived from nt sequences were more informative compared to those derived from aa and are discussed here. Isolates mostly clustered together based on their geographical origin with some deviating from this structure (Fig. 2). Examples are Slo1, Slo8, OR13, OR14, OR15 and KY4 for the MP region, and Slo8, OR14 and OR15 for the NP region. A few NJ isolates were phylogenetically distant from most other NJ isolates and formed separate clades; NJ8-1 and NJ9 for the MP region and NJ10 and NJ11 for the NP region. All analyzed geographical populations represented isolates collected from different cultivars indicating that time or genetic makeup of the host were not related to the grouping of the BlaMV isolates. In addition, topologies of MP and NP trees were different which suggested phylogenetic incongruence of the isolates.

3.2. Selection constraints, recombination and reassortment

The rate of molecular evolution differs between genes and therefore $d_{N}/d_{S}$ was calculated, showing stringent purifying
selection on both regions. Negative or purifying selection means that mutations are selected out from the population resulting in conservation of population integrity. The \( \frac{d_{NS}}{d_S} \) ratio observed for the MP was higher (0.034) compared to the NP region (0.016; Table 1) indicating stronger constraints on the latter. Codon-based maximum likelihood methods did not detect positive selection acting on any codon in either region but there were numerous codons undergoing significant negative

### Table 1

<table>
<thead>
<tr>
<th>Region</th>
<th>ENC</th>
<th>Number of negatively selected codons</th>
<th>( d_{NS} )</th>
<th>Variance (( d_{NS} ))</th>
<th>Standard deviation (( d_{NS} ))</th>
<th>( d_S )</th>
<th>Variance (( d_S ))</th>
<th>Standard deviation (( d_S ))</th>
<th>( \frac{d_{NS}}{d_S} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>MP</td>
<td>50.249</td>
<td>8 51 190</td>
<td>0.005</td>
<td>0.000</td>
<td>0.002</td>
<td>0.148</td>
<td>0.000</td>
<td>0.014</td>
<td>0.034</td>
</tr>
<tr>
<td>NP</td>
<td>45.018</td>
<td>27 83 29</td>
<td>0.003</td>
<td>0.000</td>
<td>0.001</td>
<td>0.188</td>
<td>0.000</td>
<td>0.017</td>
<td>0.016</td>
</tr>
</tbody>
</table>

* ENC = effective number of codons, calculated using DnaSP v5 (Librado and Rozas, 2009).
* Number of negatively selected codons at 0.05 significance level, obtained using SLAC, FEL and REL programs implemented in Datamonkey (http://www.datamonkey.org; Kosakovsky Pond and Frost, 2005).
* \( d_{NS} \) = average number of non-synonymous substitutions per non-synonymous site.
* \( d_S \) = average number of synonymous substitutions per synonymous site.
* Non-synonymous Analysis Program (SNAP; Korber, 2000) was used for determining the \( d_{NS} \) and \( d_S \) and their variation and standard deviations.
selection (Table 1). ENC, a widely used index for characterizing codon usage bias for expression of genes, indicates a negative correlation with the rate of synonymous substitution. The NP showed stronger codon bias (ENC of 45) compared to the MP region (50) in all other cases presented, indicating a stronger purifying selection in NP against disfavored codons.

The selective constraints across the MP and NP coding regions were analyzed by studying the distribution of the synonymous, non-synonymous and indel mutations in those regions. For the 801 codons analyzed, the number of synonymous mutations outweighed that of non-synonymous in both regions, contributing to 85% and 92% of the total mutations for the MP and NP regions, respectively. The cumulative behavior of codon mutations showed an upward slope for synonymous mutations (Fig. 3), indicative of a uniform distribution. Phylogenetic trees showed incongruent topologies and therefore, we investigated the possibility of recombination events. None was detected in the individual regions studied. When they were concatenated and screened, there were four recombination events detected by seven of the algorithms (threshold = $10^{-6}$, Table 2). Four isolates from New Jersey, NJ8-11, one from Kentucky, KY2, and one from Slovenia, Slo4, were predicted to be products of reassortments (Table 2; Fig. 4). Additional sequencing indicated that the NJ8 and KY2 sequences were artifacts of mixed infections by two isolates, leaving four potential reassortants in the population studied.

4. Discussion

BbMaV is a recently identified ophiovirus associated with the blueberry mosaic disease. Analysis of the molecular variation in 61 isolates revealed low genetic diversity (9 and 13% for nt and 4 and 5% for aa in the MP and NP regions respectively). This suggested that the population structure is highly conserved, supporting similar observations reported for other ophioviruses (Alioto et al., 2003; Navarro et al., 2005; Martin et al., 2006).

Evolutionary relationships among the isolates revealed clustering based on geographical locations with exception of some isolates grouping with the isolates from other areas (examples – NJ7 and Slo8 clustering with isolates from other locations). In contrast, previous studies on MilBVV and CPSV (Alioto et al., 2003; Navarro et al., 2005; Martin et al., 2006) did not observe geographical delineation of isolates, possibly because of the propagation mode, seed and micropropagation for MilBVV and CPSV respectively, of the primary virus hosts, lettuce and citrus. Clustering based on geographic origin with low diversity could be explained by the founder effect in which a population is established from a small number of individuals or the recent establishment of the virus in the growing areas; not allowing sufficient time for diversification. On the other hand, the deviation of a few isolates may be due to multiple introductions of virus-infected material that originated from different areas.

Commercial blueberry cultivation was initiated in North America about a century ago, based on a few genetic lineages collected from the Pine Barrens of New Jersey (Moore, 1966). BbMaV-infected material may have been transported to other growing areas, allowing the virus to become established in the presence of a virus vector (MilBVV is transmitted by the soilborne fungus, Olpidium brassicae; Lot et al., 2002). As blueberry breeding increased and cultivation spread in North America and other areas, genetically distinct variants of BbMaV started emerging. There was no diagnostic test for the virus until recently (Thekke-Veetil et al., 2014) and there are long latent periods when plants remain asymptomatic. Those factors may have allowed for material infected with multiple isolates be transported to other growing areas, shaping the current population structure of the virus. An example to support
Table 2
Reassortment events detected in concatenated nucleotide sequences of the movement (MP) and nucleocapsid protein (NP) of blueberry mosaic associated virus isolates. Only events supported by more than five different RDP4 implemented methods are presented. Details on the origin of the isolates are found in Supplementary Table 1.

<table>
<thead>
<tr>
<th>Reassortment region</th>
<th>Reassortant</th>
<th>Minor parenta</th>
<th>Major parenta</th>
<th>RDPb</th>
<th>GENECONOb</th>
<th>Bootscanc</th>
<th>Machid</th>
<th>Chimaerad</th>
<th>SiSiscanb</th>
<th>3Seqc</th>
</tr>
</thead>
</table>

a Minor and major parent = parent contributing the smaller and larger fraction of sequence.
b RDP, GENECONO, Bootscan, Mach, Chimaera, SiScan and 3seq are algorithms implemented in RDP4 (Martin et al., 2010).
c Unknown = only one parent and a recombinant need to be in the alignment for a recombination event to be detectable.

this hypothesis are the Slovenian isolates; almost all were obtained from plants propagated from 1964 New Jersey plant imports.

Analysis of evolutionary forces on MP and NP regions suggested strong but differential selection pressure acting on the genes presumably to maintain functional integrity of the proteins. The higher purifying selection on NP was also observed for CPV (Martin et al., 2006), a closely related ophiovirus. This may be due to the importance of NP in genome protection and stability, movement, and interaction with host or vector proteins (Chare and Holmes, 2004; Martin et al., 2006).

In viruses with segmented genomes, exchange of genetic segments between variants co-infecting the same plant (reassembly) is a possibility and may play an important role in evolution. The role of reassembly in the speciation of multipartite viruses has been described in both negative and positive-strand RNA viruses (Garcia-Arenal et al., 2001; Margaria et al., 2007; Miranda et al., 2000; Qui et al., 1998; White et al., 1995). Due to the perennial nature and asexual means of propagation of blueberry, there is the possibility of mixed infections of BIMAV variants that may provide the opportunity for reassembly. Phylogenetic incongruence of the two regions and the potential reassembly events detected in the BIMAV population, suggested a possible role of genetic exchange in the evolution of BIMAV.

In conclusion, this communication revealed limited molecular variation in the BIMAV populations studied and provided evidence that the virus population structure is shaped by the combined effects of purifying selection, genetic exchange and possibly virus movement through infected propagative material. Information gathered in this study will have an impact on nursery certification programs and disease management in the field through development of universal detection assays, minimizing virus escapes (false negatives in detection assays). Additionally, the highly conserved regions identified in the MP and NP regions could be used for the development of resistance through transgene approaches.

Acknowledgements

The financial support for this study was provided by the Southern Region Small Fruit Consortium (Projects #2011–13 and 2012–07), USDA-NCNP (12-8100-1572-CA), Slovenian Research Agency (P4-0072) and EU (FP7-REGPOT-CT2012-316205).

Appendix A. Supplementary data

Supplementary article associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.virol.2015.02.022.


