Characterization of 10 tobacco vein banding mosaic virus isolates from China

TONGTONG YUAN, YUNZHI SONG, KAILONG LI, QUANJUN HAN, CHANGXIANG ZHU*, FUJIANG WEN*

State Key Laboratory of Crop Biology, Shandong Key Laboratory of Crop Biology, Shandong Agricultural University, Taian, 271018, P.R. China

Received September 8, 2011; accepted December 16, 2011

Summary. – The complete RNA genome sequences of 10 tobacco vein banding mosaic virus (TVBMV) isolates from China were determined using five overlapping cDNA clones. TVBMV was divided into three groups for HC-Pro and CP and into two groups for P3 and 6K1. With more isolate sequences analyzed, the phylogenetic results suggest that TVBMV could be divided into four subgroups based on HC-Pro and CP and into three subgroups with P3 and 6K1. Nucleotide sequence diversity analysis showed geographical differentiation among the TVBMV isolates. Three of the 10 isolates were found to have undergone recombination and new recombination sites were identified in the TVBMV genome. All coding genes were under negative selection and the population appeared to have remained stable over a long period. This study also provides preliminary data on the 3’-untranslated region as potentially the best genome sequence for developing transgenic tobacco.

Keywords: tobacco vein banding mosaic virus; complete sequence; phylogenetic analysis; recombination; selection pressure

Introduction

TVBMV, which was discovered in Taiwan in 1964 (Chin, 1966; Tian et al., 2007), is a distinct member of the genus Potyvirus (Chang et al., 1994; Habera et al., 1994). The genome of TVBMV is a single-stranded positive-sense RNA with 9570 nucleotides (nt). Its large open reading frame (ORF) is predicted to encode a polyprotein of 3079 amino acids (Mr, 348.6 kDa), which is cleaved into 10 mature proteins.

It mainly infects solanaceous plants and is transmitted by aphids in a nonpersistent manner, resulting in vein clearing, vein banding and necrotic lesions on leaves (Chin, 1975; Habera et al., 1994; Zhang and Li, 2001). Its filamentous particles measure approximately 780 nm×15 nm.

TVBMV was once a major threat to tobacco production activities in North America and Taiwan, but it has rarely occurred in most regions of China (Reddick et al., 1992; Tian et al., 2007; Yu et al., 2007). Due to the lack of TVBMV-resistant tobacco (Zhang et al., 2011), TVBMV may potentially spread quickly and cause serious damages to tobacco production in commercial planting areas. This conclusion is further supported by findings that TVBMV is becoming significantly more severe, thereby requiring greater attention (Tian et al., 2007).

As evolutionary genetic analysis plays an important role in the efficient control of viral diseases, the variability of RNA viruses has been thoroughly investigated in recent years (Moreno et al., 2004; Tsompana et al., 2005; Tomitaka and Ohshima, 2006). However, research on TVBMV is very limited, with the scarce data thus preventing analysis of evolution of the entire coding regions of TVBMV. Specifically, the HC-Pro, P3, 6K1, and CP genes have been only moderately studied (Tian et al., 2007; Zhang et al., 2011). Complete genomes of only two isolates have been reported (Wang et al., 2010; Yu et al., 2007). The TVBMV isolates could be divided into three and two groups based on the analysis of HC-Pro and CP and on that of P3 and 6K1, respectively. Analysis of the CP genes of subgroup I strains suggests that they can be further divided into two groups, but the between-group and within-group genetic distances do not completely support such subgrouping.
The objective of this study was to obtain more data about
the subgrouping using more sequences. Ten new complete
nucleotide and deduced amino acid sequences of TVBMV
were collected from major tobacco production regions in
China; these were compared with published sequences of
TVBMV to determine the evolutionary relationship as well
as genetic diversity of TVBMV isolates and to identify the
most appropriate genome sequence with which to generate
tobacco cultivars.

Materials and Methods

Virus isolates. Tobacco leaves with vein banding and yellow
mosaic symptoms were extracted from Yishui (YS), Yinan (YN),
Juxian (JX), Feixian (FX), Junan (JN), Zhucheng (ZC), Pingyi (PY),
Changle (CL), Laiwu (LW), and Yiyuan (YY) regions in Shandong,
China. Ten TVBMV isolates were screened using serological char-
acterization and by amplifying a TVBMV CP conserved region of
200 bp. The CP1 (5ʹ-GATGCACAATGAGATTGTTTTTG-3ʹ) and
CP2 (5ʹ-GAGTGACAACAGCCTCAGCGGTTGTTG-3ʹ) primers were used. The isolates were named YS, YN, JX, FX, JN, ZC,
PY, CL, LW, and YY, corresponding to their origin. After purifi ca-
tion by three inoculations through
Chenopodium amaranticolor
bacterization and by amplifying a TVBMV CP conserved region of
200 bp. The CP1 (5ʹ-GATGCACAATGAGATTGTTTTTG-3ʹ) and
CP2 (5ʹ-GAGTGACAACAGCCTCAGCGGTTGTTG-3ʹ) primers were used. The isolates were named YS, YN, JX, FX, JN, ZC,
PY, CL, LW, and YY, corresponding to their origin. After purifi ca-
tion by three inoculations through
Chenopodium amaranticolor
Primer name Position (bp) Nucleotide sequence (5ʹ to 3ʹ )

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Position (bp)</th>
<th>Nucleotide sequence (5ʹ to 3ʹ )</th>
</tr>
</thead>
<tbody>
<tr>
<td>TVBMV1-5</td>
<td>1–21</td>
<td>ATACGAAAAACGAACCAAGCA</td>
</tr>
<tr>
<td>TVBMV1-3</td>
<td>2479–2509</td>
<td>AGTGTAGACACACGTAGTCGA</td>
</tr>
<tr>
<td>TVBMV2-5</td>
<td>2349–2368</td>
<td>ACAGGTACACACATTTGAAA</td>
</tr>
<tr>
<td>TVBMV2-3</td>
<td>3689–3670</td>
<td>GTTTTCATTGCTGACTCAGGA</td>
</tr>
<tr>
<td>TVBMV3-5</td>
<td>3609–3629</td>
<td>ATAGTTGGGTGTTGCAAGATCAA</td>
</tr>
<tr>
<td>TVBMV3-3</td>
<td>5789–5770</td>
<td>TGCTATAGCTGACTCTGGA</td>
</tr>
<tr>
<td>TVBMV4-5</td>
<td>5671–5690</td>
<td>CTGGACTGATGCTAGGATGAAA</td>
</tr>
<tr>
<td>TVBMV4-3</td>
<td>7091–7072</td>
<td>GCCTGCAAGCTACTGCTTTAAA</td>
</tr>
<tr>
<td>TVBMV5-5</td>
<td>6984–7004</td>
<td>ACAAGCACAACCTGATAAGAGAT</td>
</tr>
<tr>
<td>TVBMV5-3</td>
<td>9570–9551</td>
<td>CCCTCACACCAATGTATT</td>
</tr>
</tbody>
</table>

(Takara, China). The reconstructed vectors were transformed into
Escherichia coli DH5α competent cells according to the standard
transformation method. At least two independent clones of each
fragment were sequenced in both directions and the accurate nu-
cleotide sequences were determined by comparisons.

Phylogenetic analysis. Sequences were assembled using
DNAMAN (Lynnbio Software, Canada) and DNASTAR (DNASTAR
Inc., USA). The complete genome sequences of YS, YN, JX, FX, JN,
ZC, PY, CL, LW, and YY have been deposited to GenBank under
Access Nos HQ396791, JN630468, JN630471, HQ396792, JN621319,
HQ396793, JN630469, JN630470, JN630473, and JN630472. Nu-
cleotide sequence alignments were performed using CLUSTAL
W (Thompson et al., 1994). Phylogenetic analysis of the complete
genomes of the 10 isolates with YND and HN39 were conducted by
neighbor joining, minimum evolution, and maximum parsimony in
MEGA 4.1 software. Genetic distances within and between groups
were estimated using the Pamilo-Bianchi-Li method (Pamilo and
Bianchi, 1993).

Recombination analysis. The complete sequence alignments
were subjected to recombination analysis with the RDP Software
package (Martin et al., 2010). Possible recombination events were
detected using RDP, GENECONV, BOOTSCAN, MAXCHI,
CHIMAERA, 3SEQ, and SISCAN packaged in RDP with the default
settings and a Bonferroni-corrected P value cutoff of 0.05. Recom-
bination events supported by at least four programs with P values
lower than 1.0×10−4 were considered “clear”; otherwise, they
were regarded as “tentative” ones.

Analysis of selection pressure and population demographics
of coding genes. Selection pressure was evaluated using nonsyn-
onymous (dN) and synonymous (dS) substitutions with Pamilo-
Bianchi-Li (PBL) method (Pamilo and Bianchi, 1993; Li, 1993).
Confidence estimates for dN and dS values were calculated using
the bootstrap method with 500 replicates. A dN/dS ratio lower than 1
indicated that negative selection had occurred, a dN/dS ratio higher
than 1 indicated positive selection and a dN/dS ratio of 1 indicated
neutral selection. In addition, all the coding genes were subjected
to sequence polymorphism using DnaSP 5.0 (Librado and Rozas,
2009) in calculating Tajima’s D, Fu and Li’s D, Fu and Li’s F , the
dependence and the nucleotide diversity to estimate their
distributive conditions. Mismatch distribution was also estimated
and the ragged, multimodal distribution in the samples suggested
the long-term stability of the population; otherwise, the population
would have evolved under expanding conditions.

Results

Sequence identity analyses

Sequence analysis showed that the complete RNA sequence of each isolate was 9570 nt long, excluding the poly(A) tail, comprising the 146-nt 5’-UTR, the 184-nt 3’-UTR, and a large ORF (nucleotides 147–9386) encoding
one large polyprotein. The 10 isolates had similar base compositions of the entire genome: adenine, 31.66%–31.84%; uracil, 26.33%–26.65%; cytosine, 18.84%–19.18%, and guanine, 22.70%–22.83%. Consistent with previously reported data for YND and HN39, the 5’-UTR of the 10 isolates contained one or two highly conserved potybox b (UCAAGCA) motifs but no potybox a (AUAACAU). For putative polyproteins, the cleavage sites of P1, HC-Pro, and NiA-Pro were Y/S, G/G, Q/A, Q/S, Q/G, E/A, Q/S, and Q/G, identical with those of YND and HN39, except that YND had a unique NiB/CP cleavage site (Q/N). Unlike many potyviruses, the 12 TVBMV isolates had an RITC motif in HC-Pro, instead of a KITC motif, involved in aphid transmission.

Comparisons of nucleotide and amino acid sequences revealed that the YS-YY isolates shared nucleotide and amino acid identities of 96.00%–99.47% and 98.21%–99.61%, whereas with YND and HN39 shared those of 89.82%–97.45% and 95.03%–99.19%. The sequence identities of every region in the whole genome available in the GenBank were further analyzed with those described herein. For each region, all the isolates shared much lower identities with Yunnan isolates. The 3’-UTR was more conservative with the highest nucleotide identity (94.05%–100%), whereas the 5’-UTR had the lowest nucleotide identity (83.56%–100%). As for amino acid identities, the P1 protein had the most variable and least conserved region (82.14%–100%) in the entire genome.

**Phylogenetic analyses**

The complete genome sequences of the 10 TVBMV isolates were subjected to phylogenetic analysis with YND and HN39 (Fig. 1). The 12 TVBMV isolates were clustered into two groups, with YND forming a separate branch. The phylogenetic trees constructed with every region of their whole genomes yielded the same result, indicating that TVBMV sequence correlated with geographical origin.

Phylogenetic trees of the HC-Pro, P3, 6K1, and CP genes were constructed using all the available sequences to strengthen the evidence further. The phylogenetic analysis of HC-Pro, P3, 6K1, and CP (Fig. 2a–d) demonstrated that TVBMV was divided into geographical origin-specific subgroups. The nucleotide sequences of HC-Pro, P3, and 6K1 were grouped into two main branches: group I comprised all the isolates except the Yunnan isolates and group II consisted of the latter set. The phylogenetic tree of the CP gene was divided into three groups: most isolates formed group I; the isolates from Japan, United States, and Taiwan constituted group II; and the Yunnan isolates comprised group III. Group I could be divided into three subgroups for HC-Pro, whereas it was divided into two subgroups for P3, 6K1, and CP. The within-group and between-group genetic distances were calculated to confirm the rationality of the subgrouping. The results showed that for the four genes, the genetic distances between groups were all one to four times higher than those within groups. All the statistical data showed that the findings from the phylogenetic analysis were reasonable and that geographical distribution is a potentially important selection factor.

**Recombination analyses**

The 10 isolates were included in RDP to identify possible recombination events. The results showed that several isolates were predicted to have recombination events, but only the sites found in YN, ZC, and PY were obvious (Table 2). The recombination sites were at the 3’-end of CI, the 5’-end of NiB, the 3’-end of HC-Pro, the 5’-end of VPg and the central region of 6K1. Recombination events between different groups were not found.

<table>
<thead>
<tr>
<th>Isolates</th>
<th>Supporting software*</th>
<th>Major parent</th>
<th>Minor parent</th>
<th>P-value</th>
<th>Z-value</th>
<th>Recombination site</th>
</tr>
</thead>
<tbody>
<tr>
<td>YN</td>
<td>RGBM33</td>
<td>YY</td>
<td>YS</td>
<td>7.448×10^-14</td>
<td>6.39</td>
<td>5018–7101 nt</td>
</tr>
<tr>
<td>PY</td>
<td>RGBM33</td>
<td>ZC</td>
<td>YN</td>
<td>5.645×10^-27</td>
<td>6.23</td>
<td>2353–5774 nt</td>
</tr>
<tr>
<td>ZC</td>
<td>RGBMCS3</td>
<td>JN</td>
<td>CL</td>
<td>1.555×10^-11</td>
<td>6.23</td>
<td>2431–3573 nt</td>
</tr>
</tbody>
</table>

*R: RDP; G: GENECONV; B: Bootscan; M: MaxChi; C: Chimaera; S: SiScan; 3: 3Seq. The software showed the highest significant P-value.
Phylogenetic analysis of TVBMV isolates constructed with the HC-Pro, P3, 6K1, and CP nucleotide sequences. Bootstrap analysis with 1000 replicates was performed. Branches with bootstrap values lower than 50% were collapsed and only bootstrap values >70% were shown.
Genetic diversity and population demography

Coding genes were subjected to genetic diversity analysis to estimate their variability. Although, HC-Pro and NIa showed the highest (0.0545) and lowest (0.0303) values respectively, no clear difference between the genes was detected. The $d_{s}/d_{a}$ ratio for the ORF was also calculated with the values for all the coding genes being lower than 1.0, suggesting that the ORF of TVBMV was under negative selection. On the other hand, the $d_{s}/d_{a}$ ratio for P1 was considerably high, showing that the P1 gene was under the highest selection pressure. The values obtained for Tajima’s D, Fu and Li’s D, as well as Fu and Li’s F of each coding gene were negative, but the $P$ values were higher than 0.05 or 0.10, rendering the results inconclusive. The frequency distributions of the number of pairwise nucleotide differences obtained from the P1-CP sequences were all ragged and multimodal, indicating the long-term stability of the population.

Discussion

This study determined the complete genomes of 10 TVBMV isolates from China. Nucleotide and amino acid comparisons showed that the Yunnan isolates shared much lower identities with other isolates. Phylogenetic analysis also suggests that the Yunnan isolates constituted a specific group. Zhang et al. (2011) reported phyletogenetic results showing that TVBMV is divided into three groups for HC-Pro (SD, MC, and YN) and CP, into two groups for P3 and 6K1, whereas we obtained novel results with the subgroups for HC-Pro, P3, and 6K1 by including more sequences in the current study. Based on HC-Pro, TVBMV was divided into two major groups. Group I, which was divided into three subgroups, corresponded to the formerly recognized SD and MC. For P3 and 6K1, group I was clustered into two subgroups; the previously identified group I could not be further divided. Overall, the phylogenetic trees demonstrated that the clustering of TVBMV correlated with geographical origin and that the isolates from mainland China, excluding the Yunnan ones, formed a large group.

Previous research has shown that the "hot" recombination spots of TVBMV occurred in the 3'-end of HC-Pro and the central region of 6K1. Recombination events in the 5'-end of HC-Pro, the 5'-end of P3 and in the CP gene were found also. In the current study we identified new recombination sites in the 3'-termini of HC-Pro and CI as well as the 5'-termini of VPg and NIb. Recombination was not detected between TVBMV isolates of different geographical groups.

For the TVBMV population, the $d_{s}/d_{a}$ ratios for each coding gene were all lower than 1.0, indicating that the ORFs were under negative selection to preserve their protein functions. As the degree of selection pressure on different genes was associated with the functions of their encoding proteins, different genes endured different selection pressures. The HC-Pro gene was under the strongest pressure, while on the other hand, the NIa gene was under the weakest pressure.

Mismatch distribution results implied that TVBMV had existed stably over an extended period of time. As detected in this study, there are two possible reasons behind its increasing incidence in recent years, firstly, TVBMV has always induced similar systems with PVY in fields, making PVY responsible for the great losses caused by TVBMV, and secondly, favorable environments enable TVBMV to spread quickly.

The lack of TVBMV-resistant tobacco (Zhang et al., 2011) evidently prompts its immediate cultivation. Advances in the genetic transformation of crop plants and deeper understanding of the genomic components of TVBMV have allowed for the development of transgenic tobacco. Viral coat protein, replicase and movement protein genes have been typically used to produce virus-resistant crops (Dasgupta et al., 2003; Batuman et al., 2006); however, the 3'-UTR of viruses, which may obtain the same or an even more durable and stable level of resistance (Vaslin et al., 2001; Batuman et al., 2006), has been widely unexplored because its secondary structure is greatly associated with virus replication (Li et al., 2006). As the identity of the 3'-UTR is higher and more stable than that of other regions, it may be used to produce TVBMV-resistant tobacco.

Acknowledgements. This work was financially supported by the National Natural Science Foundation of China [No. 30771408] and Tobacco Monopoly Administration (Corporation) Research Found of Shandong Province, P.R. China [No. KN102].

References


