Genetic diversity and population structure of rice stripe virus in China

Tai-Yun Wei,1 Jin-Guang Yang,1 Fu-Long Liao,1 Fang-Luan Gao,1 Lian-Ming Lu,1 Xiao-Ting Zhang,1 Fan Li,1,2 Zu-Jian Wu,1 Qi-Yin Lin,1 Lian-Hui Xie1 and Han-Xin Lin1

Correspondence
Han-Xin Lin
hanxinl@yahoo.com or linhan@mcmaster.ca
Lian-Hui Xie
xielh@fjau.edu.cn

1Institute of Plant Virology, Fujian Agricultural and Forestry University, Fuzhou, Fujian 350002, PR China
2Key Laboratory of Agricultural Biodiversity and Pest Management of Ministry of Education, Yunnan Agricultural University, Kunming, PR China

Received 26 August 2008
Accepted 29 December 2008

Rice stripe virus (RSV) is one of the most economically important pathogens of rice and is repeatedly epidemic in China, Japan and Korea. The most recent outbreak of RSV in eastern China in 2000 caused significant losses and raised serious concerns. In this paper, we provide a genotyping profile of RSV field isolates and describe the population structure of RSV in China, based on the nucleotide sequences of isolates collected from different geographical regions during 1997–2004. RSV isolates could be divided into two or three subtypes, depending on which gene was analysed. The genetic distances between subtypes range from 0.050 to 0.067. The population from eastern China is composed only of subtype I/IB isolates. In contrast, the population from Yunnan province (southwest China) is composed mainly of subtype II isolates, but also contains a small proportion of subtype I/IB isolates and subtype IA isolates. However, subpopulations collected from different districts in eastern China or Yunnan province are not genetically differentiated and show frequent gene flow. RSV genes were found to be under strong negative selection. Our data suggest that the most recent outbreak of RSV in eastern China was not due to the invasion of new RSV subtype(s). The evolutionary processes contributing to the observed genetic diversity and population structure are discussed.

INTRODUCTION

Due to their error-prone RNA replication, large population size and short generation time, RNA viruses have high mutation rates (Drake & Holland, 1999). Therefore, RNA viruses exhibit high potential for genetic variation, and a large number of nucleotide variations could exist in natural populations. Analyzing the polymorphic pattern of these variations will help us to understand the phylogenetic relationships, epidemiological routes, population structures and underlying evolutionary mechanisms of RNA viruses. In turn, this information will facilitate the development of effective control strategies for plant viral diseases.

Rice stripe virus (RSV) is one of the most important plant pathogens in China. The rice stripe disease induced by RSV was first recorded in Jiangsu–Zhejiang–Shanghai (JZS) district in 1963 and was later discovered in 16 provinces (Lin et al., 1990). Outbreaks of this disease were reported in JZS district in 1966, in Taiwan in 1969, in Yunnan province in 1974, in Beijing in 1975, in Shandong in 1986 and in Liaoning in the early 1990s (Lin et al., 1990; Xie et al., 2001). The disease is endemic in Yunnan, Jiangsu, Shanghai, Shandong, Beijing and Liaoning provinces (Fig. 1). Since 2000, the disease has circulated widely in Jiangsu province and become more severe. For instance, approximately 780 000 ha rice was infected in 2002 in Jiangsu province. This increased to 957 000 ha in 2003 and 1 571 000 ha in 2004, accounting for 80 % of the rice fields and a 30–40 % yield loss (http://www.hzag.gov.cn/bcjb/200559154000.htm). The outbreak of RSV also spread to adjacent provinces, such as Henan, Zhejiang, Anhui, Shandong, Shanghai and Hebei. However, the reasons behind this outbreak are not well understood. For example, it is unclear whether the invasion was triggered by a new RSV genotype. Outside China, RSV has been reported only...
in Japan, Korea and the Far East (ex-USSR) and has been epidemic in Japan and Korea since the 1960s, causing significant losses of rice yield (Hibino, 1996).

RSV is the type member of the genus *Tenuivirus*. It mainly infects rice plants, but also infects some other species in the family Poaceae, such as wheat and maize. RSV is transmitted by a small brown planthopper, *Laodelphax striatellus* Fallen (Hemiptera, Delphacidae), in circulative–propagative and transovarial manners (Falk & Tsai, 1998). The genome of RSV is composed of four negative-sense single-stranded RNA segments, designated RNA 1, 2, 3 and 4 according to decreasing size (Fig. 2) (Takahashi *et al.*, 1993; Toriyama *et al.*, 1994; Zhu *et al.*, 1991, 1992). RNA 1 is of negative polarity, encoding the putative RNA-dependent RNA polymerase (Toriyama *et al.*, 1994). The other three segments adopt an unusual ambisense coding strategy, i.e. both the viral-sense RNA (vRNA) and viral complementary-sense RNA (vcRNA) possess coding capacity, but the functions of the translated proteins are unclear. What is known is that the vRNA 2 encodes a membrane-associated protein and the vcRNA 2 encodes a polyglycoprotein (Falk & Tsai, 1998; Takahashi *et al.*, 1993). The nucleocapsid (N) protein gene has been mapped to vcRNA 3 (Kakutani *et al.*, 1991; Zhu *et al.*, 1991), and the NS3 protein encoded by vRNA 3 could act as an RNA-silencing suppressor based on the function of the analogous NS3 of rice hoja blanca virus, another member of the genus *Tenuivirus* (Bucher *et al.*, 2003). vRNA 4 encodes a protein known as major non-capsid protein (NCP) that accumulates in infected plants and may be involved in pathogenesis (Toriyama, 1986), whilst vcRNA 4 encodes a protein that has recently been shown to be involved in movement (Xiong *et al.*, 2008).

Numerous studies have been performed in recent decades on RSV aetiology, pathogenesis, ecology, molecular biology, control strategies etc. (Falk & Tsai, 1998). Nevertheless, there are still major gaps, especially in our understanding of the genetic diversity and population structure of RSV. Our previous studies have shown the biological diversity and genetic variations of several RSV isolates (Lin *et al.*, 1999, 2001, 2002; Wei *et al.*, 2003a, b). In this report, we analysed five genes (NS2, N, NS3, NCP and NSvc4) of 136 RSV isolates collected from different areas in China during 1997–2004. Our data showed that RSV isolates could be divided into two or three subtypes, depending on which gene was analysed. The distribution of these subtypes is correlated with their geographical locations, but not with the collection years. All of the isolates collected from eastern China and Japan belong to subtype I/IB, whereas isolates from Yunnan province are much more diverse, belonging to different subtypes with subtype II being predominant.
METHODS

Virus samples. RSV samples were collected in rice fields from eight provinces (Fujian, Jiangsu, Shanghai, Zhejiang, Henan, Shandong, Beijing and Liaoning) in eastern China and from Yunnan province in southwest China during 1997–2004 (Fig. 1). A virus sample from an individual rice plant was considered as one isolate. Infected rice plants were either used for extraction of total RNA or stored at −80°C for future use. Some isolates were maintained on suitable rice varieties (e.g. Hexi 28) via transmission by small brown planthoppers, L. striatellus Fallen. The geographical locations, collection years and rice varieties of RSV isolates used in this study are listed in Supplementary Table S1 (available in JGV Online).

RT-PCR, cloning and sequencing. Extraction of total RNA from infected rice leaves, cDNA synthesis and PCR amplification were performed as described previously (Lin et al., 2001). Forward and reverse primers were designed according to the nucleotide sequences of RNAs 2, 3 and 4 of RSV isolate T (GenBank accession numbers NC_003754, NC_003776 and NC_003753, respectively): 5′-ATGGTACTCTCTTTTATCTG-3′ (nt 81–102) and 5′-CCAAA-TTCACCTAGAATG-3′ (nt 666–686) for the NS2 gene; 5′-GGT-CAGTCTAGTATCAGC-3′ (nt 1437–1457) and 5′-ACACAAAGGTGTTGATTCT-3′ (nt 2490–2504) for the N gene; 5′-ACACAAAGGTGTTGATTCT-3′ (nucleotides nt 1–15) and 5′-CTACACGACAGCTG-GAGAGCTG-3′ (nt 681–701) for the NS3 gene; 5′-ACACAAAGGTGTTGATTCT-3′ (nt 1–15) and 5′-GGTGGAAAATGTGATATGCAAT-3′ (nt 597–616) for the NCP gene; and 5′-GGTGGAAAATGTGATATGCAAT-3′ (nt 1208–1228) and 5′-ACACAAAGGTGTTGATTCT-3′ (nt 2123–2137) for the NSVc4 gene.

RT-PCR products were purified by using a QIAquick PCR extraction kit (Qiagen). The purified PCR products were inserted into the pGEM-T vector (Promega) followed by transformation into Escherichia coli DH5α. Nucleotide sequences were determined by the Shanghai Jikang Biotech Company. For each gene of each isolate, one or two clones were sequenced. The GenBank accession numbers of these sequences are listed in Supplementary Table S1.

Phylogenetic analysis. Multiple nucleotide sequence alignments were performed by using CLUSTAL W (Thompson et al., 1994). Alignments were also adjusted manually to guarantee correct reading frames. Non-coding sequences were removed before alignment. Eight RSV isolates (T, M, O, C, Y, JS-YM, JSYD-05 and SD-JN2) that had been sequenced by other laboratories were included as references (Kakutani et al., 1990, 1991; Qu et al., 1997; Wang et al., 1992; Zhu et al., 1991, 1992). Maize stripe virus (MSV), the most closely related virus to RSV in the genus Tenuivirus, was used as outgroup for phylogenetic analyses. The GenBank accession numbers for RNAs 2, 3 and 4 of MSV are U53224, S40180 and AJ969410, respectively. All of the sequence alignments used in this study are available upon request. Phylogenetic trees were reconstructed by using the neighbour-joining (NJ) method with the Kimura two-parameter model implemented in PAUP* 4.0b10 (Swofford, 2002). Gaps were treated as a fifth character state. Evaluation of statistical confidence in nodes was based on 1000 bootstrap replicates. Branches with <50% bootstrap value were collapsed.

Estimation of genetic distance and selection pressure. Genetic distances (the average number of nucleotide substitutions between two randomly selected sequences in a population) within and between subtypes were calculated by MEGA 2.1 based on the Kimura two-parameter model (Kumar et al., 2001). The dS/dN ratio is used to estimate selection pressure. dS (the average number of non-synonymous substitutions per non-synonymous site) and dN (the average number of synonymous substitutions per synonymous site) were estimated by using the Pamilo–Bianchi-Li method implemented in MEGA 2.1. SEM was computed by MEGA 2.1 using a bootstrap with 100 replicates.

RESULTS

Genotype profile of RSV field isolates

To provide a genotype profile of RSV field isolates, we performed phylogenetic analyses of RSV isolates collected from different provinces in eastern and southwestern China during 1997–2004 (Fig. 1; Supplementary Table S1). Neighbour-joining (NJ) trees were constructed by using datasets for five RSV genes: NS2 (600 bp, 55 sequences), NS3 (636 bp, 44 sequences), N (969 bp, 87 sequences), NCP (543 bp, 65 sequences) and NSvC4 (861 bp, 33 sequences). As illustrated in Figs 3 and 4, RSV isolates fell into two monophyletic clades. As all of the isolates collected from eastern China form one of the monophyletic clades, we refer to it as subtype I and to the other as subtype II. The mean genetic distance between these two subtypes ranges from 0.054 to 0.067, and those within subtypes range from 0.008 to 0.038 (Table 1). Unlike the polytomic topology of subtype I in the NS2 and NS3 gene trees, subtype I in the NCP and NSvC4 gene trees is dichotomic (Fig. 3). Noticeably, one of the sister clades only contains isolates that were collected from Yunnan province. We thereby refer to this clade as subtype IA, and to the other as subtype IB (Fig. 3c, d). Indeed, the genetic distances between subtypes IA, IB and II fall into a range (0.050–0.060) similar to those between subtypes I and II (Table 2). Therefore, for the NCP and NSvC4 genes, we divided RSV isolates into three subtypes.

Spatial and temporal distribution of RSV isolates in nature

To better understand the spatial and temporal distribution patterns, the collection years and geographical locations of RSV isolates corresponding to taxa in the phylogenetic trees were depicted (information for the N gene tree is shown in Fig. 4). Considering the geographical separation/proximity, the RSV population in China was first divided into two subpopulations, YN and E, containing isolates collected from Yunnan province and eastern China, respectively. RSV isolates collected from Yunnan province were further grouped into five districts: Baoshan–Shaba–Banqiao–Xinjie–Hetu–Jiaguan (BS), Dali–Xizhou–Fengyi–
Fig. 3. Phylogenetic analysis of four RSV genes. Trees were constructed by the NJ algorithm with the Kimura two-parameter model implemented in PAUP* 4.0b10.0. Branches were collapsed when the bootstrap value was <50%. Bootstrap values are given above the branches of each clade. The trees were rooted by using maize stripe virus (MStV) as the outgroup. Eight RSV isolates (T, M, O, C, Y, JS-YM, JSYD-05 and SD-JN2) that were sequenced by other laboratories were included as references and are underlined. Taxa on grey backgrounds indicate the subtype I, IB and IA isolates that were collected in Yunnan province. Bars, number of substitutions per site. (a) NS2 gene; (b) NS3 gene; (c) NCP gene; (d) NSvC4 gene.
Fig. 4. Phylogenetic analysis of RSV isolates based on the N gene and their spatial and temporal distributions. The left panel shows the N gene tree, constructed by the NJ algorithm with the Kimura two-parameter model implemented in PAUP* 4.0b10.0. Branches were collapsed when the bootstrap value was < 50%. Bootstrap values are given above the branches of each clade. The tree was rooted by using MStV as the outgroup. RSV isolates (T, M, C, Y, JS-YM and SD-JN2) that were sequenced by other laboratories were included as references and are underlined. Taxa on grey backgrounds indicate the subtype I, IB and IA isolates that were collected in Yunnan province. Bar, 0.005 substitutions per site. In the right panel, the collection year and site for each isolate are shown. Collecting sites in China fall into two geographically large categories, Yunnan (YN) province and eastern China. There are five districts [Baoshan–Shaba–Banqiao–Xinjie–Hetu–Jiaguan (BS), Dali–Xizhou–Fengyi–Weishan (DL), Chuxiong–Yongren–Dayao–Yaoan (CX), Kunming–Luquan–Wuding–Fumin–Yiliang–Shiling (KM) and Yuxi–Jiangchun (YX)] in YN province and six districts, Fujian, Jiangsu–Zhejiang–Shanghai, Henan, Shandong and Liaoning province (FJ, JZS, HN, SD, BJ and LN, respectively), in eastern China.
Weishan (DL), Chuxiong–Yongren–Dayao–Yaoan (CX), Kunming–Luquan–Wuding–Fumin–Yiliang–Luliang (KM) and Yuxi–Jiangchun (YX). Those from eastern China were further grouped into six districts: Fujian (FJ), Jiangsu–Zhejiang–Shanghai (JZS), Henan (HN), Shandong (SD), Beijing (BJ) and Liaoning (LN) (Figs 1, 4).

Two interesting patterns are seen in Fig. 4. Firstly, all of the isolates collected from eastern China and Japan belong to subtype I. Secondly, all of the subtype II isolates were collected from Yunnan province. However, two isolates, Y and CX2 (shown on grey backgrounds in Fig. 4), collected from Yunnan province also belong to subtype I. Therefore, isolates collected from Yunnan province are more diverse, belonging to different subtypes. A similar genetic structure was observed in the trees for the other four genes (Fig. 3).

We also examined the spatial distribution pattern of isolates collected from different districts of eastern China and Yunnan province, but failed to see any obvious pattern.

Despite examining the phylogenetic grouping of RSV isolates and their collection years closely, we are unable to find a general pattern as clear as that for spatial distribution. However, we discovered that most of the subtype IA isolates in the NCP and NSvc4 gene trees were collected in 2004 (Fig. 3c, d).

Genetic differentiation between subpopulations

To test the genetic differentiation between and within subpopulations E and YN, three statistical tests, Ks*, Z and Snn, were used (Hudson, 2000; Hudson et al., 1992). As shown in Table 3, all of the tests strongly support the hypothesis that these two subpopulations are genetically differentiated (P=0). Although, for some genes, significant genetic differentiation could also be observed within subpopulations YN and E, it was not supported by all three statistical tests. To estimate the extent of genetic differentiation, we measured the coefficient $F_{st}$, which is also an estimate of gene flow. Except for the NS2 gene, the values of $F_{st}$ between subpopulations are all $>0.33$ (Table 3), an indication of infrequent gene flow. The absolute values of $F_{st}$ for within subpopulations are all $<0.33$, suggesting frequent gene flow. This was also seen when $F_{st}$ between any two districts within YN or E was measured (data not shown).

Strong selective pressures acting on RSV genes

We also estimated the negative selection pressure acting on RSV genes. Overall, the values of the $d_n:d_s$ ratio for five genes were markedly low (0.046–0.123; Table 4), implying that all of these genes are under strong negative selection. The value of the $d_n:d_s$ ratio for the NS2 gene is slightly higher than that for the N and NS3 genes, but is 2.6 times higher than that of the NCP and NSvc4 genes. Therefore, the NCP and NSvc4 genes are subject to stricter selective constraints. Interestingly, the values of the $d_n:d_s$ ratio for genes on the same RNA segment, i.e. the NS3 and N genes on RNA 3, and the NCP and NSvc4 genes on RNA 4, are almost identical (Table 4). This suggests that the single
**Table 3.** Genetic differentiation between and within subpopulations YN and E

Subpopulation division is based on the geographical sites of origin of RSV samples. The RSV population in China was first divided into two subpopulations, Yunnan province (YN) and eastern China (E). Subpopulation YN was further divided into five districts: Baoshan (BS), Dali (DL), Chuxiong (CX), Kunming (KM) and Yuxi (YX) (Fig. 1). Subpopulation E was further divided into six districts: Liao ning (LN), Beijing (BJ), Henan (HN), Shandong (SD), Jiangsu–Zhejiang–Shanghai (JZS) and Fujian (F) (Fig. 1).

<table>
<thead>
<tr>
<th>Gene</th>
<th>Test</th>
<th>Subpopulation</th>
<th>Between YN and E</th>
<th>Within YN</th>
<th>Within E</th>
</tr>
</thead>
<tbody>
<tr>
<td>NS2</td>
<td>P-value of Ks*</td>
<td>0.000*</td>
<td>0.510</td>
<td>0.126</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P-value of Z</td>
<td>0.000*</td>
<td>0.767</td>
<td>0.403</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P-value of Snn</td>
<td>0.000*†</td>
<td>0.353</td>
<td>0.032†</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fst‡</td>
<td>0.310</td>
<td>−0.048</td>
<td>−0.010</td>
<td></td>
</tr>
<tr>
<td>NS3</td>
<td>P-value of Ks*</td>
<td>0.000†</td>
<td>0.098</td>
<td>0.232</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P-value of Z</td>
<td>0.000†</td>
<td>0.189</td>
<td>0.066</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P-value of Snn</td>
<td>0.000†</td>
<td>0.030†</td>
<td>0.033†</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fst‡</td>
<td>0.476</td>
<td>0.133</td>
<td>0.281</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>P-value of Ks*</td>
<td>0.000†</td>
<td>0.385</td>
<td>0.436</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P-value of Z</td>
<td>0.000†</td>
<td>0.197</td>
<td>0.344</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P-value of Snn</td>
<td>0.000†</td>
<td>0.206</td>
<td>0.281</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fst‡</td>
<td>0.604</td>
<td>0.069</td>
<td>0.042</td>
<td></td>
</tr>
<tr>
<td>NCP</td>
<td>P-value of Ks*</td>
<td>0.000†</td>
<td>0.648</td>
<td>0.514</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P-value of Z</td>
<td>0.000†</td>
<td>0.863</td>
<td>0.686</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P-value of Snn</td>
<td>0.000†</td>
<td>0.087</td>
<td>0.831</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fst‡</td>
<td>0.415</td>
<td>−0.043</td>
<td>−0.059</td>
<td></td>
</tr>
<tr>
<td>NSvc4</td>
<td>P-value of Ks*</td>
<td>0.000†</td>
<td>0.007†</td>
<td>0.318</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P-value of Z</td>
<td>0.000†</td>
<td>0.008†</td>
<td>0.334</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P-value of Snn</td>
<td>0.000†</td>
<td>0.090</td>
<td>0.858</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fst‡</td>
<td>0.435</td>
<td>0.282</td>
<td>−0.095</td>
<td></td>
</tr>
</tbody>
</table>

*P<0.05, which was considered as significantly rejecting the null hypothesis that there is no genetic differentiation between two subpopulations.

†Fst is a coefficient of the extent of genetic differentiation and provides an estimate of the extent of gene flow. An absolute value of Fst>0.33 suggests infrequent gene flow.

RNA segments of the divided genome may be the evolutionary unit of selection. With regard to selection pressure acting on different subtypes, there is no general pattern of which subtype is under stronger or weaker selection pressure.

**DISCUSSION**

A number of methods have been used to differentiate virus isolates in order to provide a genotyping basis for examining the genetic composition of viral populations. These include restriction fragment-length polymorphism (RFLP; Arboleda & Azzam, 2000), PCR–single-strand conformation polymorphism (PCR-SSCP; Lin et al., 2004), RNase-protection assays (Fraile et al., 1997) and molecular phylogeny (Abubakar et al., 2003). In this paper, we applied a distance-based NJ method to construct the phylogenetic trees of RSV isolates collected from different areas in China during 1997–2004. Overall, these isolates fell into two monophyletic clades, namely subtypes I and II (Figs 3, 4). However, the subtype I clade in the NSvc4 gene tree is only poorly supported (51 % bootstrap value, Fig. 3d). Furthermore, some subtype clades collapsed when the character-based maximum-parsimony method was used (data not shown). The low resolution of the phylogenetic trees of RSV isolates could be due to the low genetic diversity (0.05–0.067 between subtypes and up to 0.092 between isolates) and insufficient informative sites regarding evolution of the genome. Nevertheless, our analyses provided a genotyping profile of RSV field isolates, and this allowed us to investigate the genetic structure of RSV populations in China.

Our data show that the population collected from eastern China is composed only of subtype I or IB isolates. In contrast, the population from Yunnan province is composed of different subtypes, with subtype II predominating (Fig. 4). Subpopulations collected from different districts in eastern China or Yunnan province are not genetically differentiated and show frequent gene flow (Table 3). Particularly, subtype I/IB isolates prevailing in the outbreak sites (Jiangsu–Zhejiang–Shanghai and Henan districts) show high sequence identities to isolates collected from other provinces in eastern China (Fig. 4). These data indicate that the most recent outbreak of RSV in these provinces was not due to the invasion of a new subtype(s) from Yunnan province.

Although the prevailing genotype in Yunnan province is subtype II, we noticed a recent expansion of a ‘new’ lineage, subtype IA. Subtype IA isolates were collected from different districts in Yunnan province and mainly in the year 2004 (Fig. 3c, d; Supplementary Table S1). In fact, some of them already existed in nature before 2004. For example, isolate CHUX was collected in 2001, and the collection year of isolate Y was 1995 (Qu et al., 1997). It would be interesting to see whether this subtype expands further in the field and to compare the biological properties of isolates of subtype IA with those of subtypes IB and II.

The population structure described here for RSV in China is distinct from that of cucumber mosaic virus (CMV), but is somewhat similar to those of two insect vector-borne rice viruses, rice tungro bacilliform virus (RTBV) and rice yellow mottle virus (RYMV). The genetic composition of 11 CMV populations collected in Spain was not correlated with their geographical locations and collection years, and was described as metapopulation with local colonization, extinction and recolonization (Fraile et al., 1997). In contrast, RTBV populations also showed a spatial distribution pattern with a greater genetic diversity in the
Table 4. Estimation of nucleotide diversity of five RSV genes

Nucleotide diversity, defined here as the average number of nucleotide substitutions per site, was computed separately for nonsynonymous sites \( d_s \) and synonymous sites \( d_k \). The \( d_s : d_k \) ratio gives an estimate of selection pressure; if \( d_s : d_k < 1.0 \), negative selection is implied.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Subtype IA*</th>
<th>Subtype I/IB</th>
<th>Subtype II</th>
<th>All†</th>
</tr>
</thead>
<tbody>
<tr>
<td>NS2</td>
<td>( d_n )</td>
<td>0.006 ± 0.001</td>
<td>0.009 ± 0.002</td>
<td>0.013 ± 0.003</td>
</tr>
<tr>
<td></td>
<td>( d_k )</td>
<td>0.058 ± 0.002</td>
<td>0.073 ± 0.014</td>
<td>0.105 ± 0.015</td>
</tr>
<tr>
<td></td>
<td>( d_n : d_k )</td>
<td>0.103</td>
<td>0.123</td>
<td>0.123</td>
</tr>
<tr>
<td>NS3</td>
<td>( d_n )</td>
<td>0.010 ± 0.002</td>
<td>0.006 ± 0.001</td>
<td>0.014 ± 0.003</td>
</tr>
<tr>
<td></td>
<td>( d_k )</td>
<td>0.052 ± 0.006</td>
<td>0.083 ± 0.013</td>
<td>0.139 ± 0.019</td>
</tr>
<tr>
<td></td>
<td>( d_n : d_k )</td>
<td>0.192</td>
<td>0.072</td>
<td>0.100</td>
</tr>
<tr>
<td>N</td>
<td>( d_n )</td>
<td>0.004 ± 0.001</td>
<td>0.008 ± 0.002</td>
<td>0.011 ± 0.003</td>
</tr>
<tr>
<td></td>
<td>( d_k )</td>
<td>0.047 ± 0.006</td>
<td>0.064 ± 0.008</td>
<td>0.111 ± 0.012</td>
</tr>
<tr>
<td></td>
<td>( d_n : d_k )</td>
<td>0.085</td>
<td>0.125</td>
<td>0.099</td>
</tr>
<tr>
<td>NCP</td>
<td>( d_n )</td>
<td>0.003 ± 0.002</td>
<td>0.008 ± 0.002</td>
<td>0.007 ± 0.002</td>
</tr>
<tr>
<td></td>
<td>( d_k )</td>
<td>0.090 ± 0.017</td>
<td>0.065 ± 0.012</td>
<td>0.151 ± 0.025</td>
</tr>
<tr>
<td></td>
<td>( d_n : d_k )</td>
<td>0.033</td>
<td>0.123</td>
<td>0.046</td>
</tr>
<tr>
<td>NSvc4</td>
<td>( d_n )</td>
<td>0.004 ± 0.002</td>
<td>0.007 ± 0.002</td>
<td>0.007 ± 0.001</td>
</tr>
<tr>
<td></td>
<td>( d_k )</td>
<td>0.032 ± 0.006</td>
<td>0.067 ± 0.012</td>
<td>0.149 ± 0.015</td>
</tr>
<tr>
<td></td>
<td>( d_n : d_k )</td>
<td>0.125</td>
<td>0.104</td>
<td>0.047</td>
</tr>
</tbody>
</table>

*NA, Not applied. The subdivisions of subtype IA, IB and II were applied only to the NCP and NSvc4 genes.
†All isolates were included in the analysis.

Indonesian population than in Philippine or Vietnamese populations (Azzam et al., 2000). Similarly, the genetic diversity of the RYMV population is highest in east Africa, especially in eastern Tanzania, and decreases progressively from the east to the west of Africa (Abubakar et al., 2003).

The geographical origin of a virus can be inferred from the extent of its genetic diversity. If a viral population shows higher genetic diversity, it is normally considered more ancient (Azzam et al., 2000; Fargette et al., 2006; Gessain et al., 1992; Giri et al., 1997; Koralink et al., 1994; Moya & Garcia-Arenal, 1995). The historical record of disease may or may not be consistent with the extent of genetic diversity. For example, the higher genetic diversity of the RTBV Indonesian population is consistent with historic records of rice tungro disease in 1840 in Indonesia, but this disease was not reported in Vietnam only recently (Azzam et al., 2000). Likewise, eastern Tanzania is believed to be the centre of origin for RYMV, as the most divergent isolates were found in this area (Abubakar et al., 2003; Fargette et al., 2006). The virus was first reported in Kenya; however, Kenya is located directly north-east of Tanzania (Fargette et al., 2006). According to the population structure described in our report, Yunnan province could be the geographical origin of RSV in China. Although the report of rice stripe disease in this province occurred almost a decade after its discovery in Jiangsu–Zhejiang–Shanghai district (Lin et al., 1990; Xie et al., 2001), the disease could have been unnoticed in Yunnan province for many years. This is likely considering the fact that plant virology research in Yunnan province greatly lagged behind that in eastern China. The scenario of a Yunnan origin is reinforced by the fact that this province is well known as the ‘kingdom of plants and animals’ that may harbour primary indigenous hosts and efficient insect vectors of RSV. Analogously, Tanzania is also reputed to be a biodiversity hotspot for plants and animals, and this rich biodiversity has been proposed to account for the origin of RYMV as well as another insect-borne plant virus, cassava mosaic virus, in Africa (Fargette et al., 2006). Despite this evidence, we cannot exclude the possibility that different subtypes might have been present in eastern China a long time ago, giving rise to subtype I/IB only in recent years. Interestingly, although the occurrence of rice stripe disease dates back to the early 1990s in Japan (Hibino, 1996), the only three Japanese isolates (T, M and O) with available nucleotide sequences belong to subtype I/IB (Figs 3, 4). Apparently, more sequences, especially for Japanese and Korean isolates, are needed to fully disclose the secrets of the origin of RSV.

Why is the RSV Yunnan population composed of different subtypes, whereas the eastern China population is only composed of a single subtype? This can be explained simply by the founder effect, which is often invoked to explain the low genetic diversity of certain populations of various plant viruses, including the RYMV population in western Africa.
(Fargette et al., 2006; Garcia-Arenal et al., 2001). However, we believe that the interplay between RSV and its insect vector, L. striatellus, could explain the observed population structure of RSV in China more specifically. Unlike CMV or RYMV, which can be transmitted by contact and several different vectors (Fargette et al., 2006; Palukaitis et al., 1992), RSV is strictly transmitted by L. striatellus in persistent propagative and transovarial manners (Toriyama, 1986). Therefore, the RSV–vector interaction probably played a critical role in shaping the population structure. It is generally believed that L. striatellus lacks the ability to migrate long distances unless helped by a monsoon and therefore has evolved to adapt to the local climate (Hoshizaki, 1997). Therefore, it is possible that L. striatellus populations in Yunnan and eastern China are genetically differentiated and have different affinities for RSV subtypes. The vector for the Yunnan population may be favourable for the transmission of the predominant subtype II isolates, whereas the vector for the eastern China population is favourable for the subtype I/IB isolates. Such biased transmissibility has been reported for other plant virus–vector systems. For example, whitefly Bemisia tabaci populations from various geographical locations had different transmission efficiencies with different isolates of tomato leaf curl geminivirus or other geminiviruses (Bedford et al., 1994; McGrath & Harrison, 1995). This scenario is supported by our unpublished data indicating that four L. striatellus populations collected from Yunnan province grouped together with similar random-amplified polymorphic DNA (RAPD) patterns, and were distinct from another group containing nine insect populations collected from eastern China. The specific interaction between RSV subtypes and insect vectors is being investigated further in our laboratory.

The strict manner of insect-vector transmission, together with the narrow host range of RSV, may explain the observed low genetic diversity. Compared with the highly variable CMV isolates (genetic distance up to 0.4 between subgroup I and II isolates), RSV isolates show very low genetic diversity (0.05–0.067 between subtypes and up to 0.092 between isolates). Similarly, compared with a list of genetic diversity (0.05–0.067 between subtypes and up to 0.092 between isolates), RSV isolates show very low genetic diversity. Compared with the highly variable CMV isolates (genetic distance up to 0.4 between subgroup I and II isolates), RSV isolates show very low genetic diversity. Compared with the highly variable CMV isolates (genetic distance up to 0.4 between subgroup I and II isolates), RSV isolates show very low genetic diversity. Compared with the highly variable CMV isolates (genetic distance up to 0.4 between subgroup I and II isolates), RSV isolates show very low genetic diversity.

ACKNOWLEDGEMENTS

We are grateful to Bashan Huang, Jun Zheng, Kexian Zong, Yushan Wang, Weimin Li, Dechun Fang, Lingeng Gong, Xiaomin Li, Wenwu Shi, Yixiu Huang, Dunfan Fan, Yuxin Xu, Yunke He, Yijun Zhou, Zhaoan Chen, Maosheng Li, Xuelian Lai, Honglian Li, Daling Shen, Chongguan Pan and Guanghe Zhou for their help in collecting RSV samples. We also thank Dr Bryce Falk and Dr Dustin Johnson for their help with the writing. This work was supported by projects from the National Natural Science and Technological Foundation of China (items 3990091 and 3000002 to L.-H. X. and H.-X. L., respectively), Natural Scientific and Technological Foundation of Fujian province (item B0110014 to H.-X. L.), the Ministry of Education of China (item 020401 to H.-X. L.) and the Natural Scientific and Technological Foundation of Yunnan province (item 2003C0042M to F. L.)

REFERENCES


