Prevalence and genetic diversity of an unusual virus associated with Kobu-sho disease of gentian in Japan

Go Atumi,1 Reiko Tomita,1 Kappei Kobayashi2 and Ken-Taro Sekine1

1Iwate Biotechnology Research Center, Kitakami 024-0003, Iwate, Japan
2Faculty of Agriculture, Ehime University, Matsuyama, Ehime 790-8566, Japan

Gentian Kobu-sho-associated virus (GKaV) is a recently discovered novel virus from Kobu-sho (a hyperplastic or tumorous disorder)-affected Japanese gentians. To obtain insight into GKaV transmission and pathogenesis, the genetic diversity of the virus in the putative helicase and RNA-dependent RNA polymerase coding regions was studied. The extent of GKaV sequence diversity within single host plants differed within samples and between viral genomic regions. Phylogenetic analysis of 30 Kobu-sho-affected samples from different production areas and host cultivars revealed that GKaV populations have diverged as they became prevalent in different geographical regions. The diversification of GKaV was shown to be driven by geographical isolation rather than host adaptation; however, no geographical patterns were found. Therefore, it was not feasible to trace the pathway of GKaV spread.

From the mid-1980s, the disease known as Kobu-sho has damaged cultivated gentians (Gentiana triflora, Gentiana scabra or their cross) and led to substantial economic losses in Japan. Kobu-sho is a syndrome that causes hyperplasia and stunting in gentians (Takahashi et al., 2009). Recently, a novel virus considered to be involved in Kobu-sho (Kobayashi et al., 2013) was identified and named Gentian Kobu-sho-associated virus (GKaV). GKaV has a single large ORF that encodes a potential polyprotein. The deduced polyprotein sequence does not have significant similarity to any proteins encoded by other plant viruses, but is similar to the RNA-dependent RNA polymerase (RdRp) and RNA helicase (Hel) coding regions of the family Flaviviridae. Our earlier study indicated that the full-length sequences of two GKaV samples from Kobu-sho-affected gentians from different fields in Iwate Prefecture shared a nucleotide sequence identity of less than 90%, suggesting GKaV has a highly polymorphic nature (Kobayashi et al., 2013). Furthermore, some polymorphic sites were observed in GKaV sequences detected in single plants (Kobayashi et al., 2013). In this study, the genetic diversity of GKaV from different fields in four prefectures was studied to obtain insights into GKaV transmission and pathogenesis.

A previous study suggested that GKaV exists as heterogeneous populations consisting of closely related RNA sequences within single hosts (Kobayashi et al., 2013). To confirm this notion, the genetic diversity of GKaV in individual plants exhibiting symptoms of Kobu-sho was investigated. RNA samples were extracted from 0.1 g Kobu-sho-developing stem tissue as described previously (Kobayashi et al., 2013). A single plant was randomly selected for this analysis from each of the sampling locations including Iwate [Hachimantai (As27), Morioka (Ty1), Hanamaki (Id1) and Waga (NW)], Akita (At166), Yamagata (Yg) and Fukushima (Fs3) Prefectures (Table S1, available in JGV Online). Reverse transcription (RT) and PCR amplification were carried out using PrimeScript reverse transcriptase (TaKaRa) and, to minimize the mutations produced during PCR amplification, high-fidelity PrimeSTAR GXL DNA polymerase (TaKaRa) was used according to the manufacturers’ instructions (Smith et al., 1997). Primers were designed for the predicted conserved regions: Hel-like region (159 and 160) and RdRp-like region, 161 and 162 (Table S2) to amplify 1084 and 1060 bp regions (excluding primer sequences), respectively. The thermal cycling conditions were 98 °C for 20 s, 30 cycles of 98 °C for 10 s, 50 °C for 20 s and 68 °C for 1.5 min. The PCR products were cloned into pCR4 Blunt-TOPO vector (Life Technologies). Twenty clones were sequenced for each sample using M13 forward and reverse primers (Table S2).

To evaluate the heterogeneity of the GKaV population in a single host, the Shannon entropy values (Pawlotsky et al., 1998) were calculated using the following formula: $-\sum_j [pi \times \ln pi] / \ln N$, in which $pi$ is the frequency of each sequence in the population and $N$ is the total number of sequences.

The GenBank/EMBL/DDBJ accession numbers for the sequences of the putative helicase and RNA-dependent RNA polymerase encoding regions of GKaV are AB796346–AB796371 and AB796372–AB796398, respectively.

Two supplementary tables and one supplementary figure are available with the online version of this paper.
Shannon entropy values for GKaV populations from seven individual plants ranged from 0.196 (Ty1) to 0.844 (Id1) in the Hel-like region and from 0.066 (Yg) to 0.922 (As27) in the RdRp-like region (Table 1). These results indicated that the Hel- and RdRp-like regions were both heterogeneous in As27, NW and Fs3, whereas both genes were less heterogeneous in Ty1. Interestingly, the heterogeneities of GKaV sequences significantly differed between the Hel- and RdRp-like regions in both Id1 and Yg samples, suggesting that the complexity of the GKaV genomic sequence greatly differed, not only between samples, but also between the genomic regions. To estimate the genetic diversity of GKaV populations in single hosts, averages of pairwise genetic distances \((d)\) within the samples using Kimura’s two-parameter method were calculated (Table 1). In the Hel-like region, the \(d\) values for Ty1, Id1, NW and Yg were low, whereas those for As27, At166 and Fs3 were higher (Table 1). In the RdRp-like region, the \(d\) values in Ty1, Id1, At166 and Yg were low but those in As27, NW and Fs3 were higher (Table 1). The values of Shannon entropy and mean genetic distance collectively indicated that GKaV populations in some plants showed high diversity and heterogeneity (As27), whereas those in other locations did not (Ty1).

For comparison, the RdRp region of Broad bean wilt virus-2 (BBWV-2), which asymptomatically infects Japanese gentian populations, was amplified from the same seven samples co-infected with GKaV. The nucleotide sequences of BBWV-2 RdRp region were amplified using primers 211 and 45 (Atsumi et al., 2013) and 893 nt sequences devoid of primer sequences from 20 clones were compared. The Shannon entropy values of BBWV-2 RdRp in Id1 and Fs3 were relatively high (0.577 and 0.483, respectively), but those in the other isolates were low (between 0.000 and 0.196). The \(d\) values of BBWV-2 RdRp were significantly low except for Id1 (Table 1). These results indicated that BBWV-2 populations, except for that from Id1, were homogeneous. The low sequence heterogeneity observed in BBWV also indicated that the mutation frequencies associated with experimental technique were significantly below the levels that might have produced the sequence heterogeneity observed in GKaV populations. Although a more intensive survey of the genetic diversity of these viruses will be required to determine the extent of the diversity, the GKaV population in single hosts is as diverse as that of BBWV-2 or even more. Taken together, these results suggested that GKaVs were highly heterogeneous and genetically diverse within single hosts, as observed for many RNA viruses (Lauring & Andino, 2010).

Next, GKaV samples from different locations were phylogenetically analysed to explore the pathway of GKaV prevalence in Tohoku region, the northern part of the main island of Japan. The potential Hel region (1844 bp) and RdRp (2303 bp) encoding regions were amplified from 30 Kobu-sho-affected samples from different production areas including 20 host cultivars (Fig. 1a, b; Table S1). The 30 samples included those analyzed for genetic diversity within a single host (Table 1). Primers were designed for putative conserved regions inferred from previously determined sequences: 11 and 12 for the Hel-like region, and GKICv3-F1 and GK26-OIS4 for the RdRp-like region (Table S2). RT-PCR was performed as mentioned above. From 2008 to 2010, samples were collected from Iwate (Waga, Hachimantai, Morioka, Hanamaki and Miyako), Yamagata, Akita and Fukushima Prefectures in Japan (Fig. 1a, b, Table S1). The 30 RT-PCR products were directly sequenced using primers 11, 12 and 23 for the Hel-like region, and 29, 74, GKICv3-F1, GK51-OS1, GK36F, GK26-OIS2 and GK26-OIS4 for the RdRp-like region by Sanger sequencing (Tables S1 and S2). The consensus sequences were determined by manually examining the electropherograms using 4Peaks software (http://nucleoexes.com/index.php/4peaks).

### Table 1. Genetic heterogeneity and diversity of GKaV

Values of Shannon entropy and mean genetic distance \((d)\) were calculated for the region encompassing the Hel-like or RdRp-like regions of GKaV. For comparison, the values were also calculated for a partial region of the BBWV-2 RdRp gene.

<table>
<thead>
<tr>
<th>Location</th>
<th>Prefecture</th>
<th>City or county</th>
<th>Abbreviation</th>
<th>GKaV</th>
<th>BBWV-2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Hel-like</td>
<td>RdRp-like</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Shannon entropy</td>
<td>(d)</td>
</tr>
<tr>
<td>Iwate</td>
<td>Hachimantai</td>
<td>As27</td>
<td>0.722</td>
<td>0.055</td>
<td>0.922</td>
</tr>
<tr>
<td></td>
<td>Morioka</td>
<td>Ty1</td>
<td>0.196</td>
<td>0.000</td>
<td>0.205</td>
</tr>
<tr>
<td></td>
<td>Hanamaki</td>
<td>Id1</td>
<td>0.844</td>
<td>0.001</td>
<td>0.132</td>
</tr>
<tr>
<td></td>
<td>Waga</td>
<td>NW</td>
<td>0.695</td>
<td>0.001</td>
<td>0.477</td>
</tr>
<tr>
<td>Akita</td>
<td>Daisen</td>
<td>At166</td>
<td>0.505</td>
<td>0.054</td>
<td>0.205</td>
</tr>
<tr>
<td>Yamagata</td>
<td>Kaminoyama</td>
<td>Yg</td>
<td>0.554</td>
<td>0.004</td>
<td>0.066</td>
</tr>
<tr>
<td>Fukushima</td>
<td>Ishikawa</td>
<td>Fs3</td>
<td>0.467</td>
<td>0.042</td>
<td>0.443</td>
</tr>
</tbody>
</table>
Fig. 1. Phylogenetic trees of GKaV Hel- and RdRp-like regions. (a) Black circles indicate the locations for sample collection in Iwate Prefecture, a black triangle in Akita Prefecture, a black square in Yamagata Prefecture and a black diamond in Fukushima Prefecture. (b) Locations for sample collection within Iwate Prefecture. Blue circles, Hachimantai city; red circles, Morioka city; grey circle, Miyako city; purple circles, Hanamaki city; green circles, Waga county. (c, d) Unrooted maximum-likelihood trees were inferred from a nucleotide sequence alignment of regions encompassing the Hel-like (1667 bp) (c) or RdRp-like (1733 bp) (d) sequences. Each symbol indicates a unique field. For example, in Waga county (b), four green circles mean that the samples were collected from four fields. Details for each abbreviation are shown in Table S1. The numbers beside the nodes are bootstrap values (1000 replicates).
Sequence analysis indicated that four samples of the Hel-like region (As27, Id13, At166 and Fs3) and three samples of the RdRp-like region (As27, NW and Fs3) were highly polymorphic. For example, results from direct-sequencing of the RdRp-like region from As27 sample showed overlapping multicolour peaks throughout the region in the electrogram. Alignment of 20 cloned RdRp-like regions (1064 nt) indicated that 105 positions were heterogeneous (these clones were obtained in the analysis of genetic diversity within a single host, see above). For calculation of sequence identities, highly polymorphic complex sequences were excluded. The nucleotide sequence identities ranged from 86.7 to 100% in the Hel-like region and from 90.4 to 100% in the RdRp-like region, whereas the deduced amino acid sequence identities ranged from 97.8 to 100% in the Hel-like region and from 96.7 to 100% in the RdRp-like region. The high amino acid sequence identities suggested that GKaV isolated from gentians affected by Kubo-sho in the various locations of Tohoku region constituted a single species. The Hel- and RdRp-like sequences were conserved at the amino acid level, but the diversity was high at the nucleotide level. The ratio of non-synonymous (dN) and synonymous (dS) substitution rates for the Hel- and RdRp-like regions was estimated using SNAP (http://www.hiv.lanl.gov/content/sequence/SNAP/SNAP.html; Korber, 2000). The dNdS ratios for the Hel- and RdRp-like regions were less than 1.0 (0.0064 and 0.021, respectively). This result suggested that Hel-like and RdRp-like regions in GKaV sequences were under purifying (negative) selection.

Phylogenetic trees were constructed using the obtained sequences except for highly polymorphic sequences. GKaV Hel- or RdRp-like sequences were aligned using MUSCLE (Edgar, 2004), and the maximum-likelihood trees were inferred using MEGA5 (Fig. 1; Tamura et al., 2011). The nucleotide substitution models and rates among sites were Tamura 3-parameter and gamma distributed for the Hel- and RdRp-like regions, respectively. The majority-rule bootstrap consensus trees were constructed with a 70% cut-off value. The significance of the nodes was estimated with 1000 bootstrap replicates. The results showed that, in all samples, the cloned sequences from a single plant did not constitute a single cluster, suggesting that genetically distinct populations infected a single plant (Fig. S1). Taken together, these results suggest a lack of clustering by geographical origin, although significantly higher levels of sampling would be required to affirm this hypothesis conclusively.

The GKaV sequences were compared in a pairwise manner to examine the possibility of host adaptation (Fig. 2). The results revealed that the nucleotide identities were 99.88% (2 nt differences in 1667 nt) in the Hel-like region (‘1’ in Fig. 2a) and 99.82–99.94% (1–3 nt differences in 1733 nt) in the RdRp-like region (‘1’ in Fig. 2b) among cultivars A, B and C that were collected in Kaisawa, Waga and Iwate. In Ryosawa (Waga, Iwate), nucleotide identities were 99.88–99.94% (1–2 nt differences) in the Hel-like region (‘2’ in Fig. 2a) and 99.88% (2 nt differences) in the RdRp-like region (‘2’ in Fig. 2b) among Albireo, D and E. In Tamayama (Morioka, Iwate), nucleotide identities were 99.82% (3 nt differences) in the Hel-like region (‘3’ in Fig. 2a) and 99.48% (9 nt differences) in the RdRp-like region (‘3’ in Fig. 2b) in field A between Giovanni and Kyusuto; 99.82–99.94% (1–3 nt differences) in the Hel-like region (‘4’ in Fig. 2a) and 99.53–99.94% (1–8 nt differences) in the RdRp-like region (‘4’ in Fig. 2b) in field B among Iwate, Giovanni and Madjieru; 99.94% (1 nt differences) in the Hel-like region (‘5’ in Fig. 2a) and 99.77% (4 nt differences) in the RdRp-like region (‘5’ in Fig. 2b) in field C between Ihatovo and Albireo. These results indicated that almost identical GKaV populations infected across all cultivars when the plants were grown in the same field. Next, differences in the GKaV sequences from the same cultivars but grown in different locations were compared. GKaV sequences amplified from the samples of Iwate cultivars grown in Morioka or Hanamaki indicated that nucleotide identities in the Hel- and RdRp-like regions were 88.67 and 92.44%, respectively (‘6’ in Fig. 2, b, respectively). In cultivar Giovanni, nucleotide identities were 92.32% (Hel-like) and 91.69% (RdRp-like) between Morioka and Miyako (‘7’ in Fig. 2a, b), respectively. Furthermore, nucleotide identities were 86.80% (Hel-like) and 92.96% (RdRp-like) (‘8’ in Fig. 2a, b), respectively, between Kaisawa and Shinmachi, both of which were close...
to each other in Waga county (Fig. 1b). These results suggested that the diversity of GKaV sequences was not formed through adaptation to the host cultivar.

In this study, GKaV diversity was first analysed in single host plants, suggesting that the GKaV genome is prone to diversify through nucleotide sequence changes and, therefore, could reside in host plants as a quasi-species, as has been shown in many RNA viruses (Lozano et al., 2009; Naraghi-Arani et al., 2001; Lauring & Andino, 2010). However, it remains unknown whether a particular set of GKaV variants have been selected after multiple cycles of infection. In-depth analysis of the population dynamics of GKaV would be needed to prove the quasi-species nature of GKaV. Next, GKaV sequences were compared in samples from different locations, revealing that GKaV genomes diversified through geographical isolation rather than host adaptation. However, our data do not allow us to trace the spread of GKaV in the region, which could be determined by factors such as the mode of transmission of the disease (yet unknown) or human activities (agricultural practices, movement of plants for commercial purposes, etc.).

Acknowledgements

We thank Kazue Obara and Yukari Takahashi for technical assistance; we also thank Yasuya Iwadate and Kazumichi Fujiwara for gentian sample collection. This study was supported by the Iwate Prefecture Government.

References


Fig. 2. Matrices of pairwise nucleotide sequence identities of the Hel-like or RdRp-like regions. Pairwise comparison of nucleotide sequence identities among GKaV isolates was conducted by MEGA5 and created the colour-coded graphical matrices using Microsoft Excel 2011 based on the Hel-like (a) and RdRp-like sequences (b). The shapes and colours of the symbols adjacent to sample names refer to Fig. 1(a, b). The numbers in the coloured squares indicate the sampling locations that were discussed in the text.