A mutation ‘hot spot’ in the Schmallenberg virus M segment

Melina Fischer, Bernd Hoffmann, Katja V. Goller, Dirk Höper, Kerstin Wernike and Martin Beer

Institute of Diagnostic Virology, Friedrich-Loeffler-Institut, D-17493 Greifswald-Insel Riems, Germany

In the autumn of 2011, Schmallenberg virus (SBV), a novel orthobunyavirus of the Simbu serogroup, was identified by metagenomic analysis in Germany. SBV has since been detected in ruminants all over Europe, and investigations on phylogenetic relationships, clinical signs and epidemiology have been conducted. However, until now, only comparative sequence analysis of SBV genome segments with other species of the Simbu serogroup have been performed, and detailed data on the S and M segments, relevant for virus–host-cell interaction, have been missing. In this study, we investigated the S- and M-segment sequences obtained from 24 SBV-positive field samples from sheep, cattle and a goat collected from all over Germany. The results obtained indicated that the overall genome variability of SBV is neither regionally nor host species dependent. Nevertheless, we characterized for the first time a region of high sequence variability (a mutation ‘hot spot’) within the glycoprotein Gc encoded by the M segment.

Like all orthobunyavirus genomes, the SBV genome consists of three segments of a negative-sense ssRNA (Chowdhary et al., 2012; Gentsch & Bishop, 1979; Gentsch et al., 1980). According to size, these segments are named small (S), medium (M) and large (L). Commonly in the genus Orthobunyavirus, the S segment encodes two proteins in overlapping ORFs, a small non-structural protein (NSs) and the nucleocapsid (N) protein. The NSs protein of orthobunyaviruses acts as an interferon antagonist in mammalian cells (Elliott et al., 2013; van Knippenberg et al., 2010). Furthermore, NSs seems to play a role in the replication process in mosquitoes (Szemiel et al., 2012). A major function of the N protein is encapsidation of the viral genome. The N protein is also used for serological testing (King et al., 2011). The L segment encodes the viral RNA-dependent RNA polymerase (Chowdhary et al., 2012; Saeed et al., 2001). The glycoproteins encoded on the M segment play the most
important role in viral attachment and cell fusion. The M segment encodes a polyprotein precursor that is cleaved post-translationally into the glycoproteins Gn and Gc, as well as a non-structural protein (NSm) (in the order Gn–NSm–Gc) (Schmaljohn & Hooper, 2001). Both glycoproteins form spikes on the virus particle and are type I integral transmembrane proteins that are modified by N-linked glycosylation (Nichol et al., 2005; Schmaljohn & Hooper, 2001; Shi et al., 2005). In this study, we determined the sequences of the S- and M-segment ORFs directly from several SBV field samples originating from all over Germany and conducted comparative sequence and phylogenetic analyses to investigate their relatedness and sequence variability.

In order to investigate geographical and host-related sequence variations, we analysed the complete N protein ORF, the S-encoded NSs ORF and the M-encoded glycoprotein ORFs obtained from the original sample BH80/11-4 and 23 additional field samples from 2011 and 2012 originating from sheep (n=18), cattle (n=4) and a goat (n=1) from various federal states in Germany (Table 1). Viral RNA was extracted from the field samples using a QIAamp Viral RNA Mini kit (Qiagen) according to the manufacturer's recommendations, and SBV infection was confirmed by real-time reverse transcription PCR (RT-qPCR) (Bilk et al., 2012). Subsequently, overlapping sequence fragments covering the ORF of the S and M segments were generated using a OneStep RT-PCR kit (Qiagen). Sequencing was carried out in both directions by termination cycle sequencing using a Big Dye Terminator Mix (Applied Biosystems) with primers (sequences and protocols available on request) used for the PCR-product amplification. Cycle sequencing products were purified with a DyeEx 2.0 Spin kit (Qiagen) and sequenced with an ABI 3130 Genetic Analyzer (Applied Biosystems). For each sample, a consensus sequence was generated from the overlapping sequence fragments via Codon Code Aligner version 3.7.1 (CodonCode Corporation). Alignment of nucleotide consensus sequences and translation in amino acid comparison: x<0.005, Fig. 1a; amino acid comparison: x<0.002, Fig. 1b). A total of 23 unique nucleotide substitutions, of which 5 nt were non-synonymous, were detected in the N protein ORF alignment. However, an accumulation of substitutions was observed in the N-terminal region of the N protein (Fig. 1b), where the NSs protein is located in an overlapping ORF. Thus, the NSs ORF (nt 26–298) was analysed independently, and 14 unique nucleotide substitutions were found, resulting in 11 unique amino acid substitutions, as well as two unique nucleotide substitutions, which led to a truncated NSs protein in three samples. These substitutions were distributed equally across the NSs ORF (nucleotide comparison: x<0.01; amino acid comparison: x<0.03). A possible explanation of this finding may be the non-essential role of NSs in the replication process in vertebrates.

In contrast, in total, 306 unique nucleotide substitutions of which 184 were non-synonymous were identified in the M-segment sequence alignment. Using Pearson's r² statistics, we located one prominent mutation hot spot in the N-terminal part of the putative Gc protein of SBV ranging from nt 1483 to 1864 (x<0.0007; Fig. 1c) with a total of 121 nucleotide substitutions, of which 64 were non-synonymous. The corresponding mutation hot spot in relation to the amino acid positions ranged from aa 493 to 629 (x<0.002; Fig. 1d). Within this region, one SBV sequence showed a deletion of 12 aa (aa 532–543 in BH77/12-1) and another insertion of 2 aa (aa 573/574 in BH174/12-2) (Fig. 1e). In contrast to the mutation hot spot, highly conserved regions were found in the M segment (Fig. 1d) that seemed to correlate with functional domains. Using the TMHMM prediction program (http://www.cbs.dtu.dk/services/TMHMM/; Möller et al., 2001), a putative transmembrane domain (aa 234–253) in the Gn protein was assessed, which corresponded to a highly conserved region around aa 250 in the protein sequence (Fig. 1d). The location of the putative fusion peptide (estimated according to the position in Bunyamwera virus as aa 1058–1079; Shi et al., 2009) matched with a highly conserved region in Gc (aa ~1050–1080).

Our results clearly showed that the S segment is much more conserved among the analysed German SBV samples.
than the M segment. This is consistent with results for Akabane virus, another member of the Simbu serogroup within the orthobunyaviruses, for which it was observed that the most variable segment is the M segment (Kobayashi et al., 2007). Furthermore, in the M segment, sequence variability accumulates within the N-terminal region of the putative Gc protein. Such highly variable regions, often referred to as hypervariable regions, have been described in other RNA virus proteins such as the haemagglutinin protein of influenza virus H5N1 (Ho¨per et al., 2012), E2/NS1 of hepatitis C virus (HCV) (e.g. Driesel et al., 1994, Zhang et al., 2000) or envelope protein GP5 of porcine reproductive and respiratory syndrome virus (PRRSV, e.g. Delisle et al., 2012; Kim et al., 2013). These proteins have in common the fact that they play a role in host-cell attachment and fusion and often act as major immunogens involved in the induction of neutralizing antibodies, as is also the case with the M segment in orthobunyaviruses (Kobayashi et al., 2007). In avian influenza virus H5N1, as well as in HCV and PRRSV, these mutation hot spots or hypervariable regions seemingly play a role in the evasion of host immunity (Curran et al., 2002; Höper et al., 2012; Kim et al., 2013) and may also influence the outcome of persistent or acute infections, at least in the case of HCV (Curran et al., 2002; Zhang et al., 2000). Whilst SBV is supposed to have a very short viraemia of ~5–6 days (Hoffmann et al., 2012), SBV genome detection in certain organs (e.g. spleen and lymph nodes) is possible for at least 24 days (Wernike et al., 2012), so it is likely that the mutation hot spot may play a key role in evasion of the immune response. As shown previously for Bunyamwera virus, deletions in the N-terminal stretch within the Gc protein do not, for example, influence virus replication in cell culture, even though longer deletions in that part of the genome may lead to decreased virus assembly (Shi et al., 2009). Nevertheless, the importance and biological relevance of the highly variable region for SBV replication, immunogenicity and host-cell infection remain unclear and need to be investigated further.

In addition, all M sequences were phylogenetically analysed to identify any regional or host species-specific clusters. A maximum-likelihood tree with the HKY nucleotide

Table 1. Field samples used in this study

<table>
<thead>
<tr>
<th>Name</th>
<th>Sample*</th>
<th>Host species</th>
<th>Collection date</th>
<th>Region†</th>
<th>GenBank accession nos of S/M ORF</th>
</tr>
</thead>
<tbody>
<tr>
<td>BH 02/12-1</td>
<td>Organ sample</td>
<td>Sheep (lamb)</td>
<td>11/2011</td>
<td>NRW</td>
<td>KC108842/KC108843</td>
</tr>
<tr>
<td>BH 03/12-3</td>
<td>Organ sample</td>
<td>Sheep (lamb)</td>
<td>01/2012</td>
<td>NRW</td>
<td>KC108844/KC108845</td>
</tr>
<tr>
<td>BH 28/12-5</td>
<td>Brain</td>
<td>Sheep (lamb)</td>
<td>01/2012</td>
<td>NRW</td>
<td>KC108846/KC108847</td>
</tr>
<tr>
<td>BH 37/2-2</td>
<td>Brain</td>
<td>Sheep</td>
<td>01/2012</td>
<td>NRW</td>
<td>KC108848/KC108849</td>
</tr>
<tr>
<td>BH 59/12-8</td>
<td>Organ sample</td>
<td>Sheep</td>
<td>01/2012</td>
<td>HE</td>
<td>KC108850/KC108851</td>
</tr>
<tr>
<td>BH 77/12-1</td>
<td>Brain</td>
<td>Sheep (lamb)</td>
<td>01/2012</td>
<td>LS</td>
<td>KC108852/KC108853</td>
</tr>
<tr>
<td>BH 127/12-16a</td>
<td>Brain</td>
<td>Sheep (lamb)</td>
<td>01/2012</td>
<td>SH</td>
<td>KC108854/KC108855</td>
</tr>
<tr>
<td>BH 148/12-9</td>
<td>Parotid gland</td>
<td>Sheep (lamb)</td>
<td>01/2012</td>
<td>TH</td>
<td>KC108856/KC108857</td>
</tr>
<tr>
<td>BH 174/12-2</td>
<td>Cerebrum</td>
<td>Sheep (lamb)</td>
<td>01/2012</td>
<td>NRW</td>
<td>KC108858/KC108859</td>
</tr>
<tr>
<td>BH 197/12-3</td>
<td>Organ sample</td>
<td>Sheep</td>
<td>01/2012</td>
<td>HH</td>
<td>KC108860/KC108861</td>
</tr>
<tr>
<td>BH 198/12-5b</td>
<td>Cerebrum</td>
<td>Sheep</td>
<td>01/2012</td>
<td>SA</td>
<td>KC108862/KC108863</td>
</tr>
<tr>
<td>BH 199/12-5</td>
<td>Organ sample</td>
<td>Sheep (lamb)</td>
<td>01/2012</td>
<td>SH</td>
<td>KC108864/KC108865</td>
</tr>
<tr>
<td>BH 200/12-2</td>
<td>Brain</td>
<td>Sheep (foetus)</td>
<td>01/2012</td>
<td>BV</td>
<td>KC108866/KC108867</td>
</tr>
<tr>
<td>BH 231/12-1c</td>
<td>Brain</td>
<td>Sheep (foetus)</td>
<td>01/2012</td>
<td>RP</td>
<td>KC108868/KC108869</td>
</tr>
<tr>
<td>BH 233/12-1</td>
<td>Brain</td>
<td>Goat (goatling)</td>
<td>01/2012</td>
<td>BW</td>
<td>KC108870/KC108871</td>
</tr>
<tr>
<td>BH 237/12-4</td>
<td>Organ sample</td>
<td>Sheep (lamb)</td>
<td>02/2012</td>
<td>SH</td>
<td>KC108872/KC108873</td>
</tr>
<tr>
<td>BH 248/12-1</td>
<td>Organ sample</td>
<td>Cow</td>
<td>02/2012</td>
<td>MWP</td>
<td>KC108874/KC108875</td>
</tr>
<tr>
<td>BH 250/12-2a</td>
<td>Brain</td>
<td>Cow</td>
<td>02/2012</td>
<td>SL</td>
<td>KC108876/KC108877</td>
</tr>
<tr>
<td>BH 336/12-1</td>
<td>Organ sample</td>
<td>Sheep</td>
<td>03/2012</td>
<td>SN</td>
<td>KC108878/KC108879</td>
</tr>
<tr>
<td>BH 336/12-3</td>
<td>Organ sample</td>
<td>Sheep</td>
<td>03/2012</td>
<td>SN</td>
<td>KC108880/KC108881</td>
</tr>
<tr>
<td>BH 619/12-1</td>
<td>Serum</td>
<td>Sheep</td>
<td>09/2012</td>
<td>TH</td>
<td>KC108882/KC108883</td>
</tr>
<tr>
<td>BH 635/12-2</td>
<td>Serum</td>
<td>Cow</td>
<td>08/2012</td>
<td>RP</td>
<td>KC108884/KC108885</td>
</tr>
<tr>
<td>BH 652/12-1</td>
<td>EDTA blood</td>
<td>Cow</td>
<td>10/2012</td>
<td>SH</td>
<td>KC108886/KC108887</td>
</tr>
</tbody>
</table>

*Organ samples comprised a pool of cerebrum, cerebellum and spleen.
†Region indicates the federal state of Germany in which the respective samples were collected: BW, Baden-Württemberg; BV, Bavaria; HH, Hamburg; HE, Hesse; MWP, Mecklenburg-West Pomerania; NRW, North Rhine-Westphalia; LS, Lower Saxony; RP, Rhineland-Palatinate; SH, Schleswig-Holstein; SL, Saarland; SN, Saxony; SA, Saxony-Anhalt; TH, Thuringia.
substitution model and 1000 bootstrap replicates was calculated using MEGA5 (Tamura et al., 2006). In the resulting tree, all sequences clustered closely together (Fig. 2; for a tree including Sathuperi virus and Douglas virus as an outgroup, see Fig. S1, available in JGV Online), and, as indicated by the low bootstrap values, neither a regional nor a host species-specific cluster could be identified. In conclusion, the sequence variation observed in the M segment is probably independent
of any host species and of the spatial distribution, but may have emerged during virus replication within the individual host by selective pressures. Furthermore, arthropod-borne viruses, having an alternate two-host 'life cycle', are typically suggested to be more stable than vector-independent viruses. Genomic evolution is suggested to be slower if a virus has to adapt to two hosts (Moutailler et al., 2011; Novella et al., 1999; Strauss & Strauss, 1994). As SBV is transmitted by arthropod vectors, the virus undergoes two processes of replication cycles: one in the arthropod vector and the other in the vertebrate host, and selective pressures may act during both steps, but it remains unknown whether sequence divergence is related to the mammalian or arthropod portion of the virus life cycle. Thus, the importance of this mutation hot spot in virus replication and host-vector adaptation needs to be investigated further. Studies on sequence variability within the vector will be needed, as well as investigations concerning cross-neutralization of virus variants and the virulence of different SBV isolates. In addition, to understand better the role of the newly described mutation hot spot in the SBV Gc protein gene, further molecular analyses on its relevance in virus replication and assembly are needed, for example using reverse genetics. Furthermore, it has to be determined whether these viruses – especially those with insertions or deletions – are indeed infectious.

In summary, our study addressed, for the first time, SBV genome diversity using 20 different SBV-positive field samples from Germany. The data presented here clearly indicated that SBV genome variability is especially high within the N-terminal region of the putative Gc protein. The process of mutation within this mutation hot spot seems to be independent of host species and geographical region.

**Acknowledgements**

We thank Christian Korthase, Karin Lissek and Patrick Zitzow for excellent technical assistance. This work was supported by the
References


European Union FP7 projects ‘European Management Platform for Emerging and Re-emerging Infectious Disease Entities’ (‘EMPERIE’; no. 223498) and ‘Rapid Field Diagnostics and Screening in Veterinary Medicine’ (‘RAPIDIA-Field’, no. 289364).


