Journal of General Virology (2006), **87**, 573–580 DOI 10.1099/vir.0.81201-0

Genetic diversity of small-ruminant lentiviruses: characterization of Norwegian isolates of *Caprine arthritis encephalitis virus*

Britt Gjerset, Anne K. Storset and Espen Rimstad

Department of Food Safety and Infection Biology, Norwegian School of Veterinary Science, Box 8146 Dep, N-0033 Oslo, Norway

Correspondence
Britt Gjerset
britt.gjerset@veths.no

Received 23 May 2005
Accepted 15 November 2005

Small-ruminant lentiviruses (SRLVs), including *Caprine arthritis encephalitis virus* (CAEV) in goats and maedi-visna virus (MVV) in sheep, are lentiviruses that, despite overall similarities, show considerable genetic variation in regions of the SRLV genome. To gain further knowledge about the genetic diversity and phylogenetic relationships among field isolates of SRLVs occurring in geographically distinct areas, the full-length genomic sequence of a CAEV isolate (CAEV-1GA) and partial *env* sequences obtained from Norwegian CAEV-infected goats were determined. The genome of CAEV-1GA consisted of 8919 bp. Alignment studies indicated significant diversity from published SRLV sequences. Deletions and hypervariability in the 5′ part of the *env* gene have implications for the size of the proposed CAEV-1GA Rev protein and the encoded surface glycoprotein (SU). The variable regions in the C-terminal part of SU obtained from Norwegian CAEV isolates demonstrate higher sequence divergence than has been described previously for SRLVs. Phylogenetic analysis based on *SU* sequences gives further support for a unique group designation. The results described here reveal a distant genetic relationship between Norwegian CAEV and other SRLVs and demonstrate that there is more geographical heterogeneity among SRLVs than reported previously.

INTRODUCTION

Small-ruminant lentiviruses (SRLVs) are monocyte/macrophage-tropic lentiviruses of goats and sheep. *Caprine arthritis encephalitis virus* (CAEV) and the ovine maedi-visna virus (MVV) cause persistent infections that can induce systemic diseases affecting joints, mammary glands and respiratory and central nervous systems after long incubation periods (Narayan *et al.*, 1993). Ingestion of infected colostrum and milk or direct contact with infected animals are the major transmission modes of SRLV within flocks (reviewed by Blacklaws *et al.*, 2004). SRLV infections are widespread in most regions of the world and are associated with economic losses in small-ruminant production (reviewed by Peterhans *et al.*, 2004).

Like other lentiviruses, SRLVs are characterized by high genetic diversity, resulting from factors such as high mutational rates and rapid virus production (Wain-Hobson, 1996). The SRLV genome contains the structural genes *gag*, *pol* and *env*, in addition to the regulatory genes *rev*, *vif* and *tat*. SRLVs are closely related genetically; however, one of the hallmarks of lentiviruses is the genetic variation, found mainly in the viral *env* and *rev* genes, as well as in the long terminal repeat (LTR) region (Pyper *et al.*, 1986; Knowles *et al.*, 1991). Sequence information from SRLVs occurring in geographically distinct areas is limited, and thus the extent of diversity and the impact that these variations may have on viral properties and various methods used for diagnosis of infection are only partly known.

Prototypic full-length SRLV sequences include that of CAEV-Co, a strain originally isolated in the USA (Saltarelli *et al.*, 1990), the ovine South African SA-OMVV strain (Querat *et al.*, 1990), the British EV-1 strain (Sargan *et al.*, 1991) and several strains derived from Icelandic visna K1514 (Sonigo *et al.*, 1985; Andresson *et al.*, 1993). Recently, two more full-length or nearly full-length SRLV sequences have been published (Barros *et al.*, 2004; Shah *et al.*, 2004). In addition, numerous partial sequences of SRLVs exist. Phylogenetic analysis based on nucleotide sequences of *gag*, *pol* and *env* fragments of SRLV has led to CAEV-like viruses being found in sheep and MVV-like viruses in goats, indicating that SRLVs may be able to cross the species barrier between goats and sheep under natural conditions (Karr *et al.*, 1996; Leroux *et al.*, 1997; Zanoni, 1998; Grego *et al.*, 2002; Rolland *et al.*, 2002; Shah *et al.*, 2004).

CAEV is highly prevalent in the Norwegian goat population (Nord *et al.*, 1998). As there is little information available

The GenBank/EMBL/DDBJ accession numbers for the sequences reported in this paper are AF322109 for full-length CAEV-1GA and DQ015910–DQ015916 for 26KE.1, 26KE.2, G623, G806, G922, LM30 and 1GA partial *env* sequences, respectively.

about the genetic characteristics of SRLVs in Norway, the objective of this study was to examine the genetic properties of a Norwegian isolate of CAEV (CAEV-1GA), primarily to result in improvements in our PCR-based diagnostic methods and also to determine further its degree of genetic relatedness to reference strains from other countries. PCR amplification primers suitable for published SRLV sequences seldom work satisfactorily with CAEV-infected Norwegian goats; this could be explained by significant genomic heterogeneity. To define the variable regions in the *env* gene encoding the surface glycoprotein (SU) further, sequence information from five additional CAEV-infected goats was included in the analysis.

METHODS

Viral DNA and RNA. CAEV strain 1GA was isolated from a culture of explanted synovial cells obtained from an arthritic carpal joint of a naturally infected goat. CAEV isolate 26KE originated from an encephalitic kid and isolates G623, G806, G922 and LM30 originated from asymptomatic goats and were derived from peripheral blood mononuclear cells. Three of the isolates (G623, G922 and G806) originated from the same herd; the other infected animals were from different regions of Norway. The isolates were passaged three to four times in goat synovial membrane (GSM) cell culture following procedures described previously (Rimstad *et al.*, 1993). Genomic DNA was isolated at maximum syncytia formation by using DNAzol reagent (GibcoBRL). Total RNA from GSM cells 5 days post-infection or from cell-free supernatant was isolated by using TRIzol reagent (GibcoBRL) or a QIAamp viral RNA kit (Qiagen) as recommended by the manufacturers.

Amplification, cloning and sequencing of viral segments. Primers for PCR amplification were chosen according to assumed conserved regions of previously published SRLV sequences (Valas *et al.*, 1997; Castro *et al.*, 1999) and sequences obtained from CAEV-1GA. PCRs were performed with 0·5 μg genomic DNA and a buffer containing 2·25 mM $MgCl_2$, 200 μM each dNTP, 30 pmol each primer and 2·5 U DNA polymerase (Expand High Fidelity PCR System; Roche) in a total volume of 50 μl. The temperature profile used included an initial denaturation step at 94 °C for 2 min, then 25–35 cycles of 94 °C for 15 s, 50–55 °C for 30 s and elongation at 68–72 °C for a time dependent on fragment length, and finally, a prolonged extension for 7 min at 72 °C. The DNA fragments generated by PCR were purified with the QIAquick protocol (Qiagen) and subsequently sequenced by using a Dye Terminator kit on an ABI PRISM 377 DNA sequencer (Applied Biosystems). In parallel, when necessary for the sequencing results, the amplicons were cloned into the pCR2.1/pCR-XL TOPO vector (Invitrogen) and clones were purified with a QIAprep miniprep kit (Qiagen) before sequencing. The size of the amplified fragments ranged from 250 bp to 5·8 kbp. The sequencing was carried out by the primer-walking technique in long amplicons. The 5′-LTR region was amplified with the 5′-RACE system (GibcoBRL) to complete the sequence. The consensus CAEV-1GA sequence was determined on the basis of sequence data obtained from at least two independent clones and PCR or RT-PCR fragments. Details of the primers used are available upon request. RNA (5 μl) was used in a one-step RT-PCR (Superscript; Invitrogen) consisting of 50 °C for 30 min, then 94 °C for 2 min; further conditions were identical to those used for regular PCR. To amplify the Rev-encoding regions, primers used for RT-PCR of the *rev* transcript were 145− (5′-AGCGGCTCTCGCAGCTGGC-GCCCAAC-3′) and −8501 (5′-ATCACCAACGGCGCCACCACCA-CATT-3′). For alignment and phylogenetic analysis of the SU region, RT-PCR-generated *SU* sequences from isolates 26KE, G623, G806, G922 and LM30 were compared with the equivalent CAEV-1GA sequence. For each of the isolates, the consensus sequence was determined by using at least three clones and PCR products from independent RT-PCR procedures. The sequences reported in this paper have been given GenBank accession numbers AF322109 for full-length CAEV-1GA and DQ015910–DQ015916 for 26KE.1, 26KE.2, G623, G806, G922, LM30 and 1GA partial *env* sequences, respectively.

Sequence analysis, multiple sequence alignments and phylogenetic analysis. Sequence assembly, multiple alignments and editing were performed in the Vector NTI software package (Invitrogen). SignalP was used to predict the signal-peptide cleavage site (Nielsen *et al.*, 1997) and SimPlot 2.5 software (Lole *et al.*, 1999) was used to generate a similarity plot. Alignments of obtained nucleotide sequences from mature SU and available *SU* sequences from GenBank were used for phylogenetic analysis. Neighbour-joining (NJ) analyses were carried out by using MEGA 2.1 (Kumar *et al.*, 2001) and likelihood-mapping and maximum-likelihood (ML) analyses were performed by using the quartet-puzzling algorithm implemented in TREE-PUZZLE (Schmidt *et al.*, 2002). The analyses were conducted by using the Tamura–Nei model with parameter estimation for gamma distribution of rate variation among sites and ignoring all sites with gaps. Support was assessed through 1000 bootstrap replicates or as percentage of 1000 puzzling steps. The constructed trees were visualized by the TreeView program (Page, 1996). To investigate possible selective forces affecting the SU part of the *env* gene, a *Z*-test of selection implemented in MEGA based on the proportion of synonymous and non-synonymous substitutions was used. Estimates were derived by using a modified Nei–Gojobori method.

RESULTS

Nucleotide sequence of CAEV-1GA

The full-length genomic sequence of CAEV-1GA consisted of 8919 bp. To determine experimentally the cross-sectional variation in the quasispecies, six clones of a 1799 bp *env* sequence were compared with the original sequence. The variation was determined to be 0·15 % and was assumed to have no significant effect on the results. To determine the significance of the CAEV-1GA sequence as a replication-competent genome representative of the native virus, sequences obtained from the RT-PCR product of LTR, *pol* and *env* segments were compared with the original sequence. The RT-PCR product contained four substitutions in the U3 segment and the divergence observed in the *pol* and *env* segments was <1 %.

Comparison of the CAEV-1GA sequence with those of SRLV representatives

The obtained sequence of CAEV-1GA was aligned optimally to other full-length or nearly full-length SRLV sequences. A similarity plot (Fig. 1) was used to evaluate the genetic relatedness of the CAEV-1GA isolate to other SRLV sequences. Sequence similarity across the genome of CAEV-1GA showed highest overall similarity to CAEV-Co. The *pol* region was the most conserved, as observed previously for SRLVs (Pyper *et al.*, 1984), with a mean of 75 % similarity to the representative SRLV sequences. The predicted amino

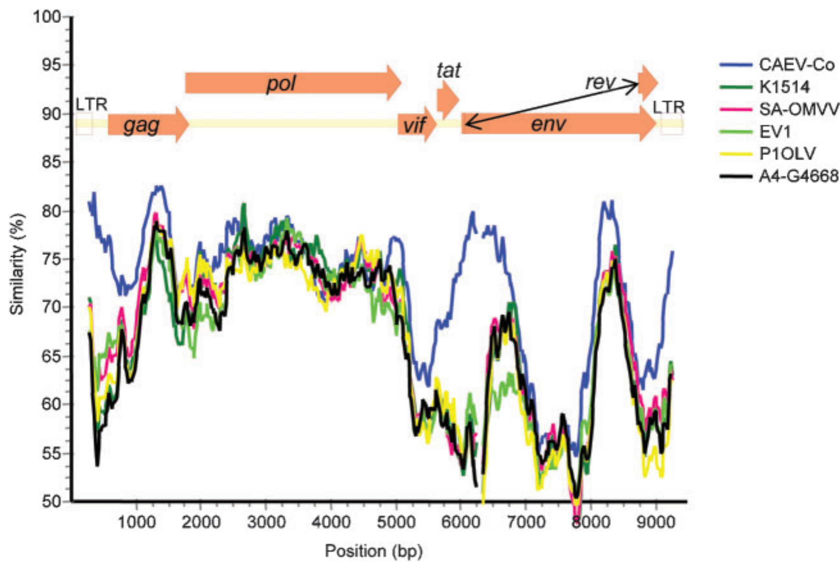

Fig. 1. Nucleotide similarity plot of full-length or nearly full-length SRLV sequences, using CAEV-1GA as the query sequence, a window of 400 nt, step size of 20 bases and sites containing gaps excluded. GenBank accession numbers are as given in the legend to Fig. 4.

acid sequence of *pol* showed a mean of 82 % similarity. High divergence was found in the LTR region as well as the 5′ part of the *env* gene (mean 54 % similarity), in the region corresponding to the surface (SU) glycoprotein encoded by the *env* gene, with mean predicted amino acid sequence similarity of <50 %. The transmembrane part of the 1GA Env glycoprotein was found to be more conserved than the surface unit (SU) and had the hydrophobicity profile found among SRLVs and the highly conserved lentiviral immunodominant epitope of the external domain (Pancino *et al.*, 1994). There was an 8 bp overlap (CAEV-1GA positions 5702–5709) between the open reading frames (ORFs) of *tat* and *vif*, a motif found to be conserved among CAEV-Co and Brazilian CAEV isolates, but not among the MVV group (Castro *et al.*, 1999).

Analysis of CAEV-1GA *rev* transcripts and the leader region of *env*

Sequence analysis identified a putative *env* gene extending from nt 6004 that could encode an Env precursor glycoprotein consisting of 849 aa, considerably shorter than those of other SRLVs (CAEV-Co, 944 aa; MVV group, 985–999 aa) (Sonigo *et al.*, 1985; Querat *et al.*, 1990; Saltarelli *et al.*, 1990; Sargan *et al.*, 1991; Barros *et al.*, 2004). An alignment with CAEV-Co was made to elucidate the variable regions and revealed deletions and hypervariability in the leader peptide (Fig. 2). A signal-peptide cleavage site was predicted between residues 37 and 38 of the Env precursor protein sequence of CAEV-1GA, implying a leader peptide 46 aa shorter than that of CAEV-Co. This truncation would not influence the size of the deduced mature Env glycoprotein. On the other hand, the proposed Rev protein of CAEV-1GA, being partly encoded by the 5′ end of the *env* gene, would be smaller than the 133 aa Rev protein of CAEV-Co (Saltarelli *et al.*, 1994). To address this, CAEV-1GA *rev* transcripts were amplified from total RNA isolated from infected GSM cells. Four RT-PCR products of different sizes were cloned and sequenced (Fig. 3). Sequence alignment allowed identification of four exons within the major 540 bp product, extending from the binding region of the 5′ primer to the splice donor at bp 382 (exon 1), bp 4898–4943 (exon 2, 3′ end of *pol*), bp 6002–6046 (exon 3, 5′ end of *env*) and bp 8287 to the primer-binding region at the 3′ end of *env* (exon 4). The 495 bp product contained exons 1, 3 and 4 and the 451 bp product contained exons 1 and 4, whilst the minor 424 bp product contained truncated forms of exon 1 (truncated at the 5′ end) and exon 4 (truncated at the 3′ end) and non-consensus splice-site motifs. This minor product was thus considered a possible result of a cryptic splicing event or a PCR-created artefact. The putative ORFs of the 540 and 495 bp products were equal and alignment studies indicated that the ORF encoded a potential Rev protein with 14 aa translated from the 5′ end of the *env* gene (exon 3) spliced to exon 4, producing a 106 aa Rev protein.

Sequence diversity in the mature SU region of Norwegian CAEV isolates

The predicted mature SU sequence of CAEV-1GA contains 515 aa with 22 potential *N*-linked glycosylation sites and 20 cysteine residues. To increase the sequence information in this region, we sequenced partial *env* genes from five other CAEV-infected goats of different geographical origin and disease status, resulting in deduced mature SU sequences of 515–516 aa with 20 cysteine residues and 21–23 potential *N*-linked glycosylation sites. Two distinct *SU* variants were obtained from the 26KE isolate, probably representing sequence variants in the virus pool, whilst isolates G623, G806, G922 and LM30 gave one consensus *SU* sequence (Fig. 2). All cysteine residues and the majority of the potential *N*-linked glycosylation sites are conserved among the Norwegian isolates. Amino acid similarities among the Norwegian isolates were 81–94 %, whereas similarities to CAEV-Co were as low as 45–46 % in the mature SU region.

```
                    Rev exon                  >                                   Leader >< SU
1GA      : --------------------MAEIRKEAKEPLIQ---GKYKKQVSRITIVILLTVRAALG-----------------------AEYITLISDPYGFSPVR : 54
26KE.1   : --------------------.EKEKEKR.LLTQ----.R.QY...KVI..TIIAIGTVQ.-----------------------.D............... : 53
26KE.2   : --------------------.EKEKEKR.LLTQ----.R.QY...KVI..TIIAIGTVQ.-----------------------T................ : 53
G623     : -------------------MAETDMEK..LVAP----...LNL..K.I..-F.AI.VG..-----------------------.D............... : 53
G806     : -------------------MAGTDMEK..LIAP----...LDL..K.I.F-..AI.VG..-----------------------................K : 53
G922     : --------------------.ERNKEKE..M...----...QY...G.SL...IIME.VQ.-----------------------VD..............K : 54
LM30     : --------------------..GTKQ.E..AILRA---...LN...KGV.L-F.VITVTC.-----------------------E................ : 54
CAEV-Co  : MDAGARYMRLTGKENWVEVT.DGEKERKR.GFTAGQQ...QP...KQIGNRNTNPCF.YKGIFLWRISLTMWILLGINMCVSAED.............IK : 100

1GA      : NVSGVPVTCVTKEFSKWGCQPIGAYPDPDLEYRNISKEILEEVYQQDWPWNTYHWPLWQMDNVVQWARQNLQDN----RKEKRDLADLLAGKIRGRFCVP : 150
26KE.1   : ...................................ED..R.....E.......................AD----I...K.I.N................ : 149
26KE.2   : ...................................EN........E.................K....AD----N...K.I.................. : 149
G623     : ................L.....EI...............K..K.......................GD----K...K.................... : 149
G806     : ...........................................K...E......................GD----K.Q.K.................... : 149
G922     : .......................................R........................G.N---........................... : 151
LM30     : ......................EI...............S.......................K.----N...K.........T............ : 150
CAEV-Co  : ...........A.........L......EI....V.Q.VVK....EN.............E..RY.LKE.M.E.QQRKNNT.EGIEE....T........ : 200
                                                                                         <      V1      >

1GA      : YPFALLECMEWCWWVKNTNAGGYGEADIRINCSRARAVSCTSEMPLASLQRVYWEKEERKNMEKMTIKPCNKNLECKNRRG-CAEGYPVPPKAELFPPAF : 249
26KE.1   : ...............S...R.......K...TK.......................D...QR.....I.....E..K..EN..-.......I..G...Y.... : 248
26KE.2   : ...............D...W.......K...TK.......................D...QR..............E..K..EN..-.......I..G...Y.... : 248
G623     : ...............G...W.............Q.......H.L....V.....SQ.DKAA..R.I.....E....TKK..-.......I..G........ : 248
G806     : ...............D.D.W.......K.....Q.K.....H......V.....N..D.TA.KR.I.....PD.........-.......I..G........ : 248
G922     : ..............RD...W.............Q.......H.I....V.....SQ.DKAD..R......K.D...EKK..-.......I..GT....... : 250
LM30     : ...............G.D.W.............Q.......N......V.....NR.DTAS....AM.....P.....R....-.......I..G........ : 249
CAEV-Co  : .......K.TK...YTAAI.NESGKAGK.K...TE.......ED.....I..A..DEKD.ES.AF.N..A.DS..R.QK.P.G.M....I.VG..II.ESM : 300
                          <        V2        >

1GA      : QDLQPKGYAYGALRG-NSKFPQRVSLRTWVKIANLTGWEKGKPAEWWNNSQQVH------WFDTTPQYHLGYVLSRAPENRSCNFTGEIRIGQHQFEYNY : 342
26KE.1   : ...K...........-.TT...G..I.....M.........N...TT.V.S.....------..NK..............G........V.....R.K... : 341
26KE.2   : ...K...........-.TT...G..I.....M.........N....T.V.S.....------..NK..............G........V.....R.K... : 341
G623     : ...............-.TT..HG..I.....M.......KE....K.V.E......------..YN......WK..Q..........K.LL...K.D... : 341
G806     : ...............-.TT...G..I.....V..M.........K.V.E......------..RNEH.....WE..Q.........K...K.K....E..... : 341
G922     : R..............-.TT...G..I.....V..M....KE....K.V.E......------..GKKS.....WE..Q......K......K.K....E.Q... : 343
LM30     : ...R.RK........-..T...G..I.....M..V.N.KD......V.M......--------.R..........K........G......K.LL...V.N... : 342
CAEV-Co  : KY.RGAKSQ..GIKDK.GELKLPLT..V...L..VSE.VN.T.PD.QDRINGSKGINGTL.GELNSMH...FA..Q--NGKW..Y....KL..ET.Q.H. : 398
                                                      < V3                                   V3 >

1GA      : TLTKNCTKEKWKEYPMWHVWRHLDQNEHLSSICFKRPRRNATQIGNSTLQGQCNRSNWTGCHCNETGI-NTTWRINGTKGAYLLNSTNGNIMVLLCW--- : 438
26KE.1   : .IAE...GNRSQD...........................A....V..S.D..........RTV................................--- : 438
26KE.2   : .IAG...GNRSQG...........................A....V..S.D..........RTV..............T..................--- : 438
G623     : ...E...QGN.SQD..........................E....V..S..............F-.P..............T..................--- : 437
G806     : ...E...QGN.SQD..........................A....V..T.S.............L-.P...T..........I..................--- :437
G922     : ...E...QGN.SQD..........................E....V..T..........Q.....L-.P.....................K..........--- : 439
LM30     : .I.E...EQRKG.....I........................A..D.V..N.............E-.....K...........................--- : 438
CAEV-Co  : KPNW...-GN.TQ..V.Q.I.N..MV..MTGE.VQ..Q.HNITV..G.IT.N.STT..D..N.SRS.N-------------..Y..SE.GLLLI..RQNS : 484
                                                                                                   < V4

1GA      : -----------NTTVAGVYESQLK-WNESLKDGDYGLCFNSTNRNCTRNGARHYVNKRVIKNDT-------------------ADHNCDSSISAIDGMV : 506
26KE.1   : -----------N.........EV.-..............K.....S.S.T..W...DR...R.A.-------------------...............I : 506
26KE.2   : -----------ND.I.......T.-.......................T..W...DR...R.AS-------------------............L...I : 506
G623     : -----------N..........NTP-.......................R..W...DR..LN.A.-------------------.........V..V...I : 505
G806     : -----------NA.........NMT-....R.................KK..W...DR..LN.A.-------------------.............V.... : 505
G922     : -----------N..........NTP-..............T.......V..W...DR..LN.A.-------------------.............V..VI : 507
LM30     : -----------NE..V......T.-.....S...F.......T.....KQ..W...DR..LN.A.-------------------.............V...I : 506
CAEV-Co  : TLTRILGTNTNW..MW.I.KNCSGCE.AT.DNTGE.TLGGVA.K..SLPHKNESNKWTCAPRQRDGKTDSLYIAGGKKFWTRIK.QFS.E.N.GQL...L : 584
                                                                          V4 >

                                    SU >
1GA      : HQQILLQRYQVIRVRAYTYGVIDMPDNYETL--PGRRRRDLAKARKKR : 552
26KE.1   : ...L.........K........M...EQF..G--IR.KK.....S.Q.. : 552
26KE.2   : ...L.F.......K........MN..EQF..G--IR.K.....R..R.. : 552
G623     : .............K.........Q..EHF..R--IR.K......Q.R.. : 551
G806     : .............K.........E..EHF..G--IR.K......Q.R.. : 551
G922     : ...L......I.K.........E..EHF..G--IR.K......Q.R.. : 553
LM30     : ...L.........K.........E...HF..G--IR.K......R.Q.. : 552
CAEV-Co  : ........K...K.........E..E..AKTRIIN.KK.E.SHK.... : 632
                                   <     V5     >
```

Fig. 2. Alignment of partial Env amino acid sequences derived from Norwegian CAEV isolates versus the CAEV-Co sequence, showing residues that differ from CAEV-1GA. Gaps represented as dashes were introduced to improve the alignment. Underlining represents potential *N*-linked glycosylation sites (NXT or NXS) and cysteine residues, which are also in bold type. The boundaries of the Rev N-terminal coding exon, leader peptide and mature SU are indicated above the alignment and variable regions V1–V5, defined previously (Valas *et al.*, 2000), are shown below the alignment.

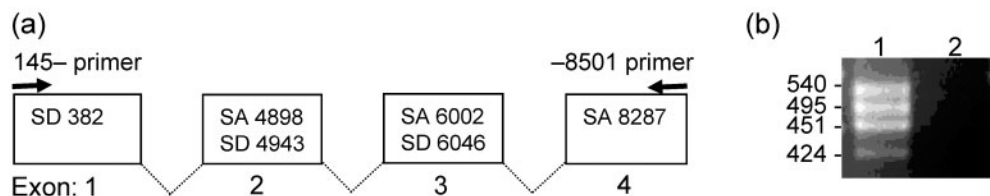

Fig. 3. Organization of CAEV-1GA *rev* transcripts. (a) Exons represented as boxes. Numbers in boxes indicate the position of splice-donor (SD) and splice-acceptor (SA) sites according to CAEV-1GA numbering. (b) RT-PCR products generated from total RNA isolated from CAEV-1GA-infected GSM cells (lane 1) or uninfected GSM cells (lane 2), using primers as shown in (a). The number on the left represents transcript length (bp).

When comparing SU sequences from Norwegian isolates with that of CAEV-Co (Fig. 2), the variable regions V1, V2 and V5 are in accordance with previous descriptions (Valas *et al.*, 2000) and, with the exception of V2, revealed small deletions of 2–4 aa. The variability in regions V3 and V4 is extensive and extends beyond the regions defined previously. The alignment revealed deletions of 6 aa in the V3 region and 22–23 aa in the extended V4 region, resulting in a size of mature SU from Norwegian CAEV isolates of 515–516 aa. Although 19 out of 20 cysteine residues in the SU sequence of Norwegian CAEV isolates remain conserved when compared with CAEV-Co, substitutions and deletions in the V4 region correspond to one cysteine residue in a different position and the loss of two cysteine residues found in the CAEV-Co sequence.

Phylogenetic comparisons of the SU region

Phylogenetic analyses were carried out with several programs to determine further the genetic relationship between mature *SU* sequences from Norwegian CAEV isolates and corresponding *SU* sequences obtainable from GenBank. Likelihood mapping was conducted to assess the phylogenetic signal in the *SU* alignment and revealed 94 % fully resolved quartet trees, indicating that the phylogenetic signal in the alignment was highly informative. Estimates for the transition/transversion ratio and the gamma-distribution parameter for rate heterogeneity were 2·20 and 0·63, respectively. The tree topologies resulting from NJ (not shown) and ML (Fig. 4) analyses were highly similar. The sequences reported here cluster together with high support values and form a unique group, suggested by the long branch that separates them from other SRLVs.

The differences between the number of synonymous and non-synonymous substitutions in the *SU* alignment were analysed with a *Z*-test of selection. The results indicated the existence of positive selection in the SU region.

DISCUSSION

SRLVs have previously been classified phylogenetically into at least six clades, with no clear separation according to host species or geographical origin (Zanoni, 1998; Rolland *et al.*, 2002). A recently proposed phylogenetic organization of SRLVs involves classification into four sequence-equidistant groups, termed A–D (Shah *et al.*, 2004). According to these analyses, sequences within *gag* and *pol* from two Norwegian goats did not cluster with any other SRLV sequences and were classified in separate groups, termed clade III or group C (Zanoni, 1998; Shah *et al.*, 2004). The result of the present analysis based on *SU* sequences and the extended number of sequences, including several obtained from virus isolates from herds in different regions of Norway, gives further support for the unique group designation. Strains of SRLVs isolated from different parts of the world have been shown to group together phylogenetically, possibly as a result of spread from a common origin (Rolland *et al.*, 2002; Shah *et al.*, 2004). The results described here demonstrate that there is larger sequence diversity and geographical heterogeneity among SRLVs than reported previously.

A *Z*-test of selection based on quantities of synonymous and non-synonymous substitutions indicated that the diversity in the SU region is a result of positive selection, meaning that amino acid-altering substitutions offer fitness advantages. The extent to which selective forces contribute to sequence diversity in SRLVs remains unclear. An indication of positive selection in the *env* region of SRLVs has been found previously (Zanoni, 1998; Rolland *et al.*, 2002). The possibility that selective forces could be related to SRLV adaptation to relatively genetically homogeneous host populations has been suggested (Bertoni *et al.*, 2000). Indeed, the domestic goat in Norway represents a population that has been kept in isolation for an extended period of time due to limited importation. The CAEV isolates examined are suggested to

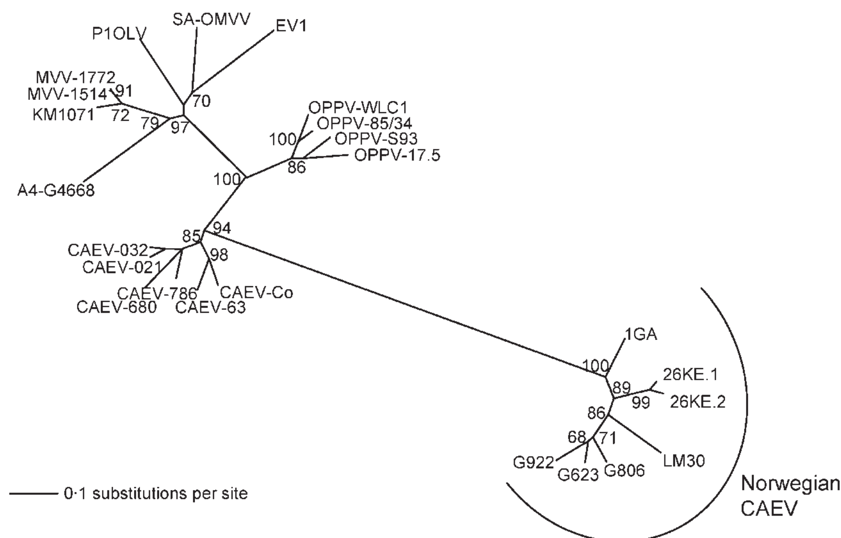

Fig. 4. Unrooted ML tree of aligned SU-encoding regions from SRLVs. The Norwegian CAEV isolates analysed are indicated. Branch lengths are drawn to scale. Support values above 65 % are shown. GenBank accession numbers of analysed *SU* sequences from CAEV and MVV isolates: CAEV-Co (M33677), CAEV-63 (M60855), CAEV-680 (AJ400718), CAEV-786 (AJ400721), CAEV-021 (AJ400719), CAEV-032 (AJ400720), EV1 (S51392), MVV-1514 (M10608), MVV-1772 (S55323), KM1071 (U51910), SA-OMVV (S31646), P1OLV (AF479638), A4-G4668 (AY445885); from ovine progressive pneumonia virus isolates: OPPV-S93 (AF338226), OPPV-85/34 (U64439), OPPV-17.5 (AY362030), OPPV-WLC1 (AY362038).

be representatives of CAEV that have persisted and co-evolved with the Norwegian domestic goat.

The genetic variability of SRLV SU has been defined to five major regions, V1–V5 (Valas *et al.*, 2000). Variant regions V1, V2 and V5 of Norwegian CAEV SU display characteristics similar to those described previously. The regions around V2 and V5 contain conserved motifs, corresponding to inner domains that may be involved in the conformational structure of SU and mediate interactions with the transmembrane glycoprotein TM (Hötzel & Cheevers, 2000, 2003). Regions V3 and V4 extend beyond the locations defined previously and demonstrate higher sequence variation than previous descriptions of SRLV mature SU. The V3–V4 region contains the majority of the potential *N*-linked glycosylation sites, hence suggesting an exposed nature for this domain and potential for immune evasion. The size of the deduced mature SU from Norwegian CAEV isolates is 30–35 aa shorter than representative CAEV mature SU sequences, which contain 546–550 aa with 22 cysteines and 20–23 potential *N*-linked glycosylation sites (Saltarelli *et al.*, 1990; Knowles *et al.*, 1991; Valas *et al.*, 1997, 2000). Conservation is found for the majority of the cysteine residues in Norwegian CAEV SU sequences, although substitutions and deletions in the V4 region lead to the loss of two residues otherwise conserved in representative SRLV sequences. The functional relevance of the deletions and diversity in the V4 region of Norwegian CAEV are of particular interest, as they suggest the size of the putative outer domains of SU to be smaller than described previously for SRLV SU and to contain extensive antigenic variations.

The C-terminal part of CAEV SU has been reported to contain the most immunogenic regions of SU (Bertoni *et al.*, 2000; Valas *et al.*, 2000). The location of a conformation-dependent neutralization epitope in connection with a cysteine-bridged loop has been proposed in the V4 region of MVV (Skraban *et al.*, 1999). The respective area in CAEV-63 SU includes a region that demonstrates rapid sequence changes during persistent infection and may be part of a variable neutralization epitope (Hötzel *et al.*, 2002). In the SRLV V4 region, the location of the cysteine loop of 12 residues is highly conserved, with a relatively conserved sequence (Sonigo *et al.*, 1985; Querat *et al.*, 1990; Saltarelli *et al.*, 1990; Sargan *et al.*, 1991; Valas *et al.*, 1997, 2000; Barros *et al.*, 2004) and is suggested to have a function analogous to that of the V3 region of human immunodeficiency virus type 1 (HIV-1) (Skraban *et al.*, 1999). A similar motif was not found in Norwegian CAEV SU sequences. The CAEV-1GA 'CFNSTNRNC' sequence at positions 464–472 of the Env amino acid sequence could be in an analogue position, as the last cysteine residue in the C-teminal part of SU is located in a conserved region (Fig. 2). The fact that this alternative sequence loop is present in isolates obtained from different regions of Norway suggests that this motif represents traits linked to Norwegian CAEV and may contribute to distinct immunogenic domains.

Genetic analysis of the *env* region showed that the variation observed in the SU amino acid sequence distinguishes Norwegian CAEV from representative SRLVs. Considering the importance of the SU region for viral pathogenesis and cell tropism, the genetic diversity in the region must be under some functional restriction. Comparing sequences from distinct SRLVs may help to identify conserved patterns of SU that are important for viral functions.

In addition to the SU region, the predicted sequence of the CAEV-1GA *env* gene suggested that the Rev-encoding region differs significantly from that of CAEV-Co Rev. In HIV-1 and SRLVs, the Rev protein regulates the switch between early and late phases of virus infection by exporting unspliced or singly spliced viral mRNA to the cytoplasm (Malim *et al.*, 1989). Functionally important regions include an arginine-rich basic domain that mediates nuclear localization, as well as RNA binding through the RNA target sequence (RRE), and a leucine-rich domain that functions as a nuclear-export signal (Schoborg & Clements, 1996; reviewed by Cullen, 1998). Both the conserved leucine-rich and arginine-rich domains, containing eight arginine residues in one stretch, were present within the deduced CAEV-1GA Rev protein. A third domain containing N-terminal residues contributes to Rev multimerization and the formation of such a Rev multimer on RRE has been reported to be required for HIV-1 Rev activity (Malim & Cullen, 1991).

The proposed CAEV-1GA Rev protein is 27 aa shorter than the corresponding protein from CAEV-Co (Saltarelli *et al.*, 1994), and the major reason for this is the small N-terminal part of 14 aa encoded by exon 3. Mutational analysis of CAEV-Co Rev has allowed identification of several residues within the N-terminal coding exon that are required for optimal Rev–RRE binding. These residues were proposed to form secondary structures that participate in interactions and possibly interact directly with the RRE (Abelson & Schoborg, 2003). A more detailed analysis of this region is needed to characterize the functional effect of CAEV-1GA N-terminal Rev protein truncation.

The 451 bp *rev* transcript lacked exon 3. Similar transcripts have been described from other CAEV strains where a Rev protein encoded from a non-AUG initiation codon yielded an isoform lacking the N-terminal region, although it was not possible to detect the Rev isoform in CAEV-infected cells or in mature viral particles (Gazit *et al.*, 1996). On the other hand, the presence of such alternative *rev* transcripts among CAEV strains may indicate a functional role in the viral life cycle for Rev proteins encoded by this gene.

In summary, both phylogenetic analysis and sequence alignment of the SU amino acid sequence revealed a distant genetic relationship between Norwegian CAEV and other SRLVs. The characterization of CAEV-1GA will facilitate more thorough studies on SRLV field isolates. Different grades of genetic variation are observed in the various regions of the SRLV genome. The sequence data obtained

allow us to locate conserved regions and give valuable information for evaluating diagnostic methods that are influenced by variations in the SRLV target sequence. Further sequence information available from SRLVs that have evolved under different conditions will give a better understanding about the genetic diversity within SRLVs, and possibly insight into common structural features.

ACKNOWLEDGEMENTS

This study was supported by grants no. 123795/122 and 143286/140 from the Norwegian Research Council. We thank Christine M. Jonassen for critical review of the manuscript.

REFERENCES

Abelson, M. L. & Schoborg, R. V. (2003). Characterization of the caprine arthritis encephalitis virus (CAEV) rev N-terminal elements required for efficient interaction with the RRE. *Virus Res* **92**, 23–35.

Andresson, Ó. S., Elser, J. E., Tobin, G. J. & 11 other authors (1993). Nucleotide sequence and biological properties of a pathogenic proviral molecular clone of neurovirulent visna virus. *Virology* **193**, 89–105.

Barros, S. C., Ramos, F., Duarte, M., Fagulha, T., Cruz, B. & Fevereiro, M. (2004). Genomic characterization of a slow/low maedi visna virus. *Virus Genes* **29**, 199–210.

Bertoni, G., Hertig, C., Zahno, M.-L. & 7 other authors (2000). B-cell epitopes of the envelope glycoprotein of caprine arthritis–encephalitis virus and antibody response in infected goats. *J Gen Virol* **81**, 2929–2940.

Blacklaws, B. A., Berriatua, E., Torsteinsdottir, S., Watt, N. J., de Andres, D., Klein, D. & Harkiss, G. D. (2004). Transmission of small ruminant lentiviruses. *Vet Microbiol* **101**, 199–208.

Castro, R. S., Greenland, T., Leite, R. C., Gouveia, A., Mornex, J.-F. & Cordier, G. (1999). Conserved sequence motifs involving the *tat* reading frame of Brazilian caprine lentiviruses indicate affiliations to both caprine arthritis–encephalitis virus and visna–maedi virus. *J Gen Virol* **80**, 1583–1589.

Cullen, B. R. (1998). Retroviruses as model systems for the study of nuclear RNA export pathways. *Virology* **249**, 203–210.

Gazit, A., Mashiah, P., Kalinski, H., Gast, A., Rosin-Abersfeld, R., Tronick, S. R. & Yaniv, A. (1996). Two species of Rev proteins, with distinct N termini, are expressed by caprine arthritis encephalitis virus. *J Virol* **70**, 2674–2677.

Grego, E., Profiti, M., Giammarioli, M., Giannino, L., Rutili, D., Woodall, C. & Rosati, S. (2002). Genetic heterogeneity of small ruminant lentiviruses involves immunodominant epitope of capsid antigen and affects sensitivity of single-strain-based immunoassay. *Clin Diagn Lab Immunol* **9**, 828–832.

Hötzel, I. & Cheevers, W. P. (2000). Sequence similarity between the envelope surface unit (SU) glycoproteins of primate and small ruminant lentiviruses. *Virus Res* **69**, 47–54.

Hötzel, I. & Cheevers, W. P. (2003). Caprine arthritis-encephalitis virus envelope surface glycoprotein regions interacting with the transmembrane glycoprotein: structural and functional parallels with human immunodeficiency virus type 1 gp120. *J Virol* **77**, 11578–11587.

Hötzel, I., Kumpula-McWhirter, N. & Cheevers, W. P. (2002). Rapid evolution of two discrete regions of the caprine arthritis-encephalitis virus envelope surface glycoprotein during persistent infection. *Virus Res* **84**, 17–25.

Karr, B. M., Chebloune, Y., Leung, K. & Narayan, O. (1996). Genetic characterization of two phenotypically distinct North American ovine lentiviruses and their possible origin from caprine arthritis-encephalitis virus. *Virology* **225**, 1–10.

Knowles, D. P., Jr, Cheevers, W. P., McGuire, T. C., Brassfield, A. L., Harwood, W. G. & Stem, T. A. (1991). Structure and genetic variability of envelope glycoproteins of two antigenic variants of caprine arthritis-encephalitis lentivirus. *J Virol* **65**, 5744–5750.

Kumar, S., Tamura, K., Jakobsen, I. B. & Nei, M. (2001). MEGA2: molecular evolutionary genetic analysis software. *Bioinformatics* **17**, 1244–1245.

Leroux, C., Chastang, J., Greenland, T. & Mornex, J. F. (1997). Genomic heterogeneity of small ruminant lentiviruses: existence of heterogeneous populations in sheep and of the same lentiviral genotypes in sheep and goats. *Arch Virol* **142**, 1125–1137.

Lole, K. S., Bollinger, R. C., Paranjape, R. S., Gadkari, D., Kulkarni, S. S., Novak, N. G., Ingersoll, R., Sheppard, H. W. & Ray, S. C. (1999). Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol* **73**, 152–160.

Malim, M. H. & Cullen, B. R. (1991). HIV-1 structural gene expression requires the binding of multiple Rev monomers to the viral RRE: implications for HIV-1 latency. *Cell* **65**, 241–248.

Malim, M. H., Hauber, J., Le, S.-Y., Maizel, J. V. & Cullen, B. R. (1989). The HIV-1 rev *trans*-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. *Nature* **338**, 254–257.

Narayan, O., Zink, M. C., Gorrell, M., Crane, S., Huso, D., Jolly, P., Saltarelli, M., Adams, R. & Clements, J. E. (1993). The lentiviruses of sheep and goats. In *The Retroviridae*, vol. 2, pp. 229–256. Edited by J. A. Levy. New York: Plenum.

Nielsen, H., Engelbrecht, J., Brunak, S. & von Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng* **10**, 1–6.

Nord, K., Rimstad, E., Storset, A. K. & Løken, T. (1998). Prevalence of antibodies against caprine arthritis-encephalitis virus in goat herds in Norway. *Small Rumin Res* **28**, 115–121.

Page, R. D. M. (1996). TreeView: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci* **12**, 357–358.

Pancino, G., Ellerbrok, H., Sitbon, M. & Sonigo, P. (1994). Conserved framework of envelope glycoproteins among lentiviruses. *Curr Top Microbiol Immunol* **188**, 77–105.

Peterhans, E., Greenland, T., Badiola, J. & 14 other authors (2004). Routes of transmission and consequences of small ruminant lentiviruses (SRLVs) infection and eradication schemes. *Vet Res* **35**, 257–274.

Pyper, J. M., Clements, J. E., Molineaux, S. M. & Narayan, O. (1984). Genetic variation among lentiviruses: homology between visna virus and caprine arthritis-encephalitis virus is confined to the 5′ *gag-pol* region and a small portion of the *env* gene. *J Virol* **51**, 713–721.

Pyper, J. M., Clements, J. E., Gonda, M. A. & Narayan, O. (1986). Sequence homology between cloned caprine arthritis encephalitis virus and visna virus, two neurotropic lentiviruses. *J Virol* **58**, 665–670.

Querat, G., Audoly, G., Sonigo, P. & Vigne, R. (1990). Nucleotide sequence analysis of SA-OMVV, a visna-related ovine lentivirus: phylogenetic history of lentiviruses. *Virology* **175**, 434–447.

Rimstad, E., East, N. E., Torten, M., Higgins, J., DeRock, E. & Pedersen, N. C. (1993). Delayed seroconversion following naturally acquired caprine arthritis-encephalitis virus infection in goats. *Am J Vet Res* **54**, 1858–1862.

Rolland, M., Mooney, J., Valas, S., Perrin, G. & Mamoun, R. Z. (2002). Characterisation of an Irish caprine lentivirus strain – SRLV phylogeny revisited. *Virus Res* **85**, 29–39.

Saltarelli, M., Querat, G., Konings, D. A. M., Vigne, R. & Clements, J. E. (1990). Nucleotide sequence and transcriptional analysis of molecular clones of CAEV which generate infectious virus. *Virology* **179**, 347–364.

Saltarelli, M. J., Schoborg, R., Pavlakis, G. N. & Clements, J. E. (1994). Identification of the caprine arthritis encephalitis virus Rev protein and its *cis*-acting Rev-responsive element. *Virology* **199**, 47–55.

Sargan, D. R., Bennet, I. D., Cousens, C., Roy, D. J., Blacklaws, B. A., Dalziel, R. G., Watt, N. J. & McConnell, I. (1991). Nucleotide sequence of EV1, a British isolate of maedi–visna virus. *J Gen Virol* **72**, 1893–1903.

Schmidt, H. A., Strimmer, K., Vingron, M. & von Haeseler, A. (2002). TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **18**, 502–504.

Schoborg, R. V. & Clements, J. E. (1996). Definition of the RRE binding and activation domains of the caprine arthritis encephalitis virus Rev protein. *Virology* **226**, 113–121.

Shah, C., Böni, J., Huder, J. B., Vogt, H.-R., Mühlherr, J., Zanoni, R., Miserez, R., Lutz, H. & Schüpbach, J. (2004). Phylogenetic analysis and reclassification of caprine and ovine lentiviruses based on 104 new isolates: evidence for regular sheep-to-goat transmission and worldwide propagation through livestock trade. *Virology* **319**, 12–26.

Skraban, R., Matthíasdóttir, S., Torsteinsdóttir, S. & 8 other authors (1999). Naturally occurring mutations within 39 amino acids in the envelope glycoprotein of maedi-visna virus alter the neutralization phenotype. *J Virol* **73**, 8064–8072.

Sonigo, P., Alizon, M., Staskus, K. & 7 other authors (1985). Nucleotide sequence of the visna lentivirus: relationship to the AIDS virus. *Cell* **42**, 369–382.

Valas, S., Benoit, C., Guionaud, C., Perrin, G. & Mamoun, R. Z. (1997). North American and French caprine arthritis-encephalitis viruses emerge from ovine maedi-visna viruses. *Virology* **237**, 307–318.

Valas, S., Benoit, C., Baudry, C., Perrin, G. & Mamoun, R. Z. (2000). Variability and immunogenicity of caprine arthritis-encephalitis virus surface glycoprotein. *J Virol* **74**, 6178–6185.

Wain-Hobson, S. (1996). Running the gamut of retroviral variation. *Trends Microbiol* **4**, 135–141.

Zanoni, R. G. (1998). Phylogenetic analysis of small ruminant lentiviruses. *J Gen Virol* **79**, 1951–1961.