Journal of General Virology (1998), **79**, 1951–1961. Printed in Great Britain

Phylogenetic analysis of small ruminant lentiviruses

R. G. Zanoni

Institute of Veterinary Virology, Laenggass-Strasse 122, CH-3012 Berne, Switzerland

Based on available sequence data, a phylogeny of small ruminant lentiviruses (SRLV) was established for *env*, *pol*, *gag* and LTR fragments using maximum likelihood, neighbour-joining and minimum evolution reconstruction techniques. To reconstruct topologies as accurately as possible, phylogenetic parameters like base composition, percentage divergences, kappa, the gamma shape parameter alpha and codon position-specific rates were estimated prior to the reconstruction of trees. Divergences between fragments of SRLV ranged from 16% in *gag* and *pol* to 22% in *env* and 35% in LTR. The codon position bias found and the ratios of synonymous to nonsynonymous substitutions were inversely related to overall divergence, indicating the existence of both negative and positive Darwinian selection in SRLV genes. The phylogenetic trees reconstructed with relative substitution rates assigned to the codon positions revealed an interesting relationship between lentiviruses from sheep and goats. Overall, at least six different clades could be differentiated, with no clear separation of SRLV strains derived from goats (caprine arthritis–encephalitis virus) or sheep (maedi–visna virus). Trees generated with fragments from different coding regions were in good agreement with each other as well as with trees generated with different phylogeny reconstruction methods. In this work, clear indications of the existence and epidemiological importance of cross-species transmission were found.

Introduction

Classical epidemiology traces the spread of infectious agents on the host side by analysing the pattern of disease or seroconversion induced in a population at risk. In contrast, molecular epidemiology, which is based on the phylogenetic analysis of genomic sequences, can be viewed as epidemiology seen from the side of the agent, telling us about its past. Given the high rate of evolution of the infectious agent compared to the host, its spread in a population is reflected in terms of genomic evolution accumulated over time since the origin of the epidemic. This is particularly true for viruses which exhibit high rates of evolution, exceeding those of mammals by nearly six orders of magnitude (Steinhauer & Holland, 1986; Li *et al.*, 1988). Therefore, it is possible to study the genomic evolution of viruses in short time frames. Phylogenetic analysis of viral genomic sequences has proven its extreme usefulness in several areas and is replacing antigenic typing by means of monoclonal antibodies (Smith, 1989). For example, the phylogenetic analysis of influenza viruses allows the reconstruction of evolution and the source of epidemics and the identification of critical reservoirs from which new epidemics might arise (Scholtissek, 1995). For both research and diagnosis of rabies, the phylogenetic analysis of genomic sequences has become an invaluable tool. Phylogenetic analysis of the sequences of strains derived from many regions of the world have revealed an interesting and plausible epidemiological pattern of relatedness between strains, the European type strains being widespread along the routes of past colonization (Sacramento *et al.*, 1992; Smith *et al.*, 1992). Since the sequences of rabies viruses exhibit clear signatures of host species adaptation (Smith *et al.*, 1995), the phylogenetic analysis of PCR fragments is being used routinely for rapid identification of the source species for human rabies cases in USA and Europe (CDC, 1995; Smith *et al.*, 1991). One of many other interesting applications of molecular epidemiology was the analysis of transmission of HIV from a dentist to several patients (Ou *et al.*, 1992).

The small ruminant lentiviruses (SRLV), caprine arthritis–encephalitis virus (CAEV) and maedi–visna virus (MVV) cause lifelong persistent infections, inducing slowly debilitating disease syndromes mainly characterized by chronic progressive arthritis, pneumonitis, mastitis, emaciation and (rarely) meningoencephalomyelitis in goats and sheep (Narayan & Cork, 1985). The application of phylogenetic analysis to CAEV and MVV sequences is promising for both epidemio-

Author for correspondence: R. G. Zanoni.
Fax +41 31 631 2534. e-mail zanoni@ivv.unibe.ch

0001-5464 © 1998 SGM

Table 1. Lentiviral sequences used for phylogenetic analysis

Abbreviations: gene, genomic localization and dataset; Co. CAEV, coordinates with respect to CAEV strain CO; CAEV, caprine arthritis–encephalitis-; MVV, maedi–visna-; OLV, ovine lenti-; BIV, bovine immunodeficiency-like-; JDV, Jembrana disease-; EIAV, equine infectious anaemia-; F-, S-, or HIV, feline-, simian- or human immunodeficiency virus.

Identification	Accession no.	Virus strain	Year	Gene	Co. CAEV*	Reference
ca35811	U35811	CAEV isolate 93036	1995	*env*(-$_L$, -$_S$)	7272–7822	Leroux *et al.* (1995)
ceacaev63	M60855	CAEV strain 75-G63	1992	*env*(-$_L$, -$_S$)	6012–8911	Knowles *et al.* (1991)
mv51910	U51910	MVV strain KM1071	1996	*env*(-$_L$, -$_S$)	5911–8881	V. Andresdottir (unpublished)
ol57506	U57506	OLV strain Sh8A	1996	*env*(-$_L$, -$_S$)	6475–8135	Chebloune *et al.* (1996)
ol57507	U57507	OLV strain Sh8B	1996	*env*(-$_L$)	6475–8135	
olu63532	U63532	OLV strain 84/28	1996	*env*(-$_L$, -$_S$)	6475–8135	Karr *et al.* (1996)
olu63651	U63651	OLV strain 85/34	1996	*env*(-$_L$, -$_S$)	6475–8135	
olu64439	U64439	OLV strain 85/34	1996	*env*(-$_L$, -$_S$)	5926–9071	Woodward *et al.* (1995)
vv35795/857	U35795/857	Visna virus isolate 664	1995	*env*(-$_L$, -$_S$)	7327–7745	Leroux *et al.* (1995)
vv35796	U35796	Visna virus isolate 676	1995	*env*(-$_L$, -$_S$)	7272–7822	
vv35858-61	U35858-61	Visna virus isolate 676	1995	*env*(-$_S$)	7589–7757	
vv35798	U35798	Visna virus isolate 683	1995	*env*(-$_L$, -$_S$)	7272–7822	
vv35799	U35799	Visna virus isolate 684	1995	*env*(-$_L$, -$_S$)	7272–7822	
vv35800	U35800	Visna virus isolate 685	1995	*env*(-$_L$, -$_S$)	7272–7822	
vv35872-74	U35872-74	Visna virus isolate 685	1995	*env*(-$_S$)	7589–7757	
vv35801	U35801	Visna virus isolate 686	1995	*env*(-$_L$, -$_S$)	7272–7822	
vv35804	U35804	Visna virus isolate 724	1995	*env*(-$_L$, -$_S$)	7272–7822	
vv35862	U35862	Visna virus isolate 678	1995	*env*(-$_S$)	7589–7757	
vv35863/5/7/8	U35863/5/7/8	Visna virus isolate 679	1995	*env*(-$_S$)	7589–7757	
vv35870-71	U35870-71	Visna virus isolate 680	1995	*env*(-$_S$)	7589–7757	
cagapr	L78446	CAEV strain 75-G63	1996	*gag*	1025–1153	A. K. Storset & E. Rimstad (unpublished)
cagapra	L78447	CAEV isolate 8062	1996	*gag*	1025–1153	
cagaprb	L78450	CAEV isolate LM30	1996	*gag*	1025–1153	
mvvgag	X54379	MVV strain ZZV 1050	1992	*gag*	876–1679	Zanoni *et al.* (1991*b*)
olgag	L78451	OLV isolate N1	1996	*gag*	1025–1153	A. K. Storset & E. Rimstad (unpublished)
vvgapr	L78453	Visna virus isolate NMV1	1996	*gag*	1025–1153	
mv52358	U52358	MVV strain KM1071	1996	LTR	8895–9093	V. Andresdottir *et al.* (unpublished)
ol39826	U39826	OLV strain CU1	1995	LTR	8872–163	Campbell & Avery (1996)
recealtr	M21924	CAEV integrated	1989	LTR	8855–195	Sherman *et al.* (1986)
ca35807	U35807	CAEV isolate 739	1995	*pol*	2410–2712	Leroux *et al.* (1995)
ca35808	U35808	CAEV isolate 93008	1995	*pol*	2410–2712	
ca35809	U35809	CAEV isolate 93017	1995	*pol*	2410–2712	
ca35810	U35810	CAEV isolate 93036	1995	*pol*	2410–2712	
ca35812	U35812	CAEV isolate 9412	1995	*pol*	2410–2712	
ca35813	U35813	CAEV isolate 9413	1995	*pol*	2410–2712	
ca35814	U35814	CAEV isolate 9414	1995	*pol*	2410–2712	
capopra	L78448	CAEV isolate 8062	1996	*pol*	2251–2695	A. K. Storset & E. Rimstad (unpublished)
capoprb	L78449	CAEV isolate LM30	1996	*pol*	2497–2695	
olpol	L78452	OLV isolate N1	1996	*pol*	2497–2695	
pol75g63	—	CAEV strain 75-G63	1991	*pol*	2497–2695	Zanoni *et al.* (1992)
polg122c	—	CAEV strain G122C	1991	*pol*	2497–2695	
recaepol	K03327	CAEV integrated	1986	*pol*	2218–2868	Chiu *et al.* (1985)
vv35674	U35674	Visna virus isolate 663	1995	*pol*	2238–2712	Leroux *et al.* (1995)
vv35675	U35675	Visna virus isolate 664	1995	*pol*	2238–2712	
vv35676	U35676	Visna virus isolate 676	1995	*pol*	2238–2712	
vv35678	U35678	Visna virus isolate 680	1995	*pol*	2238–2712	
vv35679	U35679	Visna virus isolate 684	1995	*pol*	2238–2712	
vv35680	U35680	Visna virus isolate 685	1995	*pol*	2238–2712	
vv35681	U35681	Visna virus isolate 686	1995	*pol*	2238–2712	
vv35797	U35797	Visna virus isolate 718	1995	*pol*	2410–2712	
vv35802	U35802	Visna virus isolate 678	1995	*pol*	2410–2712	

Table 1. *(cont.)*

Identification	Accession no.	Virus strain	Year	Gene	Co. CAEV*	Reference
vv35803	U35803	Visna virus isolate 721	1995	*pol*	2410–2712	Leroux *et al.* (1995)
vv35805	U35805	Visna virus isolate 724	1995	*pol*	2410–2712	Leroux *et al.* (1995)
vv35806	U35806	Visna virus isolate 732	1995	*pol*	2410–2712	Leroux *et al.* (1995)
ceavcg	M33677	CAEV strain CO	1990	whole	*a*	Saltarelli *et al.* (1990)
olvcg	M31646	OLV strain SA-OMVV	1991	whole	*b*†	Quérat *et al.* (1990)
revlxx	M10608	Icelandic visna lentivirus	1986	whole	*c*†	Sonigo *et al.* (1985)
s51392	S51392	MVV strain EV1	1991	whole	*d*†	Sargan *et al.* (1991)
bim127	K02013	BIV strain HXB3	1990	whole	2295–2597†	Garvey *et al.* (1990)
jdcg	L40990	JDV strain Tabanan/87	1995	whole	2089–2391†	Chadwick *et al.* (1995)
reu01866	M30502	EIAV strain pSPEIAV19	1993	whole	2552–2854†	Payne *et al.* (1994)
ppr	M32690	FIV isolate San Diego	1990	whole	2589–2891†	Phillips *et al.* (1990)
sivrev	U21603	SIV isolate Agm9063	1995	whole	3035–3337†	Hirsch *et al.* (1995 *b*)
hivbrucg	M36968	HIV-1 isolate BRU	1986	whole	2387–2689†	Wain-Hobson *et al.* (1985)
hiv2ben	U01866	HIV-2 isolate BEN	1990	whole	3190–3492†	Kirchhoff *et al.* (1990)

* *a*: env_L, 7272–7822; env_S, 7589–7757; *gag*, 1025–1153; LTR, 8914–9151; *pol*, 2410–2712: *b*: env_L, 7351–7898; env_S, 7668–7833; *gag*, 1033–1161; LTR, 8993–9238; *pol*, 2412–2714: *c*: env_L, 7313–7842; env_S, 7630–7777; *gag*, 1003–1131; LTR, 8925–9132; *pol*, 2370–2672: *d*: env_L, 7356–7912; env_S, 7676–7847; *gag*, 1022–1150; LTR, 8955–9183; *pol*, 2404–2706.
† Coordinates of extracted fragments on respective genomes.

logical questions and pathogenetic issues. In Switzerland, an eradication programme for CAEV was initiated in 1984, leading to a reduction of seroprevalence from more than 60% to less than 5% by 1996. Eradication was mainly based on colostrum-deprivation, strict separation of young kids and regular serological control (Zanoni *et al.*, 1991*a*, 1989). An important issue to be considered in this context is the possibility of transmission of MVV from sheep to goats. However, whereas experimental infection is possible in both directions (Banks *et al.*, 1983), cross-species transmission seems to be of minor importance under field conditions (McKenzie, 1991; Smith *et al.*, 1985). Since no comprehensive phylogenetic analysis of SRLV has been performed so far, the relationship between goat and sheep lentiviruses has yet to be established.

The aims of this work were (1) to establish a comprehensive SRLV phylogeny with advanced phylogenetic tools as a base for future studies of the molecular epidemiology of SRLV, (2) to test the hypothesis that SRLV are grouped into clearly distinguishable ecotypes, and (3) to analyse the speed and direction of evolution in SRLV in comparison with other lentiviruses.

Methods

■ Sequences. All available DNA sequences of SRLV as of June 1997 were identified with the help of the Sequence Retrieval System (Etzold *et al.*, 1996; http://srs.ebi.ac.uk:5000/) and downloaded from the EMBL nucleotide sequence database, Release 50, at the European Bioinformatics Institute (EBI), Hinxton, Cambridge, UK. Being mainly subfragments of the whole genome, the sequences were grouped according to their localization on the genome into LTR (long terminal repeat regulatory genomic domain), *gag* (gene for group antigens), *pol* (gene for polymerase and other enzymes) and *env* sequences (gene for envelope proteins). The longest overlapping fragments of these sequences were extracted in order to accommodate as many different sequences as possible in a single LTR or gene alignment. Fragments with 100% identity to another sequence in the extracted data set were excluded from the analysis. The corresponding fragments of four out of eight whole genome sequences of SRLV were also included. To reduce redundancy, only one out of five closely related or identical whole genome sequences of Icelandic MVV was included (Sonigo *et al.*, 1985). The respective data sets with the LTR, *gag* and *pol* fragments contained 7, 10 and 36 different sequences encompassing 238 bp (positions 8914–9151, part of U3 and R), 129 bp (positions 1025–1153, N-terminal part of CA) and 303 bp (positions 2410–2712, part of RT core region) with respect to the genome of CAEV strain CO (CAEV-CO; Saltarelli *et al.*, 1990). *For pol*, the corresponding fragments of a BIV (bovine immunodeficiency-like virus), JDV (Jembrana disease virus), EIAV (equine infectious anaemia virus), FIV (feline immunodeficiency virus), SIV (simian immunodeficiency virus), HIV-1 (human immunodeficiency virus type 1) and HIV-2 were also included. With the *env* fragments, two data sets with 20 different sequences encompassing 551 bp (env_L, positions 7272–7822, C-terminal part of SU) and with 33 different sequences encompassing 169 bp (env_S, positions 7589–7757) of the genome of CAEV-CO were created. A detailed description of the analysed fragments including database identification, accession number, virus strain, year of submission, genomic localization, coordinates with respect to the genome of CAEV-CO and references, is given in Table 1. All single sequence operations and pairwise homology searches were performed with the Genetics Computer Group sequence analysis software package (GCG, 1991).

■ Multiple sequence alignment. Multiple alignments of the deduced amino acid sequences of the *gag*, *pol* and *env* fragments and of the

nucleotide sequences of the LTR fragments were generated with Clustal W (Thompson *et al.*, 1994). The multiply aligned Gag, Pol and Env sequences were then used as a template for protein-guided nucleotide alignment. Including gaps, this generated aligned nucleotide sequence lengths of 252, 129, 303, 569 and 199 bp for the LTR, *gag*, *pol*, env_L and env_S fragments, respectively.

■ **Phylogenetic analysis.** Important parameters for phylogeny reconstruction are the shape parameter alpha for a discrete gamma distribution of substitution rates, which accommodates for substitution rate variation across sites (e.g. higher substitution rates in mutational hot spots; Yang, 1995; Chang, 1996), and the transition (changes within purines or pyrimidines)/transversion (changes between purines and pyrimidines) rate ratio parameter kappa. Since alpha and kappa are correlated, these parameters were estimated simultaneously by maximum likelihood using Yang's BASEML program implemented in the PAML (Phylogenetic Analysis by Maximum Likelihood) program package (Yang, 1996*a*, *b*). In separate runs of BASEML with the option G (combined analysis of different genes or codon positions) invoked and codon positions assigned in the input file, the substitution rates of the second and third codon position relative to the first codon position (C2 and C3) were estimated in addition to kappa and alpha. The substitution model assumed was Felsenstein's F84 model, which accommodates for unequal base frequencies and transition/transversion rate bias (Felsenstein & Churchill, 1996). In order to reduce the computing time of PAML and to get standard errors for the parameter estimates, all calculations were performed with user trees, which were estimated with Olsen's fastDNAml program (Olsen *et al.*, 1994). FastDNAml is derived from Felsenstein's DNAML (DNA Maximum Likelihood program) version 3.3 (Felsenstein, 1981).

The final ML trees were computed with the categories option of fastDNAml, which uses the ML estimates of both kappa and the relative substitution rates for the codon positions, if applicable. To improve the tree search efficiency, the input order of each data set was jumbled (randomized) up to ten times and local rearrangements were allowed. The statistical confidence of the topologies was assessed with 100 bootstrap replicates (Felsenstein, 1985) using the same parameters and program options.

For comparative purposes, the neighbour-joining (NJ) method (Saitou & Nei, 1987) implemented in MEGA with the Tamura–Nei gamma distance (Tamura & Nei, 1993) and the minimum evolution method (Rzhetsky & Nei, 1993) implemented in METREE (Rzhetsky & Nei, 1994) with the gamma Poisson distance were applied for the reconstruction of phylogenetic trees. The estimate of PAML for alpha was used as gamma parameter. In both cases, 500 bootstrap replicates were performed. All trees were midpoint-rooted.

Simple percentage pairwise divergences between all the sequences of the LTR domain or coding regions were calculated with MEGA (Kumar *et al.*, 1993).

■ **Estimation of synonymous and nonsynonymous substitution rates.** For all possible pairs of SRLV sequences from each coding region, the numbers of synonymous substitutions per synonymous site and those of nonsynonymous substitutions per nonsynonymous site (Nei & Gojobori, 1986) were calculated as simple proportions with MEGA. For both types of substitutions, only non-zero values were retained for the analysis. Thus, synonymous to nonsynonymous substitution rate ratios for each pair of six *gag*, 20 *pol*, 20 env_L and 31 env_S fragments, averages and standard deviations could be calculated.

■ **Statistical analysis.** Deviation of the base frequencies from equality was assessed with the Pearson χ^2 statistic. The differences found between the averages of percentage divergences in the LTR domain or in the different coding regions and between the averages of the synonymous to nonsynonymous rate ratios in the coding regions were assessed with the Kruskal–Wallis one-way analysis of variance for multiple levels and with the Mann–Whitney U test statistic for two levels. The distributions of the data were graphed as boxplots with the median, boxed 2. and 3. quartiles, whiskers (values up to 1·5 box lengths beyond box) and outside values. These statistics were performed with the use of the SYSTAT program package (Wilkinson, 1990).

The significance of differences between the likelihoods of trees estimated with different parameters or assumptions was assessed with the likelihood ratio test statistic (Yang *et al.*, 1995; Sokal & Rohlf, 1995). Briefly, the likelihood ratio test statistic for two trees is $G = 2(\ln L_1 - \ln L_2)$, where $\ln L_{1,2}$ are the log likelihood values of the trees. The distribution of G can be approximated by the χ^2 distribution with $(n-m)$ degrees of freedom, where n and m are the numbers of parameters used for the estimation of the trees. For all statistics, differences with $P < 0·01$ were considered as significant.

Results

Base composition, divergence and phylogenetic parameters

As a prerequisite for an accurate phylogeny reconstruction, parameters like base composition, percent divergences, kappa, the gamma shape parameter alpha, which accommodates for substitution rate variation across sites, and codon position-specific rates were estimated. The base composition of the LTR, *gag*, *pol* and *env* fragments analysed showed significant deviation from equality (Pearson χ^2 $P < 0·001$), with the lentivirus-typical abundance of A (Vanhemert & Berkhout, 1995) in all genomic regions which ranged from 33·8% to 39·5% (Table 2).

The average percentage pairwise divergences between fragments of SRLV were lowest in the *pol* fragments (16·3%), intermediate in the *env* fragments and highest in the LTR fragments (35·1%; Table 2). All differences except those between the *gag* and *pol* fragments were highly significant ($P < 0·001$). For the short *env* fragment, the sequences of more than one quasispecies representative within an isolate were available for five visna virus isolates (664, 676, 685, 679 and 680; Leroux *et al.*, 1995). Pairwise divergences within a single animal as assessed in these *env* sequences ranged from 0·6% to 18·6%.

The transition/transversion rate ratio parameter kappa was only high for the *gag* and *pol* fragments (4·3 and 3·5) and was below 1·5 for all other fragments (Table 3). Interestingly, the mean kappa values derived from pairwise distances were quite poor (over)estimates compared to the ML estimates of kappa. The estimates of the gamma shape parameter alpha, which is inversely related to the amount of substitution rate variation among sites, ranged from 0·22 for the *gag* fragments to 0·7 for the env_S fragments. In the rather conserved coding regions of *gag* and *pol*, lower values were found than in the more variable regions of LTR and *env*. The low values found for alpha in our data clearly indicate the existence of a relatively high variation

Table 2. Average base composition of gene fragments analysed and average percentage pairwise divergence between strains

Abbreviations: LTR, long terminal repeat fragments; *gag*, group antigen fragments; *pol*, polymerase gene fragments of SRLV; env_L, long envelope gene fragments; env_S, short envelope gene fragments.

Gene (*n*)	Average base composition (SD)*				Av. % divergence
	T	C	A	G	
LTR (7)	23·5 (1·2)	20·1 (1·7)	33·8 (2·2)	22·7 (1·2)	35·1 (13·9)
gag (10)	23·7 (2·7)	16·4 (1·8)	35·2 (2·9)	24·7 (1·3)	16·5 (4·9)
pol (29)	30·4 (1·2)	17·9 (1·6)	35·5 (2·1)	16·2 (1·7)	16·3 (6·2)
env_L (20)	21·6 (0·9)	15·0 (1·1)	39·5 (1·2)	23·9 (1·2)	22·1 (6·0)
env_S (33)	22·2 (1·5)	16·4 (2·3)	38·6 (1·9)	22·8 (2·2)	24·8 (9·0)

* Significant deviation of the frequencies of the four bases from equality for all gene fragments (Pearson χ^2 $P < 0{\cdot}001$).

Table 3. ML parameter estimates for phylogenetic analysis

All parameters were estimated with Yang's PAML program package. Genes and corresponding *n* are as noted in Table 2.

Gene	Kappa (SE)/Mean*	Alpha (SE)†	C2 (SE)‡	C3 (SE)§
LTR	0·26 (0·18)/1·27	0·56 (0·20)	NA	NA
gag	4·25 (1·79)/5·80	0·22 (0·06)	0·07 (0·07)	14·27 (NA)
pol (SRLV)	3·53 (0·56)/5·35	0·42 (0·06)	0·32 (0·07)	5·04 (0·66)
pol (Lenti.)‖	1·51 (0·16)/NA	0·73 (0·08)	0·47 (0·06)	3·88 (0·39)
env_L	1·24 (0·15)/2·85	0·62 (0·07)	0·71 (0·09)	3·72 (0·37)
env_S	1·31 (0·25)/NA	0·70 (0·13)	0·79 (0·16)	3·22 (0·56)

* Transition/transversion rate ratio parameter of the substitution model F84 (Felsenstein's DNAML). Mean, kappa estimation derived from pairwise distances.
† Shape parameter for discrete gamma distribution (variable substitution rates across sites).
‡ Substitution rate ratio of second codon position to first codon position.
§ Substitution rate ratio of third codon position to first codon position.
‖ Including bovine, equine, feline, simian and human lentiviruses as outgroup ($n = 36$).

of the substitution rate across sites. With the hypothesis that the major reason for this variation is the well-known difference in the rates at different codon positions (e.g. Nei & Gojobori, 1986), the relative rates of codon positions 2 and 3 to codon position 1 were analysed. For all coding regions, the third codon position exhibited significantly higher substitution rates than the first codon position, with relative rates (C3) ranging from 3·2 to 14·7, whereas the second codon position showed consistently lower relative rates (C2) ranging from 0·07 to 0·79 (Table 3). This is consistent with an intermediate/low/high substitution rate model at codon positions. Interestingly, the largest bias toward third codon position changes was found in the *gag* fragments, with a 200-fold higher value for C3 than for C2. The C3/C2 bias was 15·8 for the *pol* fragments but only 5·2 or 4·1 for the env_L and env_S fragments. Changes at the third codon position being largely silent, this shows considerably less, though still effective negative selection for the *env* region. Overall, the good (negative) correlation between alpha and the C3/C2 bias adds support to the hypothesis that the codon position bias might be the major reason for the low alpha values found.

Synonymous and nonsynonymous nucleotide changes

To further substantiate the finding of higher bias towards silent third codon position changes in *gag* and *pol* than in *env* fragments, the ratios of synonymous to nonsynonymous substitution rates (S/N ratios) in these coding regions were

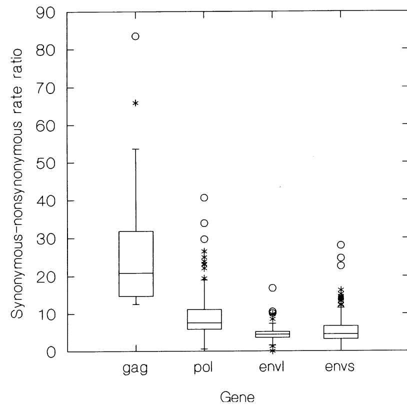

Fig. 1. Boxplots of the ratios of synonymous to nonsynonymous substitution rates for *gag*, *pol* and *env* fragments of SRLV derived from pairwise comparisons. The boxplots show the distribution of the ratios of the two substitution rates of different genes as described in Methods. Values within the 2. and 3. quartiles are boxed and medians are indicated. Whiskers (vertical bars) indicate values up to 1·5 box lengths beyond the box. *, Outliers; ○, extreme outliers; *gag*, group antigen fragments (n = 15); *pol*, polymerase gene fragments (n = 190); env_L, long envelope gene fragments (n = 190); and env_S, short envelope gene fragments (n = 465).

assessed. S/N ratios were calculated as the median and mean of the ratios found in the pairwise sequence comparisons. As expected, the median and mean S/N ratios of the *gag* fragments were significantly higher than those of the *pol* and *env* fragments. The respective median and mean (in parentheses) S/N ratios were 20·8 (29·1) for the *gag*, 7·5 (9·2) for the *pol*, 4·5 (4·6) for the env_L and 4·5 (5·2) for the env_S fragments. Boxplots showing the distribution of these ratios are given in Fig. 1. Overall differences and pairwise differences of the S/N ratios between genomic regions were highly significant with both nonparametric Kruskal–Wallis- and parametric one-way analysis of variance ($P < 0·001$). This high significance supports the finding of increased positive selection in the *env* region of SRLV as has also been shown for other lentiviruses (Mindell, 1996; Shpaer & Mullins, 1993; Li *et al*., 1988). Similarly, the average proportions of synonymous substitutions were less variable between the different genomic regions (0·57–0·72) than those of nonsynonymous substitutions, which were 0·04 in the *gag*, 0·08 in the *pol* and 0·13 in the *env* fragments (not shown; Shpaer & Mullins, 1993; Li *et al*., 1988).

Phylogeny reconstruction

For the phylogeny reconstruction of all coding regions by ML, the PAML estimates of the relative substitution rates were assigned to the codon positions as described in Methods. This resulted in significantly improved tree likelihoods, as assessed by the likelihood ratio test with two degrees of freedom ($G > 100$, $P < 0·001$; not shown). The best phylogenetic resolution was achieved in the *env* coding region, for which 20 (env_L) and 33 fragments (env_S) were analysed. The tree for the env_L fragments (569 bp in alignment) was in good agreement

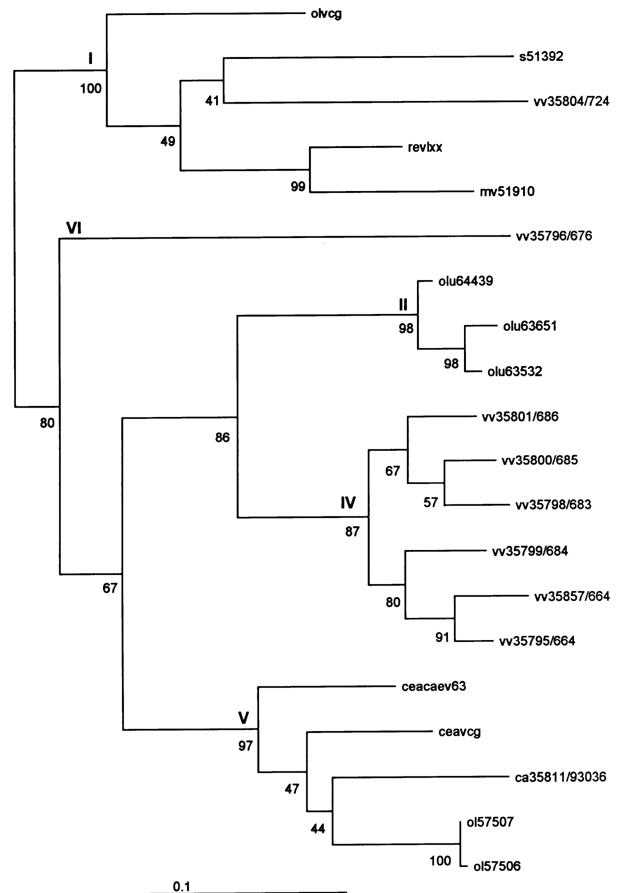

Fig. 2. Tree with long *env* fragments of SRLV generated with fastDNAml. This tree was generated with 20 *env* fragments with ten jumbles (randomization of input order of sequences) and local rearrangement allowed. The following three relative rates were assigned to the codon positions: 1, 0·71, 3·72. The estimate of kappa used was 1·24. The identification of the sequences is according to the EMBL database identification given in Table 1. For some sequences, the strain designation is shown in parenthesis for the ease of comparison with the tree generated with *pol* fragments. A tentative classification of sequence groups into clades compatible with that found in *pol* fragments is printed at the respective nodes in roman numerals. Bootstrap values derived from 100 bootstrap replicates indicating the confidence of the branching pattern are printed above the respective branches. The bar indicates the amount of evolution along the horizontal branches in substitutions per site. Vertical lines are displayed for clarity only. This tree, with a ln likelihood (lnL) of −4858·91, was the one with the most likely topology out of 698 trees examined.

with that for the env_S fragments (199 bp in alignment). The differences were mainly attributable to the inclusion of further French visna virus isolates and of closely related quasispecies representatives within French isolates in the env_S data set. Except for one isolate (664, vv35857 and vv35795), the quasispecies representatives formed monophyletic groups (not shown). The sequences clustered into four major lineages with high bootstrap support (shown for env_L fragments in Fig. 2). The first lineage (tentatively named clade I) had a quite heterogeneous composition, containing the South African ovine lentivirus (olvcg), the Edinburgh MVV EV1 (s51392),

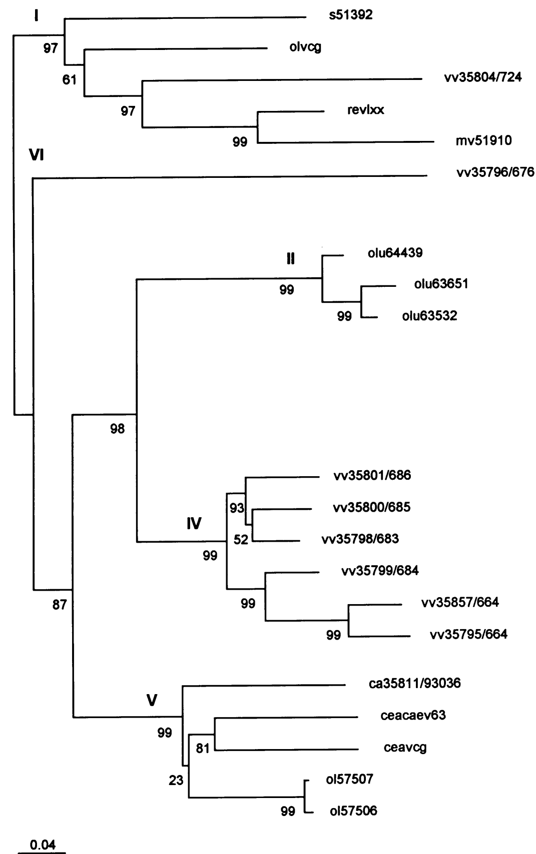

Fig. 3. Minimum evolution tree for long *env* fragments of SRLV generated with METREE. This tree, based on gamma-Poisson distances, was generated with the same *env* fragments used for Fig. 2. It was actually identical to the NJ tree. The estimate of the gamma shape parameter alpha used was 0·62. The bootstrap values of 500 bootstrap replicates are indicated.

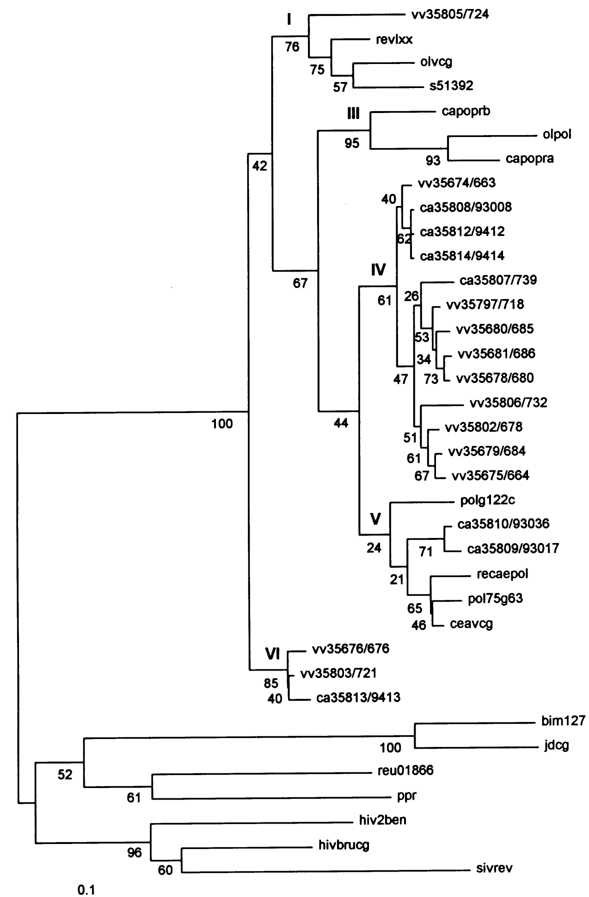

Fig. 4. Tree with *pol* fragments of lentiviruses generated with fastDNAml. This tree was generated with 36 *pol* fragments of small ruminant, bovine, equine, feline, human and simian lentiviruses with ten jumbles and local rearrangement allowed. The relative rates assigned to the codon positions were as follows: 1, 0·47, 3·88. The estimate of kappa used was 1·51. Identification of the sequences is as given in Table 1. For some sequences, the strain designation is shown in parenthesis for the ease of comparison with the tree generated with *env* fragments. A tentative classification of sequence groups into clades compatible with that found in *env* fragments is printed at the respective nodes in roman numerals. Bootstrap values derived from 100 bootstrap replicates are indicated. The bar indicates the amount of evolution along the horizontal branches in substitutions per site. This tree, with a ln likelihood (lnL) of −4279·86, was the one with the most likely topology out of 2892 trees examined.

the French visna virus isolate 724 (vv35804), the Icelandic visna lentivirus (revlxx) and the Icelandic MVV strain KM 1071 (mv51910). The second lineage (tentatively named clade II) included the North American ovine lentivirus strains OLV 85/34 and 84/28 (olu64439, olu63651 and olu63532). The third lineage (tentatively named clade IV) contained several French visna virus isolates. This lineage was resolved by both ML and minimum evolution into two subgroups. The fourth lineage (tentatively named clade V) included the North American caprine lentiviruses 75-G63 (ceacaev63) and CAEV strain CO (ceavcg), the French CAEV isolate 93036 (ca35811) and the North American ovine lentivirus strain Sh8 (A and B, ol57506 and ol57507). The French visna isolate 676 (vv35796) was either a representative of a fifth lineage (tentatively named clade VI, env_L) or was a clearly distinct member of the heterogeneous lineage 1 (env_S, not shown). The minimum evolution tree, estimated with an alpha of 0·62, was actually the NJ tree (Fig. 3). This tree was very similar to the ML tree. Bootstrap supports for the branching order were in general higher than those in ML. Minor differences in the order of strains were seen in lineages 1 and 4. Interestingly, minimum evolution grouped the sequence fragments of the two prototype North American CAEV strains 75-G63 and CO apart from the other strains of lineage 4 (bootstrap support 81%). This seems more reasonable since the other strains were from a French goat and a North American sheep.

Since lentivirus genomes do not exhibit sufficient sequence similarity for overall alignment except for the *pol* gene, only the *pol* fragments of SRLV, which are located in the core region of the RT, could be analysed together with those of other lentiviruses. As expected from the higher proportions of divergence, phylogeny reconstruction clustered the bovine, equine, feline, human and simian lentiviruses clearly apart from

small ruminant lentiviruses (Fig. 4). In contrast to minimum evolution (not shown), the ML algorithm did not group the HIV-2 fragment correctly as more closely related to SIV than to HIV-1. As far as SRLV sequences are concerned, at least five major lineages could be differentiated. The first lineage (tentatively named clade I in accordance with the tree generated with *env* fragments) contained the same ovine lentiviruses as seen for *env*, but in a different order. The second lineage (clade III) consisted of the Norwegian CAEV isolates LM30 (capoprb) and 8062 (capopra) and the OLV isolate N1 (olpol). The third lineage (clade IV) consisted of three subgroups with several French caprine and ovine lentivirus isolates. The first subgroup contained caprine and ovine isolates, which were not represented in the *env* data set. The other two subgroups were in good agreement with those seen in lineage three for *env*, as far as the isolates were represented in both data sets (isolates 685, 686, 684 and 664). A striking interrelationship between caprine and ovine isolates was most evident in the subgroups 1 and 2. The fourth lineage (clade V) was composed exclusively of sequences of caprine isolates of Swiss, French and North American origin. Also in this lineage there was good agreement between the trees generated with *env* or *pol* fragments. Lineage five (clade VI) corresponded to the putative lineage five in the *env* sequences with visna virus isolate 676 (vv35676) represented in both trees. The overall lower bootstrap support of the branching order is probably due to the lower extent of variation found in the *pol* sequences (see above). Still, there was reasonable support for clades III (95%), VI (85%) and I (76%). As for the *env* sequences, the minimum evolution tree estimated with an alpha of 0·73 coincided with the NJ tree and was very similar to that found with ML. Importantly, the bootstrap support for all clades was excellent or good with 99% for clade III, 99% for clade IV, 94% for clade V, 99% for clade VI and 82% for clade I (not shown).

The tree generated with the *gag* fragments confirmed the distant relationship of the Norwegian CAEV isolates LM30 (cagaprb) and 8062 (cagapra) to other SRLV, whereas the Norwegian visna virus isolate NMV1 (vvgapr) clustered closely with Icelandic visna lentivirus (revlxx). Other isolates were not resolved into bootstrap-supported, distinguishable clusters (not shown). The tree generated with seven LTR fragments exhibited only two lineages, one corresponding to the first lineage described above and one to the CAEV strain CO and CAEV integrated (ceavcg, cealtr; not shown).

Molecular clock

So far, all trees were estimated in the absence of a molecular clock assumption. In order to test the validity of a molecular clock assumption for the SRLV sequences analysed including both synonymous and nonsynonymous sites, the same trees were also estimated with the clock assumption (molecular clock option of BASEML) and the likelihoods were compared. As assessed by the likelihood ratio test, the likelihoods for all trees were significantly lower (LTR: $df = 6$, $P < 0{\cdot}0001$; *gag*: $df = 9$, $P = 0{\cdot}00023$; *pol*: $df = 35$, $P < 0{\cdot}0001$; env_L: $df = 22$, $P = 0{\cdot}006$; env_S: $df = 32$, $P < 0{\cdot}0001$) with the molecular clock assumption than in its absence. This indicates, that tree reconstruction with the molecular clock assumption might not be well suited for SRLV.

Discussion

In this study, a complex relationship between different SRLV from different parts of the world was found. At least six different clades could be identified, with no clear separation of SRLV strains from goats or sheep. The first clade consisted of the prototype Icelandic visna virus and related MVV strains, the second of North American ovine lentivirus strains, the third of Norwegian SRLV, the fourth of French SRLV, the fifth of North American prototype-, French- and Swiss CAEV strains and North American OLV strains and the sixth of French SRLV.

The overall divergence between strains was highest in the LTR fragments (35%), intermediate in the *env* fragments (22%) and lowest in the *pol* and *gag* fragments (16%). This level of divergence is comparable to that found in other lentiviruses like HIV (Li *et al*., 1988), SIV (Hirsch *et al*., 1995*a*) and FIV (Bachmann *et al*., 1997). But taking into consideration that SRLV might have a much longer history of evolution than HIV (Quérat *et al*., 1990), this may still indicate slower evolution of SRLV. However, both intrahost divergence (0·6% to 18·6%) and recently published data from an experimental infection with a molecular clone of CAEV seem to corroborate the notion of comparable speed of sequence evolution (Turelli *et al*., 1997), which has yet to be confirmed.

For all coding sequence fragments analysed, relatively high codon position biases were found. As expected from the codon position bias, the ratio of silent to replacement mutations was highest in the *gag* (20·8) and lowest in the *env* fragments (4·5), indicating decreasing contribution of negative and increasing contribution of positive, probably immune-driven, Darwinian selection from *gag* to *env*. The high value found for *gag* is probably an overestimation that is due to the small size of the fragment analysed. For HIV, lower ratios have been found for *gag* (5·5) and *env* (2·8) in an analysis using sequences from four defined virus subtypes (Mindell, 1996). In contrast, using many independent HIV and SIV isolates, Shpaer & Mullins (1993) found inverted ratios with higher levels of replacement mutations and correlation with virus pathogenicity in *env* sequences of human and simian immunodeficiency viruses. As mentioned for the overall divergence of SRLV, this also might be taken as an indication of slower evolution of these viruses due to either lower selection pressure (Sonigo, 1994; Shpaer & Mullins, 1993) or a lower replication level *in vivo*, although different population sizes have not been considered in these comparisons.

The relatively high variation of the substitution rate across

sites (low values of the gamma parameter alpha), which was probably due to the high codon position bias in coding genes, clearly indicates the importance of a prior analysis of aligned sequences before the application of phylogenetic analysis. This is particularly true for simple and fast algorithms like NJ, where different models of sequence evolution (distances) must be chosen according to estimates for parameters like average sequence divergence, gamma and kappa derived from actual data. Under these conditions, both ML and distance methods like NJ were shown to accurately reconstruct the phylogeny in an HIV-1 transmission cluster with exactly known phylogenetic relationships (Leitner *et al*., 1996).

The highest level of phylogenetic information for SRLV seems to be accumulated in the LTR sequences rather than in the *env* gene. However, due to the modular structure of retroviral LTRs, with a constant arrangement of possibly repeated functional elements, rather than conserved overall sequence, alignment and quantification of divergence of LTR sequences cannot be achieved properly with classical multiple alignment (Sutton *et al*., 1997; Frech *et al*., 1996). Therefore, the LTR fragments analysed in this study, which are located in the U3 and R regions, containing the consensus sequence for AP1, AML and AP4 binding factors, are not particularly suited for phylogeny reconstruction. Still, the MVV and CAEV sequences represented in this data set were clearly separated in two clusters, which could also easily be seen by visual inspection of the alignment (not shown).

The phylogeny reconstructed with the *env* and *pol* fragments is remarkable in several aspects. First, it has to be stressed that the trees were in very good agreement where sequences for the same isolates were available for both genes. This confirms that a reasonable phylogeny for SRLV could be reconstructed with the algorithms used. Second, the lack of resolution of clades III to VI according to host-species origin establishes an epidemiological concept, that SRLV may cross the species barrier with ease. This notion renders a simple ecotype classification of current sheep- and goat-derived lentivirus isolates an unsuitable model. However, based on the midpoint-rooted trees presented, it is difficult to establish a clear epidemiological history of SRLV. The high degree of evolution of geographically well-separated strains seen in the heterogeneous clade I, its exclusive composition of MVV strains and its position in the trees derived from both *env* and *pol* sequences can be taken as an indication of its ancestorship in the SRLV phylogeny. Based on this hypothesis, all other clades, including those with the prototype CAEV strains would be descendants of an ovine lentivirus. This hypothesis also finds support in recently published work (Mwaengo *et al*., 1997; Valas *et al*., 1997). On the other hand, given the close relatedness between the mixed lentiviruses of clade IV and the prototype caprine lentiviruses of clade V, it is also conceivable that goat-adapted strains could be passed between goats and sheep. Third, the apparent lack of a molecular clock in the fragments analysed does not favour the calculation of the evolutionary time frame involved, at least when both synonymous and nonsynonymous sites are included in the analysis.

In conclusion, a quite robust phylogeny of SRLV could be reconstructed. The clear indication for cross-species transmission found is epidemiologically interesting, since it is in contrast to published findings made under field conditions (McKenzie, 1991; Smith *et al*., 1985). The absence of cross-species transfer in these studies might rather be due to well-separated management of the two species than to a low risk of transmission upon close contact. This finding is a strong warning signal for eradication programmes in both species. The application of molecular epidemiological tools to sequence data arising from experimental and field work will make it possible to get a further insight into the puzzling evolution and transmission of SRLV.

This work was made possible by a generous granting of a 9 month sabbatical to R. Z. by the University Bern. The free use of all infrastructure, the high performance computing farm and the libraries of the EBI in the frame of EBI's visitor programme, which is conducted by EBI's scientific committee headed by Dr M. Ashburner, is highly appreciated. Furthermore, the kind help of the systems group headed by D. Harris for connecting a PC to the network, the kind help of Dr M. Wise in all compilations of phylogenetic software running under UNIX and useful discussions of the manuscript with Drs M. Ashburner, H. Bourhy, M. Wise, E. Peterhans and G. Bertoni are acknowledged.

References

Bachmann, M. H., Mathiason-Dubard, C., Learn, G. H., Rodrigo, A. G., Sodora, D. L., Mazzetti, P., Hoover, E. A. & Mullins, J. I. (1997). Genetic diversity of feline immunodeficiency virus: dual infection, recombination, and distinct evolutionary rates among envelope sequence clades. *Journal of Virology* **71**, 4241–4253.

Banks, K. L., Adams, D. S., McGuire, T. C. & Carlson, J. (1983). Experimental infection of sheep by caprine arthritis–encephalitis virus and goats by progressive pneumonia virus. *American Journal of Veterinary Research* **44**, 2307–2311.

Campbell, B. J. & Avery, R. J. (1996). Sequence analysis and transcriptional activity of the LTR of OLV-CU1, a North American ovine lentivirus. *Journal of General Virology* **77**, 2999–3004.

CDC (1995). Human rabies possibly of bat origin – Washington, USA, 1995. *Rabies Bulletin Europe* **19**, 10–11.

Chadwick, B. J., Coelen, R. J., Wilcox, G. E., Sammels, L. M. & Kertayadnya, G. (1995). Nucleotide sequence analysis of Jembrana disease virus: a bovine lentivirus associated with an acute disease syndrome. *Journal of General Virology* **76**, 1637–1650.

Chang, J. T. (1996). Inconsistency of evolutionary tree topology reconstruction methods when substitution rates vary across characters. *Mathematical Biosciences* **134**, 189–215.

Chebloune, Y., Karr, B., Sheffer, D., Leung, K. & Narayan, O. (1996). Variations in lentiviral gene expression in monocyte-derived macrophages from naturally infected sheep. *Journal of General Virology* **77**, 2037–2051.

Chiu, I.-M., Yaniv, A., Dahlberg, J. E., Gazit, A., Skuntz, S. F., Tronick, S. R. & Aaronson, S. A. (1985). Nucleotide sequence evidence for relationship of AIDS retrovirus to lentiviruses. *Nature* **317**, 366–368.

Etzold, T., Ulyanov, A. & Argos, P. (1996). SRS: information retrieval system for molecular biology data banks. *Methods in Enzymology* **266**, 114.

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution* **17**, 368–376.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791.

Felsenstein, J. & Churchill, G. A. (1996). A hidden Markov Model approach to variation among sites in rate of evolution. *Molecular Biology and Evolution* **13**, 93–104.

Frech, K., Brack-Werner, R. & Werner, T. (1996). Common modular structure of lentivirus LTRs. *Virology* **224**, 256–267.

Garvey, K. J., Oberste, M. S., Elser, J. E., Braun, M. J. & Gonda, M. A. (1990). Nucleotide sequence and genome organization of biologically active proviruses of the bovine immunodeficiency-like virus. *Virology* **175**, 391–409.

GCG (1991). *Program Manual for the GCG Package, version 7, April* 1991, 7th edn, 575 Science Drive, Madison, Wisconsin, USA 53711: Genetics Computer Group.

Hirsch, V., Dapolito, G., Goeken, R. & Campbell, B. J. (1995*a*). Phylogeny and natural history of the primate lentiviruses, SIV and HIV. *Current Opinion in Genetics* & *Development* **5**, 798–806.

Hirsch, V. M., Dapolito, G., Johnson, P. R., Elkins, W. R., Montali, R. J., Goldstein, S. & Brown, C. (1995*b*). Induction of AIDS by simian immunodeficiency virus from an African green monkey: species-specific variation in pathogenicity correlates with the extent of in vivo replication. *Journal of Virology* **69**, 955–967.

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.

Kirchhoff, F., Jentsch, K., Bachmann, B., Stuke, A., Laloux, C., Lueke, W., Stahl-Henning, C., Schneider, J., Nieselt, K., Eigen, M. & Hunsmann, G. (1990). A novel proviral clone of HIV-2: biological and phylogenetic relationship to other primate immunodeficiency viruses. *Virology* **177**, 305–311.

Knowles, D. P., 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. *Journal of Virology* **65**, 5744–5750.

Kumar, S., Tamura, K. & Nei, M. (1993). *MEGA*: *Molecular Evolutionary Genetics Analysis, version* 1.01. The Pennsylvania State University, University Park, PA 16802, USA.

Leitner, T., Escanilla, D., Franzen, C., Uhlen, M. & Albert, J. (1996). Accurate reconstruction of a known HIV-1 transmission history by phylogenetic tree analysis. *Proceedings of the National Academy of Sciences, USA* **93**, 10864–10869.

Leroux, C., Vuillermoz, S., Mornex, J.-F. & Greenland, T. (1995). Genomic heterogeneity in the *pol* region of ovine lentiviruses obtained from bronchoalveolar cells of infected sheep from France. *Journal of General Virology* **76**, 1533–1537.

Li, W.-H., Tanimura, M. & Sharp, P. M. (1988). Rates and dates of divergence between AIDS virus nucleotide sequences. *Molecular Biology and Evolution* **5**, 313–330.

McKenzie, J. (1991). Survey for caprine arthritis encephalitis antibodies in sheep. *Surveillance* **18**, 19–20.

Mindell, D. P. (1996). Positive selection and rates of evolution in immunodeficiency viruses from humans and chimpanzees. *Proceedings of the National Academy of Sciences, USA* **93**, 3284–3288.

Mwaengo, D. M., Grant, R. F., DeMartini, J. C. & Carlson, J. O. (1997). Envelope glycoprotein nucleotide sequence and genetic characterization of North American ovine lentiviruses. *Virology* **238**, 135–144.

Narayan, O. & Cork, L. C. (1985). Lentiviral diseases of sheep and goats: chronic pneumonia, leukoencephalomyelitis and arthritis. *Reviews of Infectious Diseases* **7**, 89–98.

Nei, M. & Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution* **3**, 418–426.

Olsen, G. J., Matsuda, H., Hagstrom, R. & Overbeek, R. (1994). FastDNAml: a tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. *Computer Applications in the Biosciences* **10**, 41–48.

Ou, C.-Y., Ciesielski, C. A., Myers, G., Bandea, C. I., Luo, C.-C., Korber, B. T. M., Mullins, J. I., Schochetman, G., Berkelman, R. L., Economou, A. N., Witte, J. J., Furman, L. J., Satten, G. A., MacInnes, K. A., Curran, J. W. & Jaffe, H. W. (1992). Molecular epidemiology of HIV transmission in a dental practice. *Science* **256**, 1165–1171.

Payne, S. L., Rausch, J., Rushlow, K., Montelaro, R. C., Issel, C., Flaherty, M., Perry, S., Sellon, D. & Fuller, F. (1994). Characterization of infectious molecular clones of equine infectious anaemia virus. *Journal of General Virology* **75**, 425–429.

Phillips, T. R., Talbott, R. L., Lamont, C., Muir, S., Lovelace, K. M. & Elder, J. H. (1990). Comparison of two host cell range variants of feline immunodeficiency virus. *Journal of Virology* **64**, 4605–4613.

Quérat, 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.

Rzhetsky, A. & Nei, M. (1993). Theoretical foundation of the minimum-evolution method of phylogenetic inference. *Molecular Biology and Evolution* **10**, 1073–1095.

Rzhetsky, A. & Nei, M. (1994). METREE: a program package for inferring and testing minimum-evolution trees. *Computer Applications in the Biosciences* **10**, 409–412.

Sacramento, D., Badrane, H., Bourhy, H. & Tordo, N. (1992). Molecular epidemiology of rabies virus in France: comparison with vaccine strains. *Journal of General Virology* **73**, 1149–1158.

Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**, 406–425.

Saltarelli, M., Quérat, 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.

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. *Journal of General Virology* **72**, 1893–1903.

Scholtissek, C. (1995). Molecular evolution of influenza viruses. *Virus Genes* **11**, 209–215.

Sherman, L., Gazit, A., Yaniv, A., Dahlberg, J. E. & Tronick, S. R. (1986). Nucleotide sequence analysis of the long terminal repeat of integrated caprine arthritis encephalitis virus. *Virus Research* **5**, 145–155.

Shpaer, E. G. & Mullins, J. I. (1993). Rates of amino acid change in the envelope protein correlate with pathogenicity of primate lentiviruses. *Journal of Molecular Evolution* **37**, 57–65.

Smith, J. S. (1989). Rabies virus epitopic variation: use in ecologic studies. *Advances in Virus Research* **36**, 215–254.

Smith, V. W., Dickson, J., Coackley, W. & Carman, H. (1985). Response of merino sheep to inoculation with a caprine retrovirus. *Veterinary Record* **117**, 61–63.

Smith, J. S., Fishbein, D. B., Rupprecht, C. E. & Clark, K. (1991). Unexplained rabies in three immigrants in the United States – a virologic investigation. *New England Journal of Medicine* **324**, 205–211.

Smith, J. S., Orciari, L. A., Yager, P. A., Seidel, H. D. & Warner, C. K. (1992). Epidemiologic and historical relationships among 87 rabies virus isolates as determined by limited sequence analysis. *Journal of Infectious Diseases* **166**, 296–307.

Smith, J. S., Orciari, L. A. & Yager, P. A. (1995). Molecular epidemiology of rabies in the United States. *Seminars in Virology* **6**, 387–400.

Sokal, R. R. & Rohlf, F. J. (1995). *Biometry: the principles and practice of statistics in biological research*, 3rd edn. New York: W. H. Freeman.

Sonigo, P. (1994). Variation and virulence of lentiviruses. *Biofutur* **135**, 8–10.

Sonigo, P., Alizon, M., Staskus, K., Klatzmann, D., Cole, S., Danos, O., Retzel, E., Tiollais, P., Haase, A. & Wain-Hobson, S. (1985). Nucleotide sequence of the visna lentivirus: relationship to the AIDS virus. *Cell* **42**, 369–382.

Steinhauer, D. S. & Holland, J. J. (1986). Direct method for quantitation of extreme polymerase error frequencies at selected single base sites in viral RNA. *Journal of Virology* **57**, 219–228.

Sutton, K. A., Lin, C. T., Harkiss, G. D., McConnell, I. & Sargan, D. R. (1997). Regulation of the long terminal repeat in visna virus by a transcription factor related to the AML/PEBP2/CBF superfamily. *Virology* **229**, 240–250.

Tamura, K. & Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution* **10**, 512–526.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673–4680.

Turelli, P., Guiguen, F., Mornex, J. F., Vigne, R. & Quérat, G. (1997). dUTPase-minus caprine arthritis–encephalitis virus is attenuated for pathogenesis and accumulates G-to-A substitutions. *Journal of Virology* **71**, 4522–4530.

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.

Vanhemert, F. J. & Berkhout, B. (1995). The tendency of lentiviral open reading frames to become A-rich: constraints imposed by viral genome organization and cellular tRNA availability. *Journal of Molecular Evolution* **41**, 132–140.

Wain-Hobson, S., Sonigo, P., Danos, O., Cole, S. & Alizon, M. (1985). Nucleotide sequence of the AIDS virus, LAV. *Cell* **40**, 9–17.

Wilkinson, L. (1990). *SYSTAT: the System for Statistics.* Evanston, IL, USA: SYSTAT Inc.

Woodward, T. M., Carlson, J. O., Delaconcha Bermejillo, A. & DeMartini, J. C. (1995). Biological and genetic changes in ovine lentivirus strains following passage in isogeneic twin lambs. *Journal of Acquired Immune Deficiency Syndromes and Human Retroviruses* **8**, 124–133.

Yang, Z. H. (1995). Evaluation of several methods for estimating phylogenetic trees when substitution rates differ over nucleotide sites. *Journal of Molecular Evolution* **40**, 689–697.

Yang, Z. H. (1996*a*). Among-site rate variation and its impact on phylogenetic analyses. *Trends in Ecology & Evolution* **11**, 367–372.

Yang, Z. (1996*b*). In *Phylogenetic Analysis by Maximum Likelihood* (*PAML*), version 1.2, pp. 1–32. Department of Integrative Biology, University of California, Berkeley, CA, USA.

Yang, Z., Goldman, N. & Friday, A. (1995). Maximum likelihood trees from DNA sequences: a peculiar statistical estimation problem. *Systematic Biology* **44**, 384–399.

Zanoni, R., Krieg, A. & Peterhans, E. (1989). Detection of antibodies to caprine arthritis–encephalitis virus by protein G enzyme-linked immunosorbent assay and immunoblotting. *Journal of Clinical Microbiology* **27**, 580–582.

Zanoni, R., Regli, J. & Peterhans, E. (1991*a*). Diagnostik und Bekämpfung der caprinen Arthritis–Encephalitis-Virus-Infektion. *Mitteilungen des Bundesamtes für Veterinärwesen* **8**, 60–61.

Zanoni, R. G., Nauta, I. M., Pauli, U. & Peterhans, E. (1991*b*). Expression in Escherichia coli and sequencing of the coding region for the capsid protein of Dutch maedi–visna virus strain ZZV 1050: application of recombinant protein in enzyme-linked immunosorbent assay for the detection of caprine and ovine lentiviruses. *Journal of Clinical Microbiology* **29**, 1290–1294.

Zanoni, R. G., Nauta, I. M., Kuhnert, P., Pauli, U., Pohl, B. & Peterhans, E. (1992). Genomic heterogeneity of small ruminant lentiviruses detected by PCR. *Veterinary Microbiology* **33**, 341–351.

Received 26 January 1998; Accepted 16 April 1998