Published online ahead of print on 25 April 2006 as DOI 10.1099/vir.0.81966-0

Genomic characterization of a novel poxvirus contributing to the decline of the red squirrel (*Sciurus vulgaris*) in the UK

Colin J. McInnes,[1] Ann R. Wood,[1] Kathryn Thomas,[1] Anthony W. Sainsbury,[2] John Gurnell,[3] F. Joshua Dein[4] and Peter F. Nettleton[1]

[1]Moredun Research Institute, Pentlands Science Park, Bush Loan, Penicuik EH26 0PZ, UK

[2]Institute of Zoology, London, UK

[3]Queen Mary, University of London, UK

[4]National Wildlife Health Center, Madison, WI, USA

Correspondence
Colin J. McInnes
mcinc@mri.sari.ac.uk

The GenBank/EMBL/DDBJ accession numbers for the sequences reported in this paper are DQ377804 and DQ377805.

© 2006 SGM

The genome of a virulent squirrelpox virus (SQPV) isolate was characterized in order to determine its relationship with other poxviruses. Restriction enzyme analysis suggested a genome length of approximately 158 kb, whilst sequence analysis of the two ends of the genome indicated a G+C composition of approximately 66 %. Two contiguous stretches of 23 and 37 kb at the left-hand and right-hand ends of the genome, respectively, were sequenced allowing the identification of at least 59 genes contained therein. The partial sequence of a further 15 genes was determined by spot sequencing of restriction fragments located across the genome. Phylogenetic analysis of 15 genes conserved in all the recognized genera of the subfamily *Chordopoxvirinae* confirmed that the SQPV does not group within the family *Parapoxvirinae*, but instead partitions on its own in a separate clade of the poxviruses. Analysis of serum from British woodland rodents failed to find any evidence of SQPV infection in wood mice or bank voles, but for the first time serum samples from grey squirrels in the USA were found to contain antibody against SQPV.

INTRODUCTION

The UK red squirrel (*Sciurus vulgaris*) has disappeared from much of mainland Britain with its geographical range now restricted mainly to the north of England and Scotland (Gurnell & Pepper, 1993). The decline of the red squirrel has been attributed to the introduction of the North American grey squirrel (*Sciurus carolinensis*) because of its ability to successfully out-compete the red species for habitat and food. However, more recently a lethal epidemic disease that is characterized by severe ulceration and formation of haemorrhagic scabs around the eyes, nose and mouth was suggested as an additional reason for the local extinction of red squirrel populations (Sainsbury & Gurnell, 1995; Sainsbury *et al.*, 2000). The putative cause of the disease was identified as a poxvirus and based on its morphological similarity with viruses belonging to the family *Parapoxvirinae*, it was classified and named *Squirrel parapoxvirus* (SPPV) (Scott *et al.*, 1981; Buller *et al.*, 2005). Recently, we reported on the phylogenetic analysis of the genes encoding the viral major membrane protein and the 30 kDa subunit of the viral RNA polymerase. This analysis did not support the classification of the virus as a parapoxvirus, but instead suggested that it most likely represents a previously unidentified genus within the poxvirus family. As a consequence, we suggested that the name of the virus should be changed from SPPV to squirrelpox virus (SQPV; Thomas *et al.*, 2003).

Experimental infections of squirrels with SQPV confirmed that it is the agent responsible for the disease seen in wild red squirrels and that conversely, the grey squirrel, although infected by the virus, is apparently clinically resistant to the disease (Tompkins *et al.*, 2002; Thomas *et al.*, 2003). A serological survey of squirrels sampled from across the UK demonstrated that 61 % of grey squirrels had antibodies to SQPV in contrast to only 2.9 % of red squirrels, the overwhelming majority of which were found dead or dying of the disease (Sainsbury *et al.*, 2000). The high seroprevalence of antibodies to the virus, but absence of disease, in the grey squirrel suggests that they may act as a reservoir host for the virus, transmitting it to susceptible red squirrels with lethal consequences, a theory that is supported by transmission studies and mathematical modelling of the dynamics of local squirrel populations (Tompkins *et al.*, 2003; Rushton *et al.*, 2005). Outbreaks of pox-like disease in red squirrels were not reported until after the introduction of the grey squirrel from America and it is thought that the virus was introduced to Britain via this route, although the virus has never been described in the USA. There is, however, the possibility that the virus is endemic to the UK and that other rodent species inhabiting the same woodland environment could be harbouring the virus.

The poxvirus family consists of two subfamilies, the *Chordopoxvirinae* and the *Entomopoxvirinae*. The subfamily *Chordopoxvirinae* can be further divided into eight genera based mainly on the animal species they infect, but also on virion morphology, antigenic cross-reactivity, genome size and gene content. Their genomes are linear double-stranded DNA molecules ranging in size from approximately 135 to over 300 kb. Within them 89 genes from the central 'core' of the genomes are conserved across all sequenced species, whereas the 'termini' vary considerably in size and gene content between the genera (Upton *et al.*, 2003;

Gubser *et al.*, 2004; Delhon *et al.*, 2004). Successful vaccination against a particular poxvirus is usually only possible using the same, but generally attenuated, virus or one from the same genus. Cross-genera vaccination is not normally successful. Transmission of the viruses also differs between genera, being spread either by aerosol, contaminated fomites, direct contact or arthropods. It was important for us to determine the classification of SQPV, because many of the assumptions previously made about the transmission and survivability of the virus and its potential for vaccine development, were based on its original classification as a *Parapoxvirus*. Here, we present a genetic map of SQPV and approximately 23 and 37 kb of contiguous sequence from the left-hand (LH) and right-hand (RH) termini, respectively. The gene content and genomic organization of SQPV is compared with *Orf virus* (ORFV), the prototypic parapoxvirus and other poxviruses. Phylogenetic analysis of 15 genes conserved across all the recognized genera of the subfamily *Chordopoxvirinae* is used to infer classification of the virus, whilst further serological studies failed to find evidence of the virus in other British woodland rodents, but provided preliminary evidence that the virus is present in grey squirrels in the USA.

METHODS

Virus. The SQPV isolate used in this study (1296/99) was originally taken from an individual red squirrel found dead, with typical SQPV lesions, during an epidemic of pox-like disease in northeast England in 1999 (Thomas *et al.*, 2003).

Isolation and cloning of viral genomic DNA. Virus DNA was isolated from scabs collected post-mortem. These were homogenized in PBS and centrifuged at 2000 ***g*** for 30 min at 4 °C. The viral cores were purified from the resulting supernatant by centrifugation through a sucrose cushion [36 % (w/w) in PBS] at 71000 ***g*** for 30 min at 4 °C. The SQPV genomic DNA was subsequently extracted and purified from these cores (Gilray *et al.*, 1998). The purified DNA was digested with the restriction endonucleases *Not*I, *Kpn*I and *Bam*HI (Roche Diagnostics) and the resulting fragments cloned into the pBluescript SK^{-} plasmid vector (Stratagene). In addition, viral genomic DNA was partially digested with *Sau*3AI (Roche Diagnostics) and the resulting fragments, 30–40 kb in size, were cloned into the SuperCos I cosmid vector (Stratagene).

Restriction endonuclease mapping of SQPV genomic DNA cloned into SuperCos I. Restriction endonuclease maps of cosmid clones were determined using a standard procedure (Rackwitz *et al.*, 1985). Briefly, cosmid clones were linearized by digestion with λ-terminase (Amersham Pharmacia Biotech). The linearized cosmid clones were partially digested with the desired restriction enzyme and labelled at one end by hybridization in solution with one or other of the ^{32}P-labelled oligonucleotides complementary to the 12 bp *cos* site. The labelled reactions were electrophoresed on a 0.5 % (w/v) agarose gel, the gel was dried onto DE-81 chromotography paper and exposed to X-ray film. The restriction maps were determined from the fragment sizes of the partially digested, end-labelled cosmid DNA.

Sequence analysis. The EZ::TN<KAN-2> Insertion kit (Epicentre) was used to randomly introduce a kanamycin resistance marker and primer binding sites into plasmid clones containing SQPV fragments. Equimolar amounts of target DNA and <KAN-2> transposon were mixed *in vitro* in transposition insertion reactions. An aliquot of each reaction mixture was used to transform One Shot TOP10 *Escherichia coli* (Invitrogen) and transformants were selected on agar plates containing 50 μg kanamycin ml<MIN>1. Double-stranded DNA templates from a selection of transformants for each of the SQPV clones were prepared and sequenced using KAN-2 FP-1 and KAN-2 RP-1 primers that correspond to the primer binding sites found on the transposon. Sequencing reactions were run on a GE Healthcare MegaBACE 500 capillary DNA sequencer using 'DYEnamic' ET Terminator chemistry (Amersham Pharmacia). DNA sequences were compared to those deposited in the GenBank/EMBL database using the FastA 3 search algorithm (Pearson & Lipman, 1988). In addition, DNA sequences were translated into all six open reading frames (ORFs) using the DNASTAR package (DNASTAR) and the predicted amino acid sequences compared with those deposited in protein databases. Further sequence

analyses were performed using the VOCS website (Ehlers *et al.*, 2002) and the ARTEMIS sequence analysis package (Rutherford *et al.*, 2000).

DNA hybridization. Double-stranded DNA probes were labelled with digoxigenin (DIG) using the nick-translation labelling kit (Roche Diagnostics). Digested viral genomic DNA was transferred from agarose gels and immobilized onto Hybond-N membrane (Amersham Pharmacia Biotech). Alternatively, 0.5 μg heat-denatured cloned DNAs were applied to Hybond-N membranes using a Hybri-dot vacuum manifold. Membranes were hybridized with the DIG-labelled probes at 65 °C for approximately 16 h, before washing. Hybridization was detected with anti-DIG antibody conjugated to alkaline phosphatase using the manufacturer's recommended procedures.

Direct ELISA. An adaptation of the ELISA described by Sainsbury *et al.* (2000) was used to detect anti-SQPV IgG in the sera from both red and grey squirrels and from wood mice and bank voles. Briefly, duplicate wells in 96-well microplates were coated overnight at 4 °C with 50 μl SQPV, and negative control, antigens. After removal of unbound antigen, test and control sera were applied, in duplicate, to wells coated with positive- and negative-antigen and incubated at 37 °C for 1 h. After washing, 100 μl protein G-horseradish peroxidase conjugate diluted 1/750 in ELISA dilution buffer [1× PBS with 1 % (w/v) BSA (Sigma)], was applied to each well and incubated at 37 °C for 1 h. After further washing, 100 μl freshly prepared orthophenylene diamine (Sigma) substrate was added to each well. The colorimetric reaction was allowed to proceed for approximately 8 min before the OD_{492} of each well was determined and the corrected optical densities for each test and control serum calculated by subtracting the mean optical density of the negative-antigen wells. An OD_{492} value of 0.2 was used as the cut-off value to discriminate between positive- and negative-antibody detection.

RESULTS

Restriction endonuclease mapping of SQPV

Restriction fragments generated by digestion of SQPV genomic DNA with single restriction enzymes were cloned into a plasmid vector and used as probes to identify overlapping cloned restriction fragments generated by different enzymes, and for isolating overlapping fragments of DNA cloned into the SuperCos I cosmid vector. Mapping of the *Bam*HI, *Kpn*I, *Hin*dIII and *Eco*RI sites was achieved by the analysis of 10 overlapping cosmid clones. A detailed restriction endonuclease map of the *Bam*HI, *Kpn*I, *Hin*dIII and *Eco*RI sites present in the SQPV genome was constructed (Fig. 1) with the corresponding sizes of the restriction fragments summarized in Table 1. The terminal restriction fragments were not cloned, but are presumed to contain the terminal hairpin loop structures found at the ends of all poxvirus genomes. However, using a Southern blot of SQPV genomic DNA digested with *Kpn*I and hybridized with *Bam*HI fragment D' (not shown), the sizes of the terminal *Kpn*I fragments J and D were estimated to be 5.3 and 9.0 kb, respectively. As a result the positions of the two ends of the genome could be estimated. The sum of the sizes of restriction fragments generated with a single enzyme indicated that the SQPV genome is approximately 158 kb in length, whilst targeted sequencing of *Bam*HI restriction fragments K and V suggested that the inverted terminal repeat (present in all poxviruses) was approximately 5 kb in length.

Genomic relationship of SQPV with other poxviruses

The genomic regions that vary most between individual poxviruses and between the different poxvirus genera are those situated at either end of the viral genome. It was therefore decided to sequence from the LH inverted terminal repeats (ITR) towards a point in the genome representing the start of the genes conserved across all the recognized chordopoxvirus genera and likewise from the RH ITR to a point representing the end of the conserved genes. As a result we sequenced a contiguous stretch of 23 kb at the LH side of the SQPV genome and a contiguous stretch of 37 kb at the RH side. The sequences have been submitted to the GenBank database under the accession numbers DQ377804 and DQ377805, respectively. The sequence was analysed for the presence of ORFs with each being scored, as to their likelihood of representing a genuine SQPV gene, using codon usage tables calculated from either the ORFV genome or the *Molluscum contagiosum virus* (MOCV) genome, these being the only two poxviruses with a similar G+C content. Putative SQPV genes were translated and the conceptual amino acid sequences used in searches of the protein databases. A summary of the proposed genes, together with their similarity to known poxvirus sequences is presented in Table 2. Although many of the genes could be recognized as orthologues of known poxvirus genes, others, particularly at the extreme ends of the genome, did not appear to have counterparts in other poxviruses and indeed did not appear to be similar to anything in the sequence databases. These sequences were subjected to further analysis to determine whether or not they contained recognizable elements that would suggest their properties and/or function. A summary of these analyses is presented in Table 3.

Southern blot analysis of SQPV genomic DNA with ORFV genomic DNA fragments as probes (not shown) had suggested that the central region of the SQPV genome was highly related to that of ORFV. To confirm this, sequence data were obtained from a number of cloned restriction fragments from across the genome. Partial sequences corresponding to the orthologues of the VACV-Copenhagen (VACV-Cop) genes *E6R*, *E9L*, *G7R*, *G9R*, *L5R*, *J1R*, *J4R*, *J5L*, *J6R*, *H2R*, *D5R*, *A16L*, *A17L*, *A18R* and *A20R* were located at points in the genome that suggested a collinear relationship with ORFV and other poxviruses. Additionally, the spacing between the sequences suggested no major insertions or deletions were present in the genome (results not shown). A summary of this data together with its relative location in the SQPV genome is presented in Fig. 2.

Phylogenetic analysis of the SQPV genome

The phylogenetic relationship between SQPV and other poxviruses was examined using 15 genes that have been shown to be conserved across all the recognized chordopoxvirus genera. These were the *F9L*, *F10L*, *F12L*, *F13L*, *F17R*, *E1L*, *E2L*, *E4L*, *A23R*, *A24R*, *A28L*, *A29L*, *A30L*, *A32L* and *A34R* genes (named according to the VACV-Cop nomenclature). The corresponding amino acid sequences derived from each gene from 21 different viruses (representing each of the recognized chordopoxvirus genera), including SQPV, were aligned and concatenated to produce a single multiple alignment representing the 15 proteins and containing 5232 aa. A neighbour-joining tree was constructed using the Neighbour program within the PHYLIP package and its' statistical significance tested using 1000 bootstrap replicates of the data. The resulting tree is presented in Fig. 3. Using *Fowlpox virus* (FWPV) as an outgroup four major groupings (including FWPV) are clearly delineated with SQPV partitioning with the other viruses containing a genome base composition of $>$60 % G+C residues. However, there is 100 % bootstrap support for the branch that separates the SQPV from both MOCV and the parapoxviruses, suggesting that SQPV diverged before the line that gave rise to MOCV and the parapoxviruses.

Serology

Wood mouse and bank vole sera, collected in 1996 from Thetford Chase (Norfolk, UK) and in 2000 and 2001 from Manor Woods (Wirral, Merseyside, UK), were tested for evidence that these species could be infected with SQPV. A total of 57 wood mouse and 131 bank vole sera were screened for anti-SQPV antibodies using an adaptation of the ELISA described by Sainsbury *et al.* (2000). The positive-control samples (seropositive grey squirrel sera) were consistently positive (corrected optical density values ranged from 1.223 to 1.482) and the negative controls (seronegative grey squirrel sera and BALB/c mouse) all had corrected optical density values less than 0.1. Likewise, all the corrected optical density values for the wood mice and bank vole sera were less than 0.1, suggesting that these animals, at least, had not been infected by the virus. Serum samples were also collected from seven grey squirrels trapped in Dane County, Wisconsin, USA, and tested by ELISA. All seven serum samples tested positive,

with readings ranging from 0.59 to 2.78 (median=2.07). This is the first time antibodies to SQPV have been detected in serum samples originating from the USA.

DISCUSSION

SQPV has been classified as a parapoxvirus, but preliminary phylogenetic studies with a number of genes from the known parapoxviruses, including those that infect red deer, reindeer and seals, cast doubt on this and would suggest that if the squirrelpox virus is a parapoxvirus it would be the most divergent member of the genus described so far (Robinson & Mercer, 1995; Becher *et al.*, 2002; Thomas *et al.*, 2003; Tikkanen *et al.*, 2004). The restriction mapping reported here suggests a genome size of approximately 158 kb, nearly 20 kb longer than any wild-type ORFV genomes previously reported and 25 kb longer than the *Bovine papular stomatitis virus* (BPSV) genome, these being the only two parapoxviruses for which there is full genomic sequence published (Menna *et al.*, 1979; Robinson *et al.*, 1982, 1987; Mercer *et al.*, 1987, 2006; McInnes *et al.*, 2001; Delhon *et al.*, 2004). Some of the extra length is due to there being approximately 8 kb of extra sequence between the end of the orthologue of the VACV-Cop *E4L* gene and the ITR at the LH end of the genome in SQPV when compared to ORFV. Part of the reason for this is that the translocation of the orthologues of the VACV *F9L* and *F10L* genes from the LH end of the genome to the RH end, which appears to be characteristic of the genus *Parapoxvirinae* (Mercer *et al.*, 1995; Ueda *et al.*, 2003; Rziha *et al.*, 2003; Delhon *et al.*, 2004), has not occurred in the SQPV genome. Instead these genes are found in a location similar to that in all the other chordopoxvirus genera. Detailed sequence analysis of the two ends of the genome also revealed little similarity between the SQPV genome and those of the genus *Parapoxvirinae*. In particular none of the genes such as the *vVEGF*, the *vIL-10* or the ‘ankyrin-like’ repeat genes found in both ORFV and BPSV are found in the corresponding regions of the SQPV genome. Indeed SQPV is the only virus, other than MOCV, that, in the terminal regions of its genome, does not encode proteins predicted to contain ankyrin-like repeats.

The majority of poxvirus genes found between the *F9L* gene (VACV-Cop nomenclature) at the LH end of poxvirus genomes and the *A34R* gene at the RH end are essential for replication of the viruses *in vivo* and *in vitro*, whereas those outwith this region are generally considered non-essential, conferring instead an advantage to the viruses in combating an immune response or dictating the range of hosts that the viruses can successfully infect (Upton *et al.*, 2003; Gubser *et al.*, 2004). Many, but not all, of the genes found near the termini are genera-specific. In the SQPV genome, the majority of genes found near the left and right termini have no recognizable counterparts in other poxviruses. The remainder may be orthologues of known poxvirus genes, but in many of these instances the genes are found in different locations from the other poxviruses and the similarities between the proteins, predicted to be encoded by these genes, and their possible orthologues are extremely low.

At the LH end of the SQPV genome we have predicted that there are nine genes between the ITR and the orthologue of the *F9L* gene. Six of these, namely *K1L*, *K2L*, *I3L*, *I4L*, *I6L* and *A1L* (nomenclature based on the *Bam*HI restriction fragment in which the start codon is located) have no counterparts in other poxviruses. Indeed they do not appear to be similar to any known sequences. The three remaining genes, *I1L*, *I2L* and *I5L*, are similar to three genes

found only in MOCV. *I1L* appears to be similar in sequence to MOCV *003L* and is found in the same relative position (Senkevich *et al.*, 1997). MOCV *003L* has a paralogue at the RH end of the MOCV genome, namely *157R*, but the function of both remains obscure. A similar paralogous gene is found at the RH end of the SQPV genome (named *W2R*); however, the protein it encodes has slightly more similarity to the MC003L protein than the MC157R protein, suggesting that SQPV *I1L* might represent an ancestral gene, which was duplicated and translocated to the RH end of the genome and that subsequently in MOCV diverged to obtain properties or a function relevant to that particular virus. The SQPV *I2L* gene may be an orthologue of the MOCV *080R* gene that encodes a protein with similarity to class I major histocompatibility complex (MHC) proteins (Senkevich & Moss, 1998). The predicted proteins encoded by each gene share a high degree of identity, but positionally the genes are found in quite different locations. Pairwise alignment of the viral proteins suggest they share approximately 21 % identity, but a similar analysis between the SQPV protein and a human MHC class I E protein suggests they share 39 % identity. This raises the question as to whether the two viral genes are in fact orthologues of each other, or whether the genes have been acquired by the two viruses independently. The human MHC class I E is a non-polymorphic molecule that is involved in signalling to natural killer (NK) cells the extent of MHC class I expression on a cell surface (Braud *et al.*, 1998). Some poxviruses are known to downregulate the expression of MHC class I at the cell surface thus making these cells susceptible to NK-mediated cytolysis. It may well be that the SQPV protein may function to subvert this NK cell-mediated killing of virus infected cells. MHC orthologue genes are also found in the genomes of Yaba-like disease virus (YLDV) and *Swinepox virus* (SWPV) but neither of these are predicted to encode a transmembrane anchor sequence, thus distinguishing them from the MOCV and SQPV genes (Lee *et al.*, 2001; Afonso *et al.*, 2002). The last SQPV gene in this region, *I5L*, is similar to, and found in a similar relative position as, the MOCV *008L* gene, the protein encoded by which is of unknown function (Senkevich *et al.*, 1996, 1997).

In addition to the six genes found to the left of the *F9L* gene that had no counterparts in other poxviruses, a further three genes at the LH side of the SQPV genome, *A7L*, *A8L* and *A9L*, located between the orthologues of the VACV *F13L* and *F15L* genes also appear to be unique to SQPV. Using InterProScan (Zdobnov & Apweiler, 2001) to suggest possible functions for the proteins encoded by the 'unique' genes led to no predictions.

At the RH end of poxvirus genomes there is a region that in *Cowpox virus* (CPXV), amongst others, encodes the major protein component of the acidophilic-type inclusion (ATI) body associated with infected cells (Funahashi *et al.*, 1988; Meyer & Rziha, 1993). The corresponding region in the other poxviruses varies considerably in size such that the gap between the orthologues of the VACV-Cop *A24R* and *A27L* genes ranges in size from approximately 450 bases in *Yaba monkey tumour virus*, SWPV and *Lumpy skin disease virus* to approximately 5850 bases in CPXV. The corresponding region in the SQPV genome is over 8200 bases in length. There are predicted to be three highly related genes within this region, the functional significance of which is unknown.

We have predicted that between the orthologue of VACV-Cop *A34R* and the ITR at the RH end of the SQPV genome there are 19 genes covering approximately 18 kb of sequence. It is likely that four of these are orthologues of the VACV-Cop genes *A35R*, *A37R*, *A41L* and *A51R*. There are a further nine SQPV genes, namely *C3R*, *C7L*, *C10R*, *C14R*, *C15R*, *C16R*, *X1R*, *X2R* and *W2R* that have some similarities to other poxvirus genes, but whether or not they are true orthologues of these genes remains to be determined. This is because in general poxvirus genomes have been shown to be collinear, whereas seven of nine genes above are found in a different genomic location from the corresponding gene in other poxviruses. For example, the SQPV *C15R* gene is predicted to encode a protein with 31 % identity to the 014 protein encoded at the opposite end of the *Myxoma virus* genome (Cameron *et al.*, 1999), whilst the SQPV C10R has some sequence and positional similarity to the VACV A38L protein, but it is transcribed from the opposite strand of DNA to the VACV protein (Johnson *et al.*, 1993). It may be that rather than representing true orthologues SQPV has evolved independently to encode a variety of proteins with functions similar to those in other poxviruses. As with the LH end of the genome there are at least six genes at the RH end, *C4R*, *C6R*, *C8R*, *C11R*, *C12R* and *W1R* that encode proteins with no similarity to anything in the databases. Analysis with InterProScan revealed little about their potential function.

No evidence has been found that would support the classification of SQPV as a parapoxvirus. The lengths and gene content of the non-conserved regions at the LH and RH sides are quite different from the parapoxviruses, with none of the genes considered to be characteristic of the parapoxviruses being found in the SQPV genome. In addition, not only does the phylogenetic analysis with the 15 proteins conserved across all of the subfamily *Chordopoxvirinae* suggest a classification separate from the parapoxviruses, the nearest poxvirus match for the remaining proteins, as judged by individual pairwise alignments, is rarely with the parapoxvirus orthologue. Indeed there is no discernable pattern as to which species or genera usually provides the closest poxvirus match. These results provide fresh impetus to the study of SQPV and the ways in which it differs from the other poxviruses. Previous assumptions about the virus, particularly the mode of transmission, based on it being a parapoxvirus should be readdressed.

The fact that antibody to SQPV was not detected in serum from wood mice and bank voles in two areas of the UK where grey squirrels are seropositive does not preclude the virus from being endemic within an, as yet, unidentified British wildlife species. However, with the serum samples collected from grey squirrels in Wisconsin, USA, being found to contain antibody against SQPV, the serological evidence, for the first time, supports the theory that the virus was introduced to the UK with the grey squirrel.

ACKNOWLEDGEMENTS

This work was funded by The Sir James Knott Trust, NERC and SEERAD. The authors would also like to thank Professor Malcolm Bennett, University of Liverpool, for providing the

sera from wood mice and bank voles and Laurie A. Baeton for collecting sera from grey squirrels in Dane County, Wisconsin, USA.

REFERENCES

Afonso, C. L., Tulman, E. R., Lu, Z., Zsak, L., Osorio, F. A., Balinsky, C., Kutish, G. F. & Rock, D. L. (2002). The genome of swinepox virus. *J Virol* **76**, 783–790.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402.

Becher, P., Konig, M., Muller, G., Siebert, U. & Thiel, H.-J. (2002). Characterisation of sealpox virus, separate member of the parapoxviruses. *Arch Virol* **147**, 1133–1140.

Braud, V. M., Allan, D. S., O'Callaghan, C. A. & 9 other authors (1998). HLA-E binds to natural killer cell receptors CD94/NKG2A, B and C. *Nature* **391**, 795–799.

Buller, R. M., Arif, B. M., Black, D. N. & 9 other authors (2005). *Poxviridae*. In *Virus Taxonomy: Eighth Report of the International Committee on Taxonomy of Viruses*, pp. 117–133. Edited by C. M. Fauquet, M. A. Mayo, J. Maniloff, U. Desselberger & L. A. Ball. San Diego, CA: Academic Press.

Cameron, C., Hota-Mitchell, S., Chen, L., Barrett, J., Cao, J. X., Macaulay, C., Willer, D., Evans, D. & McFadden, G. (1999). The complete DNA sequence of myxoma virus. *Virology* **264**, 298–318.

Delhon, G., Tulman, E. R., Afonso, C. L., Lu, Z., de la Concha-Bermejillo, A., Lehmkuhl, H. D., Piccone, M. E., Kutish, G. F. & Rock, D. L. (2004). Genomes of the parapoxviruses ORF virus and bovine papular stomatitis virus. *J Virol* **78**, 168–177.

Ehlers, A., Osborne, J., Slack, S., Roper, R. L. & Upton, C. (2002). Poxvirus orthologous clusters (POCS). *Bioinformatics* **18**, 1544–1545.

Funahashi, S., Sato, T. & Shida, H. (1988). Cloning and characterization of the gene encoding the major protein of the A-type inclusion body of cowpox virus. *J Gen Virol* **69**, 35–47.

Gilray, J. A., Nettleton, P. F., Pow, I., Lewis, C. J., Stephens, S. A., Madeley, J. D. & Reid, H. W. (1998). Restriction endonuclease profiles of orf virus isolates from the British Isles. *Vet Rec* **143**, 237–240.

Gubser, C., Hue, S., Kellam, P. & Smith, G. L. (2004). Poxvirus genomes: a phylogenetic analysis. *J Gen Virol* **85**, 105–117.

Gurnell, J. & Pepper, H. (1993). A critical look at conserving the British red squirrel Sciurus vulgaris. *Mamm Rev* **23**, 127–137.

Johnson, G. P., Goebel, S. J. & Paoletti, E. (1993). An update on the vaccinia virus genome. *Virology* **196**, 381–401.

Lee, H. J., Essani, K. & Smith, G. L. (2001). The genome sequence of Yaba-like disease virus, a yatapoxvirus. *Virology* **281**, 170–192.

McInnes, C. J., Wood, A. R., Nettleton, P. E. & Gilray, J. A. (2001). Genomic comparison of an avirulent strain of Orf virus with that of a virulent wild type isolate reveals that the Orf virus G2L gene is non-essential for replication. *Virus Genes* **22**, 141–150.

Menna, A., Wittek, R., Bachmann, P. A., Mayr, A. & Wyler, R. (1979). Physical characterization of a stomatitis papulosa virus genome: a cleavage map for the restriction endonucleases *Hin*dIII and *Eco*RI. *Arch Virol* **59**, 145–156.

Mercer, A. A., Fraser, K., Barns, G. & Robinson, A. J. (1987). The structure and cloning of orf virus DNA. *Virology* **157**, 1–12.

Mercer, A. A., Lyttle, D. J., Whelan, E. M., Fleming, S. B. & Sullivan, J. T. (1995). The establishment of a genetic map of orf virus reveals a pattern of genomic organization that is highly conserved among divergent poxviruses. *Virology* **212**, 698–704.

Mercer, A. A., Ueda, N., Friederichs, S.-M., Hofmann, K., Fraser, K. M., Bateman, T. & Fleming, S. B. (2006). Comparative analysis of genome sequences of three isolates of Orf virus reveals unexpected sequence variation. *Virus Res* **116**, 146–158.

Meyer, H. & Rziha, H. J. (1993). Characterization of the gene encoding the A-type inclusion protein of camelpox virus and sequence comparison with other orthopoxviruses. *J Gen Virol* **74**, 1679–1684.

Pearson, W. R. & Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proc Natl Acad Sci U S A* **85**, 2444–2448.

Rackwitz, H. R., Zehetner, G., Murialdo, H., Delius, H., Chai, J. H., Poustka, A., Frischauf, A. & Lehrach, H. (1985). Analysis of cosmids using linearization by phage lambda terminase. *Gene* **40**, 259–266.

Rice, P., Longden, I. & Bleasby, A. (2000). EMBOSS: the European Molecular Biology open software suite. *Trends Genet* **16**, 276–277.

Robinson, A. J. & Mercer, A. A. (1995). Parapoxvirus of red deer: evidence for its inclusion as a new member in the genus parapoxvirus. *Virology* **208**, 812–815.

Robinson, A. J., Ellis, G. & Balassu, T. (1982). The genome of orf virus: restriction endonuclease analysis of viral DNA isolated from lesions of orf in sheep. *Arch Virol* **71**, 43–55.

Robinson, A. J., Barns, G., Fraser, K., Carpenter, E. & Mercer, A. A. (1987). Conservation and variation in Orf virus genomes. *Virology* **157**, 13–23.

Rushton, S. P., Lurz, P. W., Gurnell, J., Nettleton, P., Bruemmer, C., Shirley, M. D. & Sainsbury, A. W. (2005). Disease threats posed by alien species: the role of a poxvirus in the decline of the native red squirrel in Britain. *Epidemiol Infect* Oct 20: 1–13 [Epub ahead of print]

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.-A. & Barrel, B. (2000). ARTEMIS: sequence visualisation and annotation. *Bioinformatics* **16**, 944–945.

Rziha, H. J., Bauer, B., Adam, K. H., Rottgen, M., Cottone, R., Henkel, M., Dehio, C. & Buttner, M. (2003). Relatedness and heterogeneity at the near-terminal end of the genome of a parapoxvirus bovis 1 strain (B177) compared with parapoxvirus ovis (Orf virus). *J Gen Virol* **84**, 1111–1116.

Sainsbury, A. W. & Gurnell, J. (1995). An investigation into the health and welfare of red squirrels, *Sciurus vulgaris*, involved in reintroduction studies. *Vet Rec* **137**, 367–370.

Sainsbury, A. W., Nettleton, P., Gilray, J. & Gurnell, J. (2000). Grey squirrels have high seroprevalence to a parapoxvirus associated with deaths in red squirrels. *Anim Conserv* **3**, 229–233.

Scott, A. C., Keymer, I. F. & Labram, J. (1981). Parapoxvirus infection of the red squirrel (*Sciurus vulgaris*). *Vet Rec* **109**, 202.

Senkevich, T. G. & Moss, B. (1998). Domain structure, intracellular trafficking, and β2-microglobulin binding of a major histocompatibility complex class I homolog encoded by molluscum contagiosum virus. *Virology* **250**, 397–407.

Senkevich, T. G., Bugert, J. J., Sisler, J. R., Koonin, E. V., Darai, G. & Moss, B. (1996). Genome sequence of a human tumorigenic poxvirus: prediction of specific host response-evasion genes. *Science* **273**, 813–816.

Senkevich, T. G., Koonin, E. V., Bugert, J. J., Darai, G. & Moss, B. (1997). The genome of molluscum contagiosum virus: analysis and comparison with other poxviruses. *Virology* **233**, 19–42.

Thomas, K., Tompkins, D. M., Sainsbury, A. W., Wood, A. R., Dalziel, R., Nettleton, P. F. & McInnes, C. J. (2003). A novel poxvirus lethal to red squirrels (*Sciurus vulgaris*). *J Gen Virol* **84**, 3337–3341.

Tikkanen, M. K., McInnes, C. J., Mercer, A. A., Buttner, M., Tuimala, J., Hirvela-Koski, V., Neuvonen, E. & Huovilainen, A. (2004). Recent isolates of parapoxvirus of Finnish reindeer (*Rangifer tarandus tarandus*) are closely related to bovine pseudocowpox virus. *J Gen Virol* **85**, 1413–1418.

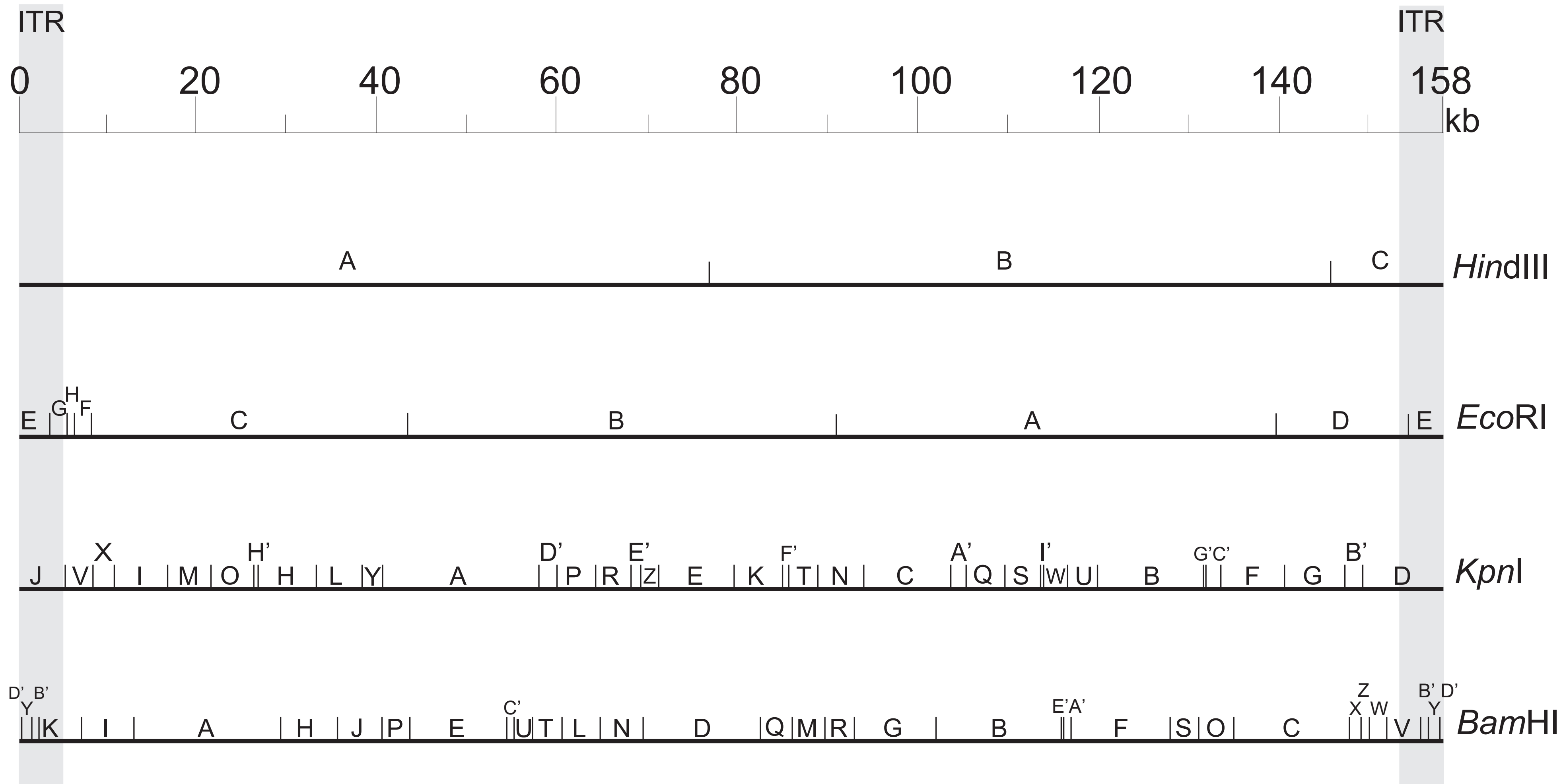
Tompkins, D. M., Sainsbury, A. W., Nettleton, P., Buxton, D. & Gurnell, J. (2002). Parapoxvirus causes a deleterious disease in red squirrels associated with UK population declines. *Proc R Soc Lond B Biol Sci* **269**, 529–533.

Tompkins, D. M., White, A. R. & Boots, M. (2003). Ecological replacement of native red squirrels by invasive greys driven by disease. *Ecol Lett* **6**, 189–196.

Ueda, N., Wise, L. M., Stacker, S. A., Fleming, S. B. & Mercer, A. A. (2003). Pseudocowpox virus encodes a homolog of vascular endothelial growth factor. *Virology* **305**, 298–309.

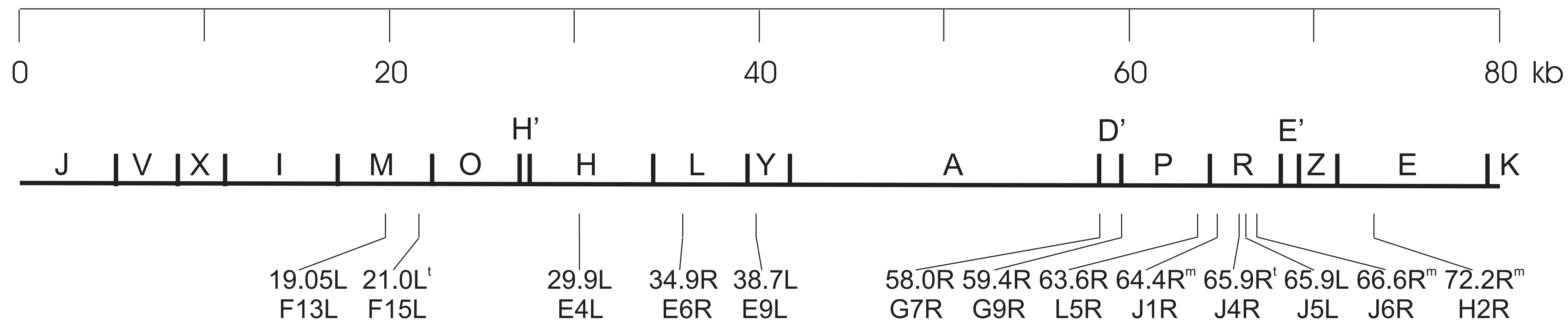
Upton, C., Slack, S., Hunter, A. L., Ehlers, A. & Roper, R. L. (2003). Poxvirus orthologous clusters: toward defining the minimum essential poxvirus genome. *J Virol* **77**, 7590–7600.

Zdobnov, E. M. & Apweiler, R. (2001). InterProScan – an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**, 847–848.

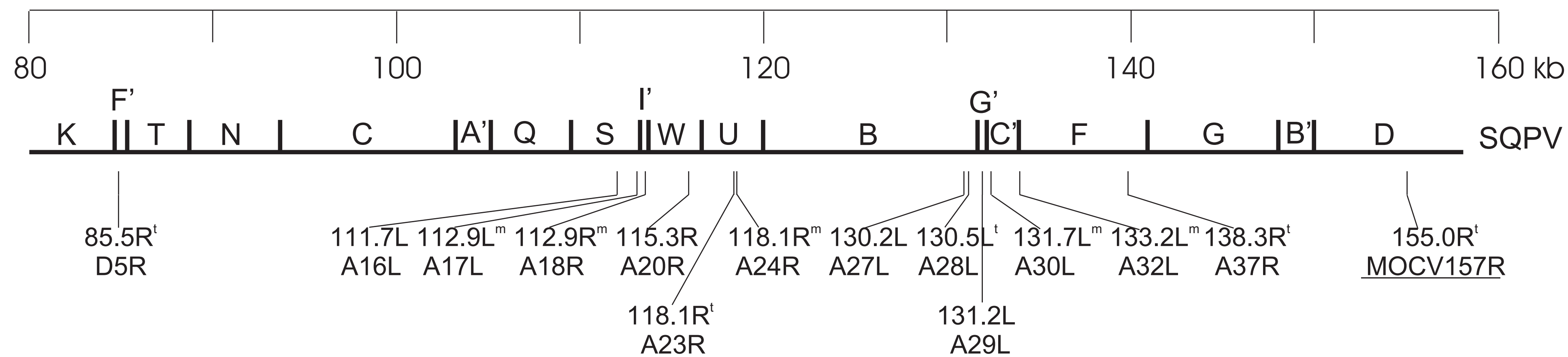


ITR
ITR
0
20
40
60
80
100
120
140
158
kb
A
B
C
HindIII
E
G
H
F
C
B
A
D
E
EcoRI
J
V
X
I
M
O
H'
H
L
Y
A
D'
P
R
E'
Z
E
K
F'
T
N
C
A'
Q
S
I'
W
U
B
G'C'
F
G
B'
D
KpnI
D' B'
Y
K
I
A
H
J
P
E
C'
U
T
L
N
D
Q
M
R
G
B
E'A'
F
S
O
C
Z
X
W
V
B' D'
Y
BamHI

Fig. 1. Schematic representation of the complete *Hin*dIII, *Eco*RI, *Kpn*I and *Bam*HI restriction endonuclease maps of SQPV. Each fragment is labelled with a letter, which corresponds to the approximate fragment size summarized in Table 1. Fragments were labelled A–Z, but where there are more than 26 fragments subsequent fragments were assigned the nomenclature A', B' etc. The inverted terminal repeats (ITR), which are approximately 5.0 kb in size, are shaded in grey.



0
20
40
60
80 kb
H'
D'
E'
J
V
X
I
M
O
H
L
Y
A
P
R
Z
E
K
19.05L
F13L
21.0L[t]
F15L
29.9L
E4L
34.9R
E6R
38.7L
E9L
58.0R
G7R
59.4R
G9R
63.6R
L5R
64.4R[m]
J1R
65.9R[t]
J4R
65.9L
J5L
66.6R[m]
J6R
72.2R[m]
H2R



80
100
120
140
160 kb
F'
I'
G'
K
T
N
C
A'
Q
S
W
U
B
C'
F
G
B'
D
SQPV
85.5R[t]
D5R
111.7L
A16L
112.9L[m]
A17L
112.9R[m]
A18R
115.3R
A20R
118.1R[t]
A23R
118.1R[m]
A24R
130.2L
A27L
130.5L[t]
A28L
131.2L
A29L
131.7L[m]
A30L
133.2L[m]
A32L
138.3R[t]
A37R
155.0R[t]
MOCV157R

Fig. 2. The *Kpn*I restriction endonuclease map of SQPV is shown together with the approximate location of a selection of genes identified by sequencing restriction fragments. The location coordinates (e.g. $64.4R^{m}$) indicates the approximate distance from the LH terminus of the SQPV genome together with the direction (L or R) in which the gene is transcribed. The suffix (^{m} or ^{t}) indicates the approximate location of either the predicted initiator methionine or the predicted stop codon. The corresponding orthologue of VACV is also indicated below the location coordinates, with the exception of MOCV157R (underlined).

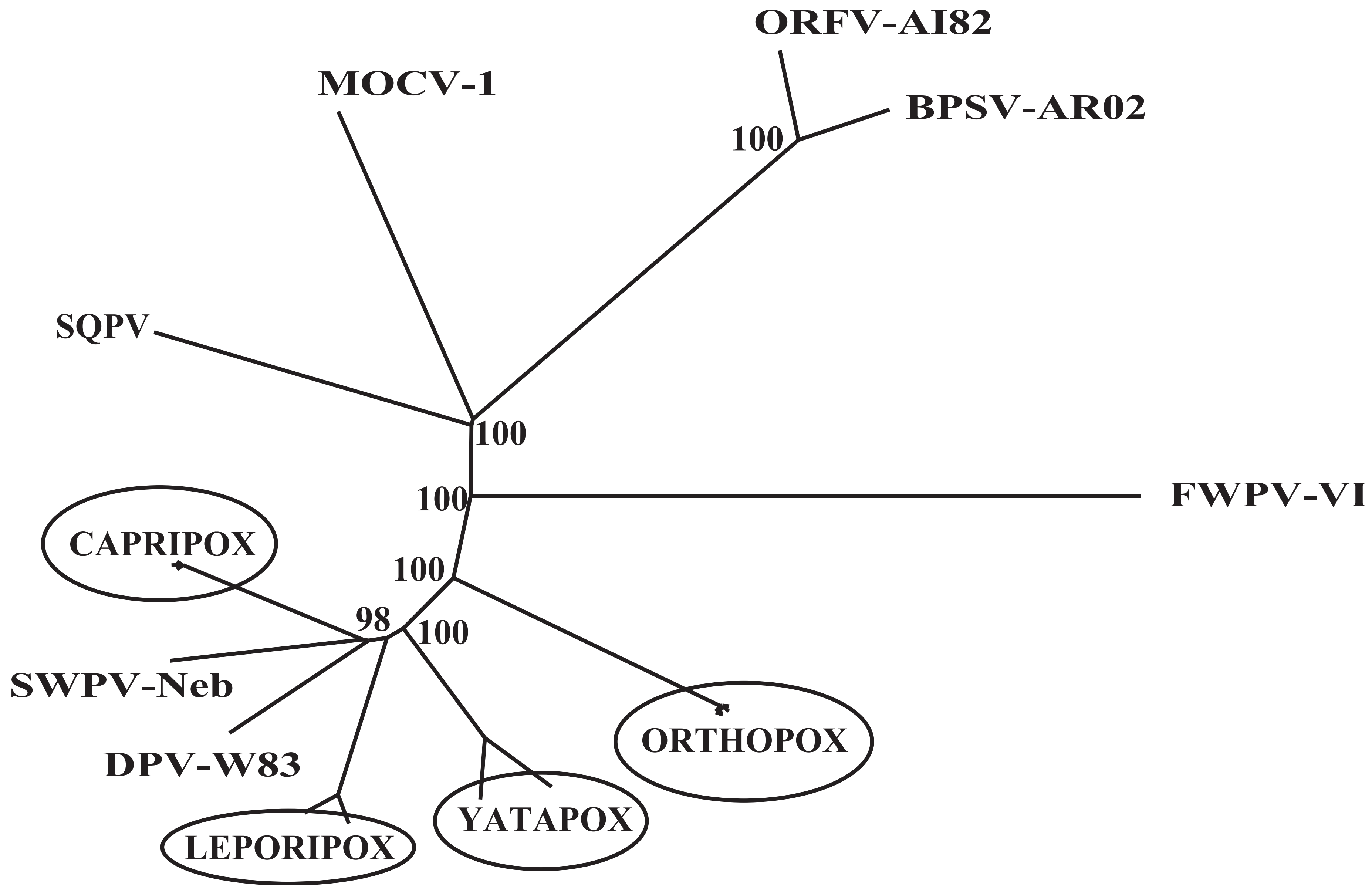

ORFV-AI82
MOCV-1
BPSV-AR02
100
SQPV
100
100
FWPV-VI
CAPRIPOX
100
98
100
SWPV-Neb
ORTHOPOX
DPV-W83
YATAPOX
LEPORIPOX
0.1

Fig. 3. Phylogenetic analysis of 15 proteins conserved across the chordopoxviruses. The conceptual amino acid sequences of the F9L, F10L, F12L, F13L, F17R, E1L, E2L, E4L, A23R, A24R, A28L, A29L, A30L, A32L and A34R proteins (named according to the VACV-Cop nomenclature) from 21 different poxviruses were aligned independently, columns with gaps removed and then concatenated to produce a single-multiple alignment file representing the 15 proteins and containing 5232 aa. A neighbour-joining tree was constructed using the Neighbour program within the PHYLIP package and its' statistical significance tested using 1000 bootstrap replicates of the data. Bootstrap values >95 % are indicated. The sequences of viruses used to construct the tree were: BPSV-AR02 (GenBank accession no. NC_005337), CMLV-M96 (NC_003391.1), CPXV-BR (NC_003663.2), DPV-W83 (NC_006966), ECTV-Mos (NC_004105), FWPV-VI (NC_002188.1), GTPV-Pellor (NC_004003), LSDV-Nee (NC_003027.1), MOCV-1 (NC_001731.1), MPXV-Zre (NC_003310.1), MYXV-Lau (NC_001132.2), ORFV-AI82 (AY386263), RPXV-Utr (AY484669), SFV-Kas (NC_001266), SPPV-NISKHI (AY077834), SWPV-Neb (NC_003389.1), VACV-Cop (M35027), VARV-Gar (Y16780.1), YLDV (NC_002642), YMTV (NC_005179) and SQPV (DQ377804 and DQ377805), all, with the exception of the SQPV, were taken from the VOCS website (www.virology.ca).

Table 1. The approximate sizes of the *Kpn*I, *Bam*HI, *Eco*RI and *Hin*dIII restriction fragments that comprise the complete genome of SQPV

Fragments contained entirely within the ITR are indicated by (×2) to emphasize they are found within both ITRs. For the location of each fragment refer to Fig. 1.

Fragments	*Kpn*I	*Bam*HI	*Eco*RI	*Hin*dIII
A	17.5	15.8	48.9	76.8*
B	11.7	14.0	47.6	69.0
C	9.6	13.8	35.1	12.4*
D	9.0*	13.3	14.1	–
E	8.4	11.35	3.9* (×2)	–
F	7.1	9.9	1.9	–
G	6.85	8.6	1.8	–
H	6.55	6.3	0.6	–
I	5.9	5.7	–	–
J	5.3*	4.75	–	–
K	5.25	4.65	–	–
L	5.05	4.6	–	–
M	4.9	4.5	–	–
N	4.85	4.1	–	–
O	4.8	3.9	–	–
P	4.65	3.15	–	–
Q	4.3	3.15	–	–
R	3.9	3.05	–	–
S	3.85	3.05	–	–
T	3.4	2.95	–	–
U	3.25	2.65	–	–
V	3.1	2.6	–	–
W	2.8	2.55	–	–
X	2.45	1.6	–	–
Y	2.25	1.15 (×2)	–	–
Z	2.05	1.1	–	–
A’	1.85	0.9	–	–
B’	1.8	0.8 (×2)	–	–
C’	1.6	0.65	–	–
D’	1.45	0.41* (×2)	–	–
E’	0.9	0.17	–	–
F′	0.7	0.027	–	–
G’	0.55	–	–	–
H’	0.25	–	–	–

I’	0.03	–	–	–
Total	**157.88**	**157.6**	**157.8**	**158.2**

*Indicates the terminal fragments, which are predicted to contain a terminal hairpin-loop structure.

Table 2. Summary of the putative homologues of previously described poxvirus genes identified in SQPV

ORF*	Nucleotide position†	Length‡	ORFV equivalent §	Identity versus ORFV equivalent (%)	Nearest poxvirus match‖	Identity versus nearest poxvirus match (%)	VACV equivalent ¶	Predicted structure/function#
LHS								
K1L	74–403c	109	–	–	–	–	–	–
K2L	509–928c	139	–	–	–	–	–	–
I1L	1079–2308c	409	–	–	MOCV-SB1_003	23	–	Immunoglobulin domain
I2L	2446–3480c	344	–	–	MOCV-SB1_080	21	–	Similar to human HLA-E
I3L	3547–4176c	209	–	–	–	–	–	–
I4L	4173–4811c	212	–	–	–	–	–	–
I5L	4845–5363c	172	–	–	MOCV-SB1_008	30	–	–
I6L	5609–5998c	129	–	–	–	–	–	–
A1L	6173–8224c	683	–	–	–	–	–	–
A2L	8251–8895c	214	131	40	RPXV-UTR_037	50	F9L	S–S bond formation pathway protein
A3L	8888–10201c	437	130	61	MOCV-SB1_017L	68	F10L	Ser/Thr protein kinase
A4L	10198–11676c	492	009	28	RPXV-UTR_039	34	F11L	–
A5L	11719–13698c	659	010	31	MOCV-SB1_019	36	F12L	Microtubule transport?

A6L	13736–14893c	385	011	50	SPPV-NISKHI_018	54	F13L	Major envelope protein/phospholipase D
A7L	14922–15215c	97	–	–	–	–	–	–
A8L	15277–15540c	87	–	–	–	–	–	–
A9L	15512–15682c	56	–	–	–	–	–	–
A10L	15729–16178c	149	–	–	SPPV-TU_025	54	F15L	–
A11L	16175–16435c	86	014	42	MOCV-SB1_026	42	–	Zinc finger/RING type
A12L	16470–18053c	527	015	23	MOCV-SB1_027	30	–	–
A13L	18102–18821c	239	016	26	MOCV-SB1_029	34	F16L	–
A14L	18893–19189	98	017	37	MPXV-ZRE_049	52	F17R	DNA-binding phosphoprotein
A15L	19194–20606c	470	018	51	LSDV-WARM_034	56	E1L	Poly(A) polymerase subunit
A16L	20603–22816c	737	019	34	MOCV-SB1_032	39	E2L	–
H1L	22881–24053c	390	020	29	MPXV-ZRE_052	33	E3L	ds RNA-binding/IFN resistance
H2L	24270–25223c	317	021	56	LSDV-1959_036	59	E4L	RNA polymerase subunit
RHS								
F1R	46–1203	385	097	57	MOCV-SB1_128	60	A23R	Intermediate transcription factor

F2R	1223–4720	1165	098	72	GTPV-Pellor_092	79	A24R	RNA polymerase subunit
F3L	4717–6207c	496	–	–	CPXV-GRI_148	22	A26L	ATI protein
F4R	6206–6478	90	–	–	–	–	–	–
S1L	7122–10898c	1258	–	–	CPXV-BR_162	21	A26L	ATI protein
O1L	10958–12958c	666	103	26	MOCV-SB1_131	25	A26L	ATI protein
O2L	12995–13219c	74	104	51	MPXV-ZRE_139	47	A27L	Fusion protein/binds to heparin sulphate
O3L	13225–13647c	140	105	52	MOCV-SB1_134	53	A28L	Membrane component of IMV
O4L	13648–14577c	309	106	49	MYXV-LAU_121	54	A29L	RNA polymerase subunit
O5L	14546–14761c	71	107	36	MOCV-SB1_136	54	A30L	Virion morphology?
O6R	14972–15457	163	–	–	VACV-WR_154	41	A31R	–
O7L	15454–16248c	264	108	70	MOCV-SB1_140	72	A32L	ATPase/DNA packaging protein
O8R	16376–16957	193	109	27	VARV-IND_140	26	A33R	EEV glycoprotein
C1R	17006–17566	186	110	25	VARV-BSH_144	29	A34R	EEV glycoprotein
C2R	17596–18129	177	111	33	MOCV-SB1_145	38	A35R	–
C3R	18177–19046	289	–	–	MOCV-SB1_144	36	–	Concanavalin-like precursor
C4R	19192–19878	228	–	–	–	–	–	–
C5R	19929–20924	331	114	29	MOCV-SB1_149	29	A37R	–

C6R	20972–21472	166	–	–	–	–	–	–
C7L	21607–22146c	179	–	–	MOCV-SB1_160	33	–	Apoptosis regulator
C8R	22217–23302	361	–	–	–	–	–	–
C9L	23470–24465c	331	–	–	ECTV-MOS_173	25	A41L	Secreted virulence protein
C10R	24620–25648	342	–	–	CPXV-BR_179	19	A38L	CD47-like
C11R	25659–25865	68	–	–	–	–	–	–
C12R	25915–26337	140	–	–	–	–	–	–
C13L	26382–27419c	345	122	26	YMTV-YLD_138	18	A51R	–
C14R	27500–28252	250	–	–	YMTV-RPY_011	21	–	Apoptosis regulator
C15R	28602–29408	268	–	–	MYXV-LAU_014	31	–	KELCH-like
C16R	29465–30565	366	–	–	MYXV-LAU_142	24	–	Ser/Thr protein kinase
X1R	30601–31446	281	124	29	CPXV-BR_163	45	–	IMV membrane component?
X2R	31241–33451	736	085	24	ORFV-SA00_085	24	–	Structural protein?
W1R	33555–33974	139	–	–	–	–	–	–
W2R	34501–35484	328	–	–	MOCV-SB1_003	26	–	Immunoglobulin domain

*The designated name of the SQPV gene based on the *Bam*HI restriction fragment in which the start codon is found together with the direction of transcription leftwards (L) or rightwards (R).

†Nucleotide coordinates of the gene from DQ377804 or DQ377805. If the gene is transcribed from the opposite strand of the DNA this is indicated by (c).

‡Length of the predicted protein in amino acids.

§The nearest ORFV protein as assessed using the BLAST tool (Altschul *et al.*, 1997) on the Poxvirus Bioinformatics Resource website (www.poxvirus.org).
||The nearest poxvirus protein as assessed using the BLAST tool on the Poxvirus Bioinformatics Resource website.
¶The VACV-Cop orthologue.
#The predicted or known structure/function of the nearest poxvirus or VACV orthologue.

Table 3. Summary of the predicted structure/function of the SQPV proteins that showed no similarity to any known proteins.

Y, Yes; N, no.

ORF*	Nucleotide position†	Length‡	SIGCLEAVE§	TMAP||	Prosite motifs¶
LHS					
K1L	74–403c	109	Y	11–39, 64–92	–
K2L	509–928c	139	N	9–27	–
I3L	3547–4176c	209	N	N	–
I4L	4173–4811c	212	Y	N	–
I6L	5609–5998c	129	N	N	–
A1L	6173–8224c	683	N	569–592	–
A7L	14922–15215c	97	N	N	–
A8L	15277–15540c	87	N	N	–
A9L	15512–15682c	56	Y	N	–
RHS					–
C4R	19192–19878	228	N	4–32	–

C6R	20972–21472	166	Y	N	164–166 MICROBODIES_CTER, 21–23 RGD
C8R	22217–23302	361	N	201–219	–
C11R	25659–25865	68	N	N	–
C12R	25915–26337	140	N	N	–
W1R	33555–33974	139	N	N	68–84 G_PROTEIN_RECEP_F1_1

*The designated name of the SQPV gene based on the *Bam*HI restriction fragment in which the start codon is found together with the direction of transcription leftwards (L) or rightwards (R).

†Nucleotide coordinates of the gene from DQ377804 or DQ377805. If the gene is transcribed from the opposite strand of the DNA this is indicated by (c).

‡Length of the predicted protein in amino acids.

§Protein coordinates of putative signal sequence predicted using the program SIGCLEAVE within the EMBOSS sequence analysis package (Rice *et al.*, 2000).

||Protein coordinates of putative transmembrane regions predicted using the program TMAP within the EMBOSS sequence analysis package.

¶Prosite motifs present within the predicted amino acid sequences.