Phylogenetic analysis of the large family of poxvirus ankyrin-repeat proteins reveals orthologue groups within and across chordopoxvirus genera

Stephanie Sonnberg, Stephen B. Fleming and Andrew A. Mercer

Correspondence
Andrew A. Mercer
andy.mercer@otago.ac.nz

Received 15 May 2011
Accepted 11 July 2011

Ankyrin-repeat (ANK) protein-interaction domains are common in cellular proteins but are relatively rare in viruses. Chordopoxviruses, however, encode a large number of ANK domain-containing ORFs of largely unknown function. Recently, a second protein-interaction domain, an F-box-like motif, was identified in several poxvirus ANK proteins. Cellular F-box proteins recruit substrates to the ubiquitination machinery of the cell, a putative function for ANK/poxviral F-box proteins. Using publicly available genome sequence data we examined all 328 predicted ANK proteins encoded by 27 chordopoxviruses that represented the eight vertebrate poxvirus genera whose members encode ANK proteins. Within these we identified 15 putative ANK protein orthologue groups within orthopoxviruses, five within parapoxviruses, 23 within avipoxviruses and seven across members of the genera Leporipoxivirus, Capripoxivirus, Yatapoxivirus, Suipoxivirus and Cervidpoxivirus. Sequence comparisons showed that members of each of these four clusters of orthologues were not closely related to members of any of the other clusters. Of these ORFs, 67% encoded a C-terminal poxviral F-box-like motif, whose absence could largely be attributed to fragmentation of ORFs. Our findings suggest that the large family of poxvirus ANK proteins arose by extensive gene duplication and divergence that occurred independently in four major genus-based groups after the groups diverged from each other. It seems likely that the ancestor ANK proteins of poxviruses contained both the N-terminal ANK repeats and a C-terminal F-box-like domain, with the latter domain subsequently being lost in a small subset of these proteins.

INTRODUCTION

Chordopoxviruses, a diverse subfamily of the family Poxviridae of DNA viruses, infect a wide range of vertebrates (Moss, 2007). Terminal regions of their large linear genomes typically encode host regulating factors and multiple ankyrin-repeat (ANK) proteins (Damon, 2007). The ANK protein-interaction domain consists of two or more ankyrin motifs and is frequently found in cellular proteins, but relatively rarely in virally encoded proteins, with chordopoxviruses a notable exception (Bork, 1993; Li et al., 2006; Mosavi et al., 2004).

The only classified chordopoxvirus not encoding ANK proteins is molluscum contagiosum virus, the sole member of the genus Molluscipoxivirus. Poxviruses of the remaining eight genera each encode numerous ANK proteins (Afonso et al., 2000, 2002, 2005; Cameron et al., 1999; Chen et al., 2003; Delhon et al., 2004; Gubser & Smith, 2002; Tulman et al., 2001, 2002, 2004, 2006). Specific ANK proteins of cowpox virus (CPX), vaccinia virus (VACV) and myxoma virus (MYXV) have been shown to affect the viruses’ host range and virulence, but the functions of the vast majority of poxvirus ANK proteins remain unknown (Bradley & Terajima, 2005; Hsiao et al., 2004, 2006; Perkus et al., 1990; Ramsey-Ewing & Moss, 1996). Recently, a second protein-interaction domain, a truncated F-box, has been identified in poxvirus ANK proteins (Chang et al., 2009; Mercer et al., 2005; Sonnberg et al., 2008; Sperling et al., 2008; van Buuren et al., 2008). Cellular F-box proteins are bipartite, containing an F-box motif that interacts with S-phase-kinase-associated protein 1 (Skp1) of the Skp1-cullin1-F-box protein (SCF1)-type ubiquitin ligase (Bai et al., 1996; Skowyra et al., 1997) as well as a second protein-interaction domain that recruits specific target proteins. Recruitment of the target proteins to the SCF1 complex by F-box proteins most commonly leads to their poly-ubiquitination and proteasomal degradation (Bai et al., 1996; Ho et al., 2008; Jin et al., 2004; Kipreos & Pagano, 2000; Skowrya et al., 1997). Previous studies revealed
most of the limited number of poxvirus ANK proteins examined contained an F-box-like motif capable of interacting with cellular SCF1 ubiquitin ligases (Blanié et al., 2009; Chang et al., 2009; Mercer et al., 2005; Sonnberg et al., 2008; Sperling et al., 2008; van Buuren et al., 2008; Werden et al., 2009). The combination of F-box-like domain and ANK domain is unique to poxviruses.

Sequencing studies have reported orthologous ANK proteins encoded in closely related poxviruses, but no comprehensive analysis of the relationships within this extensive viral protein family has been carried out. This gap in our understanding of the largest of poxvirus protein families hinders a more comprehensive analysis of the functions and origins of these proteins. We therefore surveyed the relatedness of poxvirus ANK proteins across the entire chordopoxvirus subfamily, with the goal of assessing the presence of more global orthologue groups, as well as providing information relevant to the origins of this protein family, and elucidating the significance and prevalence of the poxvirus F-box-like motif.

### RESULTS

**ANK proteins of chordopoxviruses form major homology groups within and across genera**

Previous reports of chordopoxvirus genome sequences indicated the presence of homologous ANK proteins in related poxviruses (Afonso et al., 2000, 2002, 2005; Brunetti et al., 2003; Cameron et al., 1999; Chen et al., 2003; Delhon et al., 2004; Gubser & Smith, 2002; Lee et al., 2001; Massung et al., 1994; Schelkunov et al., 2000; Tulman et al., 2001, 2002, 2004, 2006; Willer et al., 1999). Here, we conducted a comprehensive analysis of poxvirus ANK protein homologies in representative members of the eight chordopoxvirus genera that encode ANK proteins. We selected 21 species and 27 strains (Table 1) for which full genome sequences were available and examined all annotated ORFs using Pfam analysis. Each virus encoded multiple ANK ORFs ranging from four (swinepox virus, SWPV) to 51 (canarypox virus, CNPV). Altogether, 328 ANK proteins were identified (Table 1, Supplementary

<table>
<thead>
<tr>
<th>Genus</th>
<th>Species</th>
<th>Strain</th>
<th>No. ANK proteins</th>
<th>Proteins with F-box (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avipoxvirus</td>
<td>Canarypox virus (CNPV)</td>
<td>ATCC VR111</td>
<td>51</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>Fowlpox virus (FWPV)</td>
<td>Iowa</td>
<td>31</td>
<td>71</td>
</tr>
<tr>
<td>Orthopoxvirus</td>
<td>Camelpox virus (CML)</td>
<td>CMS</td>
<td>13</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>Cowpox virus (CPX)</td>
<td>Brighton Red</td>
<td>17</td>
<td>82</td>
</tr>
<tr>
<td></td>
<td>Ectromelia virus (ECT)</td>
<td>Naval</td>
<td>8</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>Horsepox virus* (HSPV)</td>
<td>MNR-76</td>
<td>31</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>Monkeypox virus (MPX)</td>
<td>USA-2003-044</td>
<td>11</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>Taterapox virus (TAV)</td>
<td>Zaire-1979-005</td>
<td>11</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>Vaccinia virus (VACV)</td>
<td>Dahomey</td>
<td>18</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>Variola virus (VARV)</td>
<td>Copenhagen</td>
<td>16</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bangladesh 1975</td>
<td>10</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Garcia 1966</td>
<td>11</td>
<td>36</td>
</tr>
<tr>
<td>Capripoxvirus</td>
<td>Sheeppox virus (SPPV)</td>
<td>TU-V02127</td>
<td>5</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>Goatpox virus (GTPV)</td>
<td>G20-LKV</td>
<td>5</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>Lumpy skin disease virus (LSDV)</td>
<td>Neethling 2490</td>
<td>5</td>
<td>80</td>
</tr>
<tr>
<td>Leporipoxvirus</td>
<td>Rabbit fibroma virus (RFV)</td>
<td>Kasza</td>
<td>4</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Myxoma virus (MYXV)</td>
<td>Lausanne</td>
<td>5</td>
<td>100</td>
</tr>
<tr>
<td>Parapoxvirus</td>
<td>Orf virus (OV)</td>
<td>NZ2</td>
<td>5</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IAB2</td>
<td>5</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SA00</td>
<td>5</td>
<td>100</td>
</tr>
<tr>
<td>Suipoxvirus</td>
<td>Bovine papular stomatitis virus (BPSV)</td>
<td>BV_AR02</td>
<td>7</td>
<td>100</td>
</tr>
<tr>
<td>Yatapoxvirus</td>
<td>Swinepox virus (SWPV)</td>
<td>Nebrask 17077-99</td>
<td>4</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Tanapox virus (TPV)</td>
<td>Davis</td>
<td>5</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>Yaba monkey tumor virus (YMTV)</td>
<td>Amano</td>
<td>4</td>
<td>100</td>
</tr>
<tr>
<td>Cervidpoxvirus</td>
<td>Deerpox virus (DPV)</td>
<td>W-1170-84</td>
<td>6</td>
<td>83</td>
</tr>
</tbody>
</table>

*Recently reclassified as VACV strain by www.poxvirus.org.
Table S1, available in JGV Online). We began an examination of the inter-relationships between the multiple proteins encoded by a single virus as well as their relationships with the ANK proteins encoded by other viruses by aligning the amino acid sequences of these 328 proteins and generating a neighbour-joining phylogenetic tree (Supplementary Fig. S1, available in JGV Online) as described in Methods. The generated tree displayed major alignment groups composed of orthopoxvirus (OPV)-encoded ANK proteins, parapoxvirus (PPV)-encoded ANK proteins, avipoxvirus-encoded ANK proteins, and a group of ANK proteins encoded by leporipoxviruses, capripoxviruses, yatapoxviruses, suipoxviruses and cervidpoxviruses (termed hereafter the leporipoxvirus-supergroup, LSG).
Seven ANK protein orthologues are encoded by members of LSG

In order to more clearly identify putative orthologue groups among the chordopoxvirus ANK proteins, further alignments were carried out within the major alignment groups. The ANK proteins encoded by the LSG viruses clustered together in the initial analysis of all 328 proteins (Supplementary Fig. S1). These 43 proteins were realigned and a phylogenetic tree was generated as described in Methods. Seven putative orthologue groups were identified (Fig. 1a, Supplementary Table S2). Each virus encoded a single representative of four to six of these orthologue groups (Fig. 1b). The orthologue groups were numbered using Roman numerals, commencing at the left-most ORF of the leporipoxvirus member of the orthologue groups. The multiple ANK proteins encoded by each virus can be considered paralogues as they appear to be gene duplications of an ancestor ORF in the respective genome. The orthologue groups are numbered with Roman numerals commencing from the left-most of the leporipoxvirus orthologues. Square boxes indicate full-length ORFs, while small rectangles indicate truncated ORFs as identified by being of less than 400 aa, or by comparison with other members of the specific orthologue group. Filled boxes indicate the presence of a C-terminal F-box-like motif and open boxes indicate the absence of this motif. Viruses and abbreviations are as in Table 1 and Supplementary Fig. S1. CPV, Capripoxviruses; LPV, leporipoxviruses; YPV, yatapoxviruses; SPV, suipoxviruses; CRV, cervidpoxviruses.

PPVs encode five to seven ANK protein orthologues

We analysed representatives of two species of PPVs, bovine papular stomatitis virus (BPSV, strain BV_AR02) and orf virus (OV) (strains NZ2, SA00 and IA82) (Delhon et al., 2004; Mercer et al., 2006). BPSV and OV display an overall amino acid identity of 71 % and are relatively distantly related when compared to members of other genera (Delhon et al., 2004). BPSV encodes seven ANK protein paralogues, which exhibited amino acid identities of 31–48 % to each other. The three isolates of OV each encoded five ANK proteins among which orthologues and paralogues were readily apparent. Among the five ANK paralogues of each OV isolate the mean amino acid identity was 35 % (range 27–44 %). The mean amino acid sequence identity of the OV orthologues was 94 % (range 81.8–99.8 %) and each had a clear orthologue among the BPSV ANK proteins (Fig. 2, Supplementary Table S5, available in JGV Online) with orthologues between OV and BPSV displaying amino acid identities of ~60 % and groupings consistent with a previous report (Hughes et al., 2010). The orthologue groups were numbered according to their genomic location commencing at the left-most ANK protein ORF of OV using Roman numerals (Fig. 2b). BPSV ANK proteins 002 and 003 displayed 30–40 % amino acid identity included in this orthologue group rather than in group I (23.7 % amino acid identity with MYXV-L005L/R and 25.1 % with RIVF_K005) another possibility that appeared in the phylogenetic analysis (Fig. 1a). The closest parologue to MYXV_L150R in the MYXV genome are the duplicated genes MYXV-L005L and MYXV-L005R of LSG group I. With the exception of ankyrin genes duplicated in the left and right terminal repeat regions of poxvirus genomes, no orthologue groups included two full-length ANK genes encoded by the same virus. MYXV-L150 may represent a unique gene duplication and may not belong to any of the LSG orthologue groups. All ANK proteins of chordopoxviruses display homologies to each other to different extents, therefore the assignment of orthologue groups with relatively low levels of homology must be tentative and will be assisted by the availability of further genome sequences and functional studies.

Fig. 1. (a) Phylogenetic relationship of 43 ANK proteins of capripoxviruses, leporipoxviruses, suipoxviruses, yatapoxviruses and deerpox virus (DPV). The amino acid sequences were aligned in MEGALIGN using CLUSTAL W. A neighbour-joining tree was generated and bootstrapped (1000 trials, seed 111). The dotted bracket indicates three proteins that may be additional members of orthologue group VII. Solid lines in the phylogenetic tree represent pairwise distances with dashed lines inserted for ease of viewing. Roman numerals refer to orthologue groups listed in (b). (b) Overview of the orthologue groups of LSG ANK proteins. Boxes represent genes encoding putative ANK protein orthologue groups arranged to represent the relative distributions of the genes in the genomes. The orthologue groups are numbered with Roman numerals commencing from the left-most of the leporipoxvirus orthologues. Square boxes indicate full-length ORFs, while small rectangles indicate truncated ORFs as identified by being of less than 400 aa, or by comparison with other members of the specific orthologue group. Filled boxes indicate the presence of a C-terminal F-box-like motif and open boxes indicate the absence of this motif. Viruses and abbreviations are as in Table 1 and Supplementary Fig. S1. CPV, Capripoxviruses; LPV, leporipoxviruses; YPV, yatapoxviruses; SPV, suipoxviruses; CRV, cervidpoxviruses.
identity to all other PPV ANK proteins and were most closely related to BPSV_129 (Fig. 2). They may represent an ORF duplication that only occurred in BPSV or two additional orthologue groups.

**OPVs encode 15 ANK protein orthologues**

We next more closely examined the relationships between the 181 ANK proteins encoded by the selected OPVs. Further alignment of only these OPV ANK proteins identified 15 putative orthologue groups (Fig. 3, Supplementary Fig. S2, Supplementary Table S6, available in JGV Online). The large number of short ORF among the OPV required a combination of phylogenetic analysis and individual pairwise alignments to assign ORF fragments to their orthologue groups. OPV orthologue groups were numbered according to genomic location commencing at the left-most ANK protein ORF of CPX strain Brighton Red (BR) using Roman numerals (Fig. 3, Supplementary Fig. S2).

CPX is considered more closely related to the ancestral OPV than any other sequenced OPV (Meyer et al., 2002; Shchelkunov et al., 2002). Consistent with this hypothesis, the three CPX strains examined here, BR, GRI-90 and Germany-91 (GER), encoded the largest set of ANK proteins within OPV (Fig. 3). CPX_BR had the largest set of ANK proteins, 15 paralogues, two of which were encoded in the inverted terminal repeat (ITR) regions and thus present in duplicate (Fig. 3). The CPX ANK proteins displayed amino acid identities of generally 7–15% to their paralogues and >95% to their orthologues (Esposito et al., 2006) (GenBank accessions nos NC_003663 and X94355) (data not shown). CPX_GER and CPX_GRI-90 differed in only two ANK proteins from strain BR: GER contained OPV orthologue XV in three fragments, while GRI-90 did not encode OPV orthologue XII (Fig. 3, Supplementary Fig. S2).

Whole-genome analysis has recently indicated that CPX_GRI-90 may be a separate species from CPX_GER and CPX_BR, an observation that is also reflected in our analysis (Hendrickson et al., 2010, Supplementary Figs S1 and S2).

The remaining nine OPVs examined encoded different subsets of the 15 ANK proteins of CPX_BR, but none...
encoded additional ANK proteins (Fig. 3). The OPV orthologues displayed amino acid identities of ~85–99% reflecting the high level of relatedness within this genus (Gubser et al., 2004) (data not shown). OPV orthologue XI was the only ANK protein present in all examined OPVs, while orthologues I and IX were fragmented only in VACV strain Copenhagen or horsepox virus (HSPV), respectively (Fig. 3, Supplementary Fig. S2, Supplementary Table S6). There was no difference in the sets of ANK proteins encoded by the two monkeypox virus (MPX) strains, however, there were several small differences between variola virus strain Garcia-1966 (VARV_G), which causes smallpox minor, and variola virus strain Bangladesh-1975 (VARV_B), which causes smallpox major (Fig. 3). The VARV_B orthologue of OPV group VII is truncated compared with other proteins in this orthologue group, and is fragmented further into two predicted ORFs in VARV_G. VARV_B encodes a fragment of OPV orthologue group II, while VARV_G encodes a fragment of OPV orthologue group VIII (Fig. 3).

**Avipoxvirus ANK proteins**

The genus *Avipoxvirus* currently comprises 10 viruses with the largest genomes among the poxvirus family (Bolte et al., 1999; Boyle, 2007; Jarmin et al., 2006). The two avipoxviruses sequenced to date, fowlpox virus (FWPV) and CNPV, display an overall amino acid sequence identity of 53% and are recognized as highly diverse members of the genus (Jarmin et al., 2006; Tulman et al., 2004). CNPV and FWPV encode 51 and 31 ANK protein ORFs, respectively (Afonso et al., 2000; Tulman et al., 2004). The significant overall sequence divergence among these proteins limited the extent to which they could be assigned to orthologue groups.

The amino acid sequences of the 82 avipoxvirus ANK proteins were aligned using CLUSTAL W and a neighbour-joining phylogenetic tree was generated. CNPV ANK protein paralogues displayed amino acid identities of 10–22%, while FWPV ANK protein paralogues were 15–25% identical at the amino acid level (data not shown). Only one probable orthologue group, containing CNPV_009 and FWPV_244, displaying an amino acid identity of 81%, could be identified, while 16 further putative orthologue groups displayed amino acid identities of 40–60% (Fig. 4, Supplementary Fig. S3, available in JGV Online). These levels of amino acid identities were similar to LSG orthologues encoded by viruses of different genera (Supplementary Table S4). Six further ANK proteins displayed amino acid identities of 30–40% (Fig. 4), which is significantly higher than the amino acid identities of ANK protein paralogues encoded in either genome. Because of the uncertainty surrounding how many orthologue groups are present in avipoxviruses, no numbering of groups was attempted here. The previously reported rearrangement of genes in the FWPV genome was also reflected in the ANK protein orthologue pairings, as were gene duplications in CNPV (Fig. 4) (Afonso et al., 2000; Tulman et al., 2004).

**ORF fragmentation accounts for the majority of F-box loss in poxvirus ANK proteins**

ANK proteins of chordopoxviruses have been shown to contain a truncated F-box motif located at the C terminus.
Fig. 4. Overview of the orthologue groups of avipoxvirus ANK proteins. Boxes represent genes encoding ANK proteins arranged to represent their relative distribution in the viral genomes. Lines between boxes indicate putative orthologue groups. Square boxes, small rectangles, filled boxes and open boxes are as described in Fig. 1. Boxed numbers indicate percentage amino acid identities between two indicated proteins. Boxed adjacent FWPV ORFs are predicted fragments of the single indicated orthologue ORF of CNPV. Boxed numbers below the FWPV schematic refer to percentage amino acid identities between the less similar fragment and the CNPV orthologue.
of the proteins, though a proportion of poxvirus ANK proteins lack the motif (Mercer et al., 2005; Sonnberg et al., 2008; Sperling et al., 2008; van Buuren et al., 2008; Werden et al., 2009). Here, we determined the prevalence of this C-terminal poxviral F-box motif in 328 chordopoxvirus ANK proteins by direct examination of the amino acid sequences. Of the 328 poxvirus ANK proteins, 219 contained a highly conserved C-terminal poxvirus F-box motif (Table 1, Supplementary Fig. S4, Supplementary Tables S2, S5, S6). We found significant differences in F-box prevalence among the different chordopoxvirus genera: OPV ANK proteins displayed the lowest F-box prevalence (mean 55%), followed by avipoxvirus, capripoxvirus, cervidpoxvirus, yatapoxvirus and PPV ANK proteins (Table 1). Of the ANK proteins encoded by leporipoxviruses and suipoxviruses, 100% contained a C-terminal F-box-like motif (Table 1, Supplementary Fig. S4). Most of the ANK proteins that lacked the poxviral F-box were encoded by OPV and showed evidence of gene fragmentation (Figs 1b, 2b, 3, 4). In several instances small consecutive ORFs encoded ANK domains and, if taken together, would constitute a coding region typical of a full-length ANK ORF. An example of this situation is provided by HSPV ORFs 013, 014 and 015, which together appear to form an orthologue of OPV group V and which in other OPVs are represented by a single large ORF such as CPX_BR 019 (Fig. 3, Supplementary Table S6). An exception in OPVs was orthologue group IX, where the full-length ANK proteins of 442–473 residues lacked the C-terminal F-box motif. When considering only ANK proteins of >400 aa lengths, the prevalence of the C-terminal poxviral F-box motif in OPV rises to 80% or higher, with the exception of VACV (Table 2).

Viruses of the genera Capripoxvirus, Leporipoxvirus, Yatapoxvirus, Suipoxvirus and Parapoxvirus encode a relatively low number of ANK proteins, almost all of which were greater than 400 residues in length and contained the C-terminal poxvirus F-box motif. The few ANK proteins of the LSG that lacked an F-box motif were truncated compared with their full-length orthologues or considerably shorter than the typical poxvirus ANK proteins (Fig. 1b) (Mercer et al., 2005). Among the 82 avipoxvirus ANK proteins, 24 lacked a C-terminal F-box, 11 of which were truncated or fragmented, while 13 were assumed to be full-length proteins (Fig. 4).

Therefore, almost all ANK proteins of greater than 400 residues in length contained a C-terminal poxvirus F-box motif, while those ANK proteins that lacked an F-box appeared to be either fragmented or truncated.

**DISCUSSION**

All chordopoxviruses except molluscum contagiosum virus encode multiple ANK proteins, ranging from four to more than 50 within each viral genome. Despite their large number, limited knowledge exists about the functions of poxvirus ANK proteins. Comparative analyses of the family of proteins have been complicated by the presence of multiple (typically seven) repeats of the ANK motif in any one protein and the presence of multiple ANK proteins encoded by any one virus. In this study, we examined the relationships between ANK proteins of 27 sequenced chordopoxviruses by comparing their amino acid sequences and determining, in a global perspective, the presence of orthologues across the different genomes. Paralogues encoded within the same genome displayed low levels of homology, suggesting variant functions. In contrast, orthologues were encoded by different viruses, and displayed high levels of homology, consistent with carrying out the same or very similar functions for the different viruses. We detected 29 orthologue groups among the ANK proteins of mammalian poxviruses, and a further 23 likely orthologues among avipoxviruses.

ANK protein orthologues of OPV displayed high levels of amino acid identities consistent with the high level of interspecies relatedness in this genus (Gubser et al., 2004; Jarmin et al., 2006). The three CPX strains examined in this study encoded the largest set of ANK proteins among the 12 OPVs. CPX contains the largest genome of OPV and has been theorized to be more closely related to the ancestor of the genus than any other sequenced OPV (Gubser et al., 2004; Shchelkunov et al., 1998). It has been proposed that CPX_GRI should be considered a separate species from CPX_BR and CPX_GER (Gubser et al., 2004). Nonwithstanding, CPX_GRI remains closely related to the OPV ancestor and encodes a near-full set of ANK proteins, though CPX_BR encodes one additional ANK ORF. The nine remaining OPVs encoded different subsets of the 15

<table>
<thead>
<tr>
<th>Species</th>
<th>Genus</th>
<th>F-box prevalence (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>In all ANK proteins</td>
</tr>
<tr>
<td>CML</td>
<td>54</td>
<td>86</td>
</tr>
<tr>
<td>CPX_BR</td>
<td>82</td>
<td>88</td>
</tr>
<tr>
<td>CPX_GER</td>
<td>74</td>
<td>87</td>
</tr>
<tr>
<td>CPX_GRI</td>
<td>88</td>
<td>88</td>
</tr>
<tr>
<td>ECT</td>
<td>75</td>
<td>78</td>
</tr>
<tr>
<td>HSPV</td>
<td>36</td>
<td>55</td>
</tr>
<tr>
<td>MPX_USA</td>
<td>73</td>
<td>80</td>
</tr>
<tr>
<td>MPX_Z</td>
<td>73</td>
<td>80</td>
</tr>
<tr>
<td>TATV</td>
<td>39</td>
<td>83</td>
</tr>
<tr>
<td>VACV</td>
<td>19</td>
<td>60</td>
</tr>
<tr>
<td>VARV_B</td>
<td>40</td>
<td>80</td>
</tr>
<tr>
<td>VARV_G</td>
<td>36</td>
<td>80</td>
</tr>
</tbody>
</table>

* >400 aa length.
CPX ANK proteins, consistent with a model of the ongoing adaptation of these viruses to new hosts occurring by gene loss from a CPX-like ancestor (Hendrickson et al., 2010).

The ANK protein orthologues of the LSG are of particular interest as they span five genera. Capripoxviruses, leporipoxviruses, yatapoxviruses, SWPV and deepox virus (DPV) have previously been reported to be more closely related to each other than is typical of the inter-genus relationships among the remaining four chordopoxvirus genera (Gubser et al., 2004; Hendrickson et al., 2010; Jarmin et al., 2006). The patterns of ANK protein ORFs in the genomes of the LSG viruses suggest that the six paralogues of DPV may represent the complete set of ANK proteins of the LSG-ancestor virus in a manner similar to that seen with CPX and other members of the genus Orthopoxvirus, with the possibility of a seventh parologue (LSG group I) that has only been detected in leporipoxviruses. This parologue may, however, represent a gene duplication event unique to leporipoxviruses. DPV isolates were previously shown to cluster with other LSG genera, and were suggested to represent a separate genus among the chordopoxviruses (Afonso et al., 2005), recently termed cervidpoxvirus (http://www.ictvonline.org), an observation and interpretation that is also reflected in our phylogenetic examination of DPV ANK proteins.

Representatives of two PPV species, OV (strains NZ2, IA82 and SA00) and BPSV (strain BV_AR02), were examined and displayed five orthologue groups. The recently published sequence of a third PPV species, pseudocowpox virus, revealed five ANK proteins, which are orthologues of the five OV ANK proteins, but does not contain orthologues of the two additional BPSV ANK proteins (Hautaniemi et al., 2010). However, further orthologues may be detected once more sequences of PPVs become available.

The avipoxvirus ANK protein orthologue groups identified in this study are likely to be incomplete due to the lack of sequencing data available for members of this genus and the high level of divergence between the two available species: FWPV and CNPV (Jarmin et al., 2006; Tulman et al., 2004). Using amino acid percentage identities and phylogenetic analysis, a number of candidate ANK protein orthologues could nevertheless be identified between the two species, both at the terminal regions of the genomes, and in internal areas that have previously been described to be the sites of rearrangement events within either genome (Gubser et al., 2004; Tulman et al., 2004). Due to the lack of studies of these ANK proteins, and of avipoxviruses in general, this is only the first step in examining the importance and function of these proteins. The maintenance of these proteins in such significant numbers (31 ANK protein ORFs in FWPV and 51 in CNPV) alone warrants their further investigation.

Recently, a second domain was identified in poxvirus ANK proteins: the C-terminal poxviral F-box. Using previous annotations and new examinations of each of the amino acid sequences, we determined that this truncated F-box is present in the majority of chordopoxvirus ANK proteins but that its prevalence varies between genera. OPV displayed the lowest level of F-box prevalence and also the highest level of ORF fragmentation. A considerable number of OPV ANK ORF give the appearance of having been separated into several fragments by the insertion of stop codons, with only one of the fragments containing a poxviral F-box, thereby reducing the F-box prevalence in this group. This fragmentation of OPV ORFs near the left and right termini of the genomes has previously been observed (Gubser et al., 2004; Gubser & Smith, 2002; Shchelkunov et al., 1998) and these regions include the locations of many of the ANK ORFs. These gene fragments are not thought to encode functional proteins and may be the result of the ongoing and relatively recent speciation (Gubser et al., 2004), however, unusually short ANK protein ORFs may not automatically be considered non-functional: the orthologues of OPV group X, which spanned only ~285 residues, includes VACV (Copenhagen) protein K1L, a known host range mediator essential for this virus’ replication in several cell lines (Bradley & Terajima, 2005; Perkus et al., 1990; Shisler & Jin, 2004). Therefore, the mere appearance of ORF fragmentation or truncation should not preclude functional studies of these putative proteins, especially since several poxvirus ANK proteins appear to have evolved functions independent of the poxviral F-box. Members of the OPV orthologue groups IX and X, and LSG group II generally show high levels of sequence conservation and are not fragmented but uniformly lack a poxviral F-box. Additionally, several poxvirus ANK proteins have been found to be able to interact with SCF1 through their poxviral F-box motif, but to also carry out functions independent of this domain (Chang et al., 2009; Johnston et al., 2005; Mohamed et al., 2009a, b; Werden et al., 2007, 2009).

Our analysis reveals that the large number of ANK proteins encoded by chordopoxviruses form distinct orthologue clusters either within genera (Orthopoxvirus, Parapoxvirus and Avipoxvirus) or across genera (Leporipoxvirus, Capripoxvirus, Suipoxvirus, Yatapoxvirus and Cervidpoxvirus). This pattern is consistent with previous, global comparisons of the inter-relationships of the nine chordopoxvirus genera (Gubser et al., 2004; Xing et al., 2006). It is likely that these orthologues will have at least similar functional roles and similar binding partners.

It is clear that extensive ANK gene duplication has occurred at some stage(s) in the evolution of these viruses and a number of different scenarios can be considered to have generated the array of poxvirus ANK proteins we have described. A possibility that we favour as most consistent with the data and the most parsimonious is that an ancestral ANK gene was acquired by an ancestor virus common to all four lineages of chordopoxviruses, but that duplication of the ANK gene occurred within each of the four lineages after they diverged from each other. Alternatively, it is also possible that the ancestor virus of each lineage independently acquired an ANK gene in four separate acquisition events after divergence of the lineages.
However, most of these proteins contain in addition to the ANK domain, a C-terminal truncated F-box motif. The apparent lack of such a domain combination in vertebrate genomes suggests that this combination was assembled in the virus(es) and would therefore require that this combination event occurred on four separate occasions. In the case of molluscipoxvirus, which does not encode any ANK proteins, it seems likely that the ancestral ANK/F-box gene was lost during development of the lineage. Alternatively, acquisition of the ancestral ANK/F-box gene may have occurred after divergence of the molluscipoxvirus lineage, however, the avipoxvirus lineage, which contains ANK/F-box proteins, is thought to have diverged prior to molluscipoxvirus (Xing et al., 2006). In some cases these ANK proteins have lost their poxviral F-box motif either through truncation of the ORF, or through gene fragmentation.

ANK-encoding genes have now been detected in a number of viruses other than poxviruses including specific members of the families Polydnaviridae, Iridoviridae, Phycodnaviridae and Mimiviridae (Kroemer & Webb, 2005; He et al., 2001; Lu et al., 1995; Raoult et al., 2004). We have not examined these proteins for any possible relationship between them and the poxvirus ANK proteins, although their lack of an F-box domain might suggest they arose independently from the poxvirus proteins. Further comparative analysis of the ANK proteins of these large DNA viruses will be of interest and along with further functional analysis of the poxvirus ANK proteins will add to our understanding of the roles of these proteins in viral pathogenicity in different hosts, host restriction and virus evolution.

**METHODS**

**Chordopoxviruses examined.** The genome sequences examined (Table 1) are available at www.poxvirus.org. Due to the controversy surrounding the classification of the three sequenced CPX strains [Brighton Red (CPX_BR), Germany-91 (CPX_GER) and GRI-90 (CPX_GRI)] all were included in the analysis (Guber et al., 2004). Three strains of OV were examined due to the large natural diversity within this species (Mercer et al., 2002, 2006). Two strains of VARV, Bangladesh-1975 (VARV_B) and Garcia-1966 (VARV_G), were included to represent variola major and variola minor, respectively (Massung et al., 1994; Shchetkunov et al., 2000). Two strains of MPX were examined: strain USA-2003-44 (MPX-USA) was isolated from a prairie dog, while strain Zaire-1979-005 (MPX-Z) was isolated from a severely ill patient (Likos et al., 2005). Two strains of vaccinia virus were examined, strain Copenhagen (VACV) is a laboratory strain (Goebel et al., 1990), while horsepox virus (HSPV) was isolated from naturally infected Mongolian horses in 1976 (Tulman et al., 2006).

**Identifying ANK domain-containing proteins.** Annotations available through the public online database www.poxvirus.org along with the Pfam protein prediction program (Finn et al., 2006) were used to identify ORFs encoding ANK proteins.

**Phylogenetic analyses.** The amino acid sequences of ANK proteins were aligned in MEGAALIGN version 8.0.2 (13) PowerPC, DNASTAR; Clewley & Arnold, 1997) using CLUSTAL W (gap penalty 10, gap length penalty 0.20, delay divergent sequence percentage 30, Gonnet series protein weight matrix) (Thompson et al., 1994). Neighbour-joining phylogenetic trees were generated using the Kimura distance formula and bootstrapped at default settings of 1000 trials and a random seed of 111.

**Identifying F-box domains.** All ANK protein sequences were visually examined for the presence of a C-terminal F-box using three F-box consensus sequences as guides (Sonnenberg et al., 2008) adapted from Kipreos & Pagano (2000); (Willems et al., 2004; Schulman et al., 2000).

**ACKNOWLEDGEMENTS**

The authors wish to thank Norihito Ueda, Lyn Wise and Otto Hayink (University of Otago), for helpful discussions and technical advice. This research was supported by funding from the Health Research Council of New Zealand and the University of Otago.

**REFERENCES**


ectromelia virus, the causative agent of mousepox. Virology 317, 165–186.


Phylogeny of poxvirus ankyrin-repeat proteins


