Genomic sequence of a clonal isolate of the vaccinia virus Lister strain employed for smallpox vaccination in France and its comparison to other orthopoxviruses

Aude Garcel, Jean-Marc Crance, Robert Drillien, Daniel Garin and Anne-Laure Favier

Correspondence
Daniel Garin
Daniel.Garin@wanadoo.fr

1Laboratoire de Virologie, CRSSA Emile Pardé, La Tronche, France
2IGBMC, CNRS, UMR 7104, Inserm U 596, F-67400 Illkirch, France
3Université Louis Pasteur, F-67000 Strasbourg, France

Received 13 November 2006
Accepted 5 March 2007

Since 1980 there has been global eradication of smallpox due to the success of the vaccination programme using vaccinia virus (VACV). During the eradication period, distinct VACV strains circulated, the Lister strain being the most commonly employed in Europe. Analysis of the safety of smallpox vaccines has suggested that they display significant heterogeneity. To gain a more detailed understanding of the diversity of VACV strains it is important to determine their genomic sequences. Although the sequences of three isolates of the Japanese Lister original strain (VACV-LO) are available, no analysis of the relationship of any Lister sequence compared to other VACV genomes has been reported. Here, we describe the sequence of a representative clonal isolate of the Lister vaccine (VACV-List) used to inoculate the French population. The coding capacity of VACV-List was compared to other VACV strains. The 201 open reading frames (ORFs) were annotated in the VACV-List genome based on protein size, genomic localization and prior characterization of many ORFs. Eleven ORFs were recognized as pseudogenes as they were truncated or fragmented counterparts of larger ORFs in other orthopoxviruses (OPVs). The VACV-List genome also contains several ORFs that have not been annotated in other VACVs but were found in other OPVs. VACV-List and VACV-LO displayed a high level of nucleotide sequence similarity. Compared to the Copenhagen strain of VACV, the VACV-List sequence diverged in three main regions, one of them corresponding to a substitution in VACV-List with coxopox virus GRI-90 strain ORFs, suggestive of prior genetic exchanges. These studies highlight the heterogeneity between VACV strains and provide a basis to better understand differences in safety and efficacy of smallpox vaccines.

INTRODUCTION

In the past, smallpox was a deadly human disease because of infection by variola virus (VARV). Complete eradication of smallpox was announced in 1980 subsequent to the WHO sponsored global vaccination programme using vaccinia virus (VACV) (Breman & Arita, 1980; Fenner, 1982). VACV and VARV both belong to the genus Orthopoxvirus (OPV) within the family Poxviridae and encode numerous common antigens, a fact that accounts for the cross protection observed upon smallpox vaccination. In recent years, VACV has attracted renewed attention as it could once again be used in mass vaccination against smallpox, due to the rise in bioterrorism threats around the world (Henderson, 1999a, b; Levy-Bruhl & Guerin, 2001; Mahalingam et al., 2004).

OPVs are large double-stranded DNA viruses (~180–225 kbp) encompassing about 200 open reading frames (ORFs). Many OPV proteins are involved in transcription or replication of the viral genome, and assembly of viral particles in the cytoplasm. A number of other proteins are involved in the inhibition of the host antiviral defence. There still remain some ORFs whose functions in the viral life cycle are unknown. Due to the complexity of poxvirus genomes, their intricate life cycle and their ability to successfully disrupt strategic steps in the host antiviral response (McFadden et al., 1995; Nash et al., 1999), questions concerning immunogenicity, virulence and host range of VACV strains have only been partially answered.
VACCINE LISTER STRAIN TO OTHER ORTHOPOXVIRUSES

Poxxvirus genome sequences can help in understanding virulence and immunogenicity particularly through the identification of spontaneous genomic modifications such as gene deletions and mutations that may entail distinct properties.

Complete genomic sequences have been reported for a number of OPVs, which are available online both at www.ncbi.nlm.nih.gov/GenBank and www.poxvirus.org. So far the sequences of ten VACV strains are public: Copenhagen (VACV-COP) (Goebel et al., 1990), two modified vaccinia Ankara isolates (VACV-MVA) (Antoine et al., 1998), Western Reserve (VACV-WR) (J. J. Esposito and others, unpublished), Tan-Tian (VACV-TIA) (Q. Jin and others, unpublished), the Japanese Lister original strain (VACV-LO) and two thermosensitive clones of VACV-LO (LC16mO and LC16m8) (Morikawa et al., 2005), the DryVax derived isolate VACV-3737 (R. K. Wilson and others, unpublished), the Acambis clone 2000 strain (J. J. Esposito and others, unpublished), the Duke strain (Li et al., 2006) and the rabbitpox virus Utrecht strain (RPXV-UTR) whose sequence demonstrated that it is a genuine VACV strain (Li et al., 2005). Very recently, the horsepox virus genome (horsepox virus MNR-76 strain, HSPV-MNR-76) was sequenced and found to be closely related to VACVs and RPXV (Tulman et al., 2006). We also felt that it would be valuable to obtain the genomic sequence of the VACV Lister strain (VACV-List), commonly used in Europe during the smallpox eradication period. As this strain was initially produced at the Lister Institute located in Elstree in the UK it is also sometimes referred to as the Elstree strain. Available records document a VACV-List stock established in the UK in 1961 on sheep and then a master seed stock prepared by two passages on calves. A master seed stock of calf lymph was sent to centres in Paris, Tokyo, Atlanta and Moscow and was also provided to manufacturers worldwide. The VACV-List used in the work described here comes from a batch of vaccine prepared in 1978 at the Institut Mérieux in Marcy l’Etienne, France. This virus, obtained from the Lister Institute in 1965, was used for routine vaccine production by scarification of calves until manufacturing was stopped.

This study presents the genomic sequence and analysis of a clonal isolate of VACV-List. The data have enabled comparison of the ORFs of the VACV-List genome to that of other VACV strains including the Japanese Lister strains that have been previously sequenced and annotated (Morikawa et al., 2005). The results highlight a number of major sequence differences within the isolates of the vaccinia virus species and provide a better understanding of the relationship between VACV-List and other OPVs.

**METHODS**

**Cells and viruses.** MRC-5 cells (diploid human lung fibroblasts; BioMerieux Laboratories) were grown in RPMI 1640 medium (Gibco). BHK-21 cells (hamster kidney cells, ATCC CCL 10) were grown in Glasgow minimum essential medium (Gibco) supplemented with 10% tryptose phosphate (Sigma) and 50 mM HEPES (Gibco). Vero cells (African green monkey kidney cells, ATCC CCL 81) were grown in minimal essential medium (Gibco). For cell culture, all media were supplemented with 10% fetal bovine serum (FBS), whereas virus infections were supplemented with 5% FBS and 1% penicillin-streptomycin (Gibco). The French smallpox vaccine Lister strain (X55-33), provided by Sanofi-Pasteur-France, was cloned by five plaque purifications on the MRC-5 cell line. Clones were propagated and tittered on MRC-5 cell monolayers. Phenotypic studies (plaque size, replication kinetics and yield) were first performed in vitro on 120 clones. Ten of these clones that displayed phenotypes typical of the diversity in the 120 clones were then selected for virulence studies performed in vivo (neuropathogenesis and virus replication in the brains of suckling mice after intra-cranial injection) and finally one clone (VACV-List-107) was selected as the reference strain because of its similarity to the uncloned vaccine (A. Garcel, J. Perino, J. M. Crance, R. Drillien, D. Garin and A. L. Favier; unpublished data). VACV-List-107 was produced on BHK-21 cells and purified using standard procedures described previously by Lee et al. (2001).

**DNA sequencing.** DNA from VACV-List-107 was cleaved by Hydroshear (GeneMachines). Using the Klenow and T4 DNA polymerases (Promega) blunt-ended DNA fragments were produced and cloned into pUC18 previously digested by Smal. Plasmid DNA extractions were performed using 96-multwell Multiscreen Plasmid (Millipore) by alkaline lysis followed by filtration. Fragments were sequenced by the Sanger method (Sanger et al., 1977) on an Applied Biosystems model 3730 Sequencer, using plasmid, primers and BigDye Terminators premix (Applied Biosystems) according to manufacturer’s protocol (Lee et al., 1997). At the end of the acquisition, raw data were analysed automatically by the Sequencing Analysis 5.0 Applied Biosystems software. Primer walking was performed to fill gaps and to confirm the order of the pre-assembled contigs. The terminal hairpin loops have not been sequenced. A major sequence of 184 030 bp in length and two minor contigs of 3170 and 3771 bp around the left inverted terminal repeats (ITR) were obtained with a minimal fivefold redundancy for each base position, and were linked through their overlapping regions to obtain a final sequence of 189 421 bp.

**DNA sequence analysis.** Coding sequences (CDS) were identified with GenoAnnot software of the GenoStar platform (Chalifa-Caspi et al., 2003, 2004). Two PROKOV matrices were constructed with a complexity level (k-uptle) of 5 and a minimal CDS length of 90 and 120 bp. Then, the CDS were determined with a score threshold of 90.0. For confirmation, predicted CDS were compared to three databases (including the GenPept, SWISS-PROT, PIR, PDF, PDB and NCBI RefSeq ‘nr’ sequence database) using the National Center for Biotechnology Information (NCBI) BLASTX (with an E-value <1,10-5) (Altschul et al., 1990). Annotation was performed by comparison with NCBI BLASTP results using five protein sequence databases, including the poxvirus database, SWISS-PROT database, ‘cog’ database and ‘nr’ database. The BLASTP analysis was performed on the total CDS. An ORF was annotated ‘similar to’ if it contained more than 98% identical amino acids; otherwise it was annotated ‘related to’.

**Dotplot analysis of OPV strains.** Dotplots were created by the Dotter program with the GenBank nucleotide sequence of VACV-List plotted on the horizontal axis and the complete nucleotide sequences of either VACV-COP, variola virus India strain (VARV-IND), cowpox virus GRI-90 strain (CPXV-GRI) or monkeypox virus Zaire strain (MPXV-ZAI) plotted on the vertical axis (Sonnhhammer & Durbin, 1995).

**VACV-List comparison to VACV-LO.** Orthologues of VACV-List ORFs were searched by nucleotide sequence alignment in the VACV-LO sequence. VACV-LO ORFs were translated using the
Clone Manager program and compared to VACV-List orthologues. Amino acid differences found to lie within different amino acid groups are reported in Table 1 and Supplementary Table S2 (available in JGV Online). These amino acid groups are: hydrophobic amino acid (A, G, I, V and L), basic amino acid (R and K), acid amino acid (D and E), aromatic amino acid (T, F and W), neutral amino acid (S and T) and hydrophilic amino acid (N and Q).

**Phylogenetic analyses of multiple OPV DNA sequences.** Three phylogenetic analyses were performed. First, the nucleotide sequences of 17 genes present in 15 OPVs were aligned (Gubser et al., 2004). For each virus, genes were concatenated to form a single sequence with the following order: E9L, I7L, I8R, G9R, J3R, J6R, H2R, H4L, H6R, D1R, D5R, D6R, D11L, D13L, A7L, A16L and A24R according to the VACV-COP nomenclature. Secondly, nine full-length nucleotide sequences of different OPV genomes were aligned [VACV-List, VACV-LO, VACV-WR, VACV-3737, VACV-COP, CPXV-GRI, cowpox virus Brighton strain (CPXV-BR), RPXV-UTR and HSPV-MNR-76]. Third, the nucleotide sequence of genes involved in the interferon (IFN)–mediated response were concatenated to form a single sequence.

**Table 1. Main amino acid differences between VACV-List and VACV-LO**

Protein lengths in VACV-List (second column) and in VACV-LO (third column) are indicated. Amino acid substitutions are indicated by their position (fourth column). Peptide alignments containing amino acid deletions and mutations are in bold characters. Conserved mutations despite the VACV-LO nucleotide sequence polymorphism are in italics.

<table>
<thead>
<tr>
<th>Protein</th>
<th>List aa</th>
<th>LO aa</th>
<th>VACV-List versus VACV-LO mutations</th>
<th>Comments</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>List009</td>
<td>140</td>
<td>140</td>
<td>S29L</td>
<td>The secreted protein is an O-glycosylated protein. The S29L change could entail a lack of one glycosylation site in VACV-LO.</td>
<td>Lin et al. (1990)</td>
</tr>
<tr>
<td>List023</td>
<td>512</td>
<td>512</td>
<td>F99S; K283E</td>
<td>The VACV-LO protein contains a glutamic acid at position 283, as in VACV-COP, which is inside the kelch-domain (aa 261–307), whereas it is replaced by a lysine in List023.</td>
<td>Pires de Miranda et al. (2003)</td>
</tr>
<tr>
<td>List055</td>
<td>190</td>
<td>190</td>
<td>N11D</td>
<td>The mutation N11D observed appears in the first α-helix (11–68 aa), which is responsible for enzyme binding to Z-DNA.</td>
<td>Kahmann et al. (2004)</td>
</tr>
<tr>
<td>List091</td>
<td>333</td>
<td>333</td>
<td>G331S</td>
<td>The methyltransferase is modified (G331S) in the domain thought to bind to List098 (RAP94) in the RNA polymerase complex.</td>
<td>Mohamed et al. (2001)</td>
</tr>
<tr>
<td>List102</td>
<td>844</td>
<td>844</td>
<td>L236P; N325S</td>
<td>The mRNA triphosphatase is not altered in the active site or in the GTP-binding domain but the substitution of a proline at position 236 (which is encoded in other OPVs) to a leucine for VACV-List, could have an impact on the structure of the protein.</td>
<td>Niles &amp; Christen (1993)</td>
</tr>
<tr>
<td>List124</td>
<td>99</td>
<td>108</td>
<td>List 79-N—____________DGSNN-84</td>
<td>Three glycosylation sites were described in the intracellular mature virus membrane protein. The first site is deleted in the VACV-List protein, which is shorter than in VACV-LO, and the second site is absent due to an amino acid divergence with VACV-LO.</td>
<td>Yeh et al. (2000)</td>
</tr>
<tr>
<td>List124 LO 79-NANTNNDSNSNNSN-93</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>List152</td>
<td>221</td>
<td>221</td>
<td>N76D</td>
<td>The protein, corresponding to VACV-COP A36, was shown to be involved in CEV production, but D76, common among A36 of VACVs is replaced by N in VACV-List.</td>
<td>Herrero-Martinez et al. (2005)</td>
</tr>
<tr>
<td>List175 LO 248-YDTYNNDTVPPTT-261; 301-IYNK-304</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*List 244-DDADLHD-249; 257-S; 300-C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*LO 243-D—D-244; 252-P; 295-Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Indicates that amino acid difference is found only with respect to the polymorphic sequence.
sequence with the following order: C12L, N1L, K3L, E3L, B13R, A46R, A52R, B8R and B18R according to the VACV-COP nomenclature, and were aligned for the same nine viruses. Sequence alignment was performed by the BioEdit program using Blossum62 similarity matrix. Pairwise gap removal or global gap removal were performed to exclude gaps from alignments and construct the two concatenated trees and the full-length genome tree, respectively. For alignments, neighbour-joining and maximum-likelihood trees were constructed using the program Phylowin (Galtier et al., 1996) with a Kimura model. The robustness of trees was evaluated by bootstrap analysis with 1000 rounds of replication. As VACV-LO was not annotated, gene sequences were extracted from the complete GenBank genome using VACV-List start and stop sequences.

**Sequence availability.** The VACV-List genome sequence has been deposited in the GenBank database under the accession no. DQ121394. The other OPV genomes used for comparison or for phylogenetic analysis were obtained from GenBank: VACV-COP, M35027 (Goebel et al., 1990); VACV-LO, AY678276 (Moriwaka et al., 2005); VACV-WR, AY234312; VACV-3737, DQ377945; VACV-DUKE, DQ439815 (Li et al., 2006); variola virus Bangladesh strain (VARV-BSH), L22579 (Massung et al., 1994); VACV-IND, X69198 (Schelkunov et al., 1995); variola virus Garcia strain (VACV-GAR), Y16780 (Schelkunov et al., 2000); RPXV-UTR, AY484669 (Li et al., 2005); MPXV-ZAI, AF380138 (Schelkunov et al., 2002); ectromelia virus Moscow strain (ECTV-MOS), AF108285 (Chen et al., 2003); camelpox virus CMS strain (CMLV-CMS), AY009089 (Gubser & Smith, 2002); camelpox virus M-96 strain (CMLV-M-96), AF438165 (Afonso et al., 2002); CPXV-BR, AF482758; CPXV-GRI, X94355; HSPV-MNR-76, DQ792504 (Tulman et al., 2006).

**RESULTS AND DISCUSSION**

**VACV-List genome content and topography**

The VACV genome is a double-stranded DNA molecule with covalently joined termini (Geshelin & Berns, 1974) and telomeric regions that are identical but inverted in sequence, designated ITRs. For our analysis, the VACV-List strain, employed as a smallpox vaccine in France, was first cloned by several rounds of plaque purification in MRC-5 cells. Viral DNA was then extracted from one isolate and sequenced using VACV-List start and stop sequences. ORFs were numbered from 001 to 285 as defined, included ORFs larger in size than the smallest size of expressed proteins and/or if they were entirely localized within a larger ORF known to be expressed. As any set of criteria carries a degree of arbitrariness, some real genes might have been missed. Each of the 201 annotated ORFs was subjected to similarity searches by BLASTX analysis. ORFs were numbered from 001 at the left end to 201 at the right end and are listed in Supplementary Table S1 (available in JGV Online). The best BLASTP alignment result was selected to annotate the ORFs. The vast majority of ORFs, annotated ‘similar to’ (n=192), displayed more than 98 % amino acid sequence identity to previously identified ORFs. All other ORFs, annotated ‘related to’ (n=9), displayed between 70 and 98 % amino acid sequence conservation as compared to known ORFs. Proteins were classified by function and those with undefined functions were annotated ‘unknown’, whereas 53 ORFs for which there is no proof of protein expression were annotated ‘putative’. It could not be excluded that some ORFs classified in the putative gene group do not encode any protein, however, this group, by definition, included ORFs larger in size than the smallest ‘expressed’ ORFs. Historically, since the first VACV-COP genome annotation, some additional ORFs have been annotated (i.e. VACV-COP A-ORF-A). For the majority of these ORFs there is no evidence that they are expressed, and using the criteria defined above many of them have been rejected. The organization of VACV-List ORFs is the same as that described for other poxviruses (Gubser et al., 2004): ORF orientation in the ITR is generally directed towards the hairpin ends and central ORFs are directed in both orientations.

At least nine ORFs could be recognized throughout the VACV-List genome as pseudogenes (ITR genes were counted only once) because their coding regions were truncated or fragmented by deletions and stop codons in a manner such that they were smaller than orthologous ORFs found in other OPVs. Except for a few exceptions, these pseudogenes (labelled with an asterisk in Supplementary Table S1 and hereafter) are also pseudogenes in other VACV strains. Of the three pseudogenes localized in the ITR, List002* (List200*) encodes a severely truncated version of the tumour necrosis factor (TNF) receptor of CPXV (CrmB), whereas List003*A-B-C-D (List199*A-B-C-D) and List004*A-B (List198*A-B) encode smaller versions of proteins of unknown function in VARV and CPXV, respectively. List014*A-B-C-D-E encodes a fragmented version of a CPXV host range gene (V025) that is entirely deleted in one VACV strain (VACV-COP) and fragmented in a different manner in others. List141*A-B-C-D-E corresponds to an A-type inclusion protein (ATI) similar to A26L of CPXV-GRI (160 kDa protein) but truncated as in all VACV strains, as well as in VARV, CMLV and MPXV strains (Amegadzie et al., 1992; Meyer & Rziha, 1993). Nevertheless, the 725 aa N-terminal fragment could be expressed as an intracellular 94 kDa protein (de Carlos & Paez, 1991). List171* lacks an initiation codon that is...
found in its VACV-COP A54L counterpart although it is uncertain whether the latter is actually expressed in the virus life cycle. List190*A-B corresponds to a serine protease inhibitor (SPI-2) and its sequence, including one stop codon, is similar to the entire VACWR195 gene of VACV-WR or to the sum of VACV-COP B13R and B14R. In vitro, SPI-2 prevents the cleavage of pro-interleukin (IL)-1β into mature interleukin-1β by inhibiting interleukin-1β converting enzyme (Kettle et al., 1997). In the murine dermal ear model this protein influences the outcome of infection (Tscharke et al., 2002), while it has no effect in the murine intranasal model (Kettle et al., 1995). List194* encodes a shortened version of a type I IFN-binding protein found in most OPVs including some VACV strains. It should be mentioned that even if a gene is fragmented, the N-terminal portion of the protein may be expressed as described for VACWR148 of VACV-WR by Chung et al. (2006) and for a portion of the ATI protein of some VACV strains that can cross-react with antiserum (Patel et al., 1986) but in most cases these truncated versions are no longer functional.

The organization of the VACV-List genome is overall the same as other VACVs (Fig. 1). ORFs within the ITRs are of course in duplicate, while all others are unique. ORFs within the middle of the VACV-List genome are highly conserved among poxviruses with most of them being involved in essential replication and structural functions, whereas host-range genes responsible for virulence and immunomodulation are localized in the variable regions at both ends. This variability is most likely related to adaptation of each OPV species to its natural host.
Comparison of VACV-List to OPVs

The nucleotide sequence of VACV-List was aligned with the consensus sequence of VACV-LO. Alignment with this strain was performed preferentially because the nucleotide sequence of the clones derived from it, namely LC16mO and LC16m8, had changed after being selected for their thermostability and low virulence (Morikawa et al., 2005). The sequence of VACV-LO is highly polymorphic since it is not a clonal isolate and it consists of quasi-species of at least eight genotypes (Morikawa et al., 2005). Our analysis showed that VACV-List is more closely related to the VACV-LO consensus sequence than to LC16mO or LC16m8. VACV-LO was used as the main smallpox vaccine in Japan during the eradication period but LC16m8, which forms pockmarks at the site of inoculation and displays temperature sensitivity, was used in children during the 1970s because of its more favourable safety profile (Sugimoto et al., 1985). One hundred and forty three ORFs of LC16mO (the parental strain of LC16m8) showed differences with VACV-List (data not shown), while there are fewer (57 ORFs) differences between VACV-List and VACV-LO (Table 1). One striking example is List012* which is truncated in the majority of VACV strains except in LC16mO, LC16m8 and RPXV-UTR. The counterpart of List012* in ectromelia virus is a virulence factor that displays E3 ubiquitin ligase activity (Huang et al., 2004). DNA sequence identity between VACV-List and VACV-LO was found to be 99.6%. Deletions and mutations were not only detected within CDS but also within promoter sequences where they could alter transcription entailing different protein levels. The majority of the nucleotide differences in CDS resulted in no amino acid change. Those nucleotide changes or deletions responsible for amino acid differences, deletions or truncated proteins are reported in Supplementary Table S2 (available in JGV Online). Because of the nucleotide sequence polymorphism of VACV-LO, some of these differences with respect to the consensus sequence are of course not relevant with respect to minor viral populations of VACV-LO. Eight mutations reported in Table 1 are in a structural or functional domain that could potentially be responsible for a protein dysfunction. Insufficient information is available on the other proteins affected by amino acid changes to speculate on their potential consequence. Activity and structural studies would have to be performed to assess the potential role of any of the amino acid substitutions in the virus cycle.

BLASTP comparisons were performed during VACV-List genome annotation to identify the closest poxvirus protein to each ORF. The majority of VACV-List proteins are similar to VACV-COP proteins, but some proteins have a higher similarity with VACV-WR or with other OPV (CPXV, CMLV or RPXV) proteins (Supplementary Table S1 available in JGV Online). Many of the latter ORFs encode putative proteins, whereas some of them belong to virulence and immunomodulation protein families (List009, List013, List162, List164, List172, List185, List190*A-B, List194*, List195 and List201). Two of them are implicated in the replication complex (List106 and List138) and three are involved in virus assembly (List109, List125 and List142).

Recently, a new type of virokine has been identified by Alejo et al. (2006) that contains a smallpox virus-encoded chemokine receptor domain (SECRET domain). This SECRET domain is an independent chemokine-binding site found in both TNF receptor orthologues CrmB, CrmD and in other SECRET domain-containing proteins (SCP). As for other VACV isolates, the VACV-List genome contains CrmB (List002* or List200*) but not CrmD. Among the SCP genes, the VACV-List genome contains SCP-1 and SCP-3 genes (List007 and List184), found in other full-length VACV genomes. The gene encoding the SCP-2 protein (V014 of CPXV-BR) is not present in the VACV-List genome nor in other VACV isolates.

Dotplots were performed to study ORF deletions and rearrangements and obtain a large-scale comparison between VACV sequences and other OPVs. The VACV-List versus VACV-COP dotplot highlights three gaps (Fig. 2a): two are due to ORFs present in VACV-List that are lacking in the VACV-COP genome. The first gap (12 000–15 500 bp) occurs because VACV-List contains five supplementary ORFs: List011, List012* and List015 encoding putative proteins, List013 encoding an IL-18-binding protein and List014* encoding a pseudogene similar to the V025 gene of CPXV-BR. The second gap (137 000–142 000 bp) corresponds to the pseudogene of the AT1 protein (List141*A-B-C-D-E) for which only two out of five ORFs have been found in VACV-COP. The third gap (179 000–183 000 bp) corresponds to a short sequence encoding different proteins in VACV-List and VACV-COP: List195 and List196 replace VACV-COP B19R, B20R and B21R, respectively. Moreover, the right end of List194* and the left end of List199*A-B-C-D are lacking, indicating that this region is the result of recombination. The List195 gene is similar to K3R of CPXV-GRI and encodes the CrmE TNF receptor that contributes to virulence (Saraiva & Alcami, 2001; Reading et al., 2002). This gene is present in the European and Japanese Lister strains as well as in the USSR and Evans strains (Reading et al., 2002) but is not found in any other sequenced VACV strain, nor in the RPXV genome. The List196 ORF is similar to CMP6L of CMLV-CMS. This ORF is also present in CPXV-GRI (as TIR) and no potential function has been described for the putative protein. Surprisingly, BLASTP alignment underlines a high amino acid similarity (more than 70% identity and more than 80% similarity) over the entire length of the 237 aa List196 sequence with a transmembrane Bax inhibitor motif-containing protein (238 aa) found in the genomes of several mammalian species including Homo sapiens. Such high sequence similarity with cellular genes suggests that the viral gene was captured from the cell genome more recently in the evolution of OPVs than other viral genes. This hypothesis is further supported by the presence of an upstream G+C rich sequence unusual in an OPV promoter. A VACV-List recombinant with a green fluor-
escent protein tag at the 3’ end of List196 failed to display fluorescence in infected cells suggesting that List196 is a pseudogene (Matthias Hebben, personal communication). Very recently, it has been shown that List196 encodes an apoptosis inhibitor (Gubser et al., 2007). The List013 and List195 ORFs encode proteins that play a role in virulence (an IL-18-binding protein for List013 and a TNF receptor for List195) (Smith et al., 2000; Reading et al., 2002) that are not found in the VACV-COP genome, whereas VACV-COP encodes an IFN-binding protein (B19) not found in VACV-List genome. In addition, a similar gap (corresponding to 179 000–183 000 bp of VACV-List) was detected using the VACV-List versus the VACV-WR dotplot (data not shown). Comparison of this region (from List194* to List199*A-B-C-D) between OPVs is presented in Fig. 3. Overall, VACV-List displays unique features in this region as compared with other OPV genome sequences. In particular, it is the only VACV strain containing orthologues of both the CPXV-GRI K3R gene (List195) and the CMLV-CMS CMP6L gene (List196). This observation could reflect either recombination events between VACV-List and the other OPVs or a distinct genealogical filiation.

VACV-COP is most closely related to VACV-List except for two ORFs (List195 and List196) that replace VACV-COP B19R, B20R and B21R as has already been noted from the dotplot comparison. The dotplot alignment performed between VACV-List and VACV-L0 (data not shown) showed that no ORFs were missing between the two genomes. Dotplots comparing VACV-List to CPXV-GRI, which is more closely related to VACV than CPXV-BR, revealed three major gaps, whereas in the case of MPXV-ZAI and VARV-IND six and seven major gaps were detected, respectively (Fig. 2b, c, d). A more detailed examination of the gap regions is presented as Supplementary material (available in JGV Online).

In conclusion, differences observed among the OPVs analysed result in clusters of ORF rearrangements corresponding to approximately 8% of the VACV-List genome. Ultimately, the distinction between OPV genomes should also be documented by differences in their proteomes.

**Phylogenetic relationships**

To establish the phylogenetic relationship of VACV-List with other OPVs, we have compared multiple nucleotide sequences conserved in several genomes. Seventeen conserved genes were selected as described by Gubser et al. (2004) and alignments were made from concatenated
sequences (Fig. 4a). Three clusters are apparent in the genus Orthopoxvirus. VARVs (VARV-IND, VARV-BSH and VARV-GAR) are clustered and closely related to CMLVs (CMLV-CMS and CMLV-M-96). ECTV-MOS, CPXV-BR and MPXV-ZAI have diverged and form an ancestral OPV cluster. As described by Gubser et al. (2004), the two strains of CPXVs, CPXV-GRI and CPXV-BR, are not closely clustered. The last cluster contains VACV strains such as VACV-List and VACV-LO and includes HSPV-MNR in the same branch.

Among virokines involved in immune evasion, proteins that interfere with the interferon pathway are the most representative in VACVs. To perform a phylogenetic analysis using this family of genes, the nine IFN virokine genes were selected to produce a nucleotide concatemer: IFN soluble-binding proteins (B8 and B18) (Alcami & Smith, 1995; Alcami et al., 2000), inhibitors of Toll-like receptor signalling (A46 and A52) (McCoy et al., 2005; Stack et al., 2005), an inhibitor of I-kB degradation (N1) (DiPerna et al., 2004), inhibitors of IL-18 (C12 and presumably B13) (Kettle et al., 1997; Tone et al., 1997; Reading & Smith, 2003) and inhibitors of IFN-induced protein activity (E3 and K3) (Langland & Jacobs, 2002). Distinct branches were obtained for CPXVs and VACVs, while VACV-COP and VACV-3737 deviated from the other VACV strains (Supplementary Fig. S2 available in JGV Online). As found in the previous alignment, HSPV-MNR-76 was associated with the group containing VACV strains. Although VACVs are clustered together according to this analysis, it has been reported that these strains do not all express the virokines at the same level (Alcami et al., 2000; Symons et al., 2002). To confirm actual differences in IFN-related responses among VACVs strains, comparative proteome analyses and activity profiles would be required.

The full genomic sequences of several VACVs (including RPXV-UTR), HSPV and CPXVs were more closely compared to examine the relationship between CPXV, HSPV and VACV strains (Fig. 4b). This is particularly interesting in view of the fact that during the 19th century VACV is thought to have replaced CPXV or possibly HSPV as a smallpox vaccine. Two major groups are distinguished in this analysis. The first one encompasses CPXV strains and the HSPV strain, while the second one comprises VACV strains (including RPXV strain). VACVs are separated into three branches: one branch groups VACV-List, VACV-LO, VACV-COP and RPXV-UTR; a second branch

Fig. 3. Comparison of OPVs in the region from List194* to List199*A-B-C-D. The ORFs transcribed rightward and leftward are presented above and below the horizontal centreline, respectively. Boxes are coloured identically when the ORFs match. Uncoloured boxes correspond to ORFs that do not match with any VACV ORF in the region described. ORFs are named according to the nomenclature adopted for each virus strain (www.poxvirus.org).
is composed of VACV-WR; and the third composed of VACV-3737, is the closest to the CPXV branch. The relationship described by Li et al. (2005) between the RPXV-UTR and VACVs is confirmed. HSPV-MNR-76 shares a closer relationship with CPXV-GRI than CPXV-BR, the latter being clearly distinct as previously mentioned. Interestingly, comparison of full-length genomes indicates that HSPV is not as closely related to VACVs as previously suggested by comparison of the 17 conserved genes (Fig. 4a) and the central conserved genomic regions (Tulman et al., 2006). Recombination events occurring with different OPVs (VACV and CPXV strains) and involving not only the core region but also the variable regions, may account for these findings. Whatever the mechanisms involved, the different methods employed suggest that HSPV constitutes an OPV clade at the crossroads between VACV and CPXV clades. Finally, our phylogenetic analysis also confirms that the two Lister strains, VACV-LO and VACV-List, are very similar.

**Conclusion**

In this report we have analysed and compared the genome structure and the potential coding regions of VACV-List to other VACV strains previously used as smallpox vaccines during the eradication period. The VACV-List genome is of course a typical OPV, including a central conserved region, variable left and right end regions and ITRs with tandem repeats. At least eight VACV-List genes contain stop codons in their CDS as compared with larger ORFs in other OPVs and were considered as pseudogenes. It would nevertheless be worthwhile for some of the larger pseudogenes to confirm that the truncated versions they encode are no longer functional and possibly correlate this with attenuation. The comparison of the two VACV Lister strains, VACV-List and VACV-LO, demonstrated a high nucleotide sequence similarity even if it cannot be excluded that single amino acid differences in some of their proteins may distinguish the two strains in precise properties. LC16mO and LC16m8, both clonal isolates derived from VACV-LO, appeared to be more distant from VACV-List than VACV-LO. Interestingly, the VACV-List genome contains several ORFs that have not been annotated in other VACVs: both List180 and List196 are found in CMLV-CMS; one ORF (List173) is also found in CPXV-BR and one ORF (List195) is also found in CPXV-GRI (K3R). Among these ORFs, List196 is suspected to be a gene of cellular origin that could have been incorporated into VACV-List or a parental CPXV relatively recently upon divergence of OPVs. In fact, the 4 kbp region overlapping List194 to List199*A-B-C-D is distinct in VACV-List compared to the other VACV strains so far sequenced and appears to correspond to a substitution with CPXV-GRI or CMLV-CMS ORFs suggesting prior genetic exchanges as previously postulated by Reading et al. (2002). This region accounts for one distinctive feature of VACV-List, in common with the USSR and Evans strains, namely that it encodes two active TNF receptors (List172 and List195), while it lacks an IFN inhibitor (VACV-COP B19R) found in other VACV strains. In this way, the Lister strain is better armed against the host TNF response, a property

![Fig. 4. OPV phylogenetic relationship. Phylogenetic analysis of OPV strains obtained by the maximum-likelihood (a) and the neighbour-joining (b) methods comparing nucleotide sequences, using a global gap removal distance method option. The bootstrap values from 1000 replica samplings and the divergence scale (substitution per site) are indicated. (a) The tree was applied on a concatenated sequence of 17 genes. (b) Phylogenetic analysis of VACV, HSPV and CPXV strains performed on the complete nucleotide genome sequence.](image-url)
that could potentially have an impact on Lister strain virulence and immune evasion. A recent reappraisal of historical data suggests that the frequency of adverse events after vaccination with the Lister strain was considerably higher than for the NYCBH strain (Kretzschmar et al., 2006). Differences between VACV genomes, particularly strategies adopted to escape the immune response, could explain these observations although it cannot be excluded that biases in the data may account for apparent variations in strain safety. Despite some notable differences, phylogenetic analysis of the VACV isolates underlines their particularly close relationship but does not allow the precise definition of a genealogical tree.

ACKNOWLEDGEMENTS

The work was supported by grants from Sanofi-Pasteur, the Service de Santé des Armées (SSA), the Délegation Générale pour l’Armement (DGA) and from the association ARAMI. We thank D. Dimova and T. Vermat for assistance with annotation, E. Pellegrini for creation of the genome map figure, M. Hebben, B. and M. Souberbille for critical reading of the manuscript.

REFERENCES


