28

MOLECULAR AND EVOLUTIONARY ANALYSIS OF A GENE CONSERVED IN MOST *ORTHOPOXVIRUSES*

Nicola Jennifer Douglass

Thesis presented for the degree of DOCTOR OF PHILOSOPHY in the Department of Medical Microbiology, University of Cape Town.

August 1993

The University of Cape Town has been given the right to reproduce this thesis in whole or in part. Copyright is held by the author.

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

University of Cape Town

for John,

CONTENTS

Acknowledgements IV

Abbreviations V

List of Figures VI

List of Tables VIII

Abstract IX

CHAPTER 1
GENERAL INTRODUCTION 1

CHAPTER 2
MATERIALS, METHODS AND SEQUENCING STRATEGIES 23

CHAPTER 3
INDEPENDENT EVOLUTION OF VARIOLA AND MONKEYPOX VIRUSES 38

CHAPTER 4
MONKEYPOX VIRUS EVOLUTION 48

CHAPTER 5
INTERRELATIONSHIPS WITHIN THE ORTHOPOXVIRUS *GENUS* 58

CHAPTER 6
CHARACTERIZATION OF VACCINIA VIRUS E5R 85

CHAPTER 7
GENERAL CONCLUSIONS 97

LITERATURE CITED 101

APPENDIX 117

ACKNOWLEDGEMENTS

The SAIMR and Polio Research Foundation are gratefully acknowledged for financial support.

The following people are acknowledged:

1. Prof. Dumbell, my supervisor, for all he has taught me;
2. Prof. Moodie, for giving me the opportunity to do this work, and for proof-reading the manuscript;
3. Anna-Lise Williamson, for her interest in my work, and encouragement;
4. Lafras Steyn, for his help with the computer work;
6. Ed Rybicki, for providing the TREECON package and helping me with the phylogenetic analysis;
7. Margie, for her help with the references;
8. Anne, for the loan of her computer;
9. Damian, for layout of the final manuscript;
10. All in the lab, especially the "pox" group, who have created a most pleasant and stimulating work atmosphere; with special thanks to Helen and Fernanda;
11. My family and friends, especially my parents, Tim and Diani, Margie and Gina for their support and encouragement; and, finally,
12. John, my husband, for his understanding, tolerance and confidence in me.

ABBREVIATIONS

α	alpha
β	beta
β-gal	beta-galactosidase
μg	microgram
%	percentage
>	greater than
A	adenine
AA	amino acid
ASF	African Swine Fever
ATP	adenosine triphophate
bp	base pair(s)
C	cytosine
CAM	chick chorioallantoic membrane
CHO	Chinese Hamster Ovary
CPE	cytopathic effect
CTP	cytidine triphosphate
Da	dalton
dC	2'-deoxyribocytidine
DNA	deoxyribonucleic acid
dT	2'-deoxyribothymidine
dTMP	2'-deoxyribothymidine monophosphate
dTTP	2'-deoxyribothymidine triphosphate
fig.	figure
FPV	fowlpox virus
G	guanine
gpt	xanthine-guanine phosphoribosyl transferase
HA	haemagglutinin
kb	kilobases
kD	kilodalton
LD_{50}	lethal dose (50%)
ml	millilitre
MPA	mycophenolic acid
ORF	open reading frame
PCR	polymerase chain reaction
PEK	pig embryo kidney
pfu	plaque forming units
pI	isoelectric point
RK	rabbit kidney
RNA	ribonucleic acid
T	thymine
TK	thymidine kinase
TTP	thymidine triphosphate

The amino acid abbrevations are listed in appendix C

LIST OF FIGURES

Fig. 1.1	*HindIII* genomic maps of 8 of the 10 recognized *Orthopoxvirus* species.	6
Fig. 2.1	Construction of plasmid pHar Rsa.	28
Fig. 2.2	*HindIII* map of vaccinia virus showing the location of the E5R gene.	28
Fig. 2.3	Construction of plasmid pGptC/B for interruption of vaccinia virus E5R.	29
Fig. 2.4	Strategy for determining the DNA sequence of variola virus Harvey in the region corresponding to vaccinia virus E5R	32
Fig. 2.5	Strategy for determining the DNA sequence of variola virus Somalia in the region corresponding to vaccinia virus E5R	32
Fig. 2.6	Strategy for determining the DNA sequence of monkeypox virus Denmark in the region corresponding to vaccinia virus E5R.	33
Fig. 2.7	Strategy for determining the DNA sequence of the vaccinia virus Dairen E5R region.	33
Fig. 2.8	Diagram showing the positions to which the primers bind, using vaccinia virus Dairen as a template.	36
Fig. 3.1	*HindIII* maps of variola and monkeypox virus genomes.	40
Fig. 3.2	Restriction enzyme maps of variola virus *HindIII* D and monkeypox virus *HindIII* E.	41
Fig. 3.3	DNA sequence comparison of variola virus, Harvey (1260bp) and monkeypox virus, Denmark (820bp).	43
Fig. 3.4	Schematic representation of the coding potential of variola virus, Harvey and monkeypox virus, Denmark in the region sequenced.	44
Fig. 3.5	DNA sequence in the region of the two 24bp deletions.	44
Fig. 3.6	Slot blot of monkeypox virus genomic DNA probed with A) a variola virus 81bp fragment corresponding to part of the major deletion in monkeypox virus and B) plasmid pDenE/B, which contains monkeypox virus DNA.	45
Fig. 4.1	Map of Western and Central Africa showing the extent of tropical rain forest and locations where cases of human monkeypox have occurred.	51
Fig. 4.2	Map of Africa showing the topology between West and Central Africa; and the places in Zaire from where monkeypox virus was isolated.	51
Fig. 4.3	Alignment of DNA sequences from different monkeypox virus isolates with variola virus, Harvey.	53
Fig. 4.4	Diagram to show the difference in corresponding DNA sequences of variola virus, monkeypox virus from West Africa and monkeypox virus from Zaire.	54
Fig. 4.5	Agarose gel electrophoresis of PCR-generated DNA fragments corresponding to the region of the major deletion in monkeypox virus.	55
Fig. 5.1	*HindIII* map of vaccinia virus, Dairen, showing the location of the E5R ORF.	62
Fig. 5.2	DNA sequence alignment of the E5R gene region of vaccinia viruses Dairen, Copenhagen and Western Reserve.	62

Fig. 5.3	Positions and sequences of DNA mismatches found amongst the vaccinia viruses Dairen, Copenhagen and Western Reserve.	63
Fig. 5.4	Generation of duplications or deletions by slipped-strand mispairing between contiguous repeats.	64
Fig. 5.5	Alignment of the DNA sequences corresponding to vaccinia virus E5R in other *Orthopoxviruses*.	67
Fig. 5.6	DNA sequence of the E4L and E5R promoter region.	70
Fig. 5.7	Matrix showing the % divergence and % similarity between pairs of DNA sequences.	71
Fig. 5.8	Phylogenetic tree constructed from the DNA sequences aligned in fig. 5.5 using TREECON, accounting for insertions and deletions.	74
Fig. 5.9	Phylogenetic tree constructed from the DNA sequences aligned in fig. 5.5 using TREECON, ignoring insertions and deletions.	74
Fig. 5.10	Unrooted tree constructed from DNA sequences aligned in fig 5.5 using clustal V.	76
Fig. 5.11	Alignment of the intergenic sequence between vaccinia virus E5R and E6R with the corresponding sequences of variola (2 strains) and monkeypox (7 strains) viruses.	78
Fig. 5.12	Alignment of *Orthopoxvirus* DNA sequences corresponding to the intergenic region between vaccinia virus E5R and E6R.	80
Fig. 6.1	Diagram showing the partial deletion of the vaccinia virus E5R gene.	87
Fig. 6.2	pGptC/B, The plasmid used for the replacement of part of the E5R ORF with gpt.	87
Fig. 6.3	Titration of the recombinant and parent vaccinia viruses in CV1 cells.	88
Fig. 6.4	*HindIII* digests of parent and recombinant viruses subject to agarose gel electrophoresis followed by hybridization to radiolabelled probes of: A) pGptMTL23, which contains the gpt gene, and B) a *ClaI-BglII* fragment corresponding to the region deleted in the recombinant virus.	89
Fig. 6.5	Northern blot analysis of RNA extracted from CV1 cells, uninfected; infected with parent vaccinia virus, Dairen; and infected with recombinant vaccinia virus, gptC/B.	90
Fig. 6.6	Lesions produced in rabbit skin by the recombinant and parent viruses.	92
Fig. 6.7	Diagram showing the charge distribution of the putative E5R gene product.	94
Fig. 6.8	Diagram showing regions of hydrophobicity and hydrophilicity within the putative E5R gene product.	94
Fig. 6.9	Deduced amino acid sequence of the E5R gene product of vaccinia virus, Dairen, aligned with the equivalent deduced sequences from the other *Orthopoxviruses*.	96

LIST OF TABLES

TABLE 1-1	Biological characteristics of different species of *Orthopoxvirus*.	4
TABLE 1-2	Human monkeypox: areas of tropical rain forest and annual numbers of cases reported in countries in Western and Central Africa: 1970-1986.	10
TABLE 2-1	*Orthopoxviruses* from which DNA was examined.	25
TABLE 2-2	Primers used during the course of this study.	34
TABLE 2-3	Primers used for DNA sequencing of the equivalent region in different *Orthopoxviruses*.	35
TABLE 4-1	Details of monkeypox virus isolates.	50

ABSTRACT

Evidence is presented to show that variola and monkeypox viruses evolved independently from a common ancestor. An open reading frame (ORF), potentially coding for a protein of 341 amino acid residues, was found to be conserved in two strains of variola virus (Harvey and Somalia), but degenerate in the Denmark strain of monkeypox virus. Monkeypox virus had a deletion of 391bp, two 24bp deletions and a single base pair deletion within the coding region of this single copy ORF. The ORF corresponds to the E5R ORF in the published sequence of the Copenhagen strain of vaccinia virus, and the DNA sequence was determined for an additional strain of vaccinia virus, Dairen. A number of other *Orthopoxviruses* were found to contain this ORF, strengthening confidence in its presence in an ancestral *Orthopoxvirus*.

The equivalent DNA sequence was determined for a number of monkeypox virus strains from West and Central Africa. The Denmark strain was identical to one from Liberia, indicating that this virus probably originated from West Africa. A third virus from West Africa, Benin, was found to have >99% base similarity and the same pattern of deletions as the other two monkeypox viruses.

The Zaire strains were identical to one another and different from the West African strains. Like the West African strains, they contained the two 24bp deletions and single base pair deletion. In place of the large deletion they had three smaller deletions of 5-, 9- and 127-bp as well as a single base pair insertion. They also had additional deletions of 1- and 2-bp and an insertion of 3bp. The West African strains have the potential to code for a truncated gene product of 107 amino acid residues, whereas the Zaire strains have no significant ORF. This clearly shows that monkeypox virus has diverged into two geographically isolated groups (Zaire and West Africa). There was >99% base similarity between the two groups, suggesting that the divergence occurred recently.

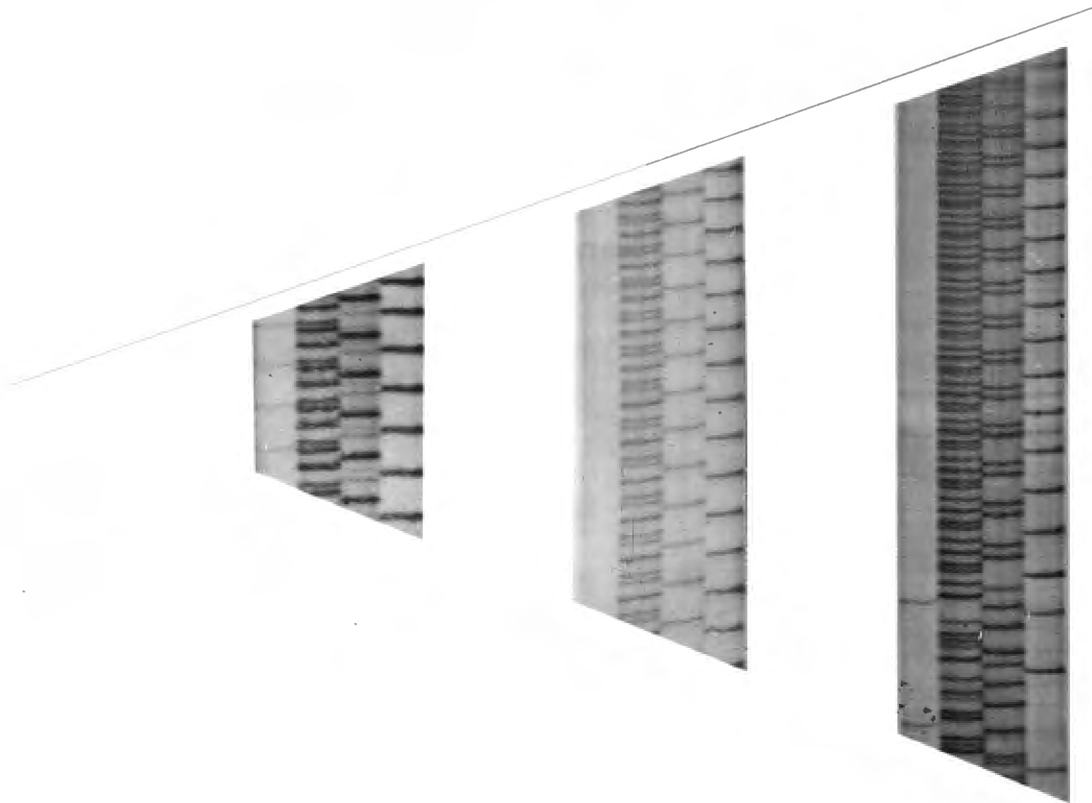
Phylogenetic analysis, by the neighbour-joining method, was undertaken on the corresponding DNA sequences from variola (2 strains), monkeypox (6 strains), vaccinia (1 strain + 2 published sequences), cowpox (2 strains), taterapox, camelpox and ectromelia viruses. For every species gerbilpox virus was the nearest neighbour, suggesting that taterapox virus may be the species most closely related to the common ancestral *Orthopoxvirus*. Within the variola and cowpox virus species there was >99% DNA sequence conservation. Between species, camelpox virus was the most closely related species to gerbilpox virus, with variola virus, and, more distantly, vaccinia virus, falling into the same group. Cowpox virus was the most diverged species examined. Ectromelia virus shared a branch with cowpox virus.

A comparison was made of the intergenic DNA sequence between this ORF and the adjacent downstream ORF. Variation was found, both within and between species, in the form of insertions and deletions. The interrelationships between the different *Orthopoxvirus* species more or less parallels that of the E5R-equivalent comparison. Some of the viruses

had clusters of direct repeats. A pentameric repeated unit was found in 2, 10 and 17 copies in camelpox, gerbilpox and ectromelia viruses respectively. Raccoon poxvirus had a 7bp unit in 13 adjacent copies. The two cowpox viruses had a more complex arrangement of repeated sequences.

It was thought that the E5R ORF may prove to be nonessential for virus replication. This was tested by interruption of the E5R gene in vaccinia virus; this did not affect the ability of the virus to form plaques in cell culture, but appeared to reduce the pathogenicity of the virus for rabbits.

The deduced amino acid sequences were analysed for conserved and variable regions within the gene, to which no specific function has yet been assigned.



CHAPTER I

GENERAL INTRODUCTION

Classification 2

The *Orthopoxviruses* 3

Origin of variola virus 6

Possible animal reservoir of variola virus 7

Monkeypox virus in captive animals 8

Biological characteristics of monkeypox and variola viruses 9

Human monkeypox 10

Incidence and transmission of monkeypox virus in humans 12

Natural reservoir of monkeypox virus 13

Whitepox viruses 14

Genome comparison of variola and monkeypox viruses 16

Other *Orthopoxviruses* closely related to variola virus 17

Thymidine kinase: selection system, enzyme and evolution 18

CHAPTER I

GENERAL INTRODUCTION

The study undertaken can be divided into four sections: 1) the possible derivation of variola virus from monkeypox virus; 2) the evolution of monkeypox virus; 3) the evolutionary relatedness of viruses within the *Orthopoxvirus* genus; and 4) characterization of a vaccinia virus gene.

The background information is from previous work which is relevant to this thesis. This will include the characterization of the *Orthopoxviruses*; work leading up to the hypothesis that monkeypox virus is a progenitor of variola virus; and previous evolutionary analyses which have been reported on poxviruses (comparing their thymidine kinase genes). An in-depth review of vaccinia virus structure, gene expression, replication and morphogenesis will not be given, as the field is now vast and a number of good review articles are available (Westwood *et al.*, 1964; Moss, 1990a; Condit and Niles, 1990; Moss, 1990b; Traktman, 1990; DeLange and McFadden, 1990; VanSlyke and Hruby. 1990). The pathogenesis of poxviruses has been reviewed by Turner and Moyer (1990) and Buller and Palumbo (1991). More recently, Gooding (1992) described some virally-encoded proteins which modulate the immune and inflammatory response. The virulence of poxviruses will not be reviewed here, but will be briefly discussed in chapter 6.

CLASSIFICATION

The initial classification of poxviruses was based on the disease that these infectious agents caused. This included the "great pox" (syphilis) and chickenpox which also produce vesicular skin lesions. Once the nature of the infectious agents was established, the poxviruses could be grouped more accurately according to the characteristics of the virus rather than the clinical manifestations caused (Fenner, 1976). By definition, the family consists of viruses with large cuboid or prolate virions 260-460 nm long containing a single double-stranded DNA molecule with a molecular weight of 130-375 X 10^6 Da (Porterfield, 1989).

There are two subfamilies of *Poxviridae*: *Chordopoxvirinae* (viruses of vertebrates) and the *Entomopoxvirinae* (insect poxviruses). The *Orthopoxviruses* are one of eight genera of *Chordopoxvirinae*. The other seven genera are *Parapoxvirus*, *Capripoxvirus*, *Suipoxvirus*, *Leporipoxvirus*, *Avipoxvirus*, *Yatapoxvirus* and *Molluscipoxvirus* (Fenner *et al.*, 1989).

THE *ORTHOPOXVIRUSES*

There are ten recognized species of *Orthopoxvirus* which are morphologically identical, serologically cross-reactive and genetically similar (Fenner *et al.*, 1989).

Biological classification

In the past the *Orthopoxviruses* were differentiated by their biological characteristics. These have been summarized in table 1-1. The morphology of the pock produced on the chick chorioallantoic membrane (CAM) has been of particular importance. An additional differentiating feature is the cytopathic effect (CPE) produced by different viruses in infected cells. Variola virus is characterized by its typical round-cell CPE in Vero, HeLa and BSC-1 cells as compared to the "strand cell" type CPE produced by vaccinia virus. Variola virus-infected cells progress to the formation of syncytial cells and small plaques 4 days after infection. In contrast, vaccinia virus produces large cytolytic plaques after 3 days. Other cell lines, such as RK13 and PEK cells, have also been used to aid the differentiation of certain *Orthopoxviruses*.

Comparing biological characteristics, camelpox and gerbilpox viruses resemble variola virus in most of the tests. The main feature which differentiates these two viruses from variola virus is the nature of their thymidine kinase enzymes. Variola virus thymidine kinase is subject to feedback inhibition by TTP, whereas the TK enzmes of camelpox and gerbilpox viruses are not sensitive to TTP (Bedson, 1982). The viruses may also be distinguished by the CPE which they produce in certain cell lines. Camelpox virus produces multinucleate giant cells in HeLa, BSC-1 and transformed human amnion cells (Bedson, 1972; Baxby, 1975); and gerbilpox virus produces cytolytic, as opposed to hypertrophic, foci in RK13 cells. Gerbilpox virus, unlike variola virus, is readily passaged in rabbit skin (Gispen, 1972).

Protein classification

Although the *Orthopoxviruses* show extensive serological cross-reactivity, three antigens have been recognised by Gispen and Brand-Saathof (1974) which allow classification of variola, monkeypox and vaccinia viruses into three groups: variola virus (vc^-, va^+, mo^-), monkeypox virus (vc^-, va^-, mo^+) and vaccinia virus (vc^+, va^+, mo^-). Rabbitpox, buffalopox and camelpox viruses resemble vaccinia virus. Extracts of cowpox virus-infected CAMs are negative for all antigens. However, Rondle and Dumble (1962) have shown that a soluble antigen present in vaccinia virus-infected cells, which cannot be detected in cowpox virus-infected cells (Gispen, 1955), is produced by cowpox virus, but only detected in extracts of infected CAM after trypsin treatment. Although the differences observed by Gispen and Brand-Saathof (1974) may be due to the presence or absence of the entire protein, the same results would be observed if the protein were truncated or had some amino acid changes which prevented it from being secreted.

TABLE 1-1

Biological Characteristics of Different Species of Orthopoxvirus[a] *(From Fenner et al., 1989)*

Character	Camelpox virus	Cowpox virus	Ectromelia virus	Monkeypox virus	Raccoon poxvirus	Tatera poxvirus	Uasin Gishu poxvirus	Vaccinia virus	Variola virus	Vole poxvirus
Pocks on CAM[b]	Small, opaque, white	Large, hem.[c]	Very small, opaque, white	Small, opaque, hem,	Very small, opaque, white	Small, opaque, white	Medium size, opaque	Strains vary; large, opaque, white, or hem.	Small, opaque white	Very small, opaque, white
Ceiling temperature (CAM)	38.5°C	40°C	39°C	39°C	?	38°C	?	41°C	37.5-38.5°C	40°C
Rabbit skin lesion	Transient, non-transmissible	Indurated, hem.	Transient, non-transmissible	Indurated, hem	Small nodule	Transient, non-transmissible	No lesion	Strains vary; indurated nodule, sometimes hem,	Transient, non-transmissible	Small nodule
Disease in monkeys	Large lesion, localised	Large lesion, localised	?	Generalized rash		?	Susceptible; no rash	Large lesion, localised	Generalized rash	?
Lethality for Mice	Low	Variable	Very high	High	High	Low	Pocks in baby mice	Strains vary; high to very high	Low	High in baby mice
Chick Embryos	Low	High	High	Medium	?	Low	?	Very high	Low	?
Type A inclusion bodies	-	+	+	-	+	?	-	-	-	?
Thymidine kinase sensitivity to TTP[d]	-	-	-	-	?	-	?	-	+	?

[a]Data on raccoon poxvirus, tatera poxvirus, Uasin Gishu poxvirus, and vole poxvirus based on a few papers on single viral isolates.
[b]CAM Chorioallantoic membrane; examined at 48 hours for vaccinia virus; 72 hours for all others.
[c]hem, Hemorrhagic.
[d]TTP, Thymidine triphosphate.

Esposito *et al.* (1977a) differentiated variola, monkeypox and vaccinia viruses by immunodiffusion of cross-absorbed sera. These three viruses have been serologically differentiated by four methods employing cross-absorption of the test sera with homologous and heterologous antigens: gel precipitation (Gispen and Brand-Saathof, 1974), immunofluorescence (Gispen *et al.*, 1976), radioimmunoassay (Hutchinson *et al*, 1977) and ELISA (Marennikova *et al*, 1981).

The three viruses (vaccinia, variola and monkeypox) have distinctive structural polypeptides in the 30-40kD range (Esposito *et al*, 1977b; Arita and Tagaya, 1977). Turner and Baxby (1979) showed that cowpox, ectromelia, camelpox, monkeypox and vaccinia viruses have distinctive surface and sub-surface polypeptides, but indistinguishable core proteins. The Moscow rat-carnivore virus (Marennikova *et al.*, 1977), which resembles cowpox virus in restriction enzyme map (Fenner *et al.*, 1989), had the same polypeptide pattern as ectromelia virus, and not cowpox virus. Harper *et al.* (1979) showed distinct patterns for the late intracellular proteins of variola, vaccinia, monkeypox and cowpox viruses. Two whitepox viruses were indistinguishable from those of variola virus; Lenny, MK-10 and Buffalopox viruses resembled vaccinia virus; and the Moscow rat carnivore virus resembled cowpox virus. Unfortunately ectromelia virus was not included in this comparison - it would be interesting to know whether ectromelia virus is similar to cowpox and Moscow rat-carnivore viruses.

The polypeptide patterns correlate with the classification based on biological tests.

DNA classification

A simpler and more reliable method of allocating a virus to a specific *Orthopoxvirus* species is that of DNA analysis. Gangemi and Sharp (1976) showed that closely related strains of vaccinia virus could be differentiated by their restriction enzyme profiles. This analysis was extended to different strains and species of *Orthopoxviruses* by Esposito *et al.* (1978). Mackett and Archard (1979), Esposito and Knight (1985) and Knight *et al.* (1992) have mapped all the *Orthopoxvirus* species, except for Uasin Gishu poxvirus. Fig. 1.1 shows the *HindIII* maps of each of the species mapped, except for volepox virus, which has more recently been mapped by Knight *et al.* (1992). Each virus has a distinctive pattern, but all have a similar central region. All the fragments show cross-hybridization between species. Although less than half the raccoon poxvirus fragments hybridize to vaccinia virus, they all hybridize to cowpox virus (Parsons and Pickup, 1987).

The overall arrangement of the genetic information is conserved in all the viruses (as shown by cross-hybridization - Mackett and Archard (1979), Esposito and Knight (1985); Knight *et al.* (1992)). The central region codes for structural proteins, proteins required for gene expression and replication, and enzymes involved in nucleic acid metabolism ie. functions essential for basic survival of the virus (Goebel *et al.*, 1990a). Proteins which are involved in host range and pathogenicity, are encoded by genes located nearer the termini

(Perkus *et al.*, 1990; Turner and Moyer, 1990). Many of the terminally located genes are nonessential (Kotwal and Moss, 1988; Perkus *et al.*, 1991).

Raccoon poxvirus and volepox virus, the only *Orthopoxviruses* naturally found in the Americas, show greater divergence from the others (Knight *et al.*, 1992). They have domains throughout their genomes which are more diverged; and the tandem repeat sequences, shared by all the other viruses, are distinct in both viruses (Knight *et al.*, 1992). Although there is terminal variation, all the *Orthopoxviruses* have DNA similarity within the noncoding end region (Esposito and Knight, 1985; Parsons and Pickup, 1987). These sequences may be important for replication of the genome.

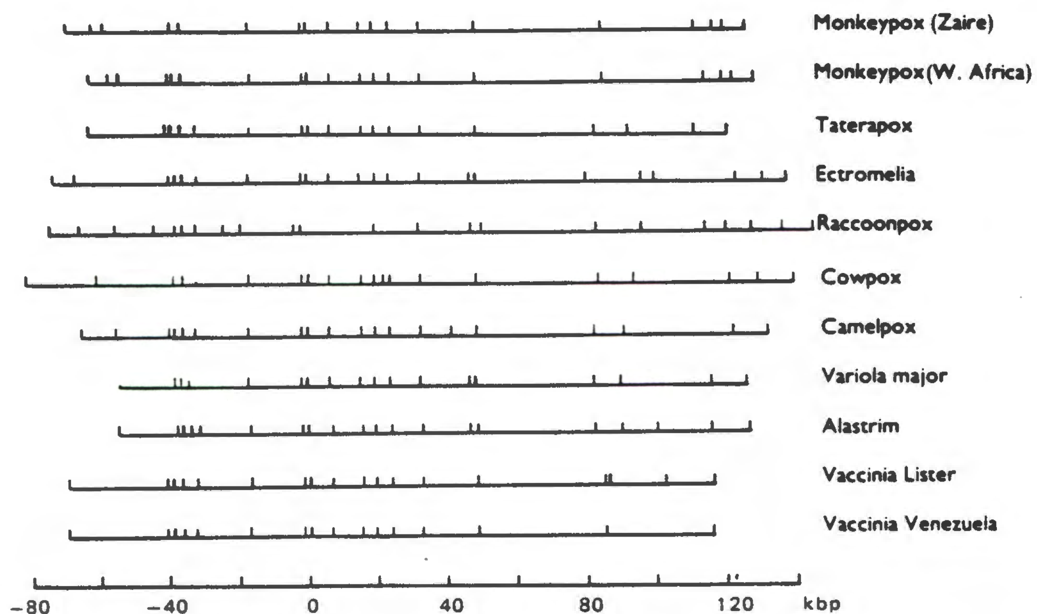


Fig. 1.1 *HindIII* genomic maps of 8 of the 10 recognized *Orthopoxvirus* species (data from Esposito and Knight, 1985).

ORIGIN OF VARIOLA VIRUS

Descriptions of smallpox have been found in ancient records. The distinctive clinical features of variola virus infection have made retrospective diagnosis possible. Smallpox is recognized as a severe, highly infectious disease with a typical rash, and which commonly results in residual facial pockmarks. The earliest records of smallpox date back to 1346 BC at the time of the "plagues" of the Hittites in Eastern Asia. The pock-marked mummified face of Ramses V who died in 1100 BC suggests that smallpox was present in Egypt at that time. Since it is unlikely that the virus originated independently in two parts of the world, either it came from East Asia, or from Africa. There is no evidence of its presence in the Americas, Australia or Southern Africa in the pre-Christian era (Fenner *et al.*, 1989).

The most widely accepted hypothesis is that variola virus originated in Eastern Asia and was spread to other parts of the world by travellers and via commercial traders. In this way

the virus may have been spread to Egypt. However, historical records made by the Roman, Egyptian, Babylonian, Assyrian and Persian civilizations record no evidence of smallpox epidemics. This lack of evidence, at a time when historical events were well documented, has led to the postulate that variola virus originated in Africa. The virus may have spread along the Nile river which served as a major communication route from Central Africa to Egypt. It has been proposed that variola virus evolved from monkeypox virus, which is endemic to West and Central Africa (Marennikova and Shelukhina, 1978).

Because variola virus is strictly limited in its host range to humans, it would require a large population to sustain transmission. Large enough settlements could have been formed about 5000 to 6000 years ago at the time of agricultural development. It is possible that variola virus originated as far back as this, probably from a virus endemic in the animal population, which infected man and evolved into variola virus.

POSSIBLE ANIMAL RESERVOIR OF VARIOLA VIRUS

A number of factors contributed towards the success of the smallpox eradication programme (Fenner *et al.*, 1988), one of which was the absence of a natural animal reservoir of variola virus. There have been seven reported cases of poxvirus infection in non-human primates between 1767 and 1949 (Arita and Henderson, 1969). None of these were confirmed variola virus infections. The first two episodes may have been caused by variola virus. In 1767 a monkey developed the disease after playing with children infected with smallpox; and, in 1841 a smallpox outbreak in a monkey population in Panama followed a human smallpox outbreak. Smallpox was reported in monkeys in 1842 and 1858 by Rayer and Furlong respectively. The next three incidences were described in more detail. Bleyer described infection of a monkey population in the Brazilian forest with smallpox in 1922; Rahman observed a poxvirus outbreak in monkeys in India in 1936; and Gispen reported a poxvirus episode amongst orangutans in a Djakarta zoo in 1949. In each of these cases the animals were severely affected and covered in numerous pustules. The second two cases specifically mention the presence of many pustules on the face, hands and soles of feet. The clinical description of these three episodes - particularly the latter two cases - do not correlate with published reports on simian infections with variola virus (Hahon, 1961; Noble and Rich, 1969). Simian smallpox differs from human smallpox clinically in that the rash is more sparse on the head, face, hands and feet; and the infection is milder. The disease described in these animals resembles more closely that described by Von Magnus in 1959, which was shown to result from a virus distinct from variola virus, monkeypox virus. Monkeypox virus produces a disease in monkeys which resembles smallpox in humans (Von Magnus *et al.*, 1959). The pustules are particularly abundant on the palms of the hands and the soles of the feet. In only one (1949) of the seven episodes mentioned was virus isolated; and in this case variola virus was not confirmed as the causative agent.

Experiments have been performed to test the ability of variola virus to be maintained in monkeys (Noble and Rich, 1969; Noble, 1970). Transmission was demonstrated in *Macaca Iris* by both contact and aerosol routes. In contrast to human smallpox, the disease was more superficial and infectivity declined upon serial passages. In one case there was an increase in number of lesions, but this monkey suffered a fractured tibia and most of the lesions appeared under the leg splint where the skin was traumatised. This monkey was the fourth in a series of passages which only died out after six generations. Variola virus can, therefore, infect susceptible monkeys, but evidence suggests that the virus will not be indefinitely maintained in these animals. In similar experiments New World monkeys from Peru and Columbia were compared to *Macaca Iris* from the Philippines for their susceptibility to variola virus infection (Noble, 1970). Although all monkeys seroconverted, the New World monkeys developed no lesions (for both variola major and variola minor). A number of sera from monkeys from South America, Africa and the Philippines were tested for poxvirus antibody, but were all negative (Noble, 1970).

No outbreaks of variola virus infection in animals in nature have been recorded since 1936. There has also been no natural case of human infection since 1977. These facts, in addition to the intense surveillance for variola virus and related animal viruses by the WHO, argue strongly against an animal reservoir of variola virus.

MONKEYPOX VIRUS IN CAPTIVE ANIMALS

Monkeypox virus was first isolated from pox-like lesions in captive cynomolgous monkeys at the State Serum Institute, Copenhagen (Von Magnus *et al.*, 1959). Two outbreaks occurred in animals imported from Singapore by air; when 20 to 30% of the animals developed clinical illness. The disease, which was nonfatal, developed only 51 and 62 days after arrival of the monkeys. It was characterized by a vesiculopustular rash. The virus resembled pox viruses morphologically and was serologically related to vaccinia virus; but the pocks produced on CAMs, and the pathogenicity of the virus for laboratory animals, classified it as a separate species of *Orthopoxvirus*. Like variola virus, the virus produced small white pocks on CAMs. But, unlike variola, and more like vaccinia virus (which can produce either haemorrhagic or nonhaemorrhagic lesions in rabbit skin), it produced large indurated haemorrhagic lesions in rabbit skin and was maintained in serial passages in rabbit skin as well as by intracerebral inoculation of adult mice. Variola virus is nontransmissible in these animals and does not produce a severe lesion in rabbit skin.

Since 1959 there have been nine additional reports of monkeypox virus outbreaks in captive animals: 1 in Denmark, 1 in France, 2 in the Netherlands, and 6 in the USA (Arita *et al.*, 1972). There have been no reported cases of outbreaks in captive animals since 1968. This may be due to increased caution with regard to the health of animals accepted by laboratories, and the breeding of monkeys in Europe and the USA. Nevertheless, considering

the frequent use of monkeys in Europe and the USA, the incidence of monkeypox virus infection is extremely low.

BIOLOGICAL CHARACTERISTICS OF MONKEYPOX AND VARIOLA VIRUSES

Evidence to suggest that monkeypox and variola viruses are ancestrally related (Marennikova and Shelukhina, 1978; Marennikova *et al.*, 1979) is based on biological characteristics. The most important features for differentiation of these two viruses are 1) host range and pathogenicity for laboratory animals, 2) effect on chick embryos and 3) behaviour in tissue culture. Nowadays, DNA restriction enzyme mapping provides a more conclusive result (Mackett and Archard, 1979; Esposito and Knight, 1985).

Host range and pathogenicity for laboratory animals

Monkeypox virus has a wide host range. Susceptible primates include humans, great apes and many species of monkeys; and non-primates include the giant anteater, rabbits, mice, chick embryos, guinea pigs, hamsters and squirrels (Cho and Wenner, 1973; Marennikova and Seluhina, 1976; Khodakevich and Jezek, 1986). In contrast, variola virus is limited to humans.

When injected intradermally into rabbit skin variola virus produces transient lesions and the virus is nontransmissible on subsequent passages. In contrast, monkeypox virus produces large indurated hemorrhagic lesions (Marennikova *et al.*, 1971; Rondle and Sayeed, 1972) and can be indefinitely passaged in rabbit skin. Monkeypox virus is lethal for 3-week-old mice when inoculated by the intracerebral route; variola virus is nonpathogenic.

Chick embryos

The pocks produced by variola and monkeypox viruses on chick chorioallantoic membranes may resemble one another if incubated at 37°C. They are nonhaemorrhagic, small and dome-shaped. Von Magnus *et al.* (1959) and Gispen *et al.* (1967) observed such pocks. Subsequently monkeypox virus has been characterised as producing a greyish pock with a haemorrhagic centre at an incubation temperature of 34.5°C (Marennikova *et al.*, 1979; Rondle and Sayeed, 1972). These haemorrhagic pocks differ from those of cowpox virus in that the erythrocytes are on the surface; for cowpox virus pocks the erythrocytes are beneath the ectodermal cells and the pocks are fully haemorrhagic.

The ceiling temperature of monkeypox virus, 39°C, lies in between those of vaccinia (41°C) and variola (37.5-38.5°C) viruses (Bedson and Dumbell, 1961; Gispen *et al.*, 1967; Marennikova *et al.*, 1971).

Cell culture

Monkeypox and variola viruses both replicate in a wide range of cells in culture. However, differences have been observed in some cell lines including HeLa, Vero and PEK cells (Marennikova *et al.*, 1971, 1972b).

HUMAN MONKEYPOX

In 1970, at the time of intense smallpox surveillance, and more than a year after the last case of smallpox had been detected in the area, human monkeypox was identified in Zaire (Ladnyj *et al.*, 1972). Virus was recovered from human skin lesions on the chorioallantoic membrane of chick embryos (Marennikova *et al.*, 1972a). After three days at 35°C haemorrhagic pocks were produced which resembled those of monkeypox virus. Like monkeypox virus, the isolate produced large necrotic lesions in rabbit skin.

TABLE I-2

Human Monkeypox: Areas of Tropical Rain Forest and Annual Numbers of Cases Reported in Countries in Western and Central Africa: 1970-1986[a]

	Cameroon	Central African Republic	Cote d'Ivoire	Liberia	Nigeria	Sierra Leone	Zaire	Totals
Population in thousands (1980):	8,554	2,290	8,247	1,871	80,555	3,296	28,532	124,791
Rain forest (1980)[b]	17,920	3,590	4,458	2,000	5,950	740	105,650	140,308
Percentage of rain forest[c]	9.5	1.9	2.4	1.1	3.2	0.4	56.2	74.7
Cases of monkeypox								
1970	-	-	-	4	-	1	1	6
1971	-	-	1	-	2	-	-	3
1972	-	-	-	-	-	-	5	5
1973	-	-	-	-	-	-	3	3
1974	-	-	-	-	-	-	1	1
1975	-	-	-	-	-	-	3	3
1976	-	-	-	-	-	-	5	5
1977	-	-	-	-	-	-	6	6
1978	-	-	-	-	1	-	12	13
1979	2	-	-	-	-	-	8	10
1980	-	-	-	-	-	-	4	4
1981	-	-	1	-	-	-	7	8
1982	-	-	-	-	-	-	40	40
1983	-	-	-	-	-	-	84	84
1984	-	6	-	-	-	-	86	92
1985	-	-	-	-	-	-	62	62
1986	-	-	-	-	-	-	59	59
Total cases:	2	6	2	4	3	1	386	404

[a] From Jezek and Fenner (1988).
[b] Thousands of hectares. Source: Food and Agriculture Organization (1981).
[c] In western and central Africa: 25.3% of total occurs in six countries in which human monkeypox has not been reported.

Human monkeypox is rare and only occurs in isolated villages in the tropical rain forests of West and Central Africa. Cases have been reported in seven countries between 1970 and 1986 (table 1-2 and fig. 4.1) (Jezek and Fenner, 1988) and in an eighth (Gabon) in 1991 (Monkeypox 1991, 1992).

Descriptions of human monkeypox incidents reported in the 1970s (Breman *et al.*, 1980) and 1980s (Jezek *et al.*, 1987a) correlate with respect to monkeypox virus producing a clinical picture in humans which resembles that of discrete ordinary-type smallpox.

Infection is believed to occur via the upper respiratory tract for person-to-person spread, and possibly through abrasions of the skin or oral cavity, in addition to via the oropharynx or nasopharynx, for animal-to-human spread. One case of congenital human monkeypox has been reported (Jezek and Fenner, 1988).

The incubation period is about 12 days. Like smallpox, there are two stages of illness, pre-eruptive and eruptive. The illness lasts for 2-4 weeks depending on its severity. The syptoms and their severity are related to the vaccination status and health status of the patient, but not sex or age.

Pre-eruptive stage: Illness usually starts with fever followed by a rash. The fever is accompanied by severe headache, backache, general malaise and prostration. Lymph node enlargement is often observed before onset of the rash. Lymphadenopathy, which is not characteristic of variola virus infection, is the major clinical difference between smallpox and human monkeypox.

Eruptive stage: Skin lesions develop simultaneously and evolve at the same rate. The number of lesions vary significantly between vaccinated and unvaccinated patients, ranging from a few to several thousand. Vaccinated patients also developed smaller lesions that the usual 0.5-1cm lesions in unvaccinated subjects and often develop an indefinite distribution of lesions compared to the usual centrifugal distribution in unvaccinated patients. The degree of fever is proportional to the number of lesions, as are the severity of symptoms and duration of illness. Like smallpox, the lesions pass through stages of macules, papules, vesicles and pustules. They then umbilicate, dry out and desquamate. A second stage of fever often occurs at the pustular stage in the second week of illness.

The severity of infection from a wildlife source (primary) does not differ from that acquired from a human source (secondary). However, more mucous membrane lesions are associated with primary cases. There is also no difference in clinical features between human monkeypox cases from different geographical areas. Similar descriptions were given for cases in Liberia, Sierra Leone and Nigeria (Foster *et al.*, 1972), Cote d'Ivoire (Breman *et al.*, 1977), Central African Republic (Khodakevich *et al.*, 1985) and Zaire (Jezek *et al.*, 1987a). This indicates that genetic differences found in geographically distinct isolates (Esposito and Knight, 1985; chapter 4 of this thesis) do not influence the pathogenicity of the virus for humans.

The prognosis of human monkeypox differs from that of smallpox. It is associated with

a crude case-fatality rate of 11% among unvaccinated patients, in contrast to 25% for smallpox. All the deaths occurred in unvaccinated children less than 8 years old. No deaths have been detected among adults or patients with a visible vaccination scar. Like smallpox, the most common sequelae of human monkeypox have been facial pockmarks.

INCIDENCE AND TRANSMISSION OF MONKEYPOX VIRUS IN HUMANS

Although smallpox and human monkeypox present a similar clinical picture, the epidemiology of the two diseases differ considerably. Smallpox was easily transmissible (20-40% secondary attack rate) and produced clinical illness in most unvaccinated persons. Sporadic cases were not seen. Human monkeypox, however, occurs sporadically in isolated villages in or near the tropical rain forests of West and Central Africa, is not easily transmissible (3.3% secondary attack rate) and is less severe with a lower case-fatality rate than smallpox. The majority of human monkeypox cases have had close contact with animals, either in the agricultural area surrounding villages or in the forest where the animals are trapped or killed.

Occasional clustering of human monkeypox episodes suggests that person-to-person transmissions do occur. These incidences could result from infection from a common animal source. Since infectivity is associated with onset of rash, cases occurring within 7 days of the primary case can be considered co-primary arising from the same nonhuman source. Cases occurring 7-21 days after exposure to a known human source are regarded as secondary cases. Jezek *et al.* (1988) reported the distribution of 338 cases in Zaire from 1981-1986. Of these 203 were primary, 42 co-primary and 93 secondary cases. Of the secondary cases 69 were first generation, 19 second generation and 5 third and fourth generation. There has been one isolated incidence of person-to-person spread for four generations (Jezek *et al.*, 1986). This decrease in incidence may partly be due to subclinical infections occurring among close contacts of monkeypox cases. The low incidence of human monkeypox in areas where the virus is endemic and contact between humans and animals is frequent, indicates that transmission from animals is infrequent. There were also no cases of human infection at the time of outbreaks in captive animals, possibly because all human contacts were vaccinated.

The number of human monkeypox cases in Zaire increased from 48 in 1970-1980 to 338 in 1981-1986. Since the secondary attack rate was constant for both periods, this increase was probably due to increased surveillance and not an increase in the transmissibility of the virus.

Of concern, is the relationship of attack rate to vaccination status. Jezek and Fenner (1988) have reported that the primary attack rate among unvaccinated individuals (1.7 per 10000) is significantly higher than that for vaccinated individuals (0.04 per 10000).

The rate was 2.6 per 10000 among unvaccinated children 5-14 years of age. Secondary attack rates are also related to vaccination status. A subclinical attack rate of 28%, assessed by serological tests, was determined for unvaccinated contacts of human monkeypox cases (Jezek and Fenner, 1988). Since vaccination has ceased there is a growing population of individuals susceptible to monkeypox virus infection. For this reason a stochastic model was designed to test the possible frequency of monkeypox virus transmission amongst a population decreasing in vaccination status; and to assess the possibility of monkeypox virus establishing itself in a human population by continuous person-to-person spread (Jezek *et al.*, 1987b). The model was based on the observed transmission frequency of monkeypox among individuals in Zaire. It produced possible chains of transmission using a randomly assigned number of contacts for different distributions of vaccination status. Using conditions present in Zaire from 1980-1984, the predicted length of transmission was as seen in real life, with one simulation continuing until the seventh generation (compared to the maximum of 5 generations observed in the field). The "worst case" scenario, assuming the entire population is unvaccinated and attack rates are high, predicted an outbreak which could last 14 generations. Even in this extreme case the incidence of monkeypox infection declined showing that monkeypox virus will not persist in the human population under present conditions of infectivity of the virus for man.

The incidence of human monkeypox is probably limited by the frequency of human contact with infected animals. Therefore, an increase in infection in the animal population with which man has contact could cause an increase in human monkeypox. This situation is unlikely to arise, since the tropical rain forests have been gradually receding, and the animal populations diminishing. Unless monkeypox virus becomes adapted to a new host which has a high population density, and which regularly comes into contact with humans, the incidence of human monkeypox should not increase explosively.

NATURAL RESERVOIR OF MONKEYPOX VIRUS

In each of the monkeypox virus outbreaks in captive animals the index case originated from either Asia or West Africa. To locate the natural reservoir of monkeypox virus the WHO conducted extensive serological surveys in both these continents. Initially, a total of 2242 monkey sera from South-East Asia and Africa (Chad, Upper Volta, Mali, Kenya, Senegal) were tested for poxvirus antibodies (Arita *et al.*, 1972), but the results were negative for all sera.

After the recognition of human monkeypox in 1970 (Ladnyj *et al.*, 1972), attention was focussed on Africa as the source of monkeypox virus. Breman *et al.* (1977) examined 195 sera collected from primates in West Africa. *Orthopoxvirus* antibodies were detected by haemagglutination inhibition (8%) and neutralization (21%) tests. These positive sera were collected from monkeys shot near two areas where human monkeypox infections had

occurred. Three of the HI-positive sera were shown to have monkeypox virus-specific antibodies by immunofluorescence following cross-adsorption of the sera with vaccinia virus antigen (Gispen *et al.*, 1976).

Since most human monkeypox cases have occurred in Zaire further serological investigations were done in this country on monkey sera (Marennikova *et al.*, 1975) as well as sera from a variety of other animals (Jezek and Fenner, 1988). Of 1331 sera from 45 species of wild animals, 227 (13.2%) gave positive HI results. The positive sera were from four species of squirrels and ten species of monkeys.

There have been no cases reported on the isolation of monkeypox virus from wild monkeys. One report (Mutombo *et al.*, 1983) describes the development of monkeypox in a child after being bitten by a chimpanzee in Zaire, but the animal could not be caught to test for virus. Due to the low incidence of monkeypox virus in wild monkeys, it is thought that the virus is maintained in a host other than monkeys, and, like man, monkeys are only incidental hosts which are not usually infected.

In 1985 monkeypox virus was isolated from a diseased wild squirrel in Zaire (Khodakevich and Jezek, 1986). Of 18 squirrels of the same species (*Funisciurus aneruthrus*) captured, another one was found to have monkeypox-specific antibodies by radioimmunoassay adsorption. Serum from 172 terrestrial rodents were tested, but none others were positive for monkeypox-specific antibodies. Another survey in Zaire (Khodakevich *et al.*, 1987b) revealed a prevalence of 24.7% monkeypox-specific antibodies in 320 squirrels of the *Funisciurus anerythrus* species, but none in 233 terrestrial rodents. A second species of squirrel, *Heliosciurus rufobrachium*, also had a high level of antibody prevalence. These two species of squirrels inhabit the agricultural area surrounding the settlement area, where they feed on oil palms and are often trapped. They could be the major source of infection for the human population (Khodakevich *et al.*, 1987a). *Heliosciurus* is also commonly found in the higher stratum of the forest, where it may be resposible for infecting monkeys. It is still uncertain whether squirrels are the major natural host of monkeypox virus. Although there is no evidence to support the maintainence of the virus in arthropods, the possibility of monkeypox virus having another natural host, probably an animal with a large population density, cannot be excluded.

In total four species of squirrels and ten species of monkeys have been found to be infected with monkeypox virus under natural conditions in the tropical rain forest regions of West and Central Africa. There is no evidence of monkeypox virus occurring naturally in any other part of the world.

WHITEPOX VIRUSES

During the search for natural reservoirs of monkeypox virus, four *Orthopoxviruses* were recovered from the kidneys of outwardly healthy animals in Zaire: 2 rodents (Marennikova

et al., 1976), a chimpanzee (Marennikova *et al.*, 1972b) and a monkey (Shelukhina *et al.*, 1975). These wild animals were shot in regions where human monkeypox cases had occurred. All the viruses were passaged on CAMs after an initial recovery of a few pocks. The small white pocks produced closely resembled these of variola virus. Characterization of ceiling temperature, reaction in rabbits, pathogenicity for mice and behaviour in PEK cell culture could not distinguish these "whitepox" viruses from variola virus. Similar whitepox viruses had previously been isolated from normal cynomolgous monkey kidney cells (Marennikova *et al.*, 1972b). These findings led to the hypothesis that whitepox viruses are variants of monkeypox virus which may represent a potential animal reservoir of variola virus.

This hypothesis was tested by passaging monkeypox virus both in hamsters (Marennikova and Shelukhina, 1978) and on CAMs (Marennikova *et al.*, 1979). Virus was isolated on CAMs from the kidneys and lungs of hamsters inoculated with monkeypox virus. Two isolates which produced dense white pocks resembling those of variola and whitepox viruses were characterized further. These viruses were lethal for chick embryos; did not produce pocks at 39°C, the ceiling temperature for monkeypox virus; replicated in PEK cells; were pathogenic for white mice; and gave no significant reaction in rabbit skin. All these characteristics are typical of variola virus and not monkeypox virus.

Marennikova *et al.* (1979) later reported their findings after passaging monkeypox virus on CAMs. They observed the emergence of white pock mutants of monkeypox virus at a level of 1-5%. Upon passage the viral population of h^+ (haemorrhagic) and h^- (white) pocks changed regularly, and this change was accompanied by a change in phenotypic character. With high titres of h^+ virus the pathogenicity for chick embryos was marked, but at 90% h^- virus the rabbit skin lesion still typified that of monkeypox virus. The authors interpreted this as a "chimeric" virus with characteristics both of monkeypox virus as well whitepox virus. However, these results could also be expected if there is a mixture of two "pure" viral populations. To show that a deletion must be occurring in the monkeypox virus genome the authors passaged "red" clones in the same manner. Upon future passages white pocks appeared and increased in number.

Orthopoxviruses which characteristically produce haemorrhagic pocks have been reported to produce white pock variants at a low frequency (±1%). Downie and Haddock (1952) first reported this for cowpox; Fenner (1958) for neurovaccinia; Gemmel and Fenner (1960) for rabbitpox; and Gispen and Brand-Saathof (1972) described a low level of white pocks produced by monkeypox virus at 35°C. This phenomenon is associated with terminal transposition events in which the left end of the genome is duplicated and replaces a portion of the right end (monkeypox: Dumbell and Archard, 1980; Esposito *et al.*, 1981; cowpox: Archard and Mackett, 1979; Archard *et al.*, 1984; Pickup *et al.* (1984); rabbitpox: Moyer and Rothe, 1980). Pickup *et al.* (1986) located a gene in cowpox virus which is in involved in the red pock phenotype. The product of this gene is a 38-kDa protein which is thought

to interact with the host inflammatory response (Palumbo *et al.*, 1989). Evidence suggests that additional genes are involved in the red pock phenotype since recombination of the 38K gene into a white pock mutant did not completely restore the red pock phenotype (Pickup *et al.*, 1986).

The white pocks described by Marennikova *et al.* (1979) arose at an unusually high frequency (5%) as compared to those referred to above (1%). Also, unlike the previous reports, the frequency at which these white pocks were produced was not constant. This indicates that the "whitepox" isolates differ from spontaneous white pock variants of monkeypox virus.

Three clones of whitepox virus passaged from monkeypox virus were characterized. One was from monkeypox virus, Copenhagen (isolated from a captive animal) and two were from Congo 8 (isolated from the first case of human monkeypox). All these viruses resembled variola and whitepox viruses in each of the differentiating biological tests described for the hamster isolates.

These results suggest that monkeypox virus is the progenitor to a variola-like virus. Since variola virus does not readily replicate in non-human hosts, the whitepox virus derived from monkeypox virus would not be maintained in the wild. The chances of a person becoming infected with a virus which has recently become altered in a wild animal (probably rodent) are remote. Alternatively, whitepox virus may emerge in humans infected with monkeypox virus. The low transmission frequency of monkeypox virus in humans makes this event highly unlikely. However, the possibility does exist. For this reason the whitepox viruses have been further characterized.

Dumbell and Kapsenberg (1982), by restriction enzyme analysis, showed the whitepox viruses isolated from cultured cynomolgous cells to be identical to variola virus strains which were in use in the laboratory at the time of isolation. It is likely that the whitepox viruses were, in fact, contaminating variola virus. Similarly, Esposito *et al.* (1985) provide evidence to suggest that all the other whitepox viruses were a result of laboratory contamination.

GENOME COMPARISON OF VARIOLA AND MONKEYPOX VIRUSES

Since DNA sequence is the blueprint that determines phenotypic character, a comparison of the viral genomes would be more informative with respect to the evolutionary pattern of the viruses.

Esposito *et al.* (1985) assessed the possibility of whitepox (variola-like) viruses arising from monkeypox virus by genome mapping. The *SacI* and *HindIII* maps provide similar information. The central region is highly conserved within all viruses mapped and the termini are variable. Differences in the termini were noted for viruses of the same species.

The whitepox viruses were identical to variola virus.

The *SacI* maps allow for differentiation of variola and monkeypox viruses. Variola virus has a distinctive terminal pattern of sites. A cluster of differences are noted in the left end of the conserved region. Variola virus has one site not present in the corresponding position in monkeypox virus; and monkeypox virus has two additional sites not found in variola virus.

The data shows that there is enough difference between variola and monkeypox viruses to rule out a simple interconversion of the two species, but the possibility still remained that monkeypox virus is a more distant ancestor of variola virus.

OTHER *ORTHOPOXVIRUSES* CLOSELY RELATED TO VARIOLA VIRUS

There are two viruses which closely resemble variola virus in their phenotypic properties: camelpox (Baxby, 1972; Bedson, 1972) and taterapox (Lourie *et al.*, 1975) viruses.

Camelpox virus.

Camelpox virus is naturally found in dromedary camels, probably throughout Africa, the Middle East and Asia. It has been isolated from camels suffering from camelpox in Iran (Ramyar and Hessami, 1972), Iraq (Al Falluji *et al.*, 1979), Kenya (Davies *et al.*, 1975), Somalia (Kriz, 1982) and the USSR (Marennikova *et al.*, 1974). Humans are rarely, if ever, infected with the virus (Jezek *et al.*, 1983).

Taterapox virus.

Taterapox viruses are presently limited to one isolate only, gerbilpox virus. This virus was isolated from a healthy wild gerbil caught in northern Dahomey (Benin) in 1968, one month after an outbreak of human smallpox in the area (Kemp *et al.*, 1974). The similarity of this virus to variola virus again caused concern about the long-term survival of variola virus as an inapparent infection of wild animals (Lourie *et al.*, 1975). It is not known whether this virus is naturally present in the gerbil population of West Africa, or whether this was an isolated incidence of variola virus entering into and adapting to another animal species.

The *HindIII* restriction enzyme maps show camelpox, gerbilpox and variola viruses to be similar, but distinct, from one another (fig. 1.1).

Ectromelia virus.

Ectromelia virus, like variola virus, produces small white pocks on chick CAMs (Bedson and Dumbell, 1961), has a narrow host range (Baxby, 1975) and has a similar distribution of *HindIII* restriction enzyme sites (Mackett and Archard, 1979). However, it is easily differentiated from variola virus by the other biological tests, and appears more closely related to cowpox virus in that it forms A-type inclusion bodies in infected cells (Baxby, 1975).

Ectromelia virus is the causative agent of mousepox in laboratory mice. It was first recognized by Marchal (1930) in Britain, and was subsequently isolated from laboratory mice in Japan (Ichihashi and Matsumoto, 1966), China, the USSR and many countries in Europe (Fenner *et al.*, 1989). Strains of virus from different parts of the world differ with respect to mouse pathogenicity (Fenner, 1958) and plaque character (Ichihashi and Matsumoto, 1966). Changes may have taken place either in the hypothetical wild animal host, or in the domestic mouse. Although the virus has not been isolated from the wild, a serological study in Britain (Kaplan *et al.*, 1980) showed that an *Orthopoxvirus*, probably either ectromelia or cowpox virus, is present in field mice and voles.

Cowpox virus.

Cowpox virus is more distantly related to variola virus. It is easily identified by the intense haemorrhagic pocks it produces on the CAM (Downie, 1939; Dumbell, 1968), the presence of large A-type inclusion bodies in infected cells (Downie, 1939) and its *HindIII* restriction enzyme map (Mackett and Archard, 1979), which differs from those of the other *Orthopoxviruses.* Like vaccinia and monkeypox viruses, cowpox virus has a broad host range. Many different animals (domestic cat, zoo and circus animals) in Europe and the USSR have been infected with the virus (Fenner *et al.*, 1989). The isolation of cowpox virus from two wild gerbils and a suslik in Turkmenia in 1974 (Marennikova *et al.*, 1978) suggests that rodents may be the natural reservoir host for this virus. Although the virus acquired its name from the disease it caused in cattle, cattle are probably mere incidental hosts.

Vaccinia virus.

Vaccinia virus was used in the eradication of smallpox (Fenner *et al.*, 1988). Although it is the *Orthopoxvirus* most studied, its origin is still obscure (Baxby, 1977).

THYMIDINE KINASE: SELECTION SYSTEM, ENZYME AND EVOLUTION

Selection system

The thymidine kinase (TK) gene interested poxvirologists primarily because of its potential use in the construction of recombinant (vaccinia) viruses (Mackett *et al.*, 1984). Conditions have been established for the selection of both TK^+ (using methotrexate) and TK^- (using 5-bromodeoxyuridine) phenotypes (Mackett *et al.*, 1985). Furthermore, interruption of the TK gene reduces the virulence of vaccinia virus (Buller *et al.*, 1985), a factor which would be favourable to a recombinant vaccine. Many foreign genes have been cloned into the TK gene of vaccinia virus, both for research purposes as well as vaccine development. The list of foreign genes expressed in vaccinia virus as well as other poxviruses is long, and will not be discussed. The numbers of proteins which had been expressed in vaccinia virus three

years ago were at least 10 cellular, 45 viral, 5 bacterial, 3 protozoan, 1 yeast and 1 bacteriophage (Hruby, 1990). Since then many more, especially cellular, proteins have been expressed in this system. A number of review articles have been published on the subject of vaccinia virus as an expression vector (Piccini and Paoletti, 1988; Hruby, 1990; Moss, 1991); and Baxby (1993) has reviewed the use of recombinant poxviruses other than (and including) vaccinia virus as vaccines.

Enzyme

Thymidine kinase (officially named ATP:thymidine 5'-phosphotransferase, EC 2.7.1.21) is associated with DNA synthesis and functions in the salvage pathway by ATP-dependent phosphorylation of thymidine to thymidine 5'-monophosphate. Since TK activity is increased in tissue and cell cultures undergoing DNA synthesis (Bucher, 1963), viral TK enzymes may be required to maintain dTTP levels in nondividing cells.

There are two groups of thymidine kinases (TKs), the herpesvirus/deoxycytidine kinases, and the poxvirus and eukaryotic TKs (Gentry, 1992). The herpesvirus TKs have a broader substrate specificity, with the ability to phosphorylate dT, dC (Jamieson and Subak-Sharpe, 1974; Jamieson *et al.*, 1974) and dTMP (Chen and Prusoff, 1978). They may utilize both ATP and CTP as phosphate donors (Kit *et al.*, 1974), and are not inhibited by dTTP (Cheng, 1976). In contrast, the poxvirus/eukaryotic enzymes specifically phosphorylate dT using ATP exclusively as the phosphate donor, and they may be subject to end product (dTTP) inhibition (Ives *et al.*, 1963). Although Bedson (1982) has shown variola (and whitepox) viruses to differ from the other *Orthopoxvirus* species in the sensitivity of their TK enzymes to dTTP inhibition (vaccinia virus is insensitive to dTTP inhibition), Kit and Dubbs (1965) have shown that purified vaccinia virus TK is sensitive to dTTP inhibition. This indicates that another component (removed during purification of the vaccinia virus enzyme) is involved in supressing the dTTP-sensitivity phenotype in vaccinia (and other *Orthopox*-) virus-infected cells ie. the difference between variola virus and the other *Orthopoxviruses* with respect to dTTP inhibition of their TK enzymes is not due to a difference in primary amino acid sequence - as Esposito and Knight (1984) suggest.

The poxvirus/eukaryotic TK enzyme is a homotetramer of about 80K (Black and Hruby, 1990a). By comparison the herpesvirus TK, which has subunits twice as large as those of the poxviruses, is a homodimer of about 80K (Kit, 1985). Somewhat different, the human mitochondrial enzyme consists of a monomer of 29K (Jansson *et al.*, 1992). At present the gene sequence has not been determined for this enzyme.

Vaccinia virus TK is the prototypic type II TK (Black and Hruby, 1990a). Its gene has been mapped to the *HindIII* J fragment (Weir *et al.*, 1982; Hruby and Ball, 1982) and sequenced (Weir and Moss, 1983; Hruby *et al.*, 1983). Comparing the deduced TK amino acid sequence with a number of other TK sequences, Black and Hruby (1990b) identified seven domains in which there was complete identity. The TK sequences compared were

from variola and monkeypox viruses (Esposito and Knight, 1984), mouse (Lin *et al.*, 1985), chicken (Kwoh and Engler, 1984), and human (Flemington *et al.*, 1987). The domains I and III have been identified as ATP-binding (Black and Hruby, 1990b) and Mg^{2+}-binding (Black and Hruby, 1992a) respectively. An arginine residue, present in corresponding positions between domains II and III in the herpes and poxvirus TKs, is thought to be important for substrate (dT) binding (Robertson and Whalley, 1988). Domain IV has been shown to to contain a glutamine amino acid residue which is essential for feedback inhibition by dTTP (Black and Hruby, 1992b). This glutamine residue is present in most type II TKs aligned by Gruidl *et al.* (1992) (including variola, vaccinia and monkeypox viruses). It is not conserved in fowlpox virus, swinepox virus, *Entomopoxvirus* and African Swine Fever virus. Whether these viral TKs are inhibited by dTTP is not known. This same domain in the type I TKs has recently been shown to contain the active site of the enzyme (Fetzer *et al.*, 1993). In the type I TK enzymes an aspartate residue is necessary for phosphorylation of the substrate. This aspartate is in the corresponding position to the glutamine residue required for dTTP inhibition in the type II TKs. The combined results of Black and Hruby (1992b) and Fetzer *et al.* (1993) suggest that dTTP inhibition occurs by direct binding of dTTP to the phosphorylating site of the enzyme.

Evolution

The thymidine kinase (TK) gene sequences have been determined for a number of poxviruses, herpesviruses, prokaryotes and eukaryotes. At present this is the most extensive sequence comparison made, both amongst the poxviruses, and between poxviruses and other organisms, to determine their evolutionary relatedness. However, the TK genes of a number of *Orthopoxvirus* species (cowpox virus, ectromelia virus, taterapox virus, raccoon poxvirus and volepox virus) have not yet been reported.

The first comparison made was one between vaccinia virus and three spontaneous TK^- mutants (Weir and Moss, 1983). All the mutants had a single-base reiteration which caused a frame shift in the coding sequence. This mutation was thought to have resulted from stuttering of the DNA polymerase enzyme during viral DNA replication. A similar effect was seen in an HA^- mutant of rabbitpox virus, except the mutant arose by deletion of a single A nucleotide (Brown *et al.*, 1991). Reversion to HA^+ was accompanied by reinsertion of the A.

Esposito and Knight (1984) reported the first DNA comparison between different *Orthopoxviruses*. They determined the DNA sequences of 1275bp in the TK region of variola and monkeypox viruses, and compared them to the vaccinia virus sequence (Weir and Moss, 1983; Hruby *et al.*, 1983). There were 41 positions of change in nucleotide sequence, of which 8 result in a change of amino acid residue. Two of the amino acid changes were unique to variola virus. Comparing these differences to a more recent alignment (Gruidl *et al.*, 1992), they are unlikely to influence the kinetic properties of the enzyme, as they are

present outside the conserved domains of the enzyme. Of the 41 positions of change, vaccinia virus differed from variola and monkeypox viruses in 21 positions, and variola and monkeypox viruses each had 10 unique differences. In each case two viruses had a common nucleotide where one virus differed. This data implies that variola and monkeypox viruses are more closely related to one another than they are to vaccinia virus.

A fourth *Orthopoxvirus* species, camelpox virus, has been sequenced in the TK gene region (Binns *et al.*, 1992). The camelpox virus DNA shows greatest similarity to that of vaccinia virus (98.3% identity), but also closely resembles variola (98.1%) and monkeypox (97.5) virus DNA.

The first phylogenetic analysis of poxviruses was made by Boyle *et al.* (1987). They had previously identified and cloned the fowlpox virus (FPV) TK gene (Boyle and Coupar, 1986). The DNA sequence of this gene was determined and the deduced amino acid sequence compared to the TK sequences of vaccinia, variola and monkeypox viruses, as well as chicken (Kwoh and Engler, 1984) and man (Bradshaw and Deininger, 1984). About 60% of the FPV amino acid sequence is homologous to those of the *Orthopoxviruses*. The TKs from chicken and man both have extended amino and carboxy termini. Comparing the entire sequence, there is 72.5% identity between the chicken and human TK sequences. The FPV TK has only about 40% identity to both of them; and the *Orthopoxvirus* TKs have about 55% amino acid identity to the vertebrate sequences.

The poxvirus TKs can be compared to the vertebrate TKs using different parameters to account for the insertions and deletions. Considering contiguous insertions or deletions as single event, and restricting the comparison to the core region, the vertebrate sequences are more closely related to those of the *Orthopoxviruses* than to fowlpox virus (and the *Orthopoxviruses* closer to vertebrates than fowlpox virus), implying that the ancestral protein was poxvirus-like. If one includes all residues, treating spaces as a 21st amino acid residue, all the poxviruses group together and the vertebrates have a separate branch. The second tree topology is more "intuitively acceptable" as it would accomodate the hypothesis that poxviruses co-evolved with their hosts. It is, however, evident that the poxviruses have a higher rate of evolutionary change than the vertebrates. If one assumes that co-evolution has taken place, the poxvirus TK sequences have a four-fold increase in rate of evolutionary change (Boyle *et al.*, 1987). It makes sense that mutations accumulate faster in organisms with shorter generation times, due to the increased number of replication cycles. Since poxviruses encode their own DNA polymerase enzyme and other proteins required for DNA replication, consideration must also be given to the relative fidelity of replication of the different organisms.

Gentry *et al.* (1988) have extended the phylogenetic analysis of the TK sequences to include shope fibroma virus of rabbits (Upton and McFadden, 1986), Chinese hamster (Lewis, 1986) and mouse (Lin *et al.*, 1985). They, too, had difficulty constructing a single meaningful tree which would correlate with co-evolution of the virus with its host. The

simplest tree implied that the fowlpox virus TK was a progenitor of the vertebrate TK. The authors showed that, by constructing two trees, one for vertebrates and one for poxviruses, and lengthening the branches of the vertebrate tree, the data fits in with an obligate virus/ host cell relationship. The positions from which fowlpox and rabbit fibroma viruses branch in the virus tree correspond to the branch points of the chicken and rodent branches respectively in the vertebrate tree. Interestingly, the three *Orthopoxviruses*, in particular, variola virus, diverged very much later than the corresponding human branch. This may represent a more recent introduction of the virus into humans. It is possible that an ancestor to variola virus co-evolved in a different host animal before entering into humans.

The display of data in two trees is useful for comparing the evolution of genes within organisms which have different rates of change. However, caution must be taken correlating virus with host evolution, as we still do not know the natural host for many of the *Orthopoxviruses*. On the other hand, this kind of analysis may indicate in which host animal a particular virus has co-evolved.

Later TK comparisons have included African swine fever (ASF) virus (Blasco *et al.*, 1990), *Capripoxvirus* (Gershon and Black, 1989), bacteriophage T4 (Valerie *et al.*, 1986), *E. coli* (Bockamp *et al.*, 1990), swinepox virus (Feller *et al.*, 1991; Schnitzlein and Tripathy, 1991), myxoma virus (Jackson and Bults, 1992) and an *Entomopoxvirus* (Gruidl *et al.*, 1992). Bockamp *et al.* (1990) derived a distance matrix tree which divided the sequences into four main groups: 1) the vertebrates, 2) the poxviruses (later shown to include swinepox and myxoma viruses), 3) ASF virus and 4) *E. coli* and bacteriophage T4. Gruidl *et al.* (1992) showed the *Entomopoxvirus Amsacta moorei (AmEPV)* amino acid sequence to be more closely related to the poxvirus sequences (±45%) than the ASF virus sequence (31.4%). This is of interest because both ASF and AmEPV can replicate in arthropod hosts and ASF has many biological features in common with the poxviruses.

All poxviruses studied have a TK gene. This indicates that the gene was acquired at or near their time of origin. The origin of poxviruses is open to speculation. They may have originated from an ancestral eukaryotic cell and later become reduced in size and function to a virus dependent on its host for further replication cycles. Alternatively, they may have originated as a small virus which gradually accumulated additional genes from its host cell. The first postulate is strengthened by viral genome comparisons. The genomes of insect (Hall and Hink, 1990) and bird (Coupar *et al.*, 1990) poxviruses are larger than those of mammalian poxviruses. Assuming the viruses originated at the time of animal divergence, their genomes have diminished as they have entered higher animal forms. A comparison of the number and nature of genes encoded by different poxviruses infecting different animals may reveal a pattern of genetic flux. The most informative genes in this respect would be those directly involved in gene expression and DNA replication. Genes which encode products related to viral pathogenicity and host range probably evolved (or were acquired) later on in the evolution of the viruses; and, being terminally located in the genome, have been subject to rearrangement.

CHAPTER 2

MATERIALS, METHODS AND SEQUENCING STRATEGIES

Viruses 24

Virus growth and DNA extraction 26

DNA techniques 26

DNA amplification by PCR 26

- **Amplification of *Orthopoxvirus* DNA equivalent to vaccinia virus E5R** 26
- **Amplification of the DNA flanking the major deletion in monkeypox virus** 27

DNA cloning 27

- **Plasmids** 27
- **Construction of plasmid pHarRsa** 27
- **The E5R gene of vaccinia virus strain Dairen** 28
- **Construction of pGptC/B** 29
- **Constructs made for DNA sequencing purposes** 30

DNA sequencing 31

Northern blot analysis of RNA 36

Cell culture techniques 36

Pathogenicity experiments 37

CHAPTER 2

MATERIALS, METHODS AND SEQUENCING STRATEGIES

VIRUSES

All viruses were from Prof.K. Dumbell's collection. Table 2.1 lists the year in which they were first isolated together with the relevant reference(s).

Monkeypox virus origins are as follows: *Human monkeypox*: Liberia I and II from Liberia; Benin and MP-82 from Nigeria; MP-266 from Sierra Leone; 86-21 and 85-240 from Bumba, Zaire (Mongala region of Equateur province), 86-112 from Bongandanga, Zaire (Mongala region of Equateur province), Z1324 from Gemena, Zaire (Sud-Ubangi region of Equateur province) and Z241 from Katako-Kombe, Zaire (Kasai-orientale). *Squirrel isolate*: 85-249 from Bumba, Zaire (Mongala region of Equateur province). *Isolates from captive animals*: Denmark from Copenhagen, Denmark; Rotterdam from Antwerp, The Netherlands; and Paris from Paris, France. Prier and Espana are from the United States.

Variola virus DNA was in the form of recombinant bacterial plasmids (Hamilton *et al.*, 1985). Harvey is a strain of variola major which was isolated in Middlesex, U.K. and Somalia is a strain of variola minor which was isolated from the last natural case of smallpox (from Somalia).

Vaccinia virus strain Dairen originated from a Japanese soldier. Reference has been made to the Copenhagen and Western Reserve (WR) strains. These originated in Ecuador and Britain respectively (Wokatsch, 1972).

Cowpox Larkin was obtained from a stock which had been passaged only three times on chick chorioallantoic membranes (CAMs), whereas cowpox Brighton has had numerous passes. Both viruses give characteristic haemorragic pocks on CAMs. They were both isolated in Britain.

Tatera poxvirus (gerbilpox) originated in Dahomey (Benin), West Africa.

Camelpox was isolated from a camel in Somalia.

Ectromelia was isolated from an outbreak of mousepox in Britain.

Raccoon poxvirus is the only *Orthopoxvirus* examined which was found naturally in the Americas (isolated in Maryland). All the others mentioned are thought to have originated in Europe, Asia or Africa.

TABLE 2-1

Orthopoxviruses from which DNA was examined

Virus	Year Isolated (Reference)
Monkeypox virus	
Denmark	1958 (von Magnus *et al.*, 1959)
Paris	1968 (Milhaud *et al.*, 1969; Arita *et al.*, 1972)
Prier	1959 (Prier and Sauer, 1960)
Espana	1967 (Arita *et al.*,1972)
Rotterdam	1965 (Peters *et al.*, 1966; Gispen *et al.*, 1967)
LiberiaI	1970 (Lourie *et al.*, 1972)
LiberiaII	1970 (Lourie *et al.*, 1972)
MP-266	1970 (Lourie *et al.*, 1972)
MP-82	1971 (Lourie *et al.*, 1972)
Benin	1978 (Breman *et al.*, 1980)
Z241	1972 (Marennikova, not published)
Z1324	1979 (Marennikova, not published)
86-21	1986 (Khodakevich *et al.*, 1987b)
85-249	1985 (Khodakevich and Jezek, 1986)
85-240	1985 (Khodakevich *et al.*, 1987b)
86-112	1986 (Khodakevich *et al.*, 1987b)
Variola virus	
Harvey	1944 (Downie and Dumbell, 1947)
Somalia	1977 (Arita, 1979)
Vaccinia virus	
Dairen	before 1954 (Tagaya *et al.*, 1961)
Cowpox virus	
Larkin	1959 (Dumbell, not published)
Brighton	1937 (Downie, 1939)
Tatera poxvirus	
Gerbilpox	1968 (Lourie *et al.*, 1975)
Camelpox virus	
Camelpox 903	1978 (Kriz, 1982)
Ectromelia virus	
Mill Hill	1959 (Marchal, 1930)
Raccoon poxvirus	
Raccoon	1964 (Herman, 1964)

VIRUS GROWTH AND DNA EXTRACTION

Standard methods were used for the inoculation of fertile eggs (Westwood *et al.*, 1957) and cell cultures (Mackett *et al.*, 1985). Virus was purified and DNA extracted as described by Dumbell and Richardson (1993). Virus prepared for future cell culture infections was resuspended in McIlvain's buffer and stored at -70°C.

Monkeypox virus Z241 DNA was PCR-amplified from a crude DNA extract from a glycerol stock of virus following a method described by Kellogg and Kwok (1990).

DNA TECHNIQUES

Unless otherwise indicated, standard molecular biology techniques described by Maniatis *et al.* (1982) or Sambrook *et al.* (1989) were used.

Large-scale preparations of plasmid DNA were made according to the method described by Greenaway and Dale (1983).

The Birnboim and Doly (1979) method of small scale plasmid DNA preparations was slightly modified by omitting the first precipitation step and precipitating with isopropanol instead of ethanol.

DNA AMPLIFICATION BY PCR

PCR amplification of DNA was performed according to standard methods (PCR Protocols: A guide to methods and applications, 1990) on either a Techne or a LEP thermocycling machine. TaqI DNA polymerase was supplied by BRL (Boehringer).

Amplification of *Orthopoxvirus* DNA equivalent to vaccinia virus E5R

Oligonucleotide primers were designed to bind to DNA sequences flanking the vaccinia virus E5R gene. Reading 5' to 3' the primer sequences are 1:GATGATTTTTCCATGGCCCATT and 2: GAGCTAGTACATGATTGAGGGT. Primer 1 (positions 52173 to 52194 of vaccinia virus Copenhagen) includes the *NcoI* site within the E4L gene, and primer 2 binds to positions 53641 to 53620 of the vaccinia virus Copenhagen strain (Goebel *et al.*, 1990b). Thirty or more cycles were performed using an annealing temperature of 50°C for 1 minute. Extension times were either 3 minutes or 1.5 minutes at 72°C.

Amplification of the DNA flanking the major deletion in monkeypox virus

Primer sequences used for differentiating the two monkeypox virus groups were 1: 5'-CCCAATTTGTTTGGAG-3' and 2: 5'-CGAACAGGCATACATG-3'. These correspond to positions 620-636, and 1102-1087 respectively in fig. 3.3. Annealing was performed at 40°C and all cycling steps were of 30 seconds duration for 40 cycles.

DNA CLONING

Plasmids

pUC18 and pUC19 were provided by Dr Lafras Steyn, Medical Microbiology, UCT, Cape Town. Their restriction enzyme maps are shown in appendix A. Plasmid pGpt07/14 was kindly provided by Dr. David Boyle, Australian Animal Health Laboratory, Geelong, Australia. Its restriction enzyme map, together with the multiple cloning site for pMTL23 (obtained from Chambers, Bartow and Minton, PHLS, Porton Down, Salisbury) is also shown in appendix A. Plasmid pAL-1 was kindly provided by Dr Mike Mackett, Manchester, U.K.

The *HindIII* D fragment of variola virus, Somalia was available as a pAT153 recombinant plasmid (Hamilton *et al.*, 1985). Cloned DNA fragments of variola virus, Harvey and monkeypox virus, Denmark, corresponding to the vaccinia virus E5R region, were available as *EcoRI-BamHI* fragments cloned into pUC19 (Pare, 1988). They are designated pHarE/B and pDenE/B repectively.

Standard methods were used for DNA cloning (Maniatis *et al.*, 1982; Sambrook *et al.*, 1989). *E. coli* cells (JM109, LKIII or DKI) were made competent and transformed as described by Chung and Miller (1988) or Sambrook *et al.*, (1989).

Construction of plasmid pHarRsa.

Fig. 2.1 diagrammatically shows the construction of a recombinant plasmid, pHarRsa, which contains 81bp of variola virus DNA corresponding to part of the major deletion in monkeypox virus. The dashes represent DNA sequences deleted in monkeypox virus. A subclone of variola virus containing the E5R equivalent ORF was digested with *HincII*, which cleaves within the ORF. The DNA was shortened bidirectionally by Exonuclease III digestion, and then cleaved with *KpnI*. Within the 6bp *KpnI* site is a 4bp *RsaI* site. After religation of the shortened plasmid, the DNA absent in monkeypox virus could be isolated on an *RsaI* fragment, due to the presence of an *RsaI* site within the stretch of DNA of interest. This *RsaI* fragment was ligated into the *SmaI* site of pUC19. It was subsequently isolated as an *EcoRI/ HindIII* fragment for probe labelling.

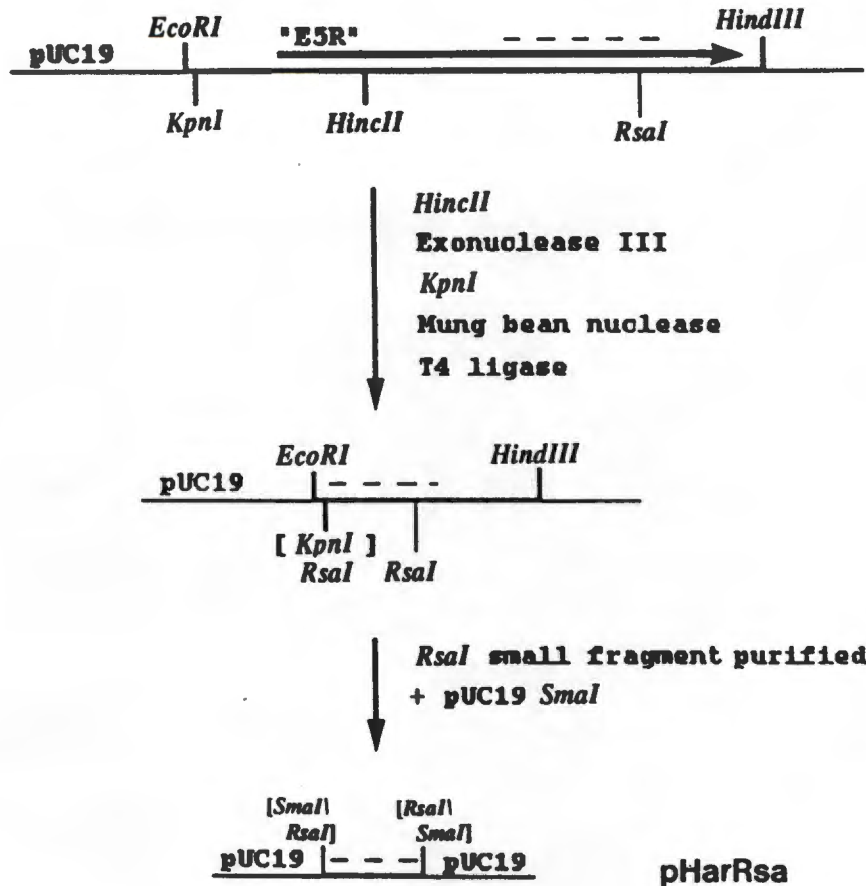


Fig. 2.1 Construction of plasmid pHarRsa

The E5R gene of vaccinia virus strain Dairen.

Fig. 2.2 shows the location of E5R in the genome. The *HindIII* E fragment was eluted from a 0.8% agarose gel (Seth, 1984), digested with *EcoRI* and "shotgun" cloned into *EcoRI*-digested pUC19. A recombinant plasmid containing the E5R gene on a 3.2kb *EcoRI* fragment was designated **pDIE3.2Eco**. This plasmid was digested with *BamHI* and re-ligated to produce the plasmid **pDIEE/B**, which contains an insert of 2.9kb.

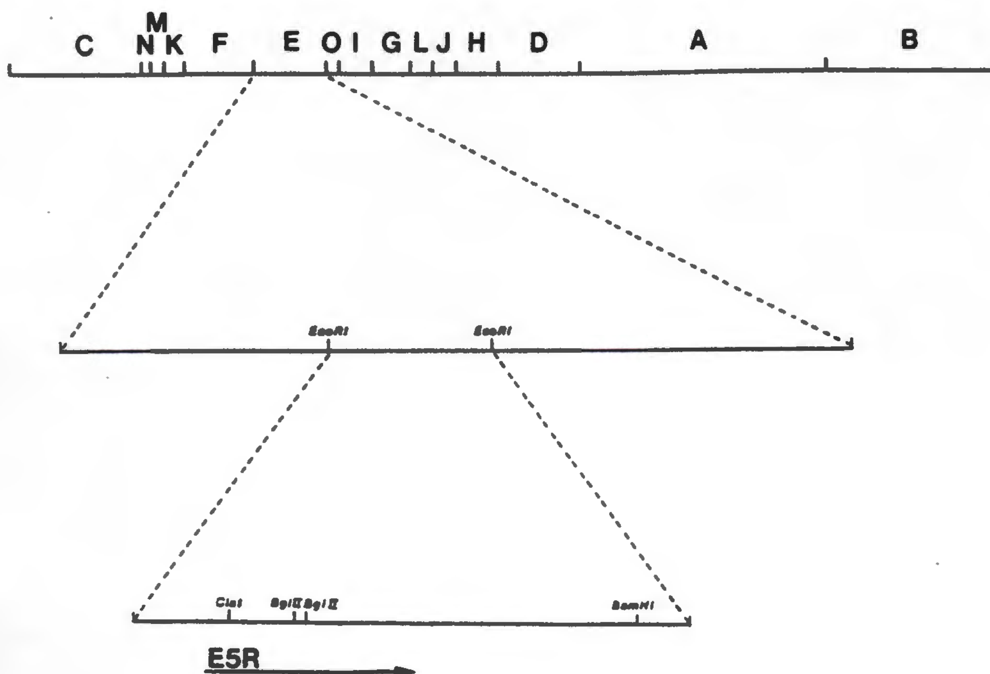


Fig. 2.2 *HindIII* map of vaccinia virus showing the location of the E5R gene.

Construction of plasmid pGptC/B

A plasmid was designed to recombine with vaccinia virus *in vivo* to produce a virus with 438bp of the E5R ORF replaced with a marker gene, gpt. The cloning strategy is shown in fig. 2.3. An *EcoRI* fragment containing the gpt gene (from the plasmid pGpt07/14) was first subcloned into the *EcoRI* site of pMTL23 (to produce **pGptMTL23**) so that it could be cleaved again with more appropriate restriction enzmes, viz. *ClaI* and *BamHI*. The vector, initially comprising the vaccinia virus 3.2kb *EcoRI* fragment harbouring the E5R gene, was digested with *BamHI* + *HindIII* and religated to produce **pDIE2.9E/B/H**. This resulted in the deletion of 0.3kb of vaccinia virus DNA as well as the multiple cloning site of pUC19. This step was included so that unique restriction enzyme sites could be generated between

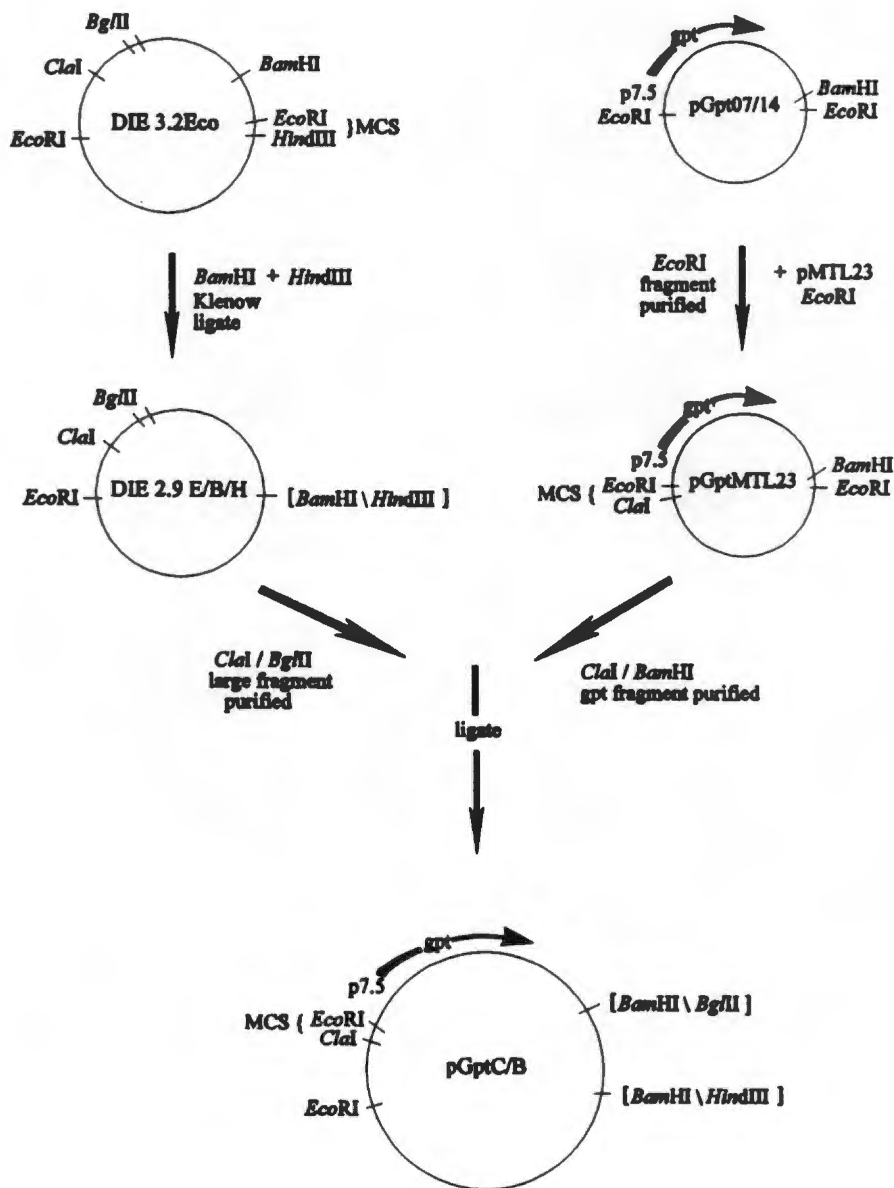


Fig. 2.3 Construction of plasmid pGptC/B for interruption of vaccinia virus E5R.

the flanking sequences in the final construct. These sites could be useful if this plasmid were to be used for future cloning of genes into vaccinia virus.

The final construct (**pGptC/B**) was made by ligation of the gpt gene isolated on a *ClaI-BamHI* fragment with the vector digested with *ClaI* and *BglII*. Three unique cloning sites - *SmaI*, *SacI* and *ClaI* - were generated between the left arm and the gpt gene.

Constructs made for DNA sequencing purposes

Initially continous DNA sequences were obtained by treating the insert to be sequenced with Exonuclease III and mung bean nuclease. Subclones were sequenced using the commercially available forward and reverse primers. In addition, subclones were constructed from appropriate restriction enzyme fragments. The sequencing strategies are diagrammatically represented in figures 2.4 to 2.7 for those sequences obtained from fragments which were subcloned. In most cases arrow labels refer to clones which were constructed. Numbers preceded by a "P" indicate that the sequence was obtained using a synthetic oligonucleotide as a primer.

Variola virus, Harvey. Plasmid **pHarE/N** was subcloned from pHarE/B by double digestion with *EcoRI* and *NcoI*, filling in with the Klenow fragment of DNA polymerase, and ligation. Plasmids **pHarE1, pHarB6**, **pHarB3** and **pHarA2** were obtained from the first cloning experiment described, using Exonuclease III and mung bean nuclease to trim the variola virus insert. Plasmid **pHarHc** is a subclone of pHarE/B digested with *HincII* and religated.

Variola virus, Somalia. **pSom4.0Eco** is the 4.0kb EcoRI fragment of the *HindIII* D fragment of variola virus, Somalia cloned into pUC18. It was obtained by ligation of the agarose gel-purified (Seth, 1984) 4.0kb *EcoRI* fragment of Somalia *HindIII* D with pUC18 digested with *EcoRI*. **pSomE/B** was subcloned from pSom3.2Eco by digestion with *BamHI* and re-ligation. **pSomE/N** is the product of *EcoRI* + *NcoI* digestion, Klenow blunt-ending and re-ligation of pSomE/B. **pSomE/C** was produced by *EcoRI* + *ClaI* digestion of pSomE/B followed by Klenow blunt-ending and religation. **pSomBg/B** was constructed from pSomE/B by digestion with *BglII* + *BamHI* and religation. **pSomHc** is a subclone of pSomE/B digested with *HincII* and religated.

Monkeypox virus, Denmark. Plasmids **pDenE/N** and **pDenHc** were constructed from pDenE/B in the same was as pHarE/N and pHarHc were constructed from pHarE/B respectively.

Vaccinia virus, Dairen. Plasmid pDIEE/B was shortened by Exonuclease III digestion from the *BamHI* site followed by SI nuclease digestion, the vector being protected by *PstI* digestion (Yanish-Perron *et al.*, 1985). Plasmids, **pDIEex12, pDIEex16, pDIEexB3**, **pDIEex24** and **pDIEex9** were generated in this experiment. Additional subclones were made: **pDIEE/N** - *EcoRI* + *NcoI* digestion of pDIEE/B followed by Klenow blunt-ending and re-ligation; **pDIE2.3Hc** - *HincII* digestion of pDIEE/B, elution of the 2.3kb fragment (Seth, 1984), and ligation into pUC19 digested with *HincII*; and **pDIEHc** - pDIEE/B digested with *HincII* and

religated. Plasmid **pDIE2.4ex** was constructed by Tonia Eeckhout (by Exonuclease III digestion from the *EcoRI* end of the fragment).

The E5R-equivalent region of all African monkeypox viruses, cowpox viruses (strains Brighton and Larkin), camelpox virus (strain 903), ectromelia virus (strain Mill Hill) and taterapox virus (gerbilpox virus). PCR-amplified DNA fragments (using primers 15 and 16, see table 2-3, page 35) were ligated into the *SmaI* site of T-tailed pUC18 or pUC19 (Marchuk *et al.*, 1991). Clones have been given the following designations: **pMPLIB** - DNA from monkeypox virus LiberiaII; **pMPBEN** - monkeypox virus Benin; **pMP86-21** - monkeypox virus Zaire 86-21; **pMP86-112** - monkeypox virus Zaire 86-21; **pMP1324** - monkeypox virus Zaire 1324; **pCOWBRI** - cowpox virus Brighton; **pCOWLAR** - cowpox virus Larkin; **pCAM903** - camelpox virus 903; **pECTMH** - ectromelia virus Mill Hill; and **pGERBIL** - taterapox virus gerbilpox virus.

Three truncated clones were obtained from PCR cloning. **pMP241** is a recombinant plasmid containing approximately 400bp of DNA from monkeypox virus Zaire 241 corresponding to the latter half of the E5R-equivalent region and the beginning of the E6R-equivalent ORF of vaccinia virus (Goebel *et al.*, 1990b). **pECTtc** is a clone of ectromelia virus DNA which contains sequences from the *NcoI* site to position 1118. **pRACB4** is a truncated clone of raccoon poxvirus DNA, containing ±500 bp which includes DNA corresponding to the latter half of the E5R-equivalent region and the beginning of the E6R-equivalent ORF of vaccinia virus (Goebel *et al.*, 1990b).

DNA SEQUENCING

Double-stranded DNA was sequenced by the Sanger dideoxy method (Sanger, 1977) using ^{35}S-dATP (Amersham) and sequenase kits (United States Biochemical). Template DNA was prepared according to a modified method of Birnboim and Doly (1979). Forward and reverse m13 primers were used as well as other synthesized oligonucleotides. Table 2.2 gives the DNA sequences of the primers used.

Figures 2.4 to 2.7 diagrammatically show the DNA sequencing strategies employed for variola virus Harvey, variola virus Somalia, monkeypox virus Denmark and vaccinia virus Dairen respectively. Arrows which are not labelled (for parts of variola virus Harvey and monkeypox virus Denmark) refer to DNA sequences which were obtained prior to registration for this degree and were written up as part of a master's thesis.

DNA sequences obtained from cloned PCR-amplified viral DNA were determined by primer walking. Table 2-2 gives the DNA sequence of all primers used and fig. 2.8 shows the corresponding position in vaccinia virus Dairen to which they bind.

The primers used for DNA sequencing of the different viral fragments are listed in table 2-3. Only part of the DNA sequence obtained from raccoon poxvirus was confirmed by sequencing in two directions. All the other sequences described in this thesis have been confirmed by sequence determination of both DNA strands.

DNA sequences were analysed using GENEPRO (Riverside Scientific, 1985, 1988), and aligned using clustal V (Higgins *et al.*, 1991). FastA (Pearson and Lipman, 1988) and Blast (Altshul *et al.*, 1990) database searches were performed through GCG (Devereux *et al.*, 1984).

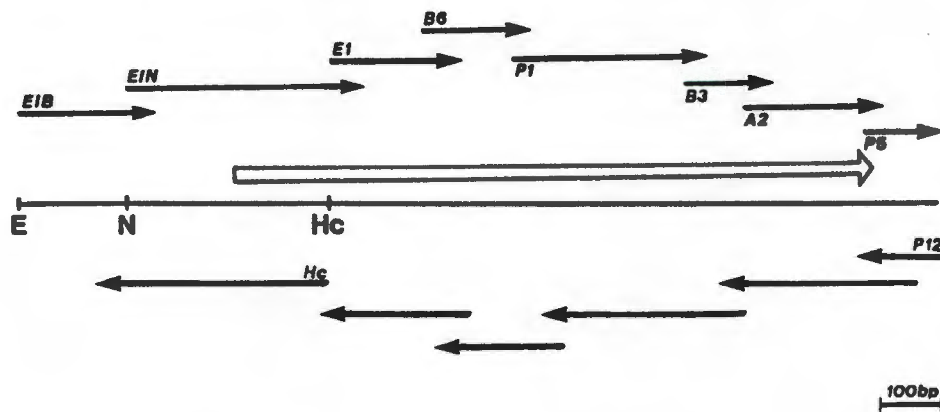


Fig. 2.4 Strategy for determining the DNA sequence of variola virus Harvey in the region corresponding to vaccinia virus E5R. The equivalent ORF is indicated by an open arrow. Filled arrows show stretches of DNA sequenced using forward or reverse primers, except for those labelled P5 and P12, which refer to primers listed in table 2-2. Labels refer to clones described in the text (omitting the pHar prefix). Unlabelled arrows refer to DNA sequences determined for a previous thesis.

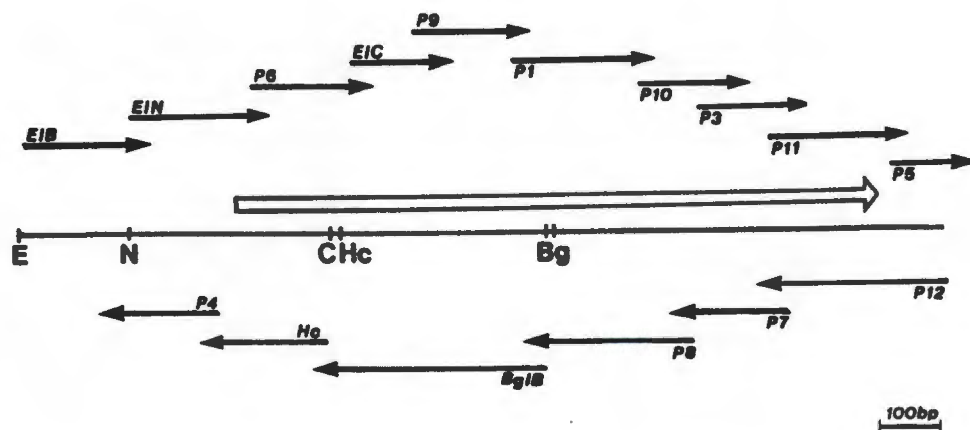


Fig. 2.5 Strategy for determining the DNA sequence of variola virus Somalia in the region corresponding to vaccinia virus E5R. The equivalent ORF is indicated by an open arrow. Filled arrows show stretches of DNA sequenced using forward or reverse primers, except for those labelled with numbers preceded by a "P", which refer to primers listed in table 2-2. Labels refer to clones described in the text (omitting the pSom prefix).

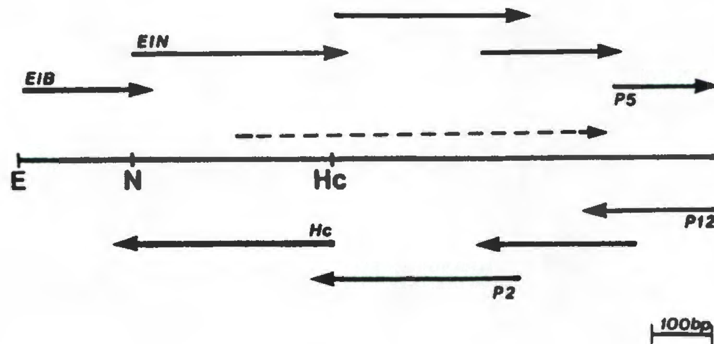


Fig. 2.6 Strategy for determining the DNA sequence of monkeypox virus Denmark in the region corresponding to vaccinia virus E5R. The region equivalent to vaccinia virus ORF E5R is indicated by a broken arrow. Filled arrows show stretches of DNA sequenced using forward or reverse primers, except for those labelled with numbers preceded by a "P", which refer to primers listed in table 2-2. Labels refer to clones described in the text (omitting the pDen prefix).

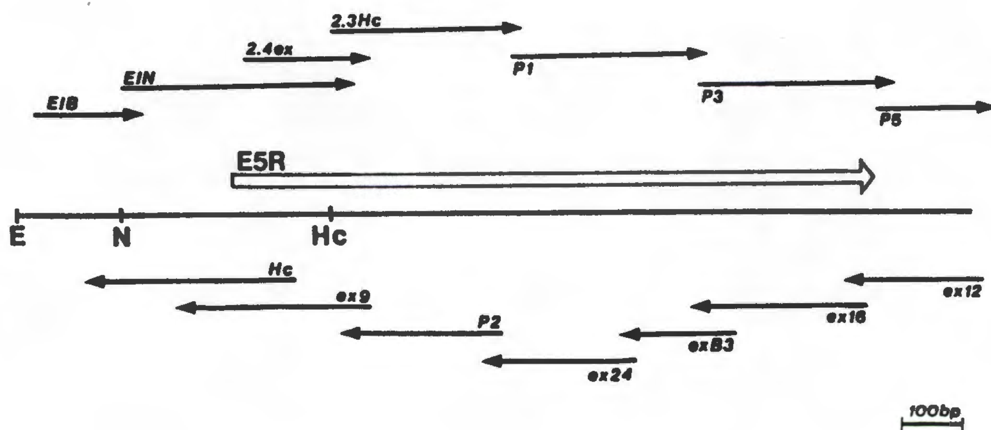


Fig. 2.7 Strategy for determining the DNA sequence of the vaccinia virus Dairen E5R region. Filled arrows show stretches of DNA sequenced using forward or reverse primers, except for those labelled with numbers preceded by a "P", which refer to primers listed in table 2-2. Labels refer to clones described in the text (omitting the pDIE prefix).

TABLE 2-2

Primers used during the course of this study. The location of primer sequences is diagrammatically shown in fig. 2.8 using vaccinia virus (Dairen) as a reference. (Primer sequences are not all identical to vaccinia virus Dairen).

Primer	DNA sequence (5' to 3')	Position in fig. 5.5 (*)
forward (for)	GTAAAACGACGGCCAGT	(pUC)
reverse (rev)	CAGGAAACAGCTATGAC	(pUC)
1	CCCAATTTGTTTGGAG	636-651
2	CTCCAAACAAATTGGG	651-636
3	CATCCCAGGAATTGG	946-960
4	ATATCAACATAATAATG	200-184
5	CTTTGTTAGTGAATAGGC	1215-(1232)
6	GGATGAGTAACTACTA	107-122
7	CGAACAGGCATACATG	1118-1103
8	CCTGGGATGTCACAC	954-940
9	TGATTAGGACTACTACT	487-503
10	GCAAGGTATCGCCTA	855-869
11	TTGTTCATGATATAGTTG	1082-1099
12	TATAAGATACTTTCTACG	(in E6R)
13	GTATCAAATCGATAGTG	329-345
14	CACTATCGATTTGATAC	345-329
15	GATGATTTTTCCATGGCCCATT	(-10)-12 (PCR primer)
16	GAGCTAGTACATGATTGAGGGT	(in E6R, PCR primer)
17	TAGAATCTTTCTTCCAC	763-747
18	TTACTCGGATTCTGTG	604-589
19	CAGATTTTCCCTTGATTC	900-883
20	CTTTGTCCTCTTCTTATC	383-366
21	GTATTGAGAAAGGAAAAG	(1257-1240, cowpox)
22	GAGGGTATCTTCTTCAG	914-898

* Primer corresponds to DNA sequence not shown in fig 5.5

TABLE 2-3

Primers used for DNA sequencing of the equivalent region in different Orthopoxviruses. Primer sequences are listed in table 2-2.

Virus	Primers used
monkeypox viruses	
LiberiaII	rev, 6, 13, 9, 1, 5, 16, 12, 7, 18, 14, 4.
Benin	rev, 6, 13, 9, 1, 5, for, 12, 7, 18, 14, 4.
Zaire 86-21	for, 6, 13, 9, 1, 11, 5, rev, 16, 12, 7, 17, 18, 14, 4.
Zaire 86-112	for, 6, 13, 9, 1, 11, 5, rev, 12, 7, 17, 18, 14, 4.
Zaire 1324	for, 6, 13, 9, 1, 11, 5, 16, 12, 7, 17, 18, 14, 4.
Zaire 241	for, 11, 5 rev, 12, 7.
cowpox viruses	
Larkin	for, 6, 9, 1, 10, 11, rev, 12, 21, 7, 19, 17, 18, 20, 4.
Brighton	for, 6, 9, 1, 10, 11, 5, rev, 12, 21, 7, 19, 17, 18, 20, 4.
ectromelia virus	
Mill Hill	for, 6, 13, 9, 1, 10, 11, rev, 12, tc*rev, 22, 17, 18, 20, 4.
taterapox virus	
gerbilpox	for, 15, 6, 13, 9, 1, 10, 11, 5, rev, 12, 7, 19, 17, 18, 20, 4.
camelpox virus	
camelpox 907	for, 6, 13, 9, 1, 10, 11, 5, rev, 7, 8, 17, 18, 14, 4.
raccoon poxvirus	
clone B4	for, rev, 12.

* tc refers to a clone which contains a truncated insert.

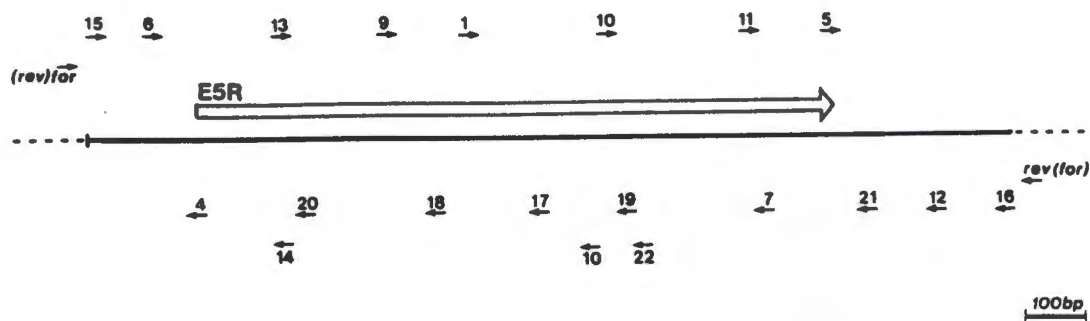


Fig. 2.8 Diagram showing the positions to which the primers bind using vaccinia virus Dairen E5R as a template. Primer sequences are listed in table 2-2.

NORTHERN BLOT ANALYSIS OF RNA

RNA was prepared as described by Howard and Smith (1989). Northern blotting was done according to Maniatis *et al.* (1982). ^{32}P-labelled DNA probes were prepared either by end-labelling (BRL) or nick translation (Amersham) followed by column purification. Using the end-labelled primer (5'-CGAACAGGCATACATG-3') hybridization was at 42°C in a solution of 25% formamide, 5 X SSPE, 2 X Denhardt's and 0.1% SDS. After overnight incubation the blot was washed in a solution of 0.1% SDS and 2 X SSPE for 20 minutes. The gpt-specific probe was prepared by nick-translation of pGptMTL23. Hybridization was left overnight at 42°C in a solution of 50% formamide, 5 X SSPE, 2 X Denhardt's and 0.1% SDS. The membrane was given three 15 minute washes at room temperature; one in 6 X SSPE, 1% SDS, and two in 2 X SSPE, 1% SDS.

The blot was stripped of its probe by boiling in 0.1% SDS for 45 minutes.

CELL CULTURE TECHNIQUES

Standard techniques were used. Cells were incubated at 37°C in a 5% CO_2 incubator. CV1 (monkey kidney), HF (human fibroblast) and RK13 (rabbit kidney) cells were available in the department and were maintained on MEM (minimal medium) containing 4% fetal bovine serum (Highveld Biologicals).

Virus infection and titrations were performed as described by Mackett *et al.*, (1985).

Marker rescue experiments

CV1 cells were infected with vaccinia virus, strain Dairen, at 0.005 pfu/cell. After adsorption for 1 hour at 37°C, the virus was removed, cells washed 3 times with Opti-MEM (Gibco Laboratories) and plasmid introduced into the cells as described by Chang and Brenner (1988) using lipofectin (BRL). After 6 hours the medium was replaced with gpt selection

medium (MEM + 1μg/ml mycophenolic acid (Sigma) + 250μg/ml xanthine (Sigma) and cells left for 48 hours. Virus-infected cells were harvested by three freeze-thaw cycles. The virus was passaged in CV1 cells under selective and nonselective conditions. Virus which only formed plaques under selection conditions was pock-purified.

As a transfection control the plasmid pAL-1 was transfected into virus-infected cells in the same manner, except no selection was applied to the infected cells. The cells were stained after 24 hours to test for ß-galactosidase expression (Chakrabarti *et al.*, 1985).

As a control for the system, pGpt07/14 was used to generate a TK^- recombinant virus.

At a later stage the selection conditions had to be modified. To reduce background false positive plaques the xanthine concentration was reduced to 25μg/ml and, to reduce cell toxicity the mycophenolic acid concentration was reduced to 0.5μg/ml.

PATHOGENICITY EXPERIMENTS

Young adult rabbits were injected intradermally with 0.1ml of virus diluted in physiological saline. 2 or 4 rabbits were injected on three separate occasions using virus titres of 10^5, 10^6 or 10^7 pfu/ml. The lesions produced were observed daily and photographed after 5 and 8 days.

CHAPTER 3

INDEPENDENT EVOLUTION OF VARIOLA AND MONKEYPOX VIRUSES

Summary 39

Introduction 39

Results and discussion 42

CHAPTER 3

INDEPENDENT EVOLUTION OF VARIOLA AND MONKEYPOX VIRUSES

SUMMARY

An open reading frame (ORF) present in the conserved region of variola and vaccinia viruses was shown to have a degenerate counterpart in monkeypox virus. Two strains of variola virus, Harvey and Somalia, were sequenced across 1260bp which included this ORF. The two sequences differed in only 1 position. The equivalent sequence in monkeypox virus could be aligned with 94% base similarity if four gaps were introduced; 24-, 1-, 391- and 24-bp in length. One base substitution results in a termination codon near the beginning of the ORF. Together, all theses changes indicate that the ORF, once present in a common ancestor, became truncated in monkeypox virus. By hybridization and PCR it was shown that no equivalent full length ORF is present within the entire monkeypox virus genome. The presence of an ORF conserved in variola virus which is degenerate in monkeypox virus indicates that monkeypox virus cannot be a direct progenitor of variola virus. Evidence suggests that the two viruses evolved independently from a common ancestor.

INTRODUCTION

This project was started by comparing the *HindIII* maps of variola and monkeypox viruses. Our aim was to locate a difference which occurred early in the evolutionary history of the two separate viral species. Such a change would most likely be found in the central conserved region of the genome, where it would be stably maintained. The termini, in constrast, are variable both within and between species, and would represent more recently acquired changes.

Secondly, we wished to identify a gene unique to variola virus. This would be evidence to suggest that monkeypox virus is not a progenitor of variola virus, since DNA not present anywhere else in the genome cannot simply be acquired. We were therefore interested in finding surplus DNA in variola virus.

The *HindIII* maps of variola and monkeypox viruses are shown in fig. 3.1 (Mackett and Archard, 1979). Focussing on the central conserved region, there is an additional *HindIII* site in variola virus between the *HindIII* M and E fragments, which lies within the corresponding *HindIII*D fragment of monkeypox virus. This difference was not investigated since the adjacent *HindIII* sites are co-incident, implying that a single base substitution event may have introduced the additional cleavage site into variola virus.

The second noticeable difference is the staggering of the *HindIII* sites dividing variola *HindIII* C and D fragments, and the corresponding monkeypox *HindIII* C and E fragments. If the two *HindIII* sites are, in fact, co-incident, then monkeypox *HindIII* C would be larger than variola *HindIII* C; and Variola *HindIII* D would be larger than monkeypox *HindIII* E. In this case there may be surplus DNA in variola *HindIII* D.

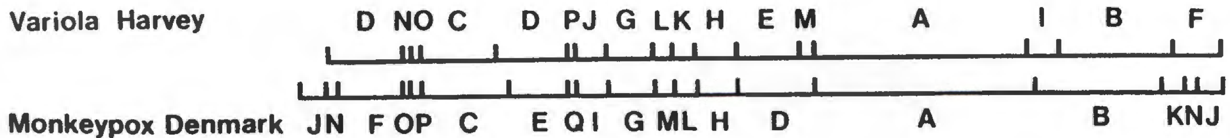


Fig. 3.1 *HindIII* maps of variola and monkeypox virus genomes (From Mackett and Archard, 1979).

Kinchington *et al.* (1984) have compared the *HindIII* C fragments by heteroduplex analysis. Three loops were formed by the heteroduplexes, one large one and two smaller ones. These loops probably represent surplus DNA in monkeypox *HindIII* C, indicating that the ends of the *HindIII* C fragments are homologous; and the *HindIII* sites separating monkeypox *HindIII* C and E; and variola *HindIII* C and D are probably in equivalent positions.

DNA sequence at the right hand end of variola virus *HindIII* D (15.3kb) matched that of the right hand of monkeypox virus *HindIII* E (14.4kb). This left a potential overlap of approximately 0.9kb between variola virus *HindIII* D and monkeypox virus *HindIII* C. A 2kb fragment of the left end of variola virus *HindIII* D was exonuclease III treated to leave the left most 0.5kb of variola virus sequence. This was used to probe a monkeypox virus *HindIII* digest and hybridized to monkeypox virus *HindIII* E and not at all to monkeypox virus *HindIII* C.

Monkeypox *HindIII* E was mapped for *EcoRI* and *BamHI* sites and compared to variola *HindIII* D (previously mapped by Hamilton, Greenaway and Dumbell, unpublished) (fig. 3.2a). Similar sized fragments were found in corresponding positions at both left and right ends. However, in the centre, variola virus has an *EcoRI-BamHI* fragment of 2.9kb which corresponds to a monkeypox fragment of only 2.4kb. These two fragments were compared to locate the precise position of the surplus DNA in variola virus. The left end was similar with respect to restriction enzyme sites; and the right end was shown to be homologous by heteroduplex formation followed by SI nuclease treatment (fig. 3.2b). The central region was sequenced in one direction and found to contain the surplus DNA in variola virus. An extra stretch of 391bp was found in variola virus, flanked by sequences homologous to monkeypox virus. This 391bp of DNA is part of two open reading frames (ORFs) on opposite strands. The one (complete) ORF is equivalent to vaccinia virus E ORF B and the other (partial) corresponds to part of the vaccinia virus E5R ORF (Goebel *et al.*, 1990b). Preliminary hybridization experiments suggested that the DNA sequence absent in monkeypox virus Denmark is also absent in other monkeypox virus isolates, but is present in other

Orthopoxvirus species.

This work has been described previously in the form of a master's thesis (Pare, 1988). The study has been extended and is the subject of the current chapter.

The full length of the vaccinia virus (Dairen) E5R ORF (chapter 5) and the corresponding DNA in both variola (Harvey) and monkeypox (Denmark) viruses has now been sequenced. In addition, DNA from a second strain of variola virus, Somalia, was sequenced. The DNA sequences were compared to investigate the likelihood of monkeypox virus being a remote, but direct, ancestor of variola virus.

A probe was constructed (pHarRsa), specific for the sequence absent in monkeypox virus Denmark, to detect whether this difference is present in other *Orthopoxviruses*. The results are shown in figure 3.6 on page 45.

Cloning and sequencing strategies are described in chapter 2 together with the methods used.

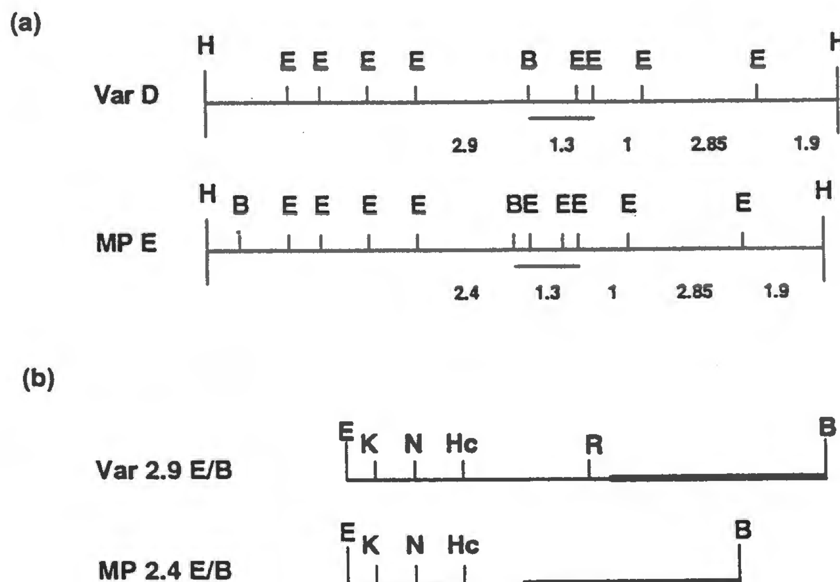


Fig. 3.2 (a) Restriction enzyme maps of variola virus *HindIII* D (Hamilton, Greenaway and Dumbell, unpublished) and monkeypox virus *HindIII* E (Douglass and Dumbell, 1992). Maps are aligned from the left to show corresponding *EcoRI* sites; numbers denote fragment sizes (in kb) to show the presence of equivalent sites at the right end. (b) Further analysis of the 2.9 and 2.4kb *EcoRI-BamHI* fragments.The bold lines denote 1.4kb stretches of DNA which were not sensitive to SI nuclease digestion of heteroduplexed material. H, E, B, K, N, Hc and R refer to restriction enzyme sites *HindIII, EcoRI, BamHI, KpnI, NcoI, HincII* and *RsaI* respectively.

RESULTS AND DISCUSSION

DNA was sequenced from the central, conserved region of the variola and monkeypox virus genomes. These sequences are located on the cross-hybridizing *HindIII* D and E fragments of variola and monkeypox viruses respectively (figure 3.1), and correspond to the E5R ORF of vaccinia virus Copenhagen (Goebel *et al.*, 1990b). Figure 3.3 shows 1260bp of variola virus Harvey DNA compared to the equivalent monkeypox virus Denmark sequence starting from a common NcoI site. The E5R-equivalent ORF occupies positions 143 to 1207. The DNA sequence shown includes the beginning of the E4L ORF (first ATG at position 136) and the short E ORF B (positions 1062-745 reading leftwards) found in vaccinia virus (Goebel et al., 1990b).

A second strain of variola virus (Somalia) was sequenced and found to contain an ORF identical to that of Harvey with the exception of a single nucleotide: the substitution of A for G at base 337 (Fig. 3.3). This change is in the wobble position and does not alter the amino acid sequence of the putative polypeptide. Harvey is a strain of variola major virus, whereas Somalia is a strain of African variola minor virus (Dumbell and Huq, 1986). Such a high degree of similarity between the two DNA sequences indicates a recent divergence of the two viruses. It also highlights the fidelity of DNA replication by the poxvirus DNA polymerase.

Aligned parts of the variola and monkeypox virus sequences have 96% base similarity allowing for four deletions of 24-, 1-, 391-, and 24-bp in the monkeypox virus sequence. However, the coding potential is different for the two viruses (Fig. 3.4). Variola virus codes for a presumptive translation product of 341 amino acids with the first ATG codon at position 185 and the stop codon at position 1208. Monkeypox virus, has the same presumptive translational initiation codon at position 185; the product would omit 8 amino acids corresponding to the 24bp deletion at position 201 to224 and a base substitution at position 263 (G to T) introduces a stop codon, thus giving a potential product of only 18 amino acid residues. The monkeypox virus sequence contains an ORF running from postions 263-1068 with an in-frame ATG at position 278, but this would presumably not be translated. A single bp deletion is at position 600 and the second 24bp deletion is at position 1155-1178.

Interestingly, the variola sequences corresponding to the 24bp deletions are flanked by short, direct repeats, of which a single copy is present in the monkeypox virus sequence (fig. 3.5). Several examples of deletions within short direct repeats have been reported between both the Copenhagen and Western Reserve (WR) strains of vaccinia virus (Smith *et al.*, 1991a) and between vaccinia virus WR and variola virus Harvey (Aguado *et al.*,1992). These repeats may be significant in the general scheme of poxvirus evolution. They more than likely are responsible for the deletions by allowing for slippage during DNA replication as described by Albertini *et al.* (1982) and Levinson and Gutman (1987).

```
Var(Harvey) CCATGGCCCATTCTATTAAGTCTTCCAAGTTGGCATCATCCACATATTGTGATAGTAATT
MP(Denmark) *********************T**************************************

61          CTCGGATATTAGTAGCGGCTACCGCCATTGATGTTTGTTCATTGGATGAGTAACTACTAA
            *********C*****T*A**************A**********T****************

121         TGTATACATTTTCCATTTATAACACGTATGTATTAACTTTGTTCATTTATATTTTTTCAT
            *************************T*****************T****************
                40.5
181         TATTATGTTGATATTAACAAAAGTGAATATATATATGTTAATAATTGTATTGTGGTTATA
            *****************T**------------------------*C**************

241         TGGCTACAATTTCATAATGAGCGGAAGTCAGTGTCCGATGATTAATGACGATAGATTTAC
            *********************TT*************T*****C***********C*****
                                                A
301         TCTGAAAAGAAAGTATCAAATCGATAGTGTAGAGTCGACAATGAAAATGGATAAGAAGAG
            *****************************C******A*******************C**T

361         GACAAAGTTTCAAAATAGAGCCAAAATGGTAAAAGAAATAAATCAGACAATAAGAGCAGC
            ************G*******T***************************G***********

421         ACAAACTCATTACGAGACATTGAAACTAGGATACATAAAATTTAAGAAAATGATTAGGAC
            *********************************T************GG************

481         TACTACTTTAGAAGATATAACAACATCTATTCCAAATATTCAGAAAATTTATAAACTATT
            *******C***********G**C***************A********C************

541         CTCGGACATTTCAGCCATTGGCAAAGTATCACAGAATCCGAGTAAAATGGCATATGCTTT
            ****T*********T***********C******************G***AT*******C-

601         ACTGCTTTACATGTTCCCCAATTTGTTTGGAGATGACCATAGATTCATTCTTTATAGAAT
            G**************T*********************************TG*********

661         GTTTCCAATGAGTAAAATCAAACACAAGATCTTCTCTCCTTTCAAACTTAATCTTATTAG
            *CA*****----------------------------------------------------

721         AATATTAGTGGAAGAAAGATTCTATAATAATGAATGCAGATCTAATAAATGGAGAATAAT
            ------------------------------------------------------------

781         TGGAACACAAGTTGATAAAATGTTGATAGTTGAATCTGATAAATATACAATAGATGCAAG
            ------------------------------------------------------------

841         GTATCGCCTAAGACCTATATATAGAATCAAGGGAAAATCTGAAGAAGATACCCTCTTCAT
            ------------------------------------------------------------

901         CAAACAGATGGTAGACCAATGTGTGACATCCCAGGAATTGGTGGAAAAAGTGTTGAAGAT
            ------------------------------------------------------------

961         ACTGTTTAGAGATTTGTTCAAGAGTGGAGAATACAAAGCGTACAGATACGATGATGACGT
            ------------------------------------------------------------

1021        AGAAAATGGATTCATTGGATTGGATAAACTAAAATTAAACATTGTTCATGATATAGTTGA
            ---------------------------------------*********************

1081        ACCATGTATGCCTGTTCGTAGGCCAGTGGCTAAGATACTGTGTAAAGAAATGGTAAATAA
            *T******************T***T***********************************

1141        ATACTTTGAGAATCCGCTGCATATTATTGGTAAGAATCTTCAAGAGTGCATTGACTTTGT
            **************------------------------**T*******************

1201        TAGTGAATAGGCATTTCATCTTTCTCCAATACTAATTCAAATTGTTAAATTAATAATGGA
            ***************C********************************************
```

Fig. 3.3 DNA sequence comparison of variola virus, Harvey (1260bp) and monkeypox virus, Denmark (820bp). The corresponding sequence for variola virus Somalia was identical to Harvey except for the A shown above the G at base 337. Asterisks denote identity of monkeypox virus with variola virus DNA and dashes indicate gaps inserted in the monkeypox virus sequence to preserve the alignment with variola virus. The first potential translational start site is indicated by an arrow and termination signal is boldly underlined. The 81-bp sequence used as a hybridization probe is underlined.

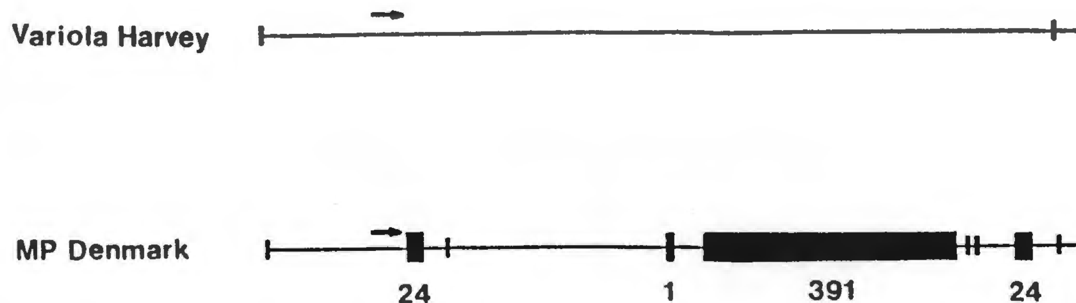


Fig. 3.4 Schematic representation of the coding potential of variola virus, Harvey and monkeypox virus, Denmark DNA in the region sequenced. Methionine start codons are indicated by arrows and termination codons by vertical lines. Deletions are denoted by closed boxes beneath which numbers refer to their size in base pairs.

VARIOLA	TGATATTAACAAAAGTGAATATATATATG**TTAATAA**TTGTATTGTGGT
MONKEYPOX	TGATA**TTAATAA**-----------------------TCGTATTGTGGT
VARIOLA	ACTTTGAGAATCCGCTGCATATTATTGGTA**AGAATC**TTCAAGAGTGCA
MONKEYPOX	ACTTTG**AGAATC**-----------------------TTTAAGAGTGCA

Fig. 3.5 DNA sequence in the region of the two 24bp deletions. Direct repeats are in bold face.

To investigate whether this degenerate monkeypox virus ORF is present in a complete form elsewhere in the genome the entire genome was probed with a radiolabelled DNA fragment (from plasmid pHarRsa) homologous to a portion of the sequence deleted in monkeypox virus (81bp underlined in Fig. 3.3). In addition, a number of other monkeypox virus isolates - 4 from captive animals, one from a case of human monkeypox in Liberia, and one from a wild squirrel in Zaire, as well as different strains of vaccinia virus (Lister, USSR, Buffalopox and rabbitpox) were probed.

The results, using stringent hybridization conditions, are shown in figure 3.6. As a control the same membrane was stripped and probed with cross hybridizing monkeypox virus DNA (plasmid pDenE/B) to show that the DNA concentration on the membrane was comparable for all samples. Clearly, the ORF is not present in its entire form in any of the isolates from captive animals or from the Liberian isolate. However, there is some hybridization of the probe to the Zaire DNA. Although the signal is not as strong as that for the vaccinia viruses, the result does suggest that there is some homology. Subsequent sequencing of the corresponding DNA in four monkeypox isolates from Zaire (see next chapter) showed that only part of the probe sequence (14bp) is present, and most of the sequence (67bp) is deleted in all of them.

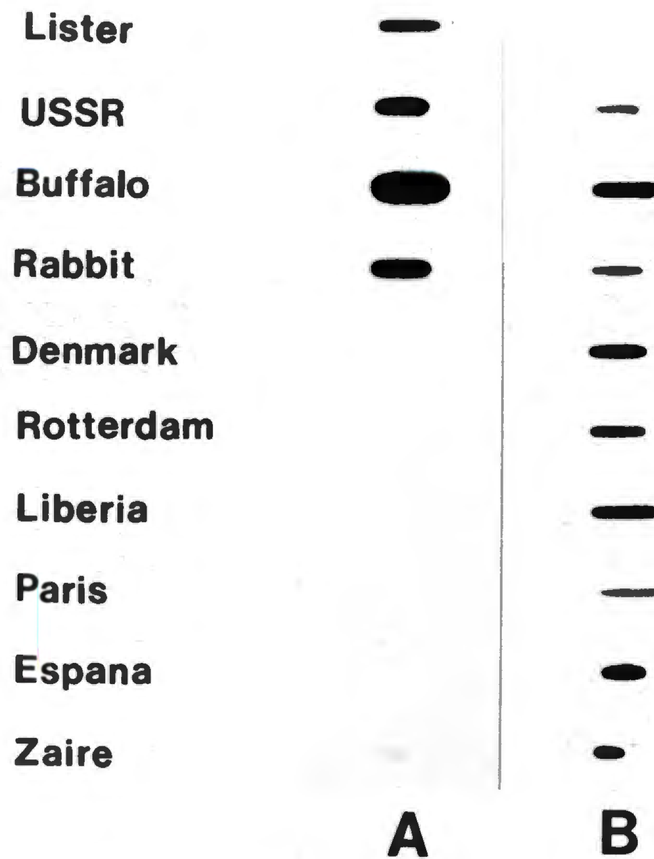


Fig. 3.6 Slot blot of monkeypox virus genomic DNA probed with A) a variola virus 81bp fragment corresponding to part of the major deletion in monkeypox virus (*RsaI* fragment from pHarRsa) and B) plasmid pDenE/B, which contains monkeypox virus DNA (2.4kb *EcoRI-BamHI* fragment referred to in fig. 3.2). Lister, USSR, Buffalopox and rabbitpox are all strains of vaccinia virus. Monkeypox viruses Denmark, Rotterdam, Espana and Paris are all isolates from captive animals; Liberia is from a human monkeypox case in Liberia and Zaire is from a squirrel in Zaire.

The DNA sequence spanning the major deletion was amplified from the Zaire isolate by PCR, using primers which bound to positions 617-632 (5'-CCCAATTTGTTTGGAG-3') and 1098-1083 (5'-CGAACAGGCATACATG-3'). The resulting amplicon was larger than the monkeypox virus Denmark fragment, but significantly shorter than the control fragment from variola virus (see next chapter). The PCR experiment, done on genomic monkeypox virus DNA, also confirmed that there was no complete copy of this ORF anywhere in the monkeypox virus genome.

There is little doubt that monkeypox and variola viruses originated from a common ancestor. All the corresponding genomic restriction enzyme fragments cross-hybridize (Mackett and Archard, 1979; Dumbell and Archard, 1980; Esposito and Knight, 1985; Esposito *et al.*, 1985) and our DNA sequence comparison, like other published reports (Esposito and Knight, 1984; Cowley and Greenaway, 1990) shows >94% base similarity between the two viruses. What is questionable is the mode of divergence of the two species from this common ancestor. Did variola virus evolve from monkeypox virus or did they evolve independently? The third alternative of monkeypox virus evolving from variola virus is highly improbable due to host range features of the viruses. The host range of variola virus is strictly limited to humans, implying that the virus had reached the end of its line in evolutionary terms. Since it is not easily maintained in another host there was little exposure of the virus to different selective pressures and therefore little likelihood of it changing to become better suited to infect different host animals. Also, historical evidence (which is not conclusive) suggests that variola virus originated in East Asia. Monkeypox virus, however, has a broad host range allowing for more variability under different selective pressures in different host systems. The virus therefore has more chance of adapting to a particular (different) host. Being confined to the tropical rain forests of West and Central Africa, monkeypox virus is unlikely to have originated anywhere other than in Africa. It is therefore more likely that monkeypox virus evolved into variola virus than vice versa. The third, and most likely alternative, is that the two viruses evolved independently.

The sequencing data shows that monkeypox virus has a degenerate form of an ORF present in variola virus. There are four separate deletions and one base substitution which have resulted in a putative gene product less than one-third the length of that in variola virus. The clearest, and most likely, scenario, is that the ORF was present in its entirety in the common ancestor and gradually accumulated mutations and became truncated in monkeypox virus. This means that, from the common ancestor, variola and monkeypox viruses evolved independently (since it is unlikely that monkeypox virus evolved from variola virus).

If variola virus were to have evolved from monkeypox virus, the ORF degenerate in monkeypox virus would have had to acquire, in addition to some base changes and three small insertions (1 X 1bp and 2 X 24bp), a major insertion of DNA which is not present elsewhere in the genome. The base substitutions and insertion of 1bp could easily be introduced through simple mutational events. The two 24bp insertions could not have arisen that simply. Although they have a repeat sequence at one end, the remaining 17- and 18-bp respectively do not resemble sequences in the immediate vicinity. It is therefore difficult to explain from where these sequences came. It appears more likely that they were originally present and later became deleted (as described earlier). Similarly, the large insertion would be more difficult to explain. Genetic elements capable of inserting themselves into DNA have been described and characterized (Lewin, 1990), but this stretch of DNA has no features of transposable elements.

Our conclusions, that variola and monkeypox viruses evolved independently from a common ancestor, are based on analysis of a single gene/pseudogene. The argument would be strengthened if similar trends were observed in other parts of the genome. Unfortunately not much monkeypox virus DNA sequencing data is available for wider comparisons to be made. The thymidine kinase gene is conserved in both monkeypox and variola viruses (Esposito and Knight, 1984) and limited sequencing data suggests that one of the host range genes present in vaccinia virus has become degenerate in both monkeypox and variola viruses (Cowley and Greenaway, 1990). Nevertheless, the location of this ORF in the conserved part of the genome and its conservation in different *Orthopoxviruses* (chapter 5) suggests that the changes present within probably represent the evolutionary pattern of the virus as a whole.

CHAPTER 4

MONKEYPOX VIRUS EVOLUTION

Summary	49
Introduction	49
Results and discussion	52

CHAPTER 4

MONKEYPOX VIRUS EVOLUTION

SUMMARY

Monkeypox viruses were differentiated into two distinct groups, West Africa and Zaire, by DNA sequence comparison and PCR amplification. A short stretch of DNA from the central conserved region of five African isolates of monkeypox virus was sequenced. This region corresponds to the vaccinia virus E5R ORF and is degenerate in monkeypox virus Denmark. Since this ORF is not conserved in monkeypox virus by functional selection it would be more variable. The monkeypox virus sequences were compared to those of variola virus. All the differences found between monkeypox virus Denmark and variola virus (previous chapter) are present in the monkeypox isolates from West Africa. The isolates from Zaire are identical to one another, but different from the West African viruses. They have three common deletions of 24-, 1- and 24-bp, but differ by having additional small deletions which abolish the ORF entirely. This indicates that monkeypox virus evolved into two groups, acquiring deletions in the genome both before and after splitting into two (West African and Zairean) lineages. The overall base similarity amongst the monkeypox virus sequences is >99%, suggesting that the divergence occurred recently.

INTRODUCTION

Monkeypox virus has been isolated from a number of different host animals and geographical locations. The isolates from outbreaks in captive animals in Europe and the USA are the only reported cases of infection outside of the tropical rain forests of West and Central Africa. In all cases, whether of human or animal origin, and irrespective of geographical origin, monkeypox virus could not be subdivided into smaller groups based on phenotypic properties.

Twenty isolates of monkeypox virus have been mapped for *HindIII* cleavage sites (Esposito and Knight, 1985). The only noticeable differences are in the terminal fragments. These differences allow monkeypox virus to be grouped according to geographical location, and not time of isolation or host animal from which the virus was isolated. The three groups formed are West Africa, Nigeria and Zaire.

A more useful enzyme for differentiating monkeypox viruses is *PstI*. The *PstI* maps differ in the conserved region as well as the termini. Again, the viruses fall into one of three groups depending on geographical origin (Richardson and Dumbell in Fenner *et. al.*, 1989).

No DNA sequence comparison has been made within the monkeypox virus species. The ORF described in chapter 3 is degenerate in monkeypox virus, and therefore would not be conserved by functional selection. For this reason it should show more variability in DNA sequence than other conserved genes, thus revealing evolutionary changes within the species.

The equivalent region was sequenced in five African isolates of monkeypox virus, one from Liberia (LiberiaII), one from Nigeria (Benin) and three from Zaire (86-21, 86-112 and 1324) (see table 4.1 and figs. 4.1 and 4.2 for virus and geographical details). A fourth Zairean strain (Z241) was partially sequenced and found to be identical to other isolates from Zaire for the 441bp sequenced. The DNA sequences have been compared to one another and to the variola and monkeypox virus sequences discussed in chapter 3.

A major difference was the size of the large deletion described in chapter 3. PCR amplification of the DNA spanning this deletion from a number of other monkeypox virus isolates provided further evidence of the occurrence of two distinct groups.

TABLE 4-1

Details of monkeypox virus isolates

Designation	year isolated	origin
Denmark	1958	captive monkey from Singapore
Paris	1968	captive chimpanzee from Sierra Leone
Prier	1959	captive monkey from Malaysia
Espana	1967	captive monkeys from India, Malaysia
LiberiaI	1970	human, Liberia
LiberiaII	1970	human, Liberia
MP-266	1970	human, Sierra Leone
MP-82	1971	human, Nigeria
Benin	1978	human, Nigeria
Z241	1972	human, Katako-kombe, Kasai, Zaire
Z1324	1979	human, Gemena, Equateur, Zaire
86-21	1986	human, Bumba, Equateur, Zaire
85-249	1985	squirrel, Bumba, Equateur, Zaire
85-240	1985	human, Bumba, Equateur, Zaire
86-112	1986	human, Bongadanga, Equateur, Zaire

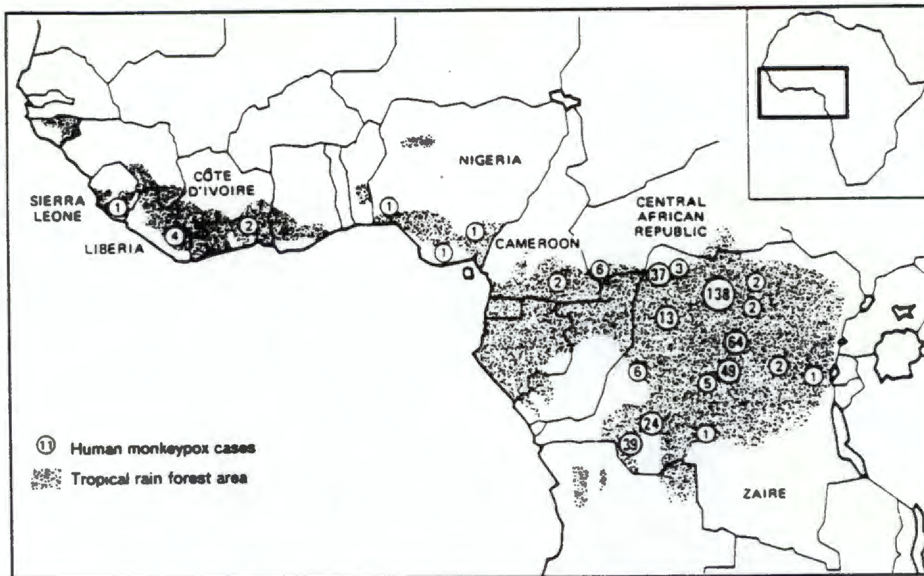


Fig. 4.1 Map of Western and central Africa, showing the extent of tropical rain forest and locations where cases of human monkeypox have occurred. Numbers refer to cases reported between 1970 and 1986. (from Jezek and Fenner, 1988.)

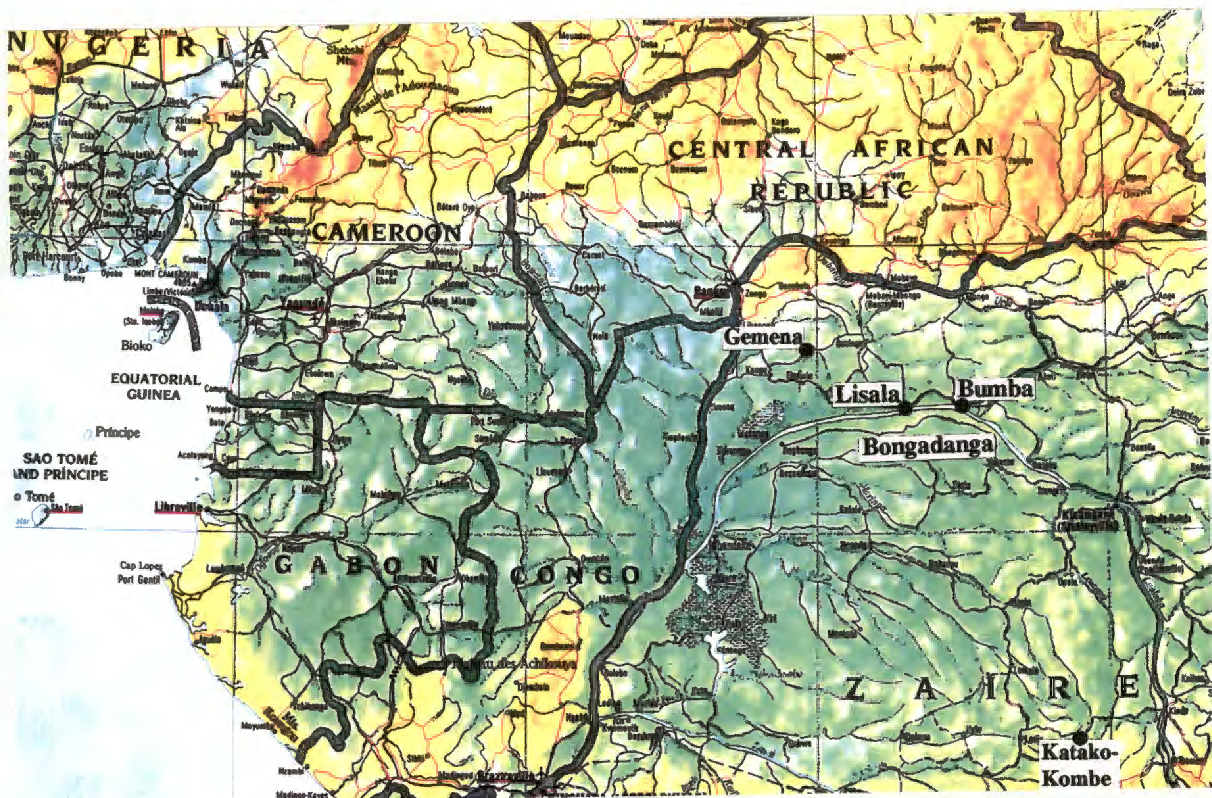


Fig. 4.2 Map of Africa showing the topology between West and Central Africa; and the places in Zaire from where monkeypox virus was isolated.

RESULTS AND DISCUSSION

Figure 4.3 summarizes the results obtained as a DNA alignment. The equivalent DNA sequences of five African isolates of monkeypox virus are compared to the corresponding variola virus Harvey and monkeypox virus Denmark sequences. This stretch of DNA corresponds to the published vaccinia virus E5R open reading frame (ORF) (Goebel *et al.*, 1990b).

The DNA sequences are aligned starting at a common *NcoI* site and ending at the variola virus termination codon. The potential E5R-equivalent coding sequence starts at position 185 and ends at position 1214. All the monkeypox viruses have a degenerate form of this gene (see figure 4.4).

Amongst the monkeypox viruses differences are found between isolates from West Africa (LiberiaII and Benin) and those from Zaire (86-21, 86-112 and Z1324). Within these two groups the sequences are identical (Zaire) or >99% similar (West Africa). An additional strain of monkeypox virus, Z241, was partially sequenced (441bp) and found to be identical to the other Zaire isolates. The Denmark strain of monkeypox virus is identical to LiberiaII.

At position 263 a transversion base substitution (G-T) generates a stop codon in all monkeypox virus sequences. The West African isolates have two open reading frames, the one running from positions 143 to 262 and the other from positions 266 to 1072. The Zaire isolates have the first ORF (positions 143 to 262), but the second ORF is reduced to 75bp by a single base pair deletion at position 308; although they have >94% DNA homology to variola virus and >99% similarity to the West African monkeypox virus isolates.

Present in all monkeypox virus sequences are the two 24bp deletions (positions 201-224 and 1155-1178) and one single base deletion (position 603). These deletions must have arisen before monkeypox virus diverged into separate groups.

All the West African isolates have the large deletion of 391bp which is present in the Denmark strain. In the corresponding position in the Zaire isolates there are three deletions of 5- (positions 697-701), 9- (positions 815-823) and 127-bp (positions 867-993) as well as an insertion of 1bp at position 1003.

Common to the Zaire isolates only are additional deletions of 1- and 2-bp and an insertion of 3bp (positions 308, 432-433 and 512-514 respectively). These changes must have occurred after the divergence of monkeypox virus into two areas of endemicity.

Fig. 4.3 (On page 53.) Alignment of DNA sequences from different monkeypox virus isolates with variola virus, Harvey. VARHAR = variola virus, Harvey; MPDEN = monkeypox virus, Denmark; MPLIB = monkeypox virus, Liberia; MPBEN = monkeypox virus, Benin; MP86-21 = monkeypox virus, Zaire 86-21; MP86-112 = monkeypox virus, Zaire 86-112; MP1324 = monkeypox virus, Zaire 1324. Asterisks above the alignment denote differences between variola and monkeypox viruses, where all the monkeypox viruses are the same. Asterisks below the alignment denote differences within the monkeypox virus species. Dashes denote gaps inserted to maintain alignment of the sequences. Potential translational start sites are indicated with an arrow and in frame termination sites are blocked. DNA sequences corresponding to primers used for PCR amplification are overlined in the variola sequence.

```
VARHAR    CCATGGCCCATTCTATTAAGTCTTCCAAGTTGGCATCATCCACATATTGTGATAGTAATTCTCGGATATTAGTAGCGGCTACCGCCATTGATGTTTGTTCATTGGATGAGTAACTACTAA  120
MPDEN     CCATGGCCCATTCTATTAAGTTTTCCAAGTTGGCATCATCCACATATTGTGATAGTAATTCTCGGATATCAGTAGTGACTACCGCCATTGATATTTGTTCATTTGATGAGTAACTACTAA
MPLIB     CCATGGCCCATTCTATTAAGTTTTCCAAGTTGGCATCATCCACATATTGTGATAGTAATTCTCGGATATCAGTAGTGACTACCGCCATTGATATTTGTTCATTTGATGAGTAACTACTAA
MPBEN     CCATGGCCCATTCTATTAAGTTTTCCAAGTTGGCATCATCCACATATTGTGATAGTAATTCTCGGATATCAGTAGTGACTACAGCCATTGATATTTGTTCATTTGATGAGTAACTACTAA
MP86-21   CCATGGCCCATTCTATTAAGTTTTCCAAGTTGGCATCATCCACATATTGTGATAGTAATTCTCGGATATCAGTAGTGACTACCGCCATTGATATTTGTTCATTTGATGAGTAACTACTAA
MP86-112  CCATGGCCCATTCTATTAAGTTTTCCAAGTTGGCATCATCCACATATTGTGATAGTAATTCTCGGATATCAGTAGTGACTACCGCCATTGATATTTGTTCATTTGATGAGTAACTACTAA
MP1324    CCATGGCCCATTCTATTAAGTTTTCCAAGTTGGCATCATCCACATATTGTGATAGTAATTCTCGGATATCAGTAGTGACTACCGCCATTGATATTTGTTCATTTGATGAGTAACTACTAA

VARHAR    TGTATACATTTTCCATTTATAACACGTATGTATTAACTTTGTTCATTTATATTTTTTCATTATATGTTGATATTAACAAAAGTGAATATATATATGTTAATAATTGTATTGTGGTTATA  240
MPDEN     TGTATACATTTTCCATTTATAACACTTATGTATTAACTTTGTTTATTTATATTTTTTCATTATATGTTGATATTAATAA------------------------TCGTATTGTGGTTATA
MPLIB     TGTATACATTTTCCATTTATAACACTTATGTATTAACTTTGTTTATTTATATTTTTTCATTATATGTTGATATTAATAA------------------------TCGTATTGTGGTTATA
MPBEN     TGTATACATTTTCCATTTATAACACTTATGTATTAACTTTGTTTATTTATATTTTTTCATTATATGTTGATATTAATAA------------------------TCGTATTGTGGTTATA
MP86-21   TGTATACATTTTCCATTTATAACACTTATGTATTAACTTTGTTTATTTATATTTTTTCATTATATGTTGATATTAATAA------------------------TCGTATTGTGGTTATA
MP86-112  TGTATACATTTTCCATTTATAACACTTATGTATTAACTTTGTTTATTTATATTTTTTCATTATATGTTGATATTAATAA------------------------TCGTATTGTGGTTATA
MP1324    TGTATACATTTTCCATTTATAACACTTATGTATTAACTTTGTTTATTTATATTTTTTCATTATATGTTGATATTAATAA------------------------TCGTATTGTGGTTATA

VARHAR    TGGCTACAATTTCATAATGAGCGGAAGTCAGTGTCCGATGATTAATGACGATAGATTTACTCTGAAAAGAAAGTATCAAATCGATAGTGTAGAGTCGACAATGAAAATGGATAAGAAGAG  360
MPDEN     TGGCTACAATTTCATAATGAGTTGAAGTCAGTGTCCTATGATCAATGACGATAGCTTTACTCTGAAAAGAAAGTATCAAATCGATAGTGCAGAGTCAACAATGAAAATGGATAAGACGAT
MPLIB     TGGCTACAATTTCATAATGAGTTGAAGTCAGTGTCCTATGATCAATGACGATAGCTTTACTCTGAAAAGAAAGTATCAAATCGATAGTGCAGAGTCAACAATGAAAATGGATAAGACGAT
MPBEN     TGGCTACAATTTCATAATGAGTTGAAGTCAGTGTCCTATGATCAATGACGATAGCTTTACTCTGAAAAGAAAGTATCAAATCGATAGTGCAGAGTCAACAATGAAAATGGATAAGACGAT
MP86-21   TGGCTACAATTTCATAATGAGTTGAAGTCAGTGTCCTATGATCAATGACGATAGCTTTACTCTGAAA-GAAAGTATCAAATCGATAGTGCAGAGTCAACAATGAAAATGGATAAGACGAT
MP86-112  TGGCTACAATTTCATAATGAGTTGAAGTCAGTGTCCTATGATCAATGACGATAGCTTTACTCTGAAA-GAAAGTATCAAATCGATAGTGCAGAGTCAACAATGAAAATGGATAAGACGAT
MP1324    TGGCTACAATTTCATAATGAGTTGAAGTCAGTGTCCTATGATCAATGACGATAGCTTTACTCTGAAA-GAAAGTATCAAATCGATAGTGCAGAGTCAACAATGAAAATGGATAAGACGAT

VARHAR    GACAAAGTTTCAAAATAGAGCCAAAATGGTAAAAGAAATAAATCAGACAATAAGAGCAGCACAAACTCATTACGAGACATTGAAACTAGGATACATAAAATTTAAGAAAATGATTAGGAC  480
MPDEN     GACAAAGTTTCAGAATAGAGTCAAAATGGTAAAAGAAATAAATCAGACGATAAGAGCAGCACAAACTCATTACGAGACATTGAAACTAGGATATATAAAATTTAAGGGAATGATTAGGAC
MPLIB     GACAAAGTTTCAGAATAGAGTCAAAATGGTAAAAGAAATAAATCAGACGATAAGAGCAGCACAAACTCATTACGAGACATTGAAACTAGGATATATAAAATTTAAGGGAATGATTAGGAC
MPBEN     GACAAAGTTTCAGAATAGAATCAAAATGGTAAAAGAAATAAATCAGACGATAAGAGCAGCACAAACTCATTACGAGACATTGAAACTAGGATATATAAAATTTAAGGGAATGATTAGGAC
MP86-21   GACAAAGTTTCAGAATAGAGTCAAAATGGTAAAAGAAATAAATCAGACGATAAGAGCAGCACAAACTCATT--GAGACATTGAAACTAGGATATATAAAATTTAAGTGAATGATTAGGAC
MP86-112  GACAAAGTTTCAGAATAGAGTCAAAATGGTAAAAGAAATAAATCAGACGATAAGAGCAGCACAAACTCATT--GAGACATTGAAACTAGGATATATAAAATTTAAGTGAATGATTAGGAC
MP1324    GACAAAGTTTCAGAATAGAGTCAAAATGGTAAAAGAAATAAATCAGACGATAAGAGCAGCACAAACTCATT--GAGACATTGAAACTAGGATATATAAAATTTAAGTGAATGATTAGGAC

VARHAR    TACTACTTTAGAAGATATAACAACATCTATT---CCAAATATTCAGAAAATTTATAAACTATTCTCGGACATTTCAGCCATTGGCAAAGTATCACAGAATCCGAGTAAAATGGCATATGC  600
MPDEN     TACTACTCTAGAAGATATAGCACCATCTATT---CCAAATAATCAGAAAACTTATAAACTATTCTCGTACATTTCAGTCATTGGCAAAGCATCACAGAATCCGAGTAAGATGATATATGC
MPLIB     TACTACTCTAGAAGATATAGCACCATCTATT---CCAAATAATCAGAAAACTTATAAACTATTCTCGTACATTTCAGTCATTGGCAAAGCATCACAGAATCCGAGTAAGATGATATATGC
MPBEN     TACTACTCTAGAAGATATAGCACCATCTATT---CCAAATAATCAGAAAACTTATAAACTATTCTCGGACATTTCAGTCATTGGCAAAGCATCACAGAATCCGAGTAAGATGATATATGC
MP86-21   TACTACTCTAGAAGATATAGCACCATCTATTATTCCAAATAATCAGAAAACTTATAAACTATTCTCGGACATTTCAGTCATTGGCAAAGCATCACAGAATCCGAGTAAGATGATATATGC
MP86-112  TACTACTCTAGAAGATATAGCACCATCTATTATTCCAAATAATCAGAAAACTTATAAACTATTCTCGGACATTTCAGTCATTGGCAAAGCATCACAGAATCCGAGTAAGATGATATATGC
MP1324    TACTACTCTAGAAGATATAGCACCATCTATTATTCCAAATAATCAGAAAACTTATAAACTATTCTCGGACATTTCAGTCATTGGCAAAGCATCACAGAATCCGAGTAAGATGATATATGC

VARHAR    TTTACTGCTTTACATGTTCCCCAATTTGTTTGGAGATGACCATAGATTCATTCTTTATAGAATGTTTCCAATGAGTAAAATCAAACACAAGATCTTCTCTCCTTTCAAACTTAATCTTAT  720
MPDEN     TC-GCTGCTTTACATGTTTCCCAATTTGTTTGGAGATGACCATAGATTCATTTGTTATAGAATGCATCCAA-------------------------------------------------
MPLIB     TC-GCTGCTTTACATGTTTCCCAATTTGTTTGGAGATGACCATAGATTCATTTGTTATAGAATGCATCCAA-------------------------------------------------
MPBEN     TC-GCTGCTTTACATGTTTCCCAATTTGTTTGGAGATGACCATAGATTCATTTGTTATAGAATGCATCCAA-------------------------------------------------
MP86-21   TC-GCTGCTTTACATGTTTCCCAATTTGTTTGGAGATGACCATAGATTCATTTGTTATAGAATGCATCCAATGAGTAAAATCAAACACAAAATCTT-----CGTTCAAACTTAATCTTAT
MP86-112  TC-GCTGCTTTACATGTTTCCCAATTTGTTTGGAGATGACCATAGATTCATTTGTTATAGAATGCATCCAATGAGTAAAATCAAACACAAAATCTT-----CGTTCAAACTTAATCTTAT
MP1324    TC-GCTGCTTTACATGTTTCCCAATTTGTTTGGAGATGACCATAGATTCATTTGTTATAGAATGCATCCAATGAGTAAAATCAAACACAAAATCTT-----CGTTCAAACTTAATCTTAT

VARHAR    TAGAATATTAGTGGAAGAAAGATTCTATAATAATGAATGCAGATCTAATAAATGGAGAATAATTGGAACACAAGTTGATAAAATGTTGATAGTTGAATCTGATAAATATACAATAGATGC  840
MPDEN     ------------------------------------------------------------------------------------------------------------------------
MPLIB     ------------------------------------------------------------------------------------------------------------------------
MPBEN     ------------------------------------------------------------------------------------------------------------------------
MP86-21   TAGAATATTAGTGGAAGAAAGATTCTATAATAATGAATGCAGAGATTATAAATGGAGAATAATTGGAACACAAGTTGATAAAATATTGATAGC---------TAAATATACAATAGATGC
MP86-112  TAGAATATTAGTGGAAGAAAGATTCTATAATAATGAATGCAGAGATTATAAATGGAGAATAATTGGAACACAAGTTGATAAAATATTGATAGC---------TAAATATACAATAGATGC
MP1324    TAGAATATTAGTGGAAGAAAGATTCTATAATAATGAATGCAGAGATTATAAATGGAGAATAATTGGAACACAAGTTGATAAAATATTGATAGC---------TAAATATACAATAGATGC

VARHAR    AAGGTATCGCCTAAGACCTATATATAGAATCAAGGGAAAATCTGAAGAAGATACCCTCTTCATCAAACAGATGGTAGACCAATGTGTGACATCCCAGGAATTGGTGGAAAAAGTGTTGAA  960
MPDEN     ------------------------------------------------------------------------------------------------------------------------
MPLIB     ------------------------------------------------------------------------------------------------------------------------
MPBEN     ------------------------------------------------------------------------------------------------------------------------
MP86-21   AATGTATCGCATAAGACCGATATATA----------------------------------------------------------------------------------------------
MP86-112  AATGTATCGCATAAGACCGATATATA----------------------------------------------------------------------------------------------
MP1324    AATGTATCGCATAAGACCGATATATA----------------------------------------------------------------------------------------------

VARHAR    GATACTGTTTAGAGATTTGTTCAAGAGTGGAGAATACAAAGC-GTACAGATACGATGATGACGTAGAAAATGGATTCATTGGATTGGATAAACTAAAATTAAACATTGTTCATGATATAG  1080
MPDEN     ---------------------------------------------------------------------------------------------------------CATTGTTCATGATATAG
MPLIB     ---------------------------------------------------------------------------------------------------------CATTGTTCATGATATAG
MPBEN     ---------------------------------------------------------------------------------------------------------CATTGTTCATGATATAG
MP86-21   ---------------------------------ATACAAAGCAGTACAGATACAATGATGATGTAGAAAATGGATTCATTGGATTGGATAAACTAAAATTAAACATTGTTCATGATATAG
MP86-112  ---------------------------------ATACAAAGCAGTACAGATACAATGATGATGTAGAAAATGGATTCATTGGATTGGATAAACTAAAATTAAACATTGTTCATGATATAG
MP1324    ---------------------------------ATACAAAGCAGTACAGATACAATGATGATGTAGAAAATGGATTCATTGGATTGGATAAACTAAAATTAAACATTGTTCATGATATAG

VARHAR    TTGAACCATGTATGCCTGTTCGTAGGCCAGTGGCTAAGATACTGTGTAAAGAAATGGTAAATAAATACTTTGAGAATCCGCTGCATATTATTGGTAAGAATCTTCAAGAGTGCATTGACT  1200
MPDEN     TTGAATCATGTATGCCTGTTCGTATGCCTGTGGCTAAGATACTGTGTAAAGAAATGGTAAATAAATACTTTGAGAATC------------------------TTTAAGAGTGCATTGACT
MPLIB     TTGAATCATGTATGCCTGTTCGTATGCCTGTGGCTAAGATACTGTGTAAAGAAATGGTAAATAAATACTTTGAGAATC------------------------TTTAAGAGTGCATTGACT
MPBEN     TTGAATCATGTATGCCTGTTCGTATGCCTGTGGCTAAGATACTGTGTAAAGAAATGGTAAATAAATACTTTGAGAATC------------------------TTTAAGAGTGCATTGACT
MP86-21   TTGAATCATGTATACCTGTTCGTATGCCTGTGGCTAAGATACTGTGTAAAGAAATGGTAAATAAATACTTTGAGAATC------------------------TTTAAGAGTGTATTGACT
MP86-112  TTGAATCATGTATACCTGTTCGTATGCCTGTGGCTAAGATACTGTGTAAAGAAATGGTAAATAAATACTTTGAGAATC------------------------TTTAAGAGTGTATTGACT
MP1324    TTGAATCATGTATACCTGTTCGTATGCCTGTGGCTAAGATACTGTGTAAAGAAATGGTAAATAAATACTTTGAGAATC------------------------TTTAAGAGTGTATTGACT

VARHAR    TTGTTAGTGAATAG  1214
MPDEN     TTGTTAGTGAATAG
MPLIB     TTGTTAGTGAATAG
MPBEN     TTGTTAGTGAATAG
MP86-21   TTGTTAGTGAATAT
MP86-112  TTGTTAGTGAATAT
MP1324    TTGTTAGTGAATAT
```

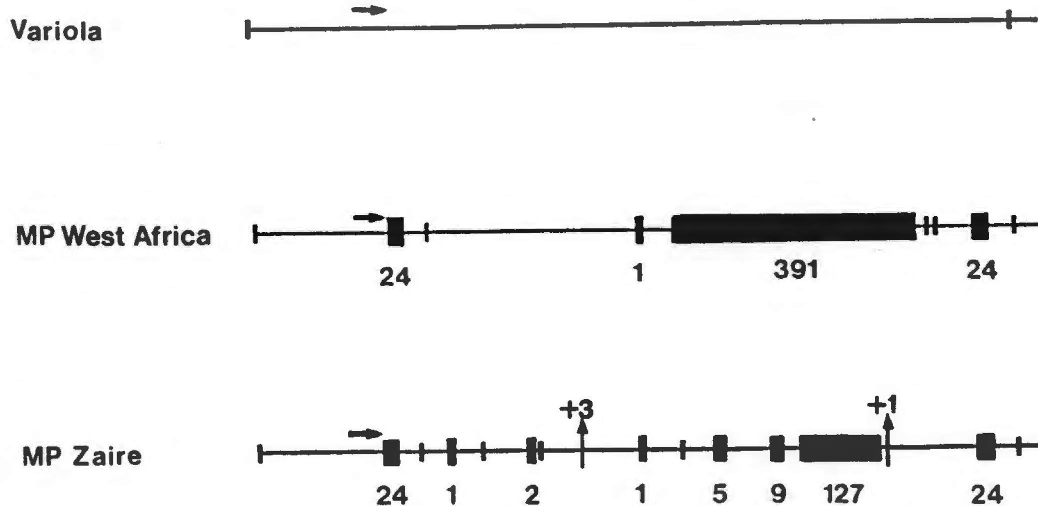


Fig. 4.4 Diagram to show the difference in corresponding DNA sequences of variola virus, monkeypox virus from West Africa and monkeypox virus from Zaire. Potential ORFs are denoted by arrows; termination codons by vertical lines; and deletions by filled boxes below which numbers refer to their sizes (in bp).

The large deletion was investigated by PCR in an additional 3 Zaire and 6 West African monkeypox virus isolates (see figure 4.5). Primers corresponding to positions 620-636 (5'-CCCAATTTGTTTGGAG-3') and 1102-1087 (5'-CGAACAGGCATACATG-3') were used to amplify DNA spanning the deletion. Fragments of either 342bp (Zaire isolates) or 91bp (West African isolates) were produced. All viruses from Zaire gave a band of 342bp, whereas those from countries further west gave a fragment size of 91bp. Each of the isolates from captive animals resembled the West African monkeypox viruses. This is additional evidence that the source of virus for the outbreaks in Europe and the United States was from West Africa. By comparison, other *Orthopoxviruses* - vaccinia viruses Dairen, rabbitpox and buffalopox, and camelpox viruses 79-241 and 2740 - gave a fragment size of 482bp (fig. 4.5).

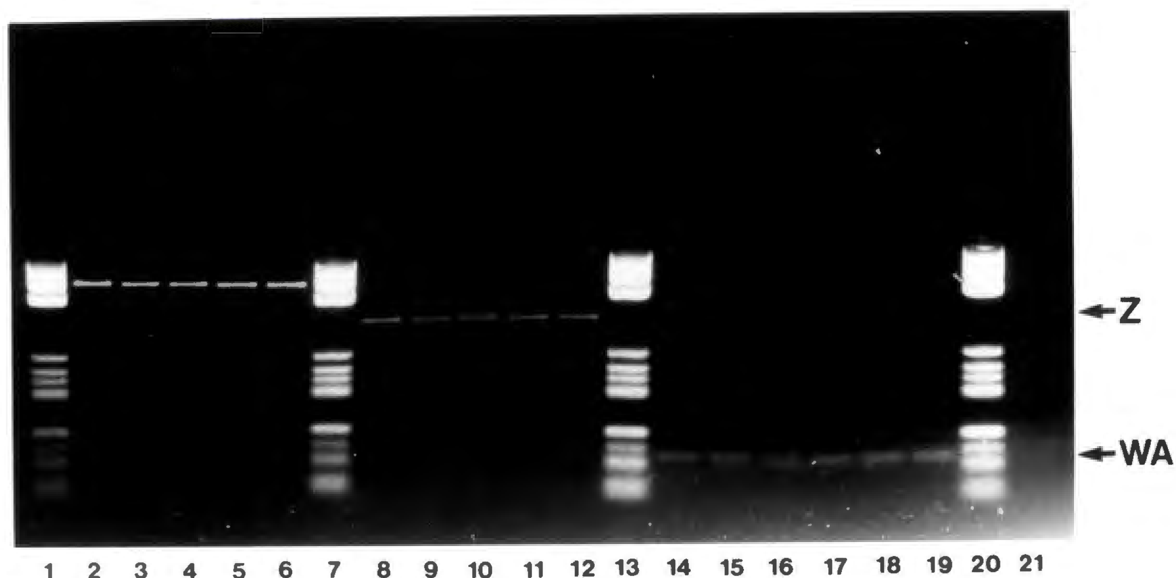


Figure 4.5 Agarose gel electrophoresis of PCR-generated DNA fragments corresponding to the region of the major deletion in monkeypox virus. WA refers to West Africa and Z refers to Zaire. PCR was performed on genomic *Orthopoxvirus* DNA as described in chapter 2, using primers corresponding to the positions underlined in figure 4.3. Lanes 2-4 = vaccinia viruses Dairen, rabbitpox and buffalopox respectively; lanes 5-6 = camelpox viruses 79-241 and 2740 respectively; lanes 8-12 = Zaire isolates of monkeypox virus, 85.249, 86.112, 85.240, 1024 and 1324 respectively; lanes 14-16 = West African isolates of monkeypox virus, Liberial, MP-82 and MP-266 respectively; lanes 17-19 = monkeypox virus isolates from captive animals, Prier, Espana and Paris respectively; lane 16 = negative control, no DNA; and lanes 1, 7, 13 and 20 are DNA markers (Boehringer marker VI, see Appendix B).

In evolutionary terms, one can speculate that the common ancestral *Orthopoxvirus* carried the gene equivalent to E5R in vaccinia virus in a complete functional form. After branching this gene was of no advantage to monkeypox virus and mutations accumulated. The function of vaccinia virus E5R is unknown. Since the gene is degenerate in monkeypox virus it probably is nonessential for vaccinia virus (this hypothesis has been investigated and discussed in chapter 6). Possible functions for nonessential genes may include host range or pathogenicity functions. Host range genes have been found clustered near the left terminus of the *Orthopoxvirus* genome within the variable region (Moyer and Rothe, 1980; Drillien *et. al.*, 1981; Gillard *et. al.*, 1986; Spehner *et al.*, 1988; Perkus *et. al.*, 1990; Chen *et. al.*, 1992). The location of E5R near the centre suggests that it is not a host range gene. However, an association of the gene with the host animal infected could explain why monkeypox virus suddenly no longer required this gene - it may have entered a different animal species. A more probable role of the E5R gene product is one of virus-host interaction where the viral product interacts with a host-specific component. Interactions may be with the host cell itself, the biochemical pathways of the cell, or the immune system. The entry of monkeypox virus into a different host may have rendered the viral gene product obsolete. In the absence of positive selection mutations would accumulate. Here the mutations are mainly in the form of deletions.

Two stages can be recognised in the order of occurrence of the deletions. Firstly, before monkeypox virus diverged into separate groups, the two 24bp deletions, 1bp deletion and base substitution causing termination near the beginning of the gene, were made. The order in which these occurred cannot be deduced.

Secondly, after monkeypox virus diverged into two separate groups (Zaire and West Africa), two deletions of 1- and 2-bp, and an insertion of 3bp were formed in the Zaire group.

The separate evolution of monkeypox virus was probably a result of deforestation. The tropical forests of West and Central Africa were at one stage continuous, containing abundant wild life. Gradually the forests have receded due to drought conditions and some parts have become isolated (see fig. 4.1). The animals inhabiting the forests (and hosting monkeypox virus) have diminished and become separated, so resulting in separation of the virus isolates.

Another feature which may be relevant to the geographical separation of the viruses is the presence of a mountain range between Nigeria and Cameroon, possibly separating the animals of West Africa from those in Central Africa (see fig. 4.2).

The large deletion may have arisen in one of two ways. Firstly, it may have been generated only after the divergence of monkeypox virus, in which case the West African group may have acquired the deletion in a single mutational event. The Zaire group, however, must have undergone at least four mutational events: three deletions of 5-, 9-, and 127-bp and one insertion of 1bp. Alternatively, one or more of the mutations in the Zaire group may have taken place before divergence of the two groups. In this case more than one mutational event would have occurred in the West African isolates.

In comparing the monkeypox and variola virus sequences, we assume that the variola sequence represents the ancestral gene. Besides the two 24-bp deletions, none of the other deletions are flanked by perfect direct repeats. The deletions in the Zaire isolates have pentameric repeats in which 4/5 bases are identical. The large deletion of the West African isolates has a direct repeat of 11/12 identical bases, with both repeats within the deleted sequence i.e. not perfectly positioned. In the true ancestor direct repeats may have been present, which were responsible for the deletions, but we cannot be sure of this, since base substitution events appear to occur infrequently.

Comparing the number of base substitutions, the divergence of monkeypox virus appears to be relatively recent. The Zaire isolates all have identical DNA sequences. An alignment of the sequences shows only 7/767 positions of nucleotide change amongst the six isolates (shown by asterisks below the alignment in fig. 4.3). In four positions the bases are identical within the two groups of viruses but differ between the Zairean and West African groups. In two of the other three positions the Zaire isolates differ from Benin and are identical to LiberiaII; and in the third position Zaire differs from LiberiaII, but is identical to Benin. Genomic maps show the West African group to be further divided into two groups - Nigeria

and West Africa. Our limited sequencing data does not separate the two groups, but the few base changes observed are consistent with the Zaire isolates being more closely related to one another (identical sequences), and the Liberia and Benin (Nigerian isolate) isolates being not as closely related to one another (3 base changes).

In comparison to the small number of base substitutions the virus appears to accumulate insertions and deletions at a remarkably high frequency. However, deletions and insertions are likely to be more disruptive of function than single base alterations, and therefore more likely to produce nonviable genomes. For this reason they probably accumulate more frequently in nonessential genes. A number of *Orthopoxvirus* genes have been disrupted by insertions or deletions (Weir and Moss, 1983 - thymidine kinase; Boursnell *et. al.*, 1988 - 37k envelope antigen-related; Brown *et. al.*, 1991 - haemagglutinin; Goebel *et. al.*, 1990a - A-type inclusion and serine protease inhibitor/haemorrage-inducing proteins of cowpox virus; Smith *et. al.*, 1991a and Aguado *et al.*, 1992 - several; and Chen *et. al.*, 1992 - CHO host range). The deletions are often flanked by short (5- or 6-bp) direct repeats (Smith *et. al.*, 1991a; Chen *et. al.*, 1992), but this is not always the case (Smith *et. al.*, 1991a). Most of these genes have been shown to be nonessential.

The occurrence of deletions and insertions has been the basis on which we have divided monkeypox virus into two groups, viz. West Africa and Zaire. The presence of small deletions and insertions in other *Orthopoxviruses* may be useful in the differentiation of viral strains, especially as the DNA sequences are so highly conserved.

CHAPTER 5

INTERRELATIONSHIPS WITHIN THE ORTHOPOXVIRUS *GENUS*

Summary 59

Introduction 60

Results and discussion 61

DNA sequence of vaccinia virus Dairen E5R and comparison with the known Copenhagen and WR sequences 61

Comparison of the equivalent region in different *Orthopoxviruses* 66

The E4L-E5R promoter region 70

Pairwise DNA comparison 71

Phylogenetic analysis of the *Orthopoxvirus* DNA sequences 73

Comparison of an intergenic DNA sequence in different *Orthopoxviruses* 77

CHAPTER 5

INTERRELATIONSHIPS WITHIN THE *ORTHOPOXVIRUS* GENUS

SUMMARY

The DNA sequence was determined for approximately 1200bp corresponding to the vaccinia virus E5R-equivalent region in six additional *Orthopoxviruses:* vaccinia virus (Dairen strain), camelpox virus (strain 903), taterapox virus (gerbilpox), ectromelia virus and two strains of cowpox virus (Larkin and Brighton). DNA comparisons of these sequences with the two variola viruses and 6 monkeypox viruses described in the previous chapters indicated that viruses of the same species have >98% base similarity and, between species, the most diverged sequences have >91% base similarity. Phylogenetic analysis of the DNA sequences showed that the 16 viruses could be divided into 7 groups: 1) three vaccinia viruses, 2) two variola viruses, 3) camelpox virus + gerbilpox virus, 4) two cowpox viruses, 5) ectromelia virus, 6) three monkeypox viruses from West Africa, and 7) three monkeypox viruses from Zaire. The camelpox and gerbilpox viruses were located on the same branch as the variola viruses. These are closely related to the vaccinia viruses. Each species had gerbilpox virus as its nearest neighbour.

A comparison was made of the intergenic DNA sequence corresponding to the region between vaccinia virus E5R and E6R in all the above-mentioned viruses as well as raccoon poxvirus. This region should show little preferential conservation of sequence due to functional constraints. Specific deletion profiles allowed for differentiation of viruses both between and within species. The monkeypox viruses were more closely related to one another as well as variola and vaccinia viruses in this region as compared to the E5R-equivalent region. The raccoon poxvirus sequence differs substantially from the others in terms of base substitutions as well as insertions and deletions. It also has a heptanucleotide sequence repeated 12 times. A repeated pentameric sequence was detected in camelpox, gerbilpox and ectromelia viruses in 2, 10 and 17 copies respectively. This sequence resembles the flanking region of PR1, an intermediate repetitive sequence found in mice. The two cowpox viruses have a more complex arrangement of repeated sequences differing between the two viral strains. These repeated sequences may represent selfish (or junk) DNA with no specific function.

INTRODUCTION

In addition to monkeypox virus, a number of other *Orthopoxviruses* closely related to variola virus are present in natural animal reservoirs. These include taterapox virus (isolated from a gerbil in Benin), camelpox virus (endemic in dromedary camels of Africa and Asia) and cowpox virus (isolated from a number of different animals, but believed to be naturally maintained in small rodents). Vaccinia virus has currently no known natural reservoir. Ectromelia virus has not been found in the wild, but commonly infects laboratory mice and resembles variola virus both in phenotype and genotype. Much has been reported on the phenotypic character of the different viruses as well as their genomic restriction enzyme maps, but their phylogenetic relationship is yet to be established.

The DNA sequencing studies of chapters 1 and 2 have been extended to other *Orthopoxviruses*. The E5R ORF of vaccinia virus, Dairen (see fig. 5.1 for position in genome) was cloned into pUC19 on an *EcoRI-BamHI* fragment (chapter 2) and sequenced. DNA and putative polypeptide sequence comparisons were made with the reported sequences of vaccinia virus Copenhagen (Goebel *et al.*, 1990b) and WR (GenBank accession no. M35027).

The equivalent regions of two strains of cowpox (Larkin and Brighton), ectromelia (Mill Hill), camelpox (903) and taterapox (gerbil) viruses were PCR amplified, cloned into pUC18 or pUC19 and sequenced (chapter 2). The ORFs were compared to vaccinia virus E5R and the overall base similarity was compared amongst all the viruses.

The DNA sequences of vaccinia (3 strains), variola (2 strains), monkeypox (6 strains), camelpox (1 strain), taterapox (1 strain), ectromelia (1 strain) and cowpox (2 strains) viruses were analysed phylogenetically by the neighbour joining method of Saitou and Nei (1987) using both clustal V (Higgins and Sharp, 1988; Higgins *et al.*, 1991) and TREECON (Van de Peer and Wachter, 1992) computer packages.

A second DNA comparison was made of the sequence corresponding to the intergenic region between vaccinia virus E5R and E6R. E6R is probably expressed early in the virus life cycle since it does not have a typical late gene motif (TAAAT) (Hangii *et al.*, 1986) or intermediate gene consensus sequence (Baldick *et al.*, 1992). The functional constraints in the intergenic region, viz. to maintain an A-T-rich early promoter region, are less rigid than those found in coding regions. One would therefore expect more variability within this region with changes being of a more random distribution. The differences found will be discussed.

Raccoon poxvirus, which originated in Maryland, America, was included in the intergenic comparison. This *Orthopoxvirus* is more different from the group of viruses from Africa and Eurasia, both in restriction enzyme map (Esposito and Knight, 1985), cross-hybridization experiments, (Cavallaro and Esposito, 1992) and in DNA sequence similarity (69%) to vaccinia virus for the hemagglutinin gene (Cavallaro and Esposito, 1992).

RESULTS AND DISCUSSION

DNA sequence of vaccinia virus Dairen E5R and comparison with the known Copenhagen and WR sequences

The position of the E5R gene on the genome of vaccinia virus is shown in fig. 5.1. This gene was cloned into pUC19 by eluting the *HindIII* E fragment from an agarose gel, digesting it with EcoRI and shotgun cloning the mixture of fragments into pUC19 digested with *EcoRI.* A recombinant plasmid containing the 3.2kb *EcoRI* fragment was digested with *BamHI* and religated to produce a 2.9kb *EcoRI-BamHI* fragment containing the E5R gene cloned into pUC19. This plasmid was used for subsequent experiments on the E5R gene.

The DNA sequence (Dairen strain) was determined for 1210bp (cloning and sequencing described in chapter 2) and is compared to the corresponding sequence in the Copenhagen (Goebel *et al.*, 1990b) and WR (GenBank accession no. M35027) strains of vaccinia virus in fig. 5.2. The first methionine codon of the Dairen E5R ORF (denoted by an arrow) is in the same position as that of WR (positions 185-187). The Copenhagen E5R has a 2bp deletion at positions 213-214, resulting in a potential start site 30bp further downstream (position 215), in the position corresponding to the second putative methionine codon of Dairen. The remaining ORF is similar for all three viruses with all of them terminating at position 1210.

The coding sequence of E5R was compared for the three vaccinia viruses. Differences are summarized in fig. 5.3. There are 17 positions in which changes are found. In each case one virus differs and the other two are identical. Dairen has 10 differences, Copenhagen has four and WR has three. (Dairen differs in 14 positions from Copenhagen and 13 positions from WR. Copenhagen and WR have 7 differences between them). Two of the changes are transversions and the other 15, transitions. This is not unusual as transitions are known to occur more frequently than transversions. Of the 17 mismatches nine are in the third position, six in the second position and two in the first position. The increased frequency of changes in the third position relative to the first and second suggests that there has been conservation of amino acid sequence. Six of the third-base changes are silent, and all the others (11) are nonsynonymous missense mutations. Two of the base substitutions (positions 845 and 846) occur in the same codon, resulting in only one amino acid change. Of the 10 amino acid changes four are conservative (positions 343, 573, 852 and 859) and six are nonconservative.

M
C NK F E O I G L J H D A B

EcoRI EcoRI

BamHI

E5R

Fig. 5.1 *HindIII* map of vaccinia virus, Dairen showing the location of the E5R ORF.

```
VACDIE  CCATGGCCCATTCTATTAAGTCTTCCAAGTTGGCATCATCCACATATTGTGATAGTAATTCTCGGATATCAGTAGCGGTTACCGCCATTGATGTTTGTTCATTGGATGAGTAACTACTAA   120
VACCOP  CCATGGCCCATTCTATTAAGTCTTCCAAGTTGGCATCATCCACATATTGTGATAGTAATTCTCGGATATCAGTAGCGGCTACCGCCATTGATGTTTGTTCATTGGATGAGTAACTACTAA
VACWR   CCATGGCCCATTCTATTAAGTCTTCCAAGTTGGCATCATCCACATATTGTGATAGTAATTCTCGGATATCAGTAGCGGTTACCGCCATTGATGTTTGTTCATTGGATGAGTAACTACTAA

VACDIE  TGTATACATTTTCCATTTATAACACTTATGTATTAACTTTGTTCATTTATATTTTTTCATTATTATGTTGATATTAACAAAAGTGAATATATATATGTTAATAATTGTATTGTGGTTATA   240
VACCOP  TGTATACATTTTCCATTTATAACACTTATGTATTAACTTTGTTCATTTATATTTTTTCATTATTATGTTGATATTAACAAAAGTGAATATAT--ATGTTAATAATTGTATTGTGGTTATA
VACWR   TGTATACATTTTCCATTTATAACACTTATGTATTAACTTTGTTCATTTATATTTTTTCATTATTATGTTGATATTAACAAAAGTGAATATATATATGTTAATAATTGTATTGTGGTTATA

VACDIE  CGGCCACAATTTTATAATGAGTGAAAGTCAGTGTCCGATGATCAATGACGATAGCTTTACTCTGAAAAGAAAGTATCAAATCGATAGTGCGGAGTCAACAATAAAAATGGATAAGAAGAG   360
VACCOP  CGGCTACAATTTCATAATGAGTGGAAGTCAGTGTCCGATGATCAATGACGATAGCTTTACTCTGAAAAGAAAGTATCAAATCGATAGTGCGGAGTCAACAATGAAAATGGATAAGAAGAG
VACWR   CGGCTACAATTTTATAATTAGTGAAAGTCAGTGTCCGATGATCAATGACGATAGCTTTACTCTGAAAAGAAAGTATCAAATCGATAGTGCGGAGTCAACAATAAAAATGGATAAGAAGAG

VACDIE  GATAAAGTTTCAGAATAGAGCCAAAATGGTAAAAGAAATAAATCAGACAATAAGAGTAGCACAAACTCATTACGAGACATTGAAACTAGGATACATAAAATTTAAGAGAATGATTAGGAC   480
VACCOP  GACAAAGTTTCAGAATAGAGCCAAAATGGTAAAAGAAATAAATCAGACAATAAGAGCAGCACAAACTCATTACGAGACATTGAAACTAGGATACATAAAATTTAAGAGAATGATTAGGAC
VACWR   GACAAAGTTTCAGAATAGAGCCAAAATGGTAAAAGAAATAAATCAGACAATAAGAGCAGCACAAACTCATTACGAGACATTGAAACTAGGATACATAAAATTTAAGAGAATGATTAGGAC

VACDIE  TACTACTCTAGAAGATATAGCACCATCTATTCCAAATAATCAGAAAACTTATAAACTATTCTCGGACATTTCAGCCATCGGCAAAGCATCACAGAATCCGAGTAAGATGGTATATGCTCT   600
VACCOP  TACTACTCTAGAAGATATAGCACCATCTATTCCAAATAATCAGAAAACTTATAAACTATTCTCGGACATTTCAGCCATCGGCAAAGCATCACAGAATCCGAGTAAGATGGTATATGCTCT
VACWR   TACTACTCTAGAAGATATAGCACCATCTATTCCAAATAATCAGAAAACTTATAAACTATTCTCGGACATTTCAGCCATCGGCAAAGCATCACGGAATCCAAGTAAGATGGTATATGCTCT

VACDIE  GCTGCTTTACATGTTTCCCAATTTGTTTGGAGATGACCATAGATTCATTCGTTATAGAATGCATCCAATGAGTAAAATCAAACACAAGATCTTCTCTCCTTTCAAACTTAATCTTATTAG   720
VACCOP  GCTGCTTTACATGTTTCCCAATTTGTTTGGAGATGATCATAGATTCATTCGTTATAGAATGCATCCAATGAGTAAAATCAAACACAAGATCTTCTCTCCTTTCAAACTTAATCTTATTAG
VACWR   GCTGCTTTACATGTTTCCCAATTTGTTTGGAGATGATCATAGATTCATTCGTTATAGAATGCATCCAATGAGTAAAATCAAACACAAGATCTTCTCTCCTTTCAAACTTAATCTTATTAG

VACDIE  AATATTAGTGGAAGAAAGATTCTATAATAATGAATGCAGATCTAATAAATGGAGAATAATTGGAACACAAGTTGATAAAATGTTGATAGCTGAATCTGATAAATATACAATAGATGCAAG   840
VACCOP  AATATTAGTGGAAGAAAGATTCTATAATAATGAATGCAGATCTAATAAATGGAGAATAATTGGAACACAAGTTGATAAAATGTTGATAGCTGAATCTGATAAATATACAATAGATGCAAG
VACWR   AATATTAGTGGAAGAAAGATTCTATAATAATGAATGCAGATCTAATAAATGGAGAATAATTGGAACACAAGTTGATAAAATGTTGATAGCTGAATCTGATAAATATACAATAGATGCAAG

VACDIE  GTATCGCCTAAGACCTATATATAGAATCAAGGGAAAATCTGAAGAAGATACCCTCTTTATCAAACAGATGGTAGAACAATGTGTGACATCCCAGGAATTGGTGGAAAAAGTGTTGAAGAT   960
VACCOP  GTATAACCTAAAACCCATGTATAGAATCAAGGGAAAATCTGAAGAAGATACCCTCTTCATCAAACAGATGGTAGAACAATGTGTGACATCCCAGGAATTGGTGGAAAAAGTGTTGAAGAT
VACWR   GTATAACCTAAAACCCATGTATAGAATCAAGGGAAAATCTGAAGAAGATACCCTCTTTATCAAACAGATGGTAGAACAATGTGTGACATCCCAGGAATTGGTGGAAAAAGTGTTGAAGAT

VACDIE  ACTGTTTAGAGATTTGTTCAAGAGTGGAGAATACAAAGCGTACAGATACGATGATGATGTAGAAAATGGATTTATTGGATTGGATACACTAAAATTAAACATTGTTCATGATATAGTTGA   1080
VACCOP  ACTGTTTAGAGATTTGTTCAAGAGTGGAGAATACAAAGCGTACAGATACGATGATGATGTAGAAAATGGATTTATTGGATTGGATACACTAAAATTAAACATTGTTCATGATATAGTTGA
VACWR   ACTGTTTAGAGATTTGTTCAAGAGTGGAGAATACAAAGCGTACAGATACGATGATGATGTAGAAAATGGATTTATTGGATTGGATACACTAAAATTAAACATTGTTCATGATATAGTTGA

VACDIE  ACCATGTATGCCTGTTCGTAGGCCAGTGGCTAAGATACTGTGTAAAGAAATGGTAAATAAATACTTTGAGAATCCGCTACATATTATTGGTAAGAATCTTCAAGAGTGCATTGACTTTGT   1200
VACCOP  ACCATGTATGCCTGTTCGTAGGCCAGTGGCTAAGATACTGTGTAAAGAAATGGTAAATAAATACTTTGAGAATCCGCTACATATTATTGGTAAAAATCTTCAAGAGTGCATTGACTTTGT
VACWR   ACCATGTATGCCTGTTCGTAGGCCAGTGGCTAAGATACTGTGTAAAGAAATGGTAAATAAATACTTTGAGAATCCGCTACATATTATTGGTAAAAATCTTCAAGAGTGCATTGACTTTGT

VACDIE  TAGTGAATAG   1210
VACCOP  TAGTGAATAG
VACWR   TAGTGAATAG
```

Fig. 5.2 DNA sequence alignment of the E5R gene region of vaccinia viruses Dairen (VACDIE), Copenhagen (VACCOP) and WR (VACWR). Asterisks indicate positions at which there are differences between the viruses. The first translational start codons are indicated by arrows and termination codons are blocked.

Position in fig. 5.2

		{——1st cluster		——}					
		245	253	259	264	343	363	417	573	580
Codon:	**DIE**	Cac	ttT	atG	gAa	atA	aTa	gTa	cAg	ccG
	COP	Tac	ttC	atG	gGa	atG	aCa	gCa	cAg	ccG
	WR	Tac	ttT	atT	gAa	atA	aCa	gCa	cGa	ccA
AA:	**DIE**	his	phe	met	glu	ile	ile	val	gln	pro
	COP	tyr	phe	met	gly	met	thr	ala	gln	pro
	WR	tyr	phe	ile	glu	ile	thr	ala	arg	pro

			{——2nd cluster			——}		
		637	845/846	852	856	859	898	1175
Codon:	**DIE**	gaC	CGc	aGa	ccT	atA	ttT	aaG
	COP	gaT	AAc	aAa	ccC	atG	ttC	aaA
	WR	gaT	AAc	aAa	ccC	atG	ttT	aaA
AA:	**DIE**	asp	arg	arg	pro	ile	phe	lys
	COP	asp	asn	lys	pro	met	phe	lys
	WR	asp	asn	lys	pro	met	phe	lys

Fig. 5.3 Positions and sequences of DNA mismatches found amongst the vaccinia viruses Dairen (DIE), Copenhagen (COP) and Western Reserve (WR). The positions correspond to those in fig. 5.2. Upper case letters refer to the positions of change and lower case letters complete the triplet to show the position of change within the codon sequence. AA refers to the amino acid residues for which the triplets code.

An interesting feature is the distribution of the base-mismatches. Most of the differences are found in one of two clusters. In the first group (positions 245 to 264) there are four positions within 20bp in which differences are found, resulting in three amino acid changes. These differences are fairly well distributed amongst the three viral strains with Dairen differing in one position (245), Copenhagen in two (253 and 264) and WR in one position (259). The nett result is one amino acid change in each virus.

The second cluster, however, shows a definite pattern. Within 15 base pairs (positions 845 to 859, fig. 5.2 and 5.3) there are five positions in which the Dairen strain differs from the other two vaccinia viruses resulting in three amino acid changes. This 15bp sequence in variola, gerbilpox and camelpox viruses (positions 864-878 in fig. 5.5) is identical to that of Dairen. Cowpox and ectromelia viruses also have the same bases as Dairen at 4 of the 5 positions. It would seem that vaccinia strains Copenhagen and WR are the ones that differ from the general consensus in this region and the results cannot be taken to confirm the reputed derivation of Dairen from a smallpox patient.

A potentially significant difference between Copenhagen and the other two vaccinia virus DNA sequences is the 2bp deletion which results in a shortening of the ORF from the 5' end. This deletion occurs within a sequence ATATATAT and could have been formed by slippage during DNA replication; or by mismatch excision and repair in nonreplicating DNA (Levinson and Gutman, 1987) as shown in fig. 5.4. The shortened putative gene product lacks a hydrophobic signal sequence (see next chapter), which may prevent secretion of the protein. This may have a dramatic effect on the function of the protein *in vivo*, where its primary function may be located outside of the cell. An alternative scenario may be that the prototype protein was translated from a start codon further downstream and functioned intracellularly; and two base pairs were duplicated in the Dairen strain. However, the creation of a signal sequence from a random DNA sequence by insertion appears less likely to occur than the destruction of an already conserved sequence.

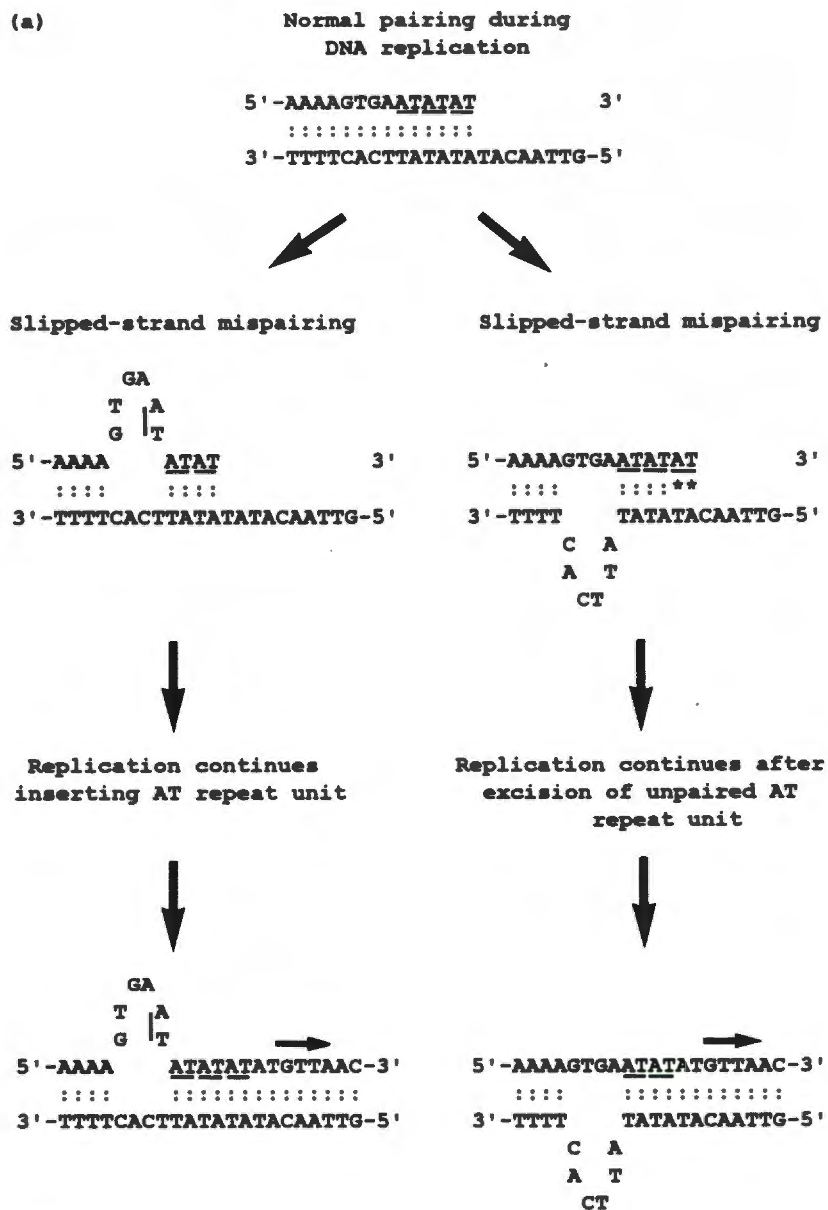


Fig. 5.4 (Legend on next page)

(b)

Normal pairing of viral DNA

```
5'-AAAAGTGAATATATATGTTAAC-3'
::::::::::::::::::::::
3'-TTTTCACTTATATATACAATTG-5'
```

Slipped-strand mispairing

```
          GA
         T  A
         G  T
5'-AAAA      ATATATGTTAAC-3'
   ::::      ::::    ::::
3'-TTTTCACTTATA      ATTG-5'
               T   A
               A   C
                TA
```

Excision/repair inserts AT repeat unit

```
          GA
         T  A
         G  T      --->
5'-AAAA      ATATATATGTTAAC-3'
   ::::      ::::::::::::::
3'-TTTTCACTTATATATACAATTG-5'
```

Fig. 5.4 (Pages 64 and 65.) Generation of duplications or deletions by slipped-strand mispairing between contiguous repeats (underlined).Arrows indicate direction of DNA synthesis. Dots denote base pairing. **(a)** Slippage during replication. Slippage in the 3'to 5' direction would result in insertion of anAT unit and slippage in the opposite direction would result in the deletion of an AT unit.The deletion shown in the right panel results from excision of the unpaired repeat unit at the 3' end of the growing strand (by the 3' to 5' exonuclease activity of DNA polymerase. **(b)** Slippage in nonreplicating DNA. Mismatched regions form single-stranded loops which may be excised and repaired. Whether an insertion or a deletion is generated will depend on which strand is excised and repaired and which strand forms the template for DNA repair (strategy from Levinson and Gutman, 1987).

Comparison of the equivalent region in different *Orthopoxviruses*

DNA from the corresponding region in five other *Orthopoxviruses*, two strains of cowpox virus (Larkin and Brighton), ectromelia virus Mill Hill, camelpox virus 903 and taterapox virus gerbilpox, was PCR amplified and cloned into T-tailed pUC18 or pUC19 (chapter 2). Primers were designed to bind to 1) a stretch of DNA including the *NcoI* site within the vaccinia virus E4L ORF (positions 52173 to 52194, Goebel *et al.*, 1990b) and 2) a DNA sequence in the E6R ORF of vaccinia virus Copenhagen (positions 53641 to 53620, Goebel *et al.*, 1990b). These primers bound to all the *Orthopoxvirus* DNA sequences examined, indicating that vaccinia virus ORFs E4L and E6R are conserved amongst the *Orthopoxviruses*. Although these two primers bound to raccoon poxvirus DNA, many of the other primers used for DNA sequencing of the E5R-equivalent region did not. This shows that the raccoon poxvirus DNA sequence corresponding to vaccinia virus ORF E5R is more diverged than that of the other *Orthopoxviruses*. The intergenic DNA sequence comparison (E5R-E6R) shows that raccoon poxvirus is widely diverged.

The DNA sequences were determined for approximately 1200bp which include the vaccinia virus E5R-equivalent ORF with its promoter region and part of the E4L-equivalent ORF. These are aligned in fig. 5.5 together with the previously described sequences of vaccinia, variola and monkeypox viruses. Gaps have been introduced (dashes) to maintain alignment of all the sequences with one another. For this reason most of the numerical positions of specific sequences differ from those of previous figures.

The first part of the sequences aligned in fig. 5.5 corresponds to the vaccinia virus ORF E4L (Goebel *et al.*, 1990b). From the portion sequenced, this gene appears well conserved in all the virus species. This is to be expected as it resembles a transcription factor (Goebel, *et al.*, 1990a) which would be essential for viral gene expression. A noticeable difference is present in the two strains of cowpox virus. There are two triplets inserted at positions 37-39 and 123-125. These may both have arisen by duplication of the adjacent sequences, which are identical to the insertions.

Fig. 5.5 (On next three pages.) DNA alignment of the sequences corresponding to vaccinia virus E5R in other *Orthopoxviruses*. VACDIE, VACCOP and VACWR refer to vaccinia viruses Dairen, Copenhagen and Western Reserve respectively; VARHAR and VARSOM refer to variola viruses Harvey and Somalia respectively; CAM903 = camelpox virus, 903; GERBIL = taterapox virus, gerbilpox virus; ECTMH = ectromelia, Mill Hill; COWLAR and COWBRI refer to cowpox viruses Larkin and Brighton respectively; and MPDEN, MPLIB, MPBEN, MP86-21, MP86-112 and MP1324 refer to monkeypox viruses Denmark, Liberia, Benin, and Zaire isolates 86-21, 86-112 and Z1324 respectively. The first potential start codons for E4L and E5R are indicated by arrows. In-frame termination codons are blocked.

```
VACDIE    CCATGGCCCATTCTATTAAGTCTTCCAAGTTGGCAT---CATCCACATATTGTGATAGTAATTCTCGGATATCAGTAGCGGTTACCGCCATTGATGTTTGTTCATTGGATGAGTAACTAC   120
VACCOP    CCATGGCCCATTCTATTAAGTCTTCCAAGTTGGCAT---CATCCACATATTGTGATAGTAATTCTCGGATATCAGTAGCGGCTACCGCCATTGATGTTTGTTCATTGGATGAGTAACTAC
VACWR     CCATGGCCCATTCTATTAAGTCTTCCAAGTTGGCAT---CATCCACATATTGTGATAGTAATTCTCGGATATCAGTAGCGGTTACCGCCATTGATGTTTGTTCATTGGATGAGTAACTAC
VARHAR    CCATGGCCCATTCTATTAAGTCTTCCAAGTTGGCAT---CATCCACATATTGTGATAGTAATTCTCGGATATTAGTAGCGGCTACCGCCATTGATGTTTGTTCATTGGATGAGTAACTAC
VARSOM    CCATGGCCCATTCTATTAAGTCTTCCAAGTTGGCAT---CATCCACATATTGTGATAGTAATTCTCGGATATTAGTAGCGGCTACCGCCATTGATGTTTGTTCATTGGATGAGTAACTAC
CAM903    CCATGGCCCATTCTATTAAGTCTTCCACGTTGGCAC---CATCCACATATTGTGATAGTAATTCTCGGATATCAGTAGCGGCTACCGCCATTGATGTTTGTTTATTGGATGAGTAACTAC
GERBIL    CCATGGCCCATTCTATTAAGTCTTCCAAGTTGGCAT---CATCCACATATTGTGATAGTAATTCTCGGATATCAGTAGCGGCTACCGCCATTAATGTTTGTTCATTGGATGAGTAACTAC
ECTMH     CCATGGCCCATTCTATTAAGTCTTCCAAGTTGGCAT---CATCCACATATTGTGATAATAATTCTCGGATATCAGTAGCGGCTACCGCCATTGATGTTTGTTCATTGGATGAGTAACTAC
COWLAR    CCATGGCCCATTCTATTAAGTCTTCCACACTGGTATTATCATCCACATATTGTGATAGTAACTCTCGGATATCAGTAGCGGCTACCGCCATTGATGTTTGTTCATTTGATGAGTAACTAC
COWBRI    CCATGGCCCATTCTATTAAGTCTTCCACGCTGGTATTATCATCCACATATTGTGATAGTAACTCTCGGATATCAGTAGCGGCTACCGCCATTGATGTTTGTTCATTTGATGAGTAACTAC
MPDEN     CCATGGCCCATTCTATTAAGTTTTCCAAGTTGGCAT---CATCCACATATTGTGATAGTAATTCTCGGATATCAGTAGTGACTACCGCCATTGATATTTGTTCATTTGATGAGTAACTAC
MPLIB     CCATGGCCCATTCTATTAAGTTTTCCAAGTTGGCAT---CATCCACATATTGTGATAGTAATTCTCGGATATCAGTAGTGACTACCGCCATTGATATTTGTTCATTTGATGAGTAACTAC
MPBEN     CCATGGCCCATTCTATTAAGTTTTCCAAGTTGGCAT---CATCCACATATTGTGATAGTAATTCTCGGATATCAGTAGTGACTACAGCCATTGATATTTGTTCATTTGATGAGTAACTAC
MP86-21   CCATGGCCCATTCTATTAAGTTTTCCAAGTTGGCAT---CATCCACATATTGTGATAGTAATTCTCGGATATCAGTAGTGACTACCGCCATTGATATTTGTTCATTTGATGAGTAACTAC
MP86-112  CCATGGCCCATTCTATTAAGTTTTCCAAGTTGGCAT---CATCCACATATTGTGATAGTAATTCTCGGATATCAGTAGTGACTACCGCCATTGATATTTGTTCATTTGATGAGTAACTAC
MP1324    CCATGGCCCATTCTATTAAGTTTTCCAAGTTGGCAT---CATCCACATATTGTGATAGTAATTCTCGGATATCAGTAGTGACTACCGCCATTGATATTTGTTCATTTGATGAGTAACTAC
          ********************* *****    *** *    ****************** *** ********** ***** *  *** ****** ** ****** *** ************

                                E4L                                                  E5R
VACDIE    TA---ATGTATACATTTTCCATTTATAACACTTATGTATTAACTTTGTTCATTTATATTTTTTCATTATTATGTTGATATTAACAAAAGTGAATATATATATGTTAATAATTGTATTGTG   240
VACCOP    TA---ATGTATACATTTTCCATTTATAACACTTATGTATTAACTTTGTTCATTTATATTTTTTCATTATTATGTTGATATTAACAAAAGTGAATATAT--ATGTTAATAATTGTATTGTG
VACWR     TA---ATGTATACATTTTCCATTTATAACACTTATGTATTAACTTTGTTCATTTATATTTTTTCATTATTATGTTGATATTAACAAAAGTGAATATATATATGTTAATAATTGTATTGTG
VARHAR    TA---ATGTATACATTTTCCATTTATAACACGTATGTATTAACTTTGTTCATTTATATTTTTTCATTATTATGTTGATATTAACAAAAGTGAATATATATATGTTAATAATTGTATTGTG
VARSOM    TA---ATGTATACATTTTCCATTTATAACACGTATGTATTAACTTTGTTCATTTATATTTTTTCATTATTATGTTGATATTAACAAAAGTGAATATATATATGTTAATAATTGTATTGTG
CAM903    TA---ATGTATACATTTTCCATTTATAACACTTATGTATTAACTTTGTTCATTTATATTTTTTCATTATTATGTTGATATTAACAAAAGTGAATATAT--ATGTTAATAATTGTATTGTG
GERBIL    TA---ATGTATACATTTTCCATTTATAACACTTATGTATTAACTTTGTTCATTTATATTTTTTCATTATTATGTTGATATTAACAAAAGTGAATATAT--ATGTTAATAATTATGTTGTG
ECTMH     TA---ATGTATACATTTTCCATTTATAACACTTATGTATTAACTTTGTTCATTTATATTTTTTCATTATTATGTTGATATTAACAAAAGTGAATATATATATGTTAATAATTGTATTGTG
COWLAR    TACTAATGAATACATTTTCCATTTATAACACTTATGTATTAACTTTGATCATTTATATTTTTTCATTATTATGTTGATATTAACAAAAGTGAATATAT--ATGTTAATAATTATGGTGTG
COWBRI    TACTAATGAATACATTTTCCATTTATAACACTTATGTATTAACTTTGATCATTTATATTTTTTCATTATTATGTTGATATTAACAAAAGTGAATATAT--ATGTTAATAATTATGGTGTG
MPDEN     TA---ATGTATACATTTTCCATTTATAACACTTATGTATTAACTTTGTTTATTTATATTTTTTCATTATTATGTTGATATTAATAA------------------------TCGTATTGTG
MPLIB     TA---ATGTATACATTTTCCATTTATAACACTTATGTATTAACTTTGTTTATTTATATTTTTTCATTATTATGTTGATATTAATAA------------------------TCGTATTGTG
MPBEN     TA---ATGTATACATTTTCCATTTATAACACTTATGTATTAACTTTGTTTATTTATATTTTTTCATTATTATGTTGATATTAATAA------------------------TCGTATTGTG
MP86-21   TA---ATGTATACATTTTCCATTTATAACACTTATGTATTAACTTTGTTTATTTATATTTTTTCATTATTATGTTGATATTAATAA------------------------TCGTATTGTG
MP86-112  TA---ATGTATACATTTTCCATTTATAACACTTATGTATTAACTTTGTTTATTTATATTTTTTCATTATTATGTTGATATTAATAA------------------------TCGTATTGTG
MP1324    TA---ATGTATACATTTTCCATTTATAACACTTATGTATTAACTTTGTTTATTTATATTTTTTCATTATTATGTTGATATTAATAA------------------------TCGTATTGTG
          **   *** ********************** **************** * ******************************** **                        *  *  ****

VACDIE    GT-TATACGGCCACAATTTTATAATGAGTGAAAG---TCAGTGTCCGATGATCAATGACGAT------AGCTTTACTCTGAAAAGAAAGTATCAAATCGATAGTGCGGAGTCAACAATAA   360
VACCOP    GT-TATACGGCTACAATTTCATAATGAGTGGAAG---TCAGTGTCCGATGATCAATGACGAT------AGCTTTACTCTGAAAAGAAAGTATCAAATCGATAGTGCGGAGTCAACAATGA
VACWR     GT-TATACGGCTACAATTTTATAATTAGTGAAAG---TCAGTGTCCGATGATCAATGACGAT------AGCTTTACTCTGAAAAGAAAGTATCAAATCGATAGTGCGGAGTCAACAATAA
VARHAR    GT-TATATGGCTACAATTTCATAATGAGCGGAAG---TCAGTGTCCGATGATTAATGACGAT------AGATTTACTCTGAAAAGAAAGTATCAAATCGATAGTGTAGAGTCGACAATGA
VARSOM    GT-TATATGGCTACAATTTCATAATGAGCGGAAG---TCAGTGTCCGATGATTAATGACGAT------AGATTTACTCTGAAAAGAAAGTATCAAATCGATAGTGTAGAGTCAACAATGA
CAM903    GT-TATACGGCTACAATTTCATAATGAGTGGAAG---TCAGTGTCCGATGATCAATGACGAT------AGCTTTACTCTGAAAAGAAAGTATCAAATCGATAGTGCAGAGTCAACAATGA
GERBIL    GT-TATACGGCTACAATTTCATAATGAGTGGAAGAAGTCAGTGTCCGATGATCAATGACGAT------AGCTTTACTCTGAAAAGAAAGTATCAAATCGATAGTGCAGAGTCAACAATGA
ECTMH     GT-TATACGGCTACAATTTCATAATGAGTGGAAG---TCAGTGTCGGATGATCAACGACGAT------AGCTTTACTTTGAAAAGAAAGTATCAAATCAATAGTGTGGAGTCAACAATGA
COWLAR    GTGTGTACGGCCACAATTTCATAATGAGTGGAAG---TCAGTGTCCTATGATCAATGACGATCTATTTAGAGTTACTTTGAAAAGAATGCATCCACTCAATAGTGAGGAGTCGAGTATGA
COWBRI    GTGTGTACGGCCACAATTTCATAATGAGTGGAAG---TCAGTGTCCTATGATCAATGACGATCTATTTAGAGTTACTTTGAAAAGAATGCATCCACTCAATAGTGAGGAGTCGAGTATGA
MPDEN     GT-TATATGGCTACAATTTCATAATGAGTTGAAG---TCAGTGTCCTATGATCAATGACGAT------AGCTTTACTCTGAAAAGAAAGTATCAAATCGATAGTGCAGAGTCAACAATGA
MPLIB     GT-TATATGGCTACAATTTCATAATGAGTTGAAG---TCAGTGTCCTATGATCAATGACGAT------AGCTTTACTCTGAAAAGAAAGTATCAAATCGATAGTGCAGAGTCAACAATGA
MPBEN     GT-TATATGGCTACAATTTCATAATGAGTTGAAG---TCAGTGTCCTATGATCAATGACGAT------AGCTTTACTCTGAAAAGAAAGTATCAAATCGATAGTGCAGAGTCAACAATGA
MP86-21   GT-TATATGGCTACAATTTCATAATGAGTTGAAG---TCAGTGTCCTATGATCAATGACGAT------AGCTTTACTCTGAAA-GAAAGTATCAAATCGATAGTGCAGAGTCAACAATGA
MP86-112  GT-TATATGGCTACAATTTCATAATGAGTTGAAG---TCAGTGTCCTATGATCAATGACGAT------AGCTTTACTCTGAAA-GAAAGTATCAAATCGATAGTGCAGAGTCAACAATGA
MP1324    GT-TATATGGCTACAATTTCATAATGAGTTGAAG---TCAGTGTCCTATGATCAATGACGAT------AGCTTTACTCTGAAA-GAAAGTATCAAATCGATAGTGCAGAGTCAACAATGA
          ** * ** *** ******* ***** **    ***    ********* ***** ** ******      **  ***** ***** *** * *** * ** ******  ***** *  ** *

VACDIE    AAATGGATAAGAAGAGGATAAAGTTTCAGAATAGAGCCAAAATGGTAAAAGAAATAAATCAGACAATAAGAGTAGCACAAACTCATTACGAGACATTGAAACTAGGATACATAAAATTTA   480
VACCOP    AAATGGATAAGAAGAGGACAAAGTTTCAGAATAGAGCCAAAATGGTAAAAGAAATAAATCAGACAATAAGAGCAGCACAAACTCATTACGAGACATTGAAACTAGGATACATAAAATTTA
VACWR     AAATGGATAAGAAGAGGACAAAGTTTCAGAATAGAGCCAAAATGGTAAAAGAAATAAATCAGACAATAAGAGCAGCACAAACTCATTACGAGACATTGAAACTAGGATACATAAAATTTA
VARHAR    AAATGGATAAGAAGAGGACAAAGTTTCAAAATAGAGCCAAAATGGTAAAAGAAATAAATCAGACAATAAGAGCAGCACAAACTCATTACGAGACATTGAAACTAGGATACATAAAATTTA
VARSOM    AAATGGATAAGAAGAGGACAAAGTTTCAAAATAGAGCCAAAATGGTAAAAGAAATAAATCAGACAATAAGAGCAGCACAAACTCATTACGAGACATTGAAACTAGGATACATAAAATTTA
CAM903    AAATGGATAAGA---GAACAAAGTTTCAGAATAGAGCCAAAATGGTAAAAGAAATAAATCAGACAATAAGAGCAGCACAAACTCATTACGAGACATTGAAACTAGGATACATAAAATTTA
GERBIL    AAATGGATAAGAAGAGGACAAAGTTTCAGAATAGAGCCAAAATGGTAAAAGAAATAAATCAGACAATAAGAGCAGCACAAACTCATTACGAGACATTGAAACTAGGATACATAAAATTTA
ECTMH     AAATGGATAAGAAGAGGACAAAGTTTCAGAATAGAGCCAAGATGGTAAAAGAAATAAATCAGACAATAAGAGCAGCCCAAACTCATTACGAGACATTGAAACTAGGATACATAAAATTTA
COWLAR    AAATAGATAAGAAGAGGACAAAGTTTCAGAATAGAGCCAAAATGGTAAAAGAAATAAATCAGACAATAAGAGCAGCACAAACTCATTACGAGACATTGAAACAAGGATACCTAAAATTTA
COWBRI    AAATAGATAAGAAGAGGACAAAGTTTCAGAATAGAGCCAAAATGGTAAAAGAAATAAATCAGACAATAAGAGCAGCACAAACTCATTACGAGACATTGAAACAAGGATACCTAAAATTTA
MPDEN     AAATGGATAAGACGATGACAAAGTTTCAGAATAGAGTCAAAATGGTAAAAGAAATAAATCAGACGATAAGAGCAGCACAAACTCATTACGAGACATTGAAACTAGGATATATAAAATTTA
MPLIB     AAATGGATAAGACGATGACAAAGTTTCAGAATAGAGTCAAAATGGTAAAAGAAATAAATCAGACGATAAGAGCAGCACAAACTCATTACGAGACATTGAAACTAGGATATATAAAATTTA
MPBEN     AAATGGATAAGACGATGACAAAGTTTCAGAATAGAATCAAAATGGTAAAAGAAATAAATCAGACGATAAGAGCAGCACAAACTCATTACGAGACATTGAAACTAGGATATATAAAATTTA
MP86-21   AAATGGATAAGACGATGACAAAGTTTCAGAATAGAGTCAAAATGGTAAAAGAAATAAATCAGACGATAAGAGCAGCACAAACTCATT--GAGACATTGAAACTAGGATATATAAAATTTA
MP86-112  AAATGGATAAGACGATGACAAAGTTTCAGAATAGAGTCAAAATGGTAAAAGAAATAAATCAGACGATAAGAGCAGCACAAACTCATT--GAGACATTGAAACTAGGATATATAAAATTTA
MP1324    AAATGGATAAGACGATGACAAAGTTTCAGAATAGAGTCAAAATGGTAAAAGAAATAAATCAGACGATAAGAGCAGCACAAACTCATT--GAGACATTGAAACTAGGATATATAAAATTTA
          **** ********     * ********** ******  *** ****************** ******* ***** ******* *** **********   ************** ******  *********
```

Fig. 5.5 (Continued over page)

```
VACDIE    AGAGAATGATTAGGACTACTACTCTAGAAGATATAGCACCATCTATT---CCAAATAATCAGAAAACTTATAAACTATTCTCGGACATTTCAGCCATCGGCAAAGCATCACAGAATCCGA  600
VACCOP    AGAGAATGATTAGGACTACTACTCTAGAAGATATAGCACCATCTATT---CCAAATAATCAGAAAACTTATAAACTATTCTCGGACATTTCAGCCATCGGCAAAGCATCACAGAATCCGA
VACWR     AGAGAATGATTAGGACTACTACTCTAGAAGATATAGCACCATCTATT---CCAAATAATCAGAAAACTTATAAACTATTCTCGGACATTTCAGCCATCGGCAAAGCATCACGGAATCCAA
VARHAR    AGAAAATGATTAGGACTACTACTTTAGAAGATATAACAACATCTATT---CCAAATATTCAGAAAATTTATAAACTATTCTCGGACATTTCAGCCATTGGCAAAGTATCACAGAATCCGA
VARSOM    AGAAAATGATTAGGACTACTACTTTAGAAGATATAACAACATCTATT---CCAAATATTCAGAAAATTTATAAACTATTCTCGGACATTTCAGCCATTGGCAAAGTATCACAGAATCCGA
CAM903    AGAAAATGATTATGACTACTACTCCAGAAGATATAGCAACATCTATT---CCAAATATTCAGAAAATTTATAAACTATTCTCGGACATTTCAGCCATTGGCAAAGCATCACAGAATCCGA
GERBIL    AGAAAATGATTAGGACTACTACTCTAGAAGATATGGCAACATCTATT---CCAAATATTCAGAAAATTTATAAACTATTCTCGGACATTTCAGCCATTGGCAAAGCATCACAGAATCCGA
ECTMH     AGAGAATGATTAGGACTACTACTCTAGAAGATATAGCAACATCTATT---CCAAATATTCAGAAAATTTATAAACTATTCTCAGACATTTCAGCCATTGGAAAAGCATCACAGAATCCGA
COWLAR    AGAGAATGATTAGAACTGCTACTCTAAAAGATATAGCACCATCTATT---CCAAAAATTCAGAAATTTTATAAACTATTCTCGGACATTTCAGCCATCAGCAAAGCATCACAGAATCCGA
COWBRI    AGAGAATGATTAGAACTGCTACTCTAAAAGATATAGCACCATCTATT---CCAAAAATTCAGAAATTTTATAAACTATTCTCGGACATTTCAGCCATCAGCAAAGCATCACAGAATCCGA
MPDEN     AGGGAATGATTAGGACTACTACTCTAGAAGATATAGCACCATCTATT---CCAAATAATCAGAAAACTTATAAACTATTCTCGTACATTTCAGTCATTGGCAAAGCATCACAGAATCCGA
MPLIB     AGGGAATGATTAGGACTACTACTCTAGAAGATATAGCACCATCTATT---CCAAATAATCAGAAAACTTATAAACTATTCTCGTACATTTCAGTCATTGGCAAAGCATCACAGAATCCGA
MPBEN     AGGGAATGATTAGGACTACTACTCTAGAAGATATAGCACCATCTATT---CCAAATAATCAGAAAACTTATAAACTATTCTCGGACATTTCAGTCATTGGCAAAGCATCACAGAATCCGA
MP86-21   AGTGAATGATTAGGACTACTACTCTAGAAGATATAGCACCATCTATTATTCCAAATAATCAGAAAACTTATAAACTATTCTCGGACATTTCAGTCATTGGCAAAGCATCACAGAATCCGA
MP86-112  AGTGAATGATTAGGACTACTACTCTAGAAGATATAGCACCATCTATTATTCCAAATAATCAGAAAACTTATAAACTATTCTCGGACATTTCAGTCATTGGCAAAGCATCACAGAATCCGA
MP1324    AGTGAATGATTAGGACTACTACTCTAGAAGATATAGCACCATCTATTATTCCAAATAATCAGAAAACTTATAAACTATTCTCGGACATTTCAGTCATTGGCAAAGCATCACAGAATCCGA
          **  ********  *** *****  * *******  ** ********    ***** * *******  ***************  ********* ***  *  **** ***** ****** *

VACDIE    GTAAGATGGTATATGCTCTGCTGCTTTACATGTTTCCCAATTTGTTTGGAGATGACCATAGATTCATTCGTTATAGAATGCATCCAATGAGTAAAATCAAACACAAGATCTTCTCTCCTT  720
VACCOP    GTAAGATGGTATATGCTCTGCTGCTTTACATGTTTCCCAATTTGTTTGGAGATGATCATAGATTCATTCGTTATAGAATGCATCCAATGAGTAAAATCAAACACAAGATCTTCTCTCCTT
VACWR     GTAAGATGGTATATGCTCTGCTGCTTTACATGTTTCCCAATTTGTTTGGAGATGATCATAGATTCATTCGTTATAGAATGCATCCAATGAGTAAAATCAAACACAAGATCTTCTCTCCTT
VARHAR    GTAAAATGGCATATGCTTTACTGCTTTACATGTTCCCCAATTTGTTTGGAGATGACCATAGATTCATTCTTTATAGAATGTTTCCAATGAGTAAAATCAAACACAAGATCTTCTCTCCTT
VARSOM    GTAAAATGGCATATGCTTTACTGCTTTACATGTTCCCCAATTTGTTTGGAGATGACCATAGATTCATTCTTTATAGAATGTTTCCAATGAGTAAAATCAAACACAAGATCTTCTCTCCTT
CAM903    TTAAGATGGTATATGCTCTACTGCTTTTCATGTTCCCCAATCTGTTTGGAGATGACCATAGATTCATTCTTTATAGAATACATCCAATGAGTAAAATCAAACACAAGATCTTCTCTCCTT
GERBIL    GTAAGATGGCATATGCTCTACTGCTTTACATGTTCCCCAATTTGTTTGGAGATGACCATAGATTCATTCTTTATAGAATGCATCCAATGAGTAAAATCAAACACAAGATCTTCTCTCCTT
ECTMH     GTAAGATGGTATATGCTCTACTGCTTTATATGTTTCCCAATTTGTTTGGAGACGATCATAGATTCATTCTTTATAGAATGCATCCAATGAGCAAAATCAAACACAAGATCTTCTCACCGT
COWLAR    GTAAGATGGTATATGCTCTACTGCTTTACATGTTACCCAATTTGTTTGGAGATGACCATAGATTCATTCGTTATAGAATGCATCCAATGAGTAAAATCAAACACAAGATCTTCTCCCCGT
COWBRI    GTAAGATGGTATATGCTCTACTGCTTTACATGTTTCCCAATTTGTTTGGAGATGACCATAGATTCATTCGTTATAGAATGCATCCAATGAGCAAAATCAAACACAAGATCTTCTCCCCGT
MPDEN     GTAAGATGATATATGCTC-GCTGCTTTACATGTTTCCCAATTTGTTTGGAGATGACCATAGATTCATTTGTTATAGAATGCATCCAA---------------------------------
MPLIB     GTAAGATGATATATGCTC-GCTGCTTTACATGTTTCCCAATTTGTTTGGAGATGACCATAGATTCATTTGTTATAGAATGCATCCAA---------------------------------
MPBEN     GTAAGATGATATATGCTC-GCTGCTTTACATGTTTCCCAATTTGTTTGGAGATGACCATAGATTCATTTGTTATAGAATGCATCCAA---------------------------------
MP86-21   GTAAGATGATATATGCTC-GCTGCTTTACATGTTTCCCAATTTGTTTGGAGATGACCATAGATTCATTTGTTATAGAATGCATCCAATGAGTAAAATCAAACACAAAATCTT-----CGT
MP86-112  GTAAGATGATATATGCTC-GCTGCTTTACATGTTTCCCAATTTGTTTGGAGATGACCATAGATTCATTTGTTATAGAATGCATCCAATGAGTAAAATCAAACACAAAATCTT-----CGT
MP1324    GTAAGATGATATATGCTC-GCTGCTTTACATGTTTCCCAATTTGTTTGGAGATGACCATAGATTCATTTGTTATAGAATGCATCCAATGAGTAAAATCAAACACAAAATCTT-----CGT
           *** ***  *******   *******  ***** ****** ********** ** ************  *********    *****

VACDIE    TCAAACTTAATCTTATTAGAATATTAGTGGAAGAAAGATTCTATAATAATGAATGCAGATCTAATAAATGGAGAATAATTGGAACACAAGTTGATAAAATGTTGATAGCTGAATCTGATA  840
VACCOP    TCAAACTTAATCTTATTAGAATATTAGTGGAAGAAAGATTCTATAATAATGAATGCAGATCTAATAAATGGAGAATAATTGGAACACAAGTTGATAAAATGTTGATAGCTGAATCTGATA
VACWR     TCAAACTTAATCTTATTAGAATATTAGTGGAAGAAAGATTCTATAATAATGAATGCAGATCTAATAAATGGAGAATAATTGGAACACAAGTTGATAAAATGTTGATAGCTGAATCTGATA
VARHAR    TCAAACTTAATCTTATTAGAATATTAGTGGAAGAAAGATTCTATAATAATGAATGCAGATCTAATAAATGGAGAATAATTGGAACACAAGTTGATAAAATGTTGATAGTTGAATCTGATA
VARSOM    TCAAACTTAATCTTATTAGAATATTAGTGGAAGAAAGATTCTATAATAATGAATGCAGATCTAATAAATGGAGAATAATTGGAACACAAGTTGATAAAATGTTGATAGTTGAATCTGATA
CAM903    TCAAACTTAATCTTATTAGAATATTAGTGGAAGAAAGATTCTATAATAATGAATGCAGATCTAATAAATGGAGAATAATTGGAACACAAGTTGATAAAATGTTGATAGCTGAATCTGATA
GERBIL    TCAAACTTAATCTTATTAGAATATTAGTGGAAGAAAGATTCTATAATAATGAATGCAGATCTAATAAATGGAGAATAATTGGAACACAAGTTGATAAAATGTTGATAGCTGAATCTGATA
ECTMH     TCAAACTTAATCTTATTAGGATATTAGTGGAAGAAAGATTCTATAATAATGAATGCAAAGATTATAGATGGAAAATAATTGGATCACAAGTTGATAAAATATTGATAGCTGAATCTGCTA
COWLAR    TCAAACTTAATCTTATTAGAATATTAGTGGAAGAAAGATTCTATAATAATGAATGCAGAGATTATAAATGGAGAATAATTGGAACACAGGTTGATAAAATATTGATAGCTGAATCTGCTA
COWBRI    TCAAACTTAATCTTATTAGAATATTAGTGGAAGAAAGATTCTATAATAATGAATGCAGAGATTATAAATGGAGAATAATTGGAACACAAGTTGATAAAATATTGATAGCTGAATCTGCTA
MPDEN     ------------------------------------------------------------------------------------------------------------------------
MPLIB     ------------------------------------------------------------------------------------------------------------------------
MPBEN     ------------------------------------------------------------------------------------------------------------------------
MP86-21   TCAAACTTAATCTTATTAGAATATTAGTGGAAGAAAGATTCTATAATAATGAATGCAGAGATTATAAATGGAGAATAATTGGAACACAAGTTGATAAAATATTGATAGC---------TA
MP86-112  TCAAACTTAATCTTATTAGAATATTAGTGGAAGAAAGATTCTATAATAATGAATGCAGAGATTATAAATGGAGAATAATTGGAACACAAGTTGATAAAATATTGATAGC---------TA
MP1324    TCAAACTTAATCTTATTAGAATATTAGTGGAAGAAAGATTCTATAATAATGAATGCAGAGATTATAAATGGAGAATAATTGGAACACAAGTTGATAAAATATTGATAGC---------TA

VACDIE    AATATACAATAGATGCAAGGTATCGCCTAAGACCTATATATAGAATCAAGGGAAAATCTGAAGAAGATACCCTCTTTATCAAACAGATGGTAGAACAATGTGTGACATCCCAGGAATTGG  960
VACCOP    AATATACAATAGATGCAAGGTATAACCTAAAACCCATGTATAGAATCAAGGGAAAATCTGAAGAAGATACCCTCTTCATCAAACAGATGGTAGAACAATGTGTGACATCCCAGGAATTGG
VACWR     AATATACAATAGATGCAAGGTATAACCTAAAACCCATGTATAGAATCAAGGGAAAATCTGAAGAAGATACCCTCTTTATCAAACAGATGGTAGAACAATGTGTGACATCCCAGGAATTGG
VARHAR    AATATACAATAGATGCAAGGTATCGCCTAAGACCTATATATAGAATCAAGGGAAAATCTGAAGAAGATACCCTCTTCATCAAACAGATGGTAGACCAATGTGTGACATCCCAGGAATTGG
VARSOM    AATATACAATAGATGCAAGGTATCGCCTAAGACCTATATATAGAATCAAGGGAAAATCTGAAGAAGATACCCTCTTCATCAAACAGATGGTAGACCAATGTGTGACATCCCAGGAATTGG
CAM903    AATATACAATAGATGCAAGGTATCGCCTAAGACCTATATATAGAATCAAGGGAGAATCTGAAGAAGATACCCTCTTCATCAAACAGATGGTAGAACAATGTGTGACATCCCAGGAATTGG
GERBIL    AATATACAATAGATGCAAGGTATCGTCTAAGACCTATATATAGAATCAAGGGAGAATCTGAAGAAGATACCCTCTTCATCAAACAGATGGTAGAACAATGTGTGACATCCCAGGAATTGG
ECTMH     AATATACAATGAATGCAATGTATCGCCTAAGACCGATACATAGAATCAAGGTAGAATCTGAAGAAGATACCCTCTTCATCAAACAGATGGTAGAAAAATGTGTAACATCCCAGGAATTGG
COWLAR    AATATACAATAGATGCAAGGTATCGCCTAAGACCTATATATAGAATCAAGGGAGAATCTGAAGAAGATACTCTCTTCATCAAACAGATGGTAGAGAAATGTGTAACGTCTCAGGAATTGG
COWBRI    AATATACAATAGATGCAAGGTATCGCCTAAGACCTATATATAGAATCAAGGGAGAATCTGAAGAAGATACCCTCTTCATCAAACAGATGGTAGAGAAATGTGTAACGTCTCAGGAATTGG
MPDEN     ------------------------------------------------------------------------------------------------------------------------
MPLIB     ------------------------------------------------------------------------------------------------------------------------
MPBEN     ------------------------------------------------------------------------------------------------------------------------
MP86-21   AATATACAATAGATGCAATGTATCGCATAAGACCGATATATA------------------------------------------------------------------------------
MP86-112  AATATACAATAGATGCAATGTATCGCATAAGACCGATATATA------------------------------------------------------------------------------
MP1324    AATATACAATAGATGCAATGTATCGCATAAGACCGATATATA------------------------------------------------------------------------------
```

Fig. 5.5 (Continued over page)

```
VACDIE    TGGAAAAAGTGTTGAAGATACTGTTTAGAGATTTGTTCAAGAGTGGAGAATACAAAGC-GTACAGATACGATGATGATGTAGAAAATGGATTTATTGGATTGGATACACTAAAATTAAAC  1080
VACCOP    TGGAAAAAGTGTTGAAGATACTGTTTAGAGATTTGTTCAAGAGTGGAGAATACAAAGC-GTACAGATACGATGATGATGTAGAAAATGGATTTATTGGATTGGATACACTAAAATTAAAC
VACWR     TGGAAAAAGTGTTGAAGATACTGTTTAGAGATTTGTTCAAGAGTGGAGAATACAAAGC-GTACAGATACGATGATGATGTAGAAAATGGATTTATTGGATTGGATACACTAAAATTAAAC
VARHAR    TGGAAAAAGTGTTGAAGATACTGTTTAGAGATTTGTTCAAGAGTGGAGAATACAAAGC-GTACAGATACGATGATGACGTAGAAAATGGATTCATTGGATTGGATAAACTAAAATTAAAC
VARSOM    TGGAAAAAGTGTTGAAGATACTGTTTAGAGATTTGTTCAAGAGTGGAGAATACAAAGC-GTACAGATACGATGATGACGTAGAAAATGGATTCATTGGATTGGATAAACTAAAATTAAAC
CAM903    TGGAAAAAGTGTTGAAGATACTGTTTAGAGATTTGTTCAAGAGTGGAGAATACAAAGC-GTACAGATACGATGATG---TAGAAAATGTATTCATTGGATTGGATAAACTAAAATTAAAC
GERBIL    TGGAAAAAGTGTTGAAGATACTGTTTAGAGATTTGTTCAAGAGTGGAGAATACAAAGC-GTACAGATACGATGATGATGTAGAAAATGGATTCATTGGATTGGATAAACTAAAATTAAAC
ECTMH     TGGAAAAA-TGTTGAAGATACTGTTTAGAGATTTGTTCAAGAGTGGAGAATACAAAGC-GTACAGGTACGATGATGATGTAGAAAATGGATTCATTGGATTGGATAAACTAAAATTAAAC
COWLAR    TGGAAAAAGTGTTGAAGATACTGTTTAAAGATTTGTTCAAGAGTGGAGAATACAAAG-ATTACAGATACGATGATGATGTAGACAATGGATTCATTGGATTGGACAAACTAAAATTAAAC
COWBRI    TGGAGAAAGTGTTGAAGATACTGTTTAAAGATTTGTTCAAGAGTGGAGAATACAAAG-ATTACAGATACGATGATGATGTAGACAATGGATTCATTGGATTGGACAAACTAAAATTAAAC
MPDEN     -----------------------------------------------------------------------------------------------------------------------C
MPLIB     -----------------------------------------------------------------------------------------------------------------------C
MPBEN     -----------------------------------------------------------------------------------------------------------------------C
MP86-21   ------------------------------------------------A-TACAAAGCAGTACAGATACAATGATGATGTAGAAAATGGATTCATTGGATTGGATAAACTAAAATTAAAC
MP86-112  ------------------------------------------------A-TACAAAGCAGTACAGATACAATGATGATGTAGAAAATGGATTCATTGGATTGGATAAACTAAAATTAAAC
MP1324    ------------------------------------------------A-TACAAAGCAGTACAGATACAATGATGATGTAGAAAATGGATTCATTGGATTGGATAAACTAAAATTAAAC
                                                                                                                                 *

VACDIE    ATTGTTCATGATATAGTTGAACCATGTATGCCTGTTCGTAGGCCAGTGGCTAAGATACTGTGTAAAGAAATGGTAAATAAATACTTTGAGAATCCGCTACATATTATTGGTAAGAATCTT  1200
VACCOP    ATTGTTCATGATATAGTTGAACCATGTATGCCTGTTCGTAGGCCAGTGGCTAAGATACTGTGTAAAGAAATGGTAAATAAATACTTTGAGAATCCGCTACATATTATTGGTAAAAATCTT
VACWR     ATTGTTCATGATATAGTTGAACCATGTATGCCTGTTCGTAGGCCAGTGGCTAAGATACTGTGTAAAGAAATGGTAAATAAATACTTTGAGAATCCGCTACATATTATTGGTAAAAATCTT
VARHAR    ATTGTTCATGATATAGTTGAACCATGTATGCCTGTTCGTAGGCCAGTGGCTAAGATACTGTGTAAAGAAATGGTAAATAAATACTTTGAGAATCCGCTGCATATTATTGGTAAGAATCTT
VARSOM    ATTGTTCATGATATAGTTGAACCATGTATGCCTGTTCGTAGGCCAGTGGCTAAGATACTGTGTAAAGAAATGGTAAATAAATACTTTGAGAATCCGCTGCATATTATTGGTAAGAATCTT
CAM903    ATTCTTCATGATATAGTTGAACCATGTATGCCTGTTCGTAGGCCAGTGGCTAAGATACTGTGTAAAGAAATGGTAAATAAATACTTTGAGAATCCGCTACATATTATTGGTAAGAATCTT
GERBIL    ATTGTTCATGATATAGTTGAACCATGTATGCCTGTTCGTAGGCCAGTGGCTAAGATACTGTGTAAAGAAATGGTAAATAAATACTTTGAGAATCCGCTACATATTATTGGTAAGAATCTT
ECTMH     ATTGTTCATGATATAGTTGAACCATGTATGCCTGTCCGCAGGCCAGTGGCTAAGATACTGTGTAAAGAAATGGTAAATAAATACTTTGAGAATCCGCTACATATTATTGGTAAGAATCTT
COWLAR    ATTGTTCACGATATAGTTGAACCATGTATGCCTGTTCGTAGGCCAGTGGCTAAGGTACTGTGTAAAGAAATGGTAAATAAATACTTTGAGAATCCTCTACATATCATTGGTAAGAATCTC
COWBRI    ATTGTTCACGATATAGTTGAACCATGTATGCCTGTTCGTAGGCCAGTGGCTAAGGTACTGTGTAAAGAAATGGTAAATAAATACTTTGAGAATCCTCTACATATCATTGGTAAGAATCTC
MPDEN     ATTGTTCATGATATAGTTGAATCATGTATGCCTGTTCGTATGCCTGTGGCTAAGATACTGTGTAAAGAAATGGTAAATAAATACTTTGAGAATC------------------------TT
MPLIB     ATTGTTCATGATATAGTTGAATCATGTATGCCTGTTCGTATGCCTGTGGCTAAGATACTGTGTAAAGAAATGGTAAATAAATACTTTGAGAATC------------------------TT
MPBEN     ATTGTTCATGATATAGTTGAATCATGTATGCCTGTTCGTATGCCTGTGGCTAAGATACTGTGTAAAGAAATGGTAAATAAATACTTTGAGAATC------------------------TT
MP86-21   ATTGTTCATGATATAGTTGAATCATGTATACCTGTTCGTATGCCTGTGGCTAAGATACTGTGTAAAGAAATGGTAAATAAATACTTTGAGAATC------------------------TT
MP86-112  ATTGTTCATGATATAGTTGAATCATGTATACCTGTTCGTATGCCTGTGGCTAAGATACTGTGTAAAGAAATGGTAAATAAATACTTTGAGAATC------------------------TT
MP1324    ATTGTTCATGATATAGTTGAATCATGTATACCTGTTCGTATGCCTGTGGCTAAGATACTGTGTAAAGAAATGGTAAATAAATACTTTGAGAATC------------------------TT
          *** **** ************ ******* ***** ** * *** ********* ***************************************                        *

                                                                                                                                 1230
VACDIE    CAAGAGTGCATTGACTTTGTTAGTGAATAG
VACCOP    CAAGAGTGCATTGACTTTGTTAGTGAATAG
VACWR     CAAGAGTGCATTGACTTTGTTAGTGAATAG
VARHAR    CAAGAGTGCATTGACTTTGTTAGTGAATAG
VARSOM    CAAGAGTGCATTGACTTTGTTAGTGAATAG
CAM903    CAAGAGTGCATTGACTTTGTTAATGAATAG
GERBIL    CAAGAGTGCATTGACTTTGTTAGTGAATAG
ECTMH     CAAGAGTGCATTGACTTTGTTAGTGAGTAG
COWLAR    CAAGAGTGCATTGACTTTGTTAGTGAGTAG
COWBRI    CAAGAGTGCATTGACTTTGTTAGTGAGTAG
MPDEN     TAAGAGTGCATTGACTTTGTTAGTGAATAG
MPLIB     TAAGAGTGCATTGACTTTGTTAGTGAATAG
MPBEN     TAAGAGTGCATTGACTTTGTTAGTGAATAG
MP86-21   TAAGAGTGTATTGACTTTGTTAGTGAATAT
MP86-112  TAAGAGTGTATTGACTTTGTTAGTGAATAT
MP1324    TAAGAGTGTATTGACTTTGTTAGTGAATAT
           ******* ************* *** **
```

Fig. 5.5

The E4L-E5R promoter region

Fig. 5.6 shows the aligned sequences corresponding to the promoter region of vaccinia virus E4L and E5R. In all sequences the first ATG of E4L occupies positions 1 to 3. E4L has a typical late promoter consesus sequence at its start - TAAAT (Cochran *et al.*, 1985; Bertholet *et al.*, 1986; Weir and Moss, 1987; Davison and Moss, 1989b). Intermediate genes also have this consensus sequence, but, in addition, have a typical core and spacer region (Baldick *et al.*, 1992). The DNA sequence upstream of the E4L coding sequence does not resemble that of intermediate promoter sequences (Baldick *et al.*, 1992).

The monkeypox virus sequences have been included in fig. 5.6 to emphasize the DNA sequence conservation in this region. The DNA sequence has been well conserved, probably because changes acceptable for one promoter region are not acceptable for the other.

The DNA sequence between the Dairen E4L and E5R (positions 4-51) is highly A/T-rich (85% over 48bp). Extending the promoter region to 76bp (for vaccinia virus Copenhagen, camelpox virus and gerbilpox virus), the A/T composition is 84%. This A/T-rich sequence is typical of early vaccinia virus promoters (Davison and Moss, 1989a). Experiments to determine the functional activity of these putative promoters have not been done. Preliminary RNA work suggests that the E5R gene is transcribed from DNA further upstream (next chapter). This may be due to the absence of an early transcriptional termination sequence (TTTTTNT) closer to the E5R ORF.

```
          E4L                                          E5R
          <-                                           ->
VACDIE    CATTTATAACACTTATGTATTAACTTTGTTCATTTATATTTTTTCATTATTATGTTGATATTAACAAAAGTGAATATATATATG
VACCOP    CATTTATAACACTTATGTATTAACTTTGTTCATTTATATTTTTTCATTATTATGTTGATATTAACAAAAGTGAATATAT--ATG
VACWR     CATTTATAACACTTATGTATTAACTTTGTTCATTTATATTTTTTCATTATTATGTTGATATTAACAAAAGTGAATATATATATG
VARHAR    CATTTATAACACGTATGTATTAACTTTGTTCATTTATATTTTTTCATTATTATGTTGATATTAACAAAAGTGAATATATATATG
VARSOM    CATTTATAACACGTATGTATTAACTTTGTTCATTTATATTTTTTCATTATTATGTTGATATTAACAAAAGTGAATATATATATG
CAM903    CATTTATAACACTTATGTATTAACTTTGTTCATTTATATTTTTTCATTATTATGTTGATATTAACAAAAGTGAATATAT--ATG
GERBIL    CATTTATAACACTTATGTATTAACTTTGTTCATTTATATTTTTTCATTATTATGTTGATATTAACAAAAGTGAATATAT--ATG
ECTMH     CATTTATAACACTTATGTATTAACTTTGTTCATTTATATTTTTTCATTATTATGTTGATATTAACAAAAGTGAATATATATATG
COWLAR    CATTTATAACACTTATGTATTAACTTTGATCATTTATATTTTTTCATTATTATGTTGATATTAACAAAAGTGAATATAT--ATG
COWBRI    CATTTATAACACTTATGTATTAACTTTGATCATTTATATTTTTTCATTATTATGTTGATATTAACAAAAGTGAATATAT--ATG
MPDEN     CATTTATAACACTTATGTATTAACTTTGTTTATTTATATTTTTTCATTATTATGTTGATATTAATAA-----------------
MPLIB     CATTTATAACACTTATGTATTAACTTTGTTTATTTATATTTTTTCATTATTATGTTGATATTAATAA-----------------
MPBEN     CATTTATAACACTTATGTATTAACTTTGTTTATTTATATTTTTTCATTATTATGTTGATATTAATAA-----------------
MP86-21   CATTTATAACACTTATGTATTAACTTTGTTTATTTATATTTTTTCATTATTATGTTGATATTAATAA-----------------
MP86-112  CATTTATAACACTTATGTATTAACTTTGTTTATTTATATTTTTTCATTATTATGTTGATATTAATAA-----------------
MP1324    CATTTATAACACTTATGTATTAACTTTGTTTATTTATATTTTTTCATTATTATGTTGATATTAATAA-----------------
          ************ *************** * ********************************* **
```

Fig 5.6 DNA sequence of the E4L and E5R promoter region. This sequence alignment is taken from positions 140 to 223 of fig. 5.5. It starts at the first translational start codon of ORF E4L (positions 1-3) and ends after the second potential start codon of the Dairen E5R ORF.

There are three positions in which differences are found between the different viruses - positions 13, 29 and 31. The difference at position 31 is conserved in the monkeypox viruses only. The other positions show a difference in one species only - variola and cowpox viruses respectively. Since the DNA sequence between the two ORFs is so highly conserved, it would be of interest to test the strength of these two putative promoters. They may be useful as a bi-directional promoter in the expression of more than one foreign gene in poxviruses, as described by Kumar and Boyle (1990), for a fowlpox virus bi-directional promoter, and Tsao *et al.* (1988), for the bi-directional promoter found upstream of vaccinia virus ORFs F16L (p25) and F17R (p11). Furthermore, it would be useful if these promoters were functional in other poxviruses, such as *Capripoxviruses* and *Avipoxviruses*, which are good potential vaccine vectors (Baxby, 1993). Promoter sequences from vaccinia virus have been shown to function in raccoon poxvirus (Esposito *et al.*, 1988), *Capripoxviruses* (A-L Williamson, personal communication), *Avipoxviruses* (Boyle and Coupar, 1988b) and *Entomopoxviruses* (Gruidl *et al.*, 1992).

Pairwise DNA Comparison

The base similarity between different pairs of viral sequences was determined for the DNA sequences shown in fig. 5.5. These values are listed in 5.7 in the form of a matrix, showing % divergence (top right) and % similarity (bottom left). Overall, the DNA sequences are highly conserved. All comparisons show >91% base similarity with the greatest difference being 8.7% between monkeypox virus, Benin and cowpox virus, Larkin. This high % of overall similarity may represent recent speciation, or strong conservation of function. It also indicates a high fidelity of replication of the poxvirus DNA polymerase.

VIRUS	% DIVERGENCE												
	1	2	3	4	5	6	7	8	9	10	11	12	13
1. VACDIE	-	1.24	1.07	3.31	3.22	4.42	4.55	4.72	5.97	5.80	2.15	2.75	4.71
2. VACCOP	**98.76**	-	0.66	3.23	3.15	3.64	3.77	4.52	6.05	5.88	2.07	2.66	4.39
3. VACWR	**98.93**	**99.34**	-	3.88	3.80	4.55	4.68	5.21	6.71	6.55	2.73	3.33	5.05
4. VARHAR	96.69	96.77	96.12	-	0.08	6.10	6.23	6.00	6.71	6.63	2.15	2.83	5.21
5. VARSOM	96.78	96.85	96.2	**99.92**	-	5.97	6.10	5.90	6.79	6.71	2.07	2.75	5.13
6. MPLIB	95.58	96.36	95.45	93.90	94.03	-	0.39	0.65	8.57	8.31	4.81	5.61	6.36
7. MPBEN	95.45	96.23	95.32	93.77	93.90	**99.61**	-	0.78	8.70	8.44	4.94	5.74	6.49
8. MPZAI	95.28	95.48	94.79	94.00	94.10	**99.35**	**99.22**	-	7.57	7.37	4.92	5.54	6.29
9. COWLAR	94.03	93.95	93.29	93.29	93.21	91.43	91.30	92.43	-	0.49	5.55	6.33	6.55
10. COWBRI	94.2	94.12	93.45	93.37	93.29	91.69	91.56	92.63	**99.5**	-	5.47	6.24	6.22
11. GERBIL	97.85	97.93	97.27	97.85	97.93	95.19	96.06	95.08	94.45	94.53	-	1.58	4.06
12. CAMEL	97.25	97.34	96.67	97.17	97.25	94.39	94.26	94.66	93.67	93.76	98.42	-	4.66
13. ECTMH	95.29	95.61	94.95	94.79	94.87	93.64	93.51	93.71	93.45	93.78	95.94	95.34	-

% BASE SIMILARITY

Fig. 5.7 Matrix showing the % Divergence (top right) and % similarity (bottom left) between pairs of DNA sequences. Intraspecies homology is indicated in bold. The nomenclature has been explained in the legend to fig. 5.5.

The base similarity within each of the monkeypox, variola, and cowpox virus species is greater than 99%. Vaccinia virus, also highly conserved (>98%), shows more intraspecies variation than the others. Although this may reflect an earlier divergence within the species, it may also be the result of recombination during the evolution of the virus (Bedson and Dumbell, 1964). More sequencing data from different strains of the same viral species, as well as sequences from different regions in the genome, is required for a valid comparison to be made.

Between species, gerbilpox and camelpox viruses are most closely related to one another (98.42%), but also show a high degree of similarity to the vaccinia and variola viruses. They are more distantly related to monkeypox virus. Since gerbilpox virus was isolated from a rodent in Benin, one may have expected a closer relationship to monkeypox virus, which also infects rodents and is found in the same part of the world. Gibbs and Fenner (1984) group gerbilpox (taterapox) with monkeypox virus, and not variola or vaccinia viruses. This analysis, based on restriction enzyme cleavage sites, is not as accurate as DNA sequence analysis. The phenotypic characters of gerbilpox virus most closely resemble those of variola virus (Lourie *et al.*, 1975).

Cowpox virus has the greatest percentage divergence. It is most similar to gerbilpox virus (94.5%) and least similar to monkeypox virus (91-92%).

Ectromelia virus cannot clearly be grouped with any other virus. Like all the other viruses, it is most similar to gerbilpox virus. The next closest is vaccinia virus (±95%). The amino acid sequence (fig. 6.7) shows six amino acid changes which are shared with cowpox virus (positions 50, 196, 197, 209, 215 and 248), whereas only four of all the other differences are shared with any of the other viruses (positions 52, 110, 118 and 159). This suggests that ectromelia virus may actually be more closely related to cowpox virus.

Phylogenetic analysis of the *Orthopoxvirus* DNA sequences

The sequences were analysed using two computer programs, clustal V (Higgins and Sharp, 1988; Higgins *et al.*, 1991) and TREECON (Van de Peer and Wachter, 1992). In both cases the neighbour-joining method (Saitou and Nei, 1987) was used for constructing phylogenetic trees. This was considered the most accurate method for obtaining the correct tree (Saitou and Imanishi, 1989; Sourdis and Nei, 1988; Ed Rybicki, personal communication). Similar results were obtained from both programs.

In all cases no correction was made for superimposed mutations. The DNA sequences are well conserved amongst the viruses and many differences are in the form of insertions and deletions, which cannot back-mutate. We therefore do not expect a large error due to superimposed mutations.

Insertions and deletions (indels) were considered important in the analyses, and trees were constructed both taking these into account and ignoring them. When they are accouted for, adjacent gaps in the sequence are treated as a single gap, regardless of its length. This is because a deletion or insertion of any length of sequence may be produced by a single mutational event.

In order to produce a tree which can be visualized on the screen and saved, using TREECON, a root has to be designated. The trees shown in figures 5.8 and 5.9 were constructed using the automatic root location, which places the root at a point where the average distance to the species on both sides is the same (midpoint rooting). Trees were constructed using a number of different sequences as the root, but the interrelationship of the viruses was the same for all roots.

The validity of the branching pattern produced by a single matrix distance calculation was tested by performing bootstrap analysis on the data. For analysis using the clustal program, 1000 bootstraps were performed, where random blocks of 111 nucleotides were compared in 1000 separate matrix calculations. Using treecon, 100 trials were done using a random seed no. of 105. A branch point with a bootstrap value of >80% was considered reliable. Lower values indicate that the sequences separated by a branch could, in fact, form part of the same group.

Figures 5.8 and 5.9 show the trees constructed in TREECON from the DNA sequences. Monkeypox virus Z1324 is identical to the other two Zaire strains, 86-112 and 86-21. It was omitted due to the lack of memory space for a 16th sequence. In fig. 5.8 insertions and deletions have been taken into account, whereas fig. 5.9 shows the tree produced not taking them into account.

In both figures the monkeypox viruses form two groups (Zaire and West Africa) well isolated from each other as well as the other viruses. But the base similarity between the two groups is >99%. This anomaly can be explained in terms of the deletions present in the monkeypox viruses. Within the DNA sequence comprising the major deletion in the West African isolates, the sequence in the Zaire isolates has a number of differences compared

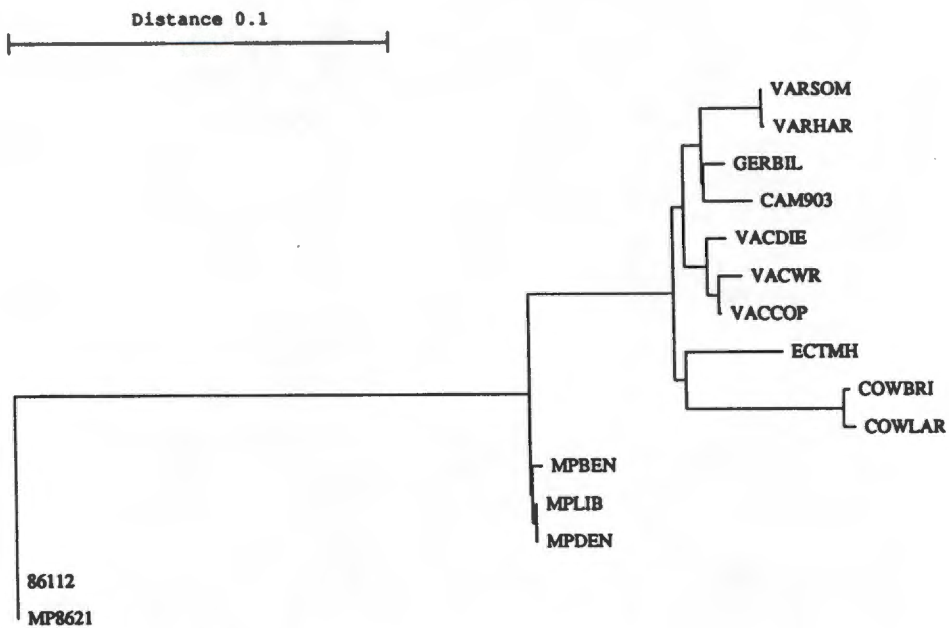


Fig. 5.8 Phylogenetic tree constructed from the DNA sequences aligned in fig. 5.5 using TREECON (Van de Peer and Wachter, 1992), taking into account insertions and deletions.

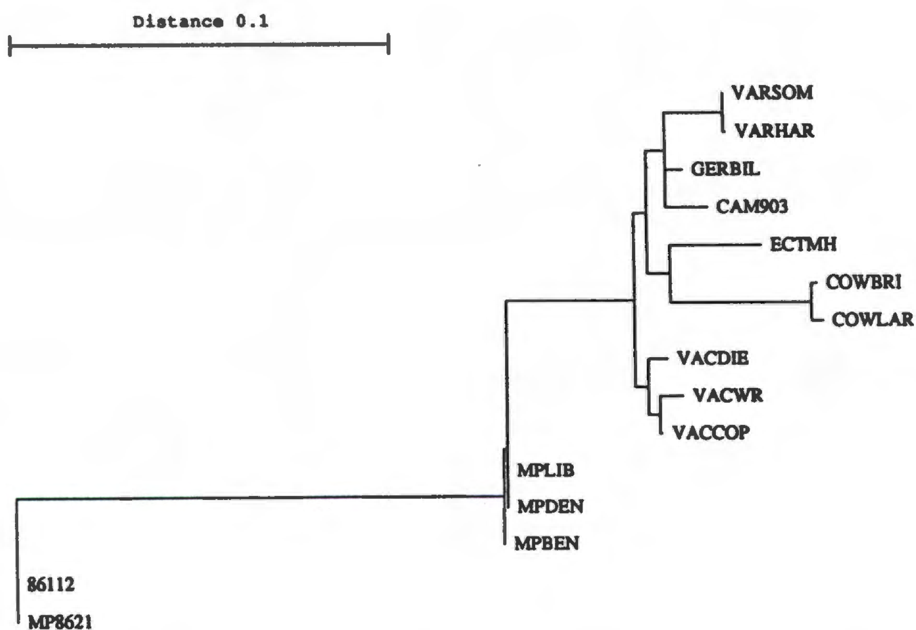


Fig. 5.9 Phylogenetic tree constructed from the DNA sequences aligned in fig. 5.5 using TREECON (Van de Peer and Wachter, 1992), ignoring insertions and deletions.

to the other viruses. It is not known whether these differences were common to the West African isolates before the deletion occurred. Comparing this region of the monkeypox viruses to the other *Orthopoxviruses*, only one difference is recognized between the West African isolates and the other *Orthopoxviruses* (one large deletion), whereas there are a number of differences between the Zaire isolates and the other *Orthopoxviruses* due to base substitutions (and smaller deletions and insertions). These differences place the Zaire monkeypox viruses far apart from the other *Orthopoxviruses*, whereas the West African isolates appear more similar to the other *Orthopoxviruses*. The large number of differences in the Zaire isolates may reflect an increase in rate of divergence which would have occurred once the E5R-equivalent ORF became nonessential for the virus. The branch lengths (divergence) of the monkeypox viruses is probably not on the same scale as the rest of the tree. Consideration must be given to the fact that this is a single gene/pseudogene comparison where the ORF is more conserved amongst the other *Orthopoxviruses* than it is in monkeypox virus. The relationship of monkeypox viruses to the other viruses may therefore be different when comparing either a gene conserved in all the viruses, or the DNA from a noncoding region which is not subject to functional constraints.

The interrelationship of the other *Orthopoxviruses* as shown in fig. 5.8 is probably fairly accurate. The two variola viruses, three vaccinia viruses and two cowpox viruses form three individual groups. The variola and cowpox virus groups were formed in all bootstrap trials and all three vaccinia viruses formed a single group in 98% of the bootstraps. Gerbilpox and camelpox viruses group reliably with the variola viruses (81% TREECON, 99.3% clustal). The branch separating the vaccinia viruses from the group comprising variola, gerbilpox and camelpox viruses has a low bootstrap value (55%). This indicates that the vaccinia viruses may actually form part of the variola/gerbilpox/camelpox virus group. Interestingly, ectromelia virus shares a branch with the cowpox viruses. These three viruses were grouped together in 92% (clustal) and 87% (TREECON) of the bootstrap trials.

Fig. 5.10 shows an unrooted tree drawn to scale from the "clustal analysis" data taking into account indels. The branching pattern of the tree resembles that of fig. 5.8. This tree shows more clearly that each virus species is more closely related to gerbilpox virus than to any other species.

Fig. 5.9 shows the tree produced when indels are not taken into account. In this case a smaller number of nucleotides are compared, as each position in which there is a deletion or insertion in any one virus is ignored. Overall, the branches are shorter, due to the recognition of fewer differences. The differences would be proportional to the number of bases compared as well as the number of indels. The grouping of cowpox and ectromelia viruses with variola, gerbilpox and camelpox viruses shows that a number of differences common to cowpox and ectromelia viruses are found in regions which are deleted in at least one of the other *Orthopoxviruses*. (There are a number of nucleotide changes common to ectromelia and cowpox viruses within the large stretch of DNA which is deleted in

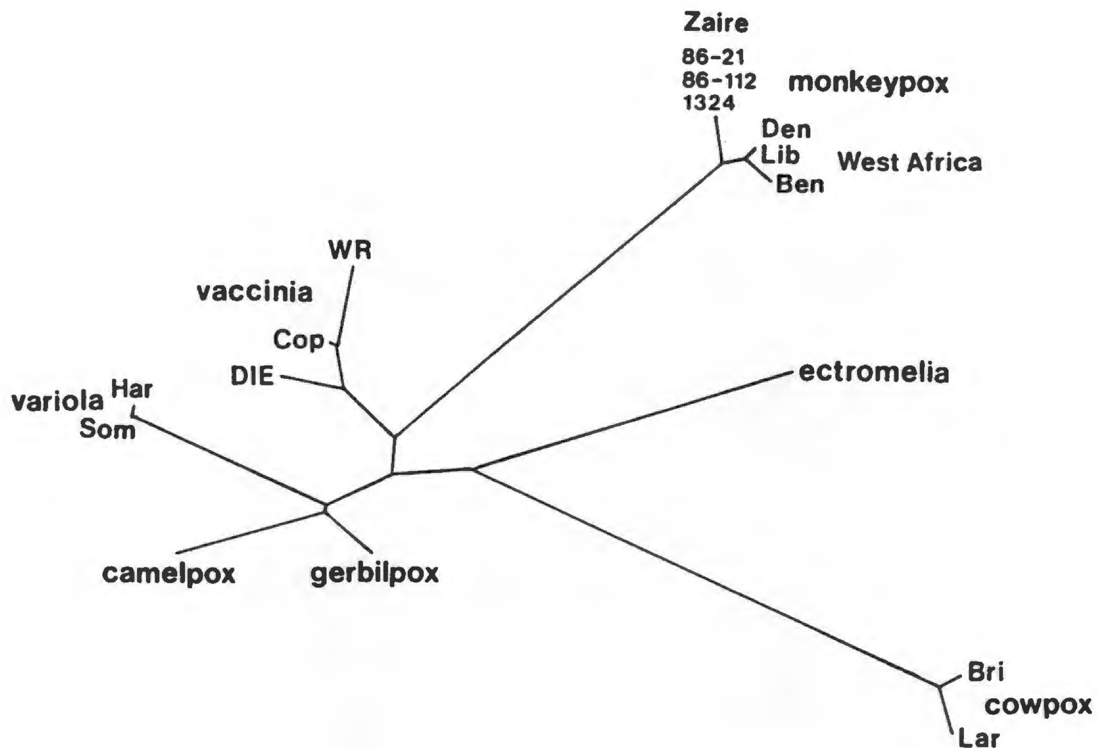


Fig. 5.10 Unrooted tree constructed from DNA sequences aligned in fig. 5.5 using clustal V (Higgins *et al.*, 1991). Insertions and deletions were taken into account.

monkeypox virus). The grouping of cowpox virus with ectromelia virus is associated with a low bootstrap value (49.1%) in this tree, implying that ectromelia virus may be on the same branch as the viruses in the variola/vaccinia group.

Phylogenetic trees show the divergence of DNA sequences as they currently exist. They could only be interpreted as "evolutionary" trees if the rates of mutation were constant overall and there were no differential selection for changes at different positions. Clearly, neither of these is realistic and the trees can only be used to suggest evolutionary relation. The relatedness of monkeypox virus to the other *Orthopoxviruses* should be tested by comparison of a DNA sequence subject to equivalent rates of mutation in the different viral species.

The relationship suggested here between gerbilpox, variola and camelpox viruses is in line with the biological properties of the viruses (Lourie *et al.*, 1975; Baxby, 1972).

Because all the species examined showed the greatest similarity to gerbilpox virus, we speculate that gerbilpox virus may be the closest species to the ancestral *Orthopoxvirus*. However, it has a narrow host range, implying adaptation to a particular host animal after divergence from an ancestral virus (which probably had a wider host range).

Although this tree (fig. 5.10) is the result of a single gene comparison, the tree for at least one other gene, the haemagglutinin gene, is similar (J. Esposito, personal communication to K. Dumbell).

Comparison of an intergenic DNA sequence in different *Orthopoxviruses*

The DNA sequence equivalent to the noncoding region between E5R and E6R of vaccinia virus was determined for all the isolates mentioned above as well as the Z241 strain of monkeypox virus and raccoon poxvirus. Two comparisons were made. The first comprises the sequences of all strains of vaccinia, variola and monkeypox viruses sequenced; and the second is a comparison of one strain each of vaccinia, variola and monkeypox (West Africa + Zaire) viruses with camelpox, taterapox and ectromelia viruses, two strains of cowpox virus and raccoon poxvirus.

The vaccinia/variola/monkeypox virus comparison is shown in fig. 5.11. The sequence starts at the E5R-equivalent termination codon and ends at the first methionine codon of vaccinia virus E6R. These two triplets will be subject to functional constraints.

Where all sequences are aligned there are four positions of base mismatch (positions 3, 9, 106 and 139). Position 3, part of the E5R termination codon, differs in monkeypox virus, which has a degenerate form of the ORF and does not require this termination codon. Of the other three base substitutions, monkeypox virus differs from variola and vaccinia viruses in two positions (9 and 106), and variola virus differs from monkeypox and vaccinia viruses at position 139.

More significant than the base substitutions is the presence of insertions or deletions, which occur in five regions of the short (152bp) stretch sequenced. These correspond to positions 53-54, 75-101, 106-109, 127-134 and 145-146 in fig. 5.11. These indels are shown in the diagram as deletions. The profile of these deletions differentiates the viruses into four main groups - vaccinia virus, variola virus, monkeypox virus West Africa and monkeypox virus Zaire.

Two features are characteristic of variola virus - the presence of a 27bp deletion from positions 75 to 101, and the absence of a 7bp deletion from positions 127 to 133.

The sequence at positions 53-54 shows variation between vaccinia, variola and monkeypox viruses. Where variola virus has two nucleotides of AT, all the monkeypox viruses have the T deleted and all three vaccinia viruses have both A and T deleted.

Monkeypox virus Zaire can be differentiated from monkeypox virus West Africa by the 3bp deletion at positions 107-109. This is the only deletion associated with a direct repeat. Within the triplet deleted in the Zaire strains there is a base mismatch between monkeypox virus West Africa and the vaccinia and variola viruses. This substitution probably occurrerd after separation of monkeypox virus from the common ancestor, but before divergence into the Zairean and West African lineages. After this split the deletion (facilitated by the direct repeat), occurred in the Zaire group of monkeypox viruses.

```
           E5R
VACDIE     TAGGCATTTCATCTTTCTCCAATACTAATTCAAATTGTTAAATTAATAATGG--ATAGTA    60
VACCOP     TAGGCATTTCATCTTTCTCCAATACTAATTCAAATTGTTAAATTAATAATGG--ATAGTA
VACWR      TAGGCATTTCATCTTTCTCCAATACTAATTCAAATTGTTAAATTAATAATGG--ATAGTA

VARHAR     TAGGCATTTCATCTTTCTCCAATACTAATTCAAATTGTTAAATTAATAATGGATATAGTA
VARSOM     TAGGCATTTCATCTTTCTCCAATACTAATTCAAATTGTTAAATTAATAATGGATATAGTA

MPDEN      TAGGCATTCCATCTTTCTCCAATACTAATTCAAATTGTTAAATTAATAATGGA-ATAGTA
MPLIB      TAGGCATTCCATCTTTCTCCAATACTAATTCAAATTGTTAAATTAATAATGGA-ATAGTA
MPBEN      TAGGCATTCCATCTTTCTCCAATACTAATTCAAATTGTTAAATTAATAATGGA-ATAGTA

MP86-21    TATGCATTCCATCTTTCTCCAATACTAATTCAAATTGTTAAATTAATAATGGA-ATAGTA
MP86-112   TATGCATTCCATCTTTCTCCAATACTAATTCAAATTGTTAAATTAATAATGGA-ATAGTA
MP1324     TATGCATTCCATCTTTCTCCAATACTAATTCAAATTGTTAAATTAATAATGGA-ATAGTA
MPZ241     TATGCATTCCATCTTTCTCCAATACTAATTCAAATTGTTAAATTAATAATGGA-ATAGTA
           ** ***** *******************************************  ******


VACDIE     TAAATAGTTATTAGTTATAAGATAGTAAAAATAATTATTAGAATAAGAGTGTAGTATCAT    120
VACCOP     TAAATAGTTATTAGTGATAAAATAGTAAAAATAATTATTAGAATAAGAGTGTAGTATCAT
VACWR      TAAATAGTTATTAGTGATAAAATAGTAAAAATAATTATTAGAATAAGAGTGTAGTATCAT

VARHAR     TAAATAGTTATTAG---------------------------AATAAGAGTGTAGTATCAT
VARSOM     TAAATAGTTATTAG---------------------------AATAAGAGTGTAGTATCAT

MPDEN      TAAATAGTTATTAGTGATAAGATAGTAAAAATAATTATTAGAATAGTAGTGTAGTATCAT
MPLIB      TAAATAGTTATTAGTGATAAGATAGTAAAAATAATTATTAGAATAGTAGTGTAGTATCAT
MPBEN      TAAATAGTTATTAGTGATAAGATAGTAAACATAATTATTAGAATAGTAGTGTAGTATCAT

MP86-21    TAAATAGTTATTAGTGATAAGATAGTAAAAATAATTATTAGAATAG---TGTAGTATCAT
MP86-112   TAAATAGTTATTAGTGATAAGATAGTAAAAATAATTATTAGAATAG---TGTAGTATCAT
MP1324     TAAATAGTTATTAGTGATAAGATAGTAAAAATAATTATTAGAATAG---TGTAGTATCAT
MPZ241     TAAATAGTTATTAGTGATAAGATAGTAAAAATAATTATTAGAATAG---TGTAGTATCAT
           ***************                          ****   ************

                                     E6R
VACDIE     AGATAA-------CTCTCTTCTAT-AAAAATG                                152
VACCOP     AGATAA-------CTCTCTTCTAT-AAAAATG
VACWR      AGATAA-------CTCTCTTCTAT-AAAAATG

VARHAR     AGATAAGTATTTACTTTCTTCTAT--AAAATG
VARSOM     AGATAAGTATTTACTTTCTTCTAT--AAAATG

MPDEN      AGATAA-------CTCTCTTCTATAAAAAATG
MPLIB      AGATAA-------CTCTCTTCTATAAAAAATG
MPBEN      AGATAA-------CTCTCTTCTATAAAAAATG

MP86-21    AGATAA-------CTCTCTTCTATAAAAAATG
MP86-112   AGATAA-------CTCTCTTCTATAAAAAATG
MP1324     AGATAA-------CTCTCTTCTAT-AAAAATG
MPZ241     AGATAA-------CTCTCTTCTATAAAAAATG
           ******       ** *********  ******
```

Fig. 5.11 Alignment of the intergenic sequence between vaccinia virus E5R and E6R with the corresponding sequences of variola (2 strains) and monkeypox (7 strains) viruses.

Within the 27bp sequence deleted in variola virus intraspecies variation is evident in the other viruses. There are two positions (76 and 81) in which Dairen differs from the Copenhagen and WR strains of vaccinia virus. This is the only difference found within the vaccinia viruses. Since the sequence is deleted in variola virus, it cannot be determined whether this change places Dairen closer to variola virus or not. At position 90 there is an A-C transition which distinguishes monkeypox virus Benin from the Liberia and Denmark strains.

The fifth region of variation is immediately upstream of the vaccinia virus E6R ORF. The differences lie in the number of adenine reidues preceding the first ATG codon. Variola viruses have a run of 4 A's, vaccinia viruses have 5 A's and most monkeypox viruses have 6 A's. Monkeypox virus 1324 has 5 and not 6 A's. This isolate is from Gemena, which is in the north-western part of Zaire. It may have lost one A independently and should not be grouped with vaccinia viruses since the rest of the DNA sequence is identical to the other Zaire isolates.

Z241 is from the southern part of the rain forest region of Zaire (Katako-Kombe), geographically more isolated from the others. Its DNA sequence is identical to the the Zaire isolates 86-21 and 86-112.

The run of A's may have been expanded or reduced by DNA slippage during replication. It is not known whether E6R is expressed early or late in infection, but the upstream DNA sequence suggests that it is an early gene. In variola virus, TAAAAT may be a weak late promoter (Davison and Moss, 1989b). If further slippage reduces the run of A's to 3, a (stronger) late promoter will be generated (TAAAT). The conversion of an early gene to a late gene may be an additional means of viral evolution where one small genetic change could result in a major alteration to the viral phenotype.

The second intergenic comparison includes raccoon poxvirus. The large number of differences between this virus and the other *Orthopoxviruses* makes a detailed comparison impossible as the relative positioning of the raccoon poxvirus nucleotides in the DNA sequence alignment of fig. 5.12 is probably not accurate. It is clear that many base substitutions, insertions and deletions have taken place since raccoon poxvirus diverged from the common *Orthopoxvirus* progenitor. Raccoon poxvirus has been shown to differ from the other *Orthopoxviruses* by restriction enzyme analysis (Esposito and Knight, 1985) cross-hybridization (Cavallaro and Esposito, 1992) and DNA sequence of the hemagglutinin gene (Cavallaro and Esposito, 1992).

The asterisks above the alignment of fig. 5.12 denote nucleotide identity of all viruses except for raccoon poxvirus. The alignment can be divided into three segments, two conserved regions from positions 1-92 and 228-312, separated by a variable region from positions 93-227. The five additional sequences (omitting raccoon poxvirus) will first be compared to those of vaccinia, variola and monkeypox viruses for the conserved regions; and the variable region will be discussed separately afterwards.

```
           **  **** ******** ***        ******** ***** *******************
VACDIE     TAGGCATTTCATCTTTCTCC-------AATACTAATTCAAATTGTTAAATTAATAATGG-    60
VARHAR     TAGGCATTTCATCTTTCTCC-------AATACTAATTCAAATTGTTAAATTAATAATGGA
MPDEN      TAGGCATTCCATCTTTCTCC-------AATACTAATTCAAATTGTTAAATTAATAATGGA
MP86-21    TATGCATTCCATCTTTCTCC-------AATACTAATTCAAATTGTTAAATTAATAATGGA
CAM903     TAGGCATTTCATCTTTCTCC-------AATACTAATTCAAATTGTTAAATTAATAATGG-
GERBIL     TAGGCATTTCATCTTTCTCC-------AATACTAATTCAAATTGTTAAATTAATAATGG-
ECTMH      TAGACATT-CATCTTTCTCCTTTCTCCAATACTAATTCAAATTGTTAAATTAATAATGG-
COWLAR     TAGACATTTCATCTTT-TCCTTTCTC-AATACTA-TTCAACTTGTTAAATTAATAATGG-
COWBRI     TAGGCATTTCATCTTT-TCCTTTCTC-AATACTA-TTCAACTTGTTAAATTAATAATGG-
RACCOON    TAGA--TT-GATTTTT--ACTTT----AGT-TTAGTAAGTTAAGTTTTCATTTCATTGCA

            ***** **** ***********
VACDIE     -ATAGTATAAA-TAGTTATTAGTT---ATAAG----------------------------    120
VARHAR     TATAGTATAAA-TAGTTATTAG--------------------------------------
MPDEN      -ATAGTATAAA-TAGTTATTAGTG---ATAAG----------------------------
MP86-21    -ATAGTATAAA-TAGTTATTAGTG---ATAAG----------------------------
CAM903     -ATAGTATAAA-TAGTTATTAGTG---ATAAGATAAG-----------------------
GERBIL     -ATAGTATAAA-TAGTTATTAGTG---ATAAGATAAAATAAGATAAGATAAGATAAGATA
ECTMH      -ATAGTATAAAATAGTTATTAGTA---GTAAGATAAGATAAGATAAGATAAGATAAGATA
COWLAR     -ATAGTTTAAAATAGTTATTAGTA---ATAAGATAGTTTAAAATAGTTATTAGTA---AT
COWBRI     -ATAGTTTAAAATAGTTATTAGTAGTAATAAGATAGTTTAAAATAGTTATTAGTAGTAAT
RACCOON    ATTATTTTAA--TTGTTAAAT-CA---ATAA--TAGTATAA-ATAGTCA-TAGAATATAG

VACDIE     ------------------------------------------------------------    180
VARHAR     ------------------------------------------------------------
MPDEN      ------------------------------------------------------------
MP86-21    ------------------------------------------------------------
CAM903     ------------------------------------------------------------
GERBIL     AGATAAGATAAGATAAGATAAG--------------------------------------
ECTMH      AGATAAGATAAGATAAGATAAGATAAGATAAGATAAGATAAGATAAGATAAGATAAG---
COWLAR     AAGATAGTTTAAAATAGTTATTAGTAATAAGATAGTTTAAAATAGTTATTAGTAATAAGA
COWBRI     AAG---------------------------------------------------------
RACCOON    AATATAGAATATAGAATATAGAATATAGAATATAGAATATAGAATATAGAATATAGAATA

VACDIE     -----------------------------------------------ATAGTA-AAAATA    240
VARHAR     ------------------------------------------------------------
MPDEN      -----------------------------------------------ATAGTA-AAAATA
MP86-21    -----------------------------------------------ATAGTA-AAAATA
CAM903     -----------------------------------------------ATAGTA-AAAATA
GERBIL     -----------------------------------------------ATAGTA-AAAATA
ECTMH      -----------------------------------------------ATAGTA-AAAATA
COWLAR     TAGTTTAAAATAGTTATAGATAGTTTAAAATAGTTATTAGTAATAAGATAGTTTAAAATA
COWBRI     -----------------------------------------------ATAGTTTAAAATA
RACCOON    TAGAATATAGAATATAGAATATAGGATAAATA--TA--AGGA-TA-TA-AGTTGTTTATA

                       ****      *          ******  *********       ** ****
VACDIE     ATTATTAG----AATAAGAGTG-------TAGTATC--ATAGATAA-------CTCTCTT    300
VARHAR     ------------AATAAGAGTG-------TAGTATC--ATAGATAAGTATTTACTTTCTT
MPDEN      ATTATTAG----AATAGTAGTG-------TAGTATC--ATAGATAA-------CTCTCTT
MP86-21    ATTATTAG----AATAG---TG-------TAGTATC--ATAGATAA-------CTCTCTT
CAM903     ATTAGTAG----AATAAGAGTG-------TAGTATC--ATAGATAAGTATTTACTCTCTT
GERBIL     ATTATTAG----AATAAGAGTG-------TAGTATCC-ATAGATAAGTATTTACTCTCTT
ECTMH      ATTATTAG----AATAAGAGTG-------TAGTATC--ATAGATAAGTATTTACTCTCTT
COWLAR     GTTAT-AG---TAATAAGATAGTTTAAAATAGTATAAAATAGATAAGTATTTACTCTCTT
COWBRI     GTTATTAGTAGTAATAAGATAGTTTAAAATAGTATAAAATAGATAAGTATTTACTCTCTT
RACCOON    --TGTT--CA-TATTAATA-A-TTTG-CCT--TTTC---TAACTC-GC-TAT-CTAAATT

           ****  ******
VACDIE     CTAT-AAAAATG                                                    312
VARHAR     CTAT--AAAATG
MPDEN      CTATAAAAAATG
MP86-21    CTATAAAAAATG
CAM903     CTAT-AAAAATG
GERBIL     CTATAAAAAATG
ECTMH      CTAT-AAAAATG
COWLAR     CTAT-AAAAATG
COWBRI     CTAT-AAAAATG
RACCOON    -TAC-AAAAATG
```

Fig. 5.12 Alignment of *Orthopoxvirus* DNA sequences corresponding to the intergenic region between vaccinia virus E5R and E6R. The termination codon of E5R is doubly overlined and the first methionine codon of E6R is denoted by an arrow. Asterisks indicate positions in which the nucleotide is identical for all viruses except raccoon poxvirus. Overlined and underlined sequences correspond to repeated elements.

Within 135 base pairs which can be aligned in the first 9 *Orthopoxviruses* compared in fig. 5.12, there are 6 positions in which there is a difference in at least one virus (95.6% base similarity). There are an additional 6 positions of dissimilarity amongst 8 viruses where a deletion is present in one of the viruses, giving a 91.5% (12/141) level of similarity. Most of these differences occur in a single species, one each in camelpox (position 245) and ectromelia (position 88), and 4 each in monkeypox (positions 3, 9, 257 and 258), and cowpox (positions 41, 260, 261 and 276) viruses. There is one position (4) where cow and ectromelia viruses share a difference. Position 84 (part of a deletion in variola virus) divides the viruses into three groups, vaccinia virus, ectromelia + cowpox viruses, and monkeypox + camelpox + gerbilpox viruses.

Each virus has a unique pattern of deletions (and/or insertions).

Camelpox virus, like vaccinia virus, has the 2bp deletion at positions 60-61 and a run of 5 A's upstream of the E6R-equivalent ORF. But, like variola virus, the 7bp sequence from positions 287-293 is present.

Gerbilpox virus most closely resembles camelpox virus, having the same features which are common to both vaccinia and variola viruses. In addition it has a unique C inserted at position 277. Like monkeypox virus, it has 6 A's upstream of the E6R ORF.

Ectromelia virus also resembles camelpox virus, but has a single base pair deletion at position 9 and an additional 7bp sequence from positions 21 to 27. Interestingly, 6 of these 7 bp are shared with both of the cowpox viruses. The cowpox viruses, however, differ at a number of other positions. Like camel, gerbil and ectromelia viruses, the cowpox viruses have the 2bp deletion from positions 60-61 (as for vaccinia virus) and the 7bp sequence from positions 287-293 (as for variola virus). There are 6 additional differences - 2 single base deletions at positions 17 and 35, 2 single base insertions at positions 234 and 252, an additional 7 bp from positions 263 to 269, and an additional 2 bp at positions 277-78. Cowpox virus also shows interspecies variation at three positions. Brighton has two 3bp insertions at positions 85-87 and 249-251; and Larkin has a single base deletion at position 246.

The overall picture shows that each species has a distinctive sequence, with some intraspecies variation. Gerbilpox virus can be regarded as having a prototypic sequence from which the other viruses diverged. The 2bp deletion at positions 60-61 is probably not a deletion, but rather an insertion in variola virus, since most viruses lack the sequence and it could easily have arisen by duplication in variola virus.

The differences divide the viruses into two main groups, cowpox viruses and "others". Monkeypox viruses are classified as part of the "others" group, as they are not sufficiently different to form a group of their own. This may be a better reflection on the positioning of monkeypox virus with respect to the other *Orthopoxviruses* than that determined by the E5R-equivalent comparison, since there is no bias with respect to maintaining an ORF.

Although ectromelia virus is grouped with the "others", it has a greater resemblance

than any of the other viruses to cowpox virus (shares 2 bp changes and 1 additional sequence of 6/7bp).

Raccoon poxvirus forms a group of its own, which is considerably different from both the cowpox viruses and "others".

The most fascinating feature of this intergenic region is the variable central segment. The variation occurs in camelpox, gerbilpox and ectromelia viruses, the two cowpox viruses and raccoon poxvirus in the form of repeated sequences. Camelpox, gerbilpox and ectromelia viruses have the same pentameric sequence - ATAAG - repeated 1, 9, and 16 times respectively. Gerbilpox and ectromelia viruses may have had an additional unit each in which a single base substitution event has occurred. The ectromelia intergenic sequence was compared to other known sequences in GenBank by a blast search (Altschul et al., 1990). An identical repeat unit is present in the sequence flanking one of the PR1 sequences (polymorphic repetitive sequence 1) of mice (Kominami *et al.*, 1983). This sequence is found in the nontranscribed spacer region downstream of the 3' end of the 28S RNA gene. Although the PR1 sequence is a moderately repetitive DNA sequence, the short repeated sequences which flank it are not the same for different PR1 elements. These short repeats are thought to be involved in the generation and dispersion of PR1 sequences (Kominami *et al.*, 1983).

Raccoon poxvirus has a 7bp sequence - TAGAATA - in thirteen adjacent copies. A blast search revealed no significant similarity of this sequence to other known sequences.

It is possible that these short repeats represent a transcription control sequence unique to certain poxviruses. They may be advantageous to the virus in specific cell types or host animals. Since they are not multiples of three (5bp in camelpox, gerbilpox and ectromelia viruses, and 7bp in raccoon poxvirus), they would not represent a relic of DNA sequences coding for a repeated amino acid sequence. A hexanucleotide sequence repeated nine times has been detected in the coding region of vaccinia virus Copenhagen B11R (Smith *et al.*, 1991). This unit is present only as a single copy in the equivalent WR gene.

They are unlikely to have arisen by laboratory selection (in cell culture or on CAMs), since the repeated sequences are not present in the vaccinia viruses, which have been passaged the most in laboratories. Although etromelia virus has only been isolated from laboratory mice, gerbilpox virus, which has the same repeated element, was isolated from a wild gerbil in Benin and has not been passaged frequently, and has not been passaged in laboratory mice at all. Similarly, camelpox virus has had very little laboratory passaging. Could these repeated sequences represent a common link between camelpox, gerbilpox and ectromelia viruses? It is conceivable that this pentameric repeat was present in a common ancestor, and gradually became deleted in certain species. This leads one to speculate on the host animal of the common ancestral virus. Unless the progenitor had some means of maintaining a period of latency, which has now been lost, the ancestral virus would require a host animal with a high density and rapid reproduction rate. These criteria are met by rodents.

The function of these repeats is not known. Since they are found both in coding and noncoding regions of the genome, they may have arisen by stuttering of the DNA polymerase during DNA replication and have no particular function. They may be equivalent to selfish DNA (Doolittle and Sapienza, 1980; Orgel and Crick, 1980). The potential to amplify and delete DNA continuously provides a means of species divergence.

The variable segment of the two cowpox viruses consists of a more complicated array of repeats. Larkin has a 28bp sequence represented 5 times (underlined). Four of these are tandemly arranged and the fifth sequence is separated by a stretch of 20bp. This 20bp sequence (dotted underline) is repeated immediately after the 28bp sequence. A second 20bp sequence adjacent to the former 20bp repeated sequence is a repeat of the sequence overlapping the fifth 28bp repeat and the first part of the second 20bp sequence (overlined).

By comparison, the Brighton strain of cowpox virus has three copies of the equivalent 28bp sequence, but with the final 3bp duplicated to make the repeated sequence 31bp in length. The first 20bp repeat present in Larkin is not present, but one copy of the second 20bp repeat occurs in the position corresponding to the final copy in Larkin (overlined).

This pattern of repeated sequences probably arose by multiple insertions and/or deletions. Larkin has only been passaged four times on CAMs, whereas Brighton has been passaged three times in guinea pig, 17 times in rabbit i.v., 13 times in rabbit by scarification and 35 times on CAMs (K.R.Dumbell, personal communication). The sequences indicate that the repeats were present in the original virus in its natural host, and, on subsequent passages in the laboratory, were lost. The simplest explanation is that, in the laboratory, selection would favour the virus which can be most efficiently replicated, which would be the one with the least amount of redundant DNA. If the repeated sequences have no function (under laboratory conditions), they may readily be deleted.

Alternatively, the sequences may influence expression of the E6R-equivalent ORF. They do not resemble known enhancer sequences or other known transcription factor binding sites, but they may represent the binding site of an as yet undetermined host or poxvirus product which influences gene regulation. The effect would be ancillary and not essential, since the sequences are not conserved amongst the *Orthopoxviruses*.

The intergenic DNA comparison shows the interrelationships amongst the viruses to parallel that determined by analysis of the E5R-equivalent region. Within species there is conservation of sequence. However, the two cowpox viruses are readily distinguished from one another, and the Zaire monkeypox viruses differ from the West African isolates by having a deletion of 3bp.

Gerbilpox and camelpox viruses are closely related to one another. The variola, vaccinia and monkeypox viruses, although distinct, form a group similar to gerbilpox and camelpox viruses, and different from the cowpox viruses. Ectromelia virus closely resembles gerbilpox virus, especially with respect to the pentameric repeats, but shares some features with cowpox virus.

The intergenic DNA sequencing data is consistent with the tree drawn in fig. 5.10 for all virus species except monkeypox virus, which is more closely related to variola and vaccinia viruses here. Since there are fewer functional constraints within this region, this sequence probably gives a more realistic representation of the relationship of monkeypox virus to the other *Orthopoxviruses.*

CHAPTER 6

CHARACTERIZATION OF VACCINIA VIRUS E5R

Summary 86

Introduction 86

Results and discussion 87

- **Interruption of vaccinia virus E5R** 87
- **RNA expression** 90
- **Role in pathogenicity** 91
- **Vaccinia virus E5R putative gene product** 93
- **Putative products of the E5R – equivalent sequence in other *Orthopox viruses*** 95

CHAPTER 6

CHARACTERIZATION OF VACCINIA VIRUS E5R

SUMMARY

The E5R gene of vaccinia virus was interrupted by replacement of 438bp near the beginning with the marker gene, gpt. The recombinant virus was viable in cell culture and on CAMs. This shows that the ORF is nonessential and may serve as a site of insertion of foreign genes. E5R-specific RNA was detected early, and not late in the life cycle of the virus. The function of the gene product is not known, but preliminary experiments suggest that it may have a role to play in viral pathogenicity. Putative structural features of the deduced amino acid sequence reveals a probable signal sequence, indicating that the protein may be secreted. Alternatively, translation may start further downstream, in which case the protein would function intracellularly. The putative gene product is highly positively charged.

INTRODUCTION

Since the function of the protein encoded by vaccinia virus E5R is not known, we can only speculate on its role in the life cycle of the virus. Because the equivalent ORF in monkeypox viruses is degenerate, this ORF was thought to be nonessential for basic viral replication, and a deletion in vaccinia virus should be non-lethal. Identification of a nonessential gene could be applicable to the use of vaccinia virus as a vector for foreign gene expression in terms of providing a site of insertion for the gene of interest. The E5R ORF is well situated for this purpose as it lies within the conserved part of the genome.

A virus construct with an interrupted E5R ORF would be of additional value as a tool for investigations into the expression and function of the ORF.

Construction of the recombinant vaccinia virus involved deletion of 438bp near the beginning of the ORF and replacement of this sequence with the bacterial gpt (xanthine-guanine phosphoribosyl transferase) gene, which served as a marker for selection of the recombinant virus (Boyle and Coupar, 1988a; Falkner and Moss, 1988).

RNA was extracted from cells infected with the parent and recombinant viruses to test for transcription of the E5R ORF and to determine whether it is expressed early and/or late in the virus life cycle. Cycloheximide, an inhibitor of translation, was added to the medium of infected cells to test for immediate early expression of RNA.

Preliminary pathogenicity experiments were done by injecting rabbits intradermally with the parent and recombinant viruses. Differences were noted in the lesions produced by the different viruses.

Structural features of the putative polypeptide were examined and related to possible functional significance.

RESULTS AND DISCUSSION

Interruption of vaccinia virus E5R

A portion (438bp) of the E5R ORF was deleted and replaced with the bacterial marker gene, xanthine-guanine phosphoribosyl transferase, (gpt) (fig. 6.1). A plasmid was constructed as shown in fig. 6.2 with the gpt gene inserted into the E5R ORF replacing the DNA sequence from positions 325-763 (fig 5.2). Methodological details are described in chapter 2.

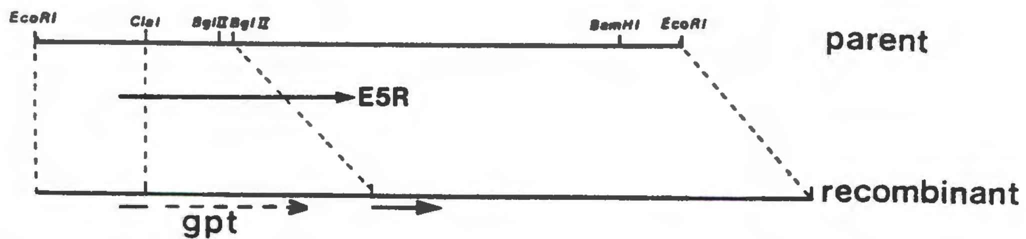


Fig. 6.1 Diagram showing the partial deletion of the vaccinia virus E5R gene. DNA between the *ClaI* and second *BglII* site was replaced with the bacterial gpt gene.

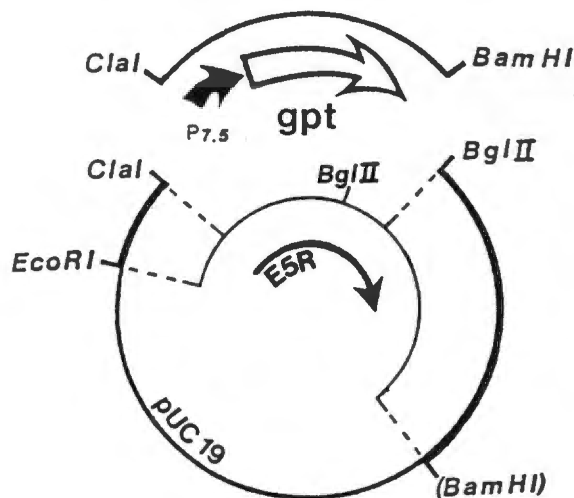


Fig. 6.2 pGptC/B,The plasmid used for the replacement of part of the E5R ORF with gpt.The construction of this plasmid is described in chapter 2 (fig. 2.3).

A recombinant virus was isolated by passage in the presence of 1µg/ml mycophenolic acid and 250µg/ml xanthine (selective for gpt activity) and pock-purified. This virus formed plaques in CV1 cells under conditions in which the parent virus did not (fig. 6.3). The presence of guanine overcame the inhibition, indicating that gpt was bypassing a block in purine metabolism.

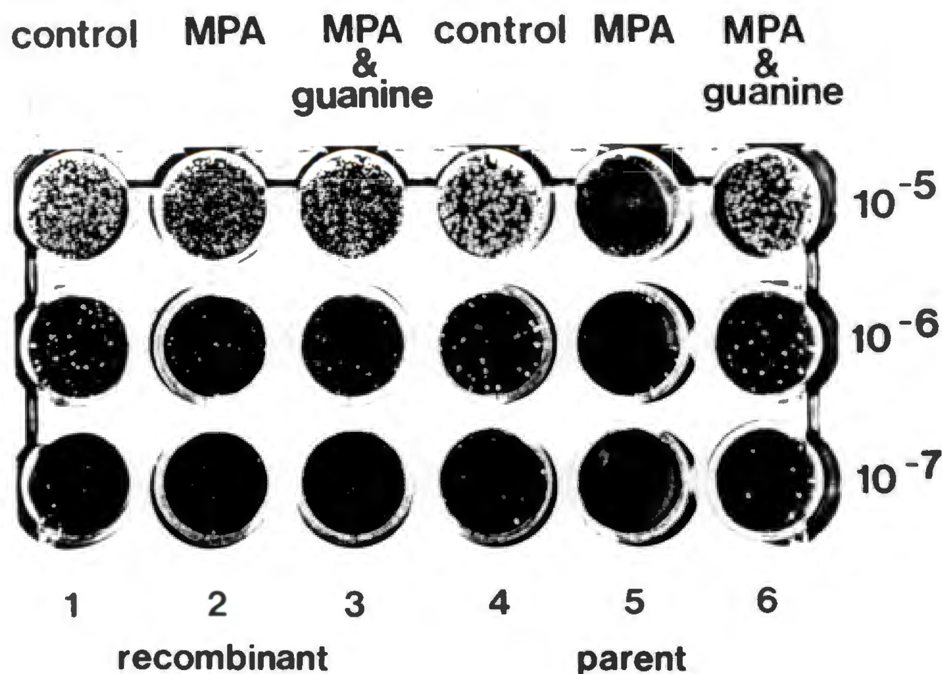


Fig. 6.3 Titration of the recombinant (lanes 1, 2 and 3) and parent (lanes 4, 5 and 6) vaccinia viruses on CV1 cells. In lanes 1 and 4 the cells were incubated in the absence of selection medium; in lanes 2 and 5 selection medium (1µg/ml mycophenolic acid and 250µg/ml xanthine) was added; and in lanes 3 and 6 selection medium + 25µg/ml guanine was added. Virus dilutions are indicated on the right.

DNA was extracted from the recombinant virus and digested with *HindIII.* The digested fragments were separated by agarose gel electrophoresis and compared to *HindIII*-digested parent vaccinia virus (fig. 6.4). A change of mobility was detected in the *HindIII* E band from 15.2kb in the parent virus to ±16.8kb in the recombinant virus (larger than the *HindIII* D fragment). Insertion of a 2kb fragment (gpt) and deletion of a 0.4kb fragment (part of E5R) would result in a nett increase of 1.6kb.

The DNA was transferred to a nylon membrane and probed with: 1) the gpt gene and 2) a DNA fragment corresponding to the portion of E5R deleted in the recombinant virus (fig. 6.4 A and B respectively). "A" shows that the gpt gene is integrated into the poxvirus genome in the correct position, and "B" shows that a portion of the E5R gene has been deleted in the recombinant virus, ie. selection is not due to integration of the entire plasmid leaving an intact E5R gene.

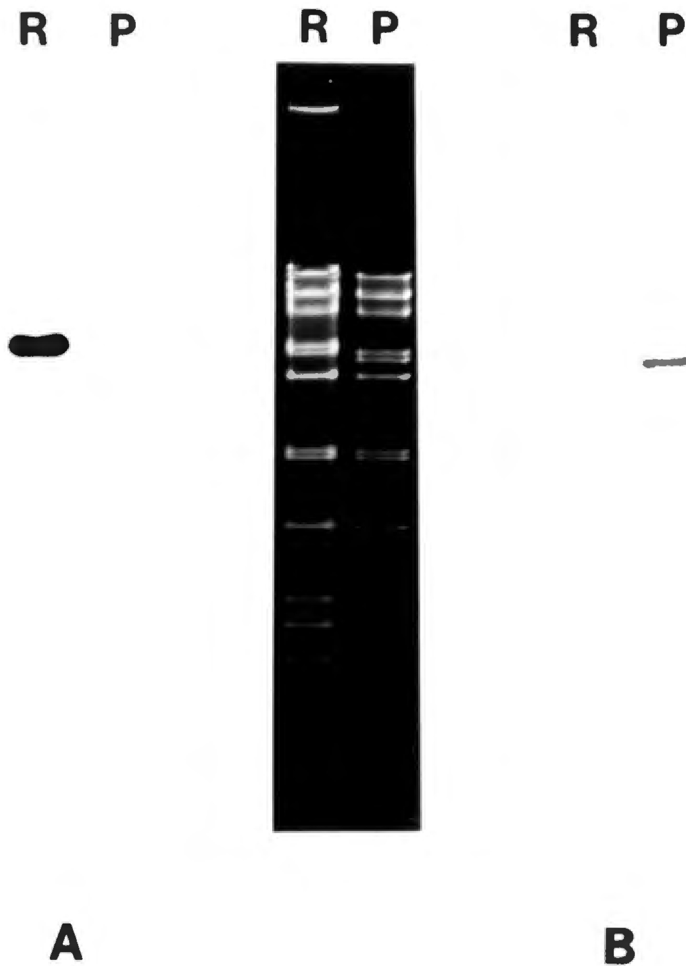


Fig. 6.4 *HindIII* digests of parent (P) and recombinant (R) viruses subject to agarose gel electrophoresis (centre), followed by hybridization to radiolabelled probes of: (A) pGptMTL23, which contains the gpt gene, and (B) a *ClaI-BglII* fragment corresponding to the region deleted in the recombinant virus.

The isolation of this recombinant virus shows that the E5R gene of vaccinia virus is nonessential for virus viability. In addition, the deletion has not altered the phenotypic properties of the virus with respect to pock formation on CAMs and plaque-forming ability in CV1, rabbit kidney or human fibroblast cells.

RNA expression of the E5R gene

RNA was extracted from CV1 cells infected with either the parent or the recombinant viruses at early and late times in the infection cycle (see chapter 2 for methods). For the early RNA extractions, cycloheximide was added to the medium both during and after infection and the cells were harvested after 6 hours. Late RNA was extracted from cells harvested 8 hours after infection.

Fig. 6.5 shows the result of Northern blot analysis of the RNA after hybridization to: A) an oligomer which binds to the coding strand of the E5R gene, downstream of the *BglII* site into which the gpt gene was cloned (positions 1102-1087 in fig. 5.2, 5'-CGAACAGGCATACATG-3'), and B) plasmid pGptMTL23 (containing the gpt gene).

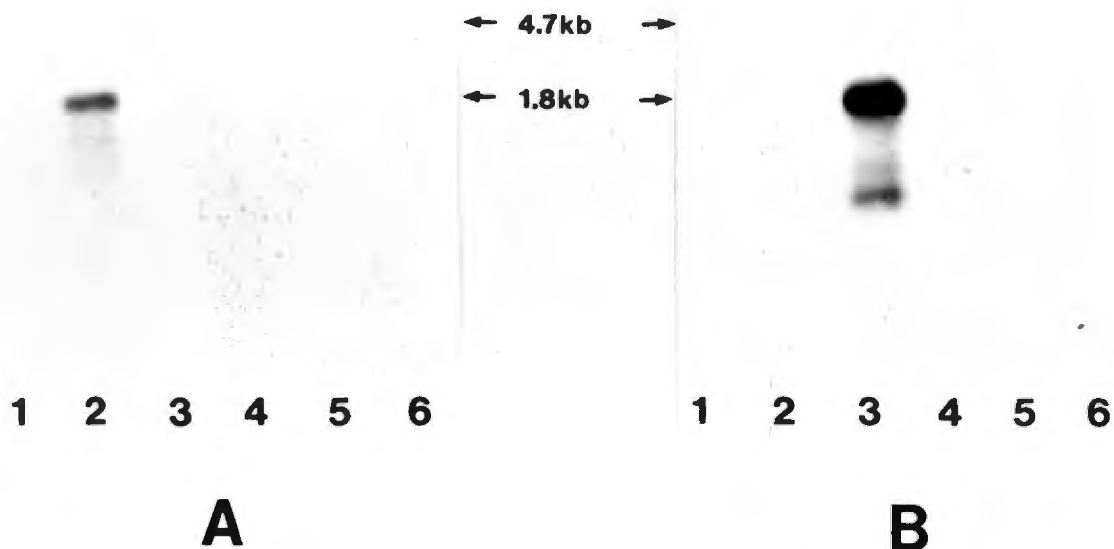


Fig. 6.5 Northern blot analysis of RNA extracted from CV1 cells, uninfected (lanes 1 and 4), infected with parent vaccinia virus, Dairen (lanes 2 and 5), and infected with recombinant virus, gptC/B (lanes 3 and 6). Immediate early RNA (lanes 1 to 3) was isolated from cells incubated in the presence of cycloheximide for 6 hours after viral infection; and late RNA (lanes 4 to 6) was isolated 8 hours after infection. A - hybridization to an end-labelled oligonucleotide, which binds downstream of the site into which the gpt fragment was inserted in the recombinant virus (5'-CGAACAGGCATACATG-3', positions 1102-1087 in fig. 5.2.). B - hybridization to the gpt gene labelled by nick-translation.

Fig. 6.5(A) shows that RNA transcribed from the E5R gene is expressed immediately after virus infection. Since RNA is expressed in the presence of cycloheximide, which inhibits translation, no prior expression of poxviral genes is required - all the necessary components for transcription are present in the virus core particle and the cell. No late expression was detected. The recombinant virus does not produce RNA from the template downstream of the gpt insertion site.

The RNA band detected is of a discrete size, approximately 2kb. Within the stretch of DNA sequenced no transcriptional termination site (TTTTTNT) was present up- or downstream of the E5R gene. Examination of the published Copenhagen sequence revealed the presence of early transcriptional termination signals at positions 51534-51540 (TTTTTTT) in the E4L ORF and 53706-53712 (TTTTTTT) within the E6R ORF. The E5R RNA product probably begins soon after the termination signal within the E4L ORF and ends at the latter termination signal, producing a transcript of <2172bp. The observed size of approximately 2kb implies that the E5R coding sequence is flanked by long stretches of RNA both upstream and downstream of the intiation and termination codons respectively. The gpt-specific probe hybridized to RNA from cells infected with the recombinant virus only, as expected (B). The size of this transcript (± 2kb) is larger than the ±1kb of the E5R sequence. This transcript will contain undetermined lengths of 5' leader and 3' poly-A sequence. The exact position of RNA initiation should be determined by primer extension and S1 nuclease protection experiments.

The role of the E5R gene product in viral pathogenicity

Exeriments in cell culture showed that the recombinant replicated at the same rate as the parent virus.

Preliminary experiments were done to determine whether the E5R gene product has any effect on the pathogenicity of the virus *in vivo.* 10^5 pfu of the recombinant and parent viruses were injected intradermally into rabbits and the lesions produced photographed after 5 and 8 days. A typical result is shown in fig. 6.6. The recombinant virus produced a smaller raised lesion which resolved faster than that of the parent virus. This may reflect a role of the gene product in increasing the time length of infectivity (if the same occurs in species other than rabbits).

A similar effect was observed in two repeat experiments, where individual rabbits were injected with either one or both viruses.

The difference observed between the two lesions probably results from interruption of the E5R gene, but this is not conclusive. The small E ORF B open reading frame, which has its coding sequence on the opposite DNA strand to E5R, was truncated at its 3' end by 88 nucleotides in the recombinant virus. It is possible that this deletion caused the difference in viral lesions. Secondly, a foreign gene, gpt, which may influence viral pathogenicity, has been inserted into the genome (However, Kerr *et al.*, 1991, have compared the virulence of

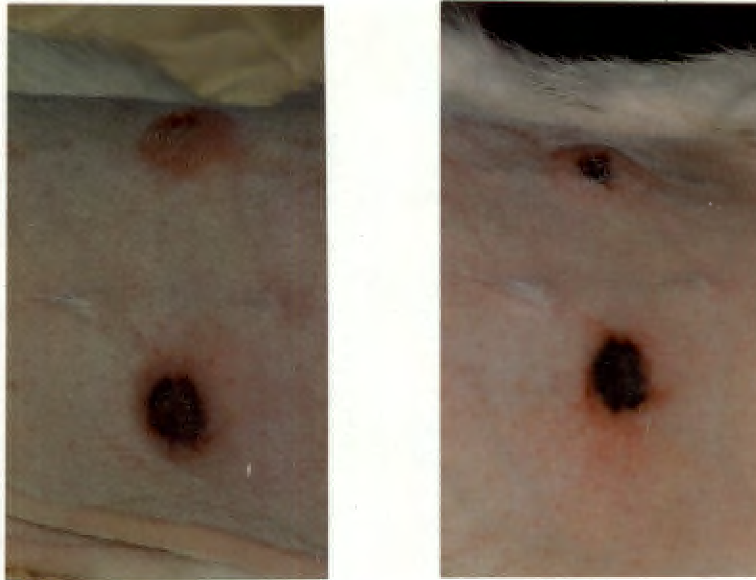


Fig 6.6 Lesions produced in rabbit skin by the recombinant (above) and parent (below) viruses. Rabbits were inoculated intradermally with 10^5 p.f.u. of virus and lesions were photographed 5 (left) and 8 (right) days later.

vaccinia virus similar in all respects except for the presence or absense of the gpt gene, and observed no difference in pathogenicity). A third factor to be considered is the presence of the first 141bp of potential coding seqence in the recombinant virus which may allow for expression of a polypeptide of 47 amino acid residues. This short sequence may have residual activity in which case the entire effect of the gene product may not have been abolished.

To confirm that the E5R gene product is responsible for the increase in severity of the lesion further recombinants ought to be made and tested. In particular, the restoration of virulence by reinsertion of the full E5R sequence into the less pathogenic recombinant should be demonstrated.

In addition to rabbit inoculation, pathogenicity experiments should also include the measurement of weight loss and determination of LD_{50} for mice. These are standard virulence tests and would allow comparisons to be made to other genes, which are known to influence viral pathogenicity.

Most nonessential genes affect viral pathogenicity for laboratory animals. A decrease in virulence has been observed for viruses containing deletions in the following genes: thymidine kinase (Buller *et al.*, 1985), haemagglutinin (Flexner *et al.*, 1987), vaccinia growth factor (Buller *et al.*, 1988), the 13.8kDa secreted protein (Kotwal *et al.*, 1989), ribonucleotide reductase (Child *et al.*, 1990), DNA ligase (Kerr *et al.*, 1991) and 3ß-hydroxysteroid dehydrogenase isomerase (Moore and Smith, 1992).

The effect of the E5R gene product on viral virulence is inconclusive. However, disruption of the gene appears to reduce the virulence of vaccinia virus for rabbits. This would be a positive factor when considering the use of this gene as a site of insertion for the construction of recombinant vaccines.

However, animal models are not always a true reflection on the effect produced in humans. The best example is that of variola virus, which is the least pathogenic *Orthopoxvirus* for laboratory animals, but the most pathogenic one for humans.

Vaccinia virus E5R putative gene product

The predicted polypeptide encoded by the E5R gene of the Dairen strain of vaccinia virus is 341 amino acids in length with a predicted molecular weight of 40436 Da. The protein has an overall positive charge (fig. 6.7) with a net charge of 16 and a pI of approximately 9.8. There is a potential signal sequence from residues 8 to 22 (overlined in fig. 6.9) with cleavage site between amino acids glycine and histidine at positions 20 and 21 respectively (von Heijne, 1986). The amino terminus is hydrophobic (fig. 6.8) and has the ability to form a beta sheet (Chou and Fasman, 1974). The hydrophobic signal sequence in the Dairen E5R amino acid sequence indicates that the putative protein probably spans the cell membrane and is secreted. Two potential glycosylation sites are present at positions 76 and 134. The site at position 76 may be glycosylated *in vivo*, an event associated with the secretion of eukaryotic proteins. There is a proline residue at position 135 which may inhibit glycosylation at position 134.

A search was made for α-helices, β-sheets and β-turns using the Chou and Fasman method of prediction of protein conformation (Chou and Fasman, 1974). There are three regions of possible α-helix formation, from positions 50-72, 233-245 and 250-260. The latter two predicted helices are separated by the sequence CVTS which is a helix breaker (<Pα> = 0.88). It is tempting to suggest that the sequence from 233-260 may form a helix-turn-helix motif which may interact with DNA (Harrison and Aggarwal, 1990). This is purely speculative.

Searches for similarity with other known protein sequences through GenBank have revealed no clues, other than the amino terminus bearing resemblance to some membrane spanning proteins.

The function of vaccinia virus E5R is open to speculation. The limited pathogenicity tests done would suggest that its absence results in faster healing. This may indicate that the gene product is involved in inhibiting the inflammatory response. Poxviruses code for a number of products with recognized anti-inflammatory functions. These include: an interleukin-1 soluble receptor homologue (Smith and Chan, 1991; Alcami and Smith, 1992; Spriggs *et al.*, 1992; Ray *et al.*, 1992); a tumour necrosis factor α receptor homologue (Smith *et al.*, 1990; Smith *et al.*, 1991b); a γ-interferon receptor homologue (Upton *et al.*, 1992); and serine protease inhibitors (Palumbo *et al.*, 1989; Kotwal and Moss,1989; Smith *et al.*, 1989).

Preliminary experiments, comparing the intracellular and secreted proteins of the parent and recombinant viruses, indicate that the E5R gene product is secreted. These results could not be confirmed as attempts at making an antibody specific to the E5R gene prod-

uct failed. A fusion protein was expressed in *E.coli* from a recombinant pUEX2 construct. This protein was purified and injected into rabbits in different ways (intradermal, subcutaneous or intramuscular). None of the antisera bound specifically to a vaccinia virus protein; most were specific to β-gal. A better approach would be to express the gene as a smaller fusion protein, using vectors such as pMal, pET or pGex which also allow for better purification of the fusion protein. Peptide synthesis was too costly an alternative.

Without a specific antibody further characterization of the protein is limited. Future work would entail the production of a specific antibody which would allow a number of questions to be answered. Once the site of action has been established, subsequent purification of the protein by immunoprecipitation would allow for functional assays to be performed.

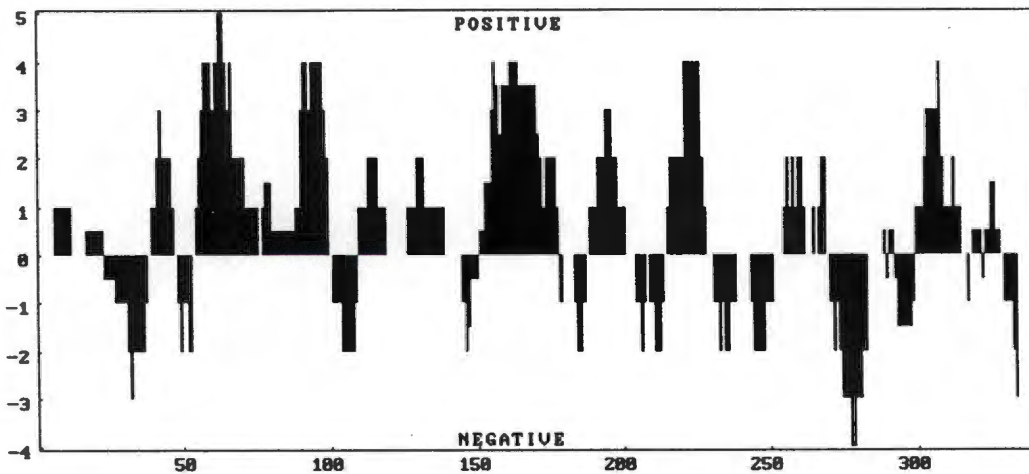


Fig. 6.7 Diagram showing the charge distribution of the putative E5R gene product.

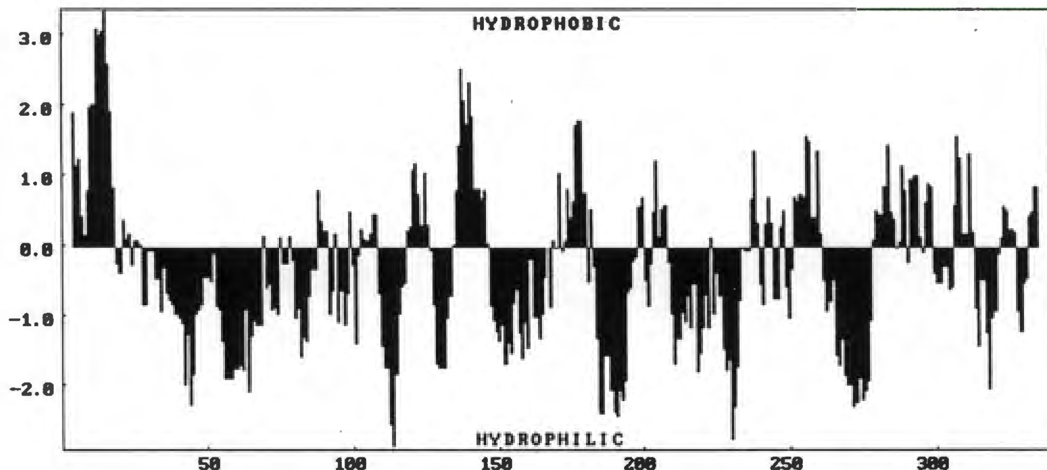


Fig. 6.8 Diagram showing regions of hydrophobicity (+ve) and hydrophilicity (-ve) within the putative E5R gene product (Kyte and Doolittle, 1982).

Putative products of the E5R-equivalent sequence in other *Orthopoxviruses*

The two variola viruses and vaccinia virus WR have ORFs which would translate to a product similar to and equal in length to the Dairen product. Reference to fig. 5.5 shows that the ectromelia sequence has a putative gene product which is truncated by 85 amino acids at the carboxy end. Camelpox virus, gerbilpox virus and vaccinia virus Copenhagen each has a 2bp deletion at positions 219-220 which introduces a frame-shift that would stop translation at position 296. The two cowpox viruses have the same 2bp deletion at positions 219-220, but also have a 1bp insertion at position 243; resulting in termination of translation at position 262.

More than 900bp 5' to the vaccinia virus stop codon (positions 1228 in fig. 5.5) are well conserved in all the viruses and an ORF is maintained in each of them of >250 codons. This occurrence in 6 species of Orthopoxvirus strongly suggests that a product is made from this sequence and that its integrity is functionally important. The present data would support this hypothesis only if translation were to start from an ATG codon downstream of the first ATG shown for vaccinia virus Dairen. Although the ATG at position 191 (fig. 5.5) appears to be in a suitable context (Kozak, 1989), it could be too close to the RNA start site to be used. It is thus important to establish where transcription begins.

Presumptive gene products were determined on the hypothesis that initiation were at the first in-frame ATG codon (after reading backwards from the termination codon at position 1228) and the corresponding amino acid sequences are aligned in fig. 6.7. This hypothesis would imply that some of the viral proteins do not have a signal sequence and are retained intracellularly.

Fig. 6.9 (On next page.) Deduced amino acid sequence of the E5R gene product of vaccinia virus Dairen aligned with the equivalent deduced sequences from the other *Orthopoxviruses*.The hydrophobic signal sequence is overlined and the cleavage site denoted by a vertical arrow.The potential glycosylation site at position 76 is blocked.Asterisks denote positions of identity, dots show conservative amino acid changes, and bold type highlights the positions in which differences are found. Despite the occurrence of deletions and insertions in many of the viral sequences at the beginning of the equivalent ORF, ORFs coding for >250 amino acid residues at the 3'end are preserved in 6 species. Translation products are shown reading back to the first available in-frame methionine. It has not been established whether these viruses have a mechanism for initiation at ATG codons downstream of the first ATG.

```
VACDIE   MLILTKVNIYMLIIVLWLYGHNFIMSE-SQCPMINDDSF--TLKRKYQIDSAESTIKMDK   60
VACCOP   ----------MLIIVLWLYGYNFIMSG-SQCPMINDDSF--TLKRKYQIDSAESTMKMDK
VACWR    MLILTKVNIYMLIIVLWLYGYNFIISE-SQCPMINDDSF--TLKRKYQIDSAESTIKMDK
VARHAR   MLILTKVNIYMLIIVLWLYGYNFIMSG-SQCPMINDDRF--TLKRKYQIDSVESTMKMDK
VARSOM   MLILTKVNIYMLIIVLWLYGYNFIMSG-SQCPMINDDRF--TLKRKYQIDSVESTMKMDK
CAM903   ----------MLIIVLWLYGYNFIMSG-SQCPMINDDSF--TLKRKYQIDSAESTMKMDK
GERBIL   ----------MLIIMLWLYGYNFIMSGRSQCPMINDDSF--TLKRKYQIDSAESTMKMDK
ECTMH    MLILTKVNIYMLIIVLWLYGYNFIMSG-SQCRMINDDSF--TLKRKYQINSVESTMKMDK
COWLAR   ------------------------MSG-SQCPMINDDLFRVTLKRMHPLNSEESSMKIDK
COWBRI   ------------------------MSG-SQCPMINDDLFRVTLKRMHPLNSEESSMKIDK
                                  *  *** ***** *  ****   ..* **..*.**

VACDIE   KRIKFQNRAKMVKEINQTIRVAQTHYETLKLGYIKFKRMIRTTTLEDIAPSIPNNQKTYK   120
VACCOP   KRTKFQNRAKMVKEINQTIRAAQTHYETLKLGYIKFKRMIRTTTLEDIAPSIPNNQKTYK
VACWR    KRTKFQNRAKMVKEINQTIRAAQTHYETLKLGYIKFKRMIRTTTLEDIAPSIPNNQKTYK
VARHAR   KRTKFQNRAKMVKEINQTIRAAQTHYETLKLGYIKFKKMIRTTTLEDITTSIPNIQKIYK
VARSOM   KRTKFQNRAKMVKEINQTIRAAQTHYETLKLGYIKFKKMIRTTTLEDITTSIPNIQKIYK
CAM903   -RTKFQNRAKMVKEINQTIRAAQTHYETLKLGYIKFKKMIMTTTPEDIATSIPNIQKIYK
GERBIL   KRTKFQNRAKMVKEINQTIRAAQTHYETLKLGYIKFKKMIRTTTLEDMATSIPNIQKIYK
ECTMH    KRTKFQNRAKMVKEINQTIRAAQTHYETLKLGYIKFKRMIRTTTLEDIATSIPNIQKIYK
COWLAR   KRTKFQNRAKMVKEINQTIRAAQTHYETLKQGYLKFKRMIRTATLKDIAPSIPKIQKFYK
COWBRI   KRTKFQNRAKMVKEINQTIRAAQTHYETLKQGYLKFKRMIRTATLKDIAPSIPKIQKFYK
          * ***************** *********  **.***.** *.*  *.. ***. ** **

VACDIE   LFSDISAIGKASQNPSKMVYALLLYMFPNLFGDDHRFIRYRMHPMSKIKHKIFSPFKLNL   180
VACCOP   LFSDISAIGKASQNPSKMVYALLLYMFPNLFGDDHRFIRYRMHPMSKIKHKIFSPFKLNL
VACWR    LFSDISAIGKASRNPSKMVYALLLYMFPNLFGDDHRFIRYRMHPMSKIKHKIFSPFKLNL
VARHAR   LFSDISAIGKVSQNPSKMAYALLLYMFPNLFGDDHRFILYRMFPMSKIKHKIFSPFKLNL
VARSOM   LFSDISAIGKVSQNPSKMAYALLLYMFPNLFGDDHRFILYRMFPMSKIKHKIFSPFKLNL
CAM903   LFSDISAIGKASQNPIKMVYALLLFMFPNLFGDDHRFILYRIHPMSKIKHKIFSPFKLNL
GERBIL   LFSDISAIGKASQNPSKMAYALLLYMFPNLFGDDHRFILYRMHPMSKIKHKIFSPFKLNL
ECTMH    LFSDISAIGKASQNPSKMVYALLLYMFPNLFGDDHRFILYRMHPMSKIKHKIFSPFKLNL
COWLAR   LFSDISAISKASQNPSKMVYALLLYMLPNLFGDDHRFIRYRMHPMSKIKHKIFSPFKLNL
COWBRI   LFSDISAISKASQNPSKMVYALLLYMFPNLFGDDHRFIRYRMHPMSKIKHKIFSPFKLNL
         ********.* *.** ** *****.*.*********** **. *****************

VACDIE   IRILVEERFYNNECRSNKWRIIGTQVDKMLIAESDKYTIDARYRLRPIYRIKGKSEEDTL   240
VACCOP   IRILVEERFYNNECRSNKWRIIGTQVDKMLIAESDKYTIDARYNLKPMYRIKGKSEEDTL
VACWR    IRILVEERFYNNECRSNKWRIIGTQVDKMLIAESDKYTIDARYNLKPMYRIKGKSEEDTL
VARHAR   IRILVEERFYNNECRSNKWRIIGTQVDKMLIVESDKYTIDARYRLRPIYRIKGKSEEDTL
VARSOM   IRILVEERFYNNECRSNKWRIIGTQVDKMLIVESDKYTIDARYRLRPIYRIKGKSEEDTL
CAM903   IRILVEERFYNNECRSNKWRIIGTQVDKMLIAESDKYTIDARYRLRPIYRIKGESEEDTL
GERBIL   IRILVEERFYNNECRSNKWRIIGTQVDKMLIAESDKYTIDARYRLRPIYRIKGESEEDTL
ECTMH    IRILVEERFYNNECKDYRWKIIGSQVDKILIAESAKYTMNAMYRLRPIHRIKVESEEDTL
COWLAR   IRILVEERFYNNECRDYKWRIIGTQVDKILIAESAKYTIDARYRLRPIYRIKGESEEDTL
COWBRI   IRILVEERFYNNECRDYKWRIIGTQVDKILIAESAKYTIDARYRLRPIYRIKGESEEDTL
         **************.  .*.***.****.** ** ***..* * *.*. ***  ******

VACDIE   FIKQMVEQCVTSQELVEKVLKILFRDLFKSGEYKAYRYDDDVENGFIGLDTLKLNIVHDI   300
VACCOP   FIKQMVEQCVTSQELVEKVLKILFRDLFKSGEYKAYRYDDDVENGFIGLDTLKLNIVHDI
VACWR    FIKQMVEQCVTSQELVEKVLKILFRDLFKSGEYKAYRYDDDVENGFIGLDTLKLNIVHDI
VARHAR   FIKQMVDQCVTSQELVEKVLKILFRDLFKSGEYKAYRYDDDVENGFIGLDKLKLNIVHDI
VARSOM   FIKQMVDQCVTSQELVEKVLKILFRDLFKSGEYKAYRYDDDVENGFIGLDKLKLNIVHDI
CAM903   FIKQMVEQCVTSQELVEKVLKILFRDLFKSGEYKAYRYDD-VENVFIGLDKLKLNILHDI
GERBIL   FIKQMVEQCVTSQELVEKVLKILFRDLFKSGEYKAYRYDDDVENGFIGLDKLKLNIVHDI
ECTMH    FIKQMVEKCVTSQELVEKC-----------------------------------------
COWLAR   FIKQMVEKCVTSQELVEKVLKILFKDLFKSGEYKDYRYDDDVDNGFIGLDKLKLNIVHDI
COWBRI   FIKQMVEKCVTSQELVEKVLKILFKDLFKSGEYKDYRYDDDVDNGFIGLDKLKLNIVHDI
         ******..**********

VACDIE   VEPCMPVRRPVAKILCKEMVNKYFENPLHIIGKNLQECIDFVSE                   344
VACCOP   VEPCMPVRRPVAKILCKEMVNKYFENPLHIIGKNLQECIDFVSE
VACWR    VEPCMPVRRPVAKILCKEMVNKYFENPLHIIGKNLQECIDFVSE
VARHAR   VEPCMPVRRPVAKILCKEMVNKYFENPLHIIGKNLQECIDFVSE
VARSOM   VEPCMPVRRPVAKILCKEMVNKYFENPLHIIGKNLQECIDFVSE
CAM903   VEPCMPVRRPVAKILCKEMVNKYFENPLHIIGKNLQECIDFVNE
GERBIL   VEPCMPVRRPVAKILCKEMVNKYFENPLHIIGKNLQECIDFVSE
ECTMH    --------------------------------------------
COWLAR   VEPCMPVRRPVAKVLCKEMVNKYFENPLHIIGKNLQECIDFVSE
COWBRI   VEPCMPVRRPVAKVLCKEMVNKYFENPLHIIGKNLQECIDFVSE
```

Fig. 6.9

CHAPTER 7

GENERAL CONCLUSIONS

CHAPTER 7

GENERAL CONCLUSIONS

The results of this work can be summarized in 32 points. The DNA sequences to which reference has been made are restricted to the region corresponding to vaccinia virus ORF E5R. The comparison using a single gene may, of course, not be representative of the entire genome.

1. 1260bp of variola virus DNA can be aligned to give 94% similarity with a sequence in monkeypox virus, if four deletions of 391-, 24-, 24- and 1-bp are introduced. These gaps effectively disrupt the ORF in monkeypox virus which corresponds to that in variola virus.
2. The deletion of 391bp is not represented elsewhere in the monkeypox virus genome.
3. The sequencing results are strong evidence that monkeypox virus is not a direct progenitor of variola virus. There is historical and biological evidence that variola virus is not ancestral to monkeypox virus. Hence the two viral species evolved independently from a common ancestor.
4. Two strains of variola virus, Harvey (variola major) and Somalia (variola minor), show strong DNA sequence conservation (1 base change in 1260bp).
5. The ORF sequenced in variola virus is 96.8% similar to the published sequence for the ORF E5R in the Copenhagen strain of vaccinia virus. An equivalent sequence was found in an additional strain of vaccinia virus of Japanese origin (Dairen).
6. Monkeypox virus has diverged into two geographically distinct groups, West Africa and Zaire, which can be readily differentiated (by PCR) by the size of the major deletion within the E5R-equivalent region. The Zaire isolates have a major deletion of 127bp, compared to 391bp in the West African isolates.
7. The Zaire isolates have the 24-, 24- and 1-bp deletions found in the West African isolates, and, in addition, have two deletions of 1- and 2-bp and an insertion of 3bp. In place of the 391bp deletion in the West African isolates, the Zaire isolates have three deletions of 5-, 9- and 127-bp as well as a single base pair insertion. Mutations have therefore accumulated both before and after separation of the virus into two lineages.
8. The DNA sequences within the two monkeypox virus groups are highly conserved. All the Zaire isolates are identical to one another, and the West African isolates show >99% nucleotide identity. Between the two groups there is >99% similarity. This suggests that the divergence occurred recently.
9. The DNA sequence corresponding to vaccinia virus E5R is conserved in all the *Orthopoxvirus* species analysed. These include three strains of vaccinia virus, WR (accession no. M35027), Copenhagen (Goebel *et al.*, 1990b) and Dairen; two strains

of variola virus, Harvey and Somalia; two strains of cowpox viruses, Larkin and Brighton; camelpox virus, 903; taterapox virus; and ectromelia virus, Mill Hill. This ORF was therefore probably present in a functional form in the common ancestor.

10. Within species there is >98% similarity. There is not as much sequence conservation within the vaccinia viruses (±98%) as there is within the variola, monkeypox and cowpox virus species (>99%). Since the origin of vaccinia virus is uncertain, we cannot conclude that this species diverged earlier than the other *Orthopoxvirus* species.
11. The species most similar to one another are gerbilpox and camelpox viruses (98%).
12. Phylogenetic analysis of the DNA sequences show that variola virus is closely related to camelpox and gerbilpox viruses.
13. Vaccinia virus forms a group of its own close to gerbilpox, camelpox and variola viruses.
14. The two cowpox viruses, Larkin and Brighton, form the most distant group.
15. Ectromelia virus shares a branch with cowpox virus, even though the two species show greater DNA sequence similarity to gerbilpox virus than they do to each other.
16. The nearest neighbour to each of the other *Orthopoxvirus* species is gerbilpox virus. This suggests that gerbilpox virus may be the virus most closely related to the common *Orthopoxvirus* ancestor.
17. Monkeypox virus appears well separated from the other *Orthopoxviruses*. This is probably not a true reflection on its evolutionary relationship between the other viruses, as monkeypox virus DNA could be expected to accumulate mutations at a faster rate than the other *Orthopoxviruses*, as the ORF in question appears not to have been required for the maintenance of monkeypox virus in nature.
18. The amino acid sequence of the putative E5R gene product has certain regions which are more highly conserved than others. All the viruses code for a protein which is highly positively charged (pI~9.8).
19. The central core region of the putative protein is more conserved than the termini.
20. The amino termini of vaccinia viruses Dairen and WR, variola viruses Harvey and Somalia, and ectromelia virus, contain a potential signal sequence, indicating that the protein may be secreted in these viruses.
21. The other viruses have frame-shifting insertions or deletions which would abort translation soon after the first ATG. More than 900bp of sequence from the 3' end of the vaccinia virus ORF coding region are highly conserved in the other viruses. This suggests that the corresponding translation product is selectively required and that the ribosomes may use an ATG later in the sequence to initiate translation. This suggestion would need to be confirmed by direct experimentation.
22. Ectromelia virus has a single base pair deletion which would cause premature termination of translation, resulting in a gene product truncated by 85 amino acid residues.

23. A recombinant vaccinia virus with a deletion of 438bp within the conserved region of the E5R coding sequence is viable, indicating that this ORF is nonessential for virus viability and may be useful as a site of insertion for foreign genes.
24. The gene is transcribed early in infection, both before viral DNA replication and protein synthesis. There is no transcription late in infection. However, the function of the E5R gene product is not known.
25. The recombinant virus does not differ from the parent virus with respect to plaque formation in cell culture and pock formation on chick chorioallantoic membranes.
26. The differing lesions produced by the parent and recombinant viruses in rabbit skin suggest that the gene product may have a role to play in viral pathogenicity.
27. DNA sequence comparison of the intergenic region of different *Orthopoxviruses* corresponding to the sequence between vaccinia virus ORFs E5R and E6R shows each species, and some strains, to have a specific pattern of deletions and insertions.
28. The DNA sequence from the intergenic region of monkeypox virus is less diverged from the other *Orthopoxviruses* thatn that of the E5R-equivalent region. DNA sequence conservation may be required for control of the E6R-equivalent ORF.
29. The intergenic sequence in raccoon poxvirus is very different from the other *Orthopoxvirus* sequences, indicating an earlier divergence from the common ancestor.
30. Within the intergenic region some viruses have repeated DNA sequences. Raccoon poxvirus has a 7bp DNA sequence present in 13 copies; and camelpox, gerbilpox and ectromelia viruses have the same 5bp sequence in 2, 10 and 17 copies respectively.
31. The two cowpox viruses contain a more complex arrangement of repeated sequences within the intergenic region. There are more repeated sequences in the Larkin strain than there are in the much passaged Brighton strain.
32. Relationships between the different *Orthopoxviruses* presented here was based on differences within a single gene. If similar results were obtained from comparisons of DNA sequences from different parts of the genome, confidence would be strengthened that evolutionary relationships could be meaningfully inferred from these results.

LITERATURE CITED

LITERATURE CITED

AGUADO, E.; SELMES, I.P. & SMITH, G.L. (1992). Nucleotide sequence of 21.8kbp fo variola major virus strain Harvey and comparison with vaccinia virus. *Journal of General Virology.* **73,** 2887-2902

ALBERTINI, A.M.; HOFER, M.; CALOS, M.P. & MILLER, J.H. (1982). On the formation of spontaneous deletions: the importance of short sequence homologies in the generation of large deletions. *Cell.* **29**, 319-328.

ALCAMI, A. & SMITH, G.L. (1992). A soluble receptor for interleukin-1ß encoded by vaccinia virus: a novel mechanism of virus modulation of the host response to infection. *Cell.* **71**, 153-167.

AL FALLUJI, M.M.; TANTAWI, H.M. & SHONY, O. (1979). Isolation, identification and characterization of camelpox virus in Iraq. *Journal of Hygiene.* **83**, 267.

ALTSCHUL, S.W.; GISH, W.; MILLER, W.; MYERS, E.W. & LIPMAN, D.J. (1990). Basic local alignment search tool. *Journal of Molecular Biology.* **215**, 403-410.

ARCHARD, L.C; MACKETT, M.; BARNES, D.E. & DUMBELL, K.R. (1984). The genome structure of cowpox virus white pock variants. *Journal of General Virology.* **65,** 875-886.

ARCHARD, L.C. & MACKETT, M. (1979). Restriction endonuclease analysis of red cowpox virus and its white pock variant. *Journal of General Virology.* **45,** 51-63.

ARITA, I. (1979). Virological evidence for the success of the smallpox eradication programme. *Nature.* **279,** 293-298.

ARITA, I.; GISPEN, R.; KALTER, S.S.; WAH, L.T.; MARENNIKOVA, S.S.; NETTER, R. & TAGAYA, I. (1972). Outbreaks of monkeypox and serological surveys in non-human primates. *Bulletin of the World Health Organization.* **46**, 625-631.

ARITA, I. & HENDERSON, D.A. (1969). Smallpox and monkeypox in non-human primates. *Bulletin of the World Health Organisation.* **39,** 277-283.

ARITA, M. & TAGAYA, I. (1980). Virion polypeptides of poxviruses. *Archives of Virology.* **63,** 209-225.

ARITA, M. & TAGAYA, I. (1977). Structural polypeptides of several strains of orthopoxvirus. *Microbiology and Immunology.* **21 (6)**, 343-346.

BALDICK, C.J.; KECK, J.G. & MOSS, B. (1992). Mutational analysis of the core, spacer, and initiator regions of vaccinia virus intermediate class promoters. *Journal of Virology.* **66,** 4710-4719.

BAUSE, E. (1983). Structural requirements of N-glycosylation of proteins. *Biochemical Journal.* **209**, 331-336.

BAXBY, D. (1972). Smallpox-like viruses from camels in Iran. *Lancet.* **2**, 1063-1065.

BAXBY, D. (1975). Identification and interrelationships of the variola/vaccinia subgroup of poxviruses. *Progress in Medical Virology.* **19**, 215-246.

BAXBY, D. (1977). The origin of vaccinia virus. *The Journal of Infectious Diseases.* **136 (3)**, 453-455.

BAXBY, D. (1993). Recombinant poxvirus vaccines. *Reviews in Medical Microbiology.* **4,** 80-88.

BEDSON, H.S. (1972). Camelpox and smallpox. *Lancet.* **2,** 1253.

BEDSON, H.S. (1982). Enzyme studies for the characterization of some orthopoxvirus isolates. *Bulletin of the World Health Organisation.* **60 (3),** 377-380.

BEDSON, H.S. & DUMBELL, K.R. (1961). The effect of temperature on the growth of pox viruses in the chick embryo. *Journal of Hygiene.* **59,** 457-469.

BEDSON, H.S. & DUMBELL, K.R. (1964). Hybrids derived from the viruses of variola major and cowpox. *Journal of Hygiene.* **62,** 147-158.

BERTHOLET, C; STOCCO, P.; VAN MEIR, E. & WITTEK, R. (1986). Functional analysis of the 5' flanking sequence of a vaccinia virus late gene. *EMBO Journal.* **5,** 1951-1957.

BINNS, M.; MUMFORD, J. & WERNERY, U. (1992). Analysis of the camelpox virus thymidine kinase gene. *British Veterinary Journal.* **148,** 541-546.

BIRNBOIM, H.C. & DOLY, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research.* **7,** 1513-1523.

BLACK, M.E. & HRUBY, D.E. (1990a). Quaternary structure of vaccinia virus thymidine kinase. *Biochemical and Biophysical Research Communications.* **169,** 1080-1086.

BLACK, M.E. & HRUBY, D.E. (1990b). Identification of the ATP-binding domain of vaccinia virus thymidine kinase. *Journal of Biological Chemistry.* **265 (29),** 17584-17592.

BLACK, M.E. & HRUBY, D.E. (1992a). Site-directed mutagenesis of a conserved domain in vaccinia virus thymidine kinase: evidence for a potential role in magnesium binding. *Journal of Biological Chemistry.* **267 (10),** 6801-6806.

BLACK, M.E. & HRUBY, D.E. (1992b). A single amino acid substitution abolishes feedback inhibition of vaccinia virus thymidine kinase. *Journal of Biological Chemistry.* **267(14),** 9743-9748.

BLASCO, R.; LOPEZ OTIN, C.; MUNOZ, M.; BOCKAMP, E.O.; SIMON MATEO, C. & VINUELA, E. (1990). Sequence and evolutionary relationships of African Swine Fever virus thymidine kinase. *Virology.* **178,** 301-304.

BOCKAMP, E.; BLASCO, R. & VINUELA, E. (1990). *Escherichia coli* thymidine kinase: nucleotide sequence of the gene and relationships to other thymidine kinases. *Gene.* **101,** 9-14.

BOURSNELL, M.E.G.; FOULDS, I.J.; CAMPBELL, J. & BINNS, M.M. (1988). Non-essentail genes in the vaccinia virus HindIII K fragment: a gene related to serine protease inhibitors and a gene related to the 37K vaccinia virus major envelope antigen. *Journal of General Virology.* **69,** 2995-3003.

BOYLE, D.B. & COUPAR, E.H. (1988a). A dominant selectable marker for the construction of recombinant poxviruses. *Gene.* **65,** 123-128.

BOYLE, D.B. & COUPAR, B.E.H. (1988b). Construction of recombinant fowlpox viruses as vectors for poultry vaccines. *Virus Research.* **10,** 343-356.

BOYLE, D.B. & COUPAR, B.E.H. (1986). Identification and cloning of the fowlpox virus thymidine kinase gene using vaccinia virus. *Journal of General Virology.* **67,** 1591-1600.

BOYLE, D.B.; COUPAR, B.E.H.; GIBBS, A.J.; SEIGMAN, L.J. & BOTH, G.W. (1987). Fowlpox virus thymidine kinase: nucleotide sequence and relationships to other thymidine kinases. *Virology.* **156,** 355-365.

BRADSHAW, H.D., Jr; DEININGER, P.L. (1984). Human thymidine kinase gene: Molecular cloning and nucleotide sequence of a cDNA expressible in mammalian cells. *Molecular and Cellular Biology.* **4,** 2316-2320.

BREMAN, J.G.; BERNADOU, J. & NAKANO, J.H. (1977). Poxvirus in West African nonhuman primates: seriological survey results. *Bulletin of the World Health Organization.* **55,** 605-612.

BREMAN, J.G.; KALISA-RUTI; STENIOWSKI, M.V.; ZANOTTO, E.; GROMYKO, A.I. & ARITA, I. (1980). Human monkeypox, 1970-79. *Bulletin of the World Health Organisation.* **58 (2),** 165-182.

BROWN, C.K.; BLOOM, D.C., & MOYER, R.W. (1991). The nature of naturally occurring mutations in the hemagglutinin gene of vaccinia virus and the sequence of immediately adjacent genes. *Virus Genes.* **5 (3),** 235-242.

BUCHER, N.L.R. (1963). Regeneration of mammalian liver. *International Review of Cytology.* **15,** 245-300.

BULLER, R.M.L. & PALUMBO, G.J. (1991). Poxvirus pathogenesis. *Microbiological Reviews.* **55 (1),** 80-122.

BULLER, R.M.L.; CHAKRABARTI, S.; COOPER, J.A.; TWARDZIK, D.R. & MOSS, B. (1988). Deletion of the vaccinia virus growth factor gene reduces virus virulence. *Journal of Virology.* **62,** 866-874.

BULLER, R.M.L.; SMITH, G.L.; CREMER, K.; NOTKINS, A.L. & MOSS, B. (1985). Decreased virulence of recombinant vaccinia virus expression vectors is associated with a thymidine kinase negative phenotype. *Nature.* **317,** 813-815.

CAVALLARO, K. & ESPOSITO, J.J. (1992). Short Communications - Sequences of the raccoon poxvirus hemagglutinin protein. *Virology.* **190,** 434-439.

CHAKRABARTI, S.; BRECHLING, K. & MOSS, B. (1985). Vaccinia virus expression vector: coexpression of beta-galactosidase provides visual screening of recombinant virus plaques. *Molecular and Cellular Biology.* **5,** 3403-3409.

CHANG, A.C.Y. & BRENNER, D.G. (1988). Cationic liposome-mediated transfection: a new method for the introduction of DNA into mammalian cells. *Focus.* **10,** 66-69.

CHEN, W.; DRILLIEN, R.; SPEHNER, D. & BULLER, R.M.L. (1992). Restricted replication of ectromelia virus in cell culture correlates with mutations in virus-encoded host range genes. *Virology.* **187,** 433-442.

CHEN, M.S. & PRUSSOF, W.H. (1978). Association of thymidylate kinase activity with pyrimidine deoxyribonucleoside kinase induced by herpes simplex virus. *Journal of Biological Chemistry.* **253,** 1325-1327.

CHENG, Y.-C. (1976). Deoxythymidine kinase induced in HeLa TK$^-$ cells by herpes simplex virus type I and type II. *Biochemica et Biophysica Acta.* **452,** 370-381.

CHILD, S.J.; PALUMBO, G.J.; BULLER, R.M.L. & HRUBY, D.E. (1990). Insertional inactivation of the large subunit of ribonucleotide reductase encoded by vaccinia virus is associated with reduced virulence *in vivo. Virology.* **174,** 625-629.

CHO, C. & WENNER, H.A. (1973). Monkeypox Virus. *Bacteriological Reviews.* **37,** 1-18.

CHOU, P.Y. & FASMAS, G.D. (1974). Prediction of protein conformation. *Biochemistry.* **13 (2)**, 222-245.

CHUNG, C.T. & MILLER, R.H. (1988). A rapid and convenient method for the preparation and storage of competent bacterial cells. *Nucleic Acids Research.* **16 (8)**, 3580.

COCHRAN, M.A.; PUCKETT, C. & MOSS, B. (1985). In vitro mutagenesis of the promoter region for a vaccinia virus gene: evidence for tandem early and late regulatory signals. *Journal of Virology.* **54,** 30-37.

CONDIT, R.C. & NILES, E.G. (1990). Orthopoxvirus genetics. *Current Topics in Microbiology and Immunology,* **163,** 2-39.

COUPAR, B.E.H.; TEO, T. & BOYLE, D.B. (1990). Restriction endonuclease mapping of the fowlpox virus genome. *Virology.* **179,** 159-167.

COWLEY, R. & GREENAWAY, P.J. (1990). Nucleotide sequence comparison of homologous genomic regions from variola, monkeypox, and vaccinia viruses. *Journal of Medical Virology.* **31,** 267-271.

DAVIES, F.G.; MUNGAT, J.N. & SHAW, T. (1975). Characteristics of a Kenyan camelpox virus. *Journal of Hygiene.* **75,** 318.

DAVISON, A.J. & MOSS, B. (1989a). Structure of vaccinia virus early promoters. *Journal of Molecular Biology.* **210,** 749-769.

DAVISON, A.J.; & MOSS, B. (1989b). The structure of vaccinia virus late promoters. *Journal of Molecular Biology.* **210,** 771-784.

DELANGE, A.M. & McFADDEN, G. (1990). The Role of telomeres in poxvirus DNA replication. *Current Topics in Microbiology and Immunology,* **163,** 71-91.

DEVEREUX, J.; HABERLI, P. & SMITHIES, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Research.* **12 (1)**, 387-395.

DOOLITTLE, W.F.& SAPIENZA, C. (1980). Selfish genes, the phenotype paradigm and genome evolution. *Nature.* **284,** 601-603.

DOUGLASS, N.J. & DUMBELL, K.R. (1992). Independent evolution of monkeypox and variola viruses. *Journal of Virology.* **66,** 7565-7567.

DOWNIE, A.W. (1939). A study of the lesions produced experimentally by cowpox virus. *Journal of Pathology and Bacteriology.* **48,** 361-379.

DOWNIE, A.W. & DUMBELL, K.R. (1947). The isolation and cultivation of variola virus on the chorio-allantois of chick embryos. *Journal of Pathology and Bacteriology.* **59,** 189-198.

DOWNIE, A.W. & HADDOCK.D.W. (1952). A variant of cowpox virus. *Lancet.* **1,** 1049-1050.

DRILLIEN, R., KOEHREN, F. & KIRN, A. (1981). Host range deletion mutant of vaccinia virus defective in human cells. *Virology.* **111,** 488-499.

DUMBELL, K.R. (1968). Laboratory aids to the control of smallpox in countries where the disease is not endemic. *Progress in Medical Virology.* **10,** 388-397.

DUMBELL, K.R. & ARCHARD, L.C. (1980). Comparison of white pock (h) mutants of monkeypox virus with parental monkeypox and with variola-like viruses isolated from animals. *Nature.* **286** (5768), 29-32.

DUMBELL, K.R. & HUQ, F. (1986) The virology of variola minor. Correlation of laboratory tests with the geographicac distribution and human virulence of variola isolates. *American Journal of Epidemiology*. **123,** 403-415.

DUMBELL, K.R. & KAPSENBERG, J.G. (1982). Laboratory investigation of two "whitepox" viruses and comparison with two variola strains from southern India. *Bulletin of the World Health Organisation*. **60** (3), 381-387.

DUMBELL, K.R. & RICHARDSON, M. (1993). Virological investigation of specimens from buffaloes affected by buffalopox in Maharashtra State, India between 1985 and 1987. *Archives of Virology*. **128**, 257-267.

ESPOSITO, J.J.; CABRADILLA, C.D.; NAKANO, J.H. & OBILESKI, J.F. (1981). Intragenomic Sequence Transposition in Monkeypox Virus. *Virology*. **109**, 231-243.

ESPOSITO, J.J. & KNIGHT, J.C. (1984). Nucleotide sequence of the thymidine kinase gene region of monkeypox and variola viruses. *Virology*. **135**, 561-567.

ESPOSITO, J.J. & KNIGHT, J.C. (1985). Orthopoxvirus DNA: A comparison of restriction profiles and maps. *Virology*. **143**, 230-251.

ESPOSITO, J.J.; KNIGHT, J.C.; SHADDOCK, J.H.; NOVEMBRE, F.J. & BAER, G.M. (1988). Successful oral rabies vaccination of raccoons with raccoon poxvirus recombinants expressing rabies virus glycoprotein. *Virology*. **165**, 313-316.

ESPOSITO, J.J.; NAKANO, J.H. & OBILESKI, J.F. (1985). Can variola-like viruses be derived from monkeypox virus? An investigation based on DNA mapping. *Bulletin of the World Health Organisation*. **63** (4), 695-703.

ESPOSITO, J.J.; OBIJESKI, J.F. & NAKANO, J.H. (1978). Orthopoxvirus DNA: strain differentiation by electrophoresis of restriction endonuclease fragmented virion DNA. *Virology*. **89,** 53-66.

ESPOSITO, J.J.; OBIJESKI, J.F. & NAKANO, J.H. (1977a). Serological relatedness of monkeypox, variola, and vaccinia viruses. *Journal of Medical Virology*. **1,** 35-47.

ESPOSITO, J.J.; OBIJESKI, J.F. & NAKANO, J.H. (1977b). The virion and soluble antigen proteins of variola, monkeypox, and vaccinia virus. *Journal of Medical Virology*. **1.** 95-110.

FALKNER, F.G. & MOSS, B. (1988). *Escherichia coli* gpt gene provides dominant selection for vaccinia virus open reading frame expression vectors. *Journal of Virology*. **62 (6),** 1849-1854.

FALKNER, F.G. & MOSS, B. (1990). Transient dominant selection of recombinant vaccinia viruses. *Journal of Virology*. **64 (6),** 3108-3111.

FELLER, J.A.; MASSUNG, R.F.; TURNER, P.C.; GIBSS, E.P.J.; BOCKAMP, E.O.; BELOSA. A., TALAVERA, A.; VINUELA, E. & MOYER, R.W. (1991). Isolation and molecular characterization of the swinepox virus thymidine kinase gene. *Virology*. **183**, 578-585.

FENNER, F., (1976). Classification and nomenclature of viruses. Second report of the International Committee on Taxonomy of Viruses. *Intervirology*. **11,** 137.

FENNER, F. (1958). The biological characters of several strains of vaccinia, cowpox and rabbitpox viruses. *Virology*. **5,** 502-529.

FENNER, F.; HENDERSON, D.A.; ARITA, I.; JEZEK, A. & LADNYI, I.D. (1988). "Smallpox and its Eradication." World Health Organization, Geneva.

FENNER, F.; WITTEK, R. & DUMBELL, K.R. (1989). The Orthopoxviruses. *Academic press Inc.*

FETZER, J.; FOLKERS, G.; MULLER, I. & KEIL, G.M. (1993). Site-directed mutagenesis in the active site of the herpes simplex virus type 1 thymidine kinase gene. *Virus Genes.* **7(2)**, 205-209.

FLEMINGTON, E.; BRADSHAW, H.D.Jr.; TRAINA-DONGE, V.; SLAGEL, V. & DEININGER, P.L. (1987). Sequence, structure and promoter characterization of the human thymidine kinase gene. *Gene.* **52**, 267-277.

FLEXNER, C.; HUGIN, A. & MOSS, B. (1987). Prevention of vaccinia virus infection in immunodeficient mice by vectior-directed IL-2 expression. *Nature.* **330**, 259-262.

FOSTER, S.O.; BRINK, E.W.; HUTCHINS, D.L.; PIFER, J.M.; LOURIE, B.; MOSER, C.R.; CUMMINGS, E.C.; KUYEYI, O.E.K.; EKE, R.E.A.; TITUS, J.B.; SMITH, E.A.; HICKS, J.W. & FOEGE, W.H. (1972). Human monkeypox. *Bulletin of the World Health Organisation.* **46**, 569-576.

GANGEMI, J.D. & SHARP, D.G. (1976). Use of a restriction endonumclese in analyzing the genomes from two different strains of vaccinia virus. *Journal of Virology.* **20,** 319-323.

GEMMEL, A. & FENNER, F. (1960).Genetic studies with mammalian poxviruses III. White (u) mutants of rabbitpox virus. *Virology.* **1**, 1219-235.

GENTRY, G.A. (1992). Viral thymidine kinases and their relatives. *Pharmacology and Therapeutics.* **54,** 319-355.

GENTRY, G.A.; RANA, S.; HUTCHINSON, J. & STARR, P. (1988). Evolution of herpes and pox viruses and their hosts: a problem with the molecular clock. *Intervirology.* **29,** 277-280.

GERSHON, P.D. & BLACK, D.N. (1989). The nucleotide sequence around the capripoxvirus thymidine kinase gene reveals a gene shared specifically with leporipoxvirus. *Journal of General Virology.* **70**, 525-533.

GIBBS, A. & FENNER, F. (1984). Methods for comparing sequence data such as restriction endonuclease maps or nucleotide sequences of viral nucleic acid molecules. *Journal of Virological Methods.* **9**, 317-324.

GILLARD, S., SPEHNER, D., DRILLIEN, R. & KIRN, A. (1986). Localization and sequence of a vaccinia virus gene required for multiplication in human cells. *Proceedings of the National Academy of Science USA.* **83**, 5573-5577.

GISPEN, R. (1955). Analysis of pox-virus antigens by means of double diffusion. A method for direct serological differentiation of cowpox. *Journal of Immunology.* **74**, 134-141.

GISPEN, R. (1972). Onderscheiding van een pokvirus afkomstig uit een woestijnmuis. *Volksgezondheid verslagen.* **24**, 134-141.

GISPEN, R., BRAND-SAATHOF, B. & HEKKER, A.C. (1976). Monkeypox-specific antibodies in human and simian sera from the Ivory Coast and Nigeria. *Bulletin of the World Health Organization.* **53**, 355-360.

GISPEN, R. & BRAND-SAATHOF, B. (1974). Three specific antigens produced in vaccinia, variola, and monkeypox infections. *The Journal of Infectious Diseases.* **129** (3) 289-295.

GISPEN, R. & BRAND-SAATHOF, B. (1972). "White" poxvirus strains from monkeys. *Bulletin of the World Health Organisation.* **46**, 585-592.

GISPEN, R., VERLINDE, J.D., ZWART,P. (1967). Histopathological and virological studies on monkeypox. *Archiv fur die Gegesamte Virusforschung.* **21,** 205-216.

GOEBEL, S.J.; JOHNSON, G.P.; PERKUS, M.E.; DAVIS, S.W.; WINSLOW, J.P. & PAOLETTI, E. (1990a). The complete DNA sequence of vaccinia virus. *Virology.* **179,** 247-266.

GOEBEL, S.J.; JOHNSON, G.P.; PERKUS, M.E.; DAVIS, S.W.; WINSLOW, J.P. & PAOLETTI, E. (1990b). Appendix to "The complete DNA sequence of vaccinia virus." *Virology.* **179,** 517-563

GOODING, L.R. (1992). Virus proteins that counteract host immune defenses. *Cell.* **71,** 5-7.

GREENAWAY, P.J. & DALE, J. (1983). Genetic Manipulation. A practical Introduction to basic techniques.

GRUIDL, M.E.; HALL, R.L. & MOYER, R.W. (1992). Mapping and molecular characterization of a functional thymidine kinase from *Amsactà moorei* Entomopoxvirus. *Virology.* **186,** 507-516.

HAHON, N. (1961). Smallpox and related poxvirus infections in the simian host. *Bacteriological reviews.* **25,** 459-476.

HALL, R.L. & HINK, W.F. (1990). Physical mapping and field inversion gel electrophoresis of *Amsacta moorei* entomopoxvirus DNA. *Archives of Virology.* **110,** 77-90.

HAMILTON, A.; KINCHINGTON, D.; GREENAWAY, P.J. & DUMBELL, K. (1985). Recombinant bacterial plasmids containing inserts of variola DNA. *Lancet.* **2,** 1356-1357.

HANGGII, M.; BANNWARTH, W. & STUNNENBERG, H.G. (1986). Conserved TAAT motif in vaccinia virus late promoters: overlapping TATA box an site of transcription initiation. *EMBO Journal.* **5,** 1071-1076.

HARPER, L.; BEDSON, H.S.; & BUCHAN, A. (1979). Identification of orthopoxviruses by polyacrylamide gel electrophoresis of intracellular polypeptides. *Virology.* **93,** 435-444.

HARRISON, S.C. & AGGARWAL A.K. (1990). DNA recognition by proteins with the helix-turn-helix motif. *Annual Review of Biochemistry.* **59,** 933-69.

HERMAN, Y.F. (1964). Isolation and characterization of a naturally occurring pox virus of raccoons. *Bacteriological Proceedings.* 117.

HIGGINS, D.G. & SHARP, R.M. (1988). Fast and senstive multiple sequence alignment on a microcomputer. *CABIOS.* **5.** 151-153.

HIGGINS, D.G.; BLEASBY, A.J. & PUCHS, R. (1991). CLUSTAL V: Improved software for multiple sequence alignment. MS submitted to *CABIOS.*

HOWARD, S.T. & SMITH, G.L. (1989). Two early vaccinia virus genes encode polypeptides related to protein kinases. *Journal of General Virology.* **70,** 3187-3201.

HRUBY, D.E. (1990). Vaccinia virus vectors: new strategies for producing recombinant vaccines. *Clinical Microbiological Reviews.* **3 (2),** 153-170.

HRUBY, D.E. & BALL, L.A. (1982). Mapping and identification of the vaccinia virus thymidine kinase gene. *Journal of Virology.* **43,** 403-409.

HRUBY, D.E.; MAKI, R.A.; MILLER, D.B. & BALL, L.A. (1983). Fine structure analysis and nucleotide sequence of the vaccinia virus thymidine kinase gene. *Proceedings of National Academy of Science USA.* **80,** 3411-3415.

HUTCHINSON, H.D., ZIEGLER, D.W., WELLS, D.E. & NAKANO, J.H.(1977). Differentiation of variola, monkeypox, and vaccinia antisera by radioimmunoassay. *Bulletin of the World Health Organisation.* **55 (5)**, 613-623.

ICHIHASHI, Y. & MATSUMOTO, S. (1968). The Relationship between poxvirus and A-Type inclusion body during double infection. *Virology.* **36,** 262-270.

ICHIHASHI, Y. & MATSUMOTO, S. (1966). Studies of the nature of marchal bodies (A-type inclusion) during ectromelia virus infection. *Virology.* **29,** 264-275.

IVES, D.H.; MORSE, P.A. (Jnr) & POTTER, V.R. (1963). Feedback inhibition of thymidine kinase by thymidine triphosphate. *Journal of Biological Chemistry.* **238 (4),** 1467-1474.

JACKSON, R.J. & BULTS, H.G. (1992). The myxoma virus thymidine kinase gene: sequence and transcriptional mapping. *Journal of General Virology.* **73,** 323-328.

JAMIESON, A.T.; GENTRY, G.A. & SUBAK-SHARPE, J.H. (1974). Induction of both thymidine and deoxycytidine kinase activity by herpes viruses. *Journal of General Virology.* **24,** 465-480.

JAMIESON, A.T. & SUBAK-SHARPE, J.H. (1974). Biochemical studies on the herpes simplex virus-specified deoxypyrimidine kinase activity. *Journal of General Virology.* **24,** 481-492.

JANSSON, O.; BOHMAN, C.; MUNCH-PETERSEN, B. & ERIKSSON, S. (1992). Mammalian thymidine kinase 2. Direct photoaffinity labeling with [^{32}P]dTTP of the enzyme from spleen, liver, heart and brain. *European Journal of Biochemistry.* **206,** 485-490.

JEZEK, Z; KRIZ, B. & ROTHBAUER, V. (1983). Camelpox and its risk to the human population. *Journal of Hygiene, Epidemiology, Microbiology and Immunology.* **27,** 29-42.

JEZEK, Z. & FENNER, F. (1988). Human monkeypox. *Monographs in Virology.* **17,** 1-140.

JEZEK, Z.; ARITA, I.; MUTOMBO, M.; DUNN, C.; NAKANO, J.H. & SZCZENIOWSKI, M. (1986) Four generations of probable person-to-person transmission of human monkeypox. *American Journal of Epidemiology.* **123 (6),** 1004-1012.

JEZEK, Z.; GRAB, B.; SZCZENIOWSKI, M.; PALUKU, K.M. & MUTOMBO, M. (1988). Clinico-epidemiological features of monkeypox patients with an animal or human source of infection. *Bulletin of the World Health Organization.* **66,** 459-464.

JEZEK, Z.; SZCZENIOWSKI, M.; PALUKU, K.M. & MUTOMBO, M. (1987a). Human monkeypox: clinical features of 282 patients. *The Journal of Infectious Diseases.* **156** (2), 293-298.

JEZEK, Z.; GRAB, B. & DIXON, H. (1987b). Stochastic model for interhuman spread of monkeypox. *American Journal of Epidemiology.* **126 (6)**,1082-1092.

KAPLAN, X.; HEALING, T.D.; EVANS, N.; HEALING, L. & PRIOR, A. (1980). Evidence of infection by viruses in small British field rodents. *Journal of Hygiene.* **84,** 285-294.

KELLOGG, D.E. & KWOK, S. (1990). In PCR Protocols: A guide to methods and applications. (pages 340-342).

KEMP, G.E.; CAUSEY, O.R.; SETZER, H.W. & MOORE, D.L.(1974). Isolation of viruses from wild mammals in West Africa 1966-1970. *Journal of Wildlife Diseases.* **10,** 279-293.

KERR, S.M.; JOHNSTON, L.H.; ODELL, M.; DUNCAN, S.A.; LAW, K.M. & SMITH, G.L. (1991). Vaccinia DNA ligase complements *Saccharomyces cerevisiae cdc9,* localizes in cytoplasmic factories and affects virulence and virus sensitivity to DNA damaging agents. *EMBO Journal.* **10 (13),** 4343-4350.

KHODAKEVICH, L. & JEZEK, Z. (1986). Isolation of monkeypox virus from wild squirrel infected in nature. *Lancet.* **1**, 98-99.

KHODAKEVICH, L; SZCZENIOWSKI, M.; MAMBU-ma-DISU; JEZEK, Z.; MARENNIKOVA, S.; NAKANO, J. & MESSINGER, D. (1987a). The role of squirrels in sustaining monkeypox virus transmission. *Tropical and Geographical Medicine.* **39**, 115-122.

KHODAKEVICH, L.; SZCZENIOWSKI, M; MAMBU-MA-DISU; JEZEK, Z; MARENNIKOVA, S.; NAKANO J. & MEIER, F. (1987b) Monkeypox virus in relation to the ecological features surrounding human settlements in Bumba Zone, Zaire. *Tropical and Geographical Medicine.* **39**, 56-63.

KHODAKEVICH, L.; WIDY-WIRSKI, R.; ARITA, I.; MARENNIKOVA, S.S.; NAKANO, J. & MEUNIER, D. (1985). Orthopoxvirose simienne de l'homme en republique Centrafricaine (Monkeypox in the Central African Republic). *Bulletin de la Societe de Pathologie Exotique.* **78,** 311-320.

KINCHINGTON, D.; DOLLERY, A.; GREENAWAY, P. & DUMBELL, K. (1984). The detection of subtle differences between different orthopoxvirus genomes by heteroduplex analysis. *Virus Research.* **1**, 351-363.

KIT, S. (1985). Thymidine kinase. *Microbiological Sciences.* **2**, 369-375.

KIT, S. & DUBBS, D.R. (1965). Properties of deoxythymidine kinase partially purified from noninfected and virus-infected mouse fibroblast cells. *Virology.* **26,** 16-27.

KIT, S.; PIEKARSKI, L.J. & DUBBS, D.R. (1963). Induction of thymidine kinase by vaccinia-infected mouse fibroblasts. *Journal of Molecular Biology.* **6**, 22-33.

KIT, S.; LEUNG, W.-C.; JORGENSEN, G.N.; TRKULA, D. & DUBBS, D.R. (1974). Thymidine kinase isozymes of normal and virus-infected cells. *Cold Spring Harbor Symposia on Quantitative Biology.* **39**, 703-715.

KNIGHT, J.C.; GOLDSMITH, C.S.; TAMIN, A.; REGNERY, R.L.; REGNERY, D.C. & ESPOSITO, J.J. (1992). Further analyses of the orthopoxviruses volepox virus and raccoon poxvirus. *Virology.* **190,** 423-433.

KOMINAMI, R.; URANO, Y.; MISHIMA, Y,; MURAMATSU, M.; MORIWAKI, K. & YOSHIKURA, H. (1983). Novel repetitive sequence families showing size and frequency polymorphism in the genomes of mice. *Journal of Molecular Biology,* **165**, 209-228.

KOTWAL, G.J. & MOSS, B. (1988). Analysis of a large cluster of nonessential genes deleted from a vaccinia virus terminal transposition mutant. *Virology.* **167,** 524-537.

KOTWAL, G.J. & MOSS, B. (1989). Vaccinia virus encodes two proteins that are structurally related to members of the plasma serine protease inhibitor superfamily. *Journal of Virology.* **63 (2)**, 600-606.

KOTWAL, G.J., HUGIN, A.W. & MOSS, B. (1989). Mapping and insertional mutagenesis of a vaccinia virus gene encoding a 13,800-Da secreted protein. *Viriology.* **171**, 579-687.

KOZAK, M. (1989). The scanning model for translation: an update. *The Journal of Cell Biology,* **108**, 229-241.

KRIZ, B. (1982). A study of camelpox in Somalia. *Journal of Comparative Pathology.* **92**, 1-8.

KUMAR, S. & BOYLE, D.B. (1990). A poxvirus bidirectional promoter element with early/late and late functions. *Virology.* **179**, 151-158.

KWOH, T.J. & ENGLER, J.A. (1984). The nucleotide sequence of the chicken thymidine kinase gene and the relationship of its predicted polypeptide to that of the vaccinia virus thymidine kinase. *Nucleic Acids Research.* **12**, 3959-3971.

KYTE, J. & DOOLITTLE, R.F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology.* **157**, 105-132.

LADNYJ, I.D.; ZIEGLER, P. & KIMA, E. (1972). A human infection caused by monkeypox virus in Basankusu Territory, Democratic Republic of the Congo. *Bulletin of the World Health Organisation.* **46,** 593-597.

LEVINSON, G. & GUTMAN, G.A. (1987). Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Molecular Biology and Evolution.* **4(3)**, 203-221.

LEWIN,B. (1990). Genes IV. *Oxford University Press.*

LEWIS, J.A. (1986). Structure and expression of the Chinese hamster thymidine kinase gene. *Molecular and Cellular Biology.* **6**, 1998-2010.

LIN, P.F.; LIEBERMAN, H.B.; YEH, D.B.; XU, T.; ZHAO, S.Y. & RUDDLE, F.H. (1985). Molecular cloning and structural analysis of murine thymidine kinase genomic and cDNA sequences. *Molecular and Cell Biology.* **5**, 3149-3156.

LOURIE, B.; BINGHAM, P.G.; EVANS, H. H.; FOSTER, S.O; NAKANO, J.H. & HERRMANN, K.L. (1972). Human infection with monkeypox virus: laboratory investigation of six cases in West Africa. *Bulletin of the World Health Organisation.* **46,** 633-639.

LOURIE, B.; NAKANO, J.H.; KEMP, G.E. & SETZER, H.W. (1975). Isolation of poxvirus from an African rodent. *The Journal of Infectious Diseases.* **132 (6),** 677-681.

MACKETT, M. & ARCHARD, L.C. (1979). Conservation and variation in orthopoxvirus genome structure. *Journal of Virology.* **45**, 683-701.

MACKETT, M.; SMITH, G.L. & MOSS, B. (1984). General method for production and selection of infectious vaccinia virus recombinants expressing foreign genes. *Journal of Virology.* **49 (3),** 857-864.

MACKETT, M.; SMITH, G.L. & MOSS, B. (1985) The construction and characterisation of vaccinia virus recombinants expressing foreign genes. (Ch 7). In *DNA Cloning Volume 2: A practical approach.* [Edition: DM Glover.] IRL Press, Oxford: Washington DC.

MANIATIAS, T.; FRITSCH, E.F. & SAMBROOK, J. (1982). Molecular cloning: a laboratory manual. *Cold Spring Harbor Laboratory.*

MARCHAL, J. (1930). Infectious Ectromelia: a hitherto undescribed virus disease of mice. *Journal of Pathology and Bacteriology.* **33,** 713-728.

MARCHUK, D.; DRUMM, M.; SAULINO, A. & COLLINS, F.S. (1991). Construction of T-vectors, a rapic and general system for direct cloning of unmodified PCR products. *Nucleic Acids Research.* **19**, 1154.

MARENNIKOVA, S.S.; GURVICH, E.B. & SHELUKHINA, E.M. (1971). Comparison of the properties of five pox virus strains isolated from monkeys. *Archiv fur die gesamte Virusforschung.* **33,** 201-210.

MARENNIKOVA, S.S.; LADNYJ, I.D.; OGORODNIKOVA, Z.I.; SHELUKHINA, E.M. & MALTSEVA, N.N. (1978). Identification and study of a poxvirus isolated from wild rodents in Turkmenia. *Archives of Virology.* **56,** 7-14.

MARENNIKOVA, S.S.; MALTSEVA, N.N.; KORNEEVA, V.I. & GARANINA, N.M. (1977). Outbreak of pox disease among carnivora (Felidae) and edentata. *Journal of Infectious Diseases.* **135**, 358-366.

MARENNIKOVA, S.S.; MALCEVA, NN & HABAHPASEVA, N.A. (1981). ELISA - a simple test for detecting and differentiating antibodies to closely related orthopoxviruses. *Bulletin of the World Health Organization.* **59**, 365-369.

MARENNIKOVA, S.S. & SELUKHINA, E.M. (1976). Susceptibility of some rodent species to monkeypox virus, and course of the infection. *Bulletin of the World Health Organisation.* **53**, 13-20.

MARENNIKOVA, S.S. & SHELUKHINA, E.M. (1978). Whitepox virus isolated from hamsters inoculated with monkeypox virus. *Nature.* **276**, 291-292.

MARENNIKOVA, S.S.; SELUHINA, E.M.; MAL'CEVA, N.N.; CIMISKJAN, K.L. & MACEVIC, G.R. (1972a) Isolation and properties of the causal agent of a new variola-like disease (monkeypox) in man. *Bulletin of the World Health Organisation.* **46,** 599-611.

MARENNIKOVA, S.S.; SELUHINA, E.M.; MALCEVA, N.N.. & LADNYJ, I.D. (1972b). Poxviruses isolated from clinically ill and asymptomatically infected monkeys and a chimpanzee. *Bulletin of the World Health Organisation.* **46,** 613-620.

MARENNIKOVA, S.S.; SHELUKHINA, E.M. & MALTSEVA, N.N. (1974). Isolation of camel pox virus and investigation of its properties. *Acta Virologica* (English Edition). **18**, 423.

MARENNIKOVA, S.S.; SHELUKHINA, E.M.; MALTSEVA, N.N. & MATSEVICH, G.R. (1979) Monkeypox virus as a source of whitepox viruses. *Intervirology.* **11,** 333-340.

MARENNIKOVA, S.S.; SHELUKHINA, E.M., SHENKMAN, L.S. (1976). "White-Wild" (Variola-like) poxviruses strains from rodents in Equatorial Africa. *Acta Virologica.* **20,** 80-82.

MARENNIKOVA, SA.S.; SHELUKHINA, SHENKMAN, E.M.; MALTSEVA, N.N. & MATSEVICH, F.E. (1975). The results of examinations of wildlife monkeys for presence of antismallpox antibody and viruses of the smallpox group *Voprosy Virusologii,* **20 (3)**, 321-326. (in Russian).

MILHAUD, C.; KLEIN, M. & VIRAT, J. (1969). Analyse d'un cas de variole du singe (monkeypox) chez le chimpanzee *(Pan troglodytes). Experimentation Animale.* **2,** 121.

Monkeypox, 1991 (1992). *Weekly Epidemiological record.* **14**, 101.

MOORE, J.B. & SMITH, G.L. (1992). Steriod hormone synthesis by a vaccinia enzyme: a new type of virus virulence factor. *EMBO Journal.* **11 (5)**, 1973-1980.

MOSS, B. (1990a). Regulation of vaccinia virus transcription. *Annual Review of Biochemistry,* **59**, 661-668.

MOSS, B. (1990b). Regulation of orthopoxvirus gene expression. *Current Topics in Microbiology and Immunology,* **163**, 41-69.

MOSS, B. (1991). Vaccinia virus: a tool for research and vaccine development. *Science.* **252,** 1662-1667.

MOYER, R.W. & ROTH, C.T. (1980). The white pock mutants of rabbit poxvirus I. Spontaneous host range mutants contain delations. *Virology.* **102**, 119-132.

MUTOMBO M.W.; ARITA, I. & JEZEK, Z. (1983) Human Monkeypox transmitted by a chimpanzee in a tropical rain-forest area of Zaire. *Lancet.* **1**, 735-737.

NOBLE, J. (Jr) (1970). A study of new and old world monkeys to determine the likelihood of a simian reservoir of smallpox. *Bulletin of the World Health Organisation.* **42,** 509-514.

NOBLE, M.D. (Jr) & RICH, J.A. (1969) Transmission of smallpox by contact and by aerosol routes in *Macaca irus. Bulletin of the World Health Organisation.* **40**, 279-286.

ORGEL, L.E. & CRICK, F.H.C. (1980). Selfish DNA: the ultimate parasite. *Nature.* **284**, 604-607.

PALUMBO, G.J.; PICKUP, D.J.; FREDRICKSON, T.N.; McINTRYE, L.J. & BULLER, R.M.L. (1989). Inhibition of an inflammatory response is mediated by a 38-kDa protein of cowpox Virus. *Virology.* **172**, 262-273

PARE, N.J. (1988). DNA Comparisons of the two orthopoxviruses monkeypox and variola. MSc. thesis, UCT.

PARSONS, B.L. & PICKUP, D.J. (1987). Tandemly repeated sequences are present at the ends of the DNA of racoonpox virus. *Virology.* **161**, 45-53

PCR protocols: A guide to methods and applications. (1990). *Academic Press, Inc.* Edited by Innis, M.A.; Gelfand, D.H.; Sninsky, J.J. & White, T.J.

PEARSON, W.R. AND LIPMAN, D.J. (1988). Improved tools for biological sequence analysis. *Proceedings of the National Academy of Science USA.* **85**, 2444-2448.

PERKUS, M.E.; GOEBEL, S.J.; DAVIS, S.W.; JOHNSON, G.P.; NORTON, E.K. & PAOLETTI, E. (1991). Deletion of 55 open reading frames from the termini of vaccinia virus. *Virology.* **180,** 406-410.

PERKUS, M.E.; GOEBEL, S.J.; DAVIS, S.W; JOHNSON, G.P.; LIMBACH, K.; NORTON, E.K. & PAOLETTI, E. (1990). Vaccinia virus host range genes. *Virology.* **179**, 276-286.

PETERS, J.C. (1966). A monkeypox-enzooty in the "Blijdorp" Zoo. *Tijdschrift voor Diergeneeskunde.* **91**, 387.

PICCINI, A. & PAOLETTI, E. (1988). Vaccinia: virus, vector, vaccine. *Advances in Virus Research.* **34,** 43-64.

PICKUP, D.J.; INK, B.S.; PARSONS, B.L.; HU, W. & JOKLIK, W.K. (1984). Spontaneous deletions and duplications of sequences in the genome of cowpox virus. *Proceedings of the National Academy of Science USA.* **81,** 6817-6821.

PICKUP, D.J.; INK, B.S.; HU, W.; RAY, C.A. & JOKLIK, W.K. (1986). Hemorrhage in lesions caused by cowpox virus is induced by a viral protein that is related to plasma protein inhibitors of serine proteases. *Proceedings of the National Academy of Science, USA.* **83**, 7698-7702.

PORTERFIELD, J.S. (1989) *Andrews' Viruses of Vertebrates.* (5th Edition). Bailliere Tindall.

PRIER, J.E. & SAUER, R.M. (1960). A pox disease of monkeys. *Annals of the New York Academy of Sciences.* **85**, 951-959.

RAMYAR, H. & HESSAMI, M. (1972). Isolation, cultivation and characterization of camel pox virus. *Zentralblatt fur Veterinarmedizin, Reihe.* **19**, 182-189.

RAY, C.A.; BLACK, R.A.; KRONHEIM, S.R.; GREENSTREET, T.A.; SLEATH, P.R.; SLAVESEN, G.S. & PICKUP, D.J. (1992) Viral inhibition of inflammation: cowpox virus encodes an inhibitor of the interleukin-1ß. *Cell.* **69,** 597-604.

RIVERSIDE SCIENTIFIC (1985, 1988). GENEPRO Version 4.20; serial no. NNN.

ROBERTSON, G.R. & WHALLEY, J.M. (1988). Evolution of the herpes thymidine kinase: identification and comparison of the equine herpesvirus 1 thymidine kinase gene reveals similarity to a cell-encoded thymidylate kinase. *Nucleic Acids Research.* **16**, 11303-11317.

RONDLE, C.J.M. & DUMBELL, K.R.D. (1962). Antigens of cowpox virus. *Journal of Hygiene.* **60**, 41-49.

RONDLE, C.J.M. & SAYEED, K.A.R. (1972). Studies on monkeypox virus. *Bulletin of the World Health Organisation.* **46,** 577-583.

SAITOU, N. & IMANISHI, T. (1989). Relative efficiencies of the Fitch-Margoliash, maximum-parsimony, maximum-likelihood, minimum-evolution, and neighbour-joining methods of phylogenetic tree construction in obtaining the correct tree. *Molecular Biology and Evolution.* **6,** 514-525.

SAITOU, N. & NEI, M. (1987). The neighbor-joining method: A new method for reconstrucing phylogenetic trees. *Molecular Biology and Evolution.* **4,** 406-425.

SAMBROOK, J.; FRITSCH, E.F., & MANIATIS, T. (1989). "Molecular cloning: A laboratory manual": Second edition. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

SANGER, F., NICKLEN, S. & COULSEN, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences. USA.* **74**, 5463-5467.

SCHNITZLEIN, W.M.; & TRIPATHY, D.N. (1991). Identification and nucleotide sequence of the thymidine kinase gene of swinepox virus. *Virology.* **181**, 727-732.

SETH, A. (1984). A new method for linker ligation. *Gene Analysis Techniques.* **1,** 99-103.

SHELUKHINA, E.M.; MALTSEVA, N.N.; SHENKMAN, L.S. & MARENNIKOVA, S.S. (1975). Properties of two isolates (MK-7-73) and MK-10-10-73) from wild monkeys. *British Veterinary Journal.* 131, 746-748.

SMITH, C.A.; DAVIS, T.; ANDERSON, D.; SOLAM, L.; BECKMAN, M.P.; JERZY, R.; DOWER, S.K.; COSMAN, D. & GOODWIN, R.G. (1990). A receptor for tumor necrosis factor defines an unusual family of cellular and viral proteins. *Science.* **248**, 1019-23.

SMITH, G.L.; CHAN, Y.S. & HOWARD, S. (1991a). Nucleotide sequence of 42 kbp of vaccinia virus strain WR from near the right inverted terminal repeat. *Journal of General Virology.* **72,** 1349-1376.

SMITH, C.A.; DAVIS, T.; WIGNALL, J.M.; DIN, W.S.; FARRAH, T; UPTON, C.; McFADDEN,G. &GOODWIN, R.G. (1991b). T2 open reading frame from shope fibroma virus encodes a soluble form of the TNF receptor. *Biochemical and Biophysical Research Communications.* **176**, 335-342.

SMITH, G.L.; HOWARD, S.T. & SANG CHAN, Y. (1989). Vaccinia virus encodes a family of genes with homology to serine proteinase inhibitors. *Journal of General Virology.* **70**, 2333-2343.

SMITH, G.L. & CHAN, Y.S. (1991). Two vaccinia virus proteins structurally related to the interleukin-1 receptor and the immunoglobulin superfamily. *Journal of General Virology.* **72**, 511-518.

SOURDIS, J. & NEI, M. (1988). Relative efficiencies of the maximum parsimony and distance-matric methods in obtaining the correct phylogenetic tree. *Molecular Biology and Evolution.* **5,** 298-311.

SPEHNER, D.; GILLARD, S.; DRILLIEN, R & KIRN, A. (1988). A cowpox virus gene required for multiplication in Chinese Hamster Ovary cells. *Journal of Virology.* **62**, 1297-1304.

SPRIGGS, M.K.; HRUBY, D.E.; MALISZEWSKI, C.R.; PICKUP, D.J.; SIMS, J.E.; BULLER, R.M.L. & VAN SLYKE, J. (1992) Vaccinia and cowpox viruses encode a novel secreted interleukin-1-binding protein. *Cell.* **71**, 145-152.

TAGAYA, I.; KITAMURA, K. & SANO, Y. (1961). Virology: A New Mutant of Dermovaccinia Virus. *Nature.* **192** (4800), 381-382.

TRAKTMAN, P. (1990). The enzymology of poxvirus DNA replication. *Current Topics in Microbiology and Immunology,* **163**, 93-123.

TSAO, H.; LIU, G.;.RUAN, L. & CHU, C.M. (1988). Construction and application of plasmids containing bidirectional promoters of vaccinia virus. *Journal of Virology.* **62 (12)**, 4832-4834.

TURNER, A. & BAXBY, D. (1979). Structural polypeptides of orthopoxvirus: their distribution in various members and location within the virion. *Journal of General Virology.* **45,** 537-545.

TURNER, P.C. & MOYER, R.W. (1990). The molecular pathogenesis of poxviruses. *Microbiology and Immunology.* **163,** 125-151

UPTON, C. & McFADDEN, G. (1986). Identification and nucleotide sequence of the thymidine kinase gene of Shope fibroma virus. *Journal of Virology.* **60**, 920-927.

UPTON, C.; MOSSMAN, K. & McFADDEN, G. (1992). Encoding of a homolog of the IFN-g receptor by myxoma virus. *Science.* **258**, 1369-1373

VALERIE, K.; STEVENS, J.; LYNCH, M.; HENDERSON, E.E. & de RIEL, J.K. (1986). Nucleotide sequence and analysis of the 58.3 to 65.5-kb early region of bacteriophage T4. *Nucleic Acids Research.* **14**, 8637-8654.

VAN DE PEER, Y. & DE WACHTER, R. (1992). TREECON: a software package for the construction and drawing of evolutionary trees. *Computer Applications in the Biosciences.* **9,** 177-182.

VANSLYKE, J.K. & HRUBY, D.E. (1990). Posttranslational modification of vaccinia virus proteins. *Current Topics in Microbiology and Immunology.* **163,** 185-206.

VON HEIJNE, G. (1986). A new method for predicting signal sequence cleavage sites. *Nucleic Acids Research.* **14 (11)**, 4683-4690.

VON MAGNUS, P.; ANDERSEN, E.K.; PETERSEN, K.B. & BIRCH-ANDERSEN, A. (1959). A pox-like disease in Cynomolgus monkeys. *Acta Pathologica et Microbiologica Scandinavica.* **46**, 156-176.

WEIR, J.P.; BAISZAR, G. & MOSS, B. (1982). Mapping of the vaccinia virus thymidine kinase gene by marker rescue and by cell-free translation of selected mRNA. *Proceedings of the National Academy of Science USA*. **79**, 1210-1214.

WEIR, J.P.; & MOSS, B. (1983). Nucleotide sequence of the vaccinia virus thymidine kinase gene and the nature of spontaneous frameshift mututation. *Journal of Virology*. **46**, 539-537.

WEIR, J.P. & MOSS, B. (1987). Determinination of the transcriptional regulatory region of a vaccinia virus late gene. *Journal of Virology*. **61**, 75-80.

WESTWOOD, J.C.N.; HARRIS, W.J.; ZWARTOUW, H.T.; TITMUSS, D.H.J. & APPLEYARD, G. (1964). Studies on the structure of vaccinia virus. *Journal of General Microbiology*. **34**, 67-78.

WESTWOOD, J.C.N.;PHIPPS, P.H. & BOULTER, E.A. (1957). The titration of vaccinia virus on the chorio-allantoic membrane of the developing chick embryo. *Journal of Hygiene*. **55**, 123-139.

WOKATSCH, R. (1972). Vaccinia Virus. In: *Strains of Human Viruses*. (Edited by Majer & Plotkin). Karger: Basel.

YANISH-PERRON, C.; VIEIRA, J.& MESSING, J. (1985). Improved M13 phage cloning vectors and host strains; nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene*. **33**, 103-119.

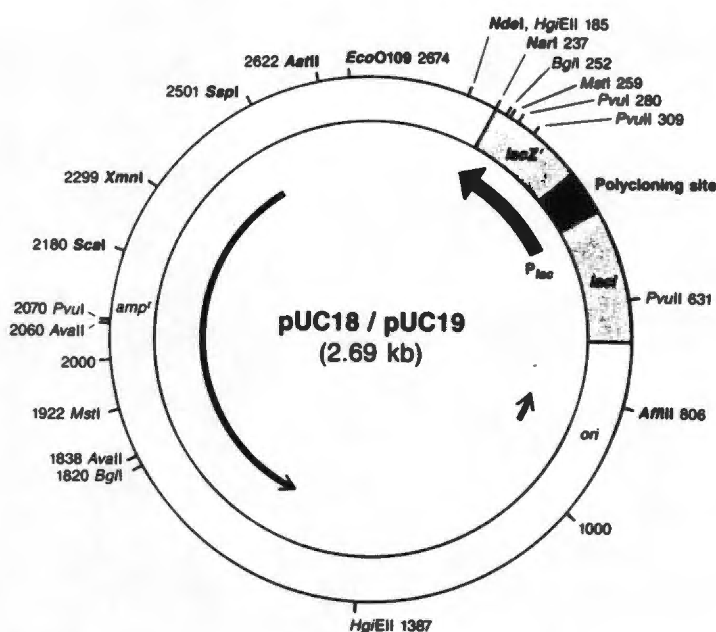
APPENDIX

Appendix A	**Vectors**	118
Appendix B	**Molecular weight marker**	120
Appendix C	**Amino acid code**	121
Appendix D	**Genetic code**	122

APPENDIX A

VECTORS

pUC18 and pUC19



Polycloning Sites

pUC18

	1	2	3	4	5	6	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	7	8
	Thr	Met	Ile	Thr	Asn	Ser	Ser	Ser	Val	Pro	Gly	Asp	Pro	Leu	Glu	Ser	Thr	Cys	Arg	His	Ala	Ser	Leu	Ala	Leu	Ala
ATG	ACC	ATG	ATT	ACG	AAT	TCG	AGC	TCG	GTA	CCC	GGG	GAT	CCT	CTA	GAG	TCG	ACC	TGC	AGG	CAT	GCA	AGC	TTG	GCA	CTG	GCC

EcoRI, SacI, KpnI, SmaI/XmaI, BamHI, XbaI, SalI/AccI/HincII, PstI, SphI, HindIII

pUC19

	1	2	3	4	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	5	6	7	8
	Thr	Met	Ile	Thr	Pro	Ser	Leu	His	Ala	Cys	Arg	Ser	Thr	Leu	Glu	Asp	Pro	Arg	Val	Pro	Ser	Ser	Asn	Ser	Leu	Ala
ATG	ACC	ATG	ATT	ACG	CCA	AGC	TTG	CAT	GCC	TGC	AGG	TCG	ACT	CTA	GAG	GAT	CCC	CGG	GTA	CCG	AGC	TCG	AAT	TCA	CTG	GCC

HindIII, SphI, PstI, SalI/AccI/HincII, XbaI, BamHI, SmaI/XmaI, KpnI, SacI, EcoRI

In pUC18, the *Eco*RI site lies immediately downstream from P_{lac}.
In pUC19, the *Hin*dIII site lies immediately downstream from P_{lac}.

pMTL23 multiple cloning site

```
                                             XhoI                                     KpnI                                      XbaI                                 AccI
                                  NruI StuI       BglII ClaI  SphI NcoI       SmaI SatI  EcoRI  HindIII PstI  MluI      SalI      NdeI
ATGACCATGATTACGCCAAGCTCGCGAGGCCTCGAGATCTATCGATGCATGCCATGGTACCCGGGAGCTCGAATTCTAGAAGCTTCTGCAGACGCGTCGACGTCATATGGATCCGATATCGCCGGCAATTCACTGGCC
 M  T  M  I  T                                                                                                                                  L  A
```

Restriction sites (left to right): NruI, StuI, XhoI, BglII, ClaI, SphI, NcoI, KpnI, SmaI, SatI, EcoRI, XbaI, HindIII, PstI, MluI, AccI, SalI, AatII, NdeI, BamHI, EcoRV, NaeI

pGpt07/14

pGpt07/14

pUC

EcoRI

B

Ecogpt

P-7.5

EcoRI

APPENDIX B

MOLECULAR-WEIGHT MARKER

DNA molecular-weight marker VI **(Boehringer Mannheim)**

Fragment sizes (bp)	
	2176
	1766
	1230
	1033
	653
	517
	453
	394
	298 (x2)
	234 (x2)
	220
	154 (x2)

APPENDIX C

AMINO ACID CODE

The single and three letter codes for the amino acids are shown below.

Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic acid	Glu	E
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

APPENDIX D

GENETIC CODE

The "universal" genetic code is shown below.

Codon	Amino Acid	Codon	Amino Acid	Codon	Amino Acid	Codon	Amino Acid
AAA	K	CAA	Q	GAA	E	TAA	*
AAC	N	CAC	H	GAC	D	TAC	Y
AAG	K	CAG	Q	GAG	E	TAG	*
AAT	N	CAT	H	GAT	D	TAT	Y
ACA	T	CCA	P	GCA	A	TCA	S
ACC	T	CCC	P	GCC	A	TCC	S
ACG	T	CCG	P	GCG	A	TCG	S
ACT	T	CCT	P	GCT	A	TCT	S
AGA	R	CGA	R	GGA	G	TGA	*
AGC	S	CGC	R	GGC	G	TGC	C
AGG	R	CGG	R	GGG	G	TGG	W
AGT	S	CGT	R	GGT	G	TGT	C
ATA	I	CTA	L	GTA	V	TTA	L
ATC	I	CTC	L	GTC	V	TTC	F
ATG	M	CTG	L	GTG	V	TTG	L
ATT	I	CTT	L	GTT	V	TTT	F