Evolutionary and Functional Studies of the Mouse Retroviral Restriction Gene, *Fv1*

Scott Anthony Ellis

A thesis submitted in part fulfilment of the requirements of the University of London for the degree of Doctor of Philosophy

March 2000

Virology Division

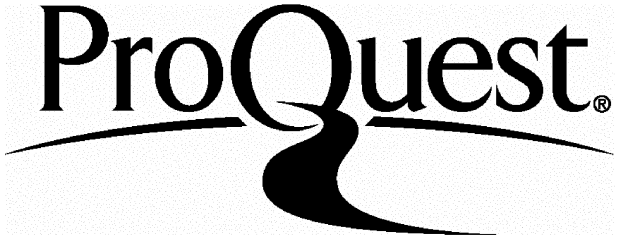
National Institute for Medical Research

The Ridgeway

Mill Hill

London

NW7 1AA



Abstract

Fv1 is a gene of mice known to restrict the replication of Murine Leukaemia virus (MLV) by blocking integration by an unknown mechanism. The gene itself is retroviral in origin, and is located on the distal part of chromosome 4.

The sequence of markers known to flank *Fv1* in the mouse was used to identify sequence from the human homologues of these 2 genes. The construction of primers to these sequences permitted the screening of 2 YAC and a PAC human genomic libraries for clones containing either of these genes. The YAC libraries were negative for both markers. The human region was finally cloned as 2 overlapping PAC clones. This region was sequenced, and the comparison of this sequence and the analogous region in rat to the *Fv1* region in mice allowed the determination of what sequence had been lost and gained during the formation of the gene.

The *Fv1* ORF of mice from across the *Mus* genus was PCR-amplified, cloned and sequenced. The analysis of this sequence has shown how *Fv1* has evolved during the speciation of the *Mus* genus. By combining this data with what is known of the distribution of endogenous MLV and *Fv1* activity among the *Mus* genus, a scheme of how and why *Fv1* has evolved activity has been proposed.

The mouse genome was screened for a 'progenitor' sequence(s) that may have given rise to *Fv1* during its germline infection of *Mus*. A mouse 129/SvJ genomic library was screened by hybridisation for the sequences bearing homology to *Fv1*. Sequences obtained were shown to be the *Fv1* gene itself, members of the murine endogenous retroviral-L (MERV-L) family, or had no known sequence homology as determined by BLAST searching. A polymorphism at an *EcoRI* site was identified in the *Fv1* gene of the 129 mouse that appears to be responsible for the presence of 2 bands that hybridise with an *Fv1*-specific probe in a 129 genomic southern blot.

$Fv1^{null}$ cell lines expressing mutant *Fv1* ORFs were established and assayed for *Fv1* activity in an assay based pseudotyped MLV. This allowed the contribution of the 3 changes known to be important between the 2 main alleles to be assessed. The first amino acid change in the *Fv1* ORF was found to be the major

determinant of *Fv1* phenotype, with the difference in the C-terminus contributing to a lesser extent. $Fv1^d$ allele, a less common allele of *Fv1*, was shown to be mediated by a single amino acid change and not by a change in the level of *Fv1* expression.

Acknowledgements

I would like to thank my supervisor, Dr. Jonathan Stoye, for his help and advice both during the course of this project and in the preparation of this thesis.

I am grateful to Paul Le Tissier and Greg Towers for their help and support in all aspects of research.

I would like to thank Paul Wilson and Tony Stevens for their technical help and support in the lab.

I would also like to acknowledge Nigel Douglas for his assistance in the construction of the phylogenetic trees.

Lastly, I would like to thank Lisa Patel for her support and limitless patience throughout the duration of this project.

Contents

Chapter 2 Materials and Methods

Chapter 3 The Cloning of the Human *FV1* Region

Chapter 4 The Evolution of *Fv1*

List of Figures

List of Tables

Abbreviations

bp	base pairs
cDNA	complementary DNA
cm	centimetre
cM	centimorgans
CNS	central nervous system
cpm	counts per minute
DEPC	diethyl pyrocarbonate
dH_2O	distilled and autoclaved water
DMEM	Dulbecco's modified Eagle's Medium
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
dpm	decays per minute
DTT	dithiothrietol
EDTA	ethylenediaminetetra-acetate
E. coli	*Escherichia coli*
EST	expressed sequence tag
FCS	foetal calf serum
Fv1	Friend virus-susceptibility-1
g	relative centrifugal force (G-number)
GDE	genetic data environment
G418	Geneticin, disodium salt
HCl	hydrochloric acid
HERV-L	human endogenous retrovirus-L
HGMP	Human Genome Mapping Project
IAP	intracisternal A-type particle
IPA	isopropanol (propan-2-ol)
IPTG	isopropylthio-β-D-galactoside
kb	kilobase
KCl	potassium chloride
kV	kilovolts
LB	Luria-Bertani broth
LMP	low melting point

LTR	long terminal repeat
Mb	megabase
MCS	multiple cloning site
MERV-L	murine endogenous retrovirus-L
ml	millilitre
MLV	murine leukaemia virus
mtDNA	mitochondrial DNA
nm	nanometre
NMR	nuclear magnetic resonance
ORF	open reading frame
PAC	P1 artificial chromosome
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PEG	polyethylene glycol
pfu	plaque forming units
pg	picogram
Phylip	Phylogeny Inference Package
pmol	picomole
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
RPA	ribonuclease protection assay
RT	reverse transcriptase
S	sedimentation unit
SDS	sodium dodecyl sulphate
ssDNA	single-stranded DNA
scnDNA	single-copy nuclear DNA
TTP	thymidine triphosphate
µl	microlitre
VSV	vesicular stomatis virus
w/v	weight/volume
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside
YAC	yeast artificial chromosome

Chapter 1 Introduction

Fv1 is a gene in mice that restricts the replication of a certain class of murine retroviruses, the Murine Leukaemia Viruses. The gene lies on the distal part of chromosome 4 and was cloned in this laboratory 4 years ago. The aim of this project was to elucidate the evolution of this gene in the *Mus* genus, and to investigate how changes that have occurred during the evolution of this gene relate to *Fv1* restriction.

This introduction will begin with a description of the Murine Leukaemia virus group (MLV), focusing chiefly on the biology of its lifecycle during infection, followed by a brief summary of the discovery and significance of endogenous retroviruses, in particular endogenous MLV. Next, the evolution of host responses to retroviral infection will be discussed, and this will be followed by short synopsis of the properties of the retroviral restriction gene *Fv1*, one example of such a host response to retroviral infection. The theme will be changed in the subsequent 2 sections, which deal more with the process of evolution. First, the evolution of retroposons- the DNA sequences present in the genome derived from reverse transcription -is discussed, from the appearance of the progenitor RT marking the end of the RNA world, to the colonisation of the genome by apparently 'junk' repetitive DNA, and the emergence of the retroviruses. This section is then followed by a summary of what we know about the evolution of the genus *Mus*. Although these 3 sections relate to each other only to a limited extent, they are all required in order to understand the nature of how *Fv1* has evolved in the *Mus* genus.

1.1 General Retroviral Biology

1.1.1 Taxonomy

Retroviruses can be divided into 2 groups depending on the complexity of the genetic organisation of the viral genome. These 2 groups are the simple and the complex retroviruses: simple retroviruses contain only the genes coding for the 3 major retroviral proteins, Gag, Pol and Env, as well as a smaller coding sequence, *pro*, encoding a protease. Additionally, the genome of these viruses sometimes

code for a dUTPase (*dut*). In complex retroviruses, this basic genetic complement is supplemented with additional regulatory accessory proteins derived from multiply spliced messages.

Alternatively, the retrovirus family can be divided into seven basic groups or genera defined by their evolutionary relatedness to one another, as determined by sequence analysis of their *pol* gene (see Table 1.1). The first five genera listed in the Table can cause oncogenic disease during the course of infection and were formerly known as the oncoretroviruses. The lentiviruses cause disease principally by killing or inducing loss of function of specific cells and tissues. All classes of vertebrates contain members of some or all of these six genera. Much less studied are the spumaviruses and no disease as yet has been ascribed to these retroviruses.

Table 1.1 Taxonomy of retroviruses

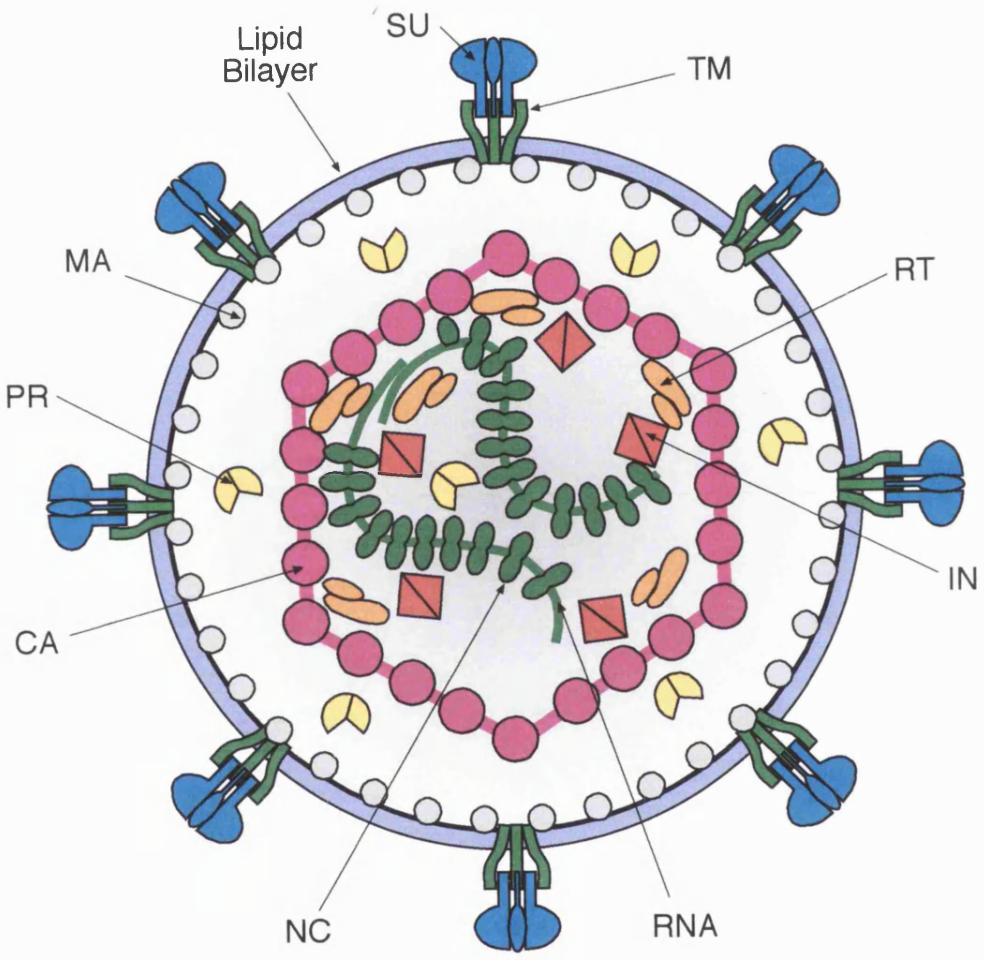
	Genus[a]	Genome	Example
1.	Avian sarcoma and leukosis viral group (ALSV)	simple	Rous sarcoma virus (RSV)
2.	Mammalian B-type viral group	simple	mouse mammary tumour virus (MMTV)
3.	Murine leukaemia-related viral group	simple	Moloney murine leukaemia (Mo-MLV)
4.	Human T-cell leukaemia-bovine leukaemia viral group	complex	human T-cell leukaemia virus (HTLV)
5.	D-type viral group	simple	Mason-Pfizer monkey virus (MPMV)
6.	Lentiviruses	complex	human immuno-deficiency virus (HIV)
7.	Spumaviruses	complex	human foamy virus HFV)

[a] Recently, the International Committee on Taxonomy of Viruses has renamed the genera: the ALSV group has become the alpharetroviruses, the B-type and D-type genera have combined as the betaretroviruses, the Murine leukaemia-related viral group (or mammalian C-type viruses) has become the gammaretroviruses and the HTLV/BLV group has become the deltaretroviruses. New genera will be named according to the Greek alphabet.

1.1.2 Virion structure

The generalised virion structure is shown in Figure 1.1. Electron microscopy has revealed some distinct differences during virion synthesis and budding, and

Figure 1.1 Schematic representation of the structure of the retrovirus particle



The lipid bilayer is host-derived and, together with the products of the *env* gene, make up the viral envelope. These products are the Transmembrane (TM) proteins and the Surface (SU) proteins, linked together by a disulphide bond. The internal structure is formed from the products of the *gag* gene, and comprises Matrix (MA), Capsid (CA) and Nucleocapsid (NC) proteins. The enzymes Reverse Transcriptase (RT) and Integrase (IN) are encoded by the *pol* gene, ans Protease (PR) is derived from the *pro* gene located between *gag* and *pol*. (Reproduced from *Retroviruses* (Coffin, J.M., Hughes, S.H. and Varmus, H., eds.))

these have been used to distinguish among various retroviral groups, leading to the original classification into 4 morphological groups (types A-D)[1]. This classification is used to a limited extent today, although it has largely been superceded by more modern molecular methods.

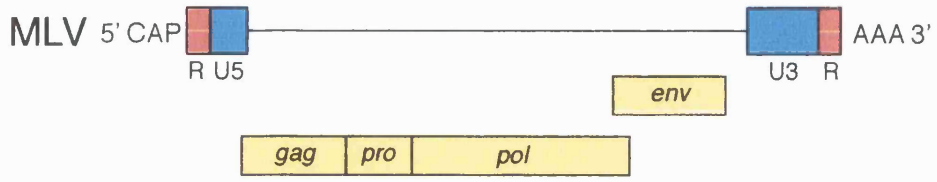
The virion is around 100nm in diameter and consists of a nucleocapsid or core surrounded by a matrix coat. The core consists of a capsid 'shell' containing the RNA genome, nucleoprotein, various enzymes and tRNA; these are required for reverse transcription and integration of the viral DNA into the host genome. This is coated with a host membrane-derived lipid envelope which has been modified by the insertion of viral envelope proteins (see Figure 1.1); these are involved in receptor-mediated cell entry and the membrane envelope serves to protect the core during the extracellular phase of the virus.

1.1.3 The retroviral genome

Retroviruses contain a plus-stranded RNA genome and replicate via a DNA intermediate which is integrated into the host genome (reviewed in [2, 3]). The processes of reverse transcription of RNA into DNA and the subsequent integration of this DNA represent the hallmarks of the Retroviridiae. They are also unique among viruses in several other respects: they have the only diploid virus genome; the only virus genome to be synthesised and processed by the cell mRNA handling machinery; the only virus genome to be associated with a specific host RNA whose function is to prime replication; and the only plus-strand RNA genome that does not serve as mRNA early after infection. Given the intimate association between the retroviral genome and the host transcription machinery, it is unsurprising to find several shared physical features with eukaryotic mRNA. Both have a 5′ capping group[4, 5] (presumably important for translation of those molecules which serve as messages). Both are modified post-translationally by methylation on the 6 position of occasional A residues[6, 7]. Not all positions are methylated, however, and the suggestion that these sites might be involved in the regulation of splicing remains to be proven. Both retroviral genomic RNAs and eukaryotic mRNAs are polyadenylated at their 3′ end as a post-translational modification by the cellular mRNA processing machinery[8, 9]. Most retroviruses have the canonical polyadenylation signal AAUAAA within 20bp of the poly(A) addition site, although in some such as the HTLV-BLV group the nearest signal is over 250bp upstream; a high degree of

secondary structure in the intervening region is thought to bring this signal near to the point of polyadenylation. It is thought that polyadenylation is important in RNA stability and the transport of RNA out of the nucleus[10].

Figure 1.2 A simple retroviral genome



The genetic map of the MLV genome encodes the *gag*, *pro*, *pol* and *env* genes. The *env* gene is in a different reading to the other 3 genes. The terminal noncoding sequence comprises 2 direct repeats (R), a unique 5' region (U5), and a unique 3' region (U3).

1.1.4 Retroviral genes

All retroviruses code for the *gag*, *pro*, *pol* and *env* genes (shown in Figure 1.2; reviewed in [11]). The *gag* gene codes for the internal structural proteins of the virion. This gene is translated as a single polyprotein which is subsequently proteolytically processed into the mature matrix (MA), capsid (CA) and nucleocapsid (NC) proteins. The *pol* gene encodes the enzymes reverse transcriptase (RT) which has DNA polymerase and associated RNase H activities, and integrase (IN) which mediates integration of the viral DNA. The *pro* gene encodes the viral protease (PR) which is responsible late in viral assembly for the proteolytic processing and maturation of the immature virion. The *env* gene encodes the surface (SU) glycoprotein and transmembrane (TM) protein of the virion which are involved in interaction with the cellular receptor and fusion between the viral membrane and the cell membrane. Another gene, *dut* (deoxyuridine triphosphatase or dU) is also present in a few retroviral groups. This enzyme enable the virus to replicate efficiently in certain cell types with low endogenous dUTPase activity[12]. The complex retroviruses also code for accessory genes, derived from multiply-spliced messages, that regulate and coordinate viral gene expression. Some of these proteins, such as the *tax* and *bel1* genes of human T-cell leukaemia virus (HTLV) and human foamy virus (HFV) respectively, act as transactivators which increase the level of transcription of the provirus, whilst others such as the *rev* gene of human immunodeficiency virus

(HIV) and the *rex* gene of HTLV serve a regulatory function by modifying the relative levels of various mRNAs. The function of others has yet to be determined.

1.1.5 The retroviral life-cycle

The basic retroviral life-cycle is outlined in Figure 1.3 below (reviewed in [3]). Briefly, the extracellular virion attaches to a specific cell-surface receptor. The virion uncoats and the core penetrates into the cell by receptor-mediated endocytosis or fusion. Once in the cytoplasm the RNA genome within the core is reverse transcribed into DNA. This DNA then moves to the nucleus, still associated with the core proteins, and is integrated into the host genome more or less randomly. The integrated viral DNA, now termed the provirus, is actively transcribed by the cellular RNA polymerases, and these transcripts are processed into genomic and messenger RNAs. The viral mRNAs are then translated into proteins, which are assembled at the cell membrane into virions that subsequently bud from the host cell. The final stage involves the proteolytic processing of the capsid proteins after budding.

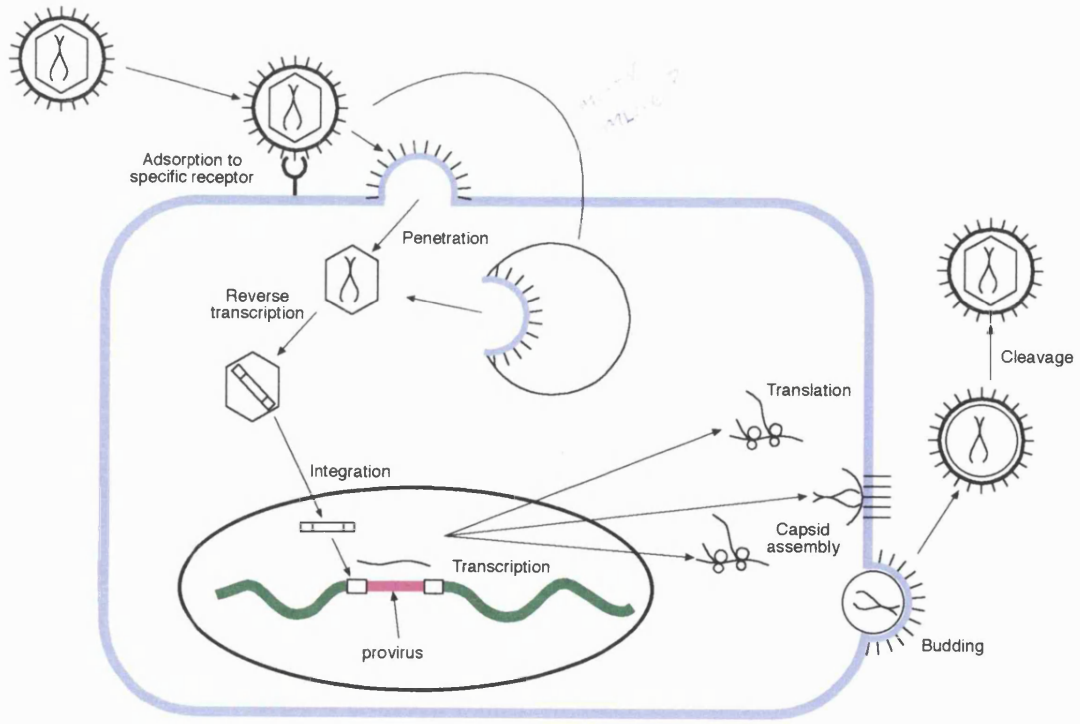


Figure 1.3 The basic retroviral life-cycle (see text for explanation)

1.2 The Murine Leukaemia viruses

The murine leukaemia viruses are simple, C-type retroviruses and are the most studied example of the mammalian C-type retrovirus family. Much of our fundamental understanding of retroviral replication comes from studies, detailed below, of the MLV family.

1.2.1 Host range of the MLV

These retroviruses can be classified according to the range of host cells they can infect. This host range is specified by their *env* gene[13]. MLV isolated from laboratory strains can be divided into 2 groups, ecotropic and non-ecotropic virus, based on this host range specificity. Ecotropic virus can infect mouse (and rat) cells only, and can be both exogenous and endogenous. This type of retrovirus can therefore subsequently be transmitted both vertically and horizontally. Ecotropic viruses use the CAT-1 receptor protein[14] for cell entry. Non-ecotropic viruses form the larger of the 2 groups, and can be subdivided into 3 classes, xenotropic, polytropic, and amphotropic viruses. Xenotropic viruses use a different receptor to ecotropic viruses and are not assocated with disease. Although these viruses are found as endogenous viruses in mouse, they can only replicate in non-murine species, except for some wild mice species[15], due to receptor incompatibility. These retroviruses probably inserted into the mouse germ line in a host background that encoded a receptor molecule with which the virus could interact to gain entry; this functional xenotropic retroviral receptor allele was subsequently lost, probably as a result of selection against xenotropic infection[16]. Xenotropism has been observed in endogenous viruses of cats and primates, as well as in the subgroup E viruses of chickens. Polytropic viruses were first identified by their ability to cause cytopathic foci on mink cells, and are therefore called mink cell focus-forming (MCF) viruses[17]. As they are able to infect both murine and non-murine cells, they are also known as dual tropic viruses. These viruses are thought to be derived from recombination events between ecotropic and non-ecotropic viruses [18], and contain an altered *env* gene relative to ecotropic viruses. Within this group is another distinct group called the modified polytropic viruses which have a characteristic deletion of 27bp in their *env* gene[19]. The xenotropic and polytropic viruses apparently use different alleles of the same receptor, *Rmc1*, which has recently been cloned by 3 groups[20-22]. The receptor has multiple transmembrane domains and

preliminary data suggest it functions as a transporter[20]. The final group of non-ecotropic MLV are the amphotropic retroviruses. This group can also infect both mouse and non-mouse cells, but use a different receptor (*Ram-1*) to the xenotropic and polytropic viruses[23]. Endogenous sequences of this group have not been found in mice[24].

The receptor for ecotropic MLV (*Rec-1*) and amphotropic MLV (*Ram-1*) are similar proteins within minimal extra- and intracellular domains and multiple transmembrane domains. The ecotropic receptor is a basic amino acid transporter[14, 25, 26], and the amphotropic receptor for MLV is a sodium-dependent inorganic phosphate transporter[27-29] which is related to a protein that serves as a phosphate transporter in *Neurospora*[30]. The gibbon ape leukaemia virus (GALV) receptor is also a sodium-dependent inorganic phosphate transporter that is related to but distinct from the amphotropic MLV receptor. The natural function of these proteins is unrelated to their role in viral infection. The GALV receptor is also used by FeLV (subgroup B) and simian sarcoma virus[31]. All of these viruses are quite closely related, and it will be interesting to see if all viruses of the same taxonomic group use receptors of similar structure.

1.2.2 Entry of the virus into the cytoplasm

Like all retroviruses, MLV gains access to the host cell by receptor-mediated fusion (section 1.1.5). The precise mechanism by which this occurs, however, remains the most poorly understood aspects of the retroviral life cycle. No active role appears to be played by the receptor itself. Fusion is probably mediated by the region of hydrophobic amino acids at the amino terminus of the TM protein in a way reminiscent of the equivalent components of other viruses e.g. influenza virus HA2, Semliki forest virus (SFV) E1 protein, and the Sendai virus fusion protein F2 (reviewed in [32]). For influenza, this occurs after endocytosis and following fusion of the virus-containing vesicles with an acidic endosome. For the majority of mammalian and avian retroviruses the mechanism of cell entry appears to be pH-independent. Entry into the cell results in the uncoating of the virion particle and penetration of the core into the cytoplasm. However, this is not true for ecotropic MLV, where fusion between the viral and cellular membrane occurs within an acidified endosomal compartment following internalisation via an endocytic pathway[33]. It is

interesting to note that ecotropic MLV entry is pH-independent on rat XC cells, perhaps due to the presence of factors on these cells (such as proteases) that negate the need for pH-induced fusion. The exact structure of the core after penetration remains undetermined. However, it must contain the retroviral genome associated with the products required for reverse transcription and integration (RT, IN, NC- see Figure 1.1), as it is difficult to see how these processes could still occur if the core's constituents were diluted into the cytoplasm. To address what components are needed for the integration reaction, a cell-free assay was developed in which amber mutations in a bacteriophage lambda genome that served as the target for integration were suppressed by integration of an MLV derivative that carried the *E. coli supF* gene[34]. Experiments using this integration system showed that the structure of the reaction product was the same as that obtained from an authentic MLV reaction. This can best be explained by the presence of a subviral structure which maintains a high degree of order during infection, and in which the retroviral DNA must be in close association with the enzymatic machinery required for integration. This was shown to be true with the finding that, in a cell-free extract from the cytoplasm and the nucleus of newly-infected cells, integration activity and viral DNA co-purified as a large (160S) nucleoprotein complex, much larger than the size expected for DNA alone[35]. This structure (the preintegration complex, or PIC) also appears to contain the major capsid protein CA (p30), as shown by the efficient immunoprecipitation of this structure with antisera specific to the CA protein. In addition, the retroviral determinants for *Fv1* restriction are known to lie in CA, and as *Fv1* is thought to interact directly with the nucleoprotein complex prior to integration (see section 1.5), this provides further evidence for the presence of CA in this complex. However, the amount of CA associated with the PIC has been found to decrease during the uncoating of the viral core, and PICs isolated from nuclear extracts have lost almost all CA proteins[36]. The presence of CA in the viral PIC is in direct contrast to the HIV preintegration complex, which appears to contain no CA[37-39]. It should be noted, however, that although this is good evidence for the presence of CA in the MLV core structure, only a subset of anti-CA sera recognise this complex in these extracts, and that the viral DNA is susceptible to nuclease attack. This indicates that the CA remaining in the MLV PIC is of a different overall conformation in the core than in the virion, and that this structure may be in a

more "relaxed" state than the virion particle. The MA protein would be expected to remain associated with the membrane, although with HIV, some of this protein appears to remain with the newly-synthesised retroviral DNA and may play an additional role in nuclear import[40].

Once in the cytoplasm, the viral RNA genome is reverse transcribed into DNA, an idea that was originally thought heretical when first proposed and for many years afterwards[41], but subsequently validated with the discovery of an RNA-dependent DNA (RT) activity associated with the virion[42, 43]. The process of reverse transcription of the retroviral genome has been exhaustively studied, and is outlined in the next section.

1.2.3 The reverse transcription reaction

Reverse transcription of the RNA retroviral genome into DNA is the keystone of retroviral biology (reviewed in [2]). Only after this event can the retrovirus integrate into the genome and initiate transcription. The reaction begins when the viral core enters the cytoplasm of the target cell. The exact structure of this core has not been well-characterised. The reaction itself generates a DNA duplex via a series of intricate steps summarised below. The linear DNA molecule represents an almost exact copy of the RNA template, the only inconsistencies being the terminal duplications in the DNA molecule known as long terminal repeats (LTRs) that are not present in the RNA (see Figure 1.4). The LTRs are identical sequences consisting of 3 elements: U3 is derived from the sequence unique to the 3′ end of the genomic RNA, R is from the sequence repeated at both ends of the genomic RNA, and U5 is derived from the sequence unique to the 5′ end of the genomic RNA. The site of transcription initiation is at the U3/R boundary, and the site of poly(A) addition is at the R/U3 boundary. U3 contains most of the transcriptional control elements of the provirus, such as the promotor and enhancer elements.

The enzyme required for the reverse transcription reaction is RT, which has 2 distinct activities: a DNA polymerase activity that can use either DNA or RNA as template, and a ribonuclease activity (termed ribonuclease H) responsible for the specific degradation of the RNA strand in DNA:RNA heteroduplexes. Although all of the enzymatic requirements for reverse transcription are provided by RT alone, it is likely that other viral proteins, such as NC, increase the efficiency of

Figure 1.4 The reverse transcription of the retroviral genome

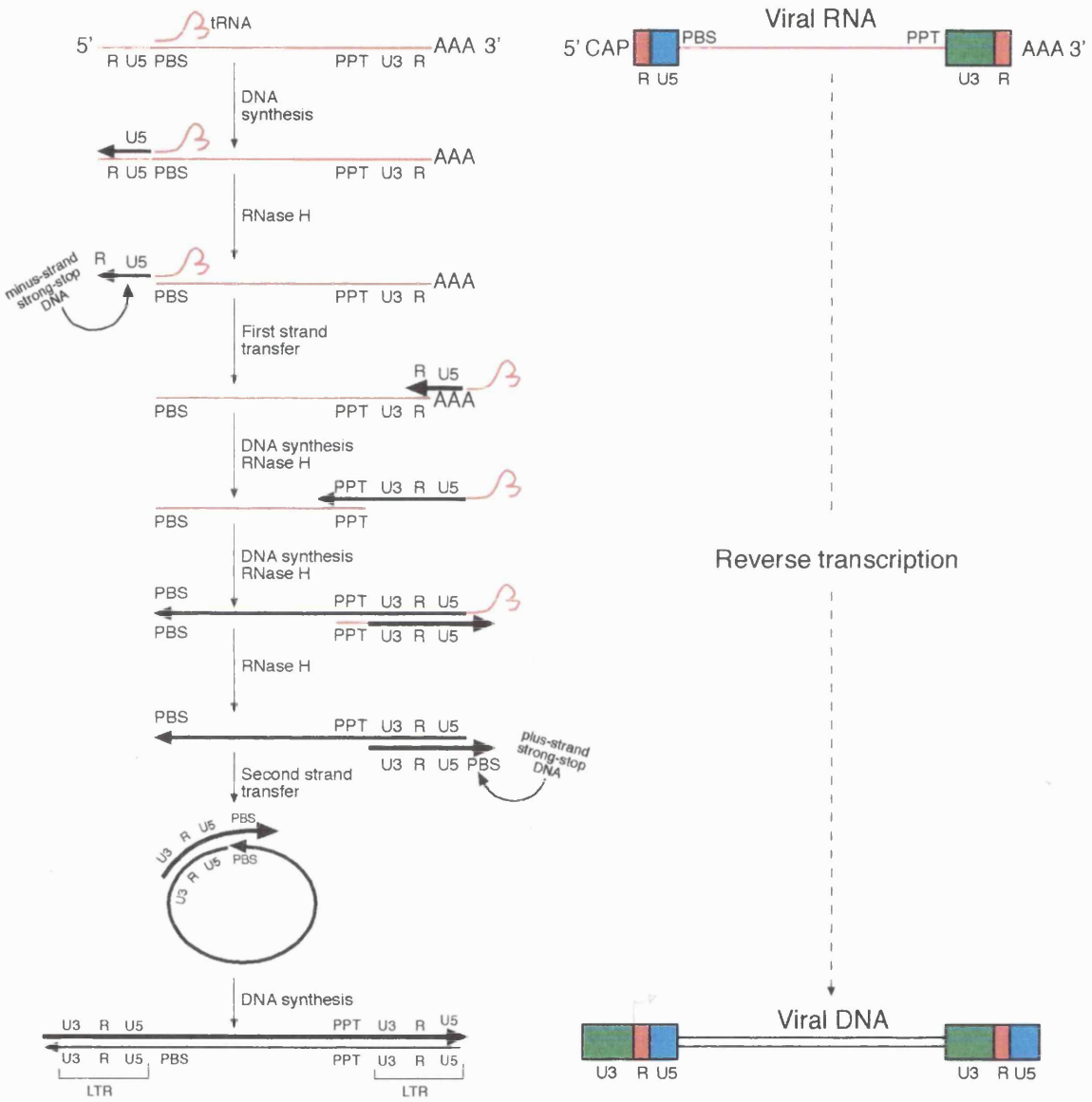


Diagram *left*: The red lines indicate RNA; the thin black lines indicate minus-strand DNA; the thick black lines indicate plus-strand DNA. See text for a description of this process.

Diagram *right*: The positions of U3, R, U5, the polypurine tract (PPT), and the primer-binding site (PBS) are indicated. The diagram shows how the terminal long terminal repeats (LTRs), the structures characteristic of the DNA form of the viral genome, are formed during the reverse transcription reaction.

the reaction. The key steps involved in this reaction are summarised below, and a schematic of the reaction itself is shown in Figure 1.4.

1. The 3′ end of a partially unwound host transfer RNA (tRNA) annealed to the primer-binding site (PBS) of the genomic RNA is used to initiate minus-strand DNA synthesis. In general, the particular tRNA used is specific to retroviruses of different classes: ASLVs use tryptophan[44], mammalian C-type viruses use proline[45, 46] (and in some instances glutamine for endogenous viruses), and both MMTVs[47] and HIV-1[48, 49] use lysine. However, this is not always the case: for instance, with the Pig Endogenous Retroviruses (PERV), PERV-A and -B groups use glycine whereas PERV-C use proline. Synthesis proceeds until the 5′ end of the genomic RNA template is reached, generating a DNA intermediate termed minus-strand strong-stop DNA (-sssDNA). As the PBS is located at the 5′ end of the genomic RNA template, the length of -sssDNA is relatively short, usually between 100-150bp long.
2. The RNase-H activity of the RT degrades the RNA strand of the RNA:-sssDNA heteroduplex. The -sssDNA strand is transferred to the 3′ end of the viral genomic RNA. This transfer is mediated by the presence of the repeated (R) sequence present in the 3′ end of the RNA genome, and the complementary sequence in the -sssDNA (present as a result of the reverse transcription across the R sequence in the 5′ end of the viral genomic RNA during the formation of -sssDNA). This annealing reaction appears to be facilitated by the NC protein.
3. As soon as the -sssDNA has annealed to the 3′ end of the genomic RNA template, minus-strand DNA synthesis continues along the length of the remaining template strand. As the DNA strand is extended, the RNA template is digested by the RNase-H activity of RT. This digestion is not complete, however, as the genomic RNA template contains a relatively RNase-H-resistant area called the polypurine tract (PPT). The RNA strand that remains undigested at this position primes plus-strand DNA synthesis. As the PPT lies directly adjacent to U3, the site where the plus-strand primer is eventually removed marks the upstream end of the LTR (and hence the end of the viral DNA). Plus-strand DNA synthesis

terminates after a portion of the tRNA is transcribed, yielding a DNA termed plus-strand strong-stop DNA (+sssDNA).

4. The tRNA is removed by the RNase-H activity of RT, allowing the complementary PBS segments of the +sssDNA and the minus-strand DNA to anneal- this is the second strand transfer.
5. Both plus- and minus-strand synthesis is completed, each strand using the other as template. The linear DNA duplex formed consists of the viral genome flanked by identical LTRs.

1.2.4 Entry of the virus into the nucleus

The exact mechanism by which the nucleoprotein complex derived from the retroviral core gains access to the nucleus is not known. Particles less than 5nm can passively diffuse through nuclear pores[50]. A particle of 160S, with an estimated diameter of 30nm[35], is therefore unlikely to enter the nucleus by passive diffusion, and may be physically too large to pass through the nuclear pore altogether. It is possible that some active transport system for nuclear import is employed, although there is no evidence for this. A more likely mechanism has been proposed based on the recognition that many, and perhaps all, oncoretroviruses require cells to be actively dividing to permit retroviral replication[51]. This cell cycle restriction of MLV infection is caused by the lack of mitosis specifically[52]. In normal, dividing cells, MLV integration can be detected within 7 hours of infection. When cells are arrested in S phase or at the G_2/M boundary, viral cell entry, uncoating and reverse transcription of the RNA genome occur at the normal rate, but integration is blocked. When assayed *in vitro*, the nucleoprotein complexes in these cells are integration-competent. When the cell-cycle block is lifted, integration occurs as the cells pass through mitosis. *In situ* hybridisation experiments show that entry of the unintegrated MLV DNA into the nucleus is mitosis-dependent[52]. Furthermore, studies involving retroviral constructs containing the *lacZ* reporter gene indicate that integration occurs into post-replication genomic DNA only, as shown by the presence of the proviral reporter construct in only half of the progeny of infected cells[53]. Therefore, it appears that for MLV, and perhaps for the other oncoretroviruses, the retroviral nucleoprotein complex gains access to the nucleus when the nuclear membrane is disassembled at mitosis.

HIV-1, and perhaps other lentiviruses, is not restricted by the cell-cycle and is able to infect non-dividing cells. Entry into the nucleus without the disassembly of the nuclear membrane is apparently due to nuclear localisation signal-mediated, energy-dependent import through the nuclear pore[54, 55].

1.2.5 Integration of the retroviral genome

In eukaryotic systems, integration is the process most unique to retrovirus replication. Other viruses use reverse transcriptase as part of their replication cycle, but no other virus has a regular mechanism for the stable integration of its genome into the host DNA. Integration is thought to be a prerequisite for the proper expression of the viral genome[56].

As well as the intermediates of retroviral DNA synthesis (see section 1.2.3), there are 4 classes of unintegrated viral DNA detectable in an infected cell[57-60]. The first class is the full-length, linear double-stranded DNA product, with an LTR at each end, which is the immediate product of reverse transcription. This is the most abundant form of DNA in infected cells, particularly during early infection, and is the only form that reaches appreciable levels in the cytoplasm. The other 3 classes are:

1. 1-LTR circles, formed by homologous recombination between the 2 LTRs of a linear viral DNA molecule;
2. 2-LTR circles, formed by the ligation of the 2 ends of the linear precursor, often with small deletions from either or both ends, or, less frequently, with larger deletions or insertions between the 2 joined ends;
3. Products of autointegration, formed by the integration of the viral ends into itself.

These circular DNAs appear in the nucleus soon after the appearance of the linear products. This delay in the appearance of the circular DNA forms is caused by the requirement of nuclear enzymes for circle formation, and the appearance of these forms of viral DNA in the nucleus can therefore be used as a marker for nuclear entry of the virus. Since all of these forms are present in the nucleus, it has been difficult to determine which is the precursor to the integrated provirus. It was originally thought that the 2-LTR circles were the precursors, with the "circle-junction" sequence formed by the 2 juxtaposed LTRs serving as the attachment site for joining the viral DNA to that of the host[61]. Later, it became clear that the circular DNAs were not involved. The integration activity purified

from the cytoplasm of infected cells, which contained no circular viral DNA, was found to be higher than that of nuclear extracts[34]. Moreover, integration activity was proportional to the amount of linear DNA. Further proof for the involvement of linear rather than circular DNA in the integration process was obtained from studies of MLV integration: an intermediate from an *in vitro* integration reaction was found to precisely match the intermediate predicted to occur during the integration of a linear precursor and not that expected from the integration of circular viral DNA[60, 62]. The circular viral DNA products, therefore, are not used as substrates in the integration process. They are presumed to result from the action of cellular enzymes on viral DNA molecules, and, as they are seen at elevated levels in infections involving integration-defective mutant viruses, they may be derived from nucleoprotein complexes that are defective for integration[34].

The mechanisms involved in the integration reaction have been proposed which involve DNA cutting and joining steps. Briefly, the 2 3′ terminal bases at each end of the newly-reverse transcribed DNA molecule are removed by the cleavage reaction of IN, leaving a 3′ OH end. This reaction may occur before entry into the nucleus. The preintegration complex, derived from the retroviral core, then enters the nucleus. This complex contains linear DNA, CA, IN, and possibly RT and NC. A staggered break is introduced into the target DNA and the strand transfer reaction then simultaneously joins the 2 ends of the viral DNA to cellular DNA about half a turn of the helix apart, with the precise spacing determined by the geometry of the IN multimer. It is the IN enzyme alone that carries out these reactions, and it has been shown that the cleavage, strand-transfer and joining reactions occur by a one-step mechanism without the involvement of a protein-DNA structural intermediate[63]. The resulting gap is filled in by cellular DNA repair mechanisms, displacing the 2 mismatched bases at the 5′ end of the provirus and ligating the remaining ends. Gap repair of the staggered join generates the duplication in the flanking DNA characteristic of integrated proviruses.

There is strong evidence for the involvement of cellular factors in the processes of integration. This will be dealt with in 1.5.7.3.

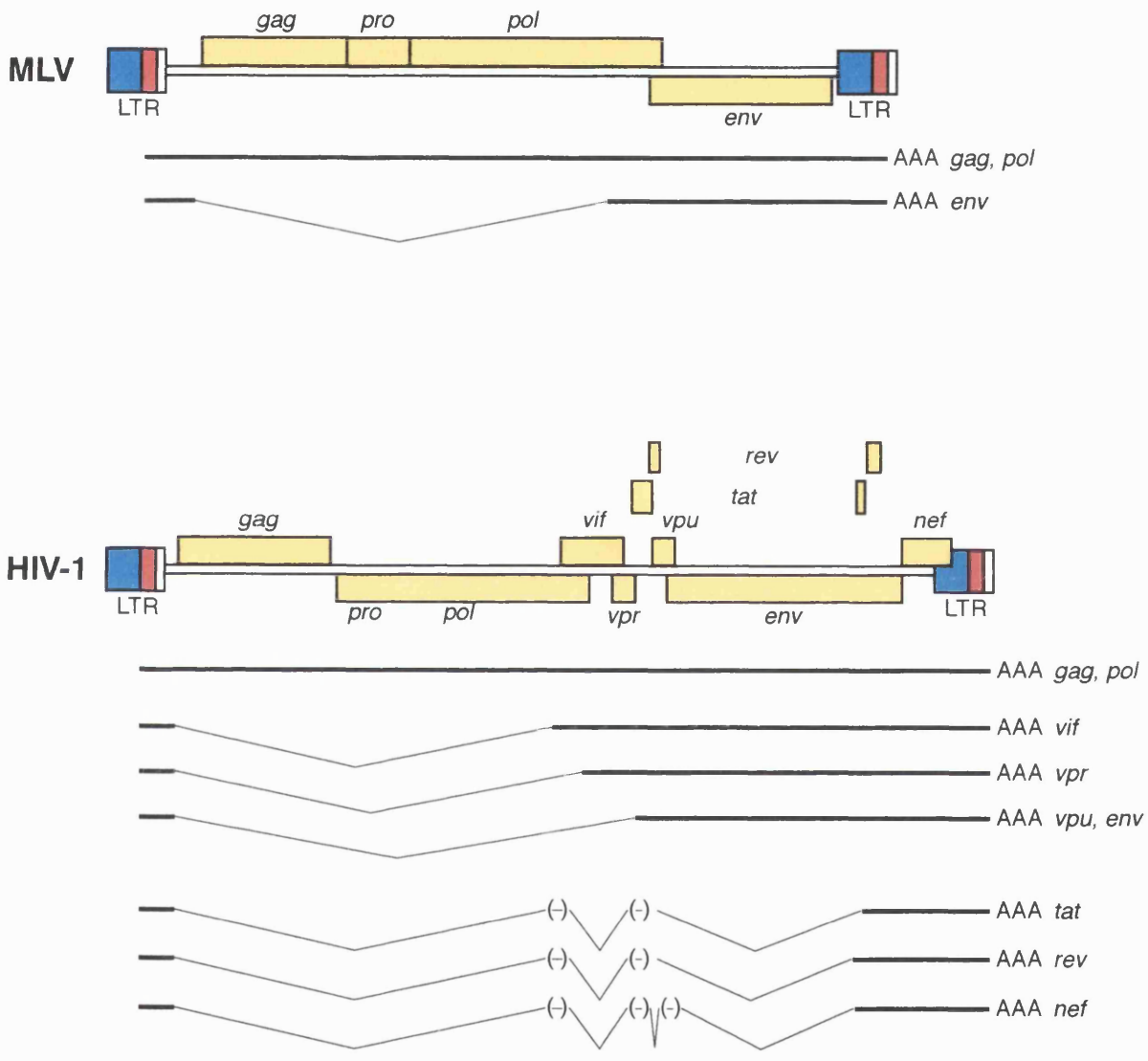
1.2.6 Expression of the MLV genome

The structure of the integrated provirus is shown in Figure 1.5. Transcription is initiated by the RNA polymerase II enzyme of the host cell, which synthesises cellular mRNAs. Evidence for this comes from the sensitivity of retroviral transcription to α-amanitin, which inhibits the function of this polymerase, as well as the presence of *cis*-acting elements encoded by the retroviral genome. Due to its absolute reliance on host transcription machinery for its replication and expression, it is not surprising to find interaction of host proteins with, and the reciprocol binding sites for these proteins in, the viral LTRs that regulate general and tissue-specific expression of the proviral DNA[64, 65].

The U3 region in the 5′ LTR contains enhancer and promoter components, and transcription is initiated at the U3-R boundary in MLV. The R (repeat) region in the 3′ LTR provides the polyadenylation signal. The leader sequence (5′ untranslated region, or UTR) of ~475bp follows the 5′ LTR, and this contains a splice donor site for the processing of mRNA. This region also contains the packaging signal ψ, but this probably extends into the coding region[66]. The position of the 3′ end of viral RNAs is determined by post-transcriptional processing. Transcription proceeds past the polyadenylation signal and the primary transcript is processed by endonucleases and polyadenylated. The only site for 3′-end processing is at the R/U5 border in the 3′ LTR. Therefore, all of the viral RNAs should have identical 3′ ends. In ASLV this is not true: a significant portion (15%) of viral transcripts fail to cleave at this position and 'read-through' into cellular sequences. Polyadenylation of these templates occurs via cellular poly(A) signals and the transcripts are packaged into virions as functional genomic RNA[67]. The primary RNA transcript serves as both the genomic RNA that is packaged into newly-synthesised virus particles and as the mRNA for the *gag*, and *pol* genes. In marked contrast with the transcription process in complex retroviruses such as HIV, the process is much more simple in MLV, where the only spliced variant of the primary transcript is for the *env* gene (see Figure 1.5), which codes for the TM and SU proteins. The mechanism that controls the frequency of splicing is not understood. All viral RNAs are transported to the cytoplasm where translation occurs.

The full-length transcript encodes *gag*, *pro* and *pol*. The products of these genes are found in different proportions relative to one another in the mature virion. To ensure the correct amounts of each protein are produced, the translation of

Figure 1.5 Alternative splicing patterns in MLV and HIV



The Figure shows the single splicing event in the simple MLV retrovirus that generates the *env* RNA, and the characteristic multiple alternative splicing events of complex retroviruses such as HIV-1. The splicing complexity in HIV-1 is increased by the alternative use of small, noncoding central exons (denoted in the Figure by parentheses).

the *gag* gene occasionally continues past its stop codon so that the Pro and Pol products are found in a proportion of the Gag polyproteins only. In MLV, this is a consequence of readthrough (or termination suppression) of the termination codon. Ribosomes bind upstream of *gag* in the leader sequence and move to the*gag* initiation codon where translation begins. At the end of *gag* the ribosome encounters an amber translational terminator and translation is stopped in 90-95% of reactions[68]. In the remaining 5-10% of cases the amber stop codon is suppressed: the first base of the amber (UAG) termination codon of the *gag* sequence is misread by the glutamine tRNA so that it functions as a glutamine-encoding CAG codon, and a glutamine is inserted at the stop site[69, 70]. This results in read-through to *pro* and *pol*, which are in the same reading frame as *gag*. In most other retroviruses, this readthrough is achieved by frameshift suppression, where suppression of the termination codon occurs by sequence-mediated tRNA slippage.

1.2.7 Virion assembly

The Gag polyprotein plays the key role in virion assembly[71]. Once at the cell membrane, this protein has the ability to direct the budding and separation of the virus-like particles from the cell surface, even in the absence of all the other viral proteins. The Env polyprotein is glycosylated during its passage through the rough endoplasmic reticulum. After travelling through the Golgi apparatus, the hydrophobic region near its carboxy terminus allows its anchorage to the plasma membrane. At the cell surface the Env protein is thought to interact with Gag, leading to the specific packaging of viral glycoproteins in the virion.

The viral RNA represents less than 1% of the total RNA of an infected cell, but is specifically packaged into virions. This specificity is due to packaging sequence(s), termed the signal (ψ) or encapsidation (E) sequence. As Gag alone has been shown to be sufficient for particle formation, and Gag-only particles contain viral RNA, it seems that the proteins needed for the specific incorporation of viral RNA in the virion are located in Gag[72, 73]. In mammalian type-C viruses, this packaging signal is located downstream of the splice donor site, thereby promoting the packaging of viral genomic DNA but not subgenomic DNA.

After budding, the virion undergoes an obligate maturation step in the formation of an infectious particle. This involves the ordered processing of the

Gag and Gag-Pro-Pol precursors by the viral protease. The protease is encoded by the *pro* gene which cleaves the Gag and Pol precursor proteins, probably after viral assembly. In this way, this mechanism ensures the correct proportions of each of the various proteins needed for the assembly of progeny virus particles are produced, provides the Pro and Pol proteins in association with Gag so they can be properly incorporated into the immature virion, and prevents the viral protease acting on cellular proteins within the cell. Post-assembly cleavage of the Gag and Pol products is thought to ensure the virion is 'primed' for replication only after release from the cell. The step can be seen by electron microscopy as the condensation of the retroviral core.

1.3 Endogenous retroviruses

As mentioned in section 1.1.5, the integration of the retrovirus into the host genome is an obligate step in the retroviral life-cycle. Once integrated, the provirus is replicated and progeny are released from the cell by budding. The retrovirus can therefore be horizontally transmitted to other host organisms. However, once integrated, the provirus is retained as a stable part of the host cell genome, and as retroviral infection does not normally kill the infected cell, the provirus is passed on to daughter cells. If the infected cell is a germ cell, and the insertion has no serious deleterious effects, this provirus can become part of the germ line. Such retroviruses are known as endogenous retroviruses, and have been found in all vertebrates in which they have been looked for, including humans[74].

1.3.1 Discovery of endogenous retroviruses

In the 1950s and 1960s, studies involving radiation- and carcinogen-induced cancers demonstrated that C-type retroviruses were frequently expressed in tumours[75]. At the same time, genetic studies led to the demonstration of genetically-transmitted, virus-associated mammary carcinomas associated with GR mice[76]. Both these observations suggested the involvement of latent, genetically-inherited proviruses in tumour formation[77]. The first direct evidence of endogenous retroviruses came from the observation that some nonproducer cell lines infected and transformed by the Bryan high-titre strain of Rous Sarcoma virus (RSV) could spontaneously start to release infectious viral particles without the addition of exogenous helper virus[273]. Almost all of the *env* gene is absent in

this RSV strain, which is consequently replication defective. Normal types of cell, it seemed, could provide the helper function to this defective virus endogenously, indicating not only the presence of endogenous retroviral sequence but the expression as well[78-81]. Endogenous retroviruses are inherited, like the host's natural genetic complement, in a stable Mendelian fashion. Although usually transcription-defective, they occur both in expressed and silent forms, as complete or partially defective viral genomes. This inactivity is associated with host-directed methylation of CG sequences in the provirus[82]. These viruses can spontaneously 'activate'[79, 83, 84] or can be induced to replicate by irradiation or by exposure to nucleotide analogues or demethylating agents[85, 86]. All endogenous retroviruses isolated so far belong to the simple category, and many are clearly related to current exogenous retroviruses.

1.3.2 Presence in the genome

Initial infection of the germline is usually followed by an amplification of the novel provirus sequence, often by an infectious mechanism identical to that seen with exogenous viruses[87]. This is not true for all sequences, for example the IAP elements: these sequences have no *env* gene, so replication is purely intracellular. By analysing the distribution and copy number of endogenous retroviral sequences among and within genera, it is apparent that they have colonised vertebrates at different times. For example, certain classes of MLV are found in some but not all subspecies of mice that have recently diverged[88], indicating a recent entry into the germline. *H*uman *e*ndogenous *r*etro*v*irus (HERV) sequences of the HERV-H family, in contrast, have been found at the same chromosomal loci in both humans and baboons, indicating the integrations took place at a point before the divergence of human and baboon, some 30 million years ago[89]. This is direct evidence that retroviral infection of vertebrates has been occurring for at least this period of time, and probably for much longer. It is sometimes helpful to think of endogenous retroviruses as either 'modern' or 'ancient'. Ancient proviruses are those sequences that inserted into the genome prior to speciation, as judged by its presence at the same loci in all individuals of a species and sometimes in related species as well. Ancient proviruses show clear signs of their age, with widespread mutations accumulated throughout their coding sequence, and none have been found to be infectious.

1.3.3 Significance of germ line infection

Because of their endogenous lifestyle, endogenous retroviruses are not just genomic copies of their exogenous relatives. Although transcriptionally active members do exist, endogenous retroviruses are predominantly inactive elements which have accrued point mutations and deletions during their residence in the host genome that make them, for the most part, replication incompetent. The few exceptions that are infectious are usually non-pathogenic; if pathogenic effects were severe, it is unlikely the endogenous virus responsible would be maintained and fixed in the host species[25]. The acquisition of endogenous retrovirus does have the potential to be extremely damaging to the host. The integration of a proviral sequence, and that of its subsequent progeny, could cause mutation by interrupting host genes. The viral sequence also contains regulatory sequences that are both cis- and trans-acting, which control the processing and expression of the retroviral genome; these sequences can also act on adjacent host sequence following integration, modifying its transcription and regulation. Examples of this include 3 mutations at the *agouti* locus that are caused by novel transcripts initiating from the LTRs of integrated intracisternal A-type particle (IAP) elements[90], and the *dilute* mutation which results from tissue-specific mis-splicing of the normal mRNA to the splice acceptor site of an integrated MLV sequence[91].

1.3.4 Human endogenous retroviruses

The human genome contains a variety of ancient endogenous retroviruses called HERV elements, but lacks any recently-acquired elements[92]. No infectious human endogenous retroviruses, therefore, have been isolated to date. HERV elements are commonly grouped according to the tRNA used for minus-strand priming, for example the HERV-L family use a leucine tRNA. This raises certain problems in cases where the primer binding site has not been determined, or when 2 different groups share the same tRNA primer. HERVs have also been divided into 2 classes based on homology in the *pol* gene sequence: those elements that share homology with mammalian C-type retroviruses comprise class I, whereas those that share homology with mammalian B- and D-type retroviruses comprise class II[93].

More recently, HERV elements with limited homology to spumaviruses have been found[94]. These retroviruses, termed HERV-L, are detectable in several

mammalian species and have expanded in primate and mouse genomes to between 100-200 copies. Although they have numerous stop codons in their coding sequence, these elements contain identifiable *gag*, *pol* and *dut* ORFs. Based on the sequence of the *pol* ORF, these elements are most closely-related to the foamy viruses. However they lack any additional accessory genes. The sequence of the cloned HERV-L element was used to isolate a murine homologue of this retroviral class[95]. Unlike the amplification of HERV-L elements, the amplification of the murine homologue of HERV-L (termed MERV-L for *mu*rine *e*ndogenous *r*etro*v*irus-L) has occurred much more recently, indicated both by the lack of this element in the related rat genome, and the 98% homology between the LTR sequences (the comparison of sequence divergence between the LTRs of a retroviral sequence is a relatively accurate way to measure how long the element has been resident in the genome, since these sequences will have diverged from one another at a fairly uniform rate due to random mutation).

No endogenous retroviruses related to the Lentivirus or BLV/HTLV groups of retroviruses have been found in the human germline to date.

1.3.5 Mouse endogenous retroviruses

There are currently 9 groups of endogenous retroviruses described in inbred strains of mice, although there undoubtedly more to be found and characterised. These groups are the B type and C type retroviruses, and the MERV-L, IAP, VL30, MuRRS (*mu*rine *r*etrovirus *r*elated *s*equence), GLN (after the glutamine tRNA primer-binding site), MuRVY (*mu*rine *r*epeated *v*irus on the *Y* chromosome) and Etn (*e*arly *t*ranspos*o*n*s*). Of these groups, only the C-type and B-type have closely related exogenous relatives, the Murine Leukaemia viruses and the Mouse Mammary Tumour viruses respectively. The MLV group will be discussed in more detail in the following sections.

1.3.6 Distribution of endogenous MLV in *Mus*

Endogenous MLV sequences have only been found in those *Mus* species closely related to laboratory strains[88]. The endogenous copies of these sequences are restricted both to specific taxonomic and geographic groups within the subgenus *Mus*, indicating that these have been acquired independently in different wild mice and have remained largely segregated in these populations. Polytropic and

xenotropic sequences are found more widely throughout the *Mus* genus, indicating the earlier acquisition of these elements into the *Mus* germline. In contrast, ecotropic sequences have only been found in a small number of the more recently diverged *Mus* species, indicating a much more recent acquisition of these sequences. The pattern seen in inbred laboratory mice seems to be the result of interbreeding of a small number of wild mice progenitors of distinct taxonomic groups[88]. This apparently recent acquisition of ecotropic virus is consistent with the proposal that, based on sequence comparisons, ecotropic MLV may have arisen from substitution of the *env* gene of a xenotropic region with the analogous region from an unknown virus which has not been fixed in the germ line[31]. Phylogenetic studies have indicated that MLV has entered the *Mus* germ line within the last 1.5 million years[88, 96].

1.3.7 Endogenous retroviruses and disease

The discovery of the oncogenic potential of retroviruses started with the discovery of retroviruses themselves at the turn of the century[97, 98]. Ellermann and Bang showed that chicken leukosis, a form of leukaemia and lymphoma, was caused by a virus, now known to be avian leukosis virus (ALV). Three years later, Rous showed the cell-free transmission of a sarcoma in chickens, by the virus we know today as Rous sarcoma virus (RSV). In 1936, John Bittner demonstrated that mammary carcinoma in mice was transmitted via milk and was caused by a filterable agent[99]. Conclusive evidence of the role of endogenous retroviruses in disease came from the study of inbred strains of mice. Several such strains (like AKR and C58) have been bred by selecting for a high incidence of thymic lymphoma, and although a viral cause was suspected, it was not until 1951 that the existence of leukaemogenic virus in these strains was demonstrated[100].

Gross showed that leukaemia could be induced by infecting new-born mice of the C3H strain, which naturally has a very low incidence of leukaemia, with cell free extracts from the organs of AKR mice, a strain with a high prevalence of this disease. Such highly-leukaemic strains of mice usually develop an 80-90% incidence of spontaneous lymphatic thymic lymphoma within 6-12 months of age[101]. High levels of MLV expression can be detected in many tissues from day 16 of gestation[102]. The disease originates in the thymus with lymphoma appearing between 6-12 months, and this spreads to the lymph nodes, spleen,

liver and sometimes the peripheral blood[103]. Although the inheritance of endogenous (ecotropic) MLV is a prerequisite for leukaemia, the actual oncogenic agent is a recombinant retrovirus known as mink cell focus-forming (MCF) virus. These viruses are formed from a complex yet consistent series of events involving at least 3 endogenous proviruses[30]. This high incidence of lymphoma probably results from the activation of a number of proto-oncogenes by MLV viral enhancer sequences after integration[104].

A second group of murine leukaemia virus was later identified from cell-free extracts of leukaemic Swiss mice[105]. As opposed to the Gross virus, which induced a chronic leukaemia with a relatively long latency period, this Friend virus was an acute transforming virus which could induce disease in adult mice, as opposed to newborn, after a short incubation period. The disease itself was characterised by a proliferation of immature mononuclear cells which invaded the spleen, liver, bone marrow, kidney, lung and also appeared in the peripheral blood. The mice would develop enlarged spleens after 2-3 weeks post-infection, and death usually followed after 1-3 months. In the latter stages of the disease, the white blood cell count was greatly elevated, and the spleen and liver were so severely enlarged that death usually resulted from splenic rupture[105, 106]. Therefore Friend virus disease did not seem to involve the thymus or lymph nodes, as with Gross virus, but predominantly involved the spleen, bone marrow and later the liver. As the major cell type involved was the erythroid cell, the disease was referred to as an erythroleukaemia[76]. As with Gross virus, the pathogenicity of Friend virus was restricted to certain strains of mice, with the Swiss and DBA/2 strains found to be most susceptible whilst C57BL seemed to be resistant to the disease. It was also observed that resistance was not absolute; certain virus-mouse combinations showed varying degrees of resistance and susceptibility. This pointed towards host genetic factors which could influence leukaemogenesis.

1.3.8 The Friend virus complex

During the initial studies of Friend virus (FV) disease, it was observed that different derivatives of the original virus produced different disease characteristics. Disease produced by one virus preparation could involve rapid splenomegaly, with marked increase of red blood cell levels in peripheral blood (or polycythaemia, hence Friend virus-P, or FV-P). Other virus preparations

would, in contrast, produce varying degrees of anaemia (hence FV-A), as in the case of the first thymus-independent disease described[105]. With the development of an *in vivo* spleen focus assay for Friend virus, it was shown that FV-P preparations gave high viral titres, whereas FV-A preparations gave much lower titres, if any at all[107]. It became apparent, therefore, that Friend virus consisted of a separable component responsible for focus formation, and was thus termed spleen focus-forming virus (SFFV). It was this component that differed between FV-A and FV-P and which led to the difference in disease phenotypes. It was also shown that Friend virus could induce lymphatic leukaemia in rats, and that when the extracts of these leukaemias were reintroduced into mice, the original erythroblastic disease reappeared. It was subsequently shown that serial passaging of the virus through rats produced a virus which could only cause lymphatic leukaemia in mice, the erythroblastic component of the virus having been lost[103]. This lymphatic component of the virus was similar to that found in the original virus by Gross, neither of which had the capacity to form spleen foci.

These findings show that the Friend virus appears to be a complex, composing separable lymphatic and SFFV components. Subsequent studies have shown that the SFFV component is replication-defective and contains a deleted recombinant *env* gene. The upstream part of the gene is related to polytropic MLVs and the downstream part is related to ecotropic MLVs, with the deletion occurring in a large central portion of this ecotropic sequence[108]. This virus is rescued by the second component of the Friend virus complex, Friend MLV, which supplies the Env proteins deficient in the SFFV component, allowing the defective virus to overcome the block in its replication. The relationship of the 2 components, therefore, is that of defective virus and helper virus. The defective SFFV component is responsible for the short-latency disease and the helper Friend MLV component is responsible for the long-latency, chronic, thymus-dependent leukaemia. The oncogenicity of the SFFV component is due to the altered *env* sequence rather than due to the acquisition of any virally-encoded sequence[109]. This is thought to be due to the ability of the gp55 product of the SFFV *env* gene to bind directly to the erythropoeitin (Epo-R) receptor, thereby triggering prolonged proliferation of infected erythroid cells[110, 111]

1.4 Evolution of host responses to retroviral infection

Studies, mainly with Friend virus but also with a variety of others such as Rauscher and Gross virus, showed that host factors could greatly influence the outcome of infection since different mouse strains were found to be susceptible whilst others were not. The following examples concern the evolution of host responses to MLV infection specifically, but due to the shared features in their life-cycle, the general principles of restriction will probably hold true for other retroviruses.

1.4.1 Cell entry and replication

There are several levels on which host genes can act to prevent retroviral infection. The most basic is at the level of cell entry: if the cell can prevent the retrovirus from entering the cell then infection will be stopped even before it has the chance to start. As discussed in section 1.2.2, retroviruses gain access to the cell by receptor-mediated fusion. If the receptor changes in a way that no longer permits functional interaction with the virion, cell entry will be blocked. An example of just this type of restriction occurs with HIV-1 in humans; some individuals possess a truncated form of the cellular co-receptor used by the virus, and these individuals show a corresponding degree of resistance to the virus[112]. Cell receptors are rarely dispensable altogether, however, as they have some part to play in the metabolism of the cell; restriction of this kind is more likely to occur by smaller point mutations that prevent specifically the interaction with the virion rather than larger mutations which are more likely to interfere with the natural receptor function. In the case of MLV, this is the most likely explanation concerning the xenotropic class: in a response to prevent further infection by this class of retrovirus, 2 mutations (a K→E change at residue 500 and a deleted T residue at position 582) in the xenotropic receptor was selected for, effectively removing the threat of infection by both endogenous and exogenous virus[113].

Physical changes to the receptor are not the only means the host can restrict receptor-mediated cell entry. The *Fv4* or *Akvr1* locus[114-116] was first identified in the Japanese mouse strain G by the observation that mice carrying the $Fv4^r$ (resistance) allele were resistant to infection by Friend virus. This gene, which acts in a dominant fashion, was subsequently found to protect mice from ecotropic but not non-ecotropic MLV infection. Further analysis showed the expression of a novel retroviral envelope protein on the cell surface of $Fv4^r$ cells

which was shown by immunological characterisation to be a gp70 molecule related to the SU protein of ecotropic viruses[117]. This prompted the speculation that this gene may be acting by the mechanism that mediates superinfection resistance[118]. For many years it has been known that the infection of a cell by one virus frequently restricts the infection of the cell by a second virus, and this has been termed superinfection resistance, or viral interference. This phenomenon is thought to be caused by the SU proteins, produced as a consequence of the initial infection, blocking the cell receptor and so preventing subsequent infection of a second virus that uses the same receptor. It was therefore suggested that the *Fv4* gene might be derived from a provirus, and its product may be an SU derivative which mediates resistance to ecotropic virus by blocking the ecotropic receptor (CAT-1) in a similar fashion[117]. The cloning of this gene has shown this to be true. *Fv4* is a truncated MLV provirus consisting of ~850bp of the 3' end of the *pol* gene, a complete *env* gene and a 3' LTR. To demonstrate that this was the genetic basis for *Fv4* resistance, ecotropic restriction was shown to be conferred to fibroblasts by the transfection of this deleted provirus[119, 120].

Any animal that expresses endogenous retrovirus has, in theory, the capacity to control infection in this manner, and it has been suggested that host restriction of retrovirus in this way might be a common mechanism for limiting virus spread[3]. There is a least one other example in mice, the *Rmcf* (*r*esistance to *m*ink *c*ell *f*ocus-forming virus) locus[121]. Several avian proviral loci also appear to have similar effects[122] and there is some evidence for this activity in cats[123].

Another way to control retroviral infection at the level of the receptor is to affect receptor expression. Efficient retroviral transduction of cells can be critically dependent on cell surface virus receptor expression levels[124]. In chicken, the *tva* gene encodes the receptor for subgroup-A viruses, *tvc* controls infection by subgroup-C and different alleles of *tvb* encode receptors for subgroups B, D and E. Therefore the inheritance of these genes will clearly influence the ability of ALVs of these different subgroups to infect chickens. In mice, NZB and AKR strains carry genes that restrict replication of Cas-Br-E MLV in the CNS and not in lymphoid organs[125, 126]. This could be controlled by differential expression of receptors, or endogenous proviral sequences that block the receptor, in these mice.

1.4.2 Target cell

In retroviral diseases that involve interactions with particular cell populations, the evolution of host responses that alter the frequency of viral target cells or the cell cycle kinetics of these cells has been seen.

Perhaps the best characterised of this type of gene is the *Fv2* locus[127, 128]. This gene affects the susceptibility of mice to erythroleukaemia by influencing the frequency and cell-cycling of late BFU-E and CFU-E, 2 classes of erythropoietin-sensitive erythroid precursors. The erythropoietin receptor itself had been excluded[129] even before it was mapped and cloned[130, 131]. However, there was evidence to suggest that this gene may code for part of the receptor complex: expressing a constitutively active form of the receptor causes disease only in $Fv2^s$ mice[129]. Moreover, deletions in the ecotropic-specific region of the retroviral envelope protein (gp55) of the SFFV component allow Friend virus to circumvent *Fv2*-mediated resistance by interacting with the erythropoeitin receptor in a way that bypasses events involving the Fv2 product[132-134], although such mutants are less active than wild-type gp55 in $Fv2^S$ mice. The *Fv2* product was therefore originally thought to be complexed with the Epo receptor in a way that blocks the interaction between the viral gp55 and the receptor in $Fv2^r$ mice. With the cloning of *Fv2*, the gene was found to encode Stk, a member of the Met subfamily of receptor kinases[135]. Susceptibility to Friend disease (Fv^{ss}) is conferred by the expression of a truncated form of this Stk (Sf-stk) that lacks an extracellular ligand-binding domain. It therefore appears that it is the $Fv2^s$ allele that encodes the positively-acting molecule, which is not expressed in resistant mice, that is required for gp55-mediated activation of the Epo receptor.

While other genes have also been shown to affect susceptibility to viral infection by modifying the frequency of target cells, the question of whether this activity has been positively selected for or is simply a chance by-product of a certain mutation is much less clear. Mice carrying the *W* and *Sl* loci show resistance to Friend virus-induced erythroleukaemia, and this has been shown to be due to the reduction of Friend virus target cells[136]. The *W* locus corresponds to the c-*kit* proto-oncogene[137], which encodes a receptor protein tyrosine kinase to stem cell factor, a molecule required for the differentiation of haematopoietic cells. Stem cell factor is encoded by the *Sl* locus itself[138]. These 2 loci therefore encode mutants of the receptor and its ligand for haematopoiesis, both of which cause a reduction in the number of Friend virus target cells. These mutations do not

however affect the lymphoid arm of haematopoiesis, and these mice still succumb to lymphoid tumours.

Mice homozygous for the nude (*nu/nu*) mutation have been found to be resistant to the development of thymic lymphoma[139]. These mice have mutations in the gene affecting lymphoid cell development and subsequently lack a normal thymus and the normal complement of T cells. These mice are also resistant to MAIDS[140] and MMTV-induced mammary tumours[141], diseases in which B-cell:T-cell interactions play a crucial role. Nude mice also show resistance to neurological disease induced by ts1 Mo-MLV[142] but not Cas-Br-E[143], a difference which may reflect the part played by the thymus in the replication of ts1 Mo-MLV but not Cas-Br-E MLV.

SCID (xid/xid) mice also show resistance to MAIDS and MMTV-induced mammary tumours[144, 145]. These mice carry a mutation in a nonreceptor protein tyrosine kinase called BTK and lack mature B and T cells because of a defect in the catalytic subunit of the DNA protein kinase involved in antigen receptor gene rearrangement[146, 147]. SCID mice do, however, develop thymic lymphomas because the microenvironment in the thymus is normal, and the animal contains the normal numbers of T-cell precursors[148].

While many mutations like these in mice do provide immunity to retroviral infection, the overall deleterious affect to the fitness of the organism suggests that this immunity more likely results as a consequence of, and not from selection for, such mutations. However, such immunity may facilitate such deleterious mutations to become fixed within populations.

1.4.3 Immune response

The role of the immune system in controlling retroviral infection has long been known. The ability of many retroviruses to cause disease in neonatal animals but not in mature animals highlights this fact. Newborn animals are not able to recognise retroviral proteins as foreign and fail to mount an effective humoral or cellular response against them. Mature animals, on the other hand, mount a vigorous cellular response against many retroviruses, producing cytotoxic T lymphocytes that can protect against disease; a vigorous humoral immune response then follows. The genes that are involved in the immune response usually affect particular retroviruses (reviewed in[149] for Friend virus), but the

mechanisms by which they do this are involved in a wide range of infectious diseases.

Several MHC genes have been shown to affect susceptibility to a variety of retroviral infections. The region that encodes the MHC genes in mouse is called H-2, and is analogous to the HLA region in humans. The *Rfv1* locus[150] was found to influence the susceptibility of mice to Friend virus-induced disease, and has been shown to be an allele of H-2D, a structural gene that encodes an MHC class I molecule. The consequence of this mutation is to increase the effectiveness of the CTL response mounted against the Friend virus, and alters the outcome of the disease in adult mice[151, 152]. Another gene prevents the development of MAIDS by a similar mechanism, although this gene remains to be identified[153]. The *Rfv2* locus corresponds to the Q/TL region of the MHC locus, but just how this locus affects Friend virus-induced disease is not known[154]. Another effect of the MHC locus has been mapped to the Aβ class II molecule, which has been shown to influence the T helper cell response to the Env glycoprotein encoded by Fr-MLV[151]. Resistance to radiation leukaemia virus (RadLV), Gross MLV, and the neurotropic MLVs has also been mapped to the MHC region[125, 155, 156].

Several non-MHC-linked genes have also been implicated in influencing the immune response to retroviral infection. The *Fv3* locus has been mapped to chromosome 15[157], and mice carrying the *Rfv3*^{r} allele are resistant to Friend virus-induced disease; these mice are able to mount a particularly vigorous antibody response[158], although the mechanism by which the locus influences antibody production is unknown. Another *Fv* gene, *Fv5*, controls the types of erythropoiesis induced by Friend virus, perhaps by controlling the rates of proliferation of late erythroid cells[159]. Mice homozygous for the *Fv5*^{a} allele develop rapid and transient anaemia, whereas those of the *Fv5*^{r} genotype suffer polycythaemia. An unmapped gene has been shown to modulate the latency period of MLV-induced neurologic disease[160]; as with the *Rfv3*^{r} allele, mice carrying this gene are also able to mount a more vigorous antibody response. The *Fv6* locus is known to govern susceptibility to early erythroleukaemia induction by helper-independent Friend leukaemia virus, since other types of leukaemia develop late in resistant mice[161]. The mode of action of this gene is also unknown.

1.5 The *Fv1* MLV-restriction gene

1.5.1 The discovery of *Fv1*

Genetic resistance to Friend virus was first observed in a screen for susceptibility of inbred mice to Friend virus-induced erythroleukaemia[162]. The results of crosses between the resistant C57BL/6 strain and the susceptible RF strain led to the conclusion that this phenotype was a result of a single autosomal gene, and that the susceptibility allele (Fv^s) was dominant over the allele for dominance (Fv^r[163]). However, it became clear that the situation was more complex than had previously been thought. It was found that, using the spleen focus assay[164], the F1 hybrids of resistant C57BL/6(Fv^r/Fv^r) x susceptible DBA/2(Fv^s/Fv^s) showed an *intermediate* susceptibility. This was also true of BALB/c mice, even though these are supposedly homozygous at all autosomal loci[165]. This indicated that more than one gene was involved in susceptibility to Friend virus. The situation was clarified in studies of the inheritance of susceptibility as measured by spleen focus-forming property of Friend virus using 2 isolates of Friend virus, the original F-S and a variant, F-B[127]. The complex pattern of inheritance of susceptibility to Friend virus was subsequently found to be due to 2 independently-segregating genes, named *Fv1* and *Fv2*. *Fv1* was responsible for relative resistance to the F-S isolate, where the allele for resistance $Fv1^r$ was dominant over the susceptibility allele, $Fv1^s$. The gene had no effect over the outcome of infection by F-B virus. *Fv2*, on the other hand, influenced the outcome of infection by both the F-S and F-B isolates, with the allele for susceptibility being dominant. As mice homozygous for the resistance allele of *Fv2* were found to be resistant to focus-formation by SFFV but not to the lymphatic component (the Friend MLV, which could be isolated from $Fv2^r$ mice under certain conditions), *Fv2* appeared to act on the SFFV component of Friend virus[103]. It was this gene that was identified originally by Odaka.

1.5.2 The main *Fv1* alleles

The replication of MLV in various cell types was investigated using a tissue culture assay, developed to study viral replication[166]. It was found that almost all MLV isolates, both from the laboratory and the wild, fell into 1 of 3 categories according to their ability to replicate in NIH Swiss and BALB/c mouse embryo cultures. One group replicated 100-1000 fold better in the NIH cells than BALB/c,

and were thus termed N-tropic. With a second group, the opposite was true, and these were termed B-tropic. In the third group, the viruses were able to replicate equally well in both cell types, and these were called NB-tropic virus. Subsequent studies using this tissue culture assay showed that a single genetic locus, with 2 alleles, was the major determinant in the mouse cells for *in vitro* infection with MLV[167]. All cell types tested in this assay resembled either the NIH Swiss or BALB/c with respect to replication patterns following MLV infection. Resistance was found to be dominant with regard to both N- and B-tropic virus (i.e. co-dominant alleles). It became clear that this genetic locus corresponded with, or was closely-linked to, the *Fv1* gene[168]. The *Fv2* allele, which had previously been found to determine sensitivity to spleen focus induction *in vivo* by both F-S and F-B variants of Friend MLV, had no effect *in vitro*. This seemed to indicate that the fibroblasts used in the assay did not express *Fv2*, which may be limited to the haematopoietic target cells of the virus. *Fv1* alone influenced the *in vitro* assay by acting on the helper Friend MLV component of the Friend virus complex[103]. In the assay, F-S virus was found to be N-tropic and F-B virus was NB-tropic, which explained why *Fv1* had no effect on infection with the F-B virus[127]. The alleles of *Fv1*, $Fv1^r$ for resistance and $Fv1^s$ for sensitive, were originally designated using the N-tropic F-S virus; once the *Fv1* gene system was known to contain co-dominant alleles for resistance, they were renamed $Fv1^b$ (susceptibility to B-tropic virus) and $Fv1^n$ (susceptibility to N-tropic virus). As the *Fv1* phenotype was found to be readily detectable both in mice and cultured cells, it was evident that the mechanism for resistance must operate on the cellular level, and does not depend on mechanisms requiring more complex cellular, immunological or hormonal interactions[169].

1.5.3 The cloning of *Fv1*

The *Fv1* gene was cloned using a positional approach[170]. Candidate YACs derived from C57BL/6J ($Fv1^b$) were identified by PCR for the presence either of the markers *Xmv9* and *Nppa*, believed to lie within 1.2 Mb of *Fv1*[171, 172]. These YACs were then tested for biological activity by transfecting into L cells and challenging transfected clonal cell lines with N-, B- and NB-tropic virus. As the L cell line was derived from an $Fv1^n$ strain of mouse, cells transfected with the YAC harbouring the $Fv1^b$ allele would restrict both N- and B-tropic virus, while those transfected cell lines containing YACs that do not contain this gene would restrict

B-tropic virus only. In this way the *Fv1*^{b} allele was associated with YAC Nppa D11, then to a cosmid derived from this YAC, and finally identified by sequencing a 6.5kb *SpeI-EcoRI* fragment of this cosmid. The *Fv1*^{n} allele was isolated subsequently from a genomic AKR lambda library.

1.5.4 Structure of the gene

The structure of the 2 main alleles of *Fv1* are shown in Figure 1.6. Sequence analysis revealed that the *Fv1*^{b} gene consists of a single ORF encoding a protein product of 459 amino acids in length and a predicted relative molecular mass of 52K. There are 2 gross changes between the 2 main alleles. Firstly, the *Fv1*^{n} allele contains an IAP element inserted within the first of a B2 repeat. Secondly, this allele contains a 1.3kb deletion that results in the removal of the C-terminal 22 amino acids and substituting instead 3 different amino acids. There are also 2 amino acid changes between the main alleles at positions 358 and 399 (marked on Figure 1.6). The polyadenylation signal used by either allele is also different: with *Fv1*^{b}, the signal is located in the second of the two B2 elements located 1.8kb downstream of the ORF. The *Fv1*^{n} allele uses the polyadenylation signal found in the 5′ LTR (R-U5 boundary) of an IAP element inserted into the first of the B2 repeats. Both alleles use the same host promoter.

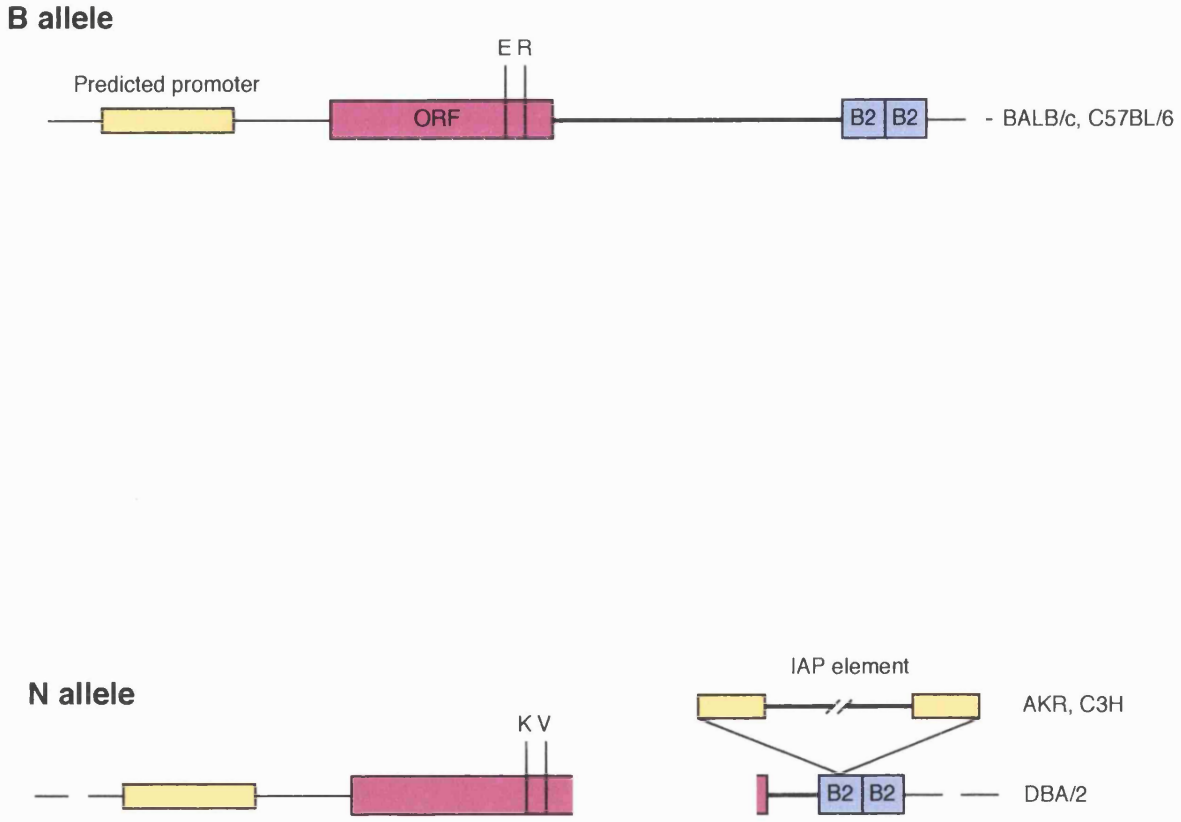
1.5.5 Other alleles of *Fv1*

Fv1^n and *Fv1*^{b} represent the 2 main alleles of *Fv1*. There are, however, other alleles:

***Fv1*^{o}**: This is a null allele found in wild mice[173]. Analysis of *Mus spretus* and *Mus musculus domesticus* (formerly *praetextus*) wild mice revealed that these mice were unable to restrict either N- or B-tropic virus, and the results of genetic crosses showed that this was due to a novel, null allele of *Fv1*.

***Fv1*^{null}**: Certain cell lines appear to have lost *Fv1* restriction. 3T3FL is a cell line derived from a Swiss mouse embryo[174], and was originally designated as *Fv1*^{b}[175]. This cell line has since been shown to be sensitive to both N- and B-tropic virus, suggesting that *Fv1* has either been lost or is no longer expressed[176, 177]. Analysis of 3T3FL in this lab suggests that these cells would have been *Fv1*^{n} (S. Best, PhD thesis), and an explanation for this discrepancy is not clear. SC-1 is another example of a cell line that has lost *Fv1* restriction[178]. In this case, the cell line shows sensitivity to viruses of both tropisms despite having

Figure 1.6 The structure of *Fv1*



The 2 main alleles of *Fv1* are shown. The polyadenylation signal is found in the B2 repeat, except in most $Fv1^n$ alleles, such as AKR, which have the intracisternal A-type particle (IAP) inserted; in these cases the polyA signal used is in the 5′ LTR of the IAP. The amino acid positions marked correspond to those residues different between the $Fv1^n$ and $Fv1^b$ alleles. The $Fv1^n$ allele has an internal deletion causing the loss of the last 22 amino acids of the open reading frame and substituting instead 3 different amino acids. In DBA mice the IAP is missing, and there is an amino acid difference in the ORF.

been originally derived from an *Fv1*^{n} embryo culture. The *Fv1* ORFs from both these cell lines was cloned and sequenced, and were found to contain no mutations that could account for this loss of function (data not shown).

Fv1^{nr} : This allele has been reported in several strains, including RF/J, NZB, NZW, SKIVE/Ei, 129 and Czech I and II[173, 179], and results in a modified *Fv1*^{n}[173]. The gene itself contains a single amino acid difference in the ORF compared to the *Fv1*^{n} allele and this alone is thought to be responsible for the modified phenotype (P. Le Tissier, unpublished data). This allele confers restriction not just to B-tropic virus but also to some N-tropic viruses.

Fv1^{d} : This allele is also an *Fv1*^{n} variant and is found in strains such as DBA/2[180]. In these strains, N-tropic viral titres are slightly lower and B-tropic viral titres are slightly higher than in *Fv1*^{n} cells. This gene was reported to only have one change from the *Fv1*^{n} allele, namely the lack of the IAP element in the downstream B2 repeat [170] (see Figure 1.6). However, my subsequent analysis showed that *Fv1*^{d} also contains a single amino acid difference in the ORF, and the modified phenotype mediated by this allele seems therefore to be due to this change alone.

1.5.6 Effects of *Fv1* restriction in the cell

Much work has been done to narrow down precisely which step(s) of the retroviral life-cycle is affected by *Fv1* restriction, with the hope that this would elucidate its mechanism of action.

1.5.6.1 Entry of the virus into the cell

Studies involving the measurement of adsorption of N- and B-tropic virus to *Fv1*^{n} and *Fv1*^{b} cell types indicated that *Fv1* did not act by preventing receptor binding of the virion[176]. Pseudotyped virus studies showed that VSV pseudotyped for N- or B-tropic MLV was able to replicate equally well in both *Fv1*^{n} and *Fv1*^{b} cells[181]; this indicated that the MLV envelope proteins were not involved in host-restriction, and so this event must occur intracellularly. This was confirmed by the demonstration of intracellular (radiolabelled) viral RNA following the infection of both *Fv1*^{n} and *Fv1*^{b} cell types with N- and B-tropic virus[182]. Comparable levels of viral genomic RNA was present in restrictive and permissive infection, demonstrating the penetration of the viral core into the cytoplasm was not affected by restriction.

1.5.6.2 *Fv1* and events after integration

To investigate whether *Fv1* effects steps in the retroviral life-cycle post-integration, total DNA from permissive infections was transfected into restrictive cell lines[183]. In these transfection experiments, restriction of the viral DNA did not occur and a productive infection was observed. The same result was achieved by microinjecting viral DNA into restrictive cells[184]. This had 2 main implications: firstly, this finding indicated that the *Fv1* product could not act upon the 'naked' viral DNA genome directly, but instead this has to be in its native form, complexed to viral proteins, to be recognised and restricted by *Fv1*. Secondly, this finding indicated that *Fv1* restriction does not operate at any of the steps after integration of the virus i.e. the transcription, translation, viral assembly or budding stages. This second point was reinforced by work involving chronically-infected mouse cells[185]. *Fv1* restriction can be overcome at high multiplicity of infection (MOI). In this way, a clonal virus-producer cell line was made from BALB/3T3 ($Fv1^b$) cells infected at a high MOI with an amphotropic N-tropic MLV. Upon super-infection with an ecotropic N-tropic MLV, *Fv1* restriction was still maintained and infection by this second virus was prevented. This could not be due to viral interference, as both viruses use different receptors (see section 1.4.1). This data indicates that *Fv1* has no effect on virus expression once integrated, and restriction must therefore occur at some point prior to this event.

1.5.6.3 Viral DNA levels in restrictive infections

After penetration of the viral core into the cytoplasm, the RNA genome is transcribed into DNA. RT levels were measured following the infection of both restrictive and permissive cell lines as an indicator of virus production in these infections. These levels, which are due to the transcription of the integrated viral genome and therefore represent the endpoint of infection, were found to be 70-100 times lower in restrictive cells than in permissive cells[186], indicating a suppression of viral synthesis. Levels of unincorporated viral DNA were found at normal levels in both restrictive and permissive cells[187]. The integration of this viral DNA, however, was detectable in permissive infections only. This indicates the *Fv1*-mediated block in the retroviral life-cycle occurs at some point after the reverse-transcription of the RNA genome into DNA, but before the integration of the genome into the host genome.

When analysed in more detail it was found that, although levels of linear viral DNA were present in normal levels in restrictive infections, levels of circular viral DNA was severely reduced[188, 189]. Furthermore, the apparent rate of nuclear import of linear DNA was found to be the same in both permissive and restrictive infection and, once in the nucleus, was not degraded preferentially in restrictive cells[190]. At the time, it was assumed that the 2-LTR circular DNA was the precursor to integration, and so the 25- to 50-fold reduction observed in resistant cells was thought to mediate restriction[188]. Today, however, we know that it is the linear viral DNA, not the circular forms, that are the precursors of integration[34]. Although linear viral DNA is reduced in the cytoplasm in restrictive cells in a fashion that appears to correlate with degree of restriction, the level of linear DNA in the nuclear fraction appears to remain the same as in permissive cells[191]. Moreover, this DNA seems to be fully integration-competent in *in vitro* integration assays. The absence of 2-LTR circular DNA, however, suggests that the preintegration complex may be prevented from physically entering into the nucleus, and may remain instead associated with the nuclear membrane. In this way the reduction in nuclear circular viral DNA is a by-product of restriction, and although it is more difficult to see why, this may also be true for the reduction in cytoplasmic linear DNA also. Alternatively, the effect of *Fv1* may be manifested in more than one place i.e. preventing nuclear import and DNA synthesis. No clear picture of restriction, however, emerges from these experiments.

1.5.7 Interaction between *Fv1* and virus

Exactly how *Fv1* acts has been the cause of speculation since the discovery of its restrictive phenotype. In order to gain more understanding of this process, much work has been done to investigate virus interactions with *Fv1* in the cell.

1.5.7.1 Viral determinants of N- or B-tropism

Ecotropic MLV can be classified according to its susceptibility to *Fv1* as either N- B- or NB-tropic, so that N-tropic virus can replicate in the presence of $Fv1^n$ but not $Fv1^b$, B-tropic virus can replicate in the presence of $Fv1^b$ but not $Fv1^n$, and NB-tropic virus can replicate in the presence of either alleles. NB-tropic virus is not a natural isolate but a product instead of the forced-passage of a single-tropism virus through a resistant cell line. Analysis of the electrophoretic mobilities of

MA, CA and envelope proteins of the dual tropic virus isolated from these experiments showed that, while MA and the envelope proteins remained the same, the CA protein had changed, as shown by its more rapid migration on a SDS polyacrylamide gel[192, 193]. This indicated that the viral determinant of *Fv1* sensitivity may lie in the CA protein. Sequence analysis of the CA coding region has revealed 2 amino acid differences between N- and B-tropic viruses, at positions 109 and 110[194, 195]. Subsequent mutagenesis studies have shown that the second change alone may be involved in N- and B-tropism[180]. The situation with NB-tropism appears more complex. Change in either of these amino acid positions is not responsible for dual-tropism. Instead, these viruses have a number of amino acid differences throughout the CA, and the changes involved in NB-tropism specifically have yet to be resolved.

1.5.7.2 Interaction with the preintegration complex

The finding that the viral determinants of *Fv1* restriction lie in CA suggests that the *Fv1* gene product interacts directly with this protein during restriction. This seems logical, since CA protein is known to be present in the preintegration complex (see section 1.2.2), and *Fv1* restriction appears to act by blocking the integration reaction. This interaction is reversible: preintegration complexes from restrictive infections can integrate as effectively in *in vitro* assay systems as those obtained from permissive infections[191]. Restriction therefore cannot be a result of cleavage, or some other irreversible physical modification.

1.5.7.3 Involvement of cellular factors

There is much data to suggest that host proteins may play a role in retroviral integration. The integration of the yeast retrovirus-like Ty3 element has been shown to require host transcription factors TFIIIB and TFIIIC *in vitro* and *in vivo*[196]. Similarly, the host factor Ini1 (IN interactor 1) has been shown to specifically bind and stimulate the *in vitro* joining activity of HIV-1 IN[197]. This protein has homology to the yeast transcription factor SNF5 and is a component of the analogous mammalian SWI/SNF complex that can remodel chromatin. Little is known about the function of Ini1 in mammalian cells, but further work will determine whether this type of interaction is specific to HIV or if it is common to other retroviruses. In MLV, integration has been shown to involve nucleoprotein complexes, termed intasomes, at the viral DNA termini which

include host proteins[198, 199]. The function in the intasome of one of these host factors is to prevent autointegration of the viral DNA. Viral preintegration complexes of Moloney murine leukaemia virus (Mo-MLV) exhibit a barrier to the autointegration reaction responsible for the third class of circular viral DNAs mentioned in section 1.2.5. This barrier can be disrupted by the treatment of the complexes with high salt, and subsequently restored by the addition of host factors provided by host cell extract[198, 200]. This indicates that host protein(s) present in cytoplasmic extracts from uninfected cells can efficiently prevent autointegration. This mechanism may also be present in other retroviruses. The barrier-to-autointegration factor (BAF) has been isolated and corresponds to a novel, highly conserved molecule of 89 amino acids[201]. An analogous host BAF in humans that prevents autointegration of the HIV preintegration complex has also been found[202]. Like Mo-MLV, HIV preintegration has also been found to preferentially integrate into a target DNA as opposed to autointegration[200, 203]. The solution structure of this 21,000 Mr protein has been solved by NMR[204] and has shown that, although this protein shows no sequence similarity to any known protein families, the topology of the protein is similar to many DNA-binding proteins. A host protein, HMG I(Y), has also been shown to be a vital component of the HIV preintegration complex[203].

Whatever cellular factors are involved in the MLV integration reaction, their presence and role in the preintegration complex make them important with regard to *Fv1* restriction. Not only are they to be found at the presumptive site of restriction, their function as a cofactor in the process of integration make them or their interaction a potential target for the *Fv1* gene product.

1.6 Reverse Transcriptase and its significance to eukaryotic genomes and the evolution of retroviruses

Fv1 is a gene now known to be the byproduct of an integration event by an ancient retrovirus around 10 million years ago. An event of this kind is not uncommon: the DNA of higher eukaryotes is literally awash with the product of retroelements and retrosequences. In the past these elements responsible, collectively termed the retroposons, have been thought of as "selfish" or "junk DNA", simply genomic parasites that are no more than a molecular burden to be endured by the genome. In the following section, the affect these elements have on the genome will be discussed, as well as the idea that, far from being a burden, these elements may play a fundamental role in evolution and genomic plasticity. In this way, *Fv1* may be put in a more accurate evolutionary context, not simply an extremely rare and unusual example of 'junk-DNA-made-good', but an example of just one product of a much more widespread and ongoing process that has been fundamental in the shaping of the genome during the course of evolution, both in the past and in the future.

1.6.1 The origin of RT

Over twenty years ago, the discovery of reverse transcriptase changed the fundamental concepts of information exchange[42, 43], allowing genetic information to flow not just from DNA to RNA, but backwards, from RNA to DNA. This finding was very important for 2 reasons. Firstly, it explained the replication of certain viruses: replication by reverse transcription is found in 4 virus groups: the animal-infecting retroviruses and hepadnaviruses, and the plant-infecting caulimoviruses and badnaviruses. The second implication of the existence of this activity was that there appeared to be a mechanism for the shift from an RNA-based system, the so-called RNA world, to one revolving around the double helix.

Since the discovery of this protein, reverse transcriptase has been found associated with a wide variety of viral and nonviral sequences: LINE elements, Ty elements, R2Bm, group II introns, and telomerases[74]. Despite the finding of this protein in so many different organisms, all reverse transcriptases reported were present in eukaryotes only, and appeared to be absent in eubacteria and the archaebacteria. This indicated 2 possibilities: either reverse transcriptase emerged in eukaryotes after their divergence from archaebacteria and

eubacteria, or reverse transcriptase emerged before all 3 groups diverged, but was subsequently lost from the eubacteria and archaebacteria soon after these 2 diverged from eukaryotes. However, both of these scenarios were negated following the discovery in 1984 of the presence of an unusual extrachromosomal DNA called msDNA (multi-copy single-stranded DNA) in the bacteria *Myxococcus xanthus*[205]. This type of DNA was found to be encoded by a locus in the bacterial genome which codes for msDNA and a single ORF. Homology searches revealed that this ORF codes for an RT similar to the polymerase of retroviruses, and the conservation of the amino acid domains between the eukaryotic and eubacterial RT suggested that these were derived from a common progenitor. This has important implications, as this suggestion places RT at a time before the divergence of eukaryotes and prokaryotes, at a critical stage in evolution during which the unique activity of RT may have been instrumental in the transition from the proposed RNA-based life to the DNA world.

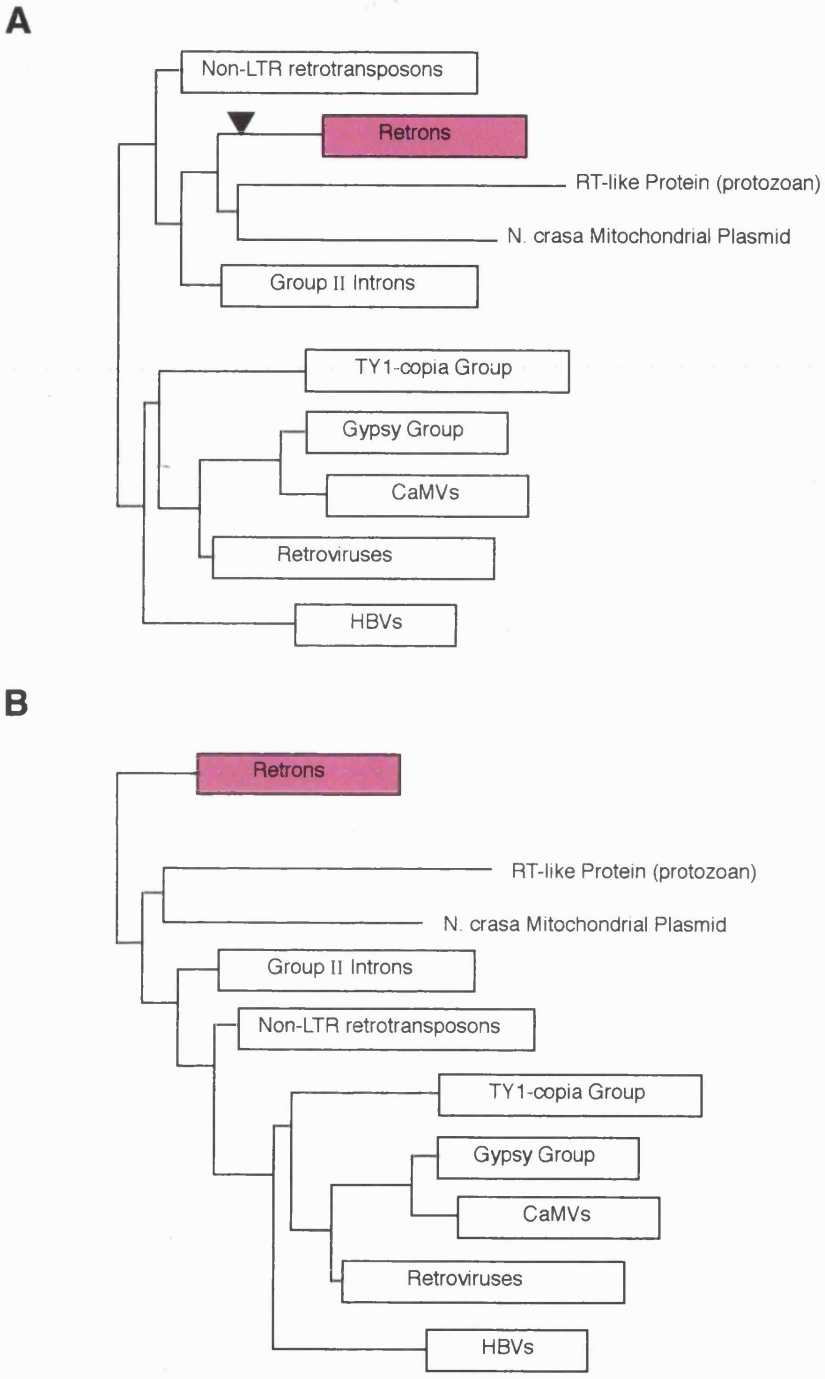
Coined originally by Walter Gilbert[206], the 'RNA World' was first described following the discovery of both informational and catalytic functions in the same molecule[207]. Before this discovery it was assumed the origin of life involved both RNA and protein, RNA to archive information and protein molecules to provide the catalytic activities required to make copies of the RNA in order to reproduce. With the discovery of self-cleaving and self-splicing RNA molecules, it now seems likely that these and other catalytic functions could also have been carried out by RNA alone. The evolution of the RNA world, it was suggested, started with RNA molecules performing the catalysis involved in self-assembly from the 'nucleotide soup'. Recombination and mutation would enable them to explore and develop new functions. The evolution of translational mechanisms enabled the synthesis of proteins, molecules that would not have novel functions but would have catalysed the same enzymatic reactions as existing RNA molecules but faster and more efficiently. Finally, DNA would have superseded RNA as a much more stable information carrier. RNA would then be relegated to the role of intermediary it has today, its 2 original roles replaced by the more efficient DNA and protein molecules. This model has been studied in several *in vitro* selection studies (reviewed in [208]), and whilst no conclusive answer as to its validity can be made, these experiments have shown the huge range of

functional activities that can be selected for and carried out by RNA. Therefore, this finding verifies the central role that RNA could play in such a scenario.
This model necessitates the emergence of reverse transcriptase activity, the appearance of which would mark the beginning of the end of the RNA world, as the vehicle for the movement from an RNA world to the DNA one we see today.

1.6.2 Where do retroviruses come from?

Howard Temin first suggested in 1970 that retroviruses evolved from cellular moveable genetic elements and that animal enveloped RNA viruses in turn evolved from retroviruses[43, 209]. The discovery of the retron-encoded bacterial RT in all myxobacteria, with an estimated maximum age of 1000 MY[210], made this the oldest RT known. This finding led Temin to suggest that this bacterial RT may be older than the separation of prokaryotes and eukaryotes[211], making the retron the possible ancestor of all retroelements. Inouye and Inouye[212] suggested that it was the acquisition of other retroelement genes such as *gag* and *int* during evolution that led to the emergence of the non-LTR retrotransposons. These elements represent an evolutionary step prior to the generation of retroviruses.
Kim *et al* demonstrated that the 'gypsy' retrotransposon in *Drosophila* is an infectious virus capable of infecting uninfected cells[213], suggesting that at the evolutionary stage of the *Drosophila*, retrotransposons acquired the ability to infect cells, in contrast to say the level of yeast cells, where retrovirus-like particles are purely intracellular. As winged insects appeared around 350 MY ago, it is possible that the oldest infectious retrovirus appeared around this time, possibly evolving from a lower form of retrotransposon that acquired an envelope gene. Transmissible retroviruses potentially could cross species barriers.
A different approach was taken by Xiong and Eickbrush[214]. These authors studied the evolutionary relationships of RT-containing retroelements from animals, plants, protozoans and bacteria to produce a phylogenetic tree. They also compared RNA-dependent RNA polymerases from various plus-stranded RNA viruses to reverse transcriptase and suggested that these 2 enzymes evolved from a common ancestor. According to their phylogenetic tree, the most probable common ancestor of current retroelements was a retrotransposable element with both *gag*-like and *pol*-like genes (see Figure 1.7A).

Figure 1.7 Phylogenetic trees based on the sequence of RT[214]



However, if their phylogenetic tree is redrawn without changing the relative distances between any 2 branches, but this time assuming that the ancestral reverse transcriptase is located on the retron branch, as seems more likely, then the tree shown in Figure 1.7B is obtained. In this scenario, the bacterial reverse transcriptases are the oldest group of retroelements, and the retrons are now considered to be the most primitive retroelements

A The sequence of RT in retroelements from animals, plants, protozoans and bacteria was used to construct a phylogenetic tree.

B The tree in A redrawn assuming the retron RT is the ancestral RT

One of the major branches has organelle and bacterial sequences (groupII introns and bacterial msDNA) that appear to have captured RT from retrotransposons that lack LTRs. In this scenario, therefore, bacterial RTs and retrons were derived from eukaryotes. On the other branch, acquisition of LTRs has given rise to 2 distinct groups of LTR retrotransposons and 3 groups of viruses (the retroviruses, the hepadnaviruses and the caulimoviruses).

The authors noted that if msDNA elements were the progenitors of all retroelements, as had been suggested[211], then the RNA viruses represent a category of retroelements whose polymerases have undergone substantial change from synthesising DNA to synthesising RNA. Due to its position on the tree, this change would have had to occur after the divergence of this group from the LTR retrotransposons and before the hepadnaviruses. If the tree is rooted such that the RNA viruses are an outgroup, as they suggest, then this "disjunction" in the tree is avoided. They believe the greater diversity in genomic organisation and sequences of the RNA viruses than any other branch of the tree, along with their presence in a wider range of prokaryotes and eukaryotes, also indicates they are older than the other retroelements. However, if their phylogenetic tree is redrawn without changing the relative distances between any 2 branches, but this time assuming that the ancestral reverse transcriptase is located on the retron branch, as seems more likely, then the tree shown in Figure 1.7B is obtained. In this scenario, the bacterial reverse transcriptases are the oldest group of retroelements, and the retrons are now considered to be the most primitive retroelements. This scenario is supported by the discovery of *Het-A* elements in *Drosophila*[215]. These elements, which are involved in the maintenance of telomeres (see later) have the hallmarks of both LTR and non-LTR transposons, and therefore look like an evolutionary intermediate of these 2 branches. The presence of this element therefore suggests a model involving the acquisition of LTRs. Also consistent with the view that retron reverse transcriptases are the most primitive of RNA-dependent DNA polymerase is the fact they do not require a separate molecule to prime the cDNA synthesis reaction but use instead a single RNA molecule as both template and primer. The tree shows the progressive evolution of the retroelements, starting with the basic reverse transcriptase element, then the evolution of the non-LTR retrotransposons following the acquisition of *gag*-like and *int*-like sequences, and subsequently the emergence of the major LTR branch following

the acquisition of this element. In this way, the revised tree now shows close agreement with the scenario outlined by Temin and others in the first part of this section.

1.6.3 The role of reverse transcriptase in the eukaryotic genome

Reverse transcriptase appears to have played an essential role during the transition from the RNA world to the DNA world of today, but what part has it played in the eukaryotic genome during its evolution? For a long time the intracellular role of RT has been assumed to be restricted to the self-reproduction of 'molecular parasites': the genome is virtually 'littered' with retroelements, sequences produced by reverse transcription reactions and subsequent integrations, that may occupy over a quarter of the human genome[216]. Since nonviral retroposition gives rise to many large pseudogene families and a great variety of retroelements provide no discernible benefit to the host, these elements have been viewed in the past as purely selfish DNA[217, 218]. It may be that RT has a role in the cell that are not yet discovered, the unavoidable byproducts of which might be the generation of such retroelements and retrosequences. If, however, there is no functional RT in the cell and the enzyme is provided by infectious retroviruses/proviruses, non-infectious retrotransposons, or non-LTR elements, then why are these elements with no apparent cellular function in the cell being tolerated rather than inactivated? Why have mechanisms that allow retroposition persisted, despite their opportunistic use by retroviruses and the potential detrimental effects of all retroposition events on the host? For example, familial hypocholesteremia has been shown to be due to the deletion of exons in the LDL receptor between *Alu* elements [219], and acholinesterasemia has been attributed to the insertion of *Alu* sequence into the cholinesterase gene [220]. Retroposons are therefore insertional mutagens, and as such their activity can have advantageous, neutral and detrimental effects in the same way as point mutations generated during DNA replication. In this way, a certain level of retroposition may be selectively favoured. This will be discussed in the following sections.

1.6.4 The Retrogenes

Retrogenes are generated by the occasional reverse transcription of mRNA into a cDNA copy, followed by their integration into the genome. If this integration

occurs in the germline, then the retrogene is inherited in future generations. The 3 hallmarks of retroposed gene copies are the presence of short direct repeats at their flanks, A-rich regions at their 3' ends, and frequent lack of introns compared to the 'founder' gene[221, 222]. The retrogene is often devoid of promoter and other regulatory elements. This inactivity and consequent lack of selective pressure leads to the acquisition of nonsense mutations and 'indels' (insertion/deletions), and has led to the term *pseudogene* for these sequences. Many genes have spawned retrogenes, generally ranging from 1-10 copies, but occasionally reaching up to several hundred copies. One such example is the single intron-containing gene for glyceraldehyde-3-phosphate dehydrogenase which has yielded more than 300 intronless retrogenes in mice and rats[223].

Many genes encoding small RNAs, such as small nuclear RNAs or SRP RNA, have also given rise to several hundred pseudogenes, referred to as short interspersed repetitive elements (SINEs). These elements are extremely abundant (10^4-10^6), and occupy a large fraction of the genome (10-15%). Among these are the *Alu*-RNAs (in turn derived from SRP RNA) which have generated primate *Alu* and rodent B1 elements. Certain tRNAs are the founders of rodent B2, rabbit C, and canoid species Can SINEs, as well as, numerous vertebrate, nonvertebrate and plant SINEs. Among these SINEs are extreme cases that reach well over 10^5 copies. The higher general copy number of tRNA-derived pseudogenes over those derived from mRNA is thought to be due to differences of each in serving as templates for reverse transcription. It is possible that most repetitive elements are being generated from one or a few master genes[224-227]. The products of these master genes are likely to be functional, as indicated by their apparent conservation over tens of millions of years.

The majority of retroposons most likely have never been and never will be functional. However, their importance lays in their *capacity* to contribute a functional gene or regulatory element (see following section). The identification of active genes with retroposon ancestry suggests an important and still ongoing contribution of retroposition in evolution, in addition to gene duplication. This latter process has long been known to be important in evolution, as the presence of an extra gene copy releases one copy from selective constraint. The spare copy can be varied without being constrained to the former function. As one would expect, the many inactive pseudogenes are the genome's testimony that this process is far from infallible.

1.6.5 Retrogenes and Exaptation

A gene is far more than simply a unit of translation. Coding sequence can be broken down into various functional and structural domains. Noncoding sequence can contain both positive and negative regulatory elements, responsible for not only for quantitative gene expression but also appropriate spatial and temporal expression, sequence motifs for nucleosome positioning, DNA methylation, polyadenylation signals and determinants of RNA stability and transport. Any mechanism of gene duplication (gene duplication or retroposition) allows the possibility of recruitment of these genetic elements into a novel molecular context and as a result, a possible variant or novel function, an evolutionary process first recognised over a decade ago and termed *exaptation.* Unlike adaptation, which involves the change of features generated by natural selection that enhances fitness, exaptation involves the change of features that enhance fitness but were not generated by natural selection for their current role[228]. Many of these sequences will have little or no affect, and under no selection will accrue mutations over time and blend into the genome- these can be termed non-aptations. A considerable fraction of the non-coding sequence that lies within or between genes may well be the result of retroposition in the past whose identity is no longer discernible. These non-aptations can occasionally be exapted as novel control elements of a gene, or even as a novel exon, increasing the availability of protein sequence motifs[229-232]. Additionally, SINEs can provide necessary splice sites[233, 234]. Therefore, any retroposed sequence in the genome that lies dormant (but gradually changing by mutation) has the potential to be exapted and contribute to the genomes fitness at any time. With the growing amount of sequencing data, there are mounting examples among many eukaryotic lineages in support of this hypothesis, namely that retroelements and retrosequences are a large pool of recruitable genetic material (reviewed in [235, 236]). The resurrection of the Θ1 globin gene (by provision of a CAAT box promoter element by an *Alu* sequence) highlights the fact that, although the likelihood of a gene being reactivated after 6-10 million years of dormancy is small[237], this process can occur after as much as 200 million years of silence[234, 238].

1.6.6 Intronless genes: an indicator of retroposition?

As mentioned briefly in the previous section, however, there are problems in assigning retroposon ancestry to sequences due to the loss of the characteristic direct repeats and polyA tails, particularly when this event has taken place more than 100 million years ago. However, the discovery of such genes makes it plausible that other and possibly most intronless genes have been derived by retroposition. Moreover, it has been suggested that the entire genome of yeast, which has very few intron-containing genes, consists of retrogenes which have themselves replaced by homologous recombination the founder genes from which they were originally derived[239-241]. The discovery of retron-associated reverse transcriptase raises the possibility that this may also be true for bacterial genomes[211].

1.6.7 Retroposition and chromosome structure/function

There is growing evidence that retrotransposed sequences cannot only be exapted as genes or regulatory control elements, but may also play an important part in chromosome structure. They can act as carriers of high density CpG dimers in regional DNA methylation[242, 243], and (conversely) specific *Alu*-binding proteins have been shown to prevent DNA methylation[244]. For this reason, the methylation state of such repetitive elements may be important in gene regulation, including genomic imprinting. Moreover, SINEs can provide binding sites for nuclear proteins, and in this way can modulate chromatin organisation[245-247], which in turn may play a role in the determination of differentiation state[248]. These repetitive elements can act indirectly, inducing recombination between otherwise unrelated DNA fragments, leading to new combinations within genes and consequently to genetic diversity[234]. However, the discovery of the direct involvement of retroelements such as *Het-A* and *TART* in the maintenance of telomeres in *Drosophila* offers the first direct evidence of transposable elements with a clear role in chromosome structure[215, 249]. This role, which is performed by the cellular reverse transcriptase enzyme telomerase in mammalian cells, may represent the first of many *bona fide* cellular roles of retroposons, not just in *Drosophila* but other organisms also.

In conclusion, retroposition is a mechanism, alongside the more conventional gene duplication by recombination process, which may constitute an extremely important evolutionary process in the generation of genetic diversity. Even

pseudogenes can be recruited as functional genes, and parts of pseudogenes or SINEs can be integrated into coding regions as functional domains or novel regulatory elements. The vehicle of retroposition, reverse transcriptase, may still exert, as it did more than 3 billion years ago, a huge influence on genomic plasticity by not only creating novel genes but by allowing the mixing of existing genes with novel regulatory elements. Altering when, where and what amount of a gene is expressed can have a profound evolutionary impact, both positive and negative. The price paid for long-term evolutionary benefit this plasticity provides for the population is frequently paid by the individual, when a deleterious genetic combination produces genetic disease.

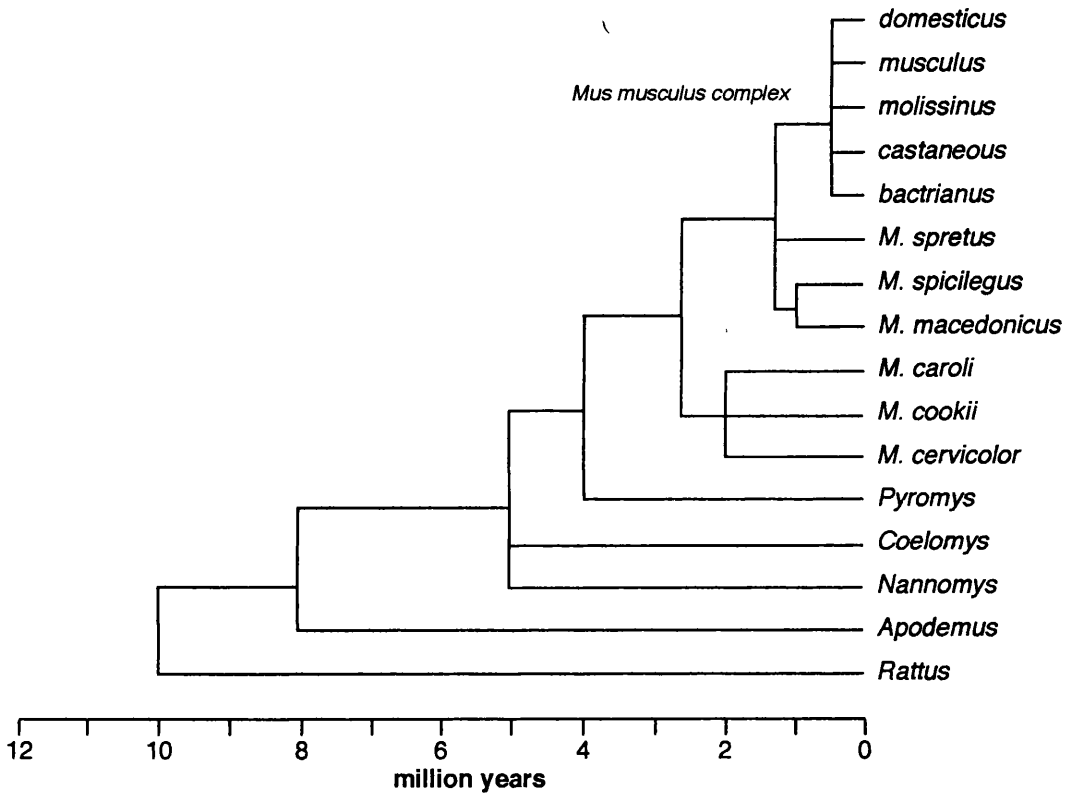
1.7 Evolution of the *Mus* genus

One of the objectives of this work is to offer an explanation of how and why the *Fv1* gene has evolved the phenotypes of the major alleles we see today. In order to do this, a basic knowledge of the evolution of the genus *Mus* is necessary. For this reason, the Introduction will finish with a brief account of the radiation and speciation of *Mus* genus.

1.7.1 Phylogeny of *Mus*

Despite the fact that the house mouse has been vital for so many aspects of fundamental science for many years, its wild populations (of the subgenus *Mus*) have been relatively poorly studied. Due to the few morphological characteristics that differentiate the wild mice, even the taxonomy of this subgenus as determined by classical approaches has been unclear. In older literature, this resulted in either minute descriptions of many species and subspecies or the oversimplification of a single polytypic taxon[250]. Within the last 2 decades, the advent of biochemical, karyological, molecular, and biometric approaches to study evolution has done much to shed light on the origin of *Mus.* The phylogenetic tree of the genus *Mus* in Figure 1.8 was obtained from allozymic data[251, 252], RFLP polymorphism data of various genes[253], mtDNA[254] and DNA/DNA hybridisation of single-copy nuclear DNA (scnDNA) data[255]. It is also consistent with the limited murid fossil record[256]. There are various levels of quasi-synchronous speciation events. The first corresponds to the separation of the subgenus *Mus* from the other subgenera. The second level is not indicated on the tree. This level refers to the separation of the Indian pigmy mice, *Mus dunni* and *Mus booduga* (also called *Leggada*), which also corresponds to the appearance of a 40 acrocentric chromosome karyotype that is characteristic of the whole *Mus* subgenus. These species cannot be placed accurately due to the lack of molecular data to precisely calibrate the split of this all-Indian offshoot of *Mus*, and therefore are not represented in the Figure. The third level corresponds to the radiation of the Asian species (*Mus caroli, Mus cookii, Mus cervicolor*) in India around 2 million years ago. The fourth level corresponds to the radiation of the West Palearctic species around 1 million years ago which have migrated as far as eastern and southern Europe and northern Africa. Although *Mus spicilegus* and *Mus macedonicus* have separated very

Figure 1.8 Phylogenetic map of the genus *Mus*311



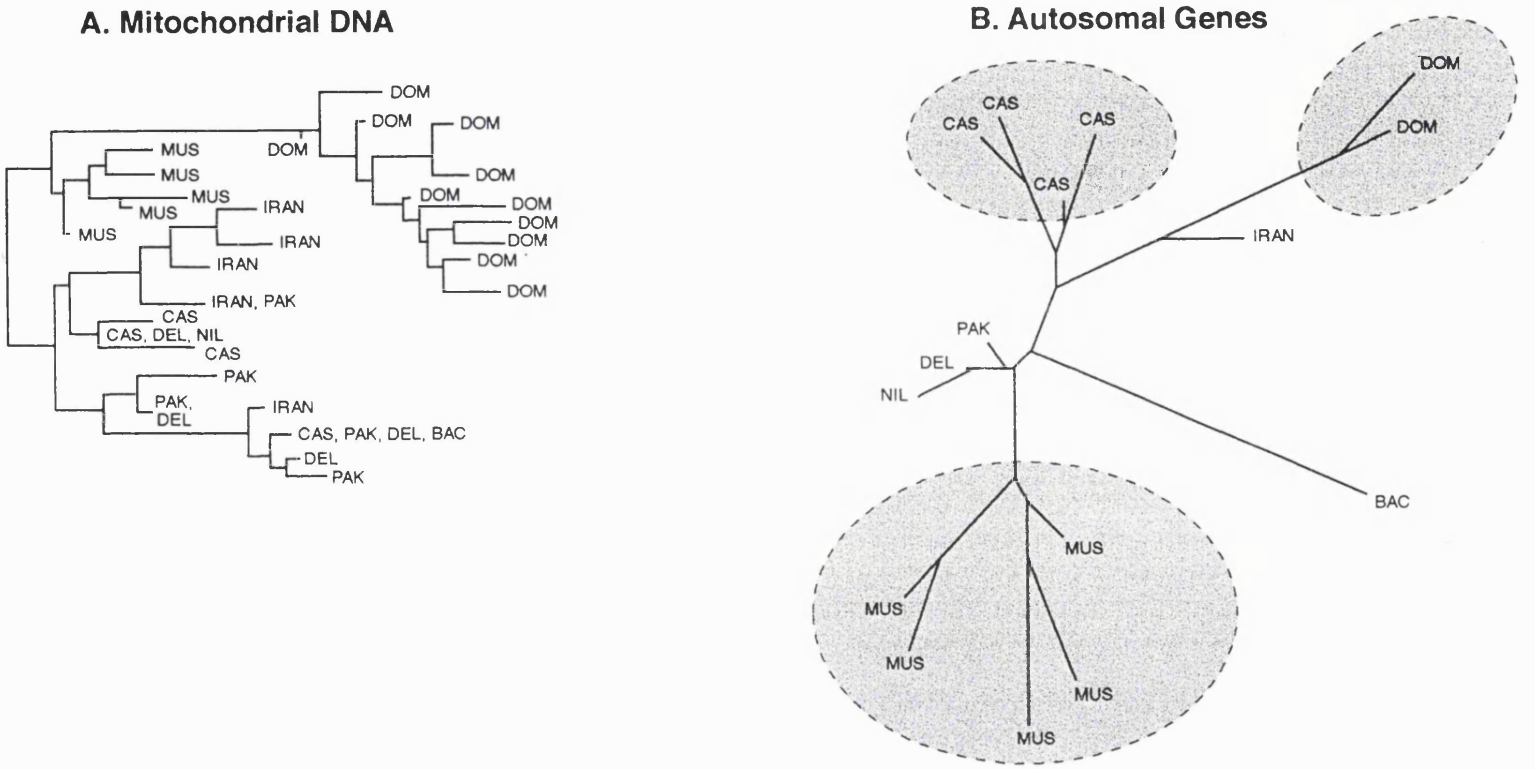
The current phylogenetic tree of the genus *Mus* is based on that of Bonhomme[279]. See text for explanation.

recently, the ancestor of these 2 species diverged quasi-simultaneously from *Mus musculus* and *Mus spretus*. The last level of radiation corresponds to the individualisation of the main subspecies of *Mus musculus* within the last 0.5 million years following its migration from the Indian subcontinent. The genus *Mus* therefore represents the classical image of a taxon that has actively radiated through geographical speciation, forming new species wherever migration is possible.

1.7.2 Partitioning in *Mus musculus*

There is considerable scope for confusion when analysing genetic characteristics in the *Mus musculus* complex. Studies of several types of genetic markers have been carried out in peripheral house mouse populations: analysis of chromosomal C-band patterns[257, 258], nuclear autosomal genes studied by protein electrophoresis[251, 252, 259-261] or by DNA RFLP[262], rDNA non-transcribed spacers RFLP[263], immunoglobulin[264-266], satellite DNA, serological studies[267], some Y chromosome sequences[268-271], mitochondrial DNA[272]. The proposed scheme of the 4 subspecies of the *Mus musculus* species in Figure 1.8 is based largely on the combinations of allele frequencies at many nuclear encoded protein loci[251]. This classification is supported by the study of mitochondrial DNA phylogeny[272b, 272d] (see Figure 1.9A) as 4 main lineages could be defined, corresponding to either *Mus musculus domesticus, musculus, castaneous,* or *bactrianus* (although *Mus musculus bactrianus* was found at low frequency in the *Mus musculus castaneous* territory). Other markers studied are compatible with this partitioning of the complex, but the picture is not clear as the phylogeny of the alleles is not known. Some of the markers appear to be diagnostic between pairs of subspecies whilst others are confined to one subspecies (reviewed in [251, 257, 276]). There is, however substantial variation within the subspecies from different geographical areas, for example between European and Asian *Mus musculus musculus* [258, 277] or between northwestern and southeastern *Mus musculus domesticus* [278]. There are no single characters in the *Mus musculus* complex that allows the partition of the peripheral populations, therefore, but combinations of characters do permit the subspecies to be clearly identified.

Figure 1.9 Phylogenetic trees of the *Mus musculus* complex based on different DNA sequences



A. Phylogenetic tree of the *Mus musculus* complex based on the sequence of mitochondrial DNA.

B. Phylogenetic tree of the *Mus musculus* complex based on the sequence of autosomal genes.

Nomenclature:

BAC	*M. m. bactrianus*	PAK	Specimen from Pakistan
CAS	*M. m. castaneous*	DEL	Specimen from Delhi
DOM	*M. m. domesticus*	NIL	Specimen from Nilgiri mountains
MUS	*M. m. musculus*		

1.7.3 Origin of *Mus musculus*

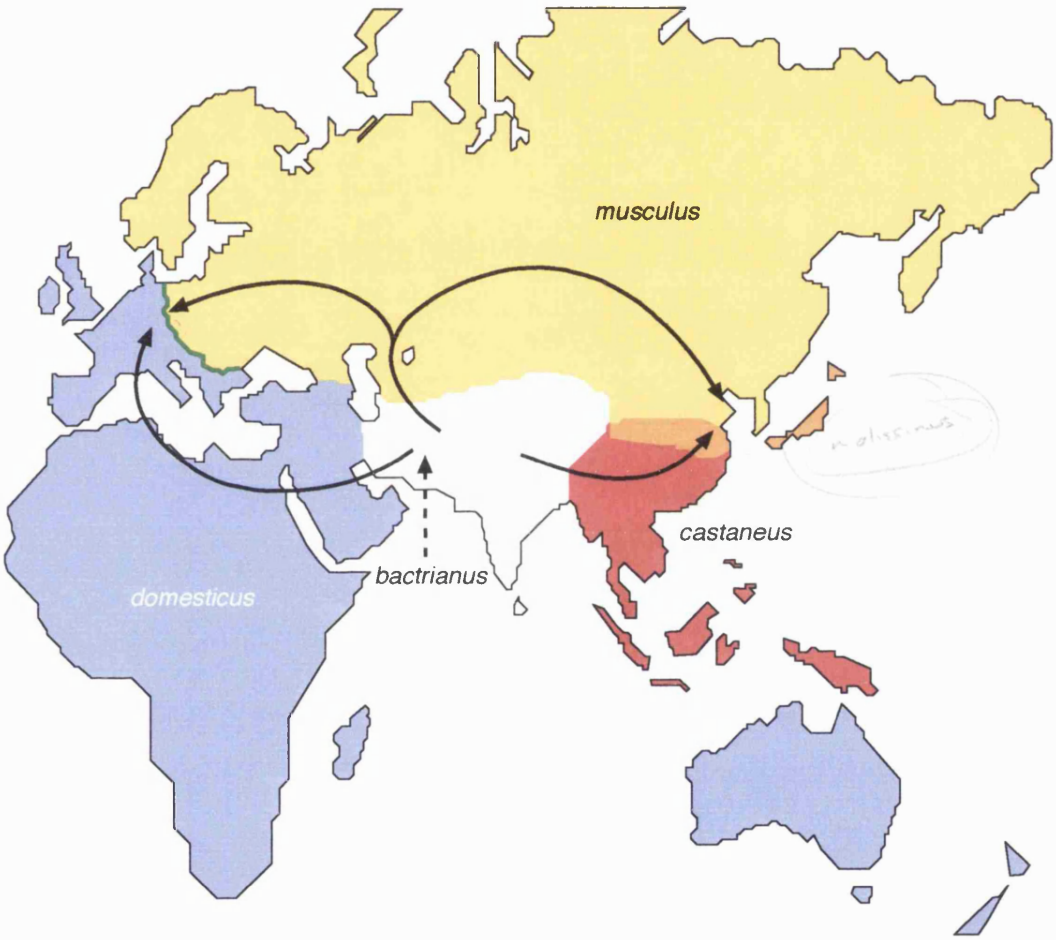
It is impossible to construct a phylogeny of these subspecies based on the distribution of individual genes, as several incompatibilities between characters become apparent[258, 276]. It therefore does not seem possible to account for the evolution of these subspecies by a simple dichotomous model of divergent populations, as mosaicism seems to be a general feature of the evolution of their genes. An explanation for this generalised genetic mosaicism emerged from the study of mouse populations from the Indian subcontinent[279]. When samples from Iran, India and Pakistan were studied it was found that they could not be classified into any of these 4 subspecies. Nor could they be considered as forming one more additional categories, although some samples had unique variants found nowhere else (either allozymes or mtDNA lineages defined by RFLP). The analysis of their allelic composition and distribution showed that these samples appear to be at the centre of a star phylogeny, with the peripheral populations at the tips (Figure 1.9B). The position of the Iranian sample, halfway on the branch leading to *Mus musculus domesticus,* is in accordance with its geographical location. These samples were also found to have very diversified mtDNA, including those previously found to be characteristic of *Mus musculus bactrianus* and *castaneous.* Therefore, these 2 subspecies do not correspond to their own monophyletic groups (for mtDNA) as found for *Mus musculus musculus* and *domesticus.* This led to the proposal that *Mus musculus* originated from northern India and its radiation led to the colonisation of the rest of the continent[279]. Moreover, the divergence times of the main mtDNA lineages indicated that this initial radiation occurred less than 0.9 million years ago, a figure supported by DNA/DNA hybridisation data[255].

The genetic data clearly shows that the 3 peripheral subspecies, *Mus musculus domesticus, musculus* and *castaneous,* correspond to 3 distinct paths of colonisation from the Indian cradle (see Figure 1.10). It is this initial radiation of *Mus musculus* occurring simultaneously that is likely to have produced the mosaicism in different genetic markers. This colonisation, which is now global, mirrors the spread of Man, and is a reflection of the commensal lifestyle of the house mouse.

1.7.4 Hybrid zones

The relatively recent expansion of the house mouse has produced zones of secondary contact between subspecies, called hybrid zones. The most well-

Figure 1.10 Geographical distribution in the Old World of the 3 main taxonomic units making up the *Mus musculus* complex.



The arrows represent plausible migration routes from the point of origin of the genus (the Indian subcontinent) during the radiation of the species. This expansion has occurred within the last 0.5 million years.

characterised of these is the transition between *Mus musculus domesticus* and *Mus musculus musculus* in Europe[271, 280-283]. Here, genes are exchanged in a 30-40km wide band between the 2 populations. Introgression from this zone is generally asymmetrical, with longer tails of introgression on the *Mus musculus musculus* side. The width of the zone and the pattern of introgression is consistent along its length. In Jutland, however, although different mtDNA haplotypes are found either side of the zone, they are all of *Mus musculus domesticus* origin[280, 283]. This hijacking of *Mus musculus musculus* mtDNA by *Mus musculus domesticus* is also found in Scandinavian mice and mice from East Holstein, and has been interpreted as the result of repeated founder effect through island hopping during the colonisation of Scandinavia by *Mus musculus musculus*[279].

Such an apparent conservation of hybrid zone width and introgression patterns suggests that the zone could be maintained by selective factors and is not simply a result of passive diffusion. Some attempts have been made to correlate the position of the zone with climatic factors[284, 285], but the climate gradient would have to be so sharp and the adaptation of both subspecies to their environments so precise that this seems unlikely. There is only indirect evidence that selection is acting on natural hybrids. Researchers have found that mice in the centre of the hybrid zone are significantly more infected by intestinal parasites than mice on either side of the zone[286], and that this susceptibility to these parasites has a genetic basis. The effect of these parasites on fitness is not known. Another indication of counterselection of the natural hybrids in the zone is found in the introgression patterns of sex chromosomes. The Y chromosome [282, 283], and at least certain parts of the X chromosome[271], introgress much less than the autosomal markers, resulting in a band only a few kilometres wide. This may be explained by the finding that certain combinations of Y chromosome on different genetic backgrounds induce problems of sex chromosome stability and sex determination in the F1s [287]. Interestingly, the Y chromosome variant that is fixed in the old inbred lab mice, which have hybrid genomes (see next section), is the same type found in the zones of natural hybridisation between *Mus musculus musculus* and *Mus musculus castaneous* in Japan[272]. In these extreme examples of genetic exchange between the *Mus musculus* subspecies, the colonisation of the Japanese archipelago resulted in at least 2 invasions paralleling those of Man. The first was by *Mus musculus castaneous* from the south and the second, by *Mus*

musculus musculus, from the north. These 2 subspecies subsequently interbred and founded the population we see today, referred to as *Mus musculus molissinus*, which possesses a dominance of *Mus musculus musculus* nuclear genes and a mosaic distribution of mtDNA. This interaction is also known to have occurred in continental China (shaded orange in Figure 1.10).

1.7.5 Laboratory mice

The commensal lifestyle of the house mouse is thought to have been adopted as a way to avoid competition with other murids by habitat partitioning[288]. Ultimately, this life-style has led to the domestication of the house mouse, first as a pet and later as a laboratory animal. As, over time, this animal has become so important to geneticists and developmental biologists, the origin of the laboratory mice, the so-called "old inbred strains" has often been discussed. They all have the same mtDNA haplotype, which is of *Mus musculus domesticus* origin [289]. Their nuclear genes in contrast show a higher level of divergence than expected given their recent derivation from a common ancestral stock[290], and this has led to speculation concerning the rate of evolution and mutation in these mice. It has been suggested that this divergence may be due to much higher rates of mutation occurring in the inbred lines, or that heterozygosity has been inadvertently selected for in the construction of the inbred lines by the selection of the most vigorous breeding pairs from each generation (to avoid inbreeding depression). Another explanation stems from the finding that these strains all carry a Y chromosome of *Mus musculus musculus* origin[268], or more precisely, the variant only found in Japanese mice and some Chinese mice[269, 271]. From the known pedigrees of laboratory mice we know that, in addition to the known inter-strain crosses that occurred early in their development, occasional introductions of unknown background have occurred, either by purchasing from specimens from pet dealers ('fancy mice') or by crossing with animals trapped from the wild. The presence of the Y chromosome of Asian origin confirms that the Oriental fancy mice has contributed to the ancestral stock of laboratory mice, through the early introduction of these types of mice (Japanese waltzing mice, for example) into European collections. In this way, the founder stock of the laboratory mouse was a combination of very divergent genomes[257].

1.8 Aims of the project

At the start of the project, *Fv1* activity had been isolated to a 6.5kb *SpeI-EcoRI* fragment, but the gene itself had yet to be cloned. The retroviral nature of the gene had therefore not been determined, and it seemed likely at this point that homologues of *Fv1* would exist in other species. The initial aim of this project was to isolate the human homologue of this gene and to see whether the retroviral restrictive nature of the gene observed in mice existed, or could be induced through mutation, in humans. To this end, large-insert human genomic libraries were screened for clones containing either or both of the human homologues of the murine genes that flanked *Fv1* in mice. From this clone(s) it was hoped the human *Fv1* homologue could be subsequently isolated.

Once the *Fv1* gene in mice had been cloned, its retroviral nature was demonstrated by sequence homology to the HERV-L family of retroviruses. Subsequently, *Fv1* was shown to be unique to members of the *Mus* genus only, and the presence of homologues in other genera therefore seemed highly unlikely. The aim of the project was then modified. By cloning and comparing *Fv1* sequences from members across this genus, the next aim was to attempt to determine how the gene had changed during the evolution of *Mus*, and to ascertain how these changes related to the development of *Fv1* activity in the mouse. To achieve this, the *Fv1* ORF was PCR-amplified and cloned from nearly 40 samples of mouse genomic DNA. The *Fv1* consensus sequence was determined for each sample, and these sequences were subjected to computer-aided analysis.

Complementary to the sequencing project, I aimed to clone a 'progenitor' viral sequence of *Fv1*; the sequence of the gene shares homology with both the HERV-L family of retroviruses from humans and the MERV-L family of retroviruses from mice, but appears distinct from both. It was hoped that, during such an infection that had given rise to the *Fv1* gene itself, another germline integration(s) had also occurred to fix a more complete example of this family of retrovirus in the mouse genome. In order to isolate such a putative sequence(s) a mouse genomic lambda library was screened by hybridisation with an *Fv1* probe at low stringency. Fragments that cross-hybridised with *Fv1* were subcloned from the clones obtained from this screen, sequenced using primers flanking the cloning site, and these sequences were subsequently used to screen nucleic acid sequence databases to identify them.

The final aim of the project was to determine how the key changes between the main alleles related to *Fv1* activity. This data also related to the study of the evolution of *Fv1,* as the sequencing project had shown how these changes had occurred during the speciation of *Mus* and the evolution of the *Fv1* phenotype. Constructs containing mutant ORFs with the 3 main differences between $Fv1^n$ and $Fv1^b$ in all possible combinations were transfected in to the *Fv1°* cell line *Mus dunni*. *Fv1* phenotype of clonal cell lines obtained from these transfections was determined using a *lacZ* assay using pseudotyped MLV. The $Fv1^d$ allele was also studied. Quantification of expression levels in tissue and sequence analysis of this allele was undertaken to determine what changes are likely to result in this modified phenotype.

Chapter 2. Materials and Methods

2.1 Solutions, enzymes and buffers

All enzymes were obtained from New England Biolabs and all chemicals were obtained from Sigma, unless otherwise stated.

Liquid cultures of *E. coli* were grown in LB (Luria-Bertani) broth as described by Sambrook *et al*[291]. LB broth is 1.0% bacto-tryptone (Oxoid), 0.5% yeast extract (Oxoid), 1.0% NaCl, pH 7.5.

Spread-plating was done onto L-agar plates containing the appropriate antibiotic. L-agar plates were made by adding 15g bacto-agar (Oxoid) to 1 litre LB broth before autoclaving. Antibiotic selection was then achieved by supplementing with 50μgml^{-1} nafcillin/ampicillin, streptomycin or kanamycin (Boehringer Mannheim GmbH) as required (media were allowed to cool to <55°C before the addition of antibiotic).

Mammalian cells were maintained in DMEM Complete media, which consists of DMEM (Gibco BRL) supplemented with 10% heat-inactivated (30 minutes at 56°C) FCS (PAA Laboratories GmbH) and 50μgml^{-1} penicillin and 80μgml^{-1} streptomycin (Life Technologies). Antibiotic selection was achieved by supplementing the media with 800μgml^{-1} G418. Stocks of cells were prepared by resuspending trypsinised cells in tissue culture freeze mix (50% FCS, 40% DMEM without supplements, 10% DMSO (Sigma)).

NZY media:

5g NaCl, 2g $MgSO_4.H_2O$, 5g yeast extract, 10g NZ amine (casein hydrolysate; (Oxoid))

-made up to 1 litre with dH_2O, pH adjusted to 7.5 with NaOH; autoclave

For NZY agar, NZY media was supplemented with 15g agar prior to autoclaving; for NZY top agar, 0.7% agar was added; for NZY top agarose, 0.7% agarose was added

RT assay mix:
24µl 1M Tris-HCl pH 7.5, 36µl 1M KCl, 1.2µl 0.5M DTT, 6µl 150mM $MnCl_2$, 5.4µl 2mM TTP, 12µl 200µgml^{-1} oligo dT-poly rA, 6µl ^{32}P-TTP (ICN; 167 Tbqmmol^{-1}), 6µl 10% Nonidet-P40, 403µl dH_2O

SES buffer
0.5% SDS
250mM EDTA, pH 8.0
150mM NaCl

SOC medium:
20g bacto-tryptone, 5g bacto-yeast extract, 0.5g NaCl, 20mM glucose
-made up to 1 litre with dH_2O, pH adjusted to 7.0 with NaOH; autoclave

SM buffer:
5.8g NaCl, 2.0g $MgSO_4.H_2O$, 50ml 1M Tris-HCl, pH 7.5, 5ml 2% (w/v) gelatin (Sigma)
-made up to 1 litre with dH_2O; autoclave

20x SSC buffer:
175.3g NaCl, 88.2g sodium citrate
-made up to 1 litre, pH adjusted to 7.0 with NaOH; autoclave

1x TE buffer:
10mM Tris-HCl, 1mM EDTA
-adjusted to pH 8.0 with NaOH; autoclave

10x TBE (electrophoresis buffer):
108g Tris-base, 55g boric acid, 9.7g EDTA
-made up to 1 litre, adjusted to pH 8.4 with NaOH

Gel loading buffer (agarose gels):
0.25% bromophenol blue, 0.25% xylene cyanol FF, 25% Ficoll

Gel loading buffer (RPA gels):
0.025% xylene cyanol, 0.025% bromophenol blue, 0.5mM EDTA

Gel loading buffer (automated sequencing gels):
50 µl 0.5M EDTA, pH 8.0, 50mg Blue dextran
-made up to 1ml with deionised formamide

DNA standards (see Appendix I):
1kb ladder from MBI Fermentas, λ*HindIII* digest from Life Technologies (for 23kb-500bp size range), ϕX174 *HaeIII* digest from Life Technologies (for 1300bp-100bp size range)

Solution I:
50mM glucose, 25mM Tris-HCl pH 8.0, 10mM EDTA

Solution II/P2:
0.2N sodium hydroxide, 1.0% SDS

Solution III:
60ml 5M potassium acetate, 11.5ml glacial acetic acid, 28.5ml dH_20
-the resulting solution is 3M with respect to potassium and 5M with respect to acetate, and is pH 5.3

Solution D:
4M guanidinium thiocyanate (Life Technologies), 17.6ml 0.75M sodium citrate, pH 7.0, 0.5% sarcosyl, 0.1M β-2-mercaptoethanol
-made with DEPC-treated dH_2O

P1:
15mM Tris-HCl pH8.0, 10mM EDTA, 100µgml^{-1} RNase A

P3: 3M potassium acetate pH 5.5

Chloroform-isoamyl alcohol:
chloroform and isoamyl alcohol added in a 24:1 ratio and mixed thoroughly

Phenol-chloroform:
phenol (saturated, pH 6.6; BDH), chloroform and isoamyl alcohol in a 25:24:1 ratio and mixed thoroughly

Denaturing (southern blotting/ colony lift/ plaque lift) solution:
1.5M NaCl, 0.5M NaOH

Neutralisation (colony lift) solution:
1.5M NaCl, 0.5M Tris-HCl pH 7.4

Neutralisation (southern blotting) solution:
1.5M NaCl, 1.0M Tris-HCl pH 7.4

Church hybridisation buffer:
70g SDS, 2ml 0.5M EDTA, 500ml 1.0M Na_2HPO_4, pH adjusted to 7.2 with orthophosphoric acid
-made up to 1 litre with dH_2O

1000x IPTG:
250mM in dH_2O

1000x X-gal:
50mgml^{-1} in dimethyl formamide

Fixative buffer (X-gal staining):
57.5ml 0.2M sodium dihydrogen phosphate, 192.5ml 0.2M disodium hydrogen phosphate, 10ml 0.25M EDTA, 500µl 2M $MgCl_2$
-made up to 500ml with dH_2O

Wash buffer (X-gal staining):
57.5ml 0.2M sodium dihydrogen phosphate, 192.5ml 0.2M disodium hydrogen phosphate, 5ml 10% sodium deoxycholate, 2.5ml BSA, 1ml 10% Nonidet-P40, 500µl 2M $MgCl_2$
-made up to 500ml with dH_2O

Stain buffer (X-gal staining):
25ml wash buffer, 0.5ml 50mgml^{-1} X-gal, 52.5mg $K_4Fe(CN)_6$, 41mg $K_3Fe(CN)_6$, 6mg spermidine, 0.4ml 2.5M NaCl
-filter-sterilised and store at 4°C for up to 2 weeks

PCR buffer:
100ml 1.0M Tris-HCl pH 8.3, 250ml 2.0M KCl, 650ml dH_2O; autoclave

T10:
10mM Tris-HCl pH 7.5; autoclave

RNase A:
1mgml^{-1} RNase A in dH_2O
-heated to 100°C for 15 minutes

Proteinase K:
10mgml^{-1} Proteinase K in dH_2O
-filter-sterilised (no pretreatment)

2.2 List of Bacterial genotypes

TG1	*supE, thi-1,* Δ(*lac-proAB*), Δ(*mcrCB-hsdSM)5,* (r^-_k, m^-_k), [F', *traD36, proAB, lacI*^{q}*ZΔM15*]
HB101	F$^-$, *hsdS20* (r^-_B, m^-_B), *recA13, ara-14, proA2, lacY1, galK2, rpsL20* (Sm^r), *xyl-5, mtl-1, supE44,* λ^-
MH1638	F$^+$, *recA1, endA1, gyrA96, thi-1, hsdR17* (r^-_k, m^-_k), *supE44, recA1*?, *R388*::TnXR, λ^-,
TOP10F' One Shot cells	F'{*lacI*^{q} Tn*10* (*Tet*^{R})}, *mcrA* Δ(*mrr-hsdRMS-mcrBC*), Φ80*lacZ*ΔM15, Δ*lacX74, deoR, recA1, araD139,* Δ(*ara-leu*)*7697, galU, galK, rpsL, endA1, nupG*
XL1 Blue MRA strain	Δ(*mcrA*)183, Δ(*mcrCB-hsdSMR-mrr*)173, *endA1, supE44 , thi-1, gyrA96, relA1, lac*
XL1 Blue MRA P2 strain	XL1 Blue MRA (P2 lysogen)

2.3 Mouse strains

Mouse strains DBA/2, CBA/J and C57BL/6 were used to obtain RNA from a variety of tissues to determine levels of *Fv1* expression. Mouse strains 129/SvEv, BALB/cJ and C57BL/6 were used to obtain DNA for southern blots. All mice were obtained in-house.

2.4 Genomic DNA samples

2.4.1 Mouse genomic DNA

Genomic DNA from a variety of mice was obtained for the evolution studies. The majority were purchased from The Jackson Laboratory (Bar Harbor, Maine) and some samples were a gift from Dr. F. Bonhomme (Laboratoire Genome et Populations, Universite de Montpellier II, CNRS URA 1493). The *Mus minutoides* sample, which is a member of the *Nannomys* group (see Figure 1.8), was a gift from Dr. B. Mock (National Cancer Institute, NIH). Details of these mice are listed in Table 2.1.

The 'CZECKI' sample is from *Mus musculus* CZECH I, the 'M.MUS' sample is from *Mus musculus* CZECH II, and the 'P.ATTECK' sample is from Peru Atteck/Ei. I have used the original strain names for all the rest.

Table 2.1 Mice DNA used in the sequencing project

Species	Name	Geographical origin	Source
Coelomys famulus	FAM	India	LGP[1]
Pyromys platythrix	PTX	India	LGP[1]
Mus minutoides	MIN	Africa	NCI[3]
Mus dunni	DUN	India	cell line[2]19
Mus cookii	COK	Thailand	LGP[1]
Mus caroli	Mus caroli	~	LGP[1]
Mus caroli	KAR	Thailand	LGP[1]
Mus spicilegus	ZYP	Yugoslavia	LGP[1]
Mus macedonicus	XBS	Bulgaria	LGP[1]
Mus spretus	SFM	France	LGP[1]
Mus spretus	M. Spretus	Spain	JL[2]
Mus musculus molissinus	MOL	Japan	LGP[1]
Mus musculus molissinus	MOLD/Rk	Japan	JL[2]
Mus musculus castaneous	CIM	India	LGP[1]
Mus musculus bactrianus	BIR	Iran	LGP[1]
Mus musculus domesticus	WMP/Pas	Tunisia	LGP[1]/JL[2]
lab strain	SEC/ReJ	~	JL[2]
lab strain	RIII/DmMob	~	JL[2]
lab strain	I/LnJ	~	JL[2]
lab strain	PRO/1ReJ	~	JL[2]
lab strain	FVB/NJ	~	JL[2]
lab strain	C57BL	~	inhouse
lab strain	SWR/J	Switzerland	JL[2]
lab strain	ST/bJ	Denmark	JL[2]
lab strain	BDP/J	~	JL[2]
lab strain	NZB/BINJ	~	JL[2]
lab strain	NZW/LacJ	~	JL[2]
lab strain	RF/J	~	JL[2]
lab strain	LG/J	~	JL[2]
Mus musculus musculus	M.MUS	Czechoslovakia	JL[2]
Mus musculus musculus	CZECKI	Czechoslovakia	JL[2]
Mus musculus musculus	MBT	Bulgaria	LGP[1]
Mus musculus domesticus	SKIVE/Ei	Denmark	JL[2]
Mus musculus domesticus	RBA/Dn	S.E. Switzerland	JL[2]
Mus musculus domesticus	SK/CamEi	Pembrookshire	JL[2]
Mus musculus domesticus	SF/CamEi	California	JL[2]
Mus musculus domesticus	P.ATTECK	Peru	JL[2]
Mus musculus domesticus	BZO	Algeria	LGP[1]
Mus musculus domesticus	WSB/Ei	Maryland, USA	JL[2]

[1]- Laboratoire Genome and Populations, Universite de Montpellier II
[2]- The Jackson Laboratory, Bar Harbor, Maine
[3]- NCI, NIH

2.4.2 Human genomic DNA

Human genomic DNA was used as a control in the screening of human genomic YAC and PAC libraries. The DNA was a gift from Dr. J.P. Stoye.

2.5 Genomic DNA libraries

All human genomic libraries were obtained from the UK HGMP Resource Centre at Hinxton, Cambridge.

2.5.1 The ICI human genomic YAC Library[292]

The library consisted of 34,500 clones with an average insert size of 350kb and a >3.5x coverage of the human genome, and was made from a human lymphoblastoid 48XXXX cell line. The primary screen consisted of 40 primary pools of DNA in LMP agarose.

2.5.2 The ICRF human genomic YAC Library[293]

The library consisted of 20,500 clones and combined 3 separate libraries:
164 plates labelled 4X1-4X164 made from a human female lymphoblastoid 48XXXX cell line with an average insert size of 600kb and a 3x coverage of autosomes and 6x coverage of the X chromosome.
24 plates labelled 4Y1-4Y24 made from a human male lymphoblastoid cell line 49XYYYY with an average insert size of 400-500kb.
26 plates labelled HD1-HD26 made from a human female lymphoblastoid, Huntingdon's disease 46XX with an average insert size of 600kb.
The primary screen consisted of 47 primary pools of DNA in LMP agarose.

2.5.3 The HGMP human genomic PAC Library (P. de Jong *et al*, Roswell Park Cancer Institute, Buffalo)

The library consisted of approximately 120,000 clones in 315 384-well microtitre plates cloned into the vector pCYCPAC2N (see Appendix 2). The source for library construction was a normal male blood donor and the average insert size was 120kb. The primary pools were coded A-U and consisted of 15 plates each.

2.5.4 The mouse 129 SvJ lambda genomic library

The library was purchased from Stratagene and was made from the spleen of a 8-week-old male mouse. DNA inserts were between 9-23kb and were cloned

into the Lambda FIX II vector. The library was made by cloning the genomic DNA partially digested with *Sau3AI* into the *XhoI* site of the Lambda FIX II vector. The library plated from this contained ~10^6 clones with an estimated 5x coverage of the mouse genome.

2.6 Preparation of DNA

2.6.1 Preparation of plasmid DNA

Plasmid 'minipreps' and 'midipreps' were carried out using a modified protocol based on the procedure described by Birnboim and Doly [294].

For 'miniprepped' DNA, 3ml of LB supplemented with the appropriate antibiotic to permit selection for the plasmid was inoculated from a single colony and incubated overnight at 37°C in a shaking incubator. 1.5ml of this overnight culture was spun for 30 seconds at 550g in an Eppendorf microfuge and resuspended in 100µl Solution I. After 5 minutes at room temperature 200µl of Solution II was added and the sample mixed by inversion. After 5 minutes on ice, 150µl of Solution III was added to the sample, which was then mixed by inversion and incubated for 5 minutes on ice. Following centrifugation at 10000g for 10 minutes the supernatant was transferred to a new tube and extracted with 500µl phenol:chloroform. After a further centrifugation at 10000g for 5 minutes the aqueous layer was carefully removed and the DNA precipitated by adding 1ml 96% ethanol followed by incubation at -20°C for 1 hour. The DNA was then pelleted by centrifugation, washed with 70% ethanol and dried. The pelleted DNA was finally resuspended in 40µl TE. The yield was usually quantified by running a digested portion of the DNA preparation on a 1.0% agarose gel with a known amount of standard DNA.

'Midipreps' were prepared in a similar way but using 20ml of a 50ml overnight culture, with the amounts of each reagent scaled up by ten times.

2.6.2 Preparation of PAC DNA

PAC DNA preparation was carried out using a modified version of the previous procedure supplied with the UK HGMP PAC human genomic library.

A single bacterial colony was inoculated into 2ml LB supplemented with kanamycin and incubated overnight. After centrifugation at 550g for 10 minutes the pellet was resuspended in 300µl P1. Following the addition of 300µl P2, the

sample was mixed by inversion and incubated at room temperature for 5 minutes. 300µl of P3 was then slowly added and the sample shaken gently. After a 5 minute incubation on ice, the sample was spun at 6000g for 10 minute at 4°C. The supernatant was carefully removed and the DNA precipitated by the addition of 0.8ml ice-cold IPA and incubation on ice for 5 minutes. The DNA was then pelleted by centrifugation at 10000g for 15 minutes at 4°C, and the resultant pellet washed with 70% ethanol. The pellet was allowed to air-dry and resuspended without pipetting in 40µl TE buffer. To check the DNA, 7µl of this was digested with *Not1* (the sites of which flank the SP6 and T7 promoter regions of the pCYPAC2N vector used in the construction of the library) and *BglII*, and the products of this digestion separated on a 0.7% agarose gel.

2.6.3 Preparation of single-stranded M13 DNA

Single-stranded DNA preparation was carried out by modification of the standard procedure described by Sambrook *et al* [291].

An overnight TG1 culture was diluted 1:50 with LB supplemented with 5mM $MgCl_2$[295] and incubated shaking at 37°C for 60 minutes. 3ml of this was then infected with 50µl of phage solution (produced by soaking a ~1mm diameter plaque in 1ml T10 overnight) and incubated for a further 3-4 hours at 37°C. 1.5ml was removed and microfuged at 10000g for 4 minutes in an Eppendorf benchtop microfuge to remove the cellular component and the supernatant transferred to a new tube. After another centrifugation at 10000g for 4 minutes, 800µl was transferred to a tube containing 200µl 20% PEG/2.5M NaCl. This was then vortexed thoroughly and allowed to stand at room temperature for 15 minutes. Following a 4 minute spin, the supernatant was discarded, the tube respun and all of the remaining supernatant removed. The resultant pellet was dissolved in 100µl TE, vortexed with 100µl phenol, and incubated at room temperature for 15 minutes. The sample was then extracted with 100µl phenol-chloroform and 100µl chloroform, and the ssDNA precipitated by the addition of 10µl 3M NaOAc and 250µl ethanol, preferably incubating at -20°C overnight. The ssDNA was then pelleted by centrifugation, washed in 70% ethanol, dried and finally resuspended in 10µl dH_2O. For automated sequencing, 1.5µl was generally used.

2.6.4 Preparation of lambda phage DNA

An overnight culture of XL1-Blue MRA strain of *E. coli* was grown in 100ml LB supplemented with 0.2% (w/v) maltose and 10mM $MgSO_4$. The culture was grown at 30°C to prevent the cells from overgrowing. The cells were then spun down at 800g in a MSE Mistral bench top centrifuge for 10 minutes and the cell pellet gently resuspended in 20 ml cold 10mM $MgSO_4$. The cells were further diluted with more 10mM $MgSO_4$ until the $O.D._{600nm}$= 0.5. Phage solution (corresponding to ~10^5 pfu) was then incubated with 600µl of these cells for 15 minutes at 37°C before adding 6.5ml NZY top agarose and plating onto a pre-warmed 15cm plate. Approximately 100µl of phage solution, prepared by incubating a single plaque plug in 1ml SM, was normally sufficient to produce confluent lysis after incubating overnight at 37°C. As a control, 600µl of cells was incubated in the absence of phage and plated. Once confluent lysis occurred the soft agarose layer was collected in a Falcon tube, 5ml 10mM Tris-HCl pH7.5/10mM $MgCl_2$ added and the contents shaken gently to dissipate the agarose pieces. After adding 200µl chloroform, the solution was mixed gently for 15 minutes on an orbital shaker. The sample was then centrifuged at 2000g for 15 minutes at 4°C. After removal of the supernatant the pellet was re-extracted with 4ml fresh buffer. EDTA and SDS were added to 10mM and 0.2% (w/v) respectively, followed by the addition of 1mg proteinase K. This phage solution was incubated at 37°C for 1 hour and then extracted with an equal volume of phenol-chloroform by gentle rocking on a orbital shaker for 10 minutes at room temperature. This was followed by centrifugation at 1000g for 5 minutes. The upper aqueous phase was then extracted with an equal volume of chloroform-isoamyl alcohol. After a further centrifugation at 1000g for 5 minutes the upper aqueous layer was transferred to a 30ml Corex tube containing 0.1 volume 3M sodium acetate (pH 5.5), the contents mixed and then placed on ice for 5 minutes. After adding 2.5 volumes of -20°C ethanol, the tubes were incubated at 4°C overnight. The DNA precipitated as a white mass which was removed using a sterile, hooked Pasteur pipette and dried in a freeze drier for 10 minutes. The DNA pellet was then resuspended in 500µl TE overnight at 4°C. If the DNA remained as a gelatinous lump, more TE was added. The quantity and quality of DNA was then determined (see section 2.6.6).

2.6.5 Preparation of genomic DNA

Tissue (typically ~1g) was chopped and homogenised in 10ml SES buffer. After adding 200µl 10mgml^{-1} proteinase K, the sample was mixed and incubated overnight at 55°C. To this , 1/10th volume of 1M Tris (pH 8.0) and 1 volume of phenol was added and the sample mixed by inversion. After centrifugation at 13000g for 10 minutes the supernatant was removed to a fresh 50ml tube and an equal volume of phenol:chloroform was added. After mixing by inversion, the sample was centrifuged for a further 10 minutes at 13000g and the supernatant was again removed to a fresh tube. After adding an equal volume of chloroform, the sample was mixed by inversion, centrifuged for a further 10 minutes at 13000g and the supernatant was removed to a final fresh tube. After the addition of 2.5 volumes of ethanol, the sample was inverted several times until the appearance of DNA was observed. This was then removed with a sterile, hooked glass Pasteur pipette and resuspended in 100µl TE. The yield and purity of the DNA was determined spectrophotometrically (see below).

2.6.6 Spectrophotometric determination of DNA

Between 2 and 10µl of the DNA sample was mixed with 400µl dH_2O and transferred to a 0.5ml quartz cuvette. The spectrophotometer was set to zero using 500µl dH_2O at $O.D._{260nm}$ and the readings were taken for the sample at both 260nm and 280nm. The yield of DNA or RNA was calculated using the guide that 1ml of a blanked solution with an A_{260nm} of 1.0 is equivalent to 50µg of double-stranded DNA or 40µg single-stranded DNA or RNA. The quality of DNA or RNA was measured by calculating the ratio of the absorbance at 260nm and 280nm: this ratio should be ~1.8 for DNA and ~2.0 for RNA. Absorbance ratios significantly lower than these values indicate there may be protein or phenol contamination in the sample which precludes accurate quantification of the nucleic acid.

2.7 Preparation of total RNA[296]

Small-scale RNA isolation from both tissues and cultured cells was carried out using a modification of the acid-guanidinium-phenol-chloroform (AGPC) method of Chomczynski and Sacchi[296].

For extraction from tissue, 1 ml of solution D was added to the freshly-dissected tissue and the sample homogenised by drawing through a wide-gauge needle

repeatedly. For extraction from cells, media was aspirated from the cells of a 10cm plate near confluency and the plate washed twice with PBS. After adding 1.8ml solution D, the plate was 'swirled' to ensure lysis of all the cells. In both cases the cell lysate was then transferred to a sterile 15ml Falcon tube. Sequentially, 0.1ml 2M sodium acetate, pH 4, 1ml phenol (water saturated), and 0.2ml chloroform-isoamyl alcohol were added with thorough mixing by inversion after the addition of each reagent. The final suspension was shaken vigorously for 10 seconds and cooled on ice for 15 minutes. After centrifugation at 10000g for 20 minutes at 4°C, RNA was present in the aqueous phase whereas the DNA and proteins were present in the interphase. The aqueous phase was transferred to a fresh tube, mixed with 1ml IPA and placed at -20°C for at least 1 hour to precipitate the RNA. After a further centrifugation at 10000g for 20 minutes the resulting pellet of RNA was dissolved in 0.3ml solution D, transferred to a 1.5ml sterile eppendorf tube and precipitated with 1 volume IPA at -20°C for 1 hour. Following centrifugation at 10000g in an Eppendorf microfuge for 10 minutes at 4°C, the RNA pellet was resuspended in 75% ethanol, sedimented, dried in a vacuum drier for 15 minutes and ultimately dissolved in 50 µl 0.5% SDS by incubation at 65°C for 10 minutes. RNA quantity and quality was assessed by measuring absorption at 260 and 280nm (see previous section).

2.8 Subcloning of DNA

2.8.1 Digestion of DNA

Digestion was carried out in a total volume of 20µl, or 50µl if digestion was overnight, using restriction endonucleases and buffers from New England Biolabs. All reactions contained 0.1 volume RNase A.

2.8.2 Purification and recovery of DNA

DNA fragments were gel-purified and recovered using the Prep-a-Gene DNA purification system (Biorad) according to a modification of the manufacturers protocol. In brief, DNA fragments were separated on agarose gels and excised. The gel slice was weighed and the amount of matrix required to bind the DNA was estimated. The slice was then dissolved in 3x(weight of slice in µg + amount of matrix required)µl binding buffer (sodium perchlorate) and the matrix added.

After 10 minutes at room temperature, the matrix was spun down by centrifugation at 10000g for 30 seconds in an Eppendorf microfuge and resuspended in 25x(amount of matrix). The matrix was pelleted as before and then resuspended in 25x(amount of matrix) wash buffer. This wash step was then repeated and the pellet dried for 2 minutes in a freeze-drier. The pellet was then resuspended in 1-2x(amount of matrix) dH_2O and incubated at 50°C for 5 minutes. The matrix was pelleted once more and the supernatant removed to a fresh tube. After a further centrifugation to remove any last traces of matrix, the DNA solution was ready to be used.

2.8.3 Dephosphorylation of cohesive termini

Dephosphorylation of cohesive termini to prevent self-ligation of vector was performed using shrimp alkaline phosphatase (Amersham Pharmacia, E70092) according to the manufacturer's protocol. 1µl of the enzyme was added directly to the restriction enzyme digest upon completion of digestion and heat inactivated at 65°C for 15 minutes after incubation at 37°C for 1 hour.

2.8.4 Ligation of DNA

DNA ligations were carried out in a total volume of 10µl using 1-2 units of bacteriophage T4 ligase (New England Biolabs) and 1x ligation buffer. Ligation reactions were incubated at room temperature for 2 hours.

2.9 Bacterial transformation with DNA

2.9.1 Heat-shock transformation of $CaCl_2$-competent cells

Competent *E. coli* cells were prepared and transformed according to the standard procedure described by Sambrook *et al* [291]. Typically 2.5µl of ligation mix was added to 100µl competent cells and the mixture incubated on ice for 30 minutes. After a 45 second heat-shock at 42°C, 1ml of SOC media was added and the cells were then incubated shaking at 37°C for 1 hour to allow expression of the selectable marker. The transformed cells were then spread onto agar plates containing the appropriate antibiotic (and 50µgml^{-1} X-gal/0.25mM IPTG if blue/white selection was required).

2.9.2 Electrotransformation of electrocompetent cells

Electrocompetent *E. coli* were prepared according to the standard procedure described by Sambrook *et al* [291]. A mixture of 5µl ligation mix and 100µl electrocompetent cells was placed in a pre-cooled electroporation cuvette and incubated on ice for 5 minutes. The cells were then electroporated (Biorad electroporation apparatus) at a voltage of 1.5kV, a resistance of 150 Ohms and a capacitance of 50 µFaradays.

Immediately after capacitor discharge, cells were transferred to 1ml pre-warmed SOC and allowed to recover for 30 minutes at 37°C prior to spread plating onto ampicillin-selective LB plates containing 50µgml^{-1} X-gal and 0.25mM IPTG.

2.10 DNA electrophoresis

For general purposes, agarose gel electrophoresis was carried out using standard procedures described by Sambrook *et al* [291], using horizontal slab gels. Gels contained 0.5µgml^{-1} ethidium bromide and were run in 1x TBE buffer at 80-120V, depending on the size of gel cast. Agarose (Life Technologies) concentration was between 0.6-2.0% depending on the size of DNA fragments being resolved. Either a λHindIII digest, a ϕX174 HaeIII digest or a 1kb marker ladder (Fermion) was run on gels for sizing purposes (~0.5µg/lane).

Polyacrylamide sequencing gels were run in 1x TBE buffer at a constant power of 60W for the desired time. Gels were cast at 6% using the Sequagel system (National Diagnostics).

2.11 Fixation of DNA to membrane

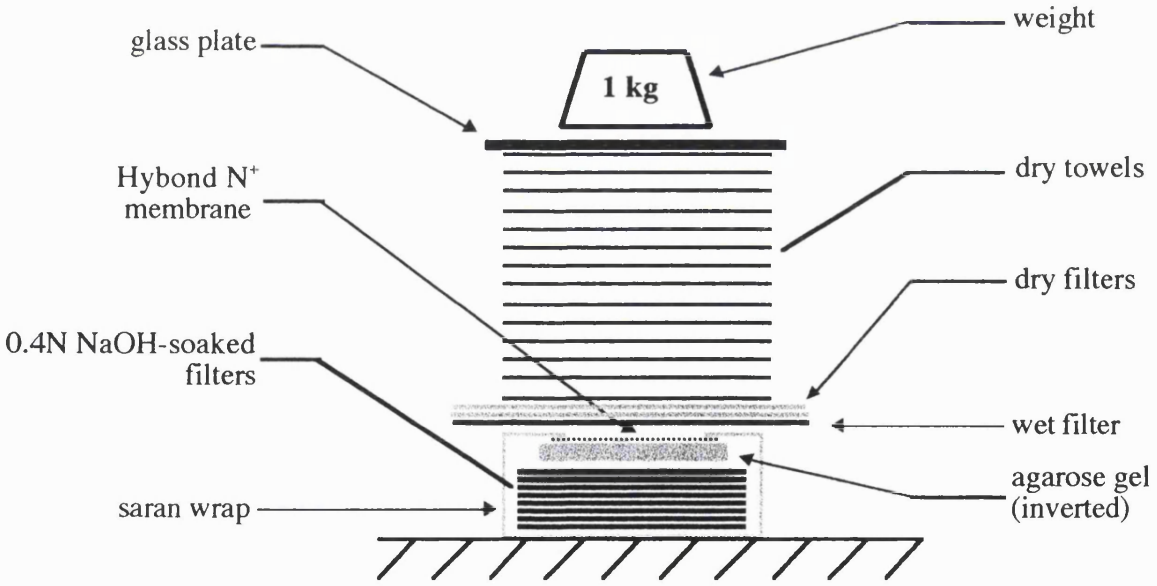
2.11.1 Southern Blots

Southern transfer of DNA separated by electrophoresis on horizontal agarose gels was by a modification of the procedure of Sambrook *et al*[291]. Dry blots were set up as shown in Figure 2.2 (next page).

Following electrophoresis the agarose gel was photographed, rinsed in dH_2O and placed inverted on 8 pieces of Whatman 3MM filter paper on a piece of saran wrap. The filter paper was pre-soaked in 0.4N NaOH as the transfer buffer and assembled carefully to avoid trapping air bubbles. Gels containing DNA fragments larger than 10 kb were depurinated first in 0.25M HCl until the bromophenol blue dye front had turned orange (~15 minutes). Hybond N$^+$

membrane (Amersham Phamacia) was then cut slightly smaller than the gel, wetted in 0.4N NaOH and positioned centrally on the gel taking care once again to avoid air bubbles. The saran wrap was wrapped around this so that each edge covered the edges of the membrane. In this way the transfer buffer could be drawn through the membrane only. A piece of damp filter paper was placed in direct contact with the membrane, and 2 more pieces of dry filter paper were placed on this. A pack of paper towels was placed on the top of this and the whole blot weighted with ~1 kg on a glass plate. Gels were generally blotted overnight, although some blots such as the PAC digests were left for up to 3 days. During dismantling the blot the position of the wells was marked on the membrane. After washing in 2x SSC to remove any agarose the blot was baked at 80°C for 2 hours. The membrane used for all Southern transfers was Hybond N^+, which under alkaline conditions requires no UV crosslinkage.

Figure 2.1 Assembling a dry Southern blot



2.11.2 Lambda phage plaque lifts

The mouse 129 genomic library was used to infect XL1 Blue MRA P2 *E. coli* cells and plated out according to the manufacturer's protocol to ensure each 15cm plate contained ~50000 plaques. These plates were then chilled at 4°C for 2 hours to prevent the top agarose from sticking to the nitrocellulose filters (PAL). The plaques were blotted onto a nitrocellulose filter for 2 minutes. For orientation

the filters and plates were marked in 3 places with a needle. The filters were then submerged in Denaturation solution for 2 minutes, followed by Neutralisation solution for 2 minutes, and finally rinsed in 0.2M Tris-HCl (pH 7.5)/0.2x SSC for 30 seconds. The filters were then baked for 2 hours at 80°C and stored at -20°C before use. The lambda library plates were stored at 4°C.

2.11.3 Bacterial colony lifts

Colony lifts, filter preparation and ultimately hybridisation was carried out according to the protocols described by Sambrook *et al*[291]. Briefly, a sterile, prenumbered nitrocellulose filter was placed on the surface of an LB agar plate, in contact with the bacterial colonies, until completely wet. The filter was marked by punching a sterile 18-guage needle through the filter into the agar at 3 asymmetric locations. With blunt-ended forceps the nitrocellulose filter was carefully peeled from the plate and laid onto the surface of a stack of 4 Whatman 3MM papers pre-soaked in 10% SDS for 3 minutes. This treatment limits the diffusion of plasmid DNA during denaturation and neutralisation, giving a sharper hybridisation signal. The nitrocellulose filter was then transferred to a second stack of filter paper soaked in denaturing solution and left for 5 minutes. At the end of this time the nitrocellulose filter was transferred to a third stack of filter paper pre-soaked in neutralising solution and left for a further 5 minutes. The filter was then laid colony side up on a dry sheet of 3MM paper and allowed to air-dry for at least 30 minutes. The filter was then sandwiched between 2 sheets of dry 3MM paper and the DNA was fixed to the nitrocellulose by baking for 1-2 hours at 80°C in an oven. The filter was then ready to hybridise immediately, or sealed into a hybridisation bag and stored at -20°C until needed.

2.12 Hybridisation

All hybridisation of radiolabelled probes was visualised by exposure to phosphorimager screens or, less commonly, by exposure to x-ray film. Phosphorimager screens offer many advantages over standard film, but the main advantages are much greater sensitivity, allowing filters to be exposed for much shorter times than film, and the ability to analyse phophorimages in the software package Imagequant. This software not only allows the image to be manipulated, for example allowing the threshold values to be optimised to reduce background or allow the detection of faint bands, but also allows the

quantification of band intensities which is essential for a quantitative protocol such as the ribonuclease protection assay.

2.12.1 Southern analysis

Filters were hybridised according to a modified procedure based on that described by Church and Gilbert[297].

Membranes were prehybridised in 10ml Church hybridisation buffer containing 50µl denatured salmon sperm DNA (10mgml^{-1}; Sigma) and incubated at 65°C in a rolling incubator (Hybaid). After 1 hour, denatured probe (heated at 95°C for 5 minutes with 50µl salmon sperm DNA and chilled on ice for 2 minutes) was added and left to hybridise overnight. The membranes were then washed initially with 100ml 2x SSC/0.1%SDS at 65°C for 20 minutes, then washed at progressively lower salt concentrations to increase the stringency. These subsequent washes were carried out at 65°C for 15 minute durations. The membranes were then sealed into hybridisation bags both to prevent drying out and for ease of handling, and typically exposed to a phosphorimager screen.

2.12.2 Lambda phage plaque lifts

The filters of the mouse genomic library were hybridised in 5ml Church solution inside hybridisation bags (up to 5 per bag) and incubated in a 65°C shaking waterbath. Template for the production of *Fv1*-specific probe used for the hybridisation of the lambda library was made by PCR-amplification using the primers GT16 and GT17 on pCI-BORF (the $Fv1^b$ allele cloned into the pCI-neo expression vector (Promega)) template. After overnight hybridisation the filters were then washed down to 0.5x SSC/0.1%SDS and exposed on a phosphorimager screen. The phosphorimage was printed at 100% magnification onto acetate sheets and, with the aid of a lightbox, these were used to identify positive plaques. Positive plaques were then picked using a Pasteur pipette, using the needle marks to orientate the plates to the filters, and stored in 1ml of SM buffer overnight at 4°C to elute the phage particles. All filters were probed in duplicate to confirm positive plaques.

2.12.3 Bacterial colony filter hybridisation

Nitrocellulose filters were used and all hybridisation was carried out in hybridisation bags (up to 5 filters/bag), using 5ml Church solution, in a 65°C

shaking waterbath. After hybridising overnight the filters were washed at 0.1x SSC/0.1% SDS and exposed on a phosphorimaging screen.

2.13 Production of radiolabelled probes

2.13.1 DNA Probes

DNA probes were made using the Redi-prime random primed labelling system (Amersham Pharmacia) following the manufacturer's protocol. Template DNA was gel purified prior to probe synthesis, and ~250ng DNA was either added directly to the reaction (if run in LMP agarose) or extracted from the gel by the Prep-a-gene method. The labelling reaction was typically incubated for 45-60 minutes. All probes were purified by passing the products of the labelling reaction down a S-300 desalting column (Promega) which removes unincorporated nucleotides. Incorporation of radiolabel was determined by measuring the ratio of incorporated label in the purified product and unincorporated label remaining on the column. Incorporation was typically 50-70%. As the amount of label added to the reaction was 1.85 Mbq, and the volume obtained after purifying through the column was 50µl, the typical specific activity of the probes was 20-25kBqµl^{-1}. 10µl of probe was added to each hybridisation.

2.13.2 Production of RNA probes (riboprobes)

Riboprobes for the ribonuclease protection assay were made using linearised plasmid containing the cloned probe sequence as template. *In vitro* transcription reactions were set up as follows: 5µl 5x buffer, 0.6µl RNasin (Life Technologies), 2.2µl 0.1M DTT, 4µl 2.5mM adenine/guanine/cytosine triphosphates, 4.5µl 0.37MBqµl^{-1} alpha-^{32}P UTP, 0.5-1.0µg linearised template DNA, 1.5µl (1 Uµl^{-1}) RNA polymerase made up to 25µl with dH_2O and incubated at 37°C. After 90 minutes, 2µl (2Uµl^{-1}) of DNase I (Boehringer Mannheim GmbH) was added to the reaction and incubated for a further 20 minutes. Following this incubation, an equal volume of Gel Loading Buffer was added and the DNA was denatured at 90°C for 5 minutes before loading onto a pre-poured 5% polyacrylamide gel (20cm x 16cm). The riboprobe was then separated by running at 25W for 100 minutes. After this time the gel was carefully dismantled and the band on the gel corresponding to the probe was visualised by exposing the gel to film for 1

minute. With the aid of carefully positioned orientation marks the autoradiograph of the gel was used to mark the position of the band on the gel: this was done by aligning the film against the gel and pushing a needle through this into the gel at each corner of the band. The band could then be excised by cutting the gel between the puncture marks made by the needle. Re-exposure of the gel to film confirmed the correct excision of the labelled RNA product. The riboprobe was eluted from the gel slice by incubation at 37°C in 250µl elution buffer (0.5M ammonium acetate/1mM EDTA/0.2%SDS) for 2 hours.

2.14 Ribonuclease protection assay

Ribonuclease protection assays were carried out using the RPA II kit (Ambion) for assays on tissue following the manufacturers protocols. Briefly, 30µg of sample RNA was mixed with 10^6 cpm high specific activity riboprobe specific to *Fv1* and $5x10^4$ cpm lower specific activity riboprobe (made with older radiolabel which was typically ~8x lower specific activity) specific to β-actin as the internal control. The specific activity of this control was further reduced by the addition of 2µl 50µM non-radioactive UTP in the labelling reaction. After incubation overnight in a Hybaid hybridisation oven at 42°C, the unprotected single-stranded RNA was digested by the addition of 2µl Solution R in 200µl digestion buffer (this corresponds to ~1ng RNase A and 20 units cloned RNase T1) and incubation at 37°C for 30 minutes. The remaining RNA was then precipitated. Following precipitation the RNA pellets were dissolved in gel loading buffer, heat-denatured and loaded onto a 5% denaturing polyacrylamide gel (42cm x 33cm). Once the gel had run for ~2.5 hours at 1500V and 65W the gel plates were dismantled and the gel transferred to filter paper. After drying the gel was then either exposed to a phosphorimager screen or to film. For every assay, control reactions were prepared containing 30µg of Yeast RNA with the same amount of riboprobe as used in the other reactions; only one of these underwent RNase digestion. These controls enable the complete digestion of single-stranded probe to be confirmed, the quality of riboprobes to be assessed and could indicate whether protected bands are real or artefactual.

The *Fv1* riboprobe comprised of 50bp of upstream sequence extending 125bp into the ORF, and 24bp of vector sequence. The full-length probe therefore was 199bp long, of which 175bp is protected in normal $Fv1^+$ cells. In the *Mus dunni* ORF there is a single base pair difference from the ($Fv1^n$ and $Fv1^b$ alleles) 78bp

into the sequence. The predicted size of the *Mus dunni* protected fragment therefore is 128bp.

2.15 Library screening

2.15.1 YAC libraries[292, 293]

The YAC human genomic libraries were screened by PCR with the primers pairs SE3/4 and SE5/6 (see Figure 3.1 and section 2.16.3). Libraries were received as pooled YAC DNA embedded in LMP agarose. In accordance with the HGMP Resource centre protocol, the agarose plugs were washed in TE for 2 hours, with shaking and 2 changes of buffer, and finally left in TE at 4°C overnight. This treatment removed all inhibitory agents from the agarose. The agarose plugs were then melted at 65°C when needed, and 1-2µl used per PCR reaction. All PCR reaction products were separated on 1.0% agarose gels, visualised and photographed under UV, and blotted overnight. These blots were then hybridised either with SE3/4 or SE5/6 probe to ensure even the faintest bands would be detected. These probes were made by random-prime labelling the products of the PCR amplification of human genomic DNA with the primers SE3/4 or SE5/6.

2.15.2 PAC libraries (P. de Jong *et al*, Roswell Park Cancer Institute, Buffalo)

The PAC human genomic library was similarly screened by PCR with the primers pairs SE3/4 and SE5/6. The library was received as primary pools of frozen live cells in media. These primary pools were defrosted, aliquoted and stored at -80°C. The working stock was then defrosted when needed and PCR was performed on 2µl of bacterial suspension. All PCR reaction products were separated on 1.0% agarose gels, visualised and photographed under UV light, and blotted overnight. These blots were then hybridised with SE3/4 or SE5/6 probes to ensure even the faintest bands would be detected.

2.15.3 Lambda phage library

The mouse 129/SvJ genomic lambda library was screened by hybridisation using the *Fv1* open reading frame as the probe. The filters were hybridised overnight and washed down to 0.5x SSC/0.1% SDS before exposing on a phosphorimaging screen. Phage plugs were taken from areas that corresponded to bright spots on

the phosphorimage. These plugs were incubated at 4°C in 1ml SM buffer. Dilutions of this phage solution was then used to infect XL1 MRA *E. coli*, and these were plated out for a secondary screen. Phage lifts were made on these plates and individual positive plaques were isolated. Plugs of these plaques were stored at 4°C in 1ml SM buffer, and used to prepare lambda DNA of positive clones.

2.16 Polymerase chain reaction

2.16.1 PCR conditions for standard reactions

PCR reactions of 25µl typically contained the following: 10mM Tris-Cl, pH8.3, 50mM KCl, 10mM dNTP, 10 pmols each primer, 1-2mM $MgCl_2$, 100-200ng template DNA and 1.25 units Taq DNA polymerase (Amplitaq; PE Applied Biosystems). When the PCR reaction products were to be subcloned the reaction mix was supplemented with 0.09 units Pfu DNA polymerase to reduce error rate. Reactions were set up without the $MgCl_2$, which was added only after the mix had reached 80°C ('hot-start').

Cycling conditions were typically a denaturation step at 96°C for 1 minute, followed by an annealing step at 58°C for 1 minute, followed by an extension for 1 minute per kb of template; this was repeated for 35 cycles and ended with a final 10 minute extension at 72°C. All reactions were carried out in an PTC-100 programmable thermocycler (MJ Research, Inc.).

2.16.2 PCR conditions for library screens

PCR reactions contained the following: 10mM Tris-Cl, pH8.3, 50mM KCl, 10mM dNTP, 10 pmols each primer, 1mM $MgCl_2$, 2µl DNA plug/cells and 1.25 units Taq DNA polymerase. Cycling conditions were as follows:

95°C for 5 minutes, 58°C for 1 minute, 72°C for 1.5 minutes	- initial step
94°C for 1 minute, 58°C for 1 minute, 72°C for 1.5 minutes	- 35 cycles
72°C for 10 minutes	-final step

2.16.3 List of PCR primers

Primers were purchased from Genosys. The following primers were used in this study (next page):

SE3	GAA GTT GAG GAC GTC TCT GG
SE4	CTC AAT CTG GAG CTC CTG AG
SE5	CAC TGT GTC AGC ACG TCT TG
SE6	GCT GTC ACC AAG GTG TTG AC
SE7	CAA CCA ACC AGT GGC AAT AG
SE8	CAA GTC TCT GTC TGA CCT TAC
SE9	CTT GTT TCC ACT CAG CTA TGG
SE10	CTT TTG ACT ACA AGC AGC ACC
SE11	CGA CAT CAG CTA ACC TCA CA
SE13	GGG TAA GCT GAA CCT TTT CC
PL26	GAA GAG CCA GCT CGA GAA AC
PL76	AAG GGC TTT GTA AGG AGA GC
PL80	GAA GCG GAA GAA GTC TCT TG
JS139	GTT TAC CAC TTA GAA CAC AG
JS140	TTT GCC GCA GAA GAT TCT GG
GT15	GGA TAT GTC GAC TCC TCC TCC TGA TTT TAA
GT16	CAA AAA GAT CTA GAT GAA TTT CCC ACG TGC G
GT17	TGG ATA GTC GAC ATC TAT ACT ATC TTG GTG AG
116	AAA TAC CAA GCA TGC C
117	GTC TCT CTG TAA ATG TGC
M13 (-40)	GTT TTC CCA GTC ACG AC
M13 (rev)	GGA AAC AGC TAT GAC CAT G
δ	AGG GGA ACT GAG AGC TCT A
γ	CCT GAA AAG GGA CCT TTG TAT ACT G

2.16.4 Direct cloning of PCR-amplified products

PCR-amplified products were cloned directly from PCR reactions using the Topo TA cloning kit (Invitrogen) according to the protocol supplied with the kit. Cloned products were used to transform TOP10F′ ultracompetent *E. coli* cells.

2.17 Automated DNA sequencing

DNA sequencing was performed using the Dye terminator sequencing kits (Perkin Elmer) according to the procedure recommended by the manufacturer. Single-stranded M13 DNA was consistently found to give better quality sequence than double-stranded template and for this reason was used as template in the majority of the automated sequencing reactions, although double-stranded column-purified (S-300 desalting columns; Promega) DNA was sequenced on occasion. In brief, 500ng-1μg of plasmid DNA was sequenced in a total volume

of 20μl using 8μl of dye-terminator reaction ready mix and 3.2 pmols primer. PCR reactions were carried out in a PTC-100 programmable thermal cycler (MJ Research, Inc.) using the following cycle conditions:
96°C for 30 seconds, 50°C for 15 seconds, 60°C for 4 minutes- 25 cycles
DNA was precipitated at -20°C for 10 minutes following the addition of 50μl ethanol and 2μl 5M sodium acetate. The pellet was then resuspended in 4μl gel loading dye and 2μl of this was loaded onto a 6% polyacrylamide gel. Samples were run on an 377 Automated sequencer (Applied Biosystems).

2.18 Transposon-based strategy to rapidly sequence large DNA fragments[298]

Day1: Competent cells of the donor *E. coli* strain MH1638 were prepared and transformed with the target plasmid according to the standard procedure described by Sambrook *et al* [291]. Transformants were selected for by plating onto LB agar plates supplemented with 50μgml^{-1} ampicillin and 40μgml^{-1} kanamycin.
Day2: A single colony of donor *E. coli* transformed with the target plasmid was streaked out onto a fresh LB agar plate supplemented with ampicillin and kanamycin. A single colony of recipient *E. coli* was streaked onto a LB agar plate. These were used as single colony stock plates.
Day3: A single colony from the transformed donor plate was used to inoculate 10ml LB broth supplemented with ampicillin and methicillin. A single colony from the recipient plate was used to inoculate 10ml LB broth with no antibiotics. Both cultures were grown overnight with no antibiotics.
Day4: 0.1ml of each of the overnight cultures was used to inoculate 10ml LB, supplemented with ampicillin and kanamycin for the donor culture and without antibiotics for the recipient. Both cultures were incubated shaking at 37°C until the $O.D._{600nm}$ reached 0.5 (typically 3-4 hours). Between 3-5ml donor culture was then pelleted by centrifugation at 1500g for 5 minutes, washed in 5ml fresh LB with no antibiotics and then pelleted again. This pellet was resuspended in 1ml of the recipient culture. This mating mixture was then plated onto a pre-dried LB agar plate and incubated for 1 hour. The cells were rinsed from the plate with 10ml LB broth and pelleted. This was then resuspended in 1ml LB and 10-100μl spread-plated onto LB agar plates containing 50μgml^{-1} methicillin, ampicillin and streptomycin. The remaining mixture was stored at 4°C for possible future use. The plates were incubated overnight at 37°C.

Day5: Individual colonies were picked into LB broth supplemented with ampicillin and kanamycin and grown at 37°C overnight.
After isolating the plasmid, the site of transposon integration could be revealed by restriction digestion; by picking and sequencing clones in which transposition had occurred in the insert fragments only, the insert could be rapidly sequenced. The sequencing primers used were to the δ and γ ends of Tn1000; the final 35bp at the δ and γ ends of Tn1000 are inverted repeats that give the following sequence with the δ and γ primers before entering the cloned DNA: AACGTA CGTTTT CGTTCC ATTGGC CCTCAA ACCCC.

2.19 Analysis of sequence data

2.19.1 Editing of raw sequence (Sequencing Analysis software)

This software package was purchased from Applied Biosystems. It was used to edit the sequence prior to contig assembly in Autoassembler.

2.19.2 Contig assembly and construction of consensus sequences (Factura and Autoassembler)

These software packages was purchased from Applied Biosystems. Factura was used to determine the valid range of accurate sequence before analysis in Autoassembler. Once imported into Autoassembler, this software was used to construct contiguous sequence by pairwise analysis of multiple sequences. Once contigs were free of ambiguities, this software was used to produce a consensus sequence.

2.19.3 Multiple sequence alignments (GDE and MAGI)

GDE (Genetic Data Environment) is the copyright of Steven Smith and the University of Illinois. GDE was used to construct multiple alignments of the mouse consensus sequences using the ClustalV method. MAGI was used to construct multiple alignments of the *Fv1*, MERV-L, HERV-L, HERV-K and HSV sequences using the ClustalW method. This sequence analysis interface (Multiple Alignment General Interface) is available at the UK HGMP Resource website, at the Sanger Centre in Cambridge (http://www.hgmp.mrc.ac.uk/Registered/Menu/).

2.19.4 Phylogeny analysis (Phylip)

This software is the copyright of Joseph Felsenstein and the University of Washington, and is available free of charge[299]. It was used as part of PIE (Phylogeny Interface Environment, also at the UK HGMP) to analyse the multiple alignments obtained from GDE to produce phylogenetic trees.

2.19.5 Calculating substitution rates (PAML)

This software is the copyright of Ziheng Yang and the University of California, and is available free of charge[300]. The number of both synonymous and non-synonymous substitutions per site (Ks and Ka) was calculated using the program codonml in the package PAML. This program uses the method of Nei and Gojobori (1986) to calculate Ks, Ka and Ks/Ka, and implements the model of Goldman and Yang (1994a).

2.19.6 Sequence similarity searches (BCM)

Sequence similarity searches were carried out using the BCM search launcher on the World Wide Web (http://kiwi.imgen.bcm.tmc.edu:8088/search-launcher/launcher.html). The search used was typically the BLASTN search which uses the nucleic acid sequence to search the Genbank, EMBL, DDBJ and PDB databases for all non-redundant sequences (Genome Research 6: 454-462).

2.20 Mouse cell culture

2.20.1 Maintenance of cell lines

Murine cell lines were maintained in DMEM Complete media and stable transfected cell lines were maintained in DMEM Complete supplemented with 800$\mu g ml^{-1}$ G418. All cell lines were split every 3-4 days to prevent the cells from becoming over-confluent. Cell lines were split by removing the media, washing the cells once with 1.5ml trypsin (Gibco BRL) and trypsinising at 37°C in a further 1ml trypsin for 3 minutes. Following trypsinisation the cells were homogenised in 3.5ml DMEM Complete media by pipetting up and down several times against the bottom of the plate. Between 4 and 10 drops of this cell suspension were then used to seed a fresh 10cm plate of media, depending on the growth rate of the particular cell line.

2.20.2 Transfection of *Mus dunni* cells and selection for clonal cell lines

Transfection protocol: Stable transfection of *Mus dunni* cells was carried out using Superfect (Qiagen) according to a modification of the manufacturer's standard protocol. Cells were grown in 6cm tissue culture plates to 20-40% confluency in 5ml DMEM Complete. The same miniprepped DNA that was used for sequencing was used to transfect mouse cells. Approximately 1μg of construct (of a minimum concentration of $0.1\mu g\mu l^{-1}$) was mixed with 150μl DMEM containing no serum or antibiotics. After mixing and spinning this down briefly, 7.5μl of Superfect was added to the DNA solution and mixed by pipetting up and down 5 times. This was then incubated at room temperature for 5-10 minutes to allow complex formation. While this was happening, the growth medium was removed and the cells washed once with PBS. After adding 1ml of DMEM Complete to the reaction tube containing the transfection complexes, this DNA complex solution was mixed by pipetting up and down twice and transferred to the cells. After incubation for 2-3 hours at 37°C and 5% CO_2, this medium was removed and the cells washed 4 times with PBS. Fresh DMEM Complete medium was then added and the cells incubated for 3 days to ensure sufficient expression of the marker gene.

Selection for stable cell lines: Following transfection and recovery, the medium was removed from the plates. The cells were then washed once with 1ml trypsin and incubated with 1ml trypsin at 37°C and 5% CO_2 for 3 minutes. After ensuring the cells had completely detached following trypsinisation, they were then resuspended in 3.5ml DMEM Complete supplemented with $40\mu gml^{-1}$ G418 and pipetted vigorously to reduce cell clumping. In order to isolate single transfected clones, the cell suspension was diluted and aliquoted: approximately 150μl (3 drops) and 750μl (15 drops) of this cell suspension was added to separate 150ml volumes of DMEM Complete supplemented with G418 and mixed thoroughly. Both of these dilutions were then transferred to 6 24-well plates, approximately 1ml/well. The media was replaced every 3 days to ensure maintenance of selection. After about 10 days, the wells were regularly examined for growth. Transfected clones from plates with less than 6 wells positive for growth in total were transferred to 6cm plates when confluent; plates with more than 6 positive wells were discarded. Transfected cell lines were maintained in G418-supplemented media.

2.20.3 Storage and reviving of stable cell lines

Cell lines were grown to near-confluency on 10cm tissue culture plates prior to freezing. The media was removed and the cells washed twice in 1.5ml trypsin before incubating at 37°C and 5% CO_2 for 3 minutes. After ensuring the cells had completely detached following trypsinisation, they were then resuspended in 5ml tissue culture freeze medium and pipetted vigorously to reduce cell clumping. This suspension was then aliquoted into 3 cryo tubes (Nunc) and transferred in an insulated box to a -80°C freezer overnight. The tubes were transferred to a liquid nitrogen store the following day. This 2-step freezing procedure reduces loss of cell viability caused by rapid freezing.

2.20.4 Virus infection of cell cultures

Supernatant from virus-producing cell lines was aliquoted into 1.5ml eppendorf tubes and stored at -80°C. These stocks were assayed by infecting the *Fv1*$^{\circ}$ *Mus dunni* cell line at various dilutions. In the case of the MLV stocks used in the RT assay, the lowest dilution that gave a productive infection was used. In the case of the pseudotyped virus used in the lacZ assay, the dilution that gave ~500 blue cells per plate was used.

Cell lines to be assayed were plated out in a 6-well tissue culture plate to a density of ~20-40% confluency. The media was removed from the cells and 1ml of DMEM Complete containing the virus and 15μg polybrene was then added. For the control well, 1ml DMEM Complete containing only polybrene was added. The cells were then incubated at 37°C and 5% CO_2 for 1 hour. This media was replaced with 3ml fresh DMEM Complete and the cells incubated for 3 days at 37°C and 5% CO_2 before being assayed.

2.21 Biological Assays for *Fv1* phenotype

2.21.1 Reverse Transcriptase assay

In order to assay cells for *Fv1* phenotype, cell lines were challenged separately with equal titres of N-tropic, B-tropic and NB-tropic MLV virus. Productive infection was then determined by measuring levels of reverse transcriptase activity in the supernatant. After infection and incubation for 3 days, 700μl of supernatant was removed and centrifuged for 30 seconds at 10000g in an Eppendorf microfuge to remove any cells. 500μl was then transferred to a

polycarbonate tube and centrifuged at 40000g for 20 minutes at 4°C in a Beckman TL-100 bench top centrifuge (Vti 100.1 rotor) to pellet viral particles. The supernatant was carefully removed and the inside of the tube dried with a tissue. 10µl of the RT assay mix was then added, mixed briefly in the bottom of the tube and incubated at 37°C for 2 hours. 5µl of this reaction mix was spotted onto DE-81 paper and allowed to air-dry. After 1 hour the filter paper was washed 3 times in 200ml 2x SSC for 10 minutes and finally in 200ml ethanol for 10 minutes. After air-drying for 30 minutes it was then exposed to film, typically for 1 hour.

2.21.2 LacZ assay using Pseudotyped MLV

Cell lines were grown in 6-well plates and challenged with N-tropic, B-tropic and NB-tropic pseudotyped MLV. After 2 days the media was removed and the cells fixed in 5ml Fixative buffer at 4°C for 10 minutes. The cells were then washed twice in 5ml Wash buffer at room temperature for 10 minutes, and finally stained in 5ml Stain buffer at 37°C for 2-16hrs. The number of blue cells was then counted.

2.21.3 Virus stocks

All virus stocks were prepared by Tony Stevens. In brief, N- and B-tropic viruses were recovered from infectious molecular clones[274] by transfection. NB-tropic virus was made by site-directed mutagenesis of B-tropic MLV (T. Stevens, unpublished). Pseudotyped virus were recovered following transfection of TEL/AF cells, which express MLV amphotropic Env and a retroviral vector expressing *lacz*[275], with N-, B- and NB-tropic Gag-Pol expression plasmids (T. Stevens, unpublished).

Chapter 3 The Cloning of the Human *FV1* Region

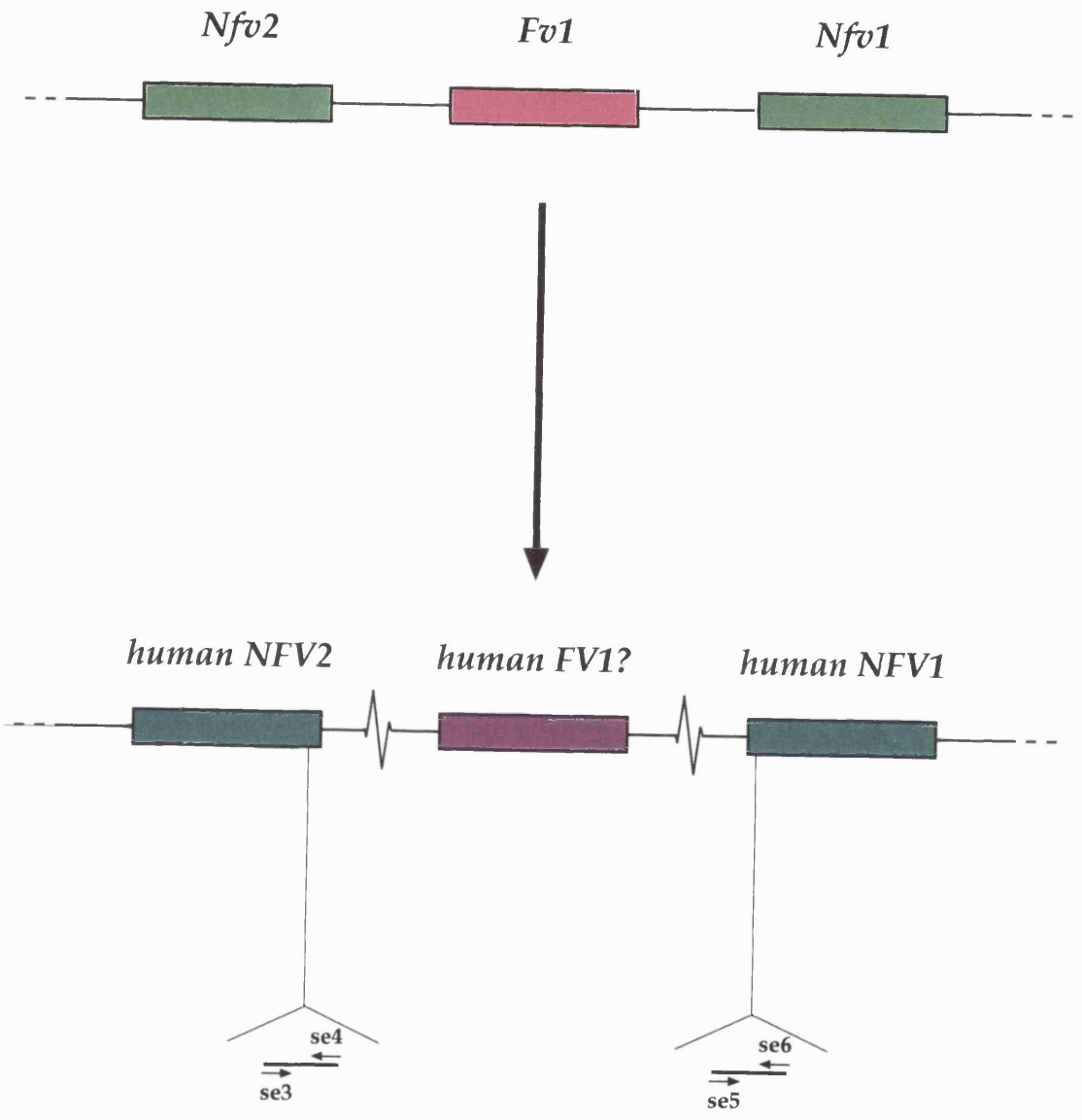
The *Fv1* gene in mice has been the subject of extensive study for over 30 years and, with its cloning in 1996, work to understand the precise mechanism of action of this gene continues. At the start of this project the gene itself had yet to be cloned but the nature of *Fv1* retroviral restriction in mice made this gene potentially an extremely interesting model to study in other organisms, most notably humans. Up until this point it was assumed to be part of the natural murine genetic complement, and as such it seemed logical to expect homologues of *Fv1* in other species. To this end, the aim of the cloning project was to isolate a fragment of human DNA containing the human homologue of *Fv1*, and to subsequently characterise this gene in humans and other species.

Initially the most important step was to select a strategy to screen human genomic libraries for the human homologue of *Fv1* that would maximise the chance of obtaining a positive clone(s). At this point *Fv1* activity had been mapped to a 6.5kb *EcoRI-SpeI* fragment only. A direct screen, using this mouse sequence to screen human clones by hybridisation, was rejected, as the *Fv1* ORF itself had not been identified. Thus, it was not known how well the gene was conserved between humans and mice, if at all: if the gene was not functional in humans, sequence divergence might have precluded identification by direct hybridisation. A PCR-based screen was determined to be the best strategy to isolate the human homologue. Sequence of the *Fv1* region revealed that the gene is flanked by 2 genes, *Nfv1* and *Nfv2*. These genes were first identified by exon trapping and RACE (*r*apid *a*mplification of *c*DNA *e*nds). Although their function remains to be defined, analysis of the EST databases indicated that there there is extensive homology between mouse and human sequences. The objective of the screening strategy was to isolate a fragment of human DNA that contained the human homologues of these flanking genes, and which would therefore contain the homologue of *Fv1*, located at some point between them.

3.1 Overview of the cloning strategy

An overview of the strategy used to isolate human clones containing the human *FV1* region is shown in Figure 3.1. The region harbouring the mouse *Fv1* gene

Figure 3.1 Flanking marker strategy to screen human genomic libraries for the human homologue of *Fv1*



BLAST searches using the sequences of murine *Nfv1* and *Nfv2* identified related EST sequences in human assumed to be the human homologues of these genes. Primers were constructed to these sequences which allowed the identification of large-insert human genomic clones containing either or both of these genes. In this way it was hoped that the human homologue of *Fv1* would be isolated.

has been well characterised (P. Le Tissier, unpublished). We know the gene lies between genes currently named *Nfv1* and *Nfv2*. As the distal half of mouse chromosome 4 shows consistent homology to human chromosome 1 for over 60 genes[301] (updated at http://www.ncbi.nim.nih.gov/Homology/), one would expect conservation of synteny between mouse and human for 3 adjacent genes within this region. Therefore, the isolation of human clones that contain the human homologues of these flanking genes should also yield the human homologue of *Fv1* itself. To this end, primer pairs were constructed to human ESTs that were identified by their homology to the mouse flanking genes; these were assumed to represent exon sequence of the human homologues of these genes.

3.2 Construction of the human primers for screening

3.2.1 BLAST searches to identify human *NFV1* and *NFV2* ESTs

The exon sequences of the flanking genes *Nfv1* and *Nfv2* in mouse were used to perform a sequence similarity search across the dbest EST database. Matches of 60-80% identity to human coding sequences were obtained (data not shown). These sequences were used to construct primer pairs (SE3/4 and SE5/6) that should allow the amplification of 400bp and 600bp products from the human *NFV2* and *NFV1* homologues respectively. In this way the 2 primer pairs could be used to screen large-insert human genomic libraries by PCR for clones containing both flanking gene homologues.

3.2.2 Testing the primers at reduced template concentrations

Once the primers SE3-6 were obtained they were tested in order to ascertain whether they would be suitable to PCR-amplify products from the limited amounts of template DNA that are found in primary PAC and YAC pools. PCR reactions were set up with each primer pair using a titration of human genomic DNA (isolated from blood; see section 2.4.2), from 100-0.1ng template DNA. A single product of the appropriate size was clearly obtained at all template dilutions (data not shown). This result indicated that these primers were specific and that the PCR reaction should be sufficiently sensitive to allow the PCR-amplification of product from the libraries tested.

3.3 Screening large-insert human genomic libraries by PCR

3.3.1 YAC libraries

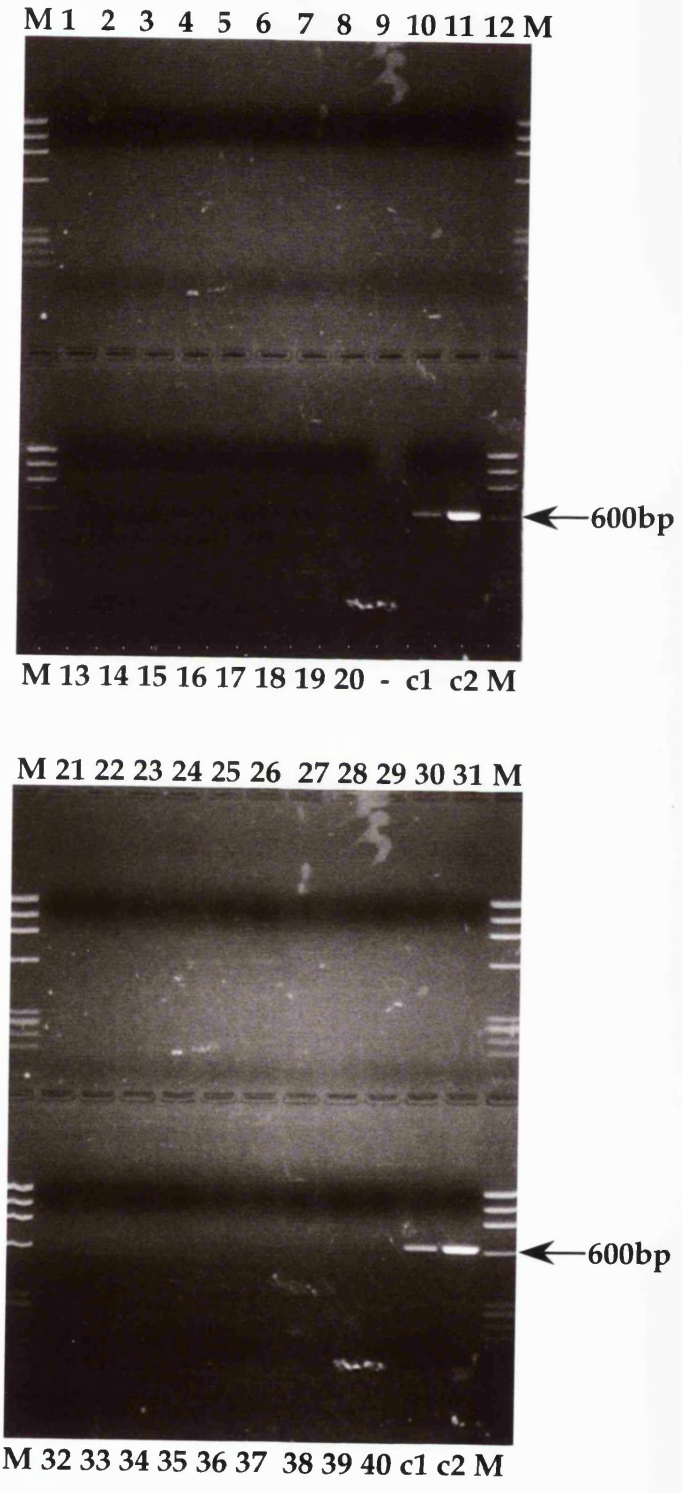
The products of the PCR screen of the ICI (Figures 3.2 and 3.3) and ICRF (data not shown) YAC libraries were run on a 1% agarose gel. No PCR-amplified product was visible in any lanes except the positive control. To ascertain whether there was any detectable product in any of the reactions, the gels for the screen of the ICI YAC library using the SE3/4 primer pair were blotted and hybridised with a probe specific for *NFV2* under low stringency conditions (Figure 3.3). This was made by random-prime labelling the SE3/4 PCR product itself. No detectable hybridisation signal was observed in any of the YAC wells. To ascertain whether the libraries themselves were satisfactory, a primer pair (116/117; a gift from Roger Buxton, NIMR) specific for the human DSC2 (desmocollin 2) gene, which is known to be in both libraries, was used to screen these libraries (data not shown). Intense bands were observed with all pools shown previously to be positive. Since both YAC libraries were shown to be negative for the *Fv1* flanking genes, a PAC library was selected for screening.

3.3.2 PAC libraries

The products of the primary PCR screen of the HGMP PAC library pools are shown in Figures 3.4 and 3.5. Bands in positive lanes were usually visible by ethidium bromide staining, but all gels run were blotted and hybridised with probes specific to both *NFV1* and *NFV2* to ensure even the faintest bands would be detected. These probes were made by random-prime labelling the SE5/6 and SE3/4 PCR products respectively. Successive screening of primary, secondary and tertiary PAC pools finally led to the isolation of 2 PAC clones that each contained both the human flanking gene homologues. These were PAC 107A22 and 277I9. Restriction digests of these 2 PACs revealed the presence of many bands unique and common to both clones (see Figure 3.6). This indicates that these were 2 overlapping but discrete clones.

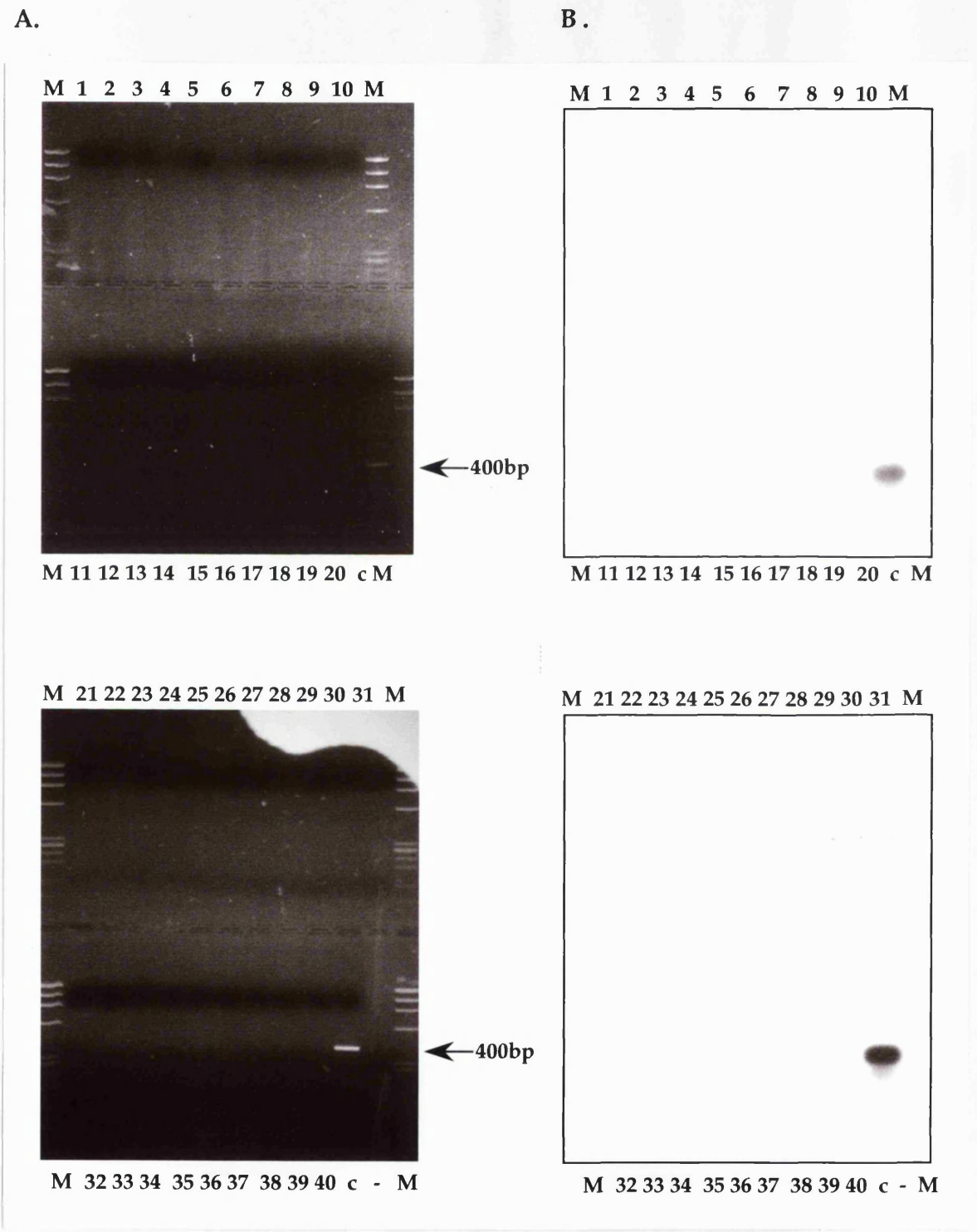
Also shown in Figure 3.6 are digests of 3 randomly-picked PAC clones (PACs 107E15, 277F8 and 277M18), and an empty PAC vector (PAC107N6). These were included as controls for a Southern blot to demonstrate hybridisation was specific for the *NFV1*$^+$ and *NFV2*$^+$ PACs only (and not due to non-specific cross-hybridisation with human insert or vector DNA; data not shown).

Figure 3.2 PCR Screen of the ICI Human Genomic YAC Library with primers SE5 and SE6



Lanes 1-40 contain the products of the PCR reaction on primary pools 1-40 respectively.
Lane c_1 is the control reaction using 1ng human genomic DNA as template.
Lane c_2 is the control reaction using 100ng human genomic DNA as template.
Lane m is the marker lane containing ϕX174 HaeIII digest.

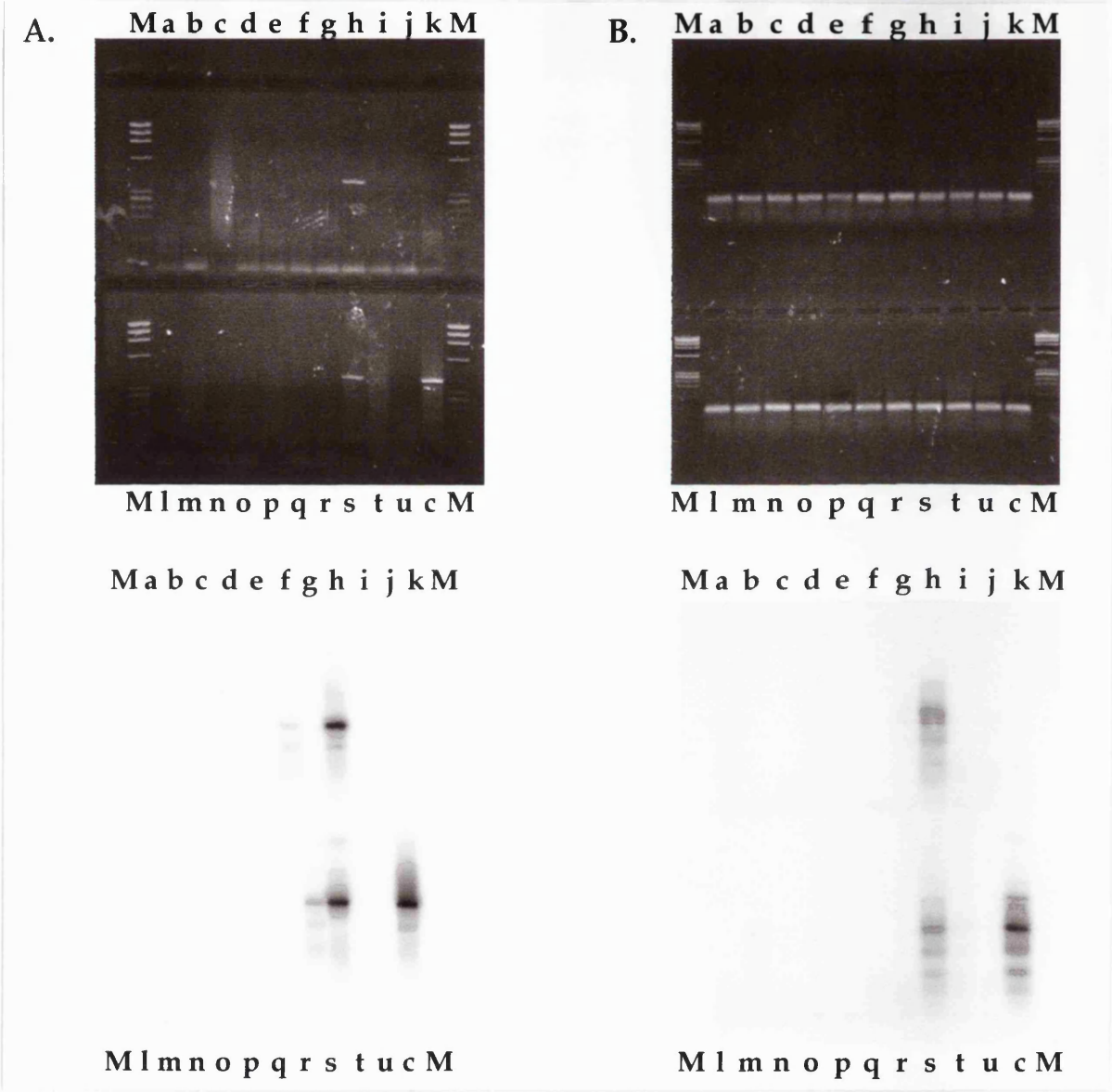
Figure 3.3 PCR Screen and Southern blot of the ICI Human Genomic YAC Library with primers SE3 and SE4



A. Lanes 1-40 contain the products of the PCR reaction with primers SE3 and SE4 on the primary ICI YAC pools. Lane c is the control reaction using 1ng human genomic DNA as template. Lane M is the marker lane containing ϕX174 *HaeIII* digest.

B. The blots of the gels in *A* were hybridised with a probe to *Nfv2* overnight, washed to 0.1x SSC and exposed to a phosphoimager screen for 1 hour.

Figure 3.4 PCR Screen and southern blot of the HGMP Human Genomic PAC Library with primers pairs SE3 and SE4, and SE5 and 6

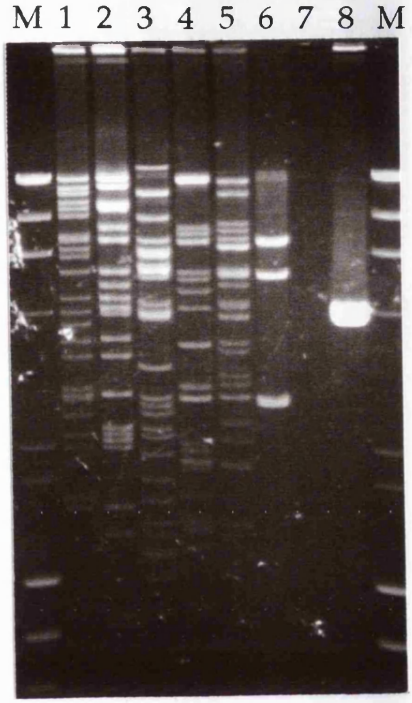


A. Lanes a to u contain the products of the PCR reaction using primers SE3 and 4 on the primary PAC pools a-u respectively.
Lane c is the control reaction using 1ng human genomic DNA as template. Lane M is the marker lane containing ϕX174 *HaeIII* digest.
The blot of the gel was hybridised overnight with the human *Nfo2*-specific probe and washed down to 0.1xSSC. It was exposed to a phosphorimager screen for 1 hour.

B. Lanes a to u contain the products of the PCR reaction using primers SE5 and 6 on the primary PAC pools a-u respectively.
Lane c is the control reaction using 1ng human genomic DNA as template. Lane M is the marker lane containing ϕX174 *HaeIII* digest.
The blot of the gel was hybridised overnight with the human *Nfo1*-specific probe and washed down to 0.1xSSC. It was exposed to a phosphorimager screen for 1 hour.

The smaller of this pair, PAC 107A22 was chosen for further characterisation of the human FV1 region.

Figure 3.5 *NotI/BglII* PAC restriction digests



The lanes were loaded as follows:

1. PAC107A22 *NotI/BglII*
2. PAC277I9 *NotI/BglII*
3. PAC107E15 *NotI/BglII*
4. PAC277F8 *NotI/BglII*
5. PAC277M18 *NotI/BglII*
6. PAC107N6 *NotI/BglII* (empty pCYPAC2N vector)
7. empty well
8. plasmid containing *Fv1* (positive control for blot)

M is the marker lane containing λ*HinDIII*

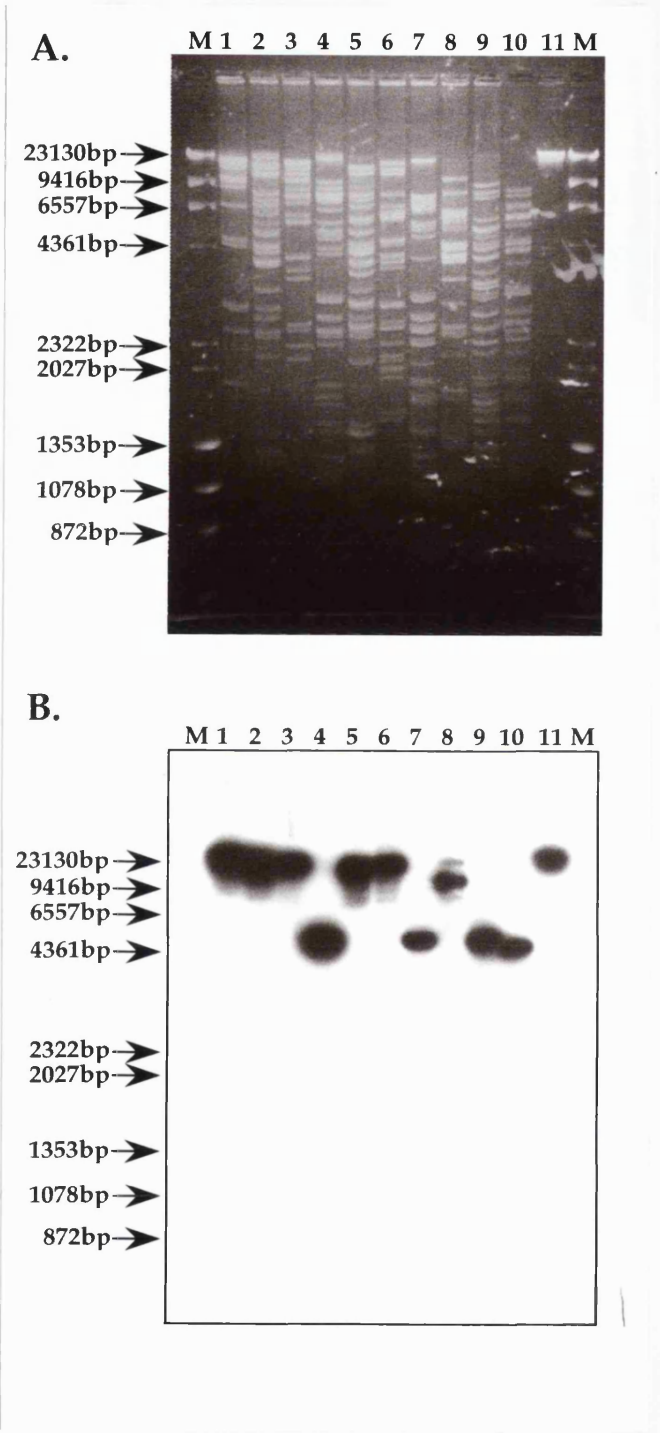
3.4 Analysis of PAC clones containing *NFV1* and *NFV2* human homologous sequence

During the isolation of these PACs, the *Fv1* ORF was identified in this lab[170]. This enabled the PACs to be screened by hybridisation with a probe specific to *Fv1* to determine whether they contained sequence with any significant homology to this gene.

3.4.1 Analysis of PACs by hybridisation

In order to ascertain whether both markers lay on a common restriction fragment, and also to verify that the PAC107A22 picked from the frozen tertiary screen plate was the correct clone, various restriction digests of this clone were run out, blotted and hybridised overnight with probes to *NFV1* and *NFV2* (Figures 3.7 and 3.8). A single, strongly-hybridising fragment was present in all the restriction digests, showing that each probe sequence lay within a single

Figure 3.6 Restriction digest and southern blot of PAC107A22 (probed for human *NFV2*)



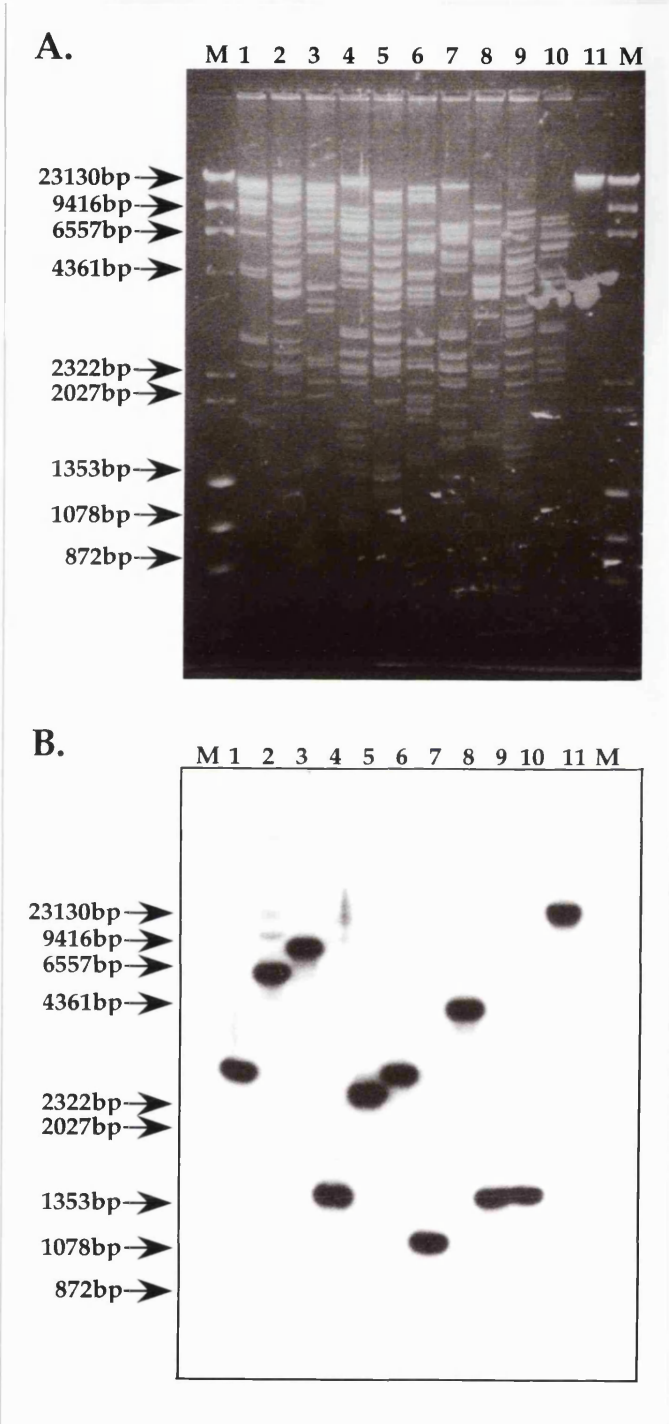
A. Lanes 1-11 contain the following:

1. PAC107A22 *BamHI*
2. PAC107A22 *HinDIII*
3. PAC107A22 *EcoRI*
4. PAC107A22 *SacI*
5. PAC107A22 *BamHI/HinDIII*
6. PAC107A22 *BamHI/EcoRI*
7. PAC107A22 *BamHI/SacI*
8. PAC107A22 *HinDIII/EcoRI*
9. PAC107A22 *HinDIII/SacI*
10. PAC107A22 *EcoRI/SacI*
11. PAC107A22 uncut

Lane M is the marker lane containing λ*HinDIII*/ϕX174 *HaeIII* digest

B. The blot was hybridised overnight with the human *Nfv1*-specific probe and washed down to 0.1xSSC. It was exposed to a phosphorimager screen overnight.

Figure 3.7 Restriction digest and southern blot of PAC107A22 (probed for human *NFV1*)



A. Lanes 1-11 contain the following:

1. PAC107A22 *BamHI*
2. PAC107A22 *HinDIII*
3. PAC107A22 *EcoRI*
4. PAC107A22 *SacI*
5. PAC107A22 *BamHI/HinDIII*
6. PAC107A22 *BamHI/EcoRI*
7. PAC107A22 *BamHI/SacI*
8. PAC107A22 *HinDIII/EcoRI*
9. PAC107A22 *HinDIII/SacI*
10. PAC107A22 *EcoRI/SacI*
11. PAC107A22 uncut

Lane M is the marker lane containing λ*HinDIII*/ϕX174 *HaeIII* digest

B. The blot was hybridised overnight with the human *Nfv1*-specific probe and washed down to 0.1xSSC. It was exposed to a phosphorimager screen overnight.

restriction fragment in each case. However, the pattern of these fragments seen in these restriction digests was different with the *NFV1*-specific and *NFV2*-specific probes, indicating that the 2 probe sequences were not located on a single restriction fragment in any of these restriction digests.

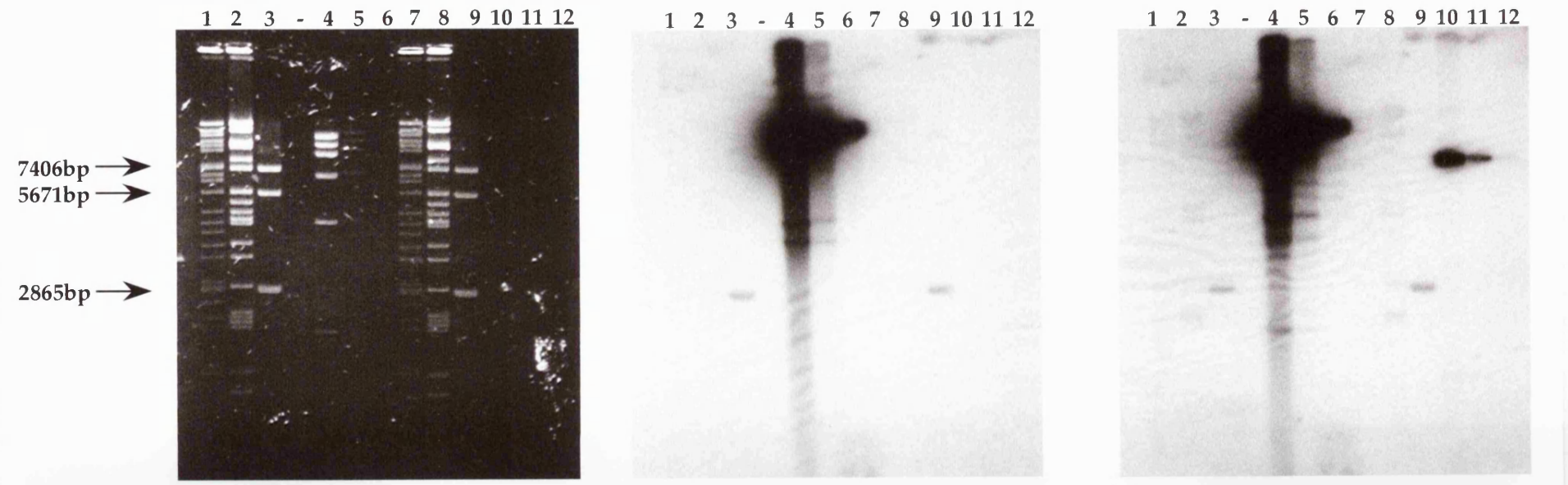
3.4.2 Probing PACs with mouse *Fv1* ORF sequence

In order to demonstrate that these clones, shown to be positive for the human homologues of the flanking genes, also contained sequence with homology to mouse *Fv1*, *NotI/BglII* restriction digests of PAC107A22, PAC277I9 and a vector-only clone, PAC107N6, were run out and blotted (Figure 3.9). Also loaded onto the gel as positive controls were *EcoRI* restriction digests of the cosmids containing the *Fv1* ORF and the marker *Nppa* (D11c11 and D11c10 respectively) at 3 dilutions. These cosmids were derived from the YAC originally found to have *Fv1* activity during the cloning of *Fv1*, YACD11[170]. This blot was hybridised overnight with a probe specific to the *Fv1* ORF (Figure 3.9b) and washed at low stringency (4x SSC). As the probe is specific to mouse DNA and the PACs contain human DNA, this stringency was chosen to ensure hybridisation of probe even to sequences of relatively low homology. The probe was made by random-prime labelling a GT16/GT17 PCR-amplified product from D11c11 template. Only a band corresponding to the *Fv1* ORF in the cosmid D11c11 hybridised intensely. Much weaker hybridisation to a single band was seen in the PAC restriction digests; as this corresponded to an identical band in PAC107N6 (vector-only) restriction digest, this hybridisation appears to be due to a very small degree of cross-hybridisation of the probe with the PCYPAC2N vector. No hybridisation, however, was evident to the insert sequences of PAC107A22 and PAC277I9. As the *Fv1* fragment from D11c11 did hybridise intensely with the probe, the hybridisation had worked well. Surprisingly, this finding suggests that the PACs contain no sequence of significant homology to *Fv1*.

3.4.3 Probing PAC clones with mouse *Nppa* sequence

Nppa (the gene for atrial natriuretic factor) has been mapped to within 4.6cM of *Fv1*[302], and was used as a marker to isolate a YAC with *Fv1* activity[16]. As the region covered by the 2 overlapping clones is relatively large (as judged by the large number of unique bands), the *Nppa* locus may be included in this area.

Figure 3.8 Digests and southern blot of PAC107A22, PAC277I9, PAC107N6 (empty pCYPAC2N vector) and cosmids containing *Fv1* and *Nppa* sequence



A. Lanes 1-12 contain the following:

1. PAC107A22 *NotI/BglII*	4. D11c11 (*Fv1*) *EcoRI* - ~100µg	7. PAC107A22 *NotI/BglII*	10. D11c10 (*Nppa*) *EcoRI* - ~10ng
2. PAC277I9 *NotI/BglII*	5. D11c11 (*Fv1*) *EcoRI* - ~10ng	8. PAC277I9 *NotI/BglII*	11. D11c10 (*Nppa*) *EcoRI* - ~1ng
3. PAC107N6 *NotI/BglII*	6. D11c11 (*Fv1*) *EcoRI* - ~1ng	9. PAC107N6 *NotI/BglII*	12. D11c10 (*Nppa*) *EcoRI* - ~100pg

B. Southern blot of gel A hybidised with an *Fv1* probe. The blot was hybridised overnight and washed down to 4x SSC.

C. Previous southern blot washed down to 0.1x SSC at 65°C for 1 hour before re-hybridising with a *Nppa* probe. The blot was hybrid overnight and washed down to 4x SSC.

To ascertain whether this was the case, Southern blots of the PAC digests were hybridised with a probe specific to *Nppa*.

The blot used in the previous section was washed at 65°C for 1 hour at 0.1x SSC to remove any non-specific hybridisation signal. It was then hybridised with a probe specific to the *Nppa* gene (Figure3.9c). This was made by random-primed labelling a purified 537bp PCR-amplified product corresponding to nt 572-1108 of the submitted sequence (Genbank number K02781; a gift from Steve Best). Still evident on the blot was the signal from the previous hybridisation between the *Fv1*-specific probe and the *Fv1* cosmid restriction digest. Following hybridisation with the *Nppa* probe, additional intense bands were seen only to the cosmid restriction digest containing the *Nppa* sequence itself. No hybridisation was evident between the probe and the restriction digest fragments of the PAC clones.

3.5 Subcloning *Fv1* region from PAC 107A22

The result from the hybridisation of *Fv1* probe to the PAC southern blots suggested that these clones did not contain an *Fv1* human homologue. Concurrent with this finding the retroviral nature of *Fv1* was identified and the gene itself was shown to be restricted to mice only. Although this now meant there was probably no human homologue of *Fv1* to be cloned, the subcloning of this region was still continued, as this region now represented the preintegration site of the event that gave rise to the *Fv1* gene. By cloning and sequencing this region it was hoped we could ascertain how the region had changed during the formation of the gene, what sequence (if any) had been lost, and exactly what had been gained. For this reason, the analogous sequence in rat (the closest $Fv1^-$ relative to the mouse) was also cloned and sequenced as well (P. Le Tissier, unpublished).

3.5.1 Subcloning largest $NFV1^+$ and $NFV2^+$ fragments

As the average insert size of the HGMP PAC library clones is 120kb, the next part of the project was to subclone the region containing the flanking markers into as small a fragment as possible to facilitate sequencing. However, the sequences homologous to the *NFV1* and *NFV2* probes do not lie on common restriction fragments with the restriction enzymes used (see section 3.4.1). In mice the region between *NFV1* and *NFV2* is around 8kb. The largest $NFV1^+$ and $NFV2^+$

fragments in the PAC clones are 20kb and 8kb respectively (see Figures 3.7 and 3.8). It was therefore possible that these 2 human fragments might overlap. In order to test this, these fragments were subcloned.

The largest bands from *BamHI* and *EcoRI* restriction digests of PAC107A22 were gel-purified and subcloned into PCRscript vector. DNA from 12 transformants of each subcloning were digested with *BamHI* or *EcoRI* as appropriate and blotted. These blots were then screened by hybridisation either with a *NFV1* (for the PAC107A22*BamHI*-derived clones) or a *NFV2* (for the PAC107A22*EcoRI*-derived clones) probe (data not shown). Inserts of positive clones showed strong hybridisation. A positive clone from each subcloning was selected and these were designated pA22B and pA22E for the *NFV1*$^+$ *BamHI* and *NFV2*$^+$ *EcoRI* fragments (in PCRscript) respectively.

3.5.2 Demonstration of overlap by hybridisation

To demonstrate that these clones overlapped (or not), a fragment (apparently) common to both was used to construct a probe which was then hybridised with digests of both clones. The presence of a hybridising band in both clones would indicate that they overlapped.

Restriction digests of pA22B and pA22E were run in parallel and blotted (Figure 3.10a). This blot was hybridised with a probe specific to the 800bp *SacI/EcoRI* pA22B fragment which appears to be common to both of the clones (marked on Figure 3.10a). This probe was constructed by random-prime labelling this gel-purified fragment from a previous digest. This probe hybridised to single unique fragments in the *BamHI/HinDIII* and *SacI/HinDIII* double digests, and to the common 800bp band in the *SacI/EcoRI* restriction digest. This shows that the PAC107A22 *BamHI* and *EcoRI* fragments subcloned do overlap. Further restriction enzyme digests of these 2 clones allowed an (overlapping) restriction map to be constructed (see Figure 3.10b).

3.5.3 Sequencing by a transposon-based strategy

As outlined in the previous section, the *FV1* region from human was subcloned into 2 overlapping clones and comprised over 28kb of sequence. To sequence this region a transposon-based strategy was chosen. This advantage of this method was that it required no further subcloning of the fragments and, once

Figure 3.9a Restriction digests and southern blot of pA22B and pA22E

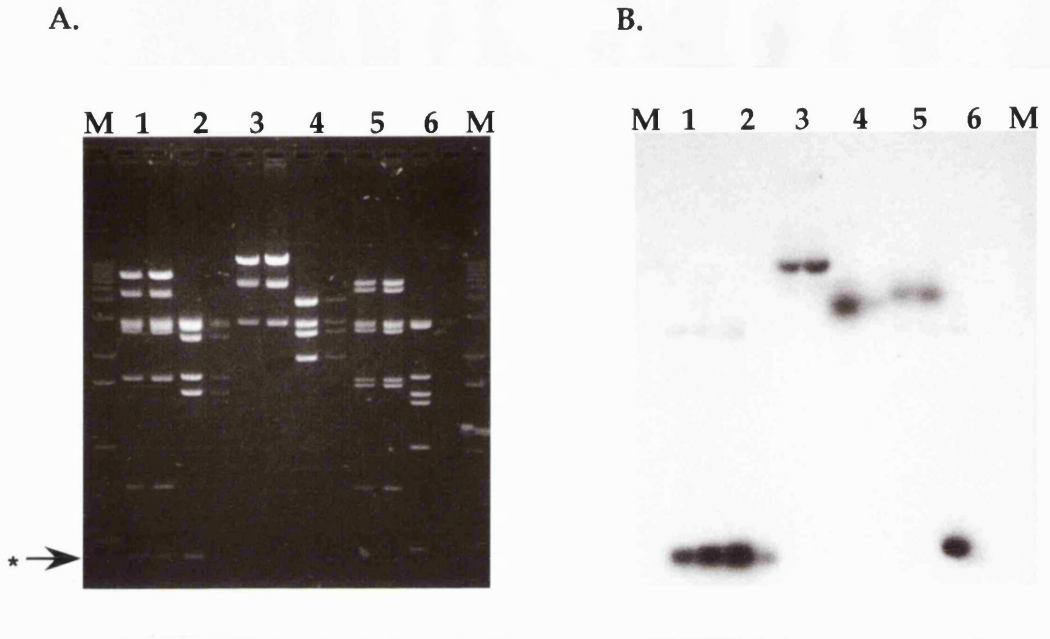
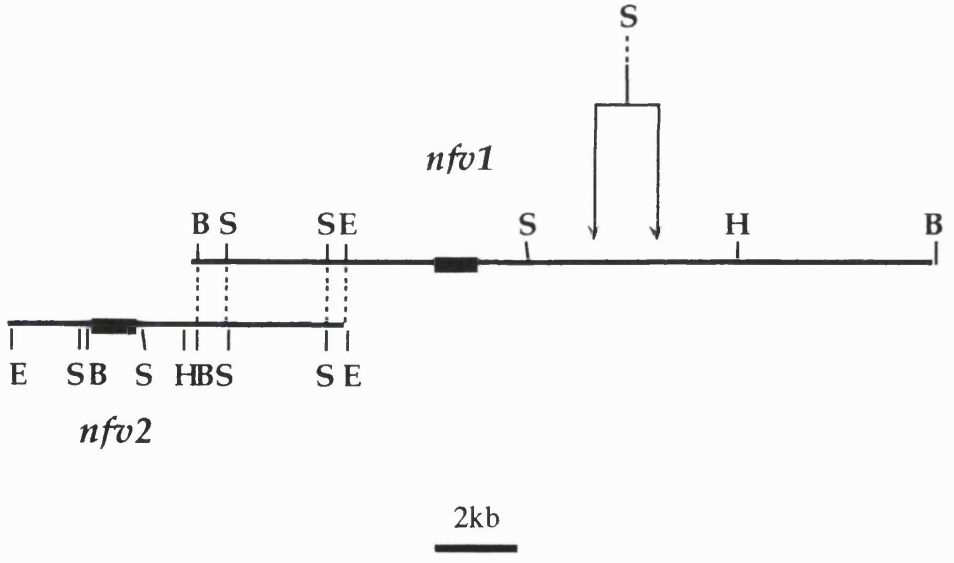


Figure 3.9b Restriction map of the *FV1* region in human



A. Lanes 1-6 contain the following digests in duplicate:
1. pA22B *SacI/EcoRI* 3. pA22B *BamHI/HinDIII* 5. pA22B *SacI/HinDIII*
2. pA22E *SacI/EcoRI* 4. pA22E *BamHI/HinDIII* 6. pA22E *SacI/HinDIII*

3 Southern blot of gel A. The probe used was constructed using the pA22B *SacI/EcoRI* fragment (marked on gel A with an asterix). The blot was hybridised overnight and washed to 0.1x SSC

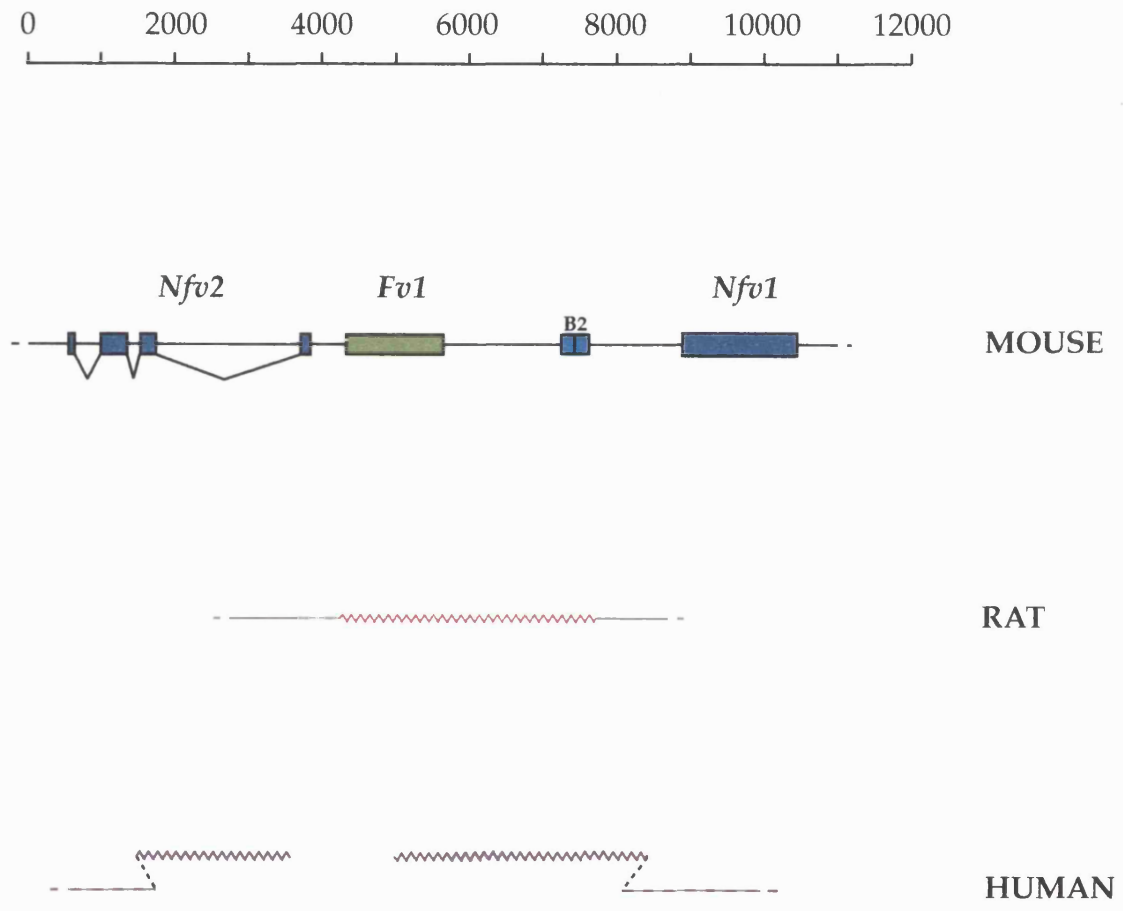
the transposon-tagged clones had been generated, enabled sequence covering the whole region to be generated simultaneously using only 2 primers. In this way the sequence of pA22B and pA22E was obtained in a short space of time by producing short sequences from across the length of the human FV1 region. These were then used to construct a contig of the region and ultimately a consensus sequence.

3.6 Comparative sequence analysis of *Fv1* region from human, rat and mouse

As stated at the start of this chapter, the region in mouse containing *Fv1* has been sequenced ([16] and P. Le Tissier, unpublished). By comparing the human sequence with the mouse sequence we can determine what is common to both and what is different. This comparison does not however allow the determination of whether a region unique to one has been gained in that organism alone, or has been lost from the other. For this reason, concurrent with the cloning and sequencing of the human region, the analogous region from rat was also cloned and sequenced (P. Le Tissier, unpublished). By comparing the sequence between 3 different organisms, one can determine exactly what the differences in this region between these 3 species are. This has shown exactly what sequences have been gained by insertion events, what sequences have been lost by deletion, and whether any rearrangements have occurred (see Figure 3.11).

The sequence from rat has a high degree of homology (80-90%) to the regions immediately upstream and downstream of the *Fv1* gene. Upstream of the gene, homology was found to extend from the *Nfv2* gene to within 100bp of the starting methionine of *Fv1*. Downsteam homology was shown to extend from the *Nfv1* gene to the point of the B2 insertion in mouse. The sequence between corresponds to around 3kb with no homology to either the mouse sequence or any sequence submited to date on EMBL, DDBJ and PDB (determined by BLAST search). The *Fv1* region from rat therefore contains the analagous regions immediately either side of the *Fv1* gene in mouse, but lacks the 3kb of sequence containing the gene itself. Instead, this region contains 3kb of unique sequence. To determine whether this sequence has been ganed in rat or has been lost in mouse, the sequence of the corresponding region isolated from human was used as a comparison. Unfortunately, this region is extremely rich in repetitive elements which not only hampered the assembly of sequence (~1.5kb of the

Figure 3.10 Comparative sequence analysis of the *Fv1* region from mouse, rat and human



central region remains unassembled due to this- see Figure 3.11) but also made the demonstration of contiguous homology difficult. In spite of this, the region in human was shown to contain patches of homology that corresponds to the regions flanking the *Fv1* gene in mouse, although these do not extend as close to the gene as with the rat. The sequence between these regions similarly shows no homology to those in the databases, or to the sequence in rat. Therefore, the comparison between the mouse, rat and human *Fv1* region has shown that the regions flanking the preintegration site appear to have remained intact during the formation of the *Fv1* gene in mouse. However, this comparison could not determine whether the unique sequence in rat, in the same position as the *Fv1* gene in mouse, has been lost during the formation of the *Fv1* gene in mouse, or has arisen since mouse and rat diverged.

Once the *Fv1* region had been cloned from human, it was shown by hybridisation that there was no sequence homology at all between this region and the *Fv1* ORF. It appeared therefore that the analogous region in humans did not contain the human homologue of *Fv1*. This was later confirmed by sequencing. While this was unexpected at the time, it was reinforced independently (see section 7.1.2). The reason for its absence from all genera except *Mus* came later with the discovery (by sequence similarity) that *Fv1* was not part of the 'natural' genetic complement of the mouse but was instead retrovirally-derived. The significance of this finding was that the *Fv1* locus now appears to be a recent addition, and as such would only be present in the genus in which the integration event had occurred. It would therefore be not only unlikely but improbable that homologues would existed in distantly-related species, such as humans. The region cloned from human therefore represents the *preintegration* site of the integration event that led to the formation of the *Fv1* gene. Its sequence could therefore be used to identify what has been gained or lost following the integration of *Fv1*. At this point in the project we decided to change direction and investigate the evolution of the *Fv1* gene.

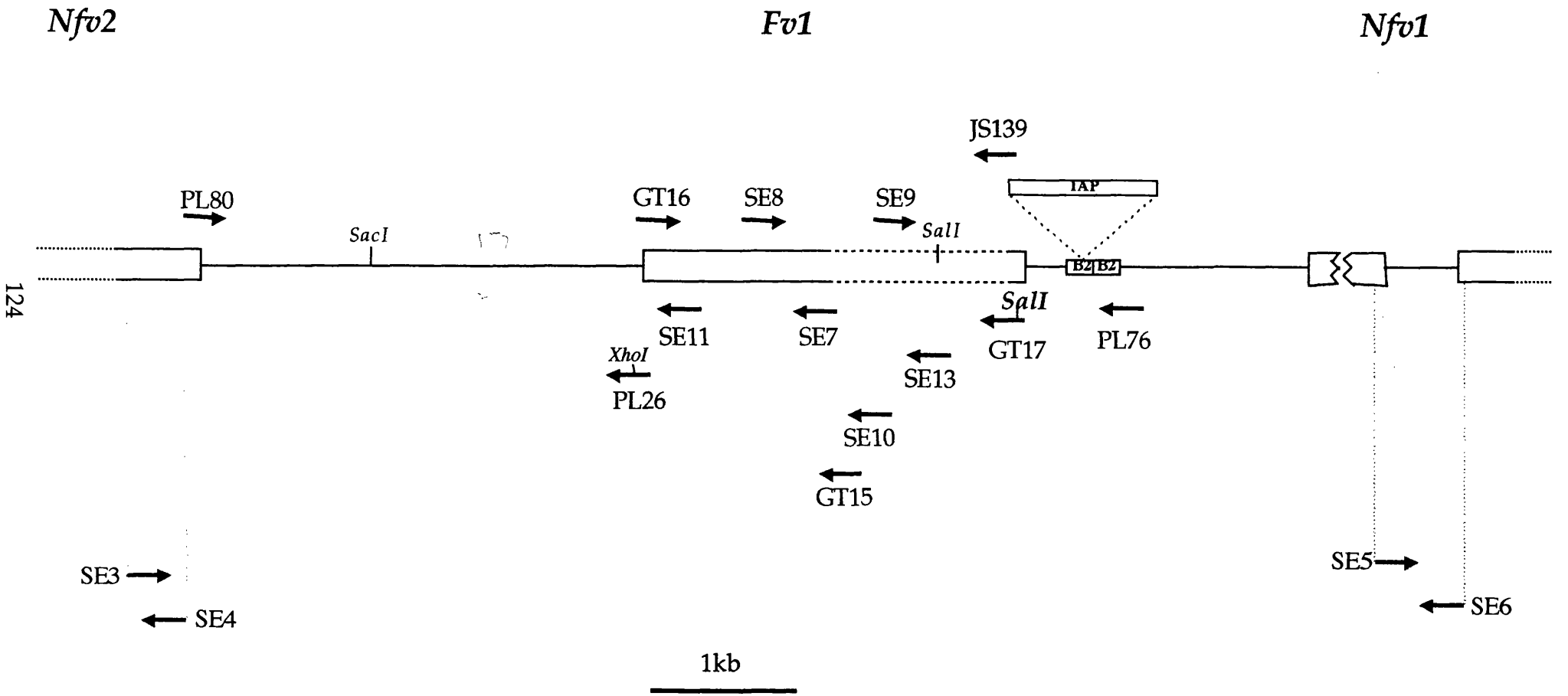
Chapter 4. The Evolution of *Fv1*

Fv1 activity appears to be restricted to laboratory strains of mice and their immediate feral relatives[173]. *Fv1* has been shown by hybridisation to be absent in all non-*Mus* species looked at, including its closest relative, the rat[170]. The development of the *Fv1* phenotype therefore appears to have arisen relatively recently in the *Mus* genus. To study precisely how *Fv1* activity has arisen in *Mus*, the *Fv1* region from members across the genus was analysed. The aim of this project was to clone the *Fv1* region from a group of mice and sequence the *Fv1* ORF (see Figure 4.1). By mapping and comparing the sequence and structural features of the gene, both common and unique among family members, it was hoped that an accurate picture of how *Fv1* has changed during the evolution of the *Mus* genus, and hence how the gene has changed during the development of *Fv1* activity, could be obtained.

4.1 PCR-amplification and subcloning of the *Fv1* ORF

The mice used in this study are shown in Table 4.1. The *Fv1* ORF was PCR-amplified from many of the mice used in this study using the primers GT16 and GT17. The PCR-amplified products obtained fell into 2 discrete sizes of ~2600bp and ~1300bp, corresponding to absence (as in the $Fv1^b$ allele) or presence (as in the $Fv1^n$ allele) of the $Fv1^n$-associated deletion (see Figures 4.1 and 4.2). For some of the samples, no PCR product was obtained using these primers due to mutations in the primer sites; in these cases, primers PL80 and GT17 were used (see Figure 4.1). GT16 and GT17 permit amplification of the sequence immediately before the 5′ end of the *Fv1* ORF to 2684bp downstream of the starting methionine in the $Fv1^b$ allele and 1340bp in the $Fv1^n$ (see Figure 4.2). The GT16 primer contains an internal *BglII* site and the GT17 primer contains an internal *SalI* site. The product of this PCR reaction was therefore subcloned as a *BglII*/*SalI* fragment into M13 phage. From sequencing the region previously from inbred mice there is known to be a *SalI* site upstream of the GT17 primer binding site, in the region that is deleted in the $Fv1^n$ allele (see Figure 4.2). Sequence analysis has revealed that in a proportion of the mice used in this study containing the undeleted form of *Fv1* (i.e. the $Fv1^b$-like form), this *SalI* site in the *Fv1* ORF has been lost by a single point mutation in the restriction site. This polymorphism was found in *Mus macedonicus*, *Mus spicilegus*, *Mus spretus* and

Figure 4.1 Primer map of the *Fv1* region



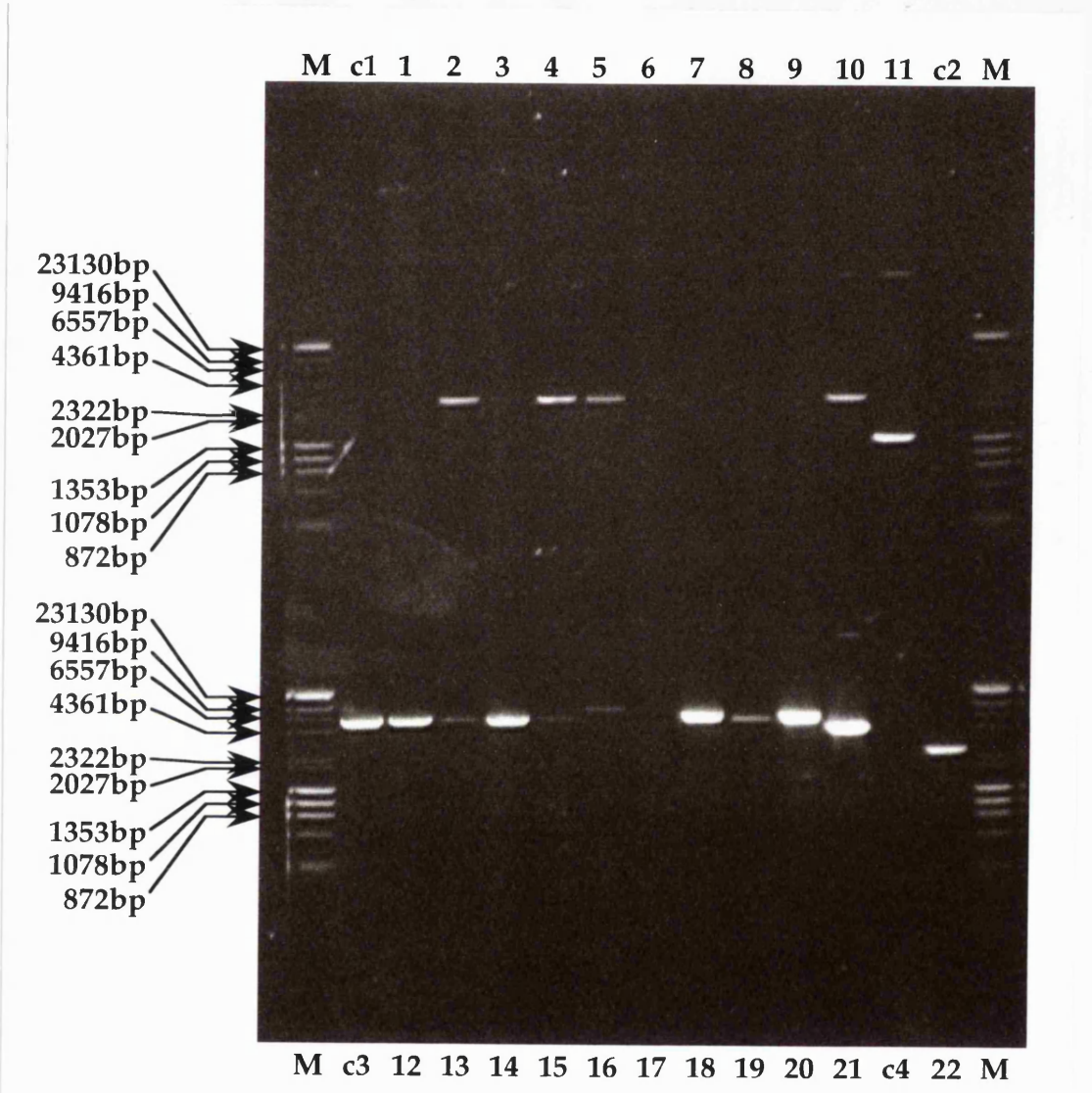
The Figure shows the primer map for the *Fv1* region in mouse. The dotted line in the *Fv1* ORF represents the region that is deleted in the *Fv1*^{n} allele. The IAP insertion is associated with the *Fv1*^{n} allele, and is inserted in a B2 tandem repeat. Also shown on the map are the primers SE3-6 designed to regions of the human homologues of *Nfv1* and *Nfv2* and the position of the corresponding regions in mouse

Table 4.1 The *Fv1* gene in the mice used in the sequencing study

Name	Species	Deletion at residue 437	*Fv1* typing data
FAM	*Coelomys famulus*	-	null[173]
PTX	*Pyromys platythrix*	-	null[173]
MIN	*Mus minutoides*	-	-
DUN	*Mus dunni*	-	null?
COK	*Mus cookii*	-	null[173]
Mus caroli	*Mus caroli*	-	null[173]
KAR	*Mus caroli*	-	-
ZYP	*Mus spicilegus*	-	-
XBS	*Mus macedonicus*	-	-
SFM	*Mus spretus*	-	-
M. Spretus	*Mus spretus*	-	null[173]
MOL	*Mus musculus molissinus*	-	-
MOLD/Rk	*Mus musculus molissinus*	-	-
CIM	*Mus musculus castaneous*	-	
BIR	*Mus musculus bactrianus*	-	
WMP/Pas	*Mus musculus domesticus**	-	
SEC/ReJ	lab strain	-	B
RIII/DmMob	lab strain	-	
I/LNJ	lab strain	-	
PRO/1ReJ	lab strain	-	B[303]
FVB/NJ	lab strain	-	B
C57BL	lab strain	-	B
SWR/J	lab strain	+	N[304]
ST/bJ	lab strain	+	N[167]
BDP/J	lab strain	+	B[303]
NZB/BlNJ	lab strain	+	NR[169]
NZW/LacJ	lab strain	+	NR[169]
RF/J	lab strain	+	NR[169]
LG/J	lab strain	+	B[303]
M.MUS	*Mus musculus musculus*	+	NR-like[173]
CZECKI	*Mus musculus musculus*	+	
MBT	*Mus musculus musculus*	+	
SKIVE/Ei	*Mus musculus domesticus*	+	NR-like[173]
RBA/Dn	*Mus musculus domesticus*	+	
SK/CamEi	*Mus musculus domesticus*	+	N[303]
SF/CamEi	*Mus musculus domesticus*	+	
P.ATTECK	*Mus musculus domesticus*	+	N[303]
BZO	*Mus musculus domesticus*	+	
WSB/Ei	*Mus musculus domesticus*	+	

*sequence in consistent with that of *Mus spretus* (see section 4.2.1).

Figure 4.2 Representative PCR of mouse genomic DNA



The following samples were PCR-amplified using the primers GT16/GT17:

1. CIM	2. ZYP	3. XBS	4. KAR	5. COK	6. PTX
7. PAH	8. FAM	9. SFM	10. MOL	11. BZO	
C1. positive control		C2. negative control			

The following samples were PCR-amplified using the primers GT17/PL80:

12.ZYP	13. XBS	14. KAR	15. COK	16. PTX	17. PAH
18. FAM	19. SFM	20. MOL	21. BZO	22. CIM	
C3. positive control		C4. negative control			

100ng C57BL/6 DNA was used in the positive control reactions, dH_20 in the negative control reactions. The marker (M) was λ*HinDIII*/ϕ*HaeIII*.

Pyromys platythrix. In these cases the PCR product could still be cloned due to the presence of a *SalI* site located in the GT17 primer itself.

With some of the more distantly-diverged mice, the *Fv1* ORF was also subcloned using the primers PL80 and GT17. These 2 primers permit the amplification of sequence 3056bp upstream of the starting methionine of the *Fv1* ORF to 2684bp downstream of the starting methionine in the $Fv1^b$ allele and 1340bp in the $Fv1^n$. This subclone therefore contained the intergenic sequence between *Nfv2* and *Fv1*, and was used both for obtaining the sequence of the *Fv1* ORF and to determine the presence of 5′ insertions in these mice (see section 4.4). From sequencing the region previously from inbred mice there are known to be 2 *SacI* sites downstream of the PL80 primer binding site. The product of this PCR reaction was therefore subcloned as a *SacI*/*SalI* fragment into M13 phage.

In order to sequence both strands the PCR-amplified product was subcloned in both M13mp18 and M13mp19. During the sequencing it was noticed that at a very low frequency (less than 10^{-3} errors/base) there was disagreement between the sequence of these 2 clones from some of the PCR reactions. This error rate was assumed to be due to the inherent infidelity of the Taq DNA polymerase, even though Pfu DNA polymerase was added for its proof-reading activity. This problem was solved by subcloning a similar PCR-amplified product from a second PCR reaction, and using this to determine the correct sequence at the ambiguous base. In a small number of instances, the third clone differed from both other 2 clones subcloned from the original PCR reaction. In these cases the consensus sequence was determined by subcloning and sequencing a forth clone. In this way, the correct sequence of the *Fv1* ORF from each of the samples was determined.

4.2 Sequence data analysis

The sequence data obtained from the automated sequencing of clones was edited in the software package Sequence Analysis (Applied Biosystems) before assembly into contigs in Autoassembler (Applied Biosystems). The sequences were aligned in GDE prior to tree-building (Phylip); these data are in Appendices 3 and 4.

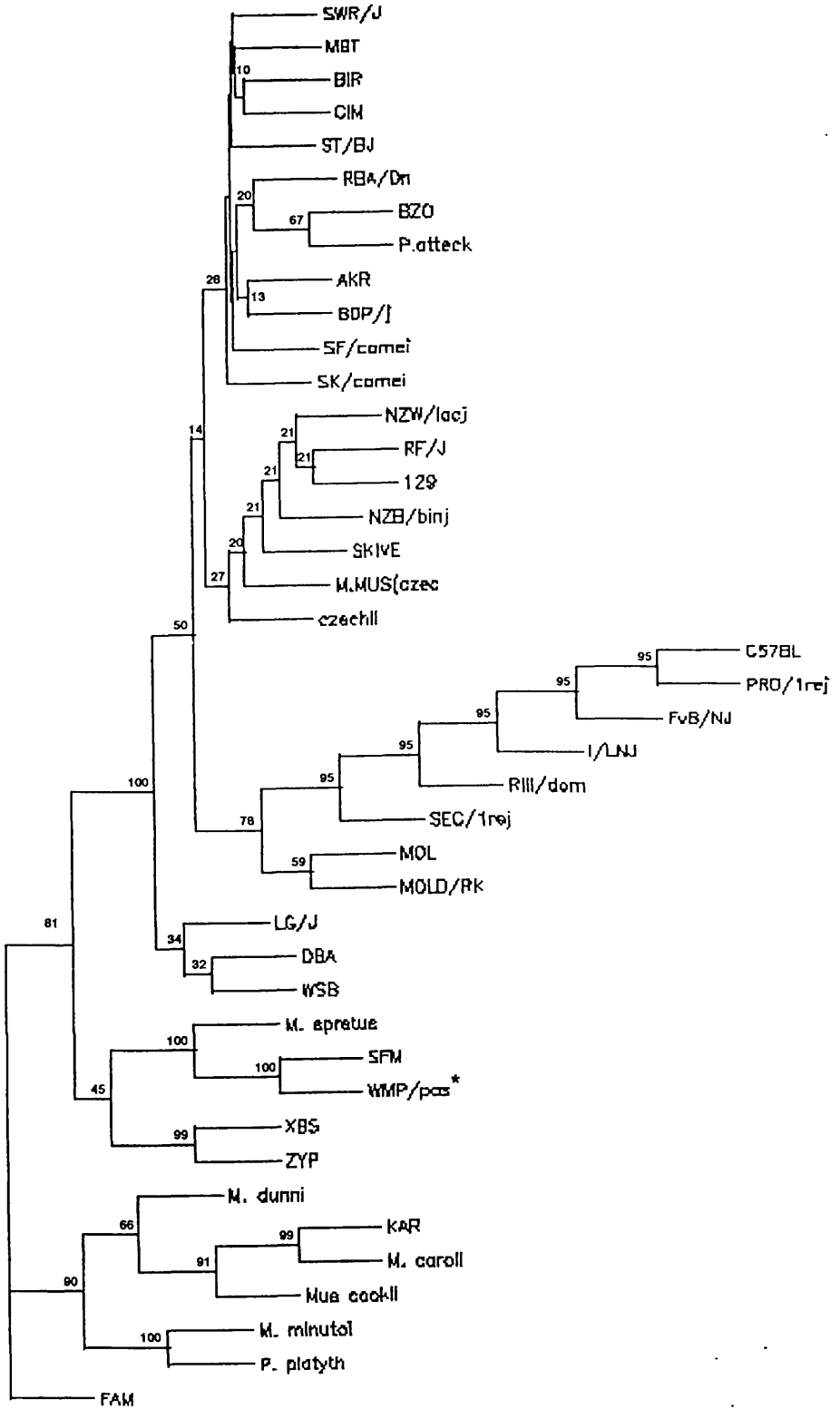
4.2.1 Construction of the *Fv1* phylogenetic tree

The DNA sequences of the *Fv1* ORFs from the different mice were aligned using GDE and this alignment was used to construct a phylogenetic tree based on neighbour-joining and maximum-likelihood methods using 100 bootstraps. As both methods yielded identical trees, only one is shown (obtained by the neighbour-joining method[305]- see Figure 4.3a). The tree was constructed using the *Fv1* ORF from *Coelomys famulus* as the outgroup. This species of mouse was the most distantly-diverged member of the genus *Mus* used in this study[279]. When the amino acid alignment was used instead of the DNA alignment, identical trees were obtained (data not shown).

The tree obtained was very similar to current phylogenetic trees for the *Mus* genus constructed from a variety of different methods, but did differ with respect to 2 branches: the *Mus dunni* and *Mus caroli/Mus cookii* lines both branch off with *Pyromys platythrix* in this tree, whereas current phylogenetic trees place this event directly after the branching of *Pyromys platythrix*[279]. In order to determine whether the tree obtained was significantly different from that published by Bonhomme[279], a Kishino-Hasegawa-Templeton statistical test[306] was carried out (data not shown). This showed that the phylogenetic tree obtained based on the sequence data was significantly different from the published tree due to these two changes.

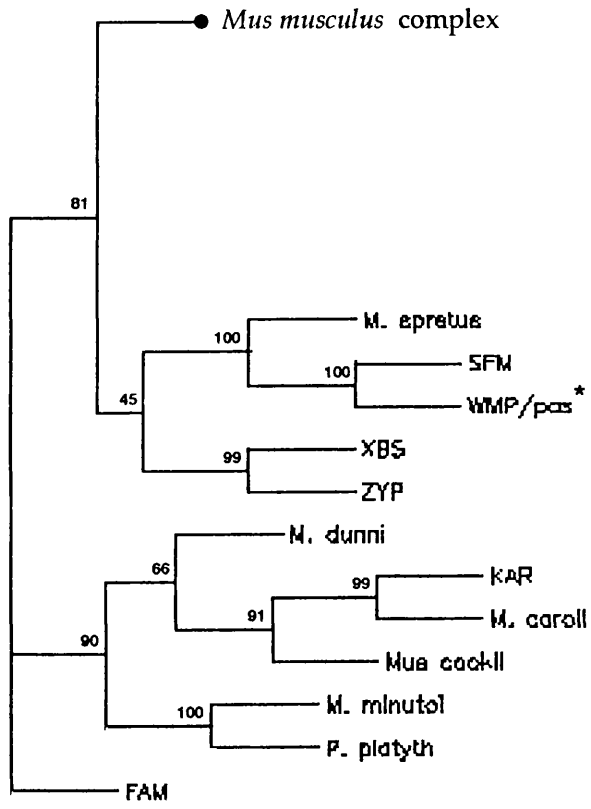
As shown in Figure 4.3, the first branching occurs between *Pyromys platythrix, Mus minutoides, Mus caroli, Mus cookii,* and the rest of the samples. The second branching takes place between *Mus macedonicus, Mus spicilegus, Mus spretus,* the *Mus musculus domesticus* sample WMP/Pas, and the *Mus musculus* complex. The WMP/Pas sequence is almost identical to the sequences for the *Mus spretus* samples, although it is distinct from these (see section 4.3). To confirm this result the WMP/Pas DNA sample was ordered from a second source (the Pasteur Institute) and the new sample used to obtain the *Fv1* sequence again. This *Fv1* sequence from the second sample was the same as the first. It is assumed, therefore, that WMP/Pas is not a true *Mus musculus domesticus* sample, but is instead a natural hybrid of *Mus musculus domesticus* and *Mus spretus.* Within the *Mus musculus* complex, the DBA-like sample WSB/Ei branches from the rest. Next, the other DBA-like sample branches; this sample also contains the $Fv1^{nr}$-associated change at position 352 (see Table 4.2). The next branching is between the $Fv1^{n}$-like and $Fv1^{b}$-like alleles; the exceptions to this are the *Mus musculus*

Figure 4.3a Phylogenetic tree based on the *Fv1* ORF sequence of the mice used in this study



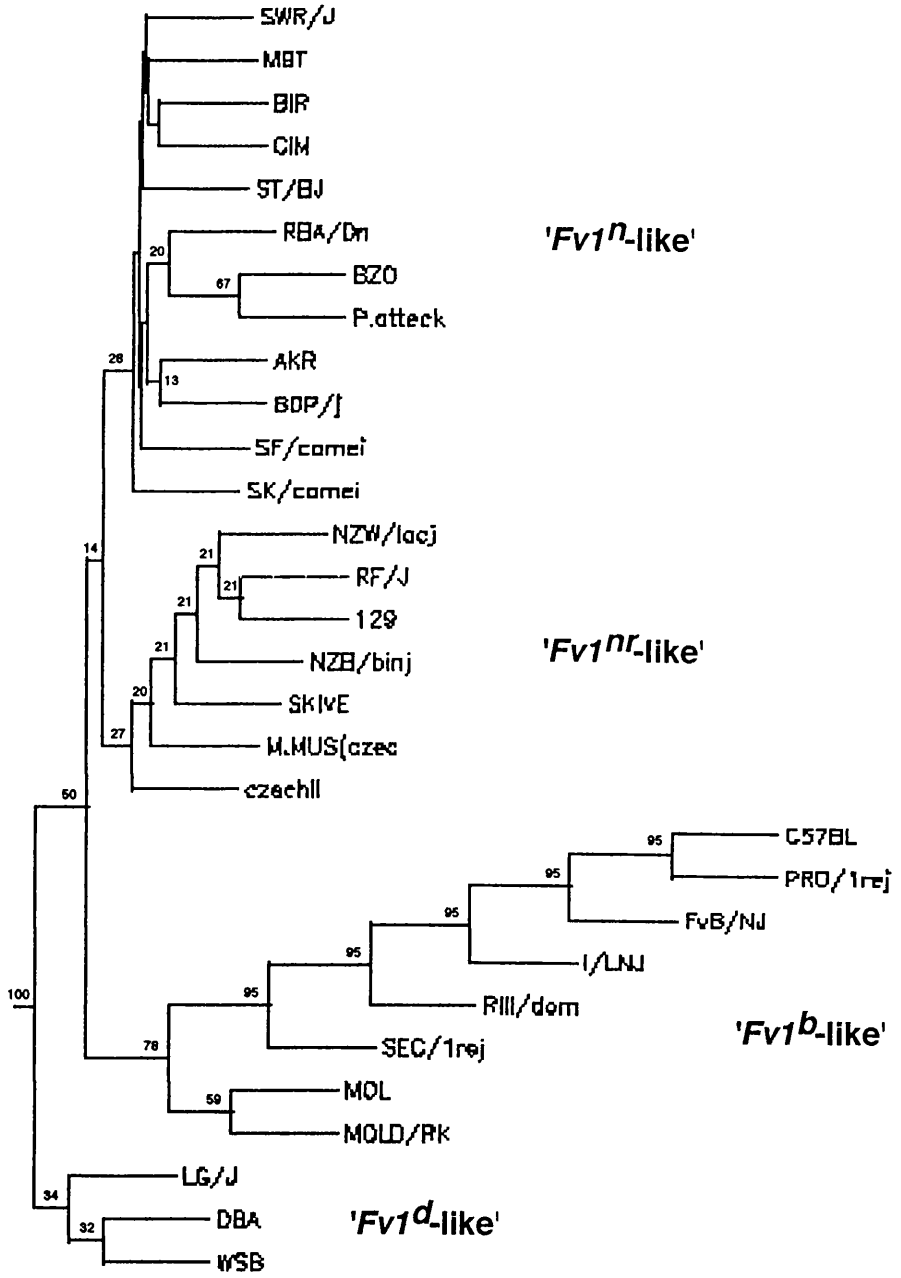
This is a neighbour-joining bootstrap tree based on the sequence of *Fv1* from the starting methionine to the deletion site of *Fv1*^{n}. The sequences were aligned using GDE and the tree constructed using PIE. The figures on each branch represent bootstrap support (from 100 replicates), with percentage values on those branches with support greater than 10 being shown.

Figure 4.3b Phylogenetic tree based on the *Fv1* ORF sequence of the mice used in this study omitting those from the *Mus musculus* complex



This is a neighbour-joining bootstrap tree based on the sequence of *Fv1* from the starting methionine to the deletion site of *Fv1*". The sequences were aligned using GDE and the tree constructed using PIE. The figures on each branch represent bootstrap support (from 100 replicates).

Figure 4.3c Phylogenetic tree based on the *Fv1* ORF sequence of the mice used in this study showing the *Mus musculus* complex only



This is a neighbour-joining bootstrap tree based on the sequence of *Fv1* from the starting methionine to the deletion site of *Fv1*". The sequences were aligned using GDE and the tree constructed using PIE. The figures on each branch represent bootstrap support (from 100 replicates), with percentage values on those branches with support greater than 10 being shown.

castaneous and *Mus musculus bactrianus* samples, which are placed in the *Fv1*^{n} group because they contain the *Fv1*^{n}-associated changes in their ORFs despite lacking the *Fv1*^{n}-associated deletion. *The Mus musculus molissinus* samples branch off with (but separately to) the *Fv1*^{b}-like alleles, probably because these sequences contain both an *Fv1*^{n} and *Fv1*^{b} change in their ORFs.

4.2.2 Distribution of changes believed to be important in determining *Fv1* phenotype

As observed in section 4.1, all *Fv1* ORFs subcloned with the primers GT16 and GT17 fell in 2 distinct sizes, 2600bp and 1300bp (see Figure 4.1). Sequencing has revealed that these sizes corresponded with either the presence or absence of the apparent deletion we see in the *Fv1*^{n} allele (see Table 4.1); until this work, it has not been known whether the smaller *Fv1*^{n} allele has arisen by deletion of the *Fv1*^{b} gene, or whether the larger *Fv1*^{b} allele has arisen as an insertion into the *Fv1*^{n} gene. *Fv1* ORFs isolated from the more distantly-diverged *Mus* species were of the larger *Fv1*^{b}-size only. The *Fv1* ORFs of *Mus musculus domesticus* and *Mus musculus musculus* subspecies were almost exclusively of the *Fv1*^{n}-size only. The only exception to this was the *Mus musculus domesticus* sample WMP/Pas; sequence analysis of this has shown this to be virtually identical to the sequences for *Mus spretus* (see Appendix 3).

The nature is of the *Fv1*^{n} and *Fv1*^{b}-specific mutations (marked in Figure 1.6) is shown in Table 4.2. The order of samples is the same as in Table 4.1. In the *Fv1*^{n} allele there is a lysine (K) residue at position 358 and a valine (V) residue at position 399, and in the *Fv1*^{b} allele there is a glutamic acid (E) residue at position 358 and an arginine (R) residue at position 399. The *Fv1*^{nr} change refers to the amino acid change at position 352 seen in the *Fv1*^{nr} allele (found in strains such as 129) which is thought to be responsible for its modified phenotype. Five out of a further six strains shown to carry this change in this study have been previously typed as having the *Fv1*^{nr} phenotype[173, 307], although the sixth (LG/J) has previously been typed as having the *Fv1*^{b} phenotype (see below). It seems, therefore, that this change alone is responsible for the *Fv1*^{nr} phenotype. Some of the wild mice species have also been typed for *Fv1* activity and have been shown to have the null (*Fv1*^{0}) allele. These mice all have the undeleted (*Fv1*^{b}-size) form of the gene, but have the *Fv1*^{n}-associated amino acids at positions 358 and 399.

Table 4.2 Genetic features of the *Fv1* ORF in the mice used in the sequencing study

Species	Residue at 270 (*Fv1*^{d} change)	Residue at 352 (*Fv1*^{m} change)	Residue at 358 (*Fv1*^{n}/*Fv1*^{b} changes)	Residue at 399	Deletion 437	IAP?
Coelomys famulus	Q	S	K	V	-	-
Pyromys platythrix	R	S	K	I	-	-
Mus minutoides	K	F	K	E	-	-
Mus dunni	Q	S	K	E	-	-
Mus cookii	Q	S	K	V	-	-
Mus caroli	Q	S	K	V	-	-
Mus caroli	Q	S	K	V	-	-
Mus spicilegus	Q	S	K	V	-	-
Mus macedonicus	Q	S	K	V	-	-
Mus spretus	Q	S	K	V	-	-
Mus spretus	Q	S	K	V	-	-
M. m. molissinus	K	F	K	R	-	-
M. m. molissinus	K	F	K	R	-	-
M. m. castaneous	K	S	K	V	-	-
M. m. bactrianus	K	S	K	V	-	-
*M. m. domesticus**	Q	S	K	V	-	+
lab strain	K	S	E	R	-	-
lab strain	K	S	E	R	-	-
lab strain	K	S	E	R	-	-
lab strain	K	S	E	R	-	-
lab strain	K	S	E	R	-	-
lab strain	K	S	E	R	-	-
lab strain	K	S	K	V	+	+
lab strain	K	S	K	V	+	+
lab strain	K	S	K	V	+	+
lab strain	K	F	K	V	+	+
lab strain	K	F	K	V	+	+
lab strain	K	F	K	V	+	+
lab strain	Q	F	K	V	+	-
M. m. musculus	K	F	K	V	+	+
M. m. musculus	K	F	K	V	+	+
M. m. musculus	K	S	K	V	+	+
M. m. domesticus	K	F	K	V	+	+
M. m. domesticus	K	S	K	V	+	+
M. m. domesticus	K	S	K	V	+	+
M. m. domesticus	K	S	K	V	+	+
M. m. domesticus	K	S	K	V	+	+
M. m. domesticus	K	S	K	V	+	+
M. m. domesticus	Q	S	K	V	+	-

*sequence in consistent with that of *Mus spretus* (see section 4.2.1).

One inconsistency was found between previous *Fv1* typing and the sequence data obtained in this study: the laboratory strains BDP/J and LG/J have previously been typed as having the *Fv1*^{b} phenotype but sequencing revealed they possess *Fv1*^{n}-like sequence. However, there has been increasing evidence that BDP/J was originally mis-typed (B. Taylor, pers. comm.), and this is thought to be true for LG/J also.

4.2.3 Maintenance of the *Fv1* ORF

The *Fv1* ORF was maintained in all but 2 mice samples, *Mus dunni* and *Mus cookii* (see Appendix 4). In the case of *Mus dunni*, the ORF is truncated due to a single base pair deletion at position 224 that causes a frameshift and premature stop. Expression of this truncated message has been confirmed by ribonuclease protection experiments (data not shown). The product of this message is predicted to be 90 amino acids in length with a novel 13 residue C-terminus due to the frameshift. Analysis of the sequence has identified 3 other frameshift mutations within 300bp of this insertion. In the case of *Mus cookii*, the *Fv1* ORF is prematurely truncated due to a single base pair transition (C→T) and a 5bp deletion at position 650 that causes the formation of a stop codon at this point. The truncated protein is predicted to be 217 residues in length with only the terminal residue changed, due to another single base pair transition at position 647. The introduction of this stop codon in the *Fv1* ORF of *Mus cookii* is the only change in the sequence that would lead to the premature truncation of the message. The lack of numerous mutations accrued over time may indicate that this truncation has occurred relatively recently.

It is also interesting to note that all of the deleted (*Fv1*^{n}-size) sequences share the same stop codon, whereas there is more variation in the undeleted sequences, which use one of four stop codons within a ~30bp region. This would seem to suggest that the *Fv1*^{n} allele may have arisen much more recently, so consequently has less allelic variation. However, although it is true that the deleted form arose more recently (shown by the presence of the undeleted form in the more distantly-diverged mice), *Fv1* activity is only found in the *Mus musculus* complex; when the stop codon of the undeleted forms of *Fv1* in these mice is examined, we find these also share the same stop codon. Therefore, based on the variation in stop codon usage, we cannot say which functional allele arose first (i.e. whether the deletion predated the *Fv1*^{b}-specific changes, or *vice*

versa), only that the undeleted form predated the deleted form (in agreement with the older mice exclusively having the undeleted form).

4.2.4 Downstream *SalI* site used for cloning

Upon sequencing these products it was found that, although this *SalI* site was found in the majority of the mice, it was not present in both of the *Mus spretus* samples, in *Mus spicilegus*, *Mus macedonicus* and *Pyromys platythrix*. As it was still present in the most distantly-diverged sample in the study, namely *Coelomys famulus*, we can infer that this *SalI* site has been conserved in all members except these 4 mice. Moreover, the loss of this site in *Mus spretus*, *Mus spicilegus* and *Mus macedonicus* was due to the same 2 base changes at positions 2865 and 2867 (see Appendix 3), and this is in accordance with the fact that these 3 members all branch from the rest of the genus *Mus* at the same time. Therefore these common differences are due to shared changes that have occurred in this branch shortly before these 3 species diverged from one another but after they split from the rest of the tree. *Pyromys platythrix*, however, diverged much earlier and as such, any difference shared between this mouse and *Mus spretus*, *Mus spicilegus* and *Mus macedonicus* would also be shared with the other members of the genus that diverged after *Pyromys platythrix*. As this is not the case, it is unsurprising to find that the loss of this site in *Pyromys platythrix* is due to a different, single mutation at position 2867.

4.3 Distribution of IAPs

One of the main structural differences between the $Fv1^n$ and $Fv1^b$ alleles is the presence of an IAP element inserted immediately downstream of the $Fv1^n$ ORF. The distribution of this particular element in the mice used in this study was determined by PCR-amplification from the *Fv1* ORF either across the site of IAP insertion (to show the absence of the insertion) or to the IAP element itself (to show its presence- see Figure 4.2). The primers used to PCR-amplify across the site of IAP insertion were GT16 and PL76. The primer amplifying from *Fv1*, GT16, is the same primer used for much of the cloning. This oligonucleotide primes just downstream from the 5′ end of the *Fv1* ORF. The other oligonucleotide, PL76, primes from the region between the B2 repeat and the gene flanking *Fv1* on the 3′ side, namely *Nfv2*. The primers used to PCR-amplify to the IAP itself were JS139 and GT16. JS139 is complementary to a portion of the

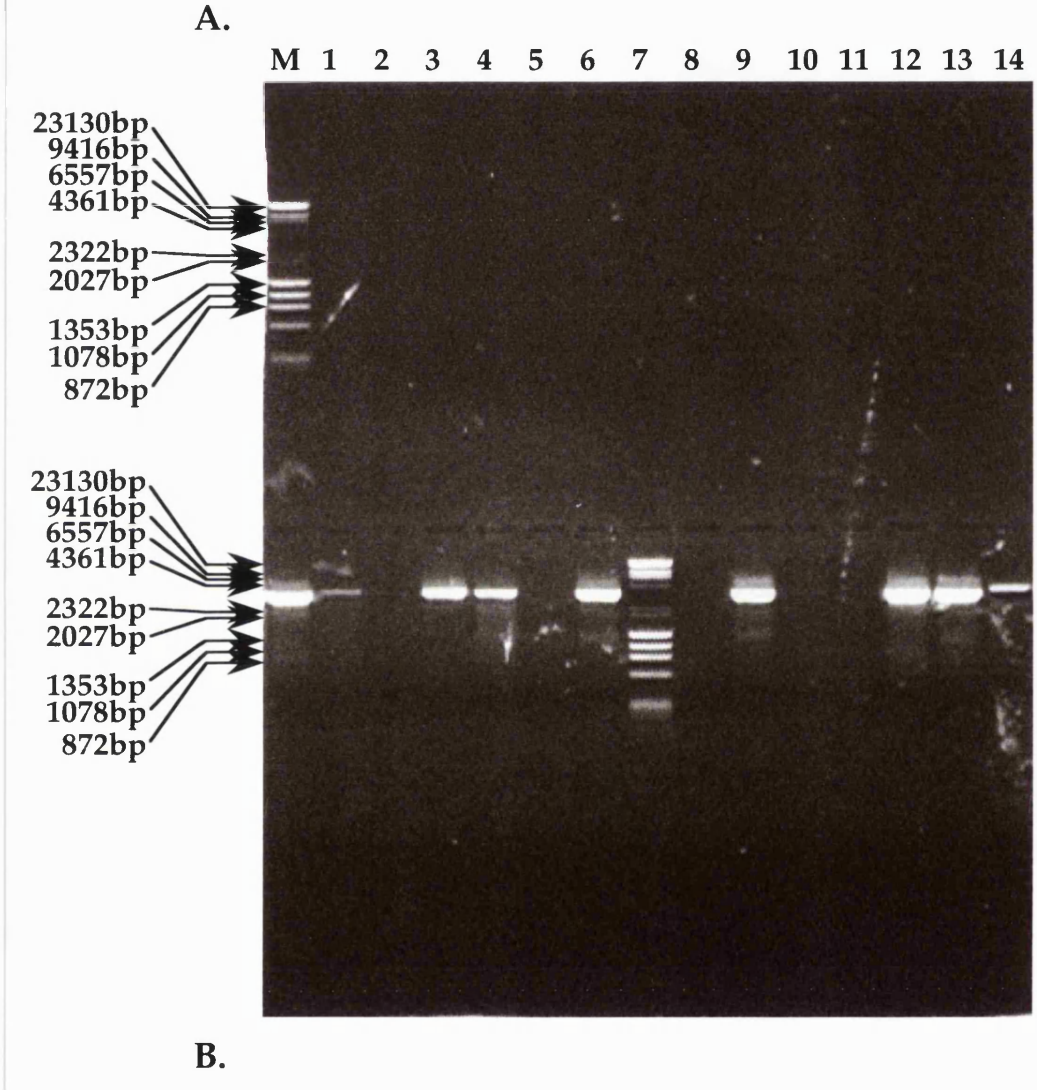
IAP LTR. Figure 4.4 shows the products of these PCR reactions on the genomic DNA of the mice used in this study. The results are summarised in Table 4.2. All but 2 of the mice harbouring an *Fv1*^{n}-size *Fv1* ORF have the corresponding *Fv1*^{n}-associated IAP element inserted immediately 3′ of the gene. The 2 mice which lacked this insertion also have a lysine→glutamine amino acid change at position 271. These 2 differences in the *Fv1*^{n} allele, the change at position 271 and the associated loss of the downstream IAP element, are characteristic of the *Fv1*^{d} allele, a less common allele of *Fv1* which has a modified *Fv1*^{n} phenotype[180]. With those mice harbouring the *Fv1*^{b}-size *Fv1* ORF, only WMP/Pas appears to have the *Fv1*^{n}-associated IAP insertion downstream of the gene. Sequence analysis of the genomic DNA-IAP junction, however, reveals that this element is different from that seen in the *Fv1*^{n} region, as shown by its site of insertion, located 504bp upstream of the *Fv1*^{n}-associated IAP insertion point. Additionally, this IAP element is in the opposite orientation to the *Fv1*^{n}-associated element, as determined by PCR and sequence data (data not shown). Therefore, all of the *Fv1*^{b}-size *Fv1* ORFs lack the *Fv1*^{n}-associated IAP element.

4.4 Investigating the sequence 5′ of the *Fv1* ORF

Fv1 is the product of a retroviral integration event. The *Fv1* ORF shows homology to part of the retroviral *gag* gene, and appears to be all that remains of the original integration event. Substantial deletions of proviral sequence have therefore taken place during the formation of the gene. To determine whether more of the virus has survived these deletion events in any of the older mice, the regions 5′ and 3′ of the gene were examined for the presence of insertions.

In order to obtain information concerning the sequence upstream of *Fv1*, the primers PL80 and GT17 were used to PCR-amplify the *Fv1* region from many of the mice (see section 4.1). The primer site for PL80 is located in exon 2 of the flanking gene upstream of *Fv1*, namely *Nfv2*. These 2 primers therefore permit the amplification of all sequence lying between *Nfv1* and the 5′ end of *Fv1*. In this way, any insertions/deletions in this sequence could be detected by an increase/decrease in product size, visualised by an appropriate shift in electrophoretic mobility on an agarose gel. This product was subcloned as a *SacI*/*SalI* fragment using the *SalI* in the GT17 primer and the *SacI* site 1325bp upstream of the starting methionine of the *Fv1* ORF (see Figure 4.1). In some mice, however, PCR-amplification of this region yielded very little product.

Figure 4.4 Determining the presence or absence of the 3' IAP in the mice



The following samples were PCR-amplified using the primers GT16/JS139 (in **A.**) and GT16/PL76 (in **B.**):

1. KAR	2. PTX	3. XBS	4. SFM	5. CIM	6. BZO
7. BALB/c	8. AKR	9.BIR	10. ZYP	11. -	12. MOL
13. FAM	14. COK				

These mice tended to be among the more-distantly diverged members of the sample, and so this problem was thought to be due to mutations in the primer site for either or both of these 2 primers. This problem was overcome by PCR-amplifying the upstream region as a shorter product using the primers PL80 and PL26; the *Fv1* ORF in these mice was subsequently amplified using the primers GT16 and GT17. As the PL26 primer contains an internal *XhoI* site and there is known to be a *SacI* site 1325bp upstream of the starting methionine (as indicated by the sequence of this region from C57BL; P. Le Tissier, unpublished), this product was subcloned as a *XhoI/SacI* fragment (see Figure 4.1). A significantly larger PCR-amplified product was only observed with *Pyromys platythrix* (Figures 4.1 and 4.5). Sequencing this has revealed an insertion of around 2kb that matches no sequences in the database (see Appendix 5).

4.5 Investigating the sequence 3′ of the *Fv1* ORF

As explained in section 4.2, the primer pair GT16 and PL76 was used to PCR-amplify across the site of IAP insertion seen characteristically in the *Fv1*^{n} allele. As the primer pair permits the sequence from the 5′ end of *Fv1* to the 3′ end of the downstream flanking gene, *Nfv1*, to be amplified, any insertions between the 2 genes would also be detected as an increase in the size of this product. As shown in Figure 4.4, no significant insert could be detected in these products. It should be noted that in those mice with the IAP insertion, product could not be amplified due to the size of the IAP, so insertions other than the IAP in these mice would not be detected. As shown in Table 4.1, however, the insertion of the IAP only occurred in those mice more-recently diverged, and so would not be expected to retain sequence of the progenitor virus not found in other members of the *Mus musculus* complex. In the more distantly-diverged mice, no insertion 3′ of the *Fv1* locus could be detected. Smaller insertions 3′ of the gene could however be detected by sequencing, and sequence similarity searches have identified these as repetitive elements (see Figure 4.5).

4.6 Analysis of synonymous and non-synonymous mutation rates

The similarity index in Table 4.3 represents the initial pairwise matrix used in the first step of the multiple alignment analysis. The percentage similarity between individual pairs of sequences has been calculated. As shown in the Table, there is very little, and in some cases no, sequence divergence between members of the

Figure 4.5 Gross changes present in the *Fv1* region of some of the mice

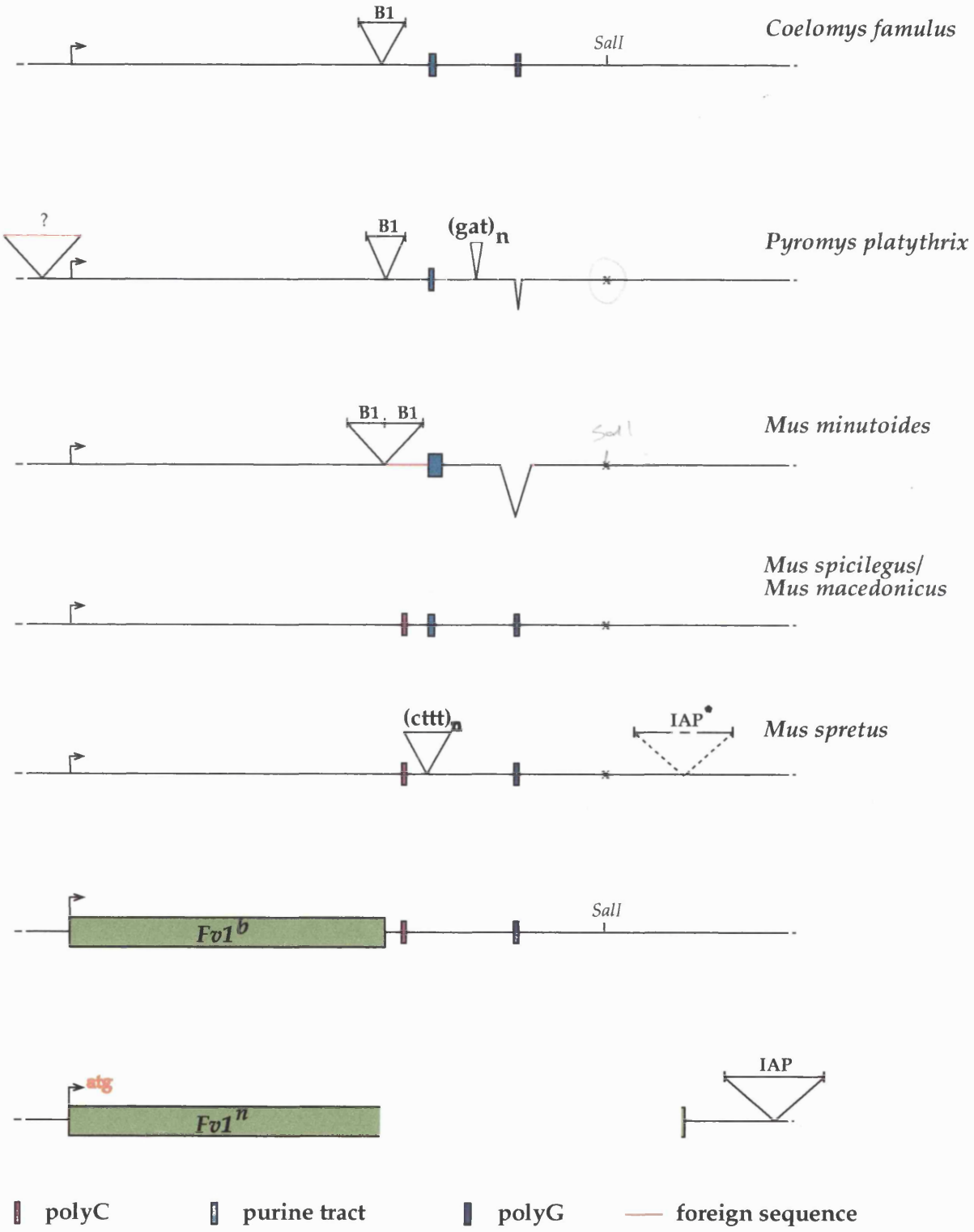


Table 4.3 Sequence pair distances of *Fv1* ORF from mice used in the sequencing project

FAM	PTX	MIN	DUN	COK	M.CAR	KAR	ZYP	XBS	SFM	M.SPRET	MOL	MOLD/RK	CIM	BIR	WMP/PAS	SEC/1REJ	RIII/DOMOB	I/LNJ	PRO/1REJ	FVB/NJ	C57BL	SWR/J	ST/BJ	BDP/J	NZB/BINJ	NZW/LACJ	RF/J	LG/J	M.MUS	CZECKI	MBT	SKIVE	RBA	SK/CAMEI	SF/CAMEI	P.ATTECK	BZO	WSB	
-	68.5	79.2	96.7	88.2	95.8	82.2	81.5	81.5	85.4	80.3	90.8	90.8	87.5	88.3	91.2	91.4	91.4	91.1	91	91.2	82.7	94.4	94.4	94.4	94.3	94.3	94.3	94.4	94.3	94.2	94.4	94.3	94.3	94.4	94.4	94.3	94.3	94.5	FAM
	-	70.2	89.7	76.6	82.7	69.7	73.5	73.6	72.5	76.4	73.6	73.6	76.3	74.5	76.5	73.3	76.6	77.4	76.7	76.5	75.6	90.9	90.9	90.9	90.8	90.8	90.8	90.7	90.8	90.6	90.9	90.8	90.9	90.9	90.9	90.9	90.8	90.9	PTX
		-	88.7	73.5	81.9	69	81.3	81.3	79.2	79.5	78.7	78.7	67.2	78.8	78.8	78.8	78.7	78.5	81.2	78.7	73.9	90.1	90.1	90.1	90.1	90.1	90.1	90.2	90.1	90	90.1	90.1	90	90.1	90.1	90	90	90.2	MIN
			-	96.4	97	91.9	97.1	96.9	96.6	95.4	95.1	95.1	95.7	95.2	95	95.1	95.1	95.1	95.1	95.1	95.1	93.6	93.6	93.6	93.5	93.5	93.5	93.6	93.5	93.4	93.6	93.5	93.5	93.6	93.6	93.5	93.5	93.7	DUN
				-	93.4	89.1	91.3	91.5	84.3	87.9	85.2	85.2	99.3	85.3	90.7	85.4	92.6	92.6	92.6	87.7	85	93.4	93.4	93.4	93.3	93.3	93.3	93.4	93.3	93.2	93.4	93.3	93.3	93.4	93.4	93.3	93.2	93.5	COK
					-	90.5	94.7	92.4	85.1	89	90	90	92.8	90.2	89.3	90.1	90.1	90	89.9	89.3	91.2	94.1	94.1	94.1	94	94	94	94.1	94	93.8	94.1	94	94	94.1	94.1	94	93.9	94.1	M.CAR
						-	87.8	87.8	80.7	82.7	85.6	85.6	90.2	85.6	85.1	85.4	85.2	85.3	85.1	85.1	83.9	91.4	91.4	91.4	91.4	91.4	91.4	91.3	91.4	91.2	91.4	91.4	91.3	91.4	91.4	91.3	91.2	91.3	KAR
							-	99.4	88.4	93.3	96.7	96.7	90.6	97.2	96.3	96.5	96.4	96.6	96.2	96.3	93.6	96.1	96.1	96.1	96	96	96	96.1	96	95.8	96.1	96	96	96.1	96.1	96	95.9	96.1	ZYP
								-	88.9	96.4	97	97	90.8	97.4	96	96.9	96.8	96.8	96.8	96	94.1	98.5	98.5	98.5	98.4	98.4	98.4	98.5	98.4	98.3	98.5	98.4	98.4	98.5	98.5	98.4	98.4	98.6	XBS
									-	94.1	86.6	86.6	81	86.7	86.3	86.5	86.5	86.6	86.6	86.4	86.6	96.4	96.4	96.4	96.4	96.4	96.4	96.4	96.4	96.2	96.4	96.4	96.4	96.4	96.4	96.4	96.3	96.5	SFM
										-	96	96	84.6	96.2	95.9	96	96	96	95.9	95.9	96	98.2	98.2	98.2	98.1	98.1	98.1	98.2	98.1	98	98.2	98.1	98.1	98.2	98.2	98.1	98.1	98.3	M.SPRET
											-	100	80.2	99.7	99.3	99.4	99.4	99.7	99.5	99.3	99.2	97.4	97.4	97.4	97.6	97.6	97.6	97.5	97.6	97.5	97.4	97.6	97.3	97.4	97.4	97.3	97.2	97.3	MOL
												-	80.2	99.7	99.3	99.4	99.4	99.7	99.5	99.3	99.2	97.4	97.4	97.4	97.6	97.6	97.6	97.5	97.6	97.5	97.4	97.6	97.3	97.4	97.4	97.3	97.2	97.3	MOLD/RK
													-	84.6	85.2	84.7	91.9	85.1	91.9	82.1	84.3	92.7	92.7	92.7	92.6	92.6	92.6	92.7	92.6	92.5	92.7	92.6	92.6	92.7	92.7	92.6	92.6	92.8	CIM
														-	99.2	99.4	99.4	99.7	99.4	99.2	99.2	97.6	97.6	97.6	97.5	97.5	97.5	97.5	97.5	97.4	97.6	97.5	97.5	97.6	97.6	97.5	97.5	97.5	BIR
															-	99.4	99.7	99.7	99.7	99.6	99.1	97.5	97.5	97.5	97.4	97.4	97.4	97.4	97.4	97.3	97.5	97.4	97.5	97.5	97.5	97.5	97.4	97.4	WMP/PAS
																-	99.8	99.6	99.6	99.4	99.3	97.5	97.5	97.5	97.5	97.5	97.5	97.4	97.5	97.3	97.5	97.5	97.5	97.5	97.5	97.5	97.4	97.5	SEC/1REJ
																	-	99.8	99.7	99.5	99.2	97.5	97.5	97.5	97.5	97.5	97.5	97.4	97.5	97.3	97.5	97.5	97.5	97.5	97.5	97.5	97.4	97.5	RIII/DOMOB
																		-	99.7	99.5	99.2	97.6	97.6	97.6	97.5	97.5	97.5	97.5	97.5	97.4	97.6	97.5	97.5	97.6	97.6	97.5	97.5	97.3	I/LNJ
																			-	99.5	99.1	97.5	97.5	97.5	97.5	97.5	97.5	97.4	97.5	97.3	97.5	97.5	97.5	97.5	97.5	97.5	97.4	97.5	PRO/1REJ
																				-	99.1	97.5	97.5	97.5	97.5	97.5	97.5	97.4	97.5	97.3	97.5	97.5	97.5	97.5	97.5	97.5	97.5	97.5	FVB/NJ
																					-	99.4	99.4	99.4	99.3	99.3	99.3	99.3	99.3	99.2	99.4	99.3	99.3	99.4	99.4	99.3	99.3	99.3	C57BL
																						-	100	100	99.9	99.9	99.9	99.9	99.9	99.8	100	99.9	99.9	100	100	99.9	99.9	99.9	SWR/J
																							-	100	99.9	99.9	99.9	99.9	99.9	99.8	100	99.9	99.9	100	100	99.9	99.9	99.9	ST/BJ
																								-	99.9	99.9	99.9	99.9	99.9	99.8	100	99.9	99.9	100	100	99.9	99.9	99.9	BDP/J
																									-	100	100	99.9	100	99.9	99.9	100	99.9	99.9	99.9	99.9	99.8	99.9	NZB/BINJ
																										-	100	99.9	100	99.9	99.9	100	99.9	99.9	99.9	99.9	99.8	99.9	NZW/LACJ
																											-	99.9	100	99.9	99.9	100	99.9	99.9	99.9	99.9	99.8	99.9	RF/J
																												-	99.9	99.8	99.9	99.9	99.8	99.9	99.9	99.8	99.7	99.9	LG/J
																													-	99.9	99.9	100	99.9	99.9	99.9	99.9	99.8	99.9	M.MUS
																														-	99.8	99.9	99.7	99.8	99.8	99.7	99.6	99.7	CZECKI
																															-	99.9	99.9	100	100	99.9	99.9	99.9	MBT
																																-	99.9	99.9	99.9	99.99	99.8	99.9	SKIVE
																																	-	99.9	99.9	99.99	99.8	99.9	RBA
																																		-	100	99.9	99.9	99.9	SK/CAMEI
																																			-	99.9	99.9	99.9	SF/CAMEI
																																				-	99.9	99.9	P.ATTECK
																																					-	99.8	BZO
																																						-	WSB

The sequence similarity index above was obtained by pairwise analysis of the *Fv1* ORF from the starting methionine to the deletion point present in the *Fv1*" allele.

Mus musculus complex, with more sequence difference observed between the more distantly diverged members of the *Mus* genus. Overall, nucleotide sequence identity among the mice used in this study varied between 67.2%-100%. In order to determine whether the sequence of *Fv1* in these species of mice has been under any kind of selection prior to the appearance of *Fv1* phenotype, the number of synonymous substitutions per site (Ks) and the number of non-synonymous substitutions per site (Ka) were measured for the more distantly-diverged species. The Ks/Ka ratios are shown in Table 4.4.

Table 4.4 Ratio of synonymous and non-synonymous mutation rates among the species of *Mus*

PTX	DUN	COK	CAR	SPRET	XBS	ZYP	C57BL	
0.971	1.473	2.367	2.336	1.546	1.328	1.388	1.504	FAM
	1.005	1.134	1.189	0.749	0.865	0.841	0.948	PTX
		1.885	1.664	1.073	1.027	1.073	1.203	DUN
			2.476	1.357	1.463	1.536	1.428	COK
				1.584	1.584	1.666	1.523	CAR
					1.473	1.617	1.511	SPRET
						-0.000	0.401	XBS
							0.427	ZYP

As a Ks/Ka ratio of 1:3 (0.333) is the value expected for neutrally drifting pseudogenes[308], a ratio greater than this is indicative of negative or purifying selection on a gene, as there has been more codon mutations that preserve the amino acid residue (synonymous mutations) than mutations that change the identity (non-synonymous mutations) in comparison to sequences under no selection. Negative selection results in a general bias for synonymous mutations. This bias will vary across the open reading frame, as different parts of the protein product will be tolerant to change to varying degrees e.g. active sites will be relatively intolerant to change, as it is more likely changes in this region will affect function (and so be selected against), whereas changes in regions further away will be less likely to have such an effect. However, by calculating the Ks/Ka ratio across the *Fv1* ORF as a whole, a general trend for maintenance of the sequence can be determined. Comparing the Ks/Ka ratio between *Fv1*

sequence of the recently-diverged C57BL and the older mice reveals that there appears to have been negative selection on the *Fv1* ORF, at least as far back as the divergence of *Coelomys famulus*, as all the ratios are above 0.33 (see Table 4.4). For FAM (*Coelomys famulus*), COK (*Mus cookii*), CAR (*Mus caroli*) and SPRET (*Mus spretus*) there appears to have been strongest selection for the ORF, with ratios between 1.428-1.523. The ratios of PTX (*Pyromys platythrix*) and DUN (*Mus dunni*) are lower, at 0.948 and 1.203 respectively. Finally, ZYP (*Mus spicilegus*) and XBS (*Mus macedonicus*) were found to have the lowest ratios (barely higher than for neutrally drifting sequences) of the group, with ratios of 0.427 and 0.401 respectively. However, as structural genes undergoing negative selection normally have a Ks/Ka ratio in the order of 5, these ratios are very low, and as such provide no conclusive evidence for negative selection.

By determining the structural and sequence features of the *Fv1* ORF of mice from across the *Mus* genus, it has been possible to determine how this gene has changed since the divergence of *Coelomys famulus* some 5MY ago. With respect to the changes specific to the 2 main alleles (the 3′ deletion and 2 amino acid changes within the ORF), it appears that the gene was originally undeleted (i.e. $Fv1^b$-like) but had the 2 amino acid residues found in the $Fv1^n$ allele. As previously suspected, the downstream insertion of the IAP element in most $Fv1^n$ alleles is a relatively recent event. Whilst there is little evidence of conservation of sequence identity among these mice (based on the synonymous:non-synonymous substitution rate), the low number of nonsense mutations (i.e. mutations that give rise to stop codons) in these mice against such a huge background of mutation does indicate selection for the maintenance of the *Fv1* ORF. No additional *Fv1*-related sequence was detected within the immediate vicinity of the locus among the more distantly-diverged mice.

Chapter 5 Isolation of *Fv1* progenitor virus

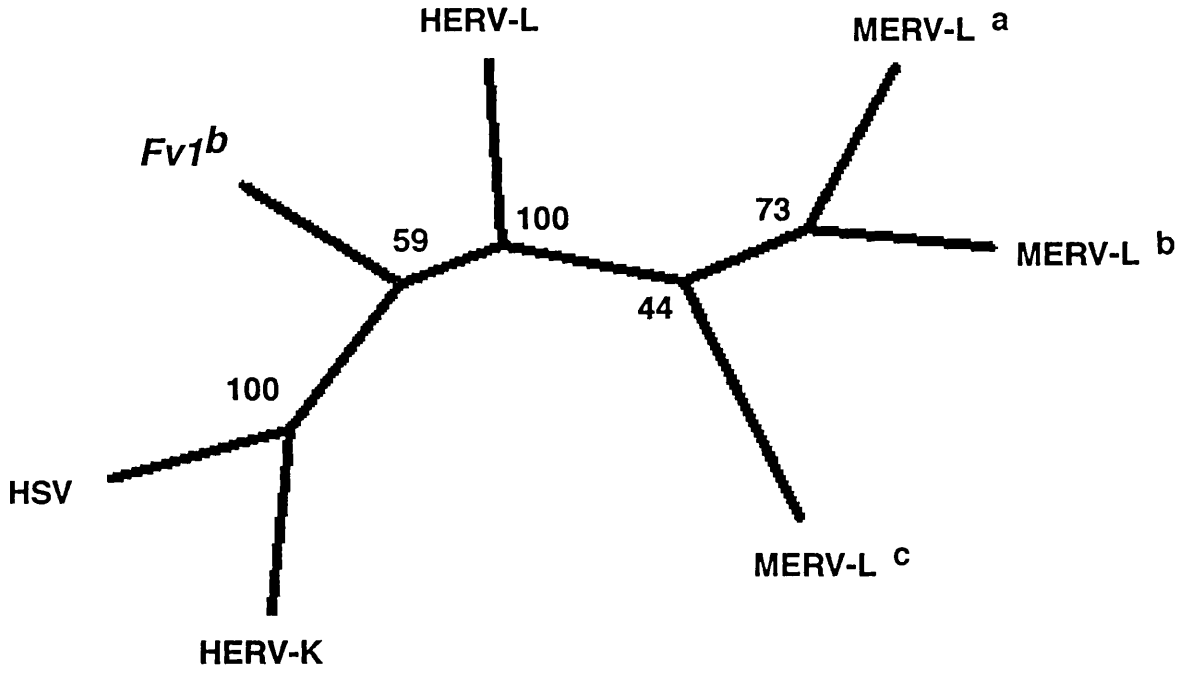
As discussed in Chapter 3, the *Fv1* gene of mice appears to have been the product of a retroviral integration event. After the gene was cloned the closest relative to *Fv1* was found to be a member of the human endogenous retrovirus family, HERV-L[94, 170]. Subsequent to the cloning of *Fv1*, the murine homologue of HERV-L, designated MERV-L, was cloned[95]. Sequence comparisons between *Fv1*, MERV-L and HERV-L (see Figure 5.1) have shown that, whilst *Fv1* appears to have a slightly higher homology to the MERV-L sequence than the HERV-L sequence, the gene itself does not appear to be directly descended from either. It seems therefore that the 'progenitor' retrovirus, the class of element whose integration created *Fv1*, was related to the MERV-L and HERV-L viruses, but was a member of neither class. Database searches have shown that sequence more homologous to *Fv1* has not been submitted, and therefore members of this progenitor virus have yet to be cloned.

The object of the following section was to search the mouse genome for *Fv1*-related sequences, with the aim of isolating a member of the retroviral class from which *Fv1* was derived. However, if the germline integration event that gave rise to the gene in the first place was a unique incident, or if other germline integrations have been subsequently lost, there may be no progenitor sequence remaining in the genome. To answer this question, a mouse genomic library in lambda phage was obtained and screened by hybridisation with the *Fv1* ORF. Clones containing sequence of highest homology were isolated. As the average insert size of these clones fell between 9-20kb, which precluded sequencing these clones directly, the sequence with *Fv1* homology was subcloned as a smaller fragment (either in pGEM®-3Zf(+) or pPCR-script^{AMP}). These clones were then sequenced using primers specific to vector sites flanking the multiple cloning site (MCS). This initial sequence was then used to search various nucleic acid databases (see section 2.17.6).

5.1 The mouse 129SvJ genomic lambda

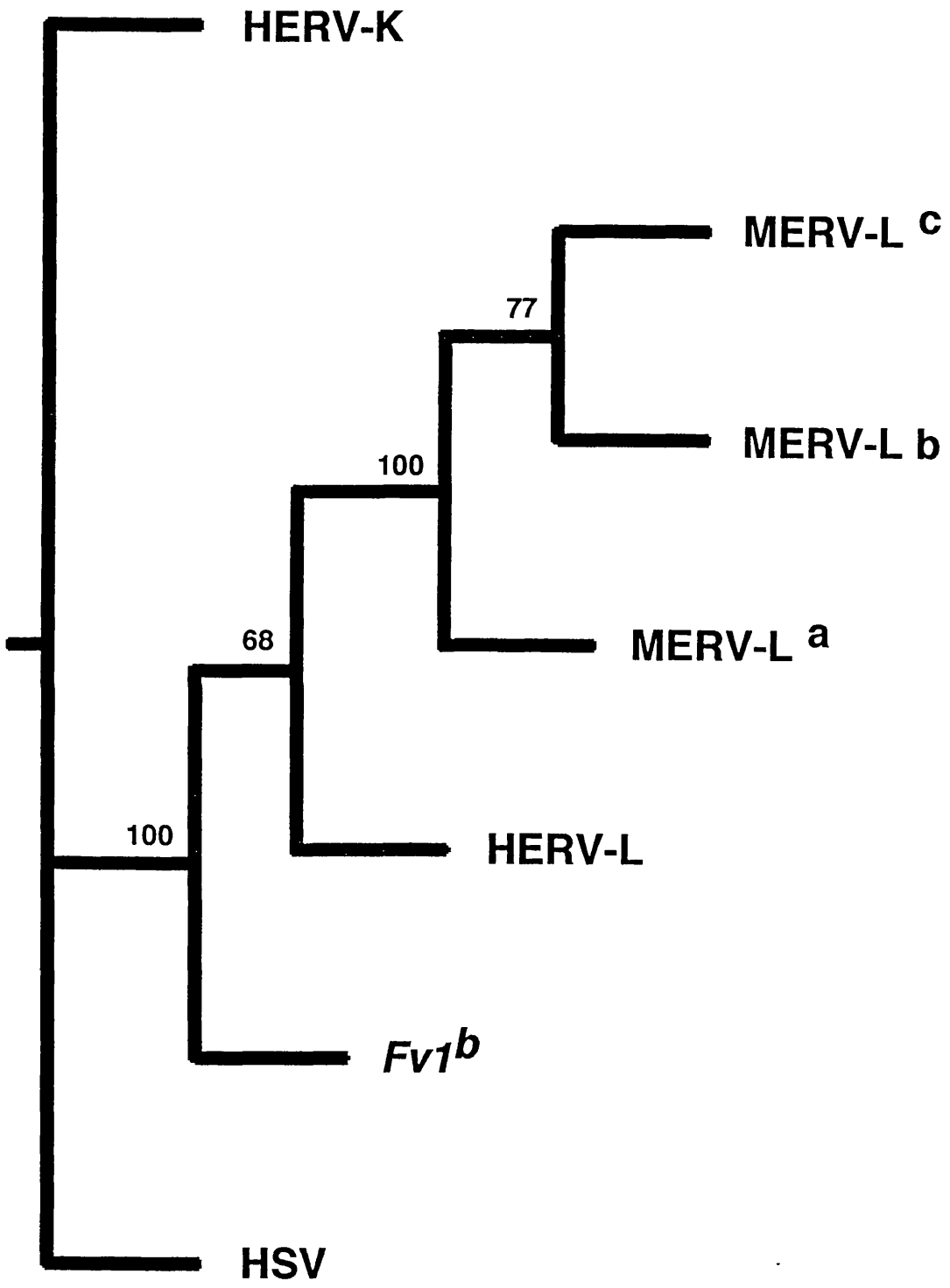
The lambda library was plated as per the manufacturer's protocol (see section 2.3.4). By plating a serial dilution of the library stock, the titre of the library was found to be $1.2 \times 10^{10} ml^{-1}$. Each 22 x 15cm plate of the library contained 90µl of a 10^{-3} dilution of the library stock. This library therefore consisted of around 1

Figure 5.1a Unrooted neighbour-joining bootstrap tree based on *Fv1*, MERV-L, HERV-L, HERV-K and HSV sequences



The sequences of *Fv1*[b] (GI:1515299), HERV-L (GI:895836), 3 MERV-L isolates (GI: 4007580[a]; 4050090[b]; 2065208[c]), HERV-K (GI: 955800) and HSV (human spumaretrovirus; GI:9629258) were aligned using MAGI (ClustalW). The tree was constructed using PIE. The figures on each branch represent bootstrap support (from 100 replicates).

Figure 5.1b Rooted neighbour-joining tree based on *Fv1*, MERV-L, HERV-L, HERV-K and HSV sequences (rooted to HERV-K)



The sequences of *Fv1*[b] (GI:1515299), HERV-L (GI:895836), 3 MERV-L isolates (GI: 4007580[a]; 4050090[b]; 2065208[c]), HERV-K (GI: 955800) and HSV (human spumaretrovirus; GI:9629258) were aligned using MAGI (ClustalW). The tree was constructed using PIE. The figures on each branch represent bootstrap support (from 100 replicates).

million clones. To facilitate the screen, phage lifts were made of each plate which consequently contained around 50000 clones per membrane. The library was then screened for sequences with some degree of shared homology to *Fv1* by hybridisation.

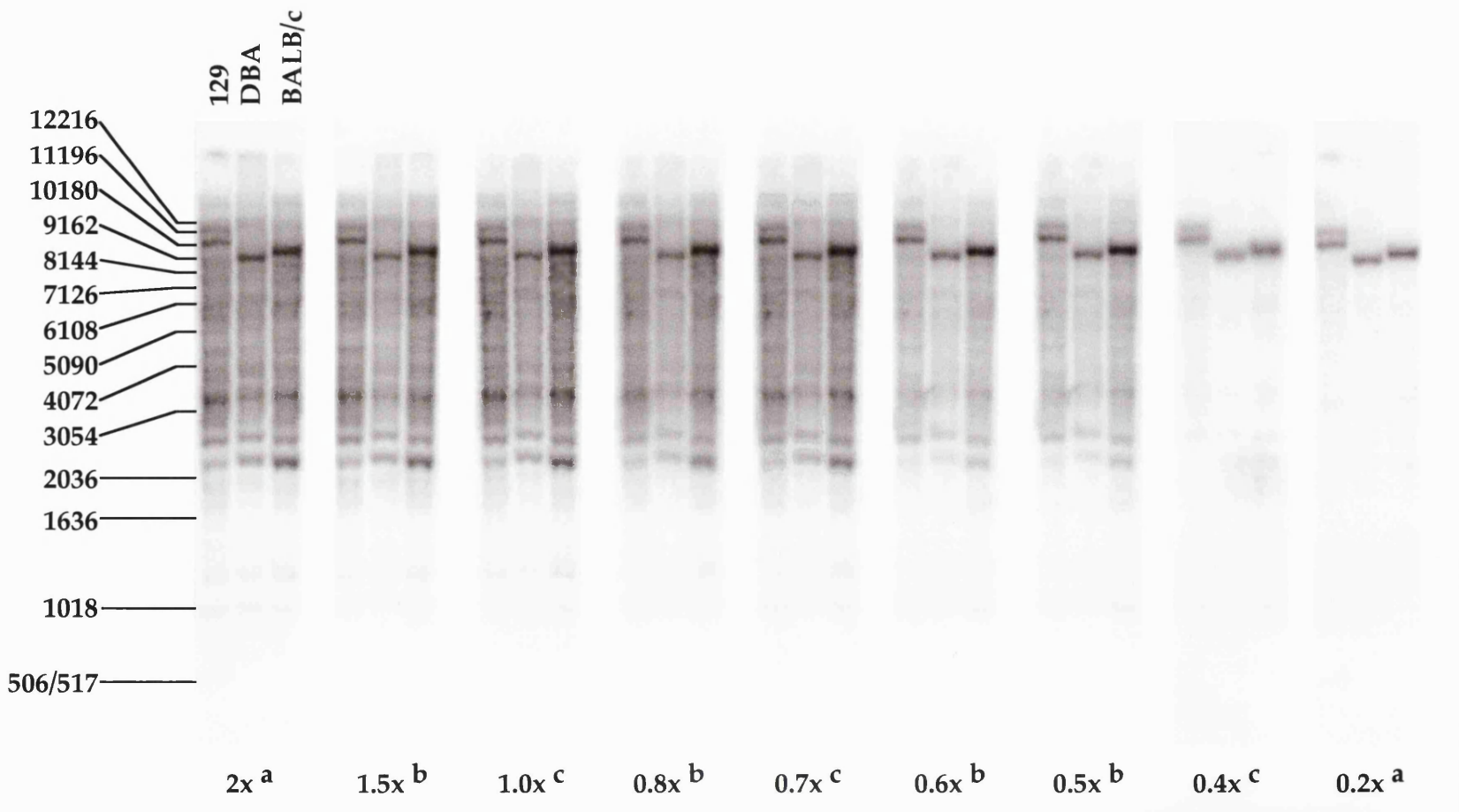
5.2 Titration of wash stringency

As the *Fv1* ORF was known to have sequence similarity to the MERV-L family, and the mouse genome is known to contain a great many of these elements, the stringency at which the blots were washed was very important; the blots had to be washed to a stringency high enough to minimise the cross-reaction to this retroviral 'background' (to allow any non-MERV sequence with *Fv1* homology not to be masked entirely), whilst allowing identification of sequences with some degree of divergence from *Fv1*. As a measure of appropriate conditions, southern blots of mouse genomic DNA were hybridised with a probe to the *Fv1* ORF and washed at progressively higher stringencies (see Figure 5.2).

30μg of mouse genomic DNA from 129, DBA and BALB/c mice was digested overnight with *EcoRI* and separated by electrophoresis on an agarose gel. 10μg of each sample were loaded onto the gel in 3 well-spaced sets. The DNA was then blotted onto a single nylon membrane which was then cut to produce 3 identical genomic southern blots, each containing digested DNA from the 3 different mice. These blots were prehybridised for 30 minutes in 10ml CHURCH buffer containing 50μl heat-denatured salmon sperm DNA and hybridised overnight with a probe generated by random-prime labelling using the $Fv1^b$ ORF as the template. This template DNA was generated by PCR-amplification of a plasmid containing the $Fv1^b$ gene (pCI-borf) using the primers GT16 and GT17 (see Figure 4.1). The blots were then washed at 65°C for 20 minutes in wash buffer containing 0.1% SDS and varying strengths of SSC. Care was taken to prevent the blots from drying out during the various washes. The blots were examined by overnight exposure to a phosphoimager screen after each wash.

The smaller bands seen at the lower wash stringencies were thought to be caused by cross-hybridisation between the probe and the MERV-L family of retroviruses. As shown in the Figure, the *Fv1*-specific probe remains hybridised to these bands at 2x SSC. As the stringency of wash buffer is increased, these bands become less intense, most notably when the salt concentration is reduced below 1x SSC. Between 0.8-0.5x SSC, these bands can still be seen, but below this

Figure 5.2 Titration of wash stringency for mouse genomic southern blots hybridised with *Fv1*



The identical blots a, b and c were hybridised overnight with a probe specific to the *Fv1*^{b} ORF and washed at progressively higher washing stringencies. The blot used at each stringency is suffixed.

concentration all signal is lost. The stringency chosen to wash the lambda library filters was 0.5x SSC to reduce the hybridisation background. At this stringency it was hoped only those sequences with the highest homology to *Fv1* would be isolated.

The bands toward the top of the gel are the *Fv1* gene itself. The probe remains hybridised even after the most stringent wash due to the high degree of homology between the gene and the probe. The band in the BALB/c lane corresponds to a ~9kb *EcoRI* fragment containing the $Fv1^b$ allele, and the band in the DBA corresponds to a ~8.5kb fragment containing the $Fv1^d$ allele (this is smaller due to the presence of the $Fv1^n$-associated deletion within the gene). The 2 bands in the 129 lane are of particular interest. Firstly, as with BALB/c and DBA, one would expect the probe to hybridise to a single fragment only. This presence of 2 bands is extremely unexpected as all 3 mice strains from which the DNA was prepared are inbred, and as such should contain only one allele at the *Fv1* locus, which would in turn correspond to only one strongly-hybridising band on a genomic southern blot. Secondly, as the $Fv1^{nr}$ allele in 129 is deleted, one would expect the $Fv1^{nr}$ allele in this strain to lie on a smaller *EcoRI* fragment than the undeleted $Fv1^b$ allele in BALB/c; as shown in Figure 5.2, the fragments in the 129 lane are ~9.2 and 11.5kb, significantly bigger than the 9kb fragment in BALB/c. It was hoped that lambda clones containing both of these fragments would be isolated from this screen, and that the subsequent analysis of the sequence would shed light on this interesting finding.

5.3 Screening the lambda library by hybridisation

The library screen was carried out as described in section 2.13.3. To ensure discrete clones were obtained, phage from plaques isolated from the first round of screening were used to re-infect XL1 Blue MRA *E. coli* and these were replated at low density (see Figure 5.3). In this way, lambda clones that cross-hybridised with the *Fv1* ORF probe were isolated on the second round of screening.

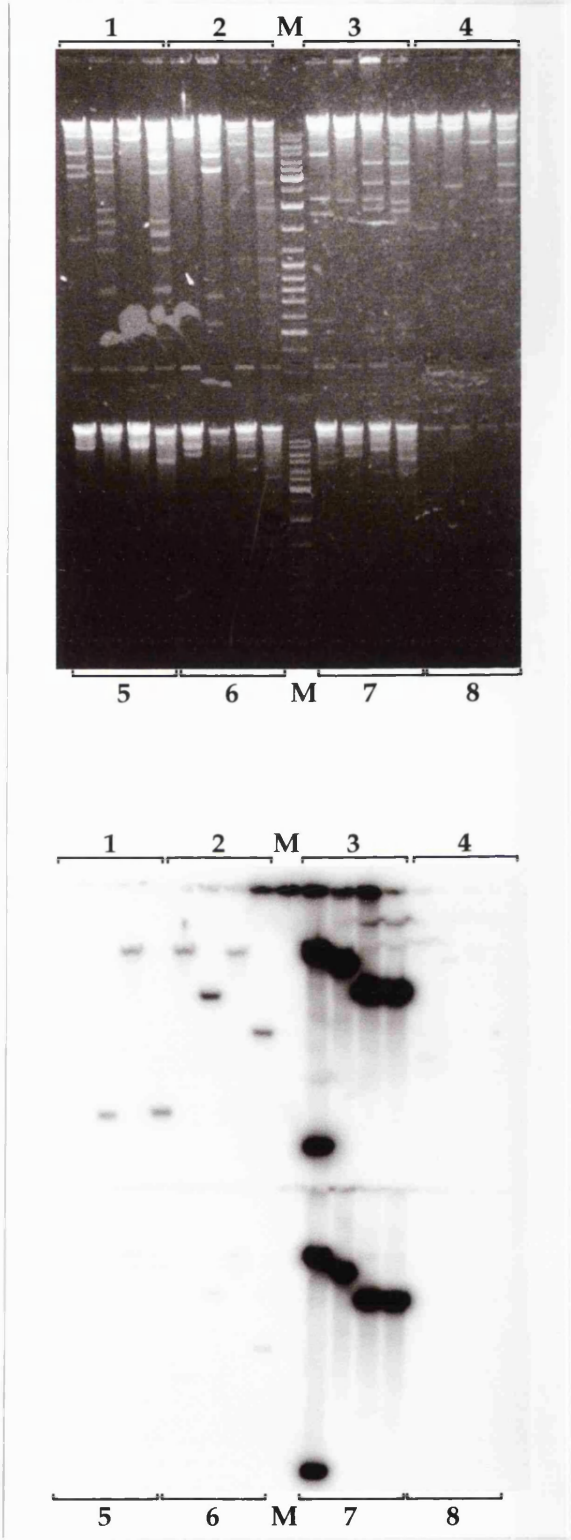
As the average insert size of the lambda clones ranged between 9-23kb, which precludes sequencing the entire insert, the sequences to which *Fv1* hybridised had to be subcloned as smaller fragments before they could be sequenced. Positive clones were digested with *EcoRI, BamHI, HinDIII, SacI,* and double digests of all 4 restriction enzymes. These digests were then separated on agarose gels, blotted on Hybond N^+ (Amersham Pharmacia) and hybridised with

a probe complementary to the Fv1 ORF (see Figure 5.4). The smaller fragments which contained the region of Fv1 homology could then be identified and subsequently subcloned into either pCR-script and pGEM(e)-T(+) vectors, which allows blue/white selection of insert-containing clones (see Figure ...)

Figure 5.3 Second round screen of clones isolated from 129/SvJ mouse genomic lambda library screen using *Fv1*-specific probe

The phage blots were hybridised overnight with a probe specific to the $Fv1^b$ ORF and washed with 0.5x SSC/0.1% SDS buffer.

Figure 5.4 Restriction digest and Southern blot analysis of various lambda clones



The lanes are loaded as *BamHI, EcoRI, HinDIII* and *EcoRI/HinDIII* digests corresponding to the following lambda clones:

1. clone 2	2. clone 3	3. 2DK1	4. 6L1D
5. 6L2	6. 6L1	7 6M1	8. 6DK1

The marker lanes M contains 200ng 1kb ladder

The blot of the gel was hybridised with a probe to *Fv1* overnight, washed to 0.1x SSC and exposed to a phosphoimager screen for 2 hours.

5.4 Subcloning of fragments that cross-hybridise with *Fv1*

Once subcloned these clones were then restriction digested and blotted, and rehybridised with the *Fv1*-specific probe to ensure the isolation of the correct fragment. A range of hybridising band intensities was seen among these clones. This observation allowed the clones to be grouped according to the intensity of their hybridised insert band, from very weak to very strong (see Table 5.1). Clones were chosen from all groups to be sequenced.

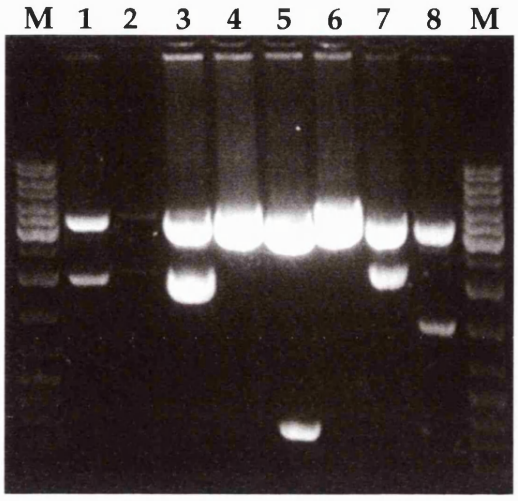
5.5 Sequence analysis

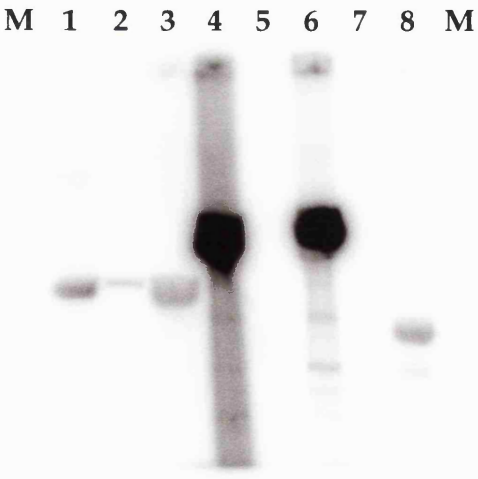
The primers used to sequence the inserts of the clones isolated were M13 (-40) forward and M13 (rev) reverse. Primer binding sites to these flank the MCS in both pGEM®-3Zf(+) and pPCR-script^{AMP}. As the insert sizes of the lambda subclones tended to be over 2kb, most of the sequence of the inserts of these clones was incomplete. However, enough sequence was produced in almost all cases to identify the sequence responsible for cross-hybridisation with the *Fv1* probe. All sequences were used to perform BLASTN searches of Genbank, EMBL, DDBJ and PDB. BLAST (Basic Local Alignment Search Tool) is a set of similarity search programs designed to explore all of the available sequence databases. BLAST uses a heuristic algorithm which seeks local as opposed to global alignments and is therefore able to detect relationships among sequences which share only isolated regions of similarity[309]. The results of these searches are summarised in Table 5.1. Those fragments found to be *Fv1* or members of the MERV-L family are shown in Figure 5.6, relative to consensus sequences of *Fv1* (X97719 and X97720) or MERV-L (Y12713).

5.5.1 Sequence similarity searches

The most strongly hybridising clones, 2DK1 and 6L10, appeared to contain the *Fv1* gene itself. Both sequences contained the *Fv1^{n}*-specific residues in the ORF (at positions 1075 and 1198) and were deleted. They also contained the *Fv1^{nr}* change at position 1058. The gene isolated in both cases was therefore therefore the *Fv1^{nr}* allele of *Fv1*. A third strongly-hybridising clone, 6M1, gave the same size restriction fragments as 2DK1 with *BamHI, EcoRI, HinDIII* and *HinDIII/EcoRI*, and was assumed to be identical to this clone. Those clones isolated that hybridised less intensely were exclusively members of the MERV-L family, while

Figure 5.5 Southern blot analysis of various lambda subclones





The lanes are marked as follows:

1. 6L1 *EcoRI/HinDIII*	2. 6DK1 *EcoRI/HinDIII*	3. clone 1 *EcoRI/HinDIII*
4. 6L10EH *EcoRI/HinDIII*	5. clone4 *EcoRI/HinDIII*	6. 2DK1 *HinDIII*
7. clone7 *BamHI*	8. 6L11B *BamHI*	

The marker lanes M contains 200ng 1kb ladder

The blot of the gel was hybridised with a probe to *Fv1* overnight, washed to 0.1x SSC and exposed to a phosphoimager screen for 30 minutes.

Table 5.1 Results of the mouse genomic library screen

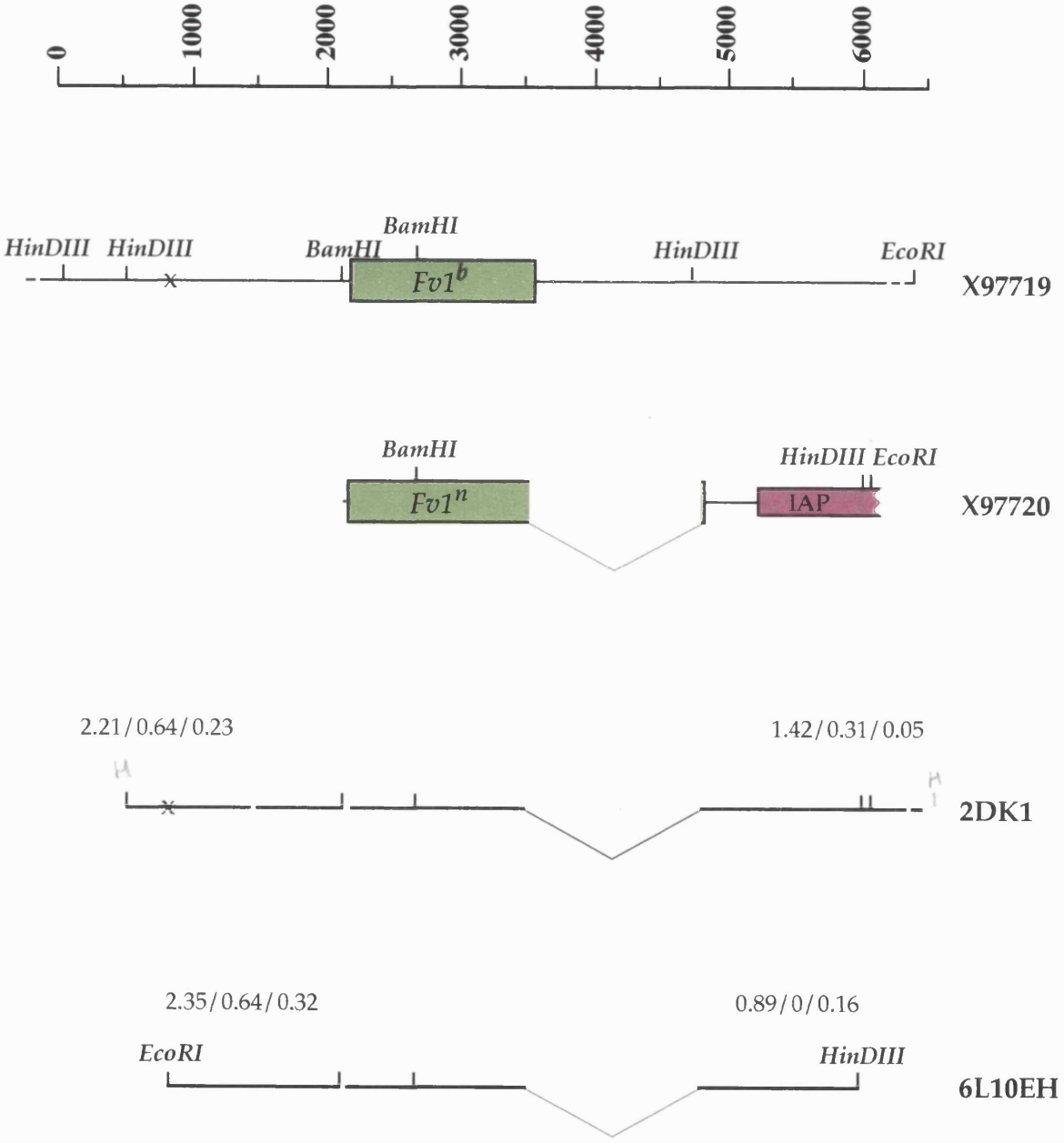
Lambda clone	Intensity of hybridisation signal[a]	Size of smallest hybridisable fragment (base pairs)	Cloning site	Subcloned?	Significant BLAST matches
2DK1	very strong	4000-4500	*HinDIII*	yes	*Fv1*^{n}
6M1	very strong	4000-4500	*HinDIII*	no[b]	
6L10	strong	3500-4000	*EcoRI/HinDIII*	yes	*Fv1*^{n}
5b	strong	2000-3000	*EcoRI/HinDIII*	yes	MERV-L
5c	strong	900	*HinDIII*	yes	MERV-L
10	strong	800	*HinDIII*	no	
6L11	moderate	1500	*BamHI*	yes	MERV-L
1	moderate	2100	*EcoRI/HinDIII*	yes	MERV-L (3′) and no match (5′)
2	moderate	900	*EcoRI*	no	
3	moderate	2000	*BamHI*	yes	MERV-L
14	moderate	1500-2000	*EcoRI/HinDIII*	yes	MERV-L
16	moderate	~1400	*BamHI*	no	
17	moderate	900	*BamHI*	no	

[a] as determined by southern blot hybridisation of lambda DNA digests with a probe to the *Fv1*^{b} ORF
[b] identical restriction fragment pattern as 2DK1, so assumed to be same sequence; not subcloned

20	moderate	~1500	*BamHI*	no	
6DK1	weak	~2000	*EcoRI/HinDIII*	no	
6L13	weak	500	*BamHI*	no	
5a	weak	2000-3000	*EcoRI/HinDIII*	yes	*alu*; no match
6	weak	n.d.[c]	/	no	
8	weak	n.d.[c]	/	no	
9	weak	1500-2000	*EcoRI/HinDIII*	yes	MERV-L
11	weak	1600-2000	*EcoRI/HinDIII*	no	
18	weak	~2100	*BamHI*	no	
19	weak	~5000	*BamHI*	no	
22	weak	~700	*EcoRI*	yes	No match
23	weak	~1300	*BamHI*	yes	GTPase?
6L1	very weak	2000	*EcoRI/HinDIII*	yes	MERV-L
6L2	very weak	n.d.[b]	/	no	
4	very weak	~600	*EcoRI*	yes	*sine*/B2; no match
12	very weak	1200	*EcoRI*	no	

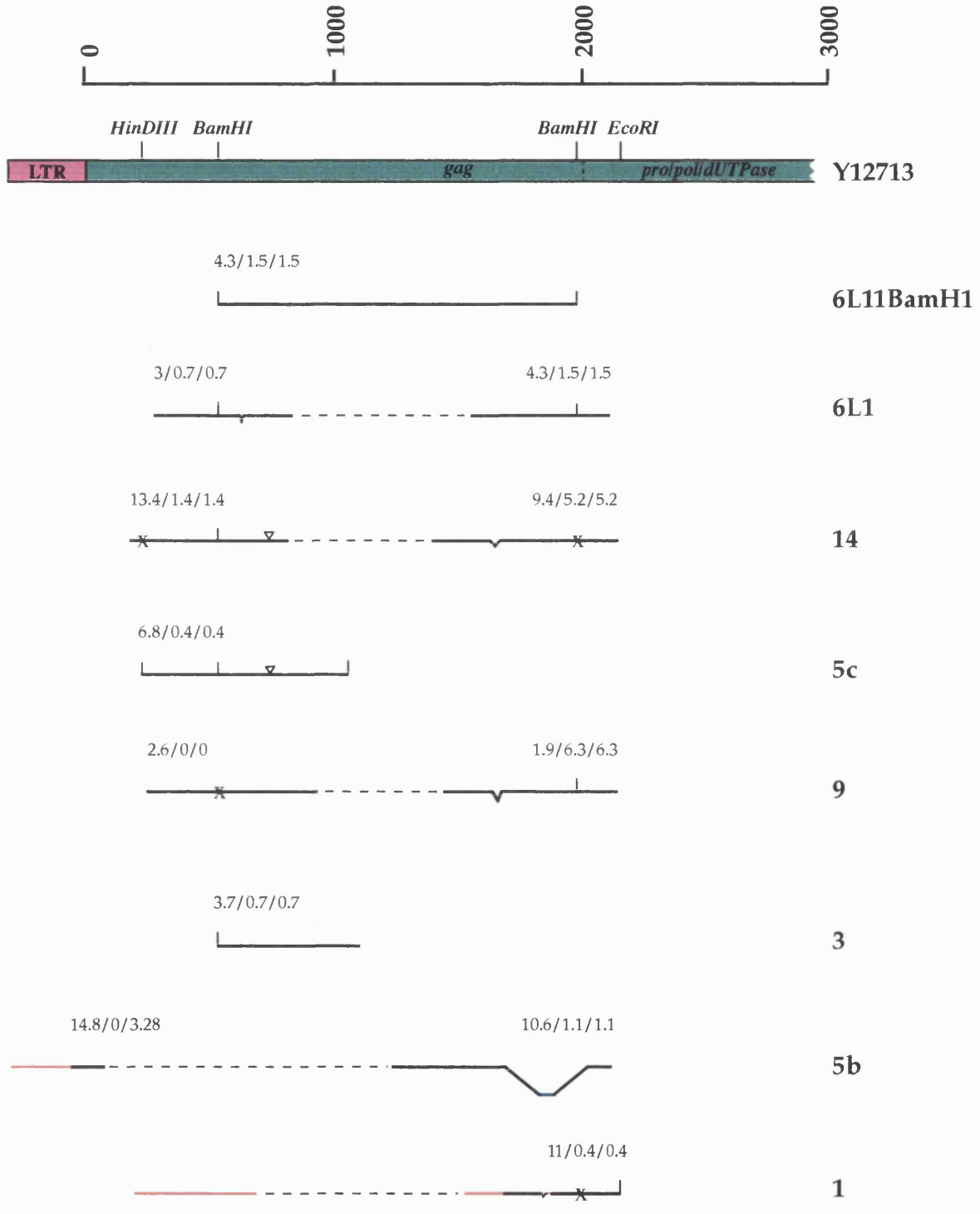
[c] not determined

Figure 5.6a Sequence comparison of the lambda subclones to the sequences of *Fv1*



The Figure shows the sequence of the 2 lambda subclones 2DK1 and 6L10EH aligned with the *Fv1*^{n} and *Fv1*^{b} submitted sequences (X97719 and X97720). The absence of the *EcoRI* restriction site in X97719 and 2DK1 is marked by 'x'. The figures above the sequence of the subclones in both Figures corresponds to the sequence divergence, deletions and insertions repectively as a percentage, compared to the submitted sequences.

Figure 5.6b Sequence comparison of the lambda subclones to the MERV-L sequence



the very weakest tended to have no known sequence similarity.
Of the 8 clones shown to have high homology to the MERV-L family, none appeared to be from the same element. These clones therefore seem to represent 8 distinct elements of the MERV-L family.

5.5.2 Sequence features

All subclones of lambda clones isolated from this screen were cloned as *BamHI*, *HinDIII*, or *EcoRI/HinDIII* fragments (see Figure 5.6) and sequenced (see Appendix 6). The subclone derived from 2DK1 was cloned as an *HinDIII* fragment; the *HinDIII* site at the 5′ end of *Fv1* could be readily identified from the sequence obtained, but the 3′ *HinDIII* site could not. This is not unusual due to the close proximity of the cloning sites to the sequencing primer sites in the vector. It should be noted that the *HinDIII* site (indicated on the far right of X97720 in Figure 5.6a) in the IAP sequence at the 3′ end is present internally in this clone. This is 176bp from the end of the readable sequence, and therefore is probably within 200bp from the 3′ terminal *HinDIII* site used to clone the fragment. This subclone must therefore be the product of a partial *HinDIII* digest of the original lambda clone; although all fragments were isolated from agarose gels prior to cloning, this degree of size selection would not be sufficient to resolve fragments of around 200bp difference. The *EcoRI* site ajacent to this site in X97720 is also present.
Surprisingly, subcloning and sequencing the hybridising fragment from clones 2DK1 and 6L10 revealed a true difference between the 2 $Fv1^{nr}$ sequences. The lambda clone 6L10 contains an *EcoRI* site 1336bp upstream of the starting methionine that was used to subclone this fragment into pGEM®-3Zf(+). Sequencing of the subclone derived from 2DK1, which encompasses this region, revealed that this *EcoRI* site is absent in this clone, due to a change in the first base pair of the recognition site (AAATTC instead of GAATTC). It therefore appears that this *EcoRI* site represents a polymorphism between these 2 clones. The absence of an *EcoRI* site is consistent with the upstream sequence of the $Fv1^{b}$ allele (X97719; this allele was isolated from mouse strain C57BL/6J). However, upstream sequence obtained from the mice used in the sequencing project (see Chapter 4) has revealed that this *EcoRI* site is present in all samples examined, namely *PTX (Pyromys platythrix)*, ZYP (*Mus spicilegus*), COK (*Mus cookii*), SFM (*Mus spretus*), MOL (*Mus musculus molissinus*), BZO (*Mus musculus domesticus*)

and MBT (*Mus musculus musculus*). It therefore seems likely that this site has been lost in clone 2DK1 (and X97719) rather than gained in clone 6L10EH as the sequence of X97719 indicates. The absence of this site in the sequence of X97719 may reflect the loss of this site either in the *Fv1*^{b} allele itself or may be limited to the *Fv1*^{b} allele of the C57BL strain only. Unfortunately, as this region was not sequenced from a *Fv1*^{b}-containing mouse strain, the exact case cannot be determined.

5.5.3 Sequence divergence

The divergence of each sequence in Figure 5.6a and 5.6b (i.e. those sequences identified as either *Fv1* or a member of the MERV-L family) from the submitted sequences (X97719 and X97720 for the *Fv1* sequences, and Y12713 for the MERV-L sequences) was calculated. Divergence was measured by assessing the amount of bases in common with the submitted sequence, the number of bases deleted between the 2 sequences and the number of bases inserted; these 3 figures, calculated as a percentage of the contiguous sequence, are shown for each sequence in the Figure.

The sequences of 2DK1 and 6L10EH that correspond to *Fv1* show very similar, low levels of divergence from the submitted sequences. Overall, both sequences diverge from the submitted sequences by less than 2%, with less than 0.5% of bases deleted and less than 0.3% of bases due to insertion. The *Fv1* ORF of 2DK1 contains no inserted or deleted bases, and is open to the termination codon used by the *Fv1*^{n} alleles. The *Fv1* ORF of 6L10EH, however, contains numerous stop codons due to the presence of 2 separate single base pair insertions, a thymidine and adenosine at 65 and 394 bp respectively from the start of the gene. However, the sequence was obtained from single-pass sequence reads only, and may therefore not represent the true consensus; the possibility that these 2 base pair insertions may both be due to sequencing errors, although unlikely, cannot be ruled out.

The sequences of the MERV-L clones showed a much greater degree and range of divergence from the published sequence. The degree of overall sequence divergence ranged from 2.2% (clone 9) to 11.2% (clones 5b and 14); deletions ranged from 0.4% (clone 5c) to 3.5% (clone 14); and insertions ranged from 0.4% (clones 1 and 5c) to 3.5% (clone 14). It seems therefore that clone 9 has the highest homology to the Y12713 submitted sequence, whilst clone 14 represents

the most divergent sequence from the submitted sequence. Although extremely similar to Y12713, clone 9 is distinct from this sequence, as shown by the presence of a small deletion in the 3' portion (see Figure 5.6b). Clone 14, on the other hand, not only has a lower degree of homology to Y12713 but also has a large deletion in the 3' portion of the sequence, and runs into 'foreign', unidentifiable sequence at the 5' end, presumably as a result of either an insertion into the MERV-L element or a deletion at the 5' terminus. With the exception of clone 3, the other sequences have deletions, insertions or loss of restriction sites which similarly mark them as being discrete elements. None of these sequences showed significantly higher homology to *Fv1* than the Y12713 MERV-L submitted sequence (data not shown).

The *Fv1*-specific probe hybridised weakly to the remaining sequences, none of which showed homology to either *Fv1* or MERV-L elements. Of these 4 sequences, 2 contained basic repetitive elements (clones 4 and 5a), and produced no other matches from the database. Clone 23 produced a good match with human GTPase genes (87% match over 181bp), while the fourth sequence, clone 22, produced no significant matches to anything in the database.

The screen of the 129/SvJ mouse genome for retroviral sequence of a higher degree of homology to *Fv1* than MERV-L identified only the gene itself and members of the MERV-L family. A southern blot of 129 genomic DNA cut with *EcoRI* identified, unexpectedly, the presence of 2 fragments in the genome that cross-hybridised with an *Fv1*-specific probe under the highest wash stringencies. These fragments were larger than expected for the undeleted form of the *Fv1* gene. Analysis of the sequences of *Fv1* isolated from the screen showed that they were almost identical to the submitted *Fv1*^{n} sequence, they contained the *Fv1*^{nr}-associated base change (marked on sequences in Appendix 6), and they had the Fv1^{n}-associated deletion. Neither sequence therefore could represent a progenitor sequence. However, both appear to be distinct from one another, as indicated by the finding that whilst 2DK1 had an intact ORF, 6L10EH appeared to be truncated due to the presence of 2 single base pair insertions. In support of this, they were also found to be polymorphic for an *EcoRI* site upstream of the gene. This site, which is present in 6L10EH (and so is linked to the allele with premature stop codons) is probably present in this region in *Fv1*^{nr}, as it was shown to be present in the sequence obtained from MBT (*Mus musculus*

musculus) and BZO (*Mus musculus domesticus*), both of which contain the *Fv1*^{n} allele, as well as the much more-distantly diverged mice. Whilst the sequence obtained 5′ of the gene did not extend far enough in this project to identify the next *EcoRI* site upstream of this one, and thus the difference the absence of this site would make to the size of the cross-reactive fragment could not be determined, it is tempting to speculate that this polymorphism could account for an additional cross-reacting band observed in the mouse genomic southern blot probed with the *Fv1* ORF (see Figure 5.2). However, this does not explain the why the 2 cross-reacting *EcoRI* fragments in 129 are larger than the undeleted *Fv1*^{b} fragment in BALB/c. As Figure 5.6a indicates, the presence of the *EcoRI* site in the 3′ IAP element in 2DK1 means that this allele (and possibly the 6L10EH allele also) must lack the 5′ *EcoRI* site (not shown in Figure 5.6a) responsible for the ~9kb *Fv1*^{b} fragment in BALB/c, in order for the allele represented by 2DK1 to be on a larger *EcoRI* fragment than the *Fv1*^{b} allele. This would mean that this allele would have to be polymorphic for the upstream *EcoRI* site. For the allele represented by 6L10EH to be on a larger *EcoRI* fragment than the *Fv1*^{b} allele, the presence of the 5′ *EcoRI* site polymorphic with 2DK1 (and the submitted X97719 sequence; shown in Figure 5.6b) would mean that this allele must lack the 3′ *EcoRI* site in the IAP. This would be the second polymorphic *EcoRI* site in this allele. To confirm the presence of this *EcoRI* polymorphism (and to rule it out as just an artefact in the lambda library), primers could be designed to the sequence flanking its position in 129 could be used to PCR-amplify across it. If the primers are equidistant from the site, cutting this product should yeild 2 bands (corresponding to the uncut fragment, and cut fragments half the size). Nevertheless, the presence of polymorphic restriction sites, and the differences in the *Fv1* ORF between these 2 clones, indicates that 129 appears to contain 2 different *Fv1* alleles. This is discussed further in Chapter 7.

Although far from exhaustive, this screen was biased towards identifying clones containing sequence with highest *Fv1* homology. The failure to detect sequence with more homology to *Fv1*, therefore, may reflect the absence of such sequence in the germline, at least in the 129/SvJ strain (and probably the inbred mice in general). This does not exclude the possibility of such sequence being present in the more distantly-related mice, such as *Apodemus*. Future work may yet identify the progenitor virus that gave rise to *Fv1*.

Chapter 6 Phenotypic studies of *Fv1*

As mentioned in the Introduction, there are 2 main alleles of *Fv1*, *Fv1*^{n} and *Fv1*^{b}. The *Fv1*^{n} allele restricts the replication of B-tropic virus, but is permissive for N-tropic MLV retroviral infection. Conversely, the *Fv1*^{b} allele restricts the replication of N-tropic virus, but is permissive for B-tropic MLV retroviral infection. The sequence of these 2 alleles shows only 3 major differences between them, and it is these that are thought to be involved in the *Fv1* phenotype. In Chapter 4 the *Fv1* gene of various mice from across the *Mus* genus was analysed to determine how this sequence has changed over time, during the speciation of the genus; the precise order and nature of mutation in the ORF, including the allele-specific mutations mentioned above, was determined, from the null alleles in the older mice to the emergence of activity in the most recently-diverged inbred mice. To assess exactly what the contribution of each of these changes is not only to *Fv1* activity we see in inbred mice today, but also what these changes may have contributed in the past, during the evolution of *Fv1* activity, mutant *Fv1* ORFs containing these 3 changes in various combinations were constructed in the pCI expression vector. These constructs were assembled using the corresponding portions of the *Fv1* gene from either the *Fv1*^{n} or *Fv1*^{b} allele from inbred mice. Although some of the mice used in the study were shown naturally to have some of these combinations of changes, these constructs were used instead to ensure differences in phenotype were due to these mutations alone, and not due to other mutations also present in these sequences. The phenotype produced by each of these mutant constructs was then examined by transfecting a null (*Fv1*$^-$) cell line with each construct and assaying for MLV restriction in a lacZ assay using pseudotyped MLV.

6.1 Transfection studies using *Fv1* ORF mutants

6.1.1 Creation of *Fv1* mutant constructs

The 3 main differences between the *Fv1*^{n} and *Fv1*^{b} alleles are shown in Figure 1.6. To ascertain exactly what the contribution of each of these changes to overall phenotype was, constructs were created in this lab to address this question (G. Towers, unpublished). The *Fv1*^{n} and *Fv1*^{b} alleles were subcloned into the expression vector pCI (Promega) and these were used as primary constructs.

Mutant constructs were then derived from these that contained the 3 changes in all 6 remaining combinations by ligating PCR products from the corresponding regions of these primary constructs (see Table 6.1). These constructs were then sequenced to confirm the nature of the mutations and also to ensure no unintended mutations had been introduced. Despite numerous attempts to sequence the entire insert, insufficient sequence data was obtained from construct MM3, and clones obtained using this construct were not assayed. The sequence data showed the presence of the correct mutations in the remaining constructs and also that only one accidental change had been introduced, found in the MM2 construct (at position 18); as this mutation caused a synonymous change at the amino acid level (GCG→GCC; both alanine), this construct was still used in this study.

Table 6.1 Mutants generated to determine the contribution of the main changes to *Fv1* phenotype

Mutant	Residue 1	Residue 2	Deletion
MM0	b	b	n
MM1	b	n	n
MM2	n	b	n
MM3	n	n	b
MM4	b	n	b
MM5	n	b	b

6.1.2 Production of cell lines expressing mutant *Fv1* ORFs

The constructs were used to transfect the *Fv1°Mus dunni* cell line. As the pCI vector contains the neomycin (*neo*) mammalian selectable marker, clonal mutants were obtained by culturing under antibiotic selection using G418. To ensure clonality, cells from each transfection experiment were seeded at limiting dilutions into 24-well tissue culture plates. Clones from wells positive for growth after long-term antibiotic selection were picked from plates with less than 6 of the 24 wells positive for growth. Once obtained, these cell lines were frozen down until required for assay. Each cell line was then assayed as described in sections 2.18.4 and 2.19.2. Table 6.2 shows the number of clonal cell lines isolated from this transfection for each construct.

Table 6.2 Number of clones isolated for each construct transfected

Mutant	MM0	MM1	MM2	MM3	MM4	MM5
No. clones isolated	20	9	8	/	21*	24

*only 7 cell lines tested to date

6.1.3 Assessment of *Fv1* phenotype

To ensure the fidelity of the assay, the pseudotyped MLV virus reagents were tested on N3T3 and B3T3 cell lines, which are homozygous for $Fv1^n$ and $Fv1^b$ respectively. The results of one such assay are shown in Table 6.3. Also shown in the Table are the results for one particular experiment to assay 7 clonal cell lines obtained from the transfection of *Mus dunni* with the MM0 construct. This construct yielded clones with a range of phenotype, from unrestricted (see clones 7 and 8) to an apparently modified $Fv1^n$ restriction pattern (see clone 9).

The results of the remaining clones that gave a demonstrable phenotype are summarised in Table 6.4. The majority of the neomycin-resistant clones isolated from the transfections of *Mus dunni* with the mutant constructs gave no detectable restrictive phenotype (data not shown).

None of the cell lines transfected with the constructs MM2 and MM4 tested to date had a measurable phenotype in this assay. From the transfection of *Mus dunni* with the MM0 construct, 5 cell lines out of 20 showed some degree of restriction to MLV infection in this assay. As shown in Table 6.3, different cell lines displayed a varying degree of restriction, presumably due to clonal variation of expression levels of the mutant gene in these cell lines. However, if this variation is due to levels of expression, you would expect the relationship in restriction between the 3 viruses to be the same in different clones; this is not clear in Table 6.3. The phenotype conferred by this constructs appears to be essentially that of the $Fv1^b$, as shown by the absolute restriction of N-tropic virus in cell line MM0.9. However, the additional restriction of B- and NB-tropic virus in most of the other clones, albeit to a lesser extent, indicates that this $Fv1^b$ pattern of restriction appears to be modified by the mutation in this construct. The effect of the MM0 combination, which is essentially an $Fv1^b$ allele with an $Fv1^n$ terminus, appears to be extend the $Fv1^b$ phenotype to moderately restrict B-tropic and to a lesser extent NB-tropic virus as well.

Table 6.3 Results of a biological assay for *Fv1* activity using pseudotyped MLV

Clone name	No. blue cells with N-tropic pseudotyped virus	No. blue cells with B-tropic pseudotyped virus	No. blue cells with NB-tropic pseudotyped virus	Level of restriction		
				N-tropic virus	B-tropic virus	NB-tropic virus
MM0.3	209	68	88	1.97-fold	1.94-fold	3.53-fold
MM0.4	163	33	266	2.53-fold	4-fold	1.17-fold
MM0.5	121	19	88	3.4-fold	6.95-fold	3.53-fold
MM0.6	32	20	362	12.88-fold	6.6-fold	0.86-fold
MM0.7	365	107	298	1.13-fold	1.23-fold	1.04-fold
MM0.8	339	148	326	1.22-fold	0.89-fold	0.95-fold
MM0.9	0	15	61	absolute	8.8-fold	5.10-fold
M. dunni	412	132	311	/	/	/
N3T3	1.4×10^4	4	3.5×10^3	0.03	33	0.09
B3T3	30	3.2×10^3	3.8×10^3	13.73	0.04	0.08

Table 6.4 Results of the biological assay for *Fv1* phenotype in cell lines expressing mutant *Fv1* ORFs

Clone name	Level of restriction		
	N-tropic virus	B-tropic virus	NB-tropic virus
MM0.4	2.53-fold	4-fold	1.17-fold
MM0.5	3.4-fold	6.95-fold	3.53-fold
MM0.6	12.88-fold	6.6-fold	0.86-fold
MM0.9	absolute	8.8-fold	5.10-fold
MM0.21	10.99-fold	24.09-fold	3.44-fold
MM1.1	absolute	6.08-fold	0.58-fold
MM1.5	113-fold	2.43-fold	0.21-fold
MM1.7	absolute	1.66-fold	0.19-fold
MM5.11	3.09-fold	46.67-fold	0.81-fold
MM2		no restriction detected	
MM4		no restriction detected	

Of the 9 cell lines obtained from the transfection of *Mus dunni* with the MM1 construct, 3 clonal cell lines showed MLV restriction. In contrast to the MM0 clonal cell lines, which exhibited a range of restriction, the phenotype of these 3 cell lines was more alike. The phenotype conferred by this construct also appears to be essentially that of the $Fv1^b$ allele, as shown by the absolute restriction of N-tropic virus in cell lines MM1.1 and MM1.7, and the high level of restriction in MM1.5. This phenotype also appears to be extended, as shown by the 6-fold reduction in B-tropic infection in MM1.1, although restriction was less than half this in the other 2 cell lines. Unlike the MM0 transfectants, however, this restriction does not extend to NB-tropic virus. In fact, these clones seem to be *more* permissive to infection by NB-tropic virus. The effect of the MM1 combination therefore, which is essentially the $Fv1^n$ allele with the first $Fv1^n$-specific residue change to the $Fv1^b$-specific residue, seems to be to produce $Fv1^b$ restriction whilst greatly diminishing $Fv1^n$ restriction, and apparently increasing sensitivity to infection by NB-tropic virus.

Of the 24 cell lines obtained from the transfection of *Mus dunni* with the MM5 construct, only a single clone showed MLV restriction. The *Fv1* phenotype of this cell line was essentially *Fv1*^{n}-like, as shown by the almost 50-fold reduction in B-tropic infection. This phenotype also seems to weakly restrict N-tropic virus, as shown by the 3-fold reduction in N-tropic infection, but seems to have no effect on NB-tropic virus. The effect of the MM5 combination therefore, which is essentially the *Fv1*^{b} allele with the first *Fv1*^{b}-specific residue change to the *Fv1*^{n}-specific residue, seems to be to produce *Fv1*^{n} restriction whilst greatly diminishing *Fv1*^{b} restriction. This construct appears not to affect NB-tropic virus. The results of the *Fv1* typing of these cell lines expressing the mutant *Fv1* ORFs are shown in Table 6.5 below.

Table 6.5 Summary of the transfection with the mutant *Fv1* constructs

construct	No. clones typed	Changes in ORF	Pattern of MLV restriction	Apparent *Fv1* phenotype
MM0	20	b b n	Restricts N-tropic; lowers B- and NB titres	*Fv1*^{b}; partial *Fv1*^{n}
MM1	9	b n n	Restricts N-tropic; lowers B-tropic titres; may increase NB-tropic titres	*Fv1*^{b}; partial *Fv1*^{n}
MM2	8	n b n	none detected	null
MM4	7	b n b	none detected	null
MM5	24	n b b	Restricts B-tropic virus; lowers N-tropic titre	*Fv1*^{n}; partial *Fv1*^{b}

6.2 Natural Variants of *Fv1*

6.2.1 The *Fv1*^{d} allele

The *Fv1*^{d} allele is a variant of the *Fv1*^{n} allele. Initially there was only 1 reported difference between the alleles, namely the presence of an IAP element downstream of the gene in the *Fv1*^{n} allele that is absent in the *Fv1*^{d} allele[170]. One possible explanation for the modified *Fv1*^{n} phenotype is that the change in the 3' untranslated region caused by the absence of this IAP element may have an effect on the *Fv1*^{d} mRNA stability or expression, and ultimately the level of

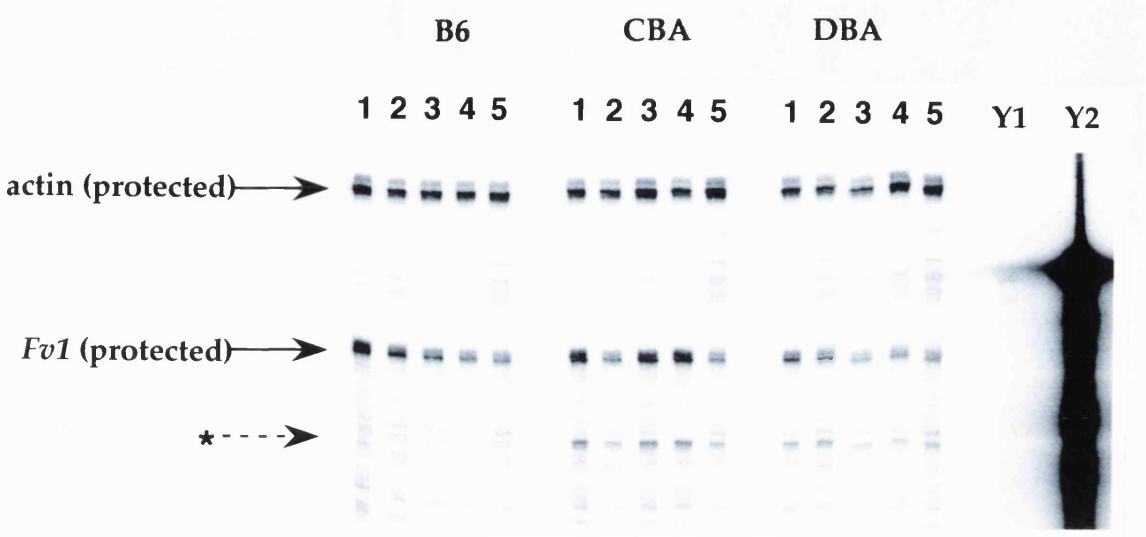
message in the cell. It may be this altered level of *Fv1* mRNA that causes the modified phenotype. To investigate this, the abundance of the $Fv1^d$ transcript in various tissues of the DBA mouse (typed as $Fv1^d$180, 310) was determined by ribonuclease protection assay and compared with the amount of message in CBA and B6 mice (which possess the $Fv1^n$ and $Fv1^b$ respectively; see Figure 6.1). In the tissues sampled, no appreciable difference in level of expression of *Fv1* mRNA was found between these 3 strains of mice.

We decided to examine how the level of $Fv1^d$ expression compared with and varied amongst other cell lines used in the lab (see Figure 6.2). As well as cell lines derived from mouse tissues, cell lines transfected with various *Fv1* constructs were also tested: B7/5 is a *Mus dunni* cell line which has been transfected with the $Fv1^b$ allele in the pCI expression vector. In this construct the *Fv1* allele has been cloned just before the starting methionine, so the 5′ untranslated region (UTR) is different to the endogenous $Fv1^b$ mRNA, and the protected fragment in this sample is subsequently shorter (157bp instead of 175bp). E12 is also a *Mus dunni* cell line which has been transfected with the genomic $Fv1^n$ *SpeI-EcoRI* fragment in the pBK expression vector, so mRNA due to expression from the natural *Fv1* promoter in this construct retains the endogenous 5′ UTR.

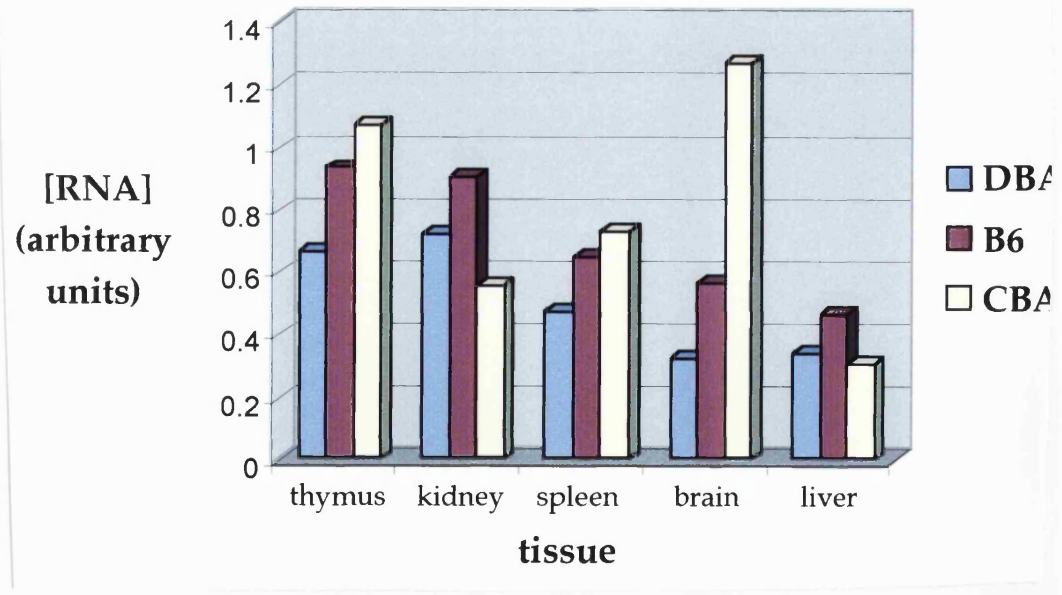
Figure 6.2B shows the relative levels of *Fv1* expression in these sampled normalised to the level of actin within the cell. In the tissues from the DBA mouse, the same pattern of expression as before was observed (see Figure 6.1B). The level of *Fv1* expression in the cell lines was found to be significantly lower in all but one of the cell lines compared to the level found in the tissues. The exception to this was B7/5, which had an apparent level of expression comparable to the level found in DBA spleen tissue. As the amount of incorporated radiolabel is proportional to the size of the protected RNA fragment, and the size of this fragment is smaller in B7/5 than in the other samples due to the absence of complementary 5′ UTR sequence, this measurement of expression level is an underestimate. The actual level of expression in the B7/5 cell line is likely to be over 10% higher than this. There was also a wide variation in expression levels between the cell lines themselves (see Figure 6.2B).

Figure 6.1 *Fv1* mRNA analysis of various tissues from DBA, CBA and B6 mice

A. Ribonuclease protection assay



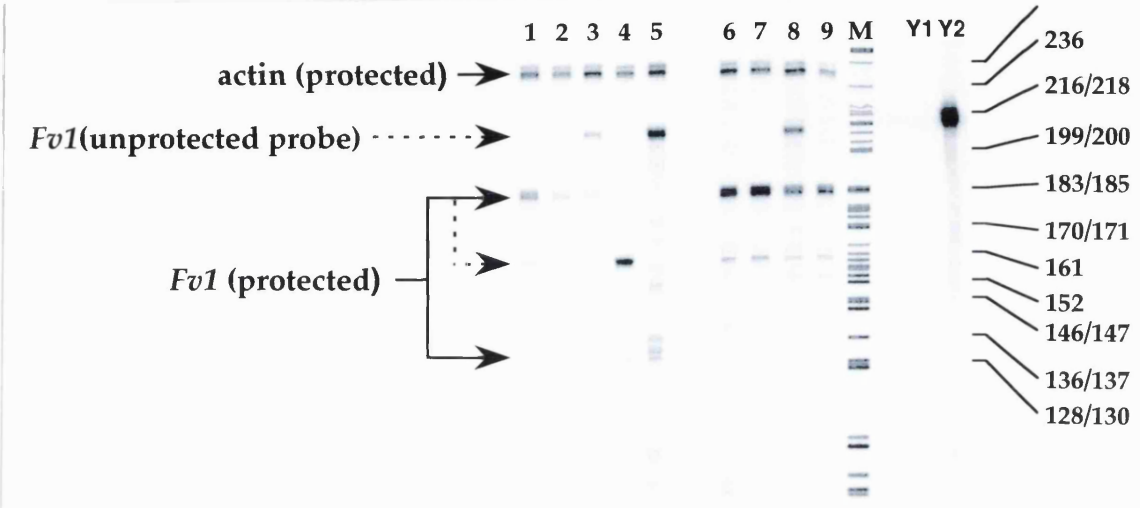
B. Quantification of *Fv1* mRNA levels



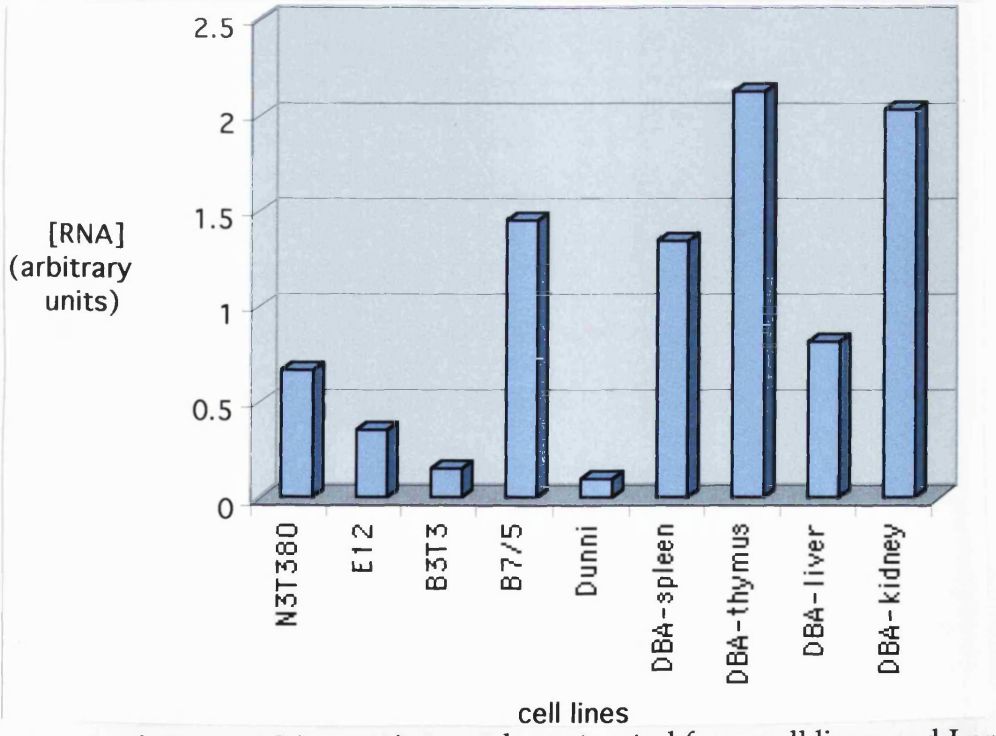
The lanes numbered in Figure 6.1A contain RNA extracted from the following tissues:
1: thymus 2: kidney 3: spleen 4: brain 5: liver
The lanes Y1 and Y2 contain 10µg Yeast RNA +/-RNase as controls reactions.
The mRNA levels in Figure 6.1B have been normalised to actin expression.

Figure 6.2 Comparison of *Fv1* mRNA levels in DBA tissues and various cell lines

A. Ribonuclease protection assay



B. Quantification of *Fv1* mRNA levels



Lanes 1-6 of Figure 6.2A contain samples extracted from cell lines and Lanes 7-10 contain samples from the tissues of a DBA mouse. The lanes were loaded as follows:

1: N3T3 3: B3T3 5: *M. dunni* 7: DBA-thymus 9: DBA-kidney
2: E12 4: B7/5 6: DBA-spleen 8: DBA-liver

The lanes Y1 and Y2 contain 10μg Yeast RNA +/-RNase as controls reactions (10% loaded).

The mRNA levels in Figure 6.2B have been normalised to actin expression.

A smaller, less intense protected fragment was seen in DBA and CBA but not B6 (marked with a * in Figure 6.1). This band was also seen in the N3T3 and E12 cell lines, but absent in B3T3, Bmyc3 and *Mus dunni* (see Figure 6.2, at the position of the protected fragment in B7/5). This band therefore appears to be specific to cells expressing *Fv1*^{n}-like alleles only, and may be due to *Fv1*^{n} mRNA secondary structure.

It appeared, therefore, that the absence of the IAP in DBA mice does not affect *Fv1* expression, and so could not be responsible for the modified *Fv1* phenotype. When the sequence of DBA-like *Fv1* ORFs (i.e. undeleted but lacking the downstream IAP element) from WSB/Ei and LG/J was analysed, an additional difference between these alleles and the *Fv1*^{n} allele in the coding sequence was detected: there is a single base pair transversion (A→C) at position 814 that changes the amino acid codon at position 270 from a lysine to a glutamine (see Table 4.2). Because of this finding the original sequence data for the *Fv1*^{d} allele cloned from DBA[170] was re-analysed, and the same change was found. It therefore seems highly likely that this amino acid change at position 270 is responsible for the modified *Fv1*^{n} phenotype associated with the *Fv1*^{d} allele, rather than by differences in expression level of the gene.

6.2.2 The *Fv1*^{nr} allele

7 of the *Fv1* ORFs isolated contain the single change associated with the *Fv1*^{nr} allele, namely a phenylalanine at position 352. These were NZB/BlNJ, NZW/LacJ, RF/J, LG/J, M.MUS, CZECHI and SKIVE/Ei. Of these, all but LG/J had been previously typed as *Fv1*^{nr} or *Fv1*^{nr}-like with respect to *Fv1* phenotype[15, 169, 173]. LG/J had been previously typed as *Fv1*^{b}, and along with BDP, is thought to have been mis-typed originally (see section 4.2.2). To test the *Fv1*^{nr} phenotype in the pseudotyped MLV assay system, the *Fv1* ORF from the NZB/binj sample was shuttled from the M13 phage vector into the pCI expression vector. The construct was then used to transfect the *Fv1*$^{\circ}$ *Mus dunni* cell line and clonal mutants were isolated by selection for antibiotic resistance in the presence of G418. Once obtained, these cell lines were frozen down until required for assay. Out of 12 clonal cell lines obtained from the transfection of *Mus dunni* with the pCI-NZB/binj construct, 2 were found to restrict MLV. These are shown in Table 6.6 (next page).

Table 6.6 Results of the biological assay for *Fv1* phenotype in cell lines transfected with the *Fv1^{nr}* allele

Clone name	Level of restriction		
	N-tropic virus	B-tropic virus	NB-tropic virus
pCI-NZB/binj.1	2.2	6.95	1.74
pCI-NZB/binj.4	1.59	50.36	3.24

As the Table shows, both clones show an *Fv1^{n}* restriction pattern. However, clone 4 appears to have a much stronger phenotype, restricting the infection by B-tropic virus over 7 times more effectively than clone 1. Again, this difference is likely to be due to clonal differences in expression of the construct in these cells. These clones cannot yet be typed for *Fv1^{nr}* activity, as the reagent required for this (namely virus pseudotyped for the Gross virus determinants) has proven difficult to clone.

6.2.3 *Mus musculus molissinus*

As shown in Table 4.1, several of the *Fv1* ORFs isolated from the mouse sequencing project had interesting changes in the coding sequence. Both *Mus musculus molissinus* samples contained an *Fv1* ORF of the undeleted, *Fv1^{b}*-like form but had the *Fv1^{n}*-associated residue at position 358 and the *Fv1^{b}*-associated residue at position 399. In this way this ORF is a natural form of the mutant MM5 (see previous section), and the phenotype of this construct will provide a good indication of the effect of this changes in the *Mus musculus molissinus* mice. However, although the biological activity of the MM5 construct in the previous section may well reflect the activity of the natural allele in *Mus musculus molissinus*, it should be noted that these 2 ORFs have a single base pair difference between them, resulting in a serine→phenylalanine change at position 352 in the *Mus musculus molissinus* allele. This residue change is the same change found in the *Fv1^{nr}* allele, and is thought to be responsible for its modified phenotype of the *Fv1^{nr}* allele. In a similar way, this difference may affect the overall phenotype in the *Mus musculus molissinus* allele, in which case the MM5 construct would not reflect the phenotype of this allele. In order to resolve this the *Fv1* ORF from the MOL sample was shuttled from the M13 phage vector into the pCI expression vector. Future work should define its biological activity.

The results of the transfection of *Mus dunni* cells with mutant *Fv1* ORFs should probably be regarded as preliminary results as the assay did not seem robust enough. It seems that the first residue change (at position 358) appears to be the most important change in determining phenotype. This is indicated by the mutants MM1 ('bnn') and MM5 ('nbb') in which the $Fv1^n$- or $Fv1^b$-specific residue at this position confers the respective *Fv1* phenotype. The results also indicate that the presence or absence of the deletion (or rather, the associated change in C-terminus) contributes to *Fv1* phenotype: it seems to mediate a partial $Fv1^n$ activity in MM0 ('bnn') that is not due to the $Fv1^n$ change at the second residue, as shown by the retention of this partial activity in MM1 ('bnn'), in which the second residue is $Fv1^b$-specific. Although it is uncertain how the change at the second residue relates to phenotype, the interaction of all 3 changes may be more complex, as indicated by the finding that some combinations appear to be non-functional (MM2 and MM4). However, it should be noted that only a small proportion of the neomycin-resistant (neo^+) transfectants obtained in the experiment showed *Fv1* activity (see Table 6.2 and 6.4). This may be due to expression of the *Fv1* constructs in these cell lines. The ribonuclease protection assays of the B7/5 cell line transfected with an $Fv1^b$ ORF cloned into the pCI expression vector showed that in these cell lines expression of *Fv1* in this vector was around the level found in normal mouse tissues. As endogenous *Fv1* has already been shown to be extremely low[170], one would assume expression from the human cytomegalovirus (CMV) major immediate-early gene enhancer/promoter region within the pCI vector, a relatively strong promoter for gene expression, would be much higher in these cell lines. It may be that expression from the CMV promoter is very weak in the *Mus dunni* cell line. If this is true, a different promoter element e.g. SV40 enhancer/early promoter, should be tested in *Mus dunni* and used instead. Alternatively, the Fv1 protein may have a deleterious effect in the cell at high levels. This would cause selection following the transfection of the construct for low-expressing cells (in which expression from the CMV promoter is reduced following integration due to regional effects). This could then account for the low number of transfectants displaying *Fv1* activity: *Fv1* expression in a larger proportion of these 'low expressers' may be below the level required to mediate an effect in the majority of the clones. In this way phenotype may be critically dependent upon a narrow range of expression levels: too high and it may have a deleterious effect in the

cell, too low and restriction may be compromised. The phenotype mediated by MM2 and MM4, therefore, may have simply not been determined due to the failure to test a functional transfectant. If this is true, a weaker promoter element or the endogenous promoter itself should be used.

Whilst the assay was potentially extremely sensitive, and proved adequate in detecting the major *Fv1* phenotypes, it was much less reliable in detecting the subtle, less marked patterns of restriction caused by some of the mutations. This was compounded by the inter-clonal variation in phenotype within clones carrying the same construct, presumably due to the difference in expression discussed above. Therefore, either a larger number of these clones, or clones with a more consistent level of expression of the *Fv1* ORF (by changing the promoter element- see above), should be assayed using this method for a more accurate determination of phenotype mediated by these constructs. Alternatively, these problems could be circumvented by using a much more robust and sensitive method, whereby expression from the construct can be determined concurrently with assessment of phenotype. Such a method has recently been developed in the lab and should provide a much more accurate assessment of the more subtle *Fv1* phenotypes (J. Stoye, pers. comm.).

The analysis of *Fv1* mRNA levels in the DBA/2 ($Fv1^d$), B6 ($Fv1^b$) and CBA ($Fv1^n$) mice showed level of expression did not differ significantly between the three strains. This indicates that the modified $Fv1^n$ phenotype of DBA/2 is not due to a difference in the level of expression of the *Fv1* gene in this mouse. Furthermore, the sequencing project identified a change in the coding region of the $Fv1^d$ allele that causes a lysine→glutamine change at residue 270 (see Chapter 4). It appears, therefore, that the modified $Fv1^n$ phenotype seen in DBA/2 is due to this amino acid change. Similarly, a single amino acid difference has been identified in the $Fv1^{nr}$ allele that is thought to be responsible for the modified $Fv1^n$ in mice with this allele (P. Le Tissier, unpublished). This is supported by the sequencing project data: all 6 of the mice previously typed as $Fv1^{nr}$ or $Fv1^{nr}$-like were found to have this change.

Chapter 7 Discussion

This project consisted of 3 main parts. In the first part we attempted to clone the *Fv1* region in human, in order to ascertain whether this region contained the human homologue of *Fv1*. To do this, 2 human genomic YAC libraries and one PAC library were screened by PCR-amplification using primers designed to the human homologues of the genes known to flank *Fv1* in mice, namely *Nfv1* and *Nfv2*. Two PAC clones were subsequently isolated that were positive for both flanking markers. Each marker was finally subcloned as separate fragments that were shown to overlap by hybridisation. In this way the human *Fv1* region was cloned as two overlapping fragments spanning 28kb. This region was subsequently sequenced using a transposon-based strategy (see section 2.17). In agreement with hybridisation data obtained during the isolation of the region, sequence analysis revealed the *Fv1* region in human does not contain a homologue of the *Fv1* gene.

The second part of this project involved identifying how different features of the *Fv1* gene contribute to overall phenotype. Transfection studies with mutant *Fv1* constructs indicate that the 3 main changes associated with either $Fv1^n$ or $Fv1^b$ phenotype do not appear to be equally important to overall phenotype. It seems the first residue change, at position 358, may be the most significant in determining phenotype. In contrast, it is not immediately clear what effect the other two changes have, but they may serve to modify the phenotype much more subtly. The $Fv1^d$ allele has two associated changes, an amino acid change at position 270 and an IAP insertion downstream of the ORF. As the presence of the IAP does not appear to modify either mRNA expression levels or stability, the *Fv1* phenotype is likely to be due to the amino acid change only. The *Fv1* alleles from mice previously typed as $Fv1^{nr}$ have only one common difference from the $Fv1^n$ allele, indicating that this change alone is probably responsible for the modified $Fv1^n$ phenotype of this allele.

The third part of the project involved attempting to establish how *Fv1* has evolved during the speciation of the genus *Mus*. To this end the *Fv1* ORF was PCR-amplified, cloned and sequenced from a variety of mice from across the *Mus* genus. Subsequent analysis of this sequence has revealed how the *Fv1* ORF has changed during the speciation of *Mus*. From the sequences isolated from the

more distantly-diverged mice we have identified the features that were present in the *Fv1* gene shortly after its appearance in *Mus*. Analysis of the sequences from members of the *Mus musculus* complex, that make up the most recently-diverged group, has enabled us to map the changes that have occurred during the evolution of the *Fv1* phenotype. From this study and from what we know of the status of endogenous MLV proviruses in this genus, both from work in this lab and from others, we have been able to outline a credible scenario which may have led to the development of both of the main *Fv1* alleles in the *Mus musculus* complex. In this scenario, the selection for *Fv1* activity is arises as a result of MLV infection. Initially the $Fv1^n$ allele was selected for as a result of germline infection by B-tropic MLV, prior to the formation of the complex. This led to the $Fv1^n$-assocated deletion at the 3' end of the gene. The $Fv1^b$ allele was then selected for by the subsequent infection by N-tropic MLV late in the formation of the complex. This then led to selection for the two amino acid changes associated with the $Fv1^b$ allele.

7.1 The *Fv1* region in human

Previous work had shown the mouse *Fv1* gene was flanked by two genes, *Nfv1* and *Nfv2* (P. Le Tissier, unpublished). Sequence database searches revealed that these genes were conserved between mice and humans. The human sequence of the genes was used to construct primers specific to both human genes. Initial attempts to PCR-amplify across the human *Fv1* region using these were unsuccessful (data not shown). We then decided to clone the region from large-insert genomic libraries using the primers to the human *NFV1* and *NFV2* sequences in a PCR-based screen

7.1.1 Problems with the human genomic libraries

Neither the ICI or the ICRF YAC libraries was positive for either *NFV1* or *NFV2*. This is not entirely surprising, as YACs containing the *Fv1* region in mice have been shown to be inherently unstable (Best, PhD thesis); during the original cloning of *Fv1*, 8 YACs were isolated that contained *Nppa* (formerly *Pnd*), *Xmv-9* and *Xmv-44*, markers shown to lie close to *Fv1* in back-crossing experiments. Of these, all showed some degree of instability. In addition, previous screening of the ICRF mouse genomic YAC library for the *Nppa* marker (estimated to lie

within 1200kb of *Fv1*) failed to isolate any YACs. Therefore, it is possible that the *Fv1* region in human also contains a homologous region of instability which could hamper its cloning.

The HGMP human genomic PAC library, however, did contain clones that contained both markers. As these clones contain inserts 3-4 times smaller than the YAC clones, these clones may lack the region of instability and so facilitate the cloning of this region. Alternatively, the problem of instability may be inherently different between PACs and YACs, due to the difference in host used to maintain and propagate the clones (i.e. bacteria and yeast), differences between the 2 cloning vectors themselves, or a combination of both.

7.1.2 Absence of *Fv1*-related sequence in humans and its significance

The southern blots of PAC clones PAC107A22 and PAC277I9 were hybridised with a probe specific to the *Fv1* ORF to identify fragments that contained the human homologue. Unexpectedly, no bands were observed, even under a low stringency. Concurrent with this result, a routine database search using the *Fv1* sequence identified a newly-submitted genomic sequence that shared a 60% homology to *Fv1* over 1300bp (the *Fv1* ORF). This clone represents a member of the human endogenous retrovirus family, HERV-L[103]. The region of the cloned HERV-L shown to have homology to *Fv1* does not encode an ORF but this is probably due to mutational decay of the ancient provirus. As *Fv1* is homologous to the region between the proviral LTR and the *pol* gene, this region seems to correspond to the *gag* gene of the HERV-L family[170]. In support of this, the *Fv1* ORF also contains the major homology region (MHR), a stretch of 20-30 amino acid residues which represent the only highly conserved sequence among retroviral capsid proteins[103] ~~(see Appendix 2)~~. This suggests a possible mechanism for *Fv1* action: as the viral determinant of *Fv1* restriction is known to map to the capsid (CA) protein of MLV, *Fv1* may be *gag*-like and share similarity with its target, perhaps allowing direct binding[311]. The evolutionary significance of the finding that *Fv1* has retroviral origins, and was not after all a 'native' host gene, was that it no longer seemed certain that there had to be homologues of *Fv1* in other species; if the integration event occurred after mouse and human diverged, for instance, then there will not be a human homologue. This seems to be the case. Although a homologue is absent in human it might be present in other species more closely related to mice, depending on when the HERV-L

integration event took place. To investigate this a zoo blot was probed with the *Fv1* ORF under low stringency[170]. Single-copy bands were present in 3 strains of mouse DNA but no reactive bands were seen in rat, cat or human samples. Therefore it seems that *Fv1* has arisen as the product of an integration event, and that this integration event occurred fairly recently, sometime after the divergence of mouse and rat, within the last 10 million years[279]. Additionally, whilst the presence of *Fv1* in *Nannomys* and *Coelomys* (shown by the amplification of the *Fv1* ORF from the *Mus minutoides* and *Coelomys famulus* samples) indicates that this integration event must have occurred some time prior to around 5 million years ago, the presence of *Fv1* in *Apodemus*[315] indicates this event has occurred before this diverged from the rest of the *Mus* genus, around 8 million years ago.

7.1.3 Comparative analysis of human and mouse regions

As mentioned in section 3.6, because the *Fv1* gene appears to be the result of a retroviral integration event that has occurred in the *Mus* genus only, the sequence of the *Fv1* region from rat and human represents the status of the region prior to integration. The analysis of the more recently-diverged rat sequence, which shows high homology with the mouse sequence, has shown us that the region immediately flanking the *Fv1* preintegration site has remained intact. This is reinforced by the demonstration of homology between the human and mouse flanking sequences, although this is less clear due to the higher degree of sequence drift between these 2 more distantly-diverged organisms, and also the wide-spread accumulation of repetitive elements in the human sequence. The 3kb of sequence from the start of *Fv1* to the end of the B2 repeat in mouse is missing in rat, which contains instead 3kb of unique sequence. It cannot be determined whether this sequence has been lost subsequent to, or more likely as a direct result of, the intregration event that gave rise to the *Fv1* gene, or whether it has been gained in rat since mouse and rat diverged. If the former is true, the events leading to the formation of the *Fv1* gene have resulted in the loss of 3kb of original sequence at the preintegration site, and the substitution of this with 3kb of retrovirally-derived sequence; if the latter possibility were true, then the integration event has resulted in minimal change in the region, as no loss or gain of sequence can be detected.

It is interesting to note that this 3kb of sequence containing *Fv1* in mouse contains homology to retroviral sequence at both termini: at the 5′ end is the *Fv1* gene, which has been shown to have homology to the *gag* gene of MERV-L and HERV-L viruses. The 3′ end shows homology to retroviral sequence as determined by the repeat filtering program RepeatMasker (Smit, A.F.A. and Green, P., unpublished; http://ftp.genome.washington.edu/RM/RepeatMasker.html), a program that identifies and masks repetitive elements within a sequence during BLAST searches (data not shown). The results from these searches are summarised in Appendix 7, and identified a sequence with 82% homology over 824bp to MYSERV, a 6550bp sequence of repetitive element submitted to the rodent repetitive element database Repbase (Genetic Information Research Institute). MYSERV represents the *pol* sequence of an endogenous retrovirus associated with the mouse *mys-1* transposon, and is most closely related to the HERV-K *pol* gene. The 824bp corresponds to the 3′ terminus of this element. Directly flanking this sequence in mouse are 196 and 146bp sequences with 77% homology to the LTRs that flank the MYSERV submitted sequence. It seems, therefore, that this sequence is a deleted form of the MYSERV submitted sequence, and appears to be unrelated to the retroviral sequence that gave rise to *Fv1*.

7.2 Search for the *Fv1* progenitor

The screen of the mouse genome failed to identify, by hybridisation, sequence with more homology to the *Fv1* gene than to MERV-L elements. This could be due to a variety of different reasons. Firstly, the screen simply may not have been exhaustive enough to identify this element, especially if it is present in the germline as a single copy. While this is impossible to rule out, the screen did identify 3 separate clones that contained *Fv1*, which is obviously also present at a single locus. It seems likely therefore that if a sequence were present in the library with a significantly higher sequence homology to *Fv1* than the MERV-L elements, it should have been readily identifiable in this screen. In agreement with this, the southern genomic blot hybridised with a probe specific to *Fv1* showed hybridisation only to the *Fv1* bands (at higher stringencies). If, however, the sequence of the progenitor had diverged to such an extent, by mutation, deletion/insertion or rearrangements, so that it was not significantly more homologous to *Fv1* than MERV-L elements (by hybridisation at least), then the

likelihood of its isolation in a screen of this type would be extremely low. If this were the case the >100 copies of MERV-L resident in the laboratory mouse genome[316] would mask the identification of the progenitor, essentially making the detection of such a sequence a hundred times more difficult. Another reason for the failure to detect progenitor sequence could be that this sequence has been lost from the genome in the period subsequent to integration. Homologous recombination between LTRs can lead to excision of such elements, and has been observed in most retrotransposons. The presence of single LTRs throughout the genome testifies to this phenomenon. Furthermore, within the MERV-L family in mouse, the closest relative of *Fv1*, this is the fate of nearly 90% of all such elements[316]. Alternatively, this sequence may not have been lost from the germline at all, because it may never have been there in the first place. While the genome is awash with a molecular fossil record of past retroviral invaders, this record is far from complete. Only when the germline itself is infected can we detect past encounters with retrotransposable elements. It is possible, although probably not likely, that a retrovirus of considerable importance in the past did not infect the germline, at least of those mice whose descendants we see today. This does not rule out the presence of this progenitor sequence in other mice species, whose progenitor's germline may have acquired it. It would be interesting to probe a genomic southern blot of mice from across the genus with the *Fv1* ORF, to see if any of the branches contain this sequence.

The 2 *Fv1* sequences isolated, 2DK1 and 6L10EH, were virtually identical. Since the purpose of sequencing these clones was for identification only, the lambda sequencing project was not as rigorous as the mouse sequencing project; each clone was isolated only once from the library, and single-pass sequence alone was obtained from these clones. As a result of this, the sequences may contain a low level of errors due to misreading of sequence traces, and this could account in part for the small amount level of apparent divergence between these two sequences. The absence of the *EcoRI* site in clone 2DK1 is, however, a true change, as the confirmed by the sequence of both strands. This result seems to explain the two bands observed in the 129 genomic southern blot hybridised with the *Fv1* ORF (see Table 5.1): the presence of the *EcoRI* site in only one of the *Fv1* alleles of 129 would produce a restriction fragment length polymorphism (RFLP) at this locus, leading to the *Fv1* alleles locating to different size *EcoRI*

fragments, and subsequently produce 2 hybridising bands on a 129 genomic southern blot. However, as already discussed in section 5.5.3, the size of these bands infers further loss of 2 further restriction sites in these clones compared to the published sequence of *Fv1*^{b}, at least 1 of which is retained in 2DK1. This is an unexpected result, as 129 represents a highly inbred strain of laboratory mouse, and as such should be homozygous for the *Fv1* allele. Such polymorphisms within any inbred or congenic strain can arise as a result of *de novo* mutation, incomplete backcrossing during congenic strain construction or genetic contamination (breeding error). Although 129/SvJ (used to make the mouse genomic library) is an inbred strain, it has been shown that, due to genetic contamination, it is more accurately classified as a recombinant congenic strain between 129/Sv and an unknown non-129 strain[317, 318]. There is also considerable variation in the 129 sub-strains as a whole due to at least 3 separate confirmed instances of genetic contamination. It is unlikely that there is still polymorphism due to this contamination, as 129/SvJ should be completley congenic after so much inbreeding. However, although this appears to be supported by the fact that this strain has been shown to be isogenic for 86 markers analysed, segregating alleles have been found in 2 other strains, 129/SvPas and 129/Sv[317]. Therefore, polymorphism due to incomplete (if inadvertant) congenic strain construction cannot be ruled out. Alternatively, these polymorphisms may have occurred spontaneously as a result of random mutation in this mouse, although this is unlikely due to the relatively low frequency this occurs.

These polymorphic restriction sites were not the only differences found between these 2 sequences. Whilst the *Fv1* gene from the 2 lambda subclones isolated were almost identical, only the ORF from 2DK1 was found to be open. It seems therefore that the 129/SvJ strain contains 2 alleles of *Fv1*. Genetic contamination has been mentioned as one possible explanation. However, another possibility is that there has been a duplication of the *Fv1* gene in this strain. Gene duplication has occurred repeatedly throughout evolution, as a way of unshackling a gene from the constraint of selection (discussed in the Introduction). This would be in accordance with the fact both alleles are *Fv1*^{nr}; if there are 2 alleles present due to genetic contamination, the probabilty is extremely low the contaminant strain would have the same, rare *Fv1* allele. Also, one would expect the restriction fragment patterns to be very similar, if not identical, between 2 isolates of the

same allele. Gene duplication, on the other hand, frequently results in the insertion of the genecopy into an entirely different locus, which would result in a difference in restriction fragment patterns compared to the original gene. This could also explain why only 1 of the alleles is open: gene duplication frequently results in the formation of pseudogenes, as the lack of negative selection for the maintenance of the duplicated gene results in the rapid accumulation of mutations.

To identify which of these scenarios is true, one could look for segregation of the 2 alleles from 2DK1 and 6L10EH: if there are 2 alleles in 129/SvJ, these will segregate in 50% of the offspring, whereas if this strain is homozygous for *Fv1* and the second band seen in the genomic southern is a result of gene duplication, the duplicated allele will fail to segregate.

7.3 Contributions of mutations to *Fv1* phenotype

The results of the *Fv1*-typing assay of transfectants expressing mutant *Fv1* ORFs in an *Fv1*$^{\circ}$ cell line gave reproducible but not always clear results: although the 'major' restrictive phenotype produced by each construct could be readily and reproducibly detected, the exact level of restriction seemed to vary between clones. The less pronounced patterns of restriction consequently proved more difficult to identify. This may simply be due to clonal differences between transfected cell lines. This is indicated by the finding that only a small proportion of all *neo*$^{+}$ clones containing the mutant constructs showed *Fv1* activity. Additionally, when expression level in a similar *Mus dunni* clone containing an *Fv1* construct (B7/5) was measured, it was found to be comparable to the low, endogenous level, even when such expression is driven from a strong CMV promoter. As discussed at the end of Chapter 6, whether the problem lies with the use of this promoter in *Mus dunni* cells, or whether the Fv1 product is toxic at higher levels in the mouse cell, clonal variation in expression levels of *Fv1* in these clones is suspected to be the most likely explanation for the variability in restriction levels measured in this assay. However, the possibility that the assay itself may be responsible cannot be ignored: whilst the assay is potentially extremely sensitive, the evaluation of *Fv1* restriction using pseudotyped MLV may well be equally sensitive to external variables such as variations in cell density at infection and the viral titre of reagents. To minimise this possibility, every effort was made to ensure the test plates contained the same number and

density of cells at infection to minimise variation between plates, a task not too difficult with careful tissue culture for each separate clone. However, the same should apply between clones in each experiment and between experiments, a task more difficult to achieve. The exact level of cell density at which to infect may also be an important factor. Determining the importance of these variables experimentally may increase the stringency and accuracy of the assay for the future.

It should also be noted that the cell line chosen to assess *Fv1* phenotype in may also be an important variable. Although the *Mus dunni* cell line used in these experiments is null with respect to *Fv1* activity, it may not be entirely inert, and as such may not be the ideal cell line for this type of experiment. Nuclease protection experiments corroborate the sequencing data to show that an *Fv1* mRNA is still produced in this cell line and is predicted to code for a truncated protein. Although any notable C-terminal truncation has been shown to abolish activity (J. Stoye, pers. comm.), the interaction between these self-similar *Fv1* products is unknown. Also, workers have shown that although these feral mice have the null allele, they all show some degree of innate resistance to MLV infection, and this has been assumed to be the result of other genes (this why MLV is mildly restricted in *Mus dunni* cells in this assay compared to N3T3 and B3T3 cells- see Table 6.3). Whether this innate resistance is due to other genes or due to the effect of the expression of the truncated endogenous Fv1 product, the only way to avoid this variable would be to use an inbred mouse strain that produces no *Fv1* product. There evidence to indicate that such strains may well exist in labs today (e.g.SC-1[178] and 3T3FL[174]- see section 1.5.5).

The results of the mutant constructs into *Mus dunni* indicates that the single most important change in determining *Fv1* phenotype appears to be the amino acid change at position 358 in the ORF, at least in the combinations MM0, MM1 and MM5. Whilst the fact that none of the MM2 and MM4 clones demonstrated a measurable level of restriction may reflect the true activity of these constructs, it is conceivable that, due to the small number of clones tested, cell lines displaying the phenotype were simply not tested.

An *Fv1*^{n}-associated lysine (K) residue at position 358 produces an *Fv1*^{n} restriction, even when the other two changes are the *Fv1*^{b}-associated changes (as in mutant MM5). Conversely, an *Fv1*^{b}-associated glutamic acid (E) residue at this position

produces an *Fv1*^{b} phenotype, again even when the other two changes are associated with the opposite allele (as in mutant MM1). The next most important contribution to phenotype appears to be the presence or absence of the *Fv1*^{n}-associated deletion which changes the 3′ end of the gene. An *Fv1*^{n}-like terminus produces a weak *Fv1*^{n} restriction whilst the opposite is true of the *Fv1*^{b}-associated terminus. In the absence of complete data for all of the constructs, it is difficult to assign an effect of the second change (at position 399 in the ORF) on *Fv1* phenotype. However, there is an indication from the data that it may have an effect, in conjunction with the *Fv1* terminus change, on the restriction of NB-tropic virus.

7.4 Evolution of *Fv1*

As indicated in the comparative analysis of the *Fv1* region from human and rat with mouse (section 7.1.3), the integration event that has given rise to the *Fv1* gene appears to have done so with minimal change at the site of integration. The gene itself corresponds to a portion of a retroviral *gag* gene. This indicates the event involved either the partial integration of the retroviral *gag* sequence originally, or the integration of the full-length retroviral genome followed by subsequent deletions removing the 5′ LTR and the 3′ LTR and the rest of the proviral genome. If the latter were true, the deletions may have occurred in a step-wise process, the intermediates of which might still be present in the more distantly-diverged mice. In the analysis of the sequences flanking the *Fv1* gene in the mice analysed, only *Pyromys platythrix* was found to have a significant insertion, 5′ of the *Fv1* gene. However, the lack of homology of this sequence to the HERV-L and MERV-L marks this most likely as the result of an independent integration event,and as such does not appear to be such an intermediate in the formation of *Fv1*. It should be noted, however, that this sequence does contain a pair of LTRs, in opposite orientation, which could mark this location as a previous site of retroviral integration (see Appendix 7).

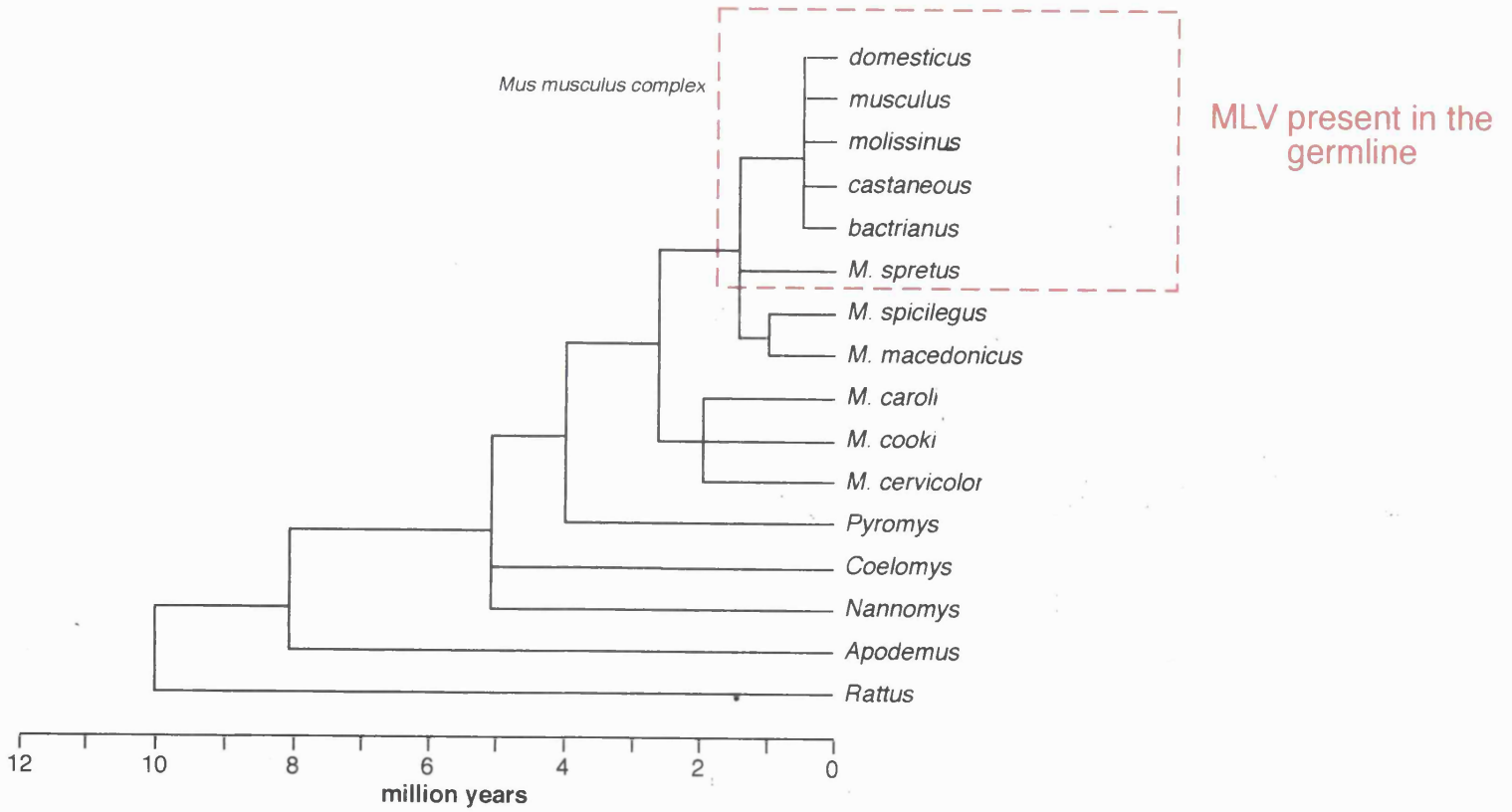
The sequencing project revealed both shared similarities and striking differences between some of the strains. Analysis of this data has provided an insight into the evolution of this gene.

7.4.1 Features of the phylogenetic tree

The multiple alignment of all the sequences of the *Fv1* ORFs was used to produce the phylogenetic tree shown in Figure 4.3. It should be stressed that the data presented in this tree is based on the *Fv1* gene alone; extreme care must be taken when inferring phylogenetic relatedness of a group of organisms from the analysis of single character sets or animals. The tree obtained therefore is strictly a genetic tree of the *Fv1* gene, not a species tree of the genus *Mus*. The analysis of a single gene or gene family to infer genealogy can be highly inaccurate and extremely misleading[276, 312]. However, the tree obtained through the analysis of *Fv1* sequence is in very close agreement with current phylogenetic trees of the genus *Mus* based on a variety of methods[279] (Figure 7.1), and as such it can be assumed that they both show the same phylogenetic relationship.

The tree deviates significantly with respect to just a single branch: the *Mus caroli/Mus cookii* samples both branch off one level earlier than in the published tree, branching off with *Pyromys platythrix* instead immediately after this species. Although current data suggests *Mus dunni* branches after *Pyromys platythrix*, the number of specimens sampled does not allow for accurate positioning, so it is not shown in Figure 7.1. The exact point of branching of *Mus dunni* therefore remains undetermined. It is interesting that the *Mus cookii* (and *Mus dunni)* ORFs are both truncated due to frameshift mutations; if the ORF was being selected for at this point, it could be argued that these were no longer under the same evolutionary pressure as the rest of the tree, and this could explain the discrepancy in the alignment of these two species compared to the current phylogenetic tree for *Mus*. The similarity between the *Mus caroli* and *Mus cookii/Mus dunni* ORFs would then have caused *Mus caroli* to be mis-grouped with them. However, as discussed later, there is no direct evidence that the ORF was under this kind of selective pressure at this point. It is more likely that the difference in the placing of these is due to the fact that the construction of this tree is based on the comparison of a single gene alone.

Figure 7.1 Phylogenetic map of the genus *Mus* showing integration of MLV provirus into the germline [279]

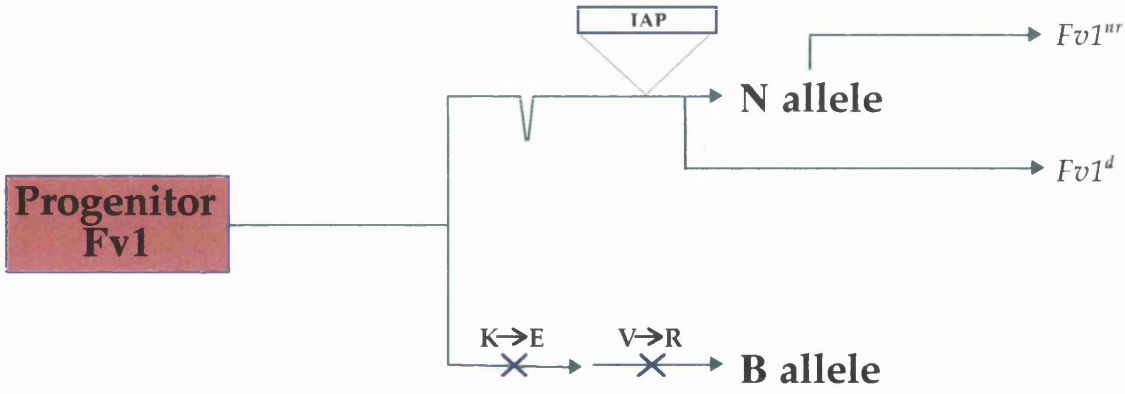


This is the current consensus of the phylogenetic tree of the genus *Mus*. Also indicated on the tree are those members that possess endogenous MLV.

7.4.2 *Fv1* gene features across the *Mus* genus

As mentioned in the Introduction, the *Fv1*^{n} allele contains an internal deletion as well as the insertion of an IAP element not seen in the *Fv1*^{b} allele. There also are two distinct changes in the ORF between these alleles. The deletion seen in the *Fv1*^{n} allele was absent in the more diverged mouse species. Analysis of the two amino acid residues known to be important for *Fv1*^{n} phenotype showed that the ten out of the eleven wild mice outside the *Mus musculus* complex had the two residues specific to the *Fv1*^{n} allele (one of the mice, *Pyromys platythrix*, has one residue specific to the *Fv1*^{n} allele, and the other is specific to neither, presumably due to random mutation after this species branched from the rest of the tree). Taking these findings together with the phylogenetic tree data, it is possible to infer a likely sequence of events that led to the two alleles (see Figure 7.2 below).

Figure 7.2 Inferred sequence of events in the formation of the *Fv1* alleles in the inbred strains



The key events leading to the formation of *Fv1* are shown. In the formation of the *Fv1*^{n} allele, these were the deletion at the 3′ end of the gene, and the insertion of the IAP element in the downstream B2 repeat. In the formation of the *Fv1*^{b} allele, 2 mutations in the ORF have occurred

Before now it has not been known how *Fv1* has changed structurally, whether the progenitor of *Fv1* had the structure of the *Fv1*^{b} allele with the *Fv1*^{n} allele arising from an internal deletion, or whether the reverse occurred, with the *Fv1*^{b} allele arising from an insertion into an *Fv1*^{n}-size progenitor. Based on the sequence data it now seems that the former was true: *Fv1* was originally *Fv1*^{b}-size, with the two *Fv1*^{n}-associated amino acids in the ORF. Over the course of evolution, the *Fv1* ORF underwent various mutations, culminating either in the change of the 2 amino acids in the ORF to form the *Fv1*^{b} allele or in the 3′ deletion

of ~1300bp to form the *Fv1*^{n} allele. This deletion left the *Fv1*^{n} allele 19 amino acids shorter than *Fv1*^{b}, and changed the last 3 amino acids.

In the wild mice, neither the changes in the *Fv1* ORF or the *Fv1*^{n}-associated deletion event have occurred. Wild mice have been reported to contain null alleles in conventional *Fv1* assays[173]. However, these mice show some small degree of MLV restriction, which has been attributed to other loci in these mixed natural populations. Further analysis of restriction using much more sensitive assays will determine whether the unchanged *Fv1* allele in these mice is truly null.

Two of the mice with *Fv1*^{n}-sized *Fv1* ORFs lacked the *Fv1*^{n}-associated IAP in the 3' part of the gene. On sequencing the *Fv1* ORFs from these two samples, WSB/Ei and LG/J, they were shown to contain the single base pair mutation associated with the *Fv1*^{d} allele. These two mice therefore contain not the *Fv1*^{n} but the *Fv1*^{d} allele. As the IAP was found to be widely distributed in both the *Mus musculus musculus* and *Mus musculus domesticus* subspecies, it is likely that the absence of the IAP in the deleted ORFs is due to loss of the element rather than a late integration event. Sequence data of the B2 repeats of the *Fv1*^{d} and *Fv1*^{n} alleles (i.e. of deleted alleles both with and without the IAP element) have shown that this excision event must have been complete (data not shown). As perfect excision events of these type of elements have not been reported, this implies that the IAP was removed either by gene conversion or recombination with the *Fv1*^{b} allele rather than by excision.

As mentioned in the Introduction, the *Fv1*^{nr} allele is manifested as a modified *Fv1*^{n} phenotype, allowing *Fv1*-mediated retroviral restriction not only of B-tropic virus but also some N-tropic virus[167, 179]. Five of the mice strains used in the sequencing study have been previously typed as *Fv1*^{nr}, and the only shared difference in the sequence of their respective *Fv1* ORFs between them and those typed as *Fv1*^{n} seems to be a single nucleotide substitution, leading to a threonine→phenylalanine change at the amino acid level at position 352. This strongly suggests that this mutation alone is required for this modified phenotype. However, other differences not analysed, such as differences in the promoter sequence and expression levels of the message, may yet be responsible for this modified phenotype. This has yet to be determined.

In the case of DBA/2, a single amino acid change and an associated loss of the IAP in the B2 repeat seem to be involved in the modified phenotype we see in

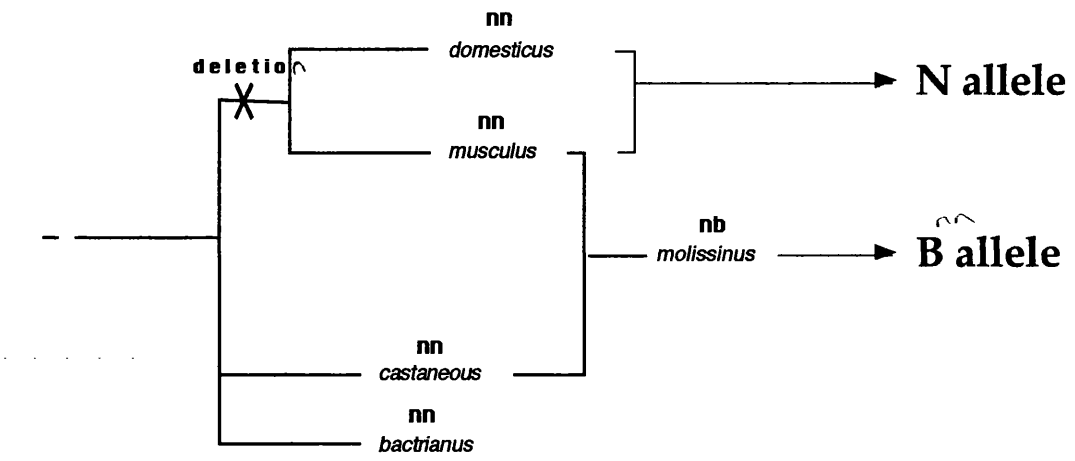
this strain, namely elevated B-tropic and lowered N-tropic viral titres. This change, which results in a lysine to glutamine amino acid change at position 270, and the associated loss of the IAP element was seen in two of the mice sequenced in this study, but, as they have yet to be typed for *Fv1* biological activity, no conclusion can be drawn. One current idea is that the loss of the IAP may alter the expression level of *Fv1* in these cells. However, as no difference in either expression levels or pattern of expression could be detected between DBA, CBA or B6 mice, it seems more likely that the modified phenotype is due to the single mutation alone and that the loss of the IAP is simply an event which has occurred at the same time as the point mutation in these mice.

It is interesting to note that the 3 most distantly-diverged mice, *Mus minutoides*, *Coelomys famulus* and *Pyromys platythrix*, all have a B1 repeat inserted at the same location, which corresponds exactly to the deletion site that gave rise to the *Fv1*^{n}-size allele (see Figure 4.5). If this element in these 3 species of *Mus* was due to a single integration event, then one would expect it to be present in subsequent species. As this is not the case then the B1 element would have to have been lost shortly after the divergence of *Pyromys platythrix* with the rest of the genus. Comparison of the sequence at the insertion point in mice lacking the repeat with those that contain it shows no deletions at this point, indicating that if the sequence was lost subsequent to insertion, it was excised perfectly. As such a specific deletion is unlikely, and the fact that the repeat in *Coelomys famulus* and *Mus minutoides* appear to be more similar to each other than *Pyromys platythrix*, it may be that the insert in these 3 mice is a result of 2 independent insertions, the first after *Coelomys* and *Nannomys* diverge from the rest of the genus, and the second after the divergence of *Pyromys*. If this is the case, then this site may favour the insertion of such elements. Deletion during such an event as this location may well have created the *Fv1*^{n}-size allele we see today.

Precisely when these different changes which have proved important in the evolution of *Fv1* function have occurred has proven more difficult to show. Computer-aided comparisons of the sequence data has shown us how the structure of *Fv1* has changed through the evolution of the genus *Mus*, and show the *Fv1*-associated changes appearing soon after the *Mus musculus* complex diverged from the rest of the tree. Events after this point in evolutionary time are much more difficult to determine. As explained in the Introduction, members of the *Mus musculus* complex, although mostly geographically distinct,

do exchange genes wherever they come into contact. As well as these natural blending of populations, *Mus musculus musculus* and *Mus musculus castaneous* have been transported by man in recent times to most parts of the world, sometimes giving rise to new hybrid populations[311-313]. This emphasises the fact that *Mus musculus* is a single polytypic species which is, due to the influence of man, probably in a stage of de-differentiation[104]. As such it is inevitably difficult to demonstrate how the members in the *Mus musculus* complex have evolved. However, from the sequence data it is possible to imagine a credible scenario (see Figure 7.3 below).

Figure 7.3 Scenario showing possible sequence of events leading to the *Fv1* phenotype in the *Mus musculus* complex



The Figure shows the sequence of change in the *Fv1* gene during the formation of the *Mus musculus* complex. The $Fv1^n$-associated deletion point, inferred from its distribution in the study, is marked, as well as the status of the allele-specific residues in each subspecies.

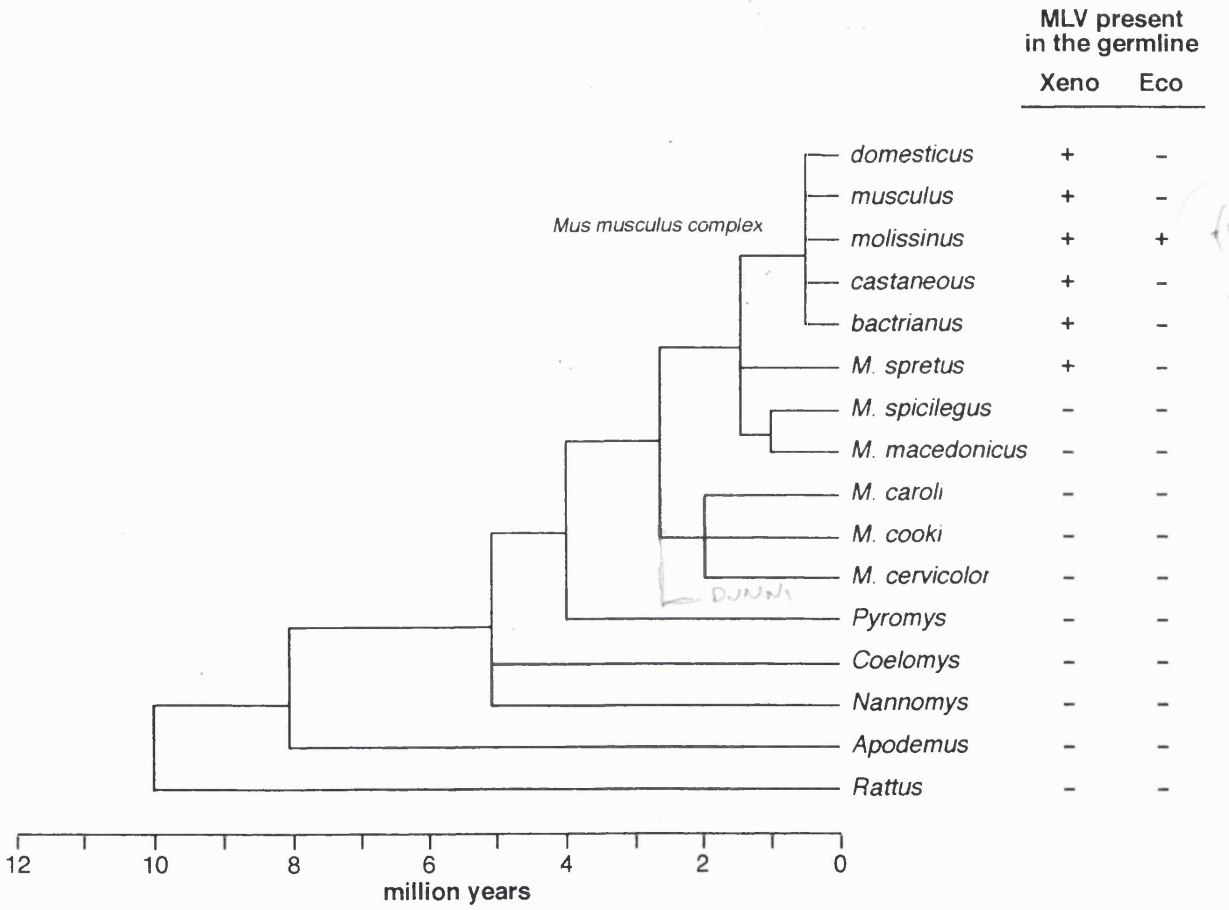
As it is not clear as to the ancestral species of many of the inbred strains, no information about the recent evolution of the *Mus musculus* complex can be gained from these, and so these have been ignored in the following scenario. In fact, the established inbred strains have been found to be a combination of very divergent genomes, including *Mus musculus domesticus* and Japanese mice[269, 289]. In the scenario the fact that the deletion is seen in both *Mus musculus musculus* and *domesticus* but *not* *Mus musculus castaneous* or *bactrianus* indicates that this event occurred after the latter two diverged from the former two, but before *Mus musculus musculus* diverged from *domesticus*. Once the deletion took

place the *Fv1*^{n} phenotype subsequently appeared and is seen in both *Mus musculus musculus and Mus musculus domesticus.* The subspecies *Mus musculus molissinus* then inherited the *Fv1*^{b}-size undeleted form of *Fv1* from *Mus musculus castaneous,* and it is in this subspecies that the first *Fv1*^{b}-allele-specific amino acid change arises. In this way, *Mus musculus molissinus* seems to represent the intermediate allele in the formation of the *Fv1*^{n} allele. This then enters the inbred strains as the Japanese component of its mixed lineage. The *Fv1*^{b} phenotype then appears subsequently in the inbred strains. The scenario assumes that the WMP/Pas *Mus musculus domesticus* strain from Tunisia has been either wrongly identified as *Mus musculus domesticus* or is a hybrid of some sort.

It seems reasonable to believe this is how *Fv1* has changed during its evolution, now we should address why: why have these changes occurred, what has been the force behind the selection for *Fv1* phenotype in mice? The answer, we believe, lies in the germline of these mice. The presence of MLV in the *Mus* germline provides a strong indication of what this selection pressure that drove the evolution of *Fv1* phenotype may have been. Around the time *Mus spretus* and the *Mus musculus* complex diverged, MLV, the virus restricted by *Fv1,* entered the *Mus* germline[96]. Soon after this event *Fv1* restriction arose in the *Mus musculus* complex, so it seems the *Fv1* phenotype evolved against a background of MLV infection, and most probably as a direct result of it. This is highlighted when examining the protein alignment of the ORF (see Appendix 4): there are a number of presumably random changes in individual species, mainly in the older mice, but at the *Mus spretus/Mus musculus* division there are six or seven distinct changes between the groups (not including the two *Fv1*^{n}/*Fv1*^{b}-specific changes and the presumptive *Fv1*^{nr} change) over a relatively short period of time. Therefore, at the same time that MLV appears in the germline, in a short period of time a large number of discrete changes occur in the *Fv1* gene, some or all of which may be important for *Fv1* phenotype. This is discussed in more detail later in this chapter.

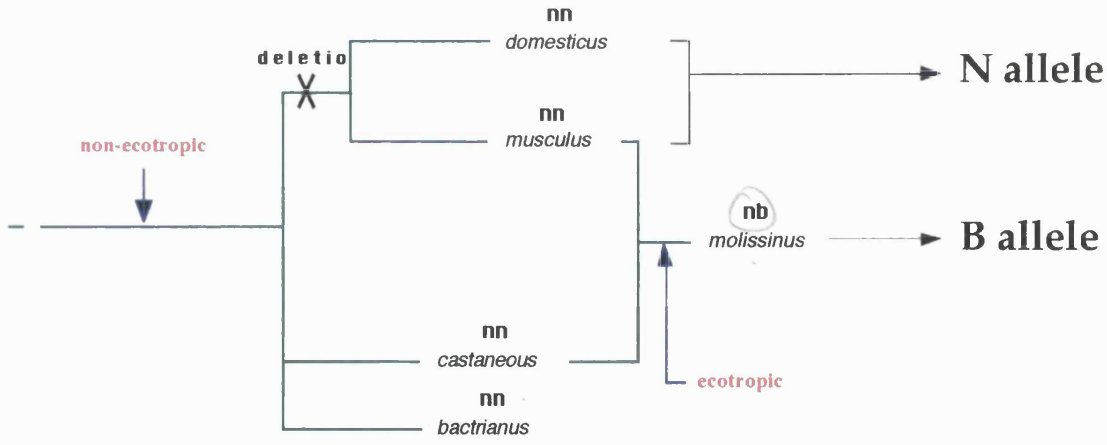
Figure 7.4 shows the status of endogenous MLV in the germline of these mice. The most widely distributed type of MLV seen in the genus *Mus* is non-ecotropic which entered the germline around the time *Mus spretus* diverged from *Mus musculus*[96]. This type of retrovirus is almost certainly B-tropic for *Fv1* restriction based on the sequence data of a large number of cloned endogenous non-

Figure 7.4 Phylogenetic map of the genus *Mus* showing the nature of endogenous MLV in the germline[279]



ecotropic MLV (A. Stevens and J. Stoye- unpublished data), and so created a selection pressure for the control of B-tropic virus- the deletion specific to *Fv1*^{n} alleles would therefore have been selected for and the *Fv1*^{n} allele would have then appeared. The selection pressure for the evolution of the *Fv1*^{b} phenotype is N-tropic retroviral infection. This has occurred more recently than non-ecotropic infection, and is shown by the presence of ecotropic AKV MLV in the *Mus musculus molissinus* germline[96, 314]. Once infected the selection pressure for N-tropic retroviral restriction caused selection for the two amino acid changes (the first of which is seen in the allele of this species), and *Fv1*^{b} phenotype appeared soon afterwards (in certain inbred strains, via the *Mus musculus molissinus* Japanese component). This scheme is summarised in Figure 7.5 (below).

Figure 7.5 Possible events leading to the formation of *Fv1* phenotype against a background of MLV infection



The Figure shows the same scheme as Figure 7.3 but has marked on it the points of MLV invasion of the *Mus* germline, based on the distribution of endogenous MLV across the genus

In the above scheme, selection for *Fv1* activity is due to infection by MLV. Is this selection for the prevention of infection by exogenous MLV, or selection for the restriction of endogenous, integrated virus, the amplification of which would cause disease? The answer is almost certainly both, as the prevention of either would increase fitness, and *Fv1* activity would prevent both.

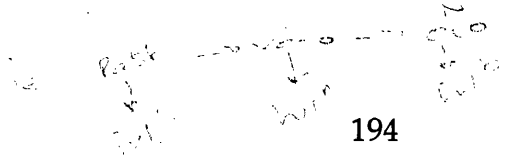
7.4.3 Evidence for the maintenance of the *Fv1* ORF

Out of 41 *Fv1* sequences analysed from 10 different species of mice, only 2 have premature stop codons. If we exclude the inbred laboratory strains of mice, in which *Fv1* activity is thought to be present, this leaves 18 sequences comprising almost 24kb of sequence. Among these sequences there have been 184 unique point mutations (not including the numerous insertions and deletions (indels) which could have also caused the introduction of stop codons). As 3/64 of random mutations will produce a stop codon in a neutrally drifting sequence, these alone should result in 9 stops among these sequences if the ORFs were under no selection. This strongly suggests that this sequence cannot be functionless in mice without *Fv1* activity, as the *Fv1* ORF appears to have been maintained intact at least since the divergence of *Coelomys famulus*, some 5 MY ago.

Further evidence for the maintenance of the *Fv1* ORF during the speciation of *Mus* came from the analysis of synonymous and non-synonymous mutation rates in the more-distantly-diverged mice. The ratios of these 2 types of mutation indicate that the *Fv1* ORF has been under negative selection, at least in *Coelomys famulus*, *Mus cookii* and *Mus caroli*, whose Ks/Ka ratios were found to be 4-5x higher than expected for a sequence under random mutation alone. However, this evidence is far from conclusive, as the ratio expected for structural genes undergoing negative selection is usually in the order of 5:1. The disagreement may reflect the fact that that the entire ORF was used, and may indicate that different parts of the gene are under different selection pressures i.e. parts of the gene may be tolerant to change (i.e. non-synonymous mutation), and mutations of this type in these regions would reduce the Ks/Ka ratio. This is supported by the fact that the mutations that have occurred in the *Fv1* ORF are not randomly distributed along its length (see Appendix 3). However, one should note that in the *Mus dunni* the Ks/Ka ratio was high (1.203) even though this sequence contains numerous stop codons, and therefore under no selection. It may be that the mutations that cause these stop codons have occurred more recently in *Mus dunni*, in which case the sequence may have been under selection until relatively recently. The finding that the *Fv1* ORF from the older mice shows signs of being under negative selection is at odds with the findings of Qi *et al*[315]. During their study, a large part of the *Fv1* ORF was cloned and sequenced from 20 mice samples. Analysis of the first 355 amino acids revealed a Ks/Ka of 0.48, a

value much lower than the trend seen in most of the *Fv1* ORFs in this study. This value, which is only marginally higher than for neutrally-drifting sequences, suggests that the *Fv1* ORF has not been under negative selection during the speciation of *Mus*. However, this figure presumably represents an average of the ratios between the sequences. Calculating the ratio in this way will be inherently inaccurate, as mutations accrued earlier will be present in more of the mice and will subsequently be weighted more than mutations unique to a sample. In this way, if more non-synonymous mutations occurred earlier, they would have a greater effect on the Ks/Ka ratio, and would lead to a much lower value.

In summary, the low level of stop codons introduced by mutation of, and the bias for synonymous mutation in, the *Fv1* ORF suggest, albeit rather inconclusively, that the *Fv1* ORF has been under negative selection during the speciation of *Mus*, prior to the evolution of 'true' *Fv1* activity. This implies that the *Fv1* allele seen in the more distantly-diverged mice has some activity that has been selected for at least from the time of the divergence of *Coelomys* some 5 million years ago, and possibly since its integration some 8-10 million years ago. Just what this activity may have been (and may still be?), is subject to speculation. It may be that the allele seen in the wild mice may restrict MLV, albeit to a lesser extent than the $Fv1^n$ and $Fv1^b$ alleles, and it may be this allele that is responsible for the low degree of innate innate restriction seen in these feral mice. This will be determined in the future using more sensitive *Fv1* assays with constructs containing *Fv1* from these mice (see the Future Work section). Another possibilty is that MLV is not the first retrovirus to be restricted by *Fv1*; it may be that the *Fv1* ORF in feral mice has been able to immunise mice against a retroviral invader in the distant past which is no longer seen today. In this scenario the *Fv1* ORF of feral mice would have been under selection for as long as this retrovirus infected mice, and selection would have ceased once the retrovirus became extinct, allowing the *Fv1* sequences from feral mice to accrue mutation naturally. This could account both for the maintenance of the intact *Fv1* ORF among the feral mice and the relatively low synonymous codon bias seen in these mice. A more unlikely explanation is that *Fv1* has another, as yet undetermined, function in the mouse, a role unrelated to retroviral restriction. However, the low Ks:Ka ratio seen in the feral mice would seem to argue against this.



The results of the *Fv1* phenotype assay using mutant constructs have important implications on the scheme of the evolution of the *Fv1* gene. The original sequencing of the two main alleles of *Fv1* highlighted the 3 diffences are shown in Figure 1.6, and these have been assumed to be required for *Fv1* phenotype. If this were true, it is difficult to see how the original sequence could have had activity after integration, and would only have gained activity after gaining the changes we see in the two alleles we see today. From the information gained from this study, namely that activity can be caused by the presence of a lysine or glutamic acid at just one residue (position 358), and that the gene probably did have one of these upon integration (namely the *Fv1*^{n}-associated lysine), it now seems possible that *Fv1* activity was acquired with the acquisition of this original sequence. However, the vast majority of the *Mus* genus have no activity, certainly all members outside the *Mus musculus* complex tested to date. The original *Fv1* activity, however, may not have been great, but would have provided the necessary selective advantage, under the conditions outlined in the scenario in Figure 7.5, for maintenance of this element. Subsequent selection for mutations that improve upon this activity would then have led to the changes we see today. The common changes mentioned earlier, observed between the *Mus musculus* complex *Fv1* ORF and the rest of the genus, may be important in this way. In most members of the *Mus* genus these changes have not occurred. This original, 'baseline' activity may still remain but current *Fv1* assays may not be sensitive enough to identify it clearly. However, feral mice have been observed to have an innate ability to restrict MLV to a small extent: could this be the original level of activity? Alternatively, it is important to note that in this project, only changes in the *Fv1* ORF have been studied; changes we see in phenotype have been equated to changes we see in the protein, and this appears to be true. However, a gene is nothing without the promoter. We know the integration event that gave rise to the gene did not provide *Fv1* with its own promoter, and the gene therefore uses a mouse promoter for expression. But it is likely that this promoter may not have been adequate originally: as mentioned previously, there is evidence to suggest *Fv1* activity may acutely sensitive to the level of *Fv1* expression. So, just as the *Fv1* ORF has been shown to have undergone great change to fine-tune its function, the same could well be true of the promoter too. It may be that most members of the *Mus* genus have a functional *Fv1* gene (to some degree) after all, but need some crucial changes in the promoter to realise

this potential. Future work, examining the promoter in the same way as the *Fv1* ORF has been, from mice from across the genus, will answer this.

7.5 Future work

The next step in the study of *Fv1* evolution would be to carry out a similar project as the sequencing project in Chapter 4, but based on phenotype. The *Fv1* ORFs cloned during the sequencing project should be typed for *Fv1* activity. In particular the MM3 construct; this *Fv1* ORF contains the changes most probably found in the original gene, so if this were found to have a measurable activity with MLV, it would make a stronger case for the maintenance of the *Fv1* ORF in older mice. Also, the assessment of phenotype for the *Fv1* ORF cloned from MOL and LG/J would be interesting: MOL appears to represent the intermediate in the formation of the $Fv1^b$ allele, being undeleted and having the Fv1b-specific residue at position 358 and the $Fv1^n$-associated residue at position 399. LG/J, on the other hand, has the $Fv1^n$-associated changes as well as both the changes associated with the $Fv1^d$ and $Fv1^{nr}$ alleles. By assessing the phenotype of these and the other *Fv1* ORFs from across the *Mus* genus, the effect on *Fv1* activity of the progression of changes that have occurred in this gene during the speciation of *Mus* could be assessed. This would provide information not only concerning the physical structure of the gene, which residues have what effects, but also would ascertain whether this sequence did have any measurable activity upon integration, and how exactly this activity was developed over time. However, the assay would have to be extremely sensitive to detect the slightest changes in phenotype, and for the reasons mentioned previously the current *Fv1* assay used for this work is probably not sensitive or robust enough. Fortunately, a new method of assessing phenotype involving the use of retroviral constructs expressing fluorescent proteins, quantified by FACS analysis, has been developed in this laboratory, and seems to be far superior to any other method (J. Stoye, pers. comm.). Using this method, it may be possible to determine the evolution of *Fv1* phenotype in mice.

References

1. Bernhard, W. Electron microscopy of tumor cells and tumor viruses: A review, *Cancer Res.* **18**, 491-509 (1958).

2. Telesnitsky, A. & Goff, S.P. Reverse transcriptase and the generation of retroviral DNA in *Retroviruses* (eds. Coffin, J.M., Hughes, S.H. & Varmus, H.E.) 121-160 (Cold Spring Harbor Laboratory Press, New York, 1997).

3. Coffin, J.M. Retroviridiae: the viruses and their replication, in *Fields Virology* (eds. Fields, B.N. & al, e.) 1767-1847 (Lippincott-Raven, 1996).

4. Keith, J. & Fraenkel-Conrat, H. Identification of the 5' end of Rous sarcoma virus RNA, *Proc. Natl. Acad. Sci.* **72**, 3347-3350 (1975).

5. Furuichi, Y., Shatkin, A.J., Stavnezer, E. & Bishop, J.M. Blocked, methylated 5' terminal sequence in avian sarcoma virus RNA, *Nature* **257**, 618-620 (1975).

6. Dimock, K. & Stoltzfus, C.M. Sequence specificity of internal methylation in B77 avian sarcoma virus RNA subunits, *Biochemistry* **16**, 471-488 (1977).

7. Beemon, K.L. & Keith, J. Localisation of N6-methyladenosine in the Rous sarcoma virus genome, *J. Mol. Biol.* **113**, 165-179 (1977).

8. Lai, M.M.C. & Duesberg, P.H. Adenylic acid-rich sequence in RNAs of Rous sarcoma virus and Rauscher mouse leukaemia virus, *Nature* **235**, 383-386 (1972).

9. Gillespie, D., Marshall, S. & Gallo, R.C. RNA of RNA tumour viruses contains poly A, *Nat. New Biol.* **236**, 227-231 (1972).

10. Atwater, J.A., Misdom, R. & Verma, I.M. Regulated mRNA stability, *Annu. Rev. Genet.* **24**, 519-41 (1990).

11. Vogt, V.M. Retroviral virions and genomes, in *Retroviruses* (eds. Coffin, J.M., Hughes, S.H. & Varmus, H.E.) 27-70 (Cold Spring Harbor Laboratory Press, New York, 1997).

12. Lerner, D.L., Wagaman, P.C., Philips, T.R., Prospero-Garcia, O., Henriksen, S.J., Fox, H.S., Bloom, F.E. & Elder, J.H.. Increased mutation frequency of FIV lacking functional deoxyuridine triphosphatase, *Proc. Natl. Acad. Sci.* **92**, 7480-4 (1995).

13. Bukrinsky, M.I., Sharova, N., McDonald, T.L., Pushkarskaya, T., Tarpley, W.G. & Stevenson, M.. Association of integrase, matrix, and reverse transcriptase antigens of human immunodeficiency type 1 with viral nucleic acids following acute infection, *Proc. Natl. Acad. Sci.* **90**, 6125-9 (1993).

14. Karageorgos, L., Li, P. & Burrell, C. Characterisation of HIV replication complexes early after cell-to-cell infection, *Aids Res. Hum. Retroviruses* **9**, 817-23 (1993).

15. Bukrinsky, M.I. & Haffar, O.K. HIV-1 nuclear import: matrix protein is back on center stage, this time together with Vpr, *Mol. Med.* **4**, 138-43 (1998).

16. Temin, H.M. Nature of the provirus of Rous sarcoma, *Natl. Cancer Inst. Monogr.* **17**, 557-570 (1964).

17. Baltimore, D. RNA-dependent DNA polymerase in virions of RNA tumour viruses, *Nature* **226**, 1209-11 (1970).

18. Temin, H.M. & Mizutani, S. RNA-dependent DNA polymerase in virions of Rous sarcoma virus, *Nature* **226**, 1211-3 (1970).

19. Harada, F., Sawyer, R.C. & Dahlberg, J.E. A primer ribonucleic acid for initiation of in vitro Rous sarcoma virus deoxyribonucleic acid synthesis, *J. Biol. Chem.* **250**, 3487-97 (1975).

20. Peters, G.F., Harada, F., Dahlberg, J.E., Panet, A., Haseltine, W.A. & Baltimore, D. Low-molecular weight RNAs of Moloney murine leukemia virus: Identification of the primer for RNA-directed DNA synthesis, *J. Virol.* **21**, 1031-41 (1977).

21. Harada, F., Peters, G.G. & Dahlberg, J.E. The primer tRNA for Moloney murine leukemia virus DNA syntheis: Nucleotide sequence and aminoacylation of $tRNA^{pro}$, *J. Biol. Chem.* **254**, 10979-85 (1979).

22. Peters, G.G. & Glover, C. tRNAs and priming of RNA-directed DNA synthesis in mouse mammary tumor virus, *J. Virol.* **35**, 31-40 (1980).

23. Jiang, M., Mak, J., Ladha, A., Cohen, E., Klein, M., Rovinski, B. & Kleiman, L.. Identification of tRNAs incorporated into wild-type and mutant human immunodeficiency virus type 1, *J. Virol.* **67**, 3246-53 (1993).

24. Ratner, L., Haseltine, W., Patarca, R., Livak, K.L., Starcich, B., Josephs, S., Duran, D.R., Rafalski, J.A., Whitehorn, E.A., Baumeister, K., Ifvanoff, L., Petteway, S.R., Pearson, M.L., Lautenberger, J.A., Papas, T.S., Ghreyeb, J., Chang, N.T., Gallo, R.C. & Wong-Staal, S. Complete nucleotide sequences of the AIDS virus, HTLV-III, *Nature* **313**, 277-84 (1985).

25. Coffin, J.M. Retroviridae and their replication, in *Virology (B.N.Fields and D.M.Knipe, eds). Raven Press, N.Y.*, pp 1437-1500 (1990).

26. Albritton, L.M., Tseng, L., Scadden, D. & Cunningham, J.M. A putative murine ecotropic receptor gene encodes a multiple membrane-spanning protein and confers susceptibility to virus infection, *Cell* **57**, 659-666 (1989).
27. Kozak, C.A. Susceptibility of wild mouse cells to exogenous infection with xenotropic leukemia viruses: control by a single dominant locus on chromosome 1 *J. Virol.* **55**, 690-5 (1985).
28. Best, S., Le Tissier, P.R. & Stoye, J.P. Endogenous retroviruses and the evolution of resistance to retroviral infections, *Trends in Microbiol.* **5**, 313-8 (1997).
29. Hartley, J.W., Wolford, N.K., Old, L.J. & Rowe, W.P. A new class of murine leukemia virus associated with the development of spontaneous lymphomas, *Proc. Natl. Acad. Sci. USA* **74**, 789-92 (1977).
30. Stoye, J.P., Moroni, C. & Coffin, J.M. Virological events leading to spontaneous AKR thymomas, *J. Virol.* **65**, 1273-1285 (1991).
31. Stoye, J.P. & Coffin, J.M. The four classes of endogenous murine leukemia viruses: structural relationships and potential for recombination, *J. Virol.* **61**, 2659-2669 (1987).
32. Tailor, C.S., Nouri, A., Lee, C.G., Kozak, C. & Kabat, D. Cloning and characterisation of a cell surface receptor for xenotropic and polytropic murine leukemia viruses, *Proc. Natl. Acad. Sci. USA* **96**, 927-932 (1999).
33. Yang, Y.-L., Lei, G., Xu, S., Holland, C.A., Kitamura, T., Hunter, K., Cunningham, J.M. The receptors for the polytropic and xenotropic murine leukemia viruses are encoded by a single gene that maps to *Rmc1*, *Nature Genetics* **21**, 216-9 (1999).
34. Battini, J.-L., Rasko, J.E.J. & Miller, A.D. A human cell-surface receptor for xenotropic and polytropic murine leukemia viruses: Possible role in G protein-coupled signal transduction, *Proc. Natl. Acad. Sci. USA* **96**, 1385-90 (1999).
35. Chien, M.L., Foster, J.L., Douglas, J.L. & Garcia, J.V. The amphotropic murine leukemia virus receptor gene encodes a 71-kilodalton protein that is induced by phosphate depletion, *J. Virol* **71**, 4564-70 (1997).
36. O'Neill, R.R., Hartley, J.W., Repaske, R. & Kozak, C.A. Amphotropic proviral envelope sequences are absent from the Mus germ line, *J. Virol* **61**, 2225-31 (1987).
37. Kim, J.W., Closs, E.I., Albritton, L.M. & Cunningham, J.M. Transport of cationic amino acids by the mouse ecotropic retrovirus receptor, *Nature* **352** (1991).

38. Wang, H., Kavanaugh, M.P., North, R.A. & Kabat, D. Cell-surface receptor for ecotropic murine retroviruses is a basic amino-acid transporter, *Nature* **352**, 729-731 (1991).

39. Miller, D.G., Edwards, R.H. & Miller, A.D. Cloning of the cellular receptor for amphotropic murine retroviruses reveals homology to that for gibbon ape leukaemia virus, *Proc. Natl. Acad. Sci. USA* **91**, 78-82 (1994).

40. van Zeijl, M., Johann, S.V., Closs, E., Cunningham, J., Eddy, R., Shows, T.B., o'Hara, B.. A human amphotropic retrovirus receptor is a second member of the gibbon ape leukaemia virus receptor family, *Proc. Natl. Acad. Sci. USA* **91**, 1168-1172 (1994).

41. Kavanaugh, M.P., Miller, D.G., Zhang, W., Law, W., Kozak, S.L., Kabat, D., Miller, A.D. Cell-surface receptors for gibbon ape leukemia virus and amphotropic murine retrovirus are inducible sodium-dependent phosphate symporters,. *Proceedings Of The National Academy Of Sciences Of The United States Of America* **91**, 7071-5 (1994).

42. Johann, S.V., Gibbons, J.J. & O'Hara, B. GLVR1, a receptor for gibbon ape leukemia virus, is homologous to a phosphate permease of Neurospora crassa and is expressed at high levels in the brain and thymus, *J. Virol* **66**, 1635-40 (1992).

43. Tailor, C.S., Takeuchi, Y., O'Hara, B., Johann, S.V., Weiss, R.A., Collins, M.K.. Mutation of amino acids within the gibbon ape leukemia virus (GALV) receptor differentially affects feline leukemia virus subgroup B, simian sarcoma-associated virus, and GALV infections, *J. Virol* **67**, 6737-41 (1993).

44. White, J., Kielian, M. & Helenius, A. Membrane fusion proteins of enveloped animal viruses, *Q. Rev. Biophysics* **16**, 151-95 (1983).

45. Bolander, F.F., Jr. Requirements for mouse mammary tumour virus internalisation in mouse epithelial cells, *J. Gen. Virol.* **77**, 793-796 (1996).

46. Brown, P.O., Bowerman, B., Varmus, H.E. & Bishop, Correct integration of retroviral DNA *in vitro*, J.M. *Cell* **49**, 347-356 (1987).

47. Bowerman, B., Brown, P.O., Bishop, J.M. & Varmus, H.E. A nucleoprotein complex mediates the integration of retroviral DNA, *Genes Dev.* **3**, 469-478 (1989).

48. Fassati, A. & Goff, S.P. Characterisation of intracellular reverse transcription complexes of Moloney Murine Leukemia Virus, *J. Virol.* **73**, 8919-8925 (1999).

49. Farnet, C.M. & Haseltine, W.A. Determination of proteins involved in the human immunodeficiency virus type 1 preintegration complex, *J. Virol.* **65**, 1910-5 (1991).
50. Keminer, O. & Peters, R. Permeability of single nuclear pores, *Biophysical Journal* **77**, 217-28 (1999).
51. Temin, H.M. & Rubin, H. Characteristics of an assay for Rous sarcoma virus and Rous sarcoma cells in tissue culture, *Virology* **6**, 669-688 (1958).
52. Roe, T., Reynolds, T.C., Yu, G. & Brown, P.O. Integration of murine leukemia virus DNA depends on mitosis, *The Embo J.* **12**, 2099-2108 (1993).
53. Hajihosseini, M., Ivachev, L. & Price J. Evidence that retroviruses integrate into post-replication host DNA, *EMBO J.* **12**, 4969-74 (1993).
54. Lewis, P.F. & Emerman, M. Passage through mitosis is required for oncoretroviruses but not for the human immunodeficiency virus, *J. Virol.* **68**, 510-6 (1998).
55. Efthymiadis, A., Briggs, L.J. & Jans, D.A. The HIV-1 Tat nuclear localisation sequence confers novel nuclear import properties, *J. Biol. Chem.* **273**, 1623-8 (1998).
56. Fincham, V.J. & Wyke, J.A. Differences between cellular integration sites of transcribed and nontranscribed Rous sarcoma proviruses, *J. Virol* **65**, 461-3 (1991).
57. Shoemaker, C., Goff, S., Gilboa, E., Paskind, M., Mitra, S.W. & Baltimore, D.. Structure of a cloned circular Moloney murine leukemia virus DNA molecule containing an inverted segment:implication for retrovirus integration, *Proc. Natl. Acad. Sci.* **77**, 3932-6 (1980).
58. Swanstrom, R., DeLorbe, W.J., Bishop, J.M. & Varmus, H.E. Nucleotide sequence of cloned unintegrated avian sarcoma virus DNA: viral DNA contains direct and inverted repeats similar to those in transposable elements, *Proc. Natl. Acad. Sci.* **78**, 124-8 (1981).
59. Shoemaker, C., Hoffman, J., Goff, S. & Baltimore, D. Intramolecular integration within Moloney murine leukemia virus DNA, *J. Virol.* **40**, 164-72 (1981).
60. Fujiwara, T. & Mizuuchi, K. Retroviral DNA integration: Structure of an integration intermediate, *Cell* **54**, 497-504 (1988).
61. Panganiban, A.T. & Temin, H.M. Circles with two tandem LTRs are precursors to integrated retrovirus DNA, *Cell* **36**, 673-9 (1984).

62. Brown, P.O., Bowerman, B., Varmus, H.E. & Bishop, J.M. Retroviral integration: Structure of the initial covalent product and its precursor, and a role for the viral IN protein, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 2525-2529 (1989).

63. Engelman, A., Mizuuchi, K. & Craigie, R. HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer, *Cell* **67**, 1211-1221 (1991).

64. Manley, N.R., O'Connell, M., Sun, W., Speck, N.A. & Hopkins, N. Two factors that bind to highly conserved sequences in mammalian type C retroviral enhancers, *J. Virol* **67**, 1967-75 (1993).

65. Yoshimura, F.K. & Diem, K. Characterisation of nuclear protein binding to a site in the long terminal repeat of a murine leukemia virus: comparison with the NFAT complex, *J. Virol.* **69**, 994-1000 (1995).

66. Adam, M.A. & Miller, A.D. Identification of a signal in a murine retrovirus that is sufficient for packaging of nonretroviral RNA into virions, *J. Virol* **62**, 3802-6 (1988).

67. Swain, A. & Coffin, J.M. Polyadenylation at correct sites in the genome RNA is not required for retrovirus replication or genome encapsidation, *J. Virol.* **63**, 3301-6 (1989).

68. Jamjoom, G.A., Naso, R.B. & Arlinghaus, R.B. Further characterisation of intracellular precursor polyproteins of Rauscher leukemia virus, *Virology* **78**, 11-34 (1977).

69. Yoshinaka, Y., Katoh, I., Copeland, T.D. & Oroszlan, S. Translational readthrough of an amber termination codon during synthesis of feline leukemia virus protease, *J. Virol.* **55**, 870-3 (1985).

70. Yoshinaka, Y., Katoh, I., Copeland, T.D. & Oroszlan, S. Murine leukemia virus protease is encoded by the gag-pol gene and is synthesised through suppression of an amber termination codon, *Proc. Natl. Acad. Sci.* **82**, 1618-22 (1985).

71. Krausslich, H.D. & Welker, R. Intracellular transport of retroviral capsid components, *Curr. Top. Microbiol. Immunol.* **214**, 25-63 (1996).

72. Shields, A., Witte, O.N., Rothenberg, E. & Baltimore, D. High frequency of aberrant expression of Moloney murine leukemia virus in clonal infections, *Cell* **14**, 601-9 (1978).

73. Sakalian, M., Wills, J.M. & Vogt, V.M. Efficiency and selectivity of RNA packaging by Rous sarcoma virus Gag deletion mutants, *J.Virol.* **68**, 5969-81 (1994).
74. Boeke, J.D. & Stoye, J.P. Retrotransposons, endogenous retroviruses, and the evolution of retroelements, in *Retroviruses* (eds. Coffin, J.M., Hughes, S.H. & Varmus, H.E.) 343-436 (Cold Spring Harbor Press, New York, 1997).
75. Gross, L. *Oncogenic viruses* (Pergamon Press, Oxford, United Kingdom, 1970).
76. Muhlbock, O. Note on a new inbred mouse strain GR/A, *Eur. J. Cancer* **1**, 123-4 (1955).
77. Lwoff, A. Tumor viruses and the cancer problem: a summation of the conference, *Cancer Res.* **20**, 820-9 (1960).
78. Dougherty, R.M., Di Stefano, H.S. & Roth, F.K. Virus particles and viral antigens in chicken tissues free of infectious avian leukosis virus, *Proceedings Of The National Academy Of Sciences Of The United States Of America* **58**, 808-17 (1967).
79. Vogt, P.K. A virus released by nonproducing Rous sarcoma cells, *Proceedings Of The National Academy Of Sciences Of The United States Of America* **58**, 801-8 (1967).
80. Hanafusa, H., Miyamoto, T. & Hanafusa, T. A cell-associated factor essential for formation of an infectious form of Rous sarcoma virus, *Proceedings Of The National Academy Of Sciences Of The United States Of America* **66**, 314-21 (1970).
81. Weiss, R.A. & Payne, L.N. The heritable nature of the factor in chicken cells which acts as a helper virus for Rous sarcoma virus, *Virology* **45**, 508-15 (1971).
82. Groudine, M., Eisenman, R. & Weintraub, H. Chromatin structure of endogenous retroviral genes and activation by an inhibitor of DNA methylation, *Nature* **292**, 311-7 (1981).
83. Levy, J.A., Kazan, P.L., Reilly, C.A. & Finkel, M.P. FBJ osteosarcoma virus in tissue culture. III. Isolation and characterization of non-virus-producing FBJ-transformed cells, *J. Virol* **26**, 11-5 (1978).
84. Pincus, T. The endogenous murine type C viruses, in *Molecular biology of RNA tumor viruses* (ed. Stephenson, J.) 77-130 (Academic Press, New York, 1980).
85. Rowe, W.P., Hartley, J.W., Lander, M.R., Pugh, W.E. & Teich, N. Noninfectious AKR mouse embryo cell lines in which each cell has the capacity to

be activated to produce infectious murine leukemia virus, *Virology* **46**, 866-76 (1971).

86. Weiss, R.A., Friis, R.R., Katz, E. & Vogt, P.K. Induction of avian tumor viruses in normal cells by physical and chemical carcinogens, *Virology* **46**, 920-38 (1971).

87. Lock, L.F., Keshet, E., Gilbert, D.J., Jenkins, N.A. & Copeland, N.G. Studies of the mechanism of spontaneous germline ecotropic provirus acquisition in mice, *Embo Journal* **7**, 4169-77 (1988).

88. Kozak, C.A. & O'Neill, R.R. Diverse mouse origins of xenotropic, mink cell focus-forming, and two types of ecotropic proviral genes, *J. Virol* **61**, 3082-3088 (1987).

89. Goodchild, N.L., Wilkinson, D.A. & Mager, D.L. Recent evolutionary expansion of a subfamily of RTVL-H human endogenous retrovirus-like elements, *Virology* **196**, 778-88 (1993).

90. Siracusa, L.D. The agouti gene: turned on to yellow [published erratum appears in Trends Genet 1995 Mar;11(3):116], *Trends In Genetics* **10**, 423-8 (1994).

91. Seperack, P.K., Mercer, J.A., Strobel, M.C., Copeland, N.G. & Jenkins, N.A. Retroviral sequences located within an intron of the dilute gene alter dilute expression in a tissue-specific manner, *Embo Journal* **14**, 2326-32 (1995).

92. Wilkinson, D.A., Mager, D.L. & Leong, J.-A. Endogenous human retroviruses, in *The retroviridae* (ed. Levy, J.A.) 465-534 (Plenum Press, New York, 1994).

93. Horn, T.M., Huebner, K., Croce, C. & Callahan, R. Chromosomal locations of members of a family of novel endogenous human retroviral genomes, *J. Virol* **58**, 955-9 (1986).

94. Cordonnier, A., Casella, J.-F. & Heidman, T. Isolation of novel human endogenous retrovirus-like elements with foamy virus-related *pol* sequence, *J. Virol* **69**, 5890-5897 (1995).

95. Benit, L., de Parseval, N., Casella, J. -F., Callebaut, I., Cordonnier, A. & Heidmann, T. Cloning of a new murine endogenous retrovirus, MuERV-L, with strong similarity to the human HERV-L element and a gag coding sequence closely related to the *Fv1* restriction gene, *J. Virol* **71**, 5652-5657 (1997).

96. Bonhomme, F. & Geunet, J.-L. The laboratory mouse and its wild relatives, *The Laboratory Mouse* , 1577-1597 (1989).

97. Ellerman, V. & Bang, O. Experimentelle Leukamie bei Huhnern, *Zentralbl. Bakteriol. Parisitenkd. Infectionskr. Hyg. Abt. Orig.* **46**, 595-609 (1908).
98. Rous, P. A sarcoma of the fowl transmissible by an agent separable from the tumor cells, *J. Exp. Med.* **13**, 397-411 (1911).
99. Bittner, J.J. Some possible effects of nursing on the mammary gland tumour incidence in mice, *Science* **84**, 162-162 (1936).
100. Gross, L. "Spontaneous" leukemia developing in C3H mice following inoculation in infancy, with AK-leukemic extracts, or AK-embryos, *Proc. Soc. Exp. Biol. Med.* **76**, 27-32 (1951).
101. Mayer, A., Duran-Reynals, M.L. & Lilly, F. *Fv-1* regulation of lymphoma development and of thymic ecotropic and xenotropic MuLV expression in mice of the AKR/J x RF/J cross, *Cell* **15**, 429-35 (1978).
102. Rowe, W.P. & Pincus, T. Quantitative studies of naturally occurring murine leukemia virus infection of AKR mice, *Journal Of Experimental Medicine* **135**, 429-36 (1972).
103. Lilly, F. & Pincus, T. Genetic control of murine viral leukaemogenesis, *Adv. Cancer. Res.* , 231-77 (1973).
104. Kung, H.J., Boerkoel, C. & Carter, T.H. Retroviral mutagenesis of cellular oncogenes: a review with insights into the mechanisms of insertional activation, *Current Topics In Microbiology And Immunology* **171**, 1-25 (1991).
105. Friend, C. Cell-free transmission in adult mice of a disease having the character of a leukaemia, *J. Exp. Med.* **105**, 307-318 (1957).
106. Teich, N., Wyke, J., Mak, T., Bernstein, A. & Hardy, W. Pathogenesis of retrovirus-induced disease, in *RNA tumour viruses* (eds. Weiss, R., Teich, N., Varmus, H. & Coffin, J.) 785-998 (Cold Spring Harbor Laboratory Press, New York, 1984).
107. Mirand, E.A., Steeves, R.A., Avila, L. & Grace, J.T. Spleen focus formation by polycythemic strains of Friend leukemia virus, *Proc. Soc. Exp. Biol. Med.* **127**, 900-4 (1968).
108. van Beveren, C., Coffin, J. & Hughes, S. Nucleotide sequences complemented with functional and structural analysis, in *RNA tumor viruses* (eds. Weiss, R., Teich, N., Varmus, H. & Coffin, J.M.) 567-1148 (Cold Spring Harbor Press, New York, Cold Spring Harbor, 1985).

109. Coffin, J. Structure of the retroviral genome, in *RNA tumor viruses* (eds. Weiss, R., Teich, N., Varmus, H. & Coffin, J.M.) 261-368 (Cold Spring Harbor Press, New York, Cold Spring Harbor, 1984).
110. Li, J.P., D'Andrea, A.D., Lodish, H.F. & Baltimore, D. Activation of cell growth by binding of Friend spleen focus-forming virus gp55 glycoprotein to the erythropoeitin receptor, *Nature* **343**, 762-4 (1990).
111. Constantinescu, S.S., Wu, H., Liu, X., Beyer, W., Fallon, A. & Lodish, H.F. The anemic Friend virus gp55 envelope protein induces erythroid differentiation in fetal liver colony-forming units-erythroid, *Blood* **91**, 1163-72 (1998).
112. Liu, R., Paxton, W.A., Choe, S., Ceradini, D., Martin, S.R., Horuk, R., MacDonald, M.E., Stuhlmann, H., Koup, R.A., Landau, N.R. Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection, *Cell* **86**, 367-77 (1996).
113. Marin, M., Tailor, C.S., Nouri, A., Kozak, S.L. & Kabat, D. Polymorphisms of the cell surface receptor control mouse susceptibilities to xenotropic and polytropic leukemia viruses, *J. Virol.* **73**, 9362-9368 (1999).
114. Suzuki, S. *Fv4*: A new gene affecting the splenomegaly induction by Friend leukaemia virus, *Jpn. J. Exp.* **45**, 473-78 (1975).
115. Gardner, M.B., Rasheed, S., Pal, B.K., Estes, J.D. & O'Brien, S.J. *Akvr-1*, a dominant murine leukaemia virus restriction gene, is polymorphic in leukaemia-prone wild mice, *Proc. Natl. Acad. Sci.* **77**, 531-35 (1980).
116. Odaka, T., Ikeda, H., Yoshikura, H., Moriwaki, K. & Suzuki, S. *Fv4*: Gene controlling resistance to NB-tropic Friend murine leukaemia virus, *J. Natl. Cancer Inst.* **67**, 1123-27 (1981).
117. Ikeda, H. & Odaka, T. Distribution in wild mice, introduction into genetic background of BALB/c mice, and mapping of chromosomes, *Virology* **133**, 65-76 (1984).
118. Hunter, E. Viral entry and receptors, in *Retroviruses* (eds. Coffin, J.M., Hughes, S.H. & Varmus, H.E.) 71-120 (Cold Spring Harbor Laboratory Press, New York, 1997).
119. Ikeda, H. & Sugimura, H. *Fv-4* resistance gene: a truncated endogenous murine leukemia virus with ecotropic interference properties. *J. Virol.* **63**, 5405-12 (1989).
120. Limjoco, T.I., Dickie, P., Ikeda, H. & Silver, J. Transgenic *Fv-4* mice resistant to Friend virus, *J. Virol.* **67**, 4163-4168 (1993).

122. Robinson, H.L., Astrin, S.M., Senior, A.M. & Salazar, S.H. Host susceptibility to endogenous viruses: Defective glycoprotein-expressing proviruses interfere with infections, *J. Virol.* **40**, 745-51 (1981).
123. McDougall, A.S., Terry, A., Tzavaras, T., Cheney, C., Rojko, J. & Neil, J.C.. Defective endogenous proviruses are expressed in feline lymphoid cells: evidence for a role in natural resistance to subgroup B feline leukemia viruses *J. Virol.* **68**, 2151-60 (1994).
124. Kurre, P., Kiem, H.-P.,Morris, J., Heyward, S., Battini, J.-L. & Miller, A.D.. Efficient transduction by an amphotropic retrovirus vector is dependent on high-level expression of the cell surface virus receptor, *J. Virol.* **73**, 495-500 (1999).
125. Hoffman, P.M. & Morse, H.C. Genetic determinants of neurological disease induced by Cas-Br-E murine leukemia virus, *J. Virol.* **53**, 40-3 (1985).
126. McAtee, F.J. & Portis, J.L. Monoclonal antibodies specific for wild mouse nuerotropic retrovirus: Detection of comparable levels of virus replication in mouse strains susceptible and resistant to paralytic disease, *J. Virol.* **56**, 1018-22 (1985).
127. Lilly, F. *Fv2*: Identification and location of a second gene governing the spleen focus response to Friend leukaemia virus in mice, *J. Natl. Cancer Inst.* **45**, 163-169 (1970).
128. Odaka, T. Inheritence of susceptibility to Friend mouse leukaemia virus, *Int. J. Cancer* **6**, 18-23 (1970).
129. Hoetlin, M.E., Kozak, S.L., Lilly, F., Chakraborti, A., Kozak, C.A. & Kabat, D.. Activation of erythropoeitin receptors by Friend virus gp55 and by erythropoietin and down-modulation by the *Fv-2*^{r} resistance gene, *Proc. Natl. Acad. Sci.* **87**, 9985-89 (1990).
130. Budarf, M., Huebner, K., Emanuel, B., Croce, C.M., Copeland, N.G., Jenkins, N.A. & D'Andrea, A.D.. Assignment of the erythropoietin receptor (EPOR) gene to mouse chromosome 9 and human chromosome 19, *Genomics* **9**, 575-8 (1990).
131. Youssoufian, H., Zon, L.I., Orkin, S.H., D'Andrea, A.D. & Lodish, H.F. Structure and transcription of the mouse erythropoietin receptor gene, *Mol. Cell Bio.* **10**, 3675-82 (1990).
132. Majumdar, M.K., Cho, C.-L., Fox, M.T., Eckner, K.L., Kozak, S., Kabat, D. & Geib, R.W. Mutations in the *env* gene of Friend spleen focus forming virus

overcome *Fv-2*^{r}-mediated resistance to Friend virus-induced erythroleukaemia, *J. Virol.* **66**, 3652-660 (1992).

133. Kozak, S.L., M.E. Hoetlin, Ferro, F.E., Majumdar, M.K., Geib, R.W. & Kabat, D. A Friend virus mutant that overcomes *Fv-2*^{r} resistance encodes a small glycoprotein that dimerises, is processed to the cell surfaces, and specifically activates erythropoeitin receptors, *J. Virol.* **67**, 2611-20 (1993).

134. Hoetlin, M.E., Ferro, F.E., Geib, R.W., Fox, M.T., Kozak, S.L. & Kabat, D. Deletions in one domain of the Friend virus-encoded membrane glycoprotein overcome host range restrictions for erythroleukemia, *J. Virol.* **69**, 856-63 (1995).

135. Persons, D.A., Paulsom, R.F., Loyd, M.R., Herley, M.T., Bodner, S.M., Bernstein, A., Correll, P.H. & Ney, P.A. *Fv2* encodes a truncated form of the Stk receptor tyrosine kinase, *Nat. Gen.* **23**, 159-165 (1999).

136. MacDonald, M.E., Mak, T.W. & Bernstein, A. Erythroleukemia induction by replication-competent type C viruses cloned from the anemia- and polycythemia-inducing isolates of Friend leukemia virus, *J. Exp. Med.* **151**, 1493-1503 (1980).

137. Chabot, B., Stephenson, D.A., Chapman, V.M., Besmer, P. & Bernstein, A. The proto-oncogene c-*kit* encoding a transmembrane tyrosine kinase maps to the mouse *W* locus, *Nature* **335**, 88-9 (1988).

138. Huang, E., Nocka, K., Beier, D.R., Chu. T.-Y., Buck, J., Lahm, W.-H., Wellner, D., Leder, P., Besmer, P.. The hematopoietic gromth factor KL is encoded by the *Sl* locus and is the ligand of the c-*kit* receptor, the gene product of the W locus, *Cell* **63**, 225-33 (1990).

139. Reth, M. .Immunodeficiency. Trapping the nude mouse gene,*Curr. Biol.* **5**, 18-20 (1995).

140. Mosier, D.E., Yetter, R.A. & Morse III, H.C. Retroviral induction of acute lymphoproliferative disease and profound immunosuppression in adult C57BL/6 mice, *J. Exp. Med.* **161**, 766-84 (1985).

141. Tsubura, A., Inaba, M., Imai, S., Murakami, A., Oyaizu, N., Yazumizu, R., Ohnishi, Y., Tanaka, H., Morii, S. & Ikehara, S. Intervention of T-cells in transportation of mouse mammary tumour virus (milk factor) to mammary gland cells in vivo, *Cancer Res.* **48**, 6555-59 (1988).

142. Prasad, G., Stoica, G. & Wong, P.K.Y. The role of the thymus in the pathogenesis of hind-limb paralysis induced by *ts1*, a mutant of Moloney murine leukemia virus-TB, *Virology* **169**, 332-40 (1989).

143. Oldstone, M.B.A., Lampert, P.W., Lee, S. & Dixon, F.J. Pathogenesis of the slow disease of the central nervous system associated with VM 1504 E virus, *Am. J. Clin. Pathol.* **88**, 193-206 (1977).

144. Tsukada, S., Rawlings, D.J. & Witte, O.N. Role of Bruton's tyrosine kinase in immunodeficiency, *Curr. Opin. Immunol.* **6**, 623-30 (1994).

145. Khan, W.N., Alt, F.W., Gerstein, R.M., Malynn, B.A., Larsson, I., Rathbun, G., Davidson, L., Muller, S., Kantor, A.B., Herzenberg, L.A., Rosen, F.S. & Sideras, P. Defective B cell development and function in BTK-deficient mice, *Immunity* **3**, 283-99 (1995).

146. Blunt, T., Finnie, N.J., Taccioli, G.E., Smith, G.C., Demengeot, G., Gottlieb, T.M., Mizuta, R., Varchese, A.J., Alt, F.W., Jeggo, P.A. & Jackson. S.P. Defective DNA-dependent protein kinase activity is linked to V(D)J recombination and DNA repair defects associated with the murine scid mutation, *Cell* **80**, 813-23 (1995).

147. Kirchgessner, C.U., Patil, C.K., Evans, J.W., Cuomo, C.A., Fried, L.M., Carter, T., Oettinger, M.A. & Brown, J.M. DNA-dependent kinase (p350) as a candidate gene for the SCID defect, *Science* **267**, 1178-83 (1995).

148. Custer, R.P., Bosma, G.C. & Bosma, M.J. Severe combined immunodeficiency (SCID) in the mouse. Pathology, reconstitution, neoplasms, *Am. J. Pathol.* **120**, 464-77 (1985).

149. Hasenkrug, K.J. & Chesebro, B. Immunity to retroviral infection: The Friend virus model, *Proc. Natl. Acad. Sci. USA* **94**, 7811-7816 (1997).

150. Chesebro, B. & Wehrly, K. *Rfv-1* and *Rfv-2*, two H2-associated genes that influence the recovery from Friend leukemia virus-induced splenomegaly, *J. Immunol.* **120**, 1081-85 (1978).

151. Miyazawa, M., Nishio, J., Wehrly, K. & Chesebro, B. Influence of MHC genes on spontaneous recovery from Friend retrovirus-induced leukemia, *J. Immunol.* **148**, 644-47 (1992).

152. Miyazawa, M., Nishio, J., Wehrly, K., Jay, G., Melvold, R.W. & Chesebro, B. Detailed mapping of the *Rfv1* genes that influences spontaneous recovery from Friend virus-induced leukemia, *Eur. J. Immunogenet.* **19**, 159-64 (1992).

153. Makino, M., Morse III, H.C., Fredrickson, T.N. & Hartley, J.W. H-2-associated and background gene influence the development of murine-retrovirus-induced immonodeficiency syndrome, *J. Immonol.* **144**, 4347-55 (1990).

154. Miyazawa, M., Nishio, J., Wehrly, K., David, C.S. & Chesebro, B. Spontaneous recovery from Friend retrovirus-induced leukemia. Mapping of the Rfv2 gene in the Q/TL region of the mouse MHC, *J. Immunol.* **148**, 1964-67 (1992).
155. Lonai, P. & Haran-Ghera, N. Resistance genes to murine leukemia in the I immune response gene region of the H-2 complex, *J. Exp. Med.* **141**, 882-89 (1977).
156. Hoffman, P.M., Robbins, D.S. & Morse, H.C. Role of immunity in age-related resistance to paralysis after murine leukemia virus infection, *J. Virol.* **52**, 734-38 (1984).
157. Hasenkrug, K.J., Valenzuela, A., Letts, V.A., Nishio, J., Chesebro, B. & Frankel, W.N. Chromosome mapping of *Rfv3*, a host resistance gene to Friend leukemia retrovirus, *J. Virol.* **69**, 2610-17 (1995).
158. Chesebro, B. & Wehrly, K. Identification of a non-H-2 gene (*Rfv-3*) influencing recovery from viremia amd leukemia induced by Friend virus complex, *Proc. Natl. Acad. Sci.* **76**, 425-29 (1979).
159. Shibuya, T., Niho, Y. & Mak, T.W. Erythroleukemia induction by Friend leukemia virus. A host gene locus controlling early anemia or polycythemia and the rate of proliferation of late erythroid cells, *J. Exp. Med.* **156**, 398-414 (1982).
160. Buller, R.S., Wehrly, K., Portis, J.L. & Chesebro, B. Host genes conferring a resistance to a central nervous system disease induced by a polytropic recombinant Friend murine retrovirus, *J. Virol.* **64**, 493-98 (1990).
161. Shibuya, T. & Mak, T.W. Host control of susceptibility to erythroleukemia and to the types of leukemia induced by Friend murine leukemia virus: initial and late stages, *Cell* **31**, 483-93 (1982).
162. Odaka, T. & Yamamoto, T. Inheritence of susceptibility to Friend mouse leukemia virus, *Japan J. Exp. Med.* **32**, 405-413 (1962).
163. Odaka, T. Inheritence of susceptibility to Friend mouse leukemia virus, *J. Virol.* **3**, 543-8 (1969).
164. Axelrad, A. & Steeves, R.A. Assay for Friend leukemia virus: rapid quantitative method based on enumeration of macroscopic spleen foci in mice, *J. Virol.* **24**, 513-8 (1964).
165. Lilly, F. Susceptibilty to two strains of Friend leukemia virus in mice, *Science* **155**, 461-2 (1967).
166. Hartley, J.W., Rowe, W.P. & Huebner, R.J. Host-range restrictions of murine leukemia viruses in mouse embryo cell cultures., *J. Virol.* **5**, 221-225 (1970).

167. Pincus, T., Hartley, J.W. & Rowe, W.P. A major genetic locus affecting resistance to infection with murine leukemia viruses. I. Tissue culture studies of naturally occuring viruses, *J. Exp. Med.* **133**, 1219-1233 (1971).
168. Pincus, T., Hartley, J.W. & Rowe, W.P. A major genetic locus affecting resistance to infection with murine leukemia viruses. II. apparent identity to a major locus described for resistance to Friend murine leukemia virus, *J. Exp. Med.* **133**, 1234-41 (1971).
169. Jolicoeur, P. The *Fv-1* gene of the mouse and its control of murine leukemia virus replication, *Current Topics in Microbiology and Immunology* **86**, 67-122 (1979).
170. Best, S., Tissier, P.L., Towers, G. & Stoye, J.P. Positional cloning of the mouse retrovirus restriction gene *Fv1*, *Nature* **382**, 826-829 (1996).
171. Frankel, W.N., Stoye, J.P., Taylor, B.A. & Coffin, J.M. Genetic analysis of endogenous xenotropic murine leukemia viruses: association with two common mouse mutations and the viral restriction locus *Fv-1*, *J. Virol.* **63**, 1763-1774 (1989).
172. Stoye, J.P., Kaushik, N., Jeremiah, S. & Best, S. Genetic map of the region surrounding the retrovirus restriction locus, *Fv1*, on mouse Chromosome 4., *Mammalian Genome.* **6**, 31-36 (1995).
173. Kozak, C.A. Analysis of wild-derived mice for the *Fv-1* and *Fv-2* murine leukemia virus restriction loci. A novel wild mouse *Fv-1* allele responsible for the lack of host range restriction, *J. Virol.* **55**, 281-285 (1985).
174. Todaro, G.J. & Green, H. Quantitative studies of the growth of the mouse embryo cells in culture and their development into established cell lines, *J. Cell Biology* **17**, 299-313 (1963).
175. Todaro, G.J. "Spontaneous" release of type C virus from clonal lines of "spontaneously" transformed Balb/3T3 cells, *Nature New Biology* **240**, 157-60 (1972).
176. Krontiris, T.G. Host restriction of Friend Leukemia virus. Role of the outer coat, *Proc. Natl. Acad. Sci. USA* **70**, 2549-2553 (1973).
177. Gisselbrecht, S., Bassin, R.H., Gerwin, B.I. & Rein, A. Dual susceptibility of a 3T3 mouse cell line to infection by N- and B-tropic murine leukemia virus: apparent lack of expression of the *Fv-1* gene, *Int. J. Cancer* **14**, 103-13 (1974).
178. Hartley, J.W. & Rowe, W.P. Clonal cell lines from a feral mouse embryo which lack host-range restrictions for murine leukemia viruses, *Virology* **65**, 128-34 (1975).

179. Steeves, R. & Lilly, F. Interactions between host and viral genomes in mouse leukemia, *Annual Review of Genetics* **11**, 277-296 (1977).
180. Kozak, C.A. & Chakraborti, A. Single amino acid changes in the Murine Leukemia virus capsid protein gene define the target of Fv1 resistance, *Virology* **225** (1996).
181. Huang, A.S., Besmer, P., Chu, L. & Baltimore, D. Growth of pseudotypes of Vesicular Stomatitis virus with N-tropic Murine Leukemia virus coats in cells resistant to N-tropic viruses, *J. Virol.* **12**, 659-662 (1973).
182. Sveda, M.M., Fields, B.N. & Soeiro, R. Host restriction of leukemia virus: fate of input virion RNA, *Cell* **2**, 271-77 (1974).
183. Hsu, I.-C., Yang, W.K., Tennant, R.W. & Brown, A. Transfection of *Fv-1* permissive and restrictive mouse cells with integrated DNA of murine leukemia viruses, *Proc. Natl. Acad. Sci. USA* **75**, 1451-5 (1978).
184. Chinsky, J., Soeiro, R. & Kopchick, J. *Fv-1* host cell restriction of Friend leukemia virus: microinjection of unintegrated viral DNA, *J. Virol.* **50**, 271-4 (1984).
185. Duran-Troise, G., Bassin, R.H., Wallace, B.F. & Rein, A. Balb/3T3 cells chronically infected with N-tropic murine leukemia virus continue to express *Fv-1^b* restriction., *Virology* **112**, 795-799 (1981).
186. Jolicoeur, P. & Baltimore, D. Effect of Fv-1 gene product on synthesis of N-tropic and B-tropic murine leukemia viral RNA, *Cell* **7**, 33-39 (1976).
187. Jolicoeur, P. & Baltimore, D. Effect of Fv-1 gene product on proviral DNA formation and integration in cells infected with murine leukemia viruses, *Proc. Natl. Acad. Sci. USA* **73**, 2236-2240 (1976).
188. Jolicoeur, P. & Rassart, E. Effect of Fv-1 gene product on synthesis of linear and supercoiled viral DNA in cells infected with murine leukemia virus, *J. Virol.* **33**, 183-195 (1980).
189. Yang, W.K., Kiggans, J.O., Yang, D.M., Ou, C.Y., Tennant, R.W., Brown, A. & R.H. Bassin Synthesis and circularization of N- and B-tropic retroviral DNA in Fv-1 permissive and restrictive mouse cells, *Proc. Natl. Acad. Sci. USA* **77**, 2994-2998 (1980).
190. Jolicoeur, P. & Rassart, E. Fate of unintegrated viral DNA in Fv-1 permissive and resistant mouse cells infected with mouse leukemia virus, *J. Virol.* **37**, 609-619 (1981).

191. Pryciak, P.M. & Varmus, H.E. *Fv-1* restriction and its effects on murine leukemia virus integration in vivo and in vitro, *J. Virol.* **66**, 5959-5966 (1992).
192. Hopkins, N., Schindler, J. & Hynes, R. Six NB-tropic leukemia viruses derived from a B-tropic virus of BALB/c have altered p30, *J. Virol.* **21**, 309-318 (1977).
193. Schindler, J., R. Hynes & Hopkins, N. Evidence for recombination between N-, and B-tropic murine leukemia viruses: analysis of three virion proteins by sodium dodecyl sulfate-polyacrylamide gel electrophoresis, *J. Virol.* **23**, 700-707 (1977).
194. Ou, C.-Y., Boone, L.R., Koh, C.-K., Tennant, R.W. & Yang, W.K. Nucleotide sequences of gag-pol regions that determine the host range property of BALB/c N-tropic and B-tropic murine leukemia viruses, *J. Virol.* **48**, 779-84 (1983).
195. DesGroseillers, L. & Jolicoeur, P. Physical mapping of the Fv-1 tropism host range determinant of BALB/c murine leukemia viruses, *J. Virol.* **48**, 685-696 (1983).
196. Kirchner, J., Connolly, C.M. & Sandmeyer, S.B. Requirement of RNA polymerase III transcription factors for in vitro position-specific integration of a retroviruslike element, *Science* **267**, 1488-91 (1998).
197. Morozov, A., Yung, E. & Kalpana, G.V. Structure-function analysis of integrase interactor 1/hSNF5L1 reveals differential properties of two repeat motifs present in the highly conserved region, *Proc. Natl. Acad. Sci. USA* **95**, 1120-5 (1998).
198. Lee, M.S. & Craigie, R. Protection of retroviral DNA from autointegration: involvement of a cellular factor, *Proc. Natl. Acad. Sci. USA* **91**, 9823-7 (1994).
199. Wei, S.-Q., Mizuuchi, K. & Craigie, R. A large nucleoprotein assembly at the ends of the viral DNA mediates retroviral DNA integration, *EMBO J.* **16**, 7511-20 (1997).
200. Li, L., Farnet, C.M., Anderson, W.F. & Bushman, F.D. Modulation of activity of Moloney murine leukemia virus preintegration complexes by host factors *in vitro*, *J. Virol.* **72**, 2125-31 (1998).
201. Lee, M.S. & Craigie, R. A previously unidentified host protein protects retroviral DNA from autointegration, *Proc. Natl. Acad. Sci. USA* **95**, 1528-33 (1998).

202. Chen, H. & Engelman, A. The barrier-to-autointegration protein is a host factor for HIV type I integration, *Proc. Natl. Acad. Sci. USA* **95**, 15270-4 (1998).
203. Farnet, C.M. & Bushman, F.D. HIV-1 cDNA integration: requirement of HMG I(Y) protein for function of preintegration complexes *in vitro*, *Cell* **88**, 483-92 (1997).
204. Cai, M., Huang, Y., Zheng, R., Wei, S.Q., Gherlando, R., Lee, M.S., Craigie, R., Gronenborn, A.M. & Chloe, G.M. Solution structure of the cellular factor BAF responsible for retroviral DNA from autointegration, *Nat. Struct. Biol.* **5**, 903-9 (1998).
205. Yee, T., Furuichi, T., Inouye, S. & Inouye, M. Multy-copy single-stranded DNA isolated from a gram negative bacterium, *Myxococcus xanthus*, *Cell* **38**, 203-209 (1984).
206. Gilbert, W. The RNA world, *Nature* **319**, 618 (1986).
207. Westheimer, F.H. Polyribonucleic acids as enzymes [news], *Nature* **319**, 534 (1986).
208. Hirao, I. & Ellington, A.D. Re-creating the RNA world, *Current Biology* **5**, 1017-22 (1995).
209. Temin, H. The DNA provirus hypothesis, *Science* **192**, 1075-80 (1976).
210. Lampson, B.C., Inouye, S. & Inouye, M. msDNA of bacteria, *Prog. Nucleic Acid Res. Mol. Biol.* **40**, 1-24 (1991).
211. Temin, H.M. Reverse transcriptases. Retrons in bacteria, *Nature* **339**, 254-5 (1989).
212. Inouye, M. & Inouye, S. Retrons and multicopy single-stranded DNA, *J. Bacteriol.* **174**, 2419-24 (1992).
213. Kim, A., Terzian, C., Santamoria, C., Pelisson, A., Prud-homme, B. & Buchton, A. Retroviruses in invertebrates: the gypsy retrotransposon is apparently an infectious retrovirus of Drosophila melanogaster, *Proc. Natl. Acad. Sci. USA* 91, 1285-9 (1994).
214. Xiong, Y. & Eickbusch, T.H. Origin and evolution of retroelements based upon their reverse transcriptase sequences, *EMBO J.* **9**, 3353-62 (1990).
215. Danilevskaya, K.L., Arkhipova, I.R., Traverse, K.L. & Pardue, M.L. Promoting in tandem: the promotor for telomere transposon Het-A and implications for the evolution of retroviral LTRs, *Cell* **88**, 647-55 (1997).
216. Smit, A.F.A. Interspersed repeats and other mementos of transposable elements in mammalian genomes, *Curr. Opin. Genet. Devel.* **6**, 734-48 (1996).

217. Doolittle, W.F. & Sapienza, C. Selfish genes, the phenotype paradigm and genome evolution, *Nature* **284**, 601-3 (1980).
218. Orgel, L.E. & Crick, F.H.C. Selfish DNA: the ultimate parasite, *Nature* **284**, 604-607 (1980).
219. Lehrman, M.A., Schneider, W.J., Sudhof, T.C., Brown, M.S., Goldstein, J.L. & Russell, D.W.. Mutation in LDL receptor: Alu-Alu recombination deletes exons encoding transmembrane and cytoplasmic domains, *Science* **227**, 140-6 (1985).
220. Muratani, K., Hada, T., Yamamoto, Y., Kaneko, T., Shigeto, Y., Ohue, T., Furuyami, J. & Higashino, K. Inactivation of the cholinesterase gene by Alu insertion: possible mechanism for human gene transposition, *Proc. Natl. Acad. Sci. USA* **83**, 11315-19 (1991).
221. Weiner, A.M., Deninger, P.L. & Efstratiadis, A. Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information, *Annu. Rev. Biochem.* **55**, 631-661 (1986).
222. Rogers, J. *Int. Rev. Cytol.* **93**, 187-279 (1985).
223. Piechaczyk, M., Blanchard, J.M., Riaad-El Sabouty, S., Dani, C., Marty, L. & Jeanteur, P. Unusual abundance of vertebrate 3-phosphate dehydrogenase pseudogenes, *Nature* **312**, 469-71 (1984).
224. Deininger, P.L., Batzer, M.A., Hutchison, C.A.. & Edgell, M.H. Master genes in mammalian repetitive DNA amplification, *Trends In Genetics* **8**, 307-11 (1992).
225. Denininger, P.L. & Batzer, M.A. The impact of short interspresed elements (SINEs) on the host genome, (ed. Maraia, R.) 43-60 (R.G. Landes, Austin, TX, 1995).
226. Deininger, P.L., Tiedge, H., Kim, J. & Brosius, J. Evolution, expression, and possible function of a master gene for amplification of an interspersed repeated DNA family in rodents, *Progress In Nucleic Acid Research And Molecular Biology* **52**, 67-88 (1996).
227. Schmid, C. & Maraia, R. Transcriptional regulation and transpositional selection of active SINE sequences, *Current Opinion In Genetics And Development* **2**, 874-82 (1992).
228. Gould, S.J. & Vrba, E. Exaption- a missing term in the science of form*Paleobiology* **8**, 4-15 (1982).

229. Parma, J., Christophe, D., Pohl, V. & Vassart, G. Structural organization of the 5' region of the thyroglobulin gene. Evidence for intron loss and "exonization" during evolution, *J. Mol. Biol.* **196**, 769-779 (1987).
230. Keese, P.K. & Gibbs, A. Origins of genes: "big bang" or continuous creation? *Proc. Natl. Acad. Sci. USA* **89**, 9849-9893 (1992).
231. White, S.H. & Jacobs, R.E. The evolution of proteins from random amino acid sequences. I. Evidence from the lengthwise distribution of amino acids in modern protein sequences, *J. Mol. Evol.* **36**, 79-95 (1993).
232. White, S.H. The evolution of proteins from random amino acid sequences: II. Evidence from the statistical distributions of the lengths of modern protein sequences, *J. Mol. Evol.* **38**, 383-394 (1994).
233. Makalowski, W., Mitchell, G.A. & Labuda, D. *Alu* sequences in the coding regions of mRNA: a source of protein variability, *Trends Genet.* **10**, 188-93 (1994).
234. Makalowski, W. The impact of short interspersed elements (SINEs) on the host genome (ed. Maraia, R.), 81-104 (R.G. Landes, Austin, TX, 1995).
235. Ivanov, V.A. & Il'in, Y.V. Reverse transcriptases and their biological role (a review), *Molecular Biology* **29**, 143-156 (1995).
236. Brosius, J. & Tiedge, H. Reverse transcriptase: mediator of genomic plasticity, *Virus genes* **11**, 163-79 (1996).
237. Marshall, C.R., Raff, E.C. & Raff, R.A. Dollo's law and the death and resurrection of genes, *Proc. Natl. Acad. Sci. USA* **91**, 12282-87 (1994).
238. Kim, J.H., Yu, C.-Y., Bailey, A., Hardison, R. & Shen, C.-K.J. Unique sequence organization and erythroid cell-specific nuclear factor-binding of mammalian theta 1 globin promoters, *Nucleic Acids Res* **17**, 5687-700 (1989).
239. Baltimore, D. Retroviruses and retrotransposons: the role of reverse transcription in shaping the eukaryotic genome, *Cell* **40**, 481-2 (1985).
240. Derr, L.K. & Strathern, J.N. A role for reverse transcripts in gene conversion, *Nature* **361**, 170-3 (1993).
241. Fink, G.R. Pseudogenes in yeast? *Cell* **49**, 5-6 (1987).
242. Humphrey, G.W. The impact of short interspresed elements (SINEs) on the host genome (ed. Maraia, R.), 195-222 (R.G. Landes, Austin, TX, 1995).
243. Schmid, C.W. & Rubin, C.M. The impact of short interspresed elements (SINEs) on the host genome (ed. Maraia, R.), 105-23 (R.G. Landes, Austin, TX, 1995).

244. Chesnokov, I.N. & Schmid, C.W. Specific Alu binding protein from human sperm chromatin prevents DNA methylation, *J. Biol. Chem.* **270**, 18539-42 (1995).
245. Englander, E.W., Wolffe, A.P. & Howard, B.H. Nucleosome interactions with a human Alu element. Transcriptional repression and effects of template methylation, *J. Biol. Chem.* **268**, 19565-73 (1993).
246. Englander, E.W. & Howard, B.H. Nucleosome positioning by human Alu elements in chromatin, *J. Biol. Chem.* **270**, 10091-6 (1995).
247. Kaplan, F.S., Murray, J., Sylvester, J.E., Gonsales, I. L., O'Connor, J.P., Doering, J.L., Muenke, M., Emanuel, B.L. & Zasloff, M.A. The topographic organization of repetitive DNA in the human nucleolus, *Genomics* **15**, 123-32 (1993).
248. Howard, B.H., Russanova, V.R. & Englander, E.W. (The impact of short interspersed elements (SINEs) on the host genome (ed. Maraia, R.), 133-41 (R.G. Landes, Austin, TX, 1995).
249. Pardue, M.L., Olga, N., Danilevskaya, K.L., Slot, F. & Traverse, K.L. Drosophila telomeres: new views on chromosome evolution, *Trends in Genetics* **12**, 48-51 (1996).
250. Schwarz, E. & Schwarz, H.K. The wild and commensal stocks of the house mouse, *Mus musculus* Linnaeus, *J. Mammal.* **24**, 59-72 (1943).
251. Bonhomme, F., Catalan, J., Britton-Davidian, J., Chapman, V.M., Moriwaki, K., Nevo, E. & Thaler, L. Biochemical diversity and evolution in the genus *Mus*, *Biochemical Genetics* **22**, 275-303 (1984).
252. Sage, R.D. Wild mic, in *The Mouse in Biomedical Research* (eds. Foster, H.L., Small, J.D. & Fox, J.G.) 39-90 (Academic, New York, 1981).
253. Jouvin-Marche, E., Cuddihy, A., Butler, S., Hansen, J.N., Fitch, W.M. & Rudikoff, S. Modern evolution of a single-copy gene: the immunoglobulin C kappa locus in wild mice, *Molecular Biology And Evolution* **5**, 500-11 (1988).
254. Fort, P., Bonhomme, F., Darlu, P., Piechaczyk, M., Jeanteur, P. & Thaler, L. Clonal divergence of mitochondrial DNA versus populational evolution of nuclear genome, *Evol. Theor.* **7**, 81-90 (1984).
255. She, J.X., Bonhomme, F., Boursot, P., Thaler, L. & Catzeflis, F. Molecular phylogenies of the genus Mus: comparative analysis of electophoretic, scnDNA hybridisation, and mtDNA RFLP data, *Biol. J. Linn. Soc.* **41**, 83-103 (1990).

256. Auffray, J.-C., Vanlerberghe, F. & Britton-Davidian, J. House mouse progression in Eurasia: a palaeontological and archaeozoological approach, *Biol. J. Linn. Soc.* **41**, 13-25 (1990).

257. Moriwaki, K., Miyashita, N. & Yonekawa, H. Genetic survey of the origin of laboratory mice and its implication in genetic monitoring, in *The Contribution of Laboratory Animal Science to the Welfare of Man and Animals* (eds. Archibold, J., Ditchfield, J. & Rowsell, H.C.) 237-247 (Gustav Fisher Verlag, Stuttgart, 1985).

258. Moriwaki, K., Miyashita, N., Suzuki, H., Kurihara, Y. & Yonekawa, H. Genetic features of major geographical isolates of *Mus musculus*, in *The Wild Mouse in Immunology* (eds. Potter, M., Nadeau, J.H. & Cancro, M.P.) 55-61 (Springer-Verlag, Berlin, 1986).

259. Karn, R.C. & Dlouhy, S.R. Salivary androgen-binding protein variation in *Mus* and other rodents, *J. Heredity* **82**, 453-458 (1991).

260. Miyashita, N., Moriwaki, K., Minezawa, M., Yonekawa, H., Bonhomme, F., Migita, S., Yu, Z.C., Lu, D.Y., Cho, W.S. & Thohari, M. Allelic constitution of the hemoglobin beta chain in wild populations of the house mouse, Mus musculus, *Biochemical Genetics* **23**, 975-86 (1985).

261. Watanabe, T., Miyashita, N., Moriwaki, K. & Hilgers, J. Evolutionary relationships between laboratory mice and subspecies of *Mus musculus* based on the genetic study of pancreatic proteinase loci, *Prt-1*, *Prt-2*, *Prt-3*, and *Prt-6*, *Biochemical Genetics* **25**, 239-51 (1987).

262. Nobuhara, H., Kuida, K., Furutani, M., Shiroishi, T., Moriwaki, K., Yanagi, Y. & Tada, T.Polymorphism of T-cell receptor genes among laboratory and wild mice: diverse origins of laboratory mice, *Immunogenetics* **30**, 405-13 (1989).

263. Suzuki, H., Miyashita, N., Moriwaki, K., Kominami, R., Muramatsu, M., Kanehisa, T., Bonhomme, F., Petras, M.L., Yu, Z.C. & Lu, D.Y. Evolutionary implication of heterogeneity of the nontranscribed spacer region of ribosomal DNA repeating units in various subspecies of *Mus musculus*, *Molecular Biology And Evolution* **3**, 126-37 (1986).

264. Cazenave, P.A., Marche, P.N., Jouvin-Marche, E., Voegtlé, D., Bonhomme, F., Bandeira, A. & Coutinho, A. V beta 17 gene polymorphism in wild-derived mouse strains: two amino acid substitutions in the V beta 17 region greatly alter T cell receptor specificity, *Cell* **63**, 717-28 (1990).

265. Morgado, M.G., Jouvin-Marche, E., Gris-Liebe, C., Bonhomme, F., Anand, R., Talwar, G.P. & Cazenave, P.A. Restriction fragment length polymorphism

and evolution of the mouse immunoglobulin constant region gamma loci, *Immunogenetics* **38**, 184-92 (1993).
266. Tutter, A. & Riblet, R. Evolution of the immunoglobulin heavy chain variable region (Igh-V) locus in the genus *Mus*, *Immunogenetics* **30**, 315-29 (1989).
267. Redi, C.A., Garagna, S., Della Valle, G., Bottiroli, G., Dell'Orto, P., Viale, G., Peverali, F.A., Raimondi, E. & Forejt, J. Differences in the organization and chromosomal allocation of satellite DNA between the European long tailed house mice *Mus domesticus* and *Mus musculus*, *Chromosoma* **99**, 11-7 (1990).
268. Bishop, C.E., Boursot, P., Baron, B., Bonhomme, F. & Hatat, D. Most classical *Mus musculus domesticus* laboratory mouse strains carry a *Mus musculus domesticus* Y chromosome, *Nature* **315**, 70-72 (1985).
269. Nagamine, C.M., Y. Nishioka, K. Moriwaki, P. Boursot, F. Bonhomme & Lau, Y. F. C. The musculus-type Y chromosome of the laboratory mouse is of Asian origin, *Mamm. Genome* **3**, 84-91 (1992).
270. Tucker, P.K., Lee, B.K. & Eicher, E.M. Y chromosome evolution in the subgenus *Mus* (genus *Mus*), *Genetics* **122**, 169-79 (1989).
271. Tucker, P.K., Lee, B.K., Lundrigan, B.L. & Eicher, E.M. Geographic origin of the Y chromosomes in old inbred strains of mice, *Mammalian Genome* **3**, 254-61 (1992).
272a. Yonekawa, H., Moriwaki, K., Gotoh, O., Hayashi, J.I., Watanabe, J., Miyashita, N., Petras, M.L. & Tagashira, Y. Evolutionary relationships among five subspecies of Mus musculus based on restriction enzyme cleavage patterns of mitochondrial DNA, *Genetics* **98**, 801-16 (1981).
272b. Yonekawa, H., Moriwaki, K., Gohoh, O., Miyashita, N., Matsushima, Y., Shi, L., Cho, W.S., Zhen, X. L. & Tagashira Y. Hybrid origin of Japanese mice *Mus musculus molissinus*: evidence from restriction analysis and mitochondrial DNA, *Mol. Biol. Evol.* **5**, 63-78 (1988).
272c. Yonekawa, H., Moriwaki, K., Gotoh, O., Miyashita, N. & Migita, S. Origin of laboratory mice deduced from restriction patterns of mitochondrial DNA, *Differentiation* **22**, 222-226 (1982).
272d. Yonekawa, H., Moriwaki, K., Gotoh, O., Watanabe, J. & Hayashi, J.-I. Relationship between laboratory mice and the subspecies *Mus musculus domesticus* based on restriction endonuclease cleavage patterns of mitochondrial DNA, *Jpn. J. Genet.* **55**, 289-296 (1980).

273. Lerner, T.L. & Hanafusa, H. DNA sequence of the Bryan high-titer strain of Rous sarcoma virus: extent of *env* deletion and possible genealogical relationship with other viral strains, *J. Virol* **49**, 549-56 (1984).
274. Boone, L.R., F.E. Myer, D.M. Yang, C.-Y. Ou, C.K. Koh, L.E. Roberson, R.W. Tennant & Yang, W.K. Reversal of *Fv-1* host range by in vitro restriction endonuclease fragment exchange between molecular clones of N-tropic and B-tropic murine leukemia virus genomes, *J. Virol.* **48**, 110-119 (1983).
275. Cosset, F.L., Morling, F.J., Takeuchi, Y., Weiss, R.A., Collins, M.K. & Russell, S.J. Retroviral retargeting by envelopes expressing an N-terminal binding domain, *J. Virol* **69**, 6314-22 (1995).
276. Bonhomme, F. Evolutionary relationships in the Genus *Mus*, *Current Topics in Microbiology and Immunology* **127**, 19-34 (1986).
277. Frisman, L.V., Korobitsina, K.V., Yakimenko, L.V., Bokshtein, F.M. & Muntyanu, A.I. Genetic differentiation of the U.S.S.R. house mice: electrophoretic study of proteins, *Biol. J. Linn. Soc.* **41**, 65-72 (1990).
278. Britton-Davidian, J. Genetic differentiation in *M. m. domesticus* populations from Europe, the Middle East and North Africa: geographic patterns and colonization events, *Biol. J. Linn. Soc.* **41**, 27-45 (1990).
279. Boursot, P., Auffray, J.-C., Britton-Davidian, J. & Bonhomme, F. The evolution of house mice, *Annu. Rev. Ecol. Syst.* **24.**, 119-52 (1993).
280. Ferris, S.D., Sage, R.D., Huang, C.M., Nielsen, J.T., Ritte, U. & Wilson, A.C. Flow of mitochondrial DNA across a species boundary, *Proceedings Of The National Academy Of Sciences Of The United States Of America* **80**, 2290-4 (1983).
281. Vanlerberghe, F., Boursot, P., Nielsen, J.T. & Bonhomme, F. A steep cline for mitochondrial DNA in Danish mice, *Genetical Research* **52**, 185-93 (1988).
282. Vanlerberghe, F., Dod, B., Boursot, P., Bellis, M. & Bonhomme, F. Absence of Y-chromosome introgression across the hybrid zone between *Mus musculus domesticus* and *Mus musculus musculus*, *Genetical Research* **48**, 191-7 (1986).
283. Vanlerberghe, F., Boursot, P., Catalan, J., Gerasimov, S., Bonhomme, F., Botev, B.A. & Thaler, L. [Genetic analysis of the hybridization zone between two subspecies *Mus musculus domesticus* and *Mus musculus musculus* in Bulgaria], *Genome* **30**, 427-37 (1988).
284. Bousot, P., Bonhomme, F., Britton-Davidian, J., Catalan, J. & Yonekawa, H. Introgression differentielle des genomes nucleaires et mitochondriaux chez deux semi-especes de souris, *C. R. Acad. Sci. Paris* **299**, 365-370 (1984).

285. Klein, J., Tichy, H. & Figueroa, F. On the origin of mice, *Ann. Univ. Chile* **5**, 91-120 (1987).
286. Moulia, C., Aussel, J.-P., Bonhomme, F., Boursot, P., Nielson, J. T. & Renaud, F. Wormy mice in a hybrid zone: a genetic control of susceptibility to parasitic infection, *J. Evol. Biol.* **4**, 679-687 (1991).
287. Nallaseth, F.S. Sequence instability and functional inactivation of murine Y chromosomes can occur on a specific genetic background, *Molecular Biology And Evolution* **9**, 331-65 (1992).
288. Auffray, J.-C., Tchernov, E. & Nevo, E. Origine du commensalisme de la souris domestique (*Mus musculus domesticus*) vis-a-vis de l'homme, *C. R. Acad. Sci. Paris* **307**, 517-522 (1988).
289. Ferris, S.D., Sage, R.D. & Wilson, A.C. Evidence from mtDNA sequences that common laboratory strains of inbred mice are descended from a single female, *Nature* **295**, 163-65 (1982).
290. Fitch, W.M. & Atchley, W.R. Evolution in inbred strains of mice appears rapid, *Science* **228**, 1169-75 (1985).
291. Sambrook, F. & Maniatis *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbour Laboratories, Cold Spring Harbour NY, USA, 1989).
292. Anand, R., Riley, J.H., Butler, R., Smith, J.C. & Markham, A.F. A 3.5 genome equivalent multi access YAC library: construction, characterisation, screening and storage, *Nucl. Acid Res.* **18**, 1951-1955 (1990).
293. Larin, Z., Monaco, A.P. & Lehrach, H. Yeast artificial chromosome libraries containing large inserts from mouse and human DNA, *Proc. Natl. Acad. Sci. USA* **88**, 4123-4127 (1991).
294. Birnboim, H.C. & Doly, J. A rapid alkaline extraction procedure for screening recombinant plasmid DNA, *Nucleic Acids Res.* **7**, 1513 (1979).
295. Reddy, P. & McKenny, K. Improved method for the production of M13 Phage and single-stranded DNA for DNA sequencing, *Biotechniques* **20**, 854-860 (1996).
296. Chomczynski, P. & Sacchi, N. Single-step method of RNA isolation by guanidinium thiocyanate-phenol-chloroform extraction, *Anal. Biochem.* **162**, 156-159 (1986).
297. Church, G.M. & Gilbert, W. Genomic sequencing, *Proc. Natl. Acad. Sci. USA* **81**, 1991 (1984).

298. Morgan, B.A., Conlan, F.L., Manzanares, M., Milar, J.B.A., Kanuga, N., Sharpe. J., Krumlauf, R., Smith, J.C., Stevens, G. & Sedgewick, S.G. Transposon tools for recombinant DNA manipulation: Characterisation of transcriptional regulators from yeast, xenopus and mouse, *Proc. Natl. Acad. Sci.* **93**, 2801-2806 (1996).

299. Felsenstein, J. PHYLIP- phylogeny inference package, version 3.4, *Cladistics* **5**, 164-166 (1989).

300. Yang, Z. Phylogenetic analysis by maximum likelihood (PAML). , version 1.2 edition. (Department of Integrative Biology, University of California at Berkeley, Berkeley, 1996).

301. DeBry, R.W. & Seldin, M.F. Human/mouse homology relationships, *Genomics* **33**, 337-351 (1996).

302. Mullins, J.J., Zeng, Q. & Gross, K.W. Mapping of the mouse atrial natriuretic factor gene: evidence for tight linkage to the *Fv1* locus, *Hypertension* **9**, 518-521 (1987).

303. Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees, *Mol. Biol. Evol.* **4**, 406-425 (1987).

304. Kishino, H. & Hasegawa, M. Evaluation of the maximum-likelihood estimate of the evolutionary tree topologies from DNA-sequence data, and the branching order in hominoidea, *J. Mol. Evol.* **29**, 170-179 (1989).

305. Jolicoeur, P. & Baltimore, D. Effect of the *Fv-1* locus on the titration of murine leukemia viruses, *J. Virol.* **16**, 1593-1598 (1975).

306. Endo, T., Ikeo, K. & Gojobori, T. Large-scale search for genes on which positive selection may operate, *Mol. Biol. Evol.* **13**, 685-690 (1996).

307. Altschul, S.F., W., G., W, M., E.W., M. & D.J., L. Basic local alignment search tool, *J. Mol. Biol.* **215**, 403-410 (1990).

308. Pincus, T., Hartley, J.W. & Rowe, W.P. A major genetic locus affecting resistance to infection with murine leukemia viruses. IV. Dose-response relationships in *Fv-1*-sensitive and resistant cell cultures, *Virology* **65**, 333-342 (1975).

309. Goff, S.P. Operating under a Gag order: a block against incoming virus by the *Fv1* gene, *Cell* **86**, 691-3 (1996).

310. Berry, R.J. Genetical processes in wild mouse populations, *Current Topics in Microbiology and Immunology* **127**, 86-94 (1986).

311. Bonhomme, F., Miyashita, N., Boursot, P., Catalan, N. & Moriwaki, K. Genetical variation and polyphyletic origin in Japanese *Mus musculus*, *Heredity* **63**, 299-308 (1989).
312. Gardner, M.B., Kozak, C.A. & O'Brien, S.J. The Lake Casitas wild mouse: evolving genetic resistance to retroviral disease, *Trends Genet.* **7**, 22-27 (1991).
313. Wheeler, L.L. & Selander, R.K. Genetic variations in populations of the house mouse, Mus musculus in the Hawaiian islands, In *Studies in genetics VII* **7213**, 269-296 (1972).
314. Inaguma, Y., Miyashita, N., Moriwaki, K., Huai, W. C., Mei-Lei, J., Xinquiao, H. & Ikeda, H. Acquisition of two endogenous ecotropic murine leukemia viruses in distinct Asian wild mouse populations, *J. Virol.* **65**, 1796-1802 (1991).
315. Qi, C.-F., Bonhomme, F., Buckler-White, A., Buckler, C., Orth, A., Lander, M.R., Chattopadhyay, S.K. & Morse III, H.C. Molecular phylogeny of *Fv1*, *Mammalian Genome* **9**, 1049-1055 (1998).
316. Bénit, L., Lallemand, J.B., Casella, J.F., Philippe, H. & Heidmann, T. ERV-L elements: a family of endogenous retrovirus-like elements active throughout the evolution of mammals, *J. Virol* **73**, 3301-8 (1999).
317. Simpson, E.M., Linder, C.C., Sargent, E.E., Davisson, M.T., Mobraaten. L.E. & Sharp, J.J. Genetic variation among 129 substrains and its importance for targeted mutagenesis in mice, *Nature Genetics* **16**, 19-27 (1997).
318. Threadgill, D.W., Yee, D., Latin, A., Nadeau, J.H. & Magnuson, T. Genealogy of the 129 inbred strains: 129/SvJ is a contaminated inbred strain, *Mammalian Genome* **8**, 390-393 (1997).
319. Lander, M.R. & Chattopadhyay, S.K. A *Mus dunni* cell line that lacks sequences closely related to endogenous murine leukemia viruses and can be infected by ectropic, amphotropic, xenotropic, and mink cell focus-forming viruses, *J. Virol* **52**, 695-8 (1984).

Appendix 1 DNA Size Markers

λ HinDIII (Life Technologies):

23,130
9,416
6,557
4,361
2,322
2,027
567
125

φx-174 HaeIII (Life Technologies):

1,353
1,078
872
603
310
281
271
234
194
118
72

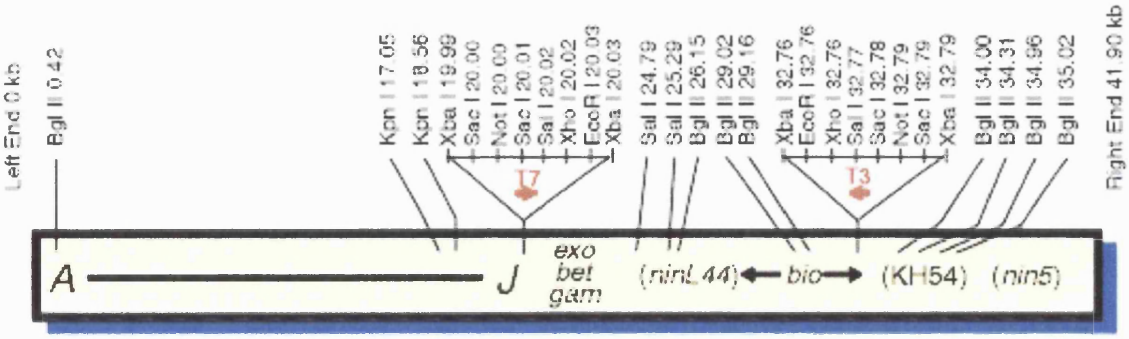
1kb Ladder (MBI Fermentas):

10000
8000
6000
5000
4000
3500
3000*
2500
2000
1500
1200
1031*
900
800
700
600
500*
400
300
200
100

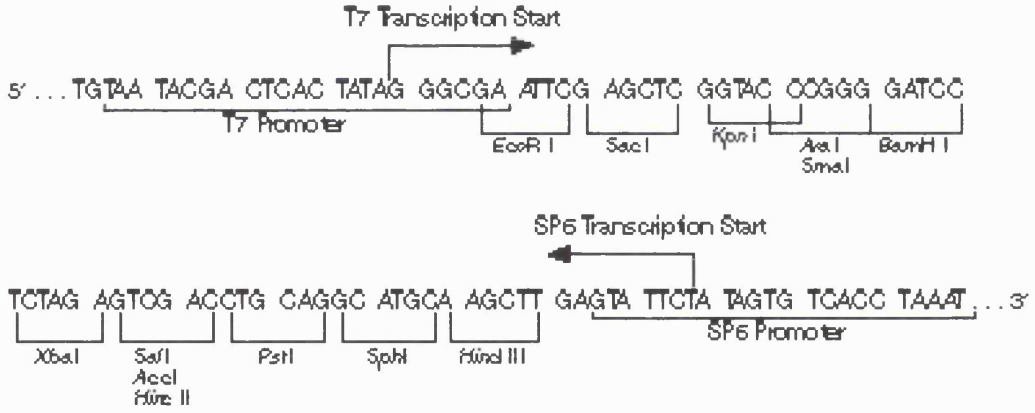
* These bands are brighter than the rest to serve as references

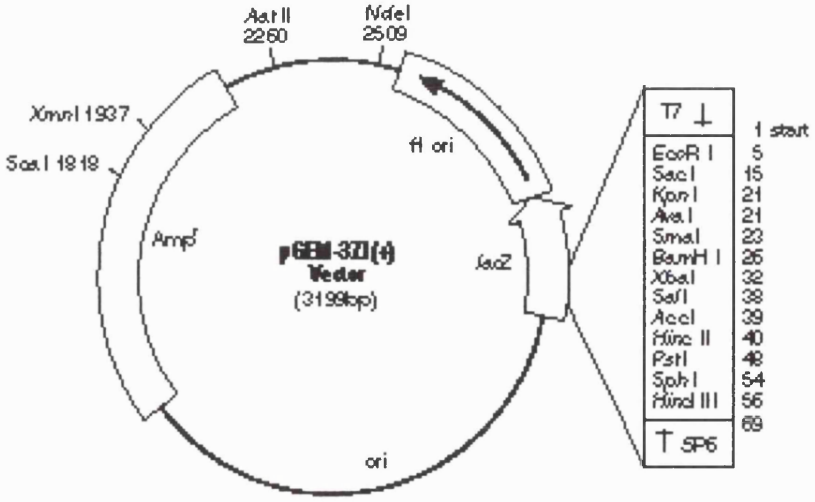
Appendix 2 Cloning and Expression Vectors

Lambda FIX II cloning vector (Stratagene)

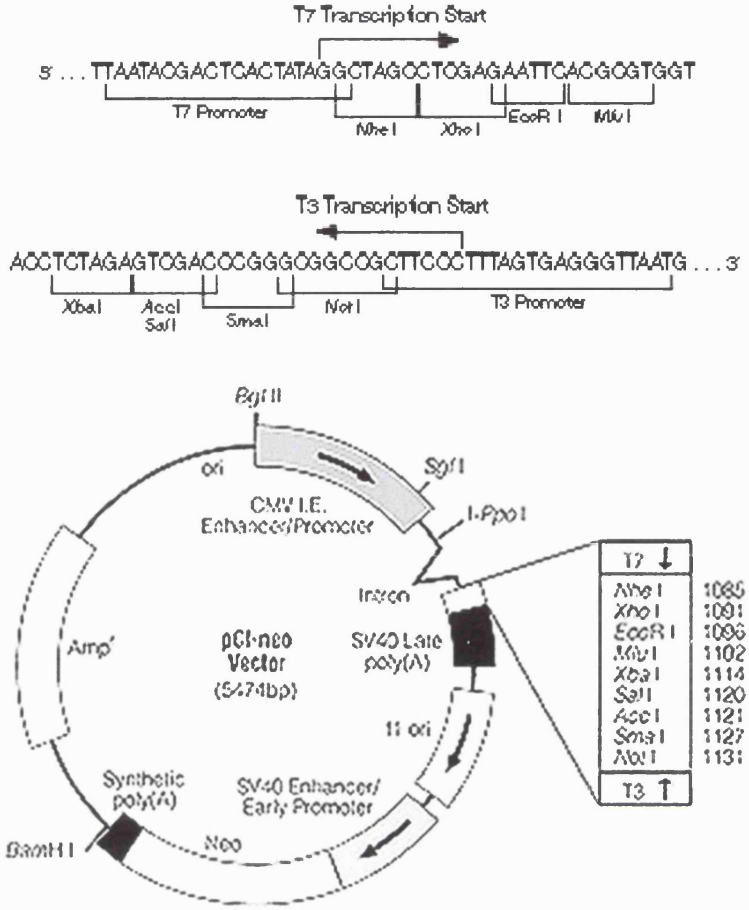


pGEM®-3Zf(+) cloning vector (Promega)

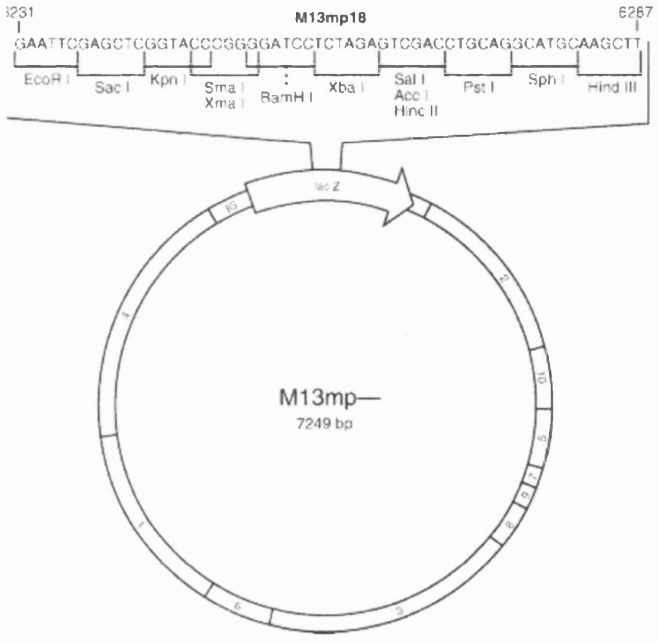




pCI-neo expression vector (Promega)

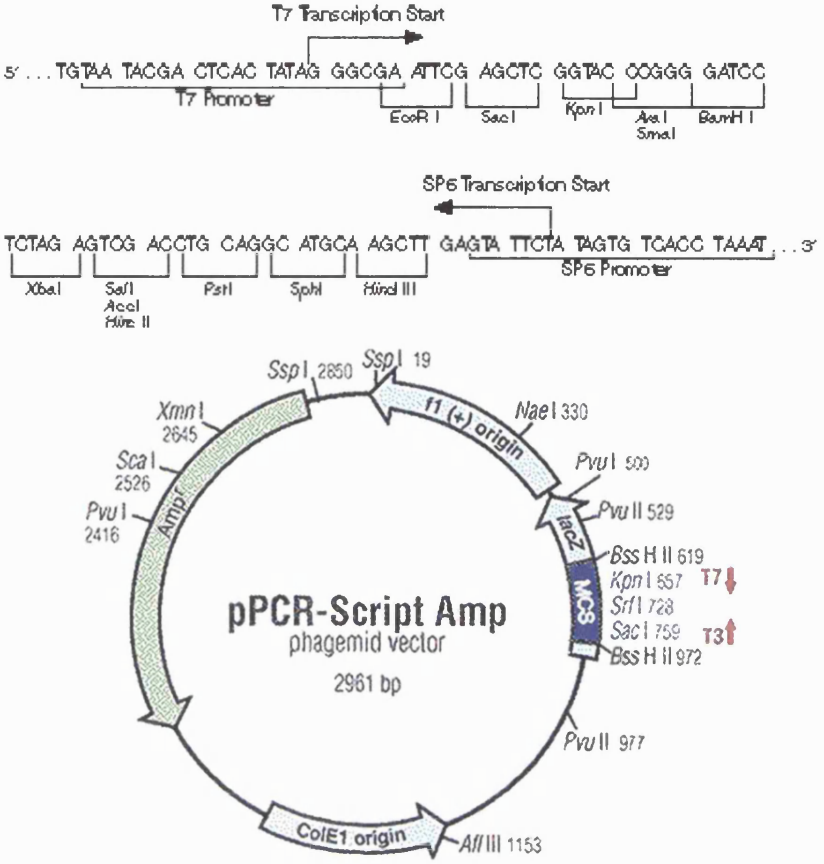


M13mp18* bacteriophage cloning vector (Amersham Phamacia)

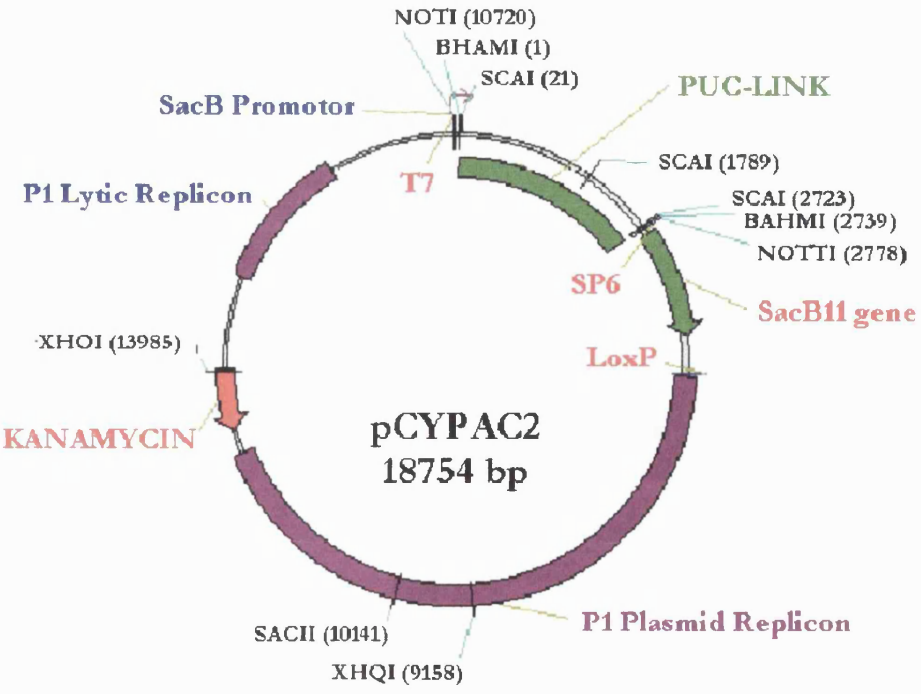


* M13mp19 is identical to M13mp18, except the MCS shown is in the opposite orientation

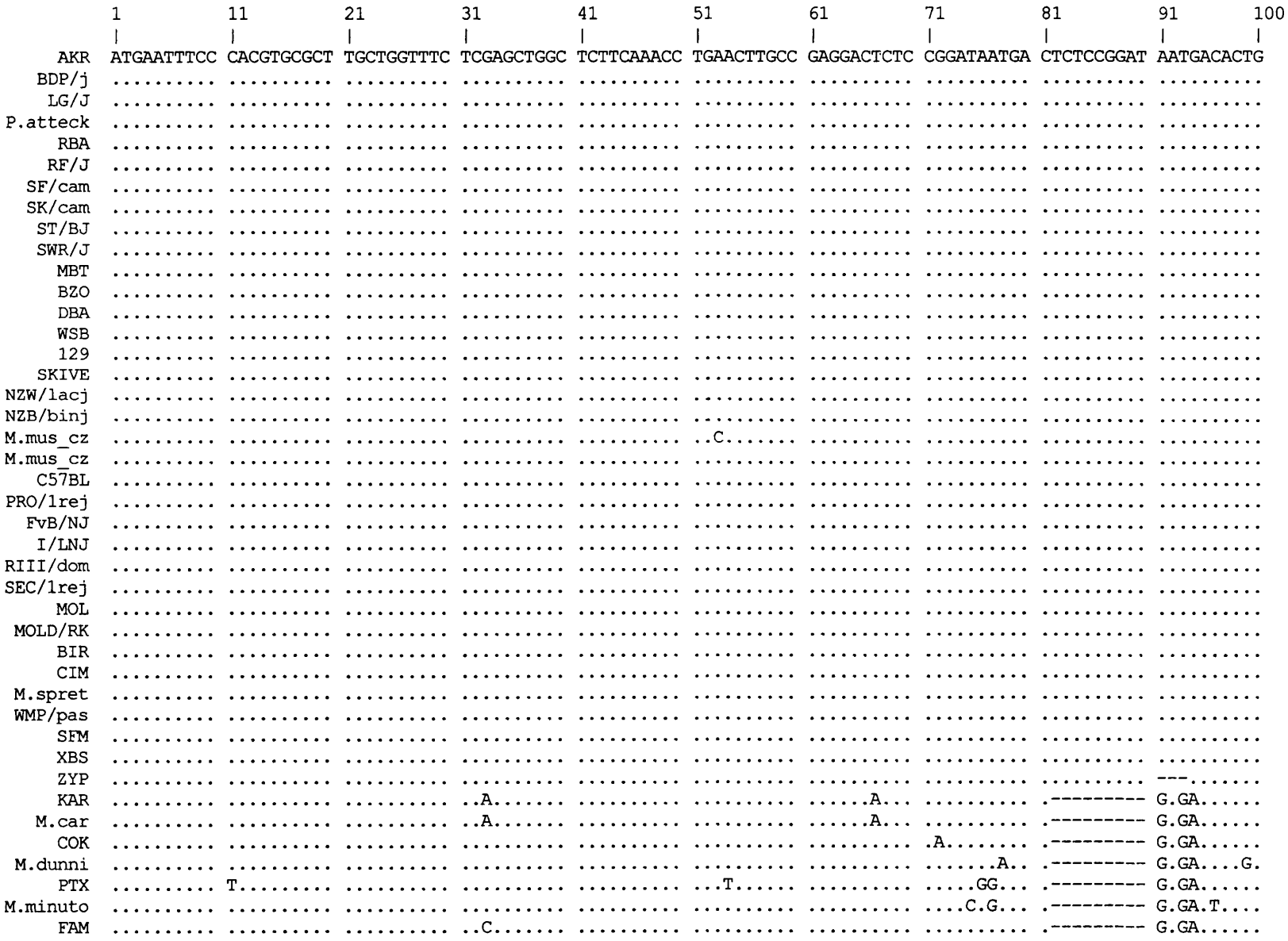
pPCR-script^{AMP} phagemid cloning vector(Stratagene)



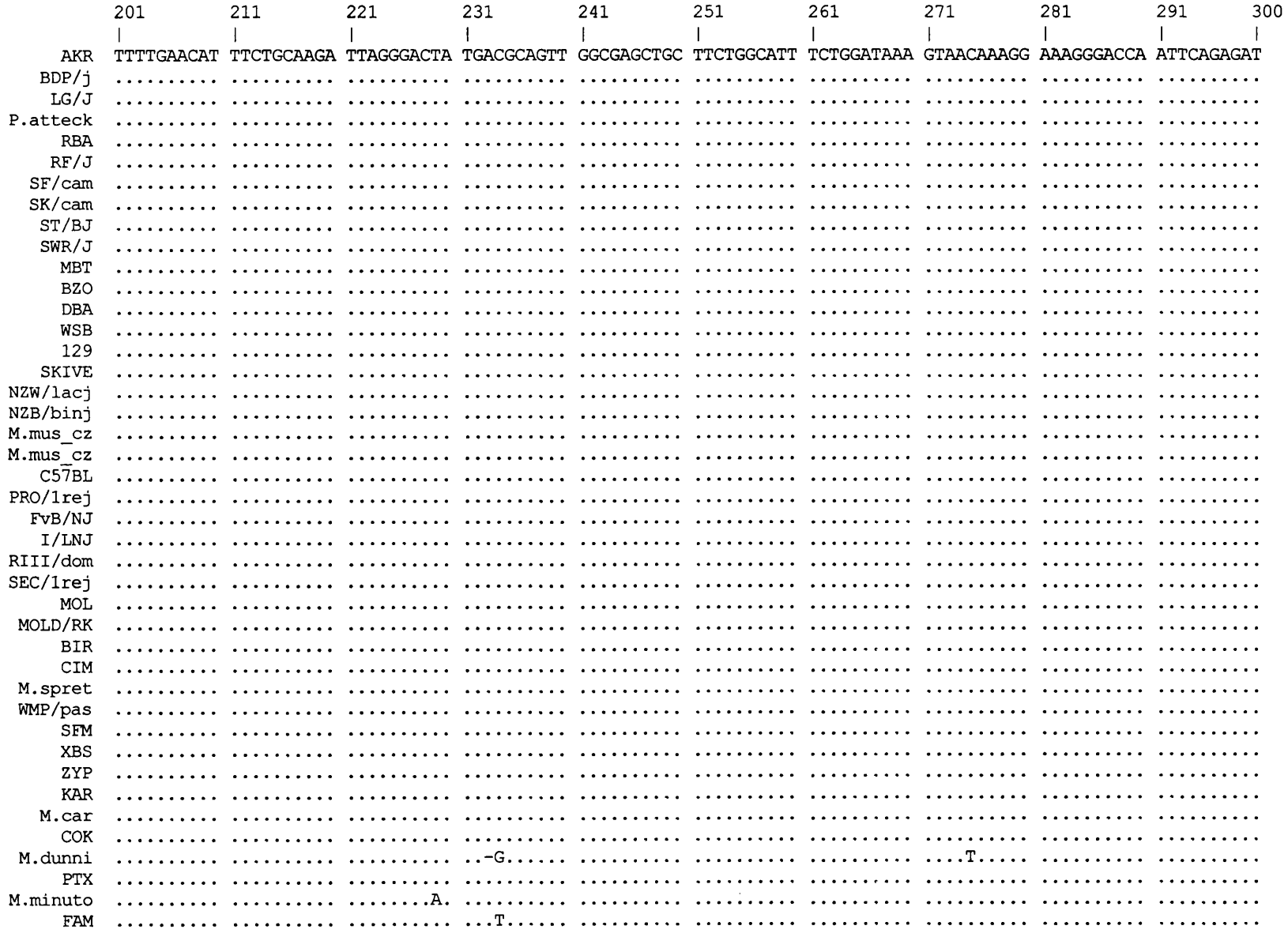
PAC cloning vector pCYCPAC2N

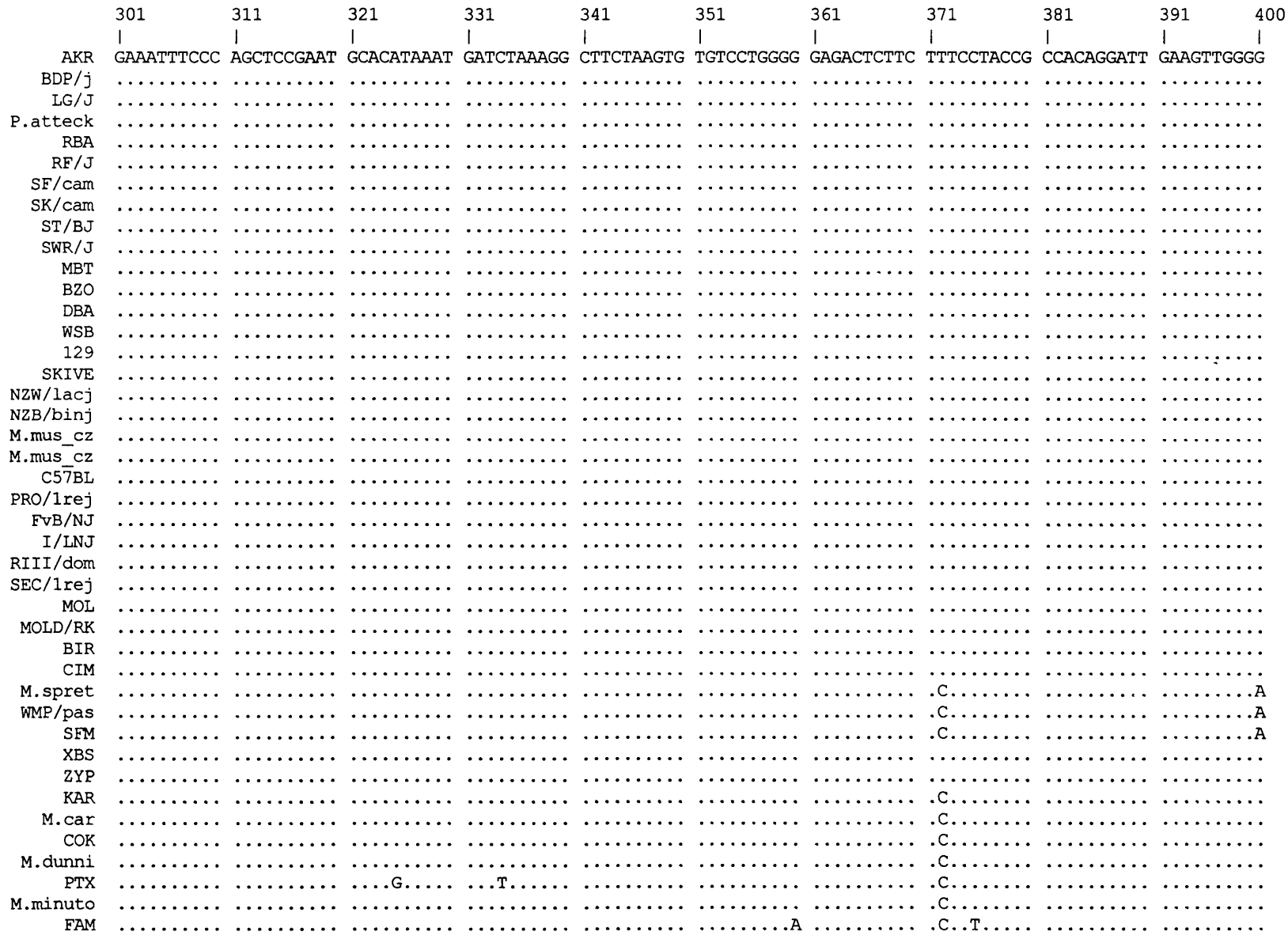


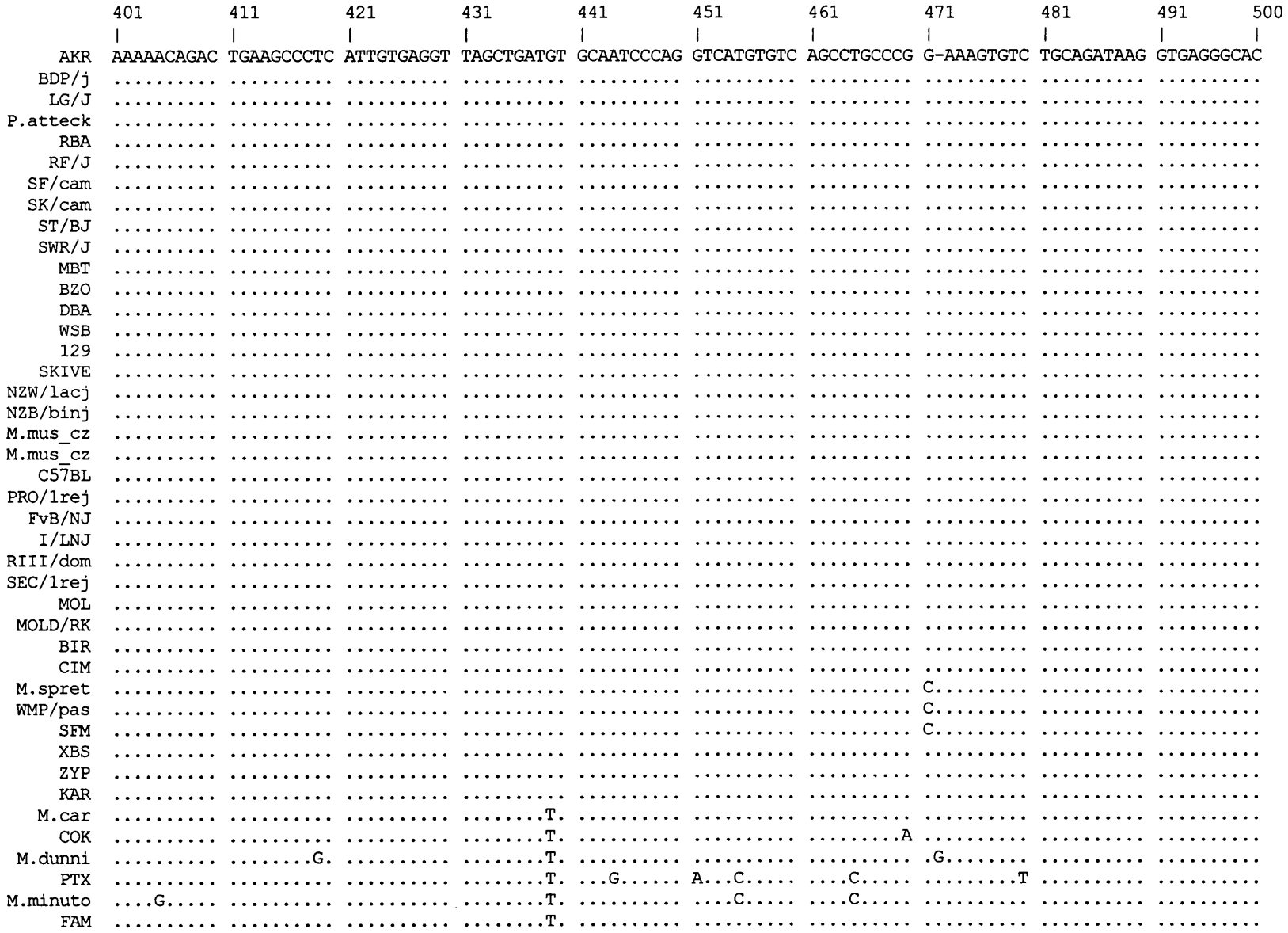
Appendix 3 Multiple sequence alignments using the *Fv1* sequence from the mice used in the sequencing project

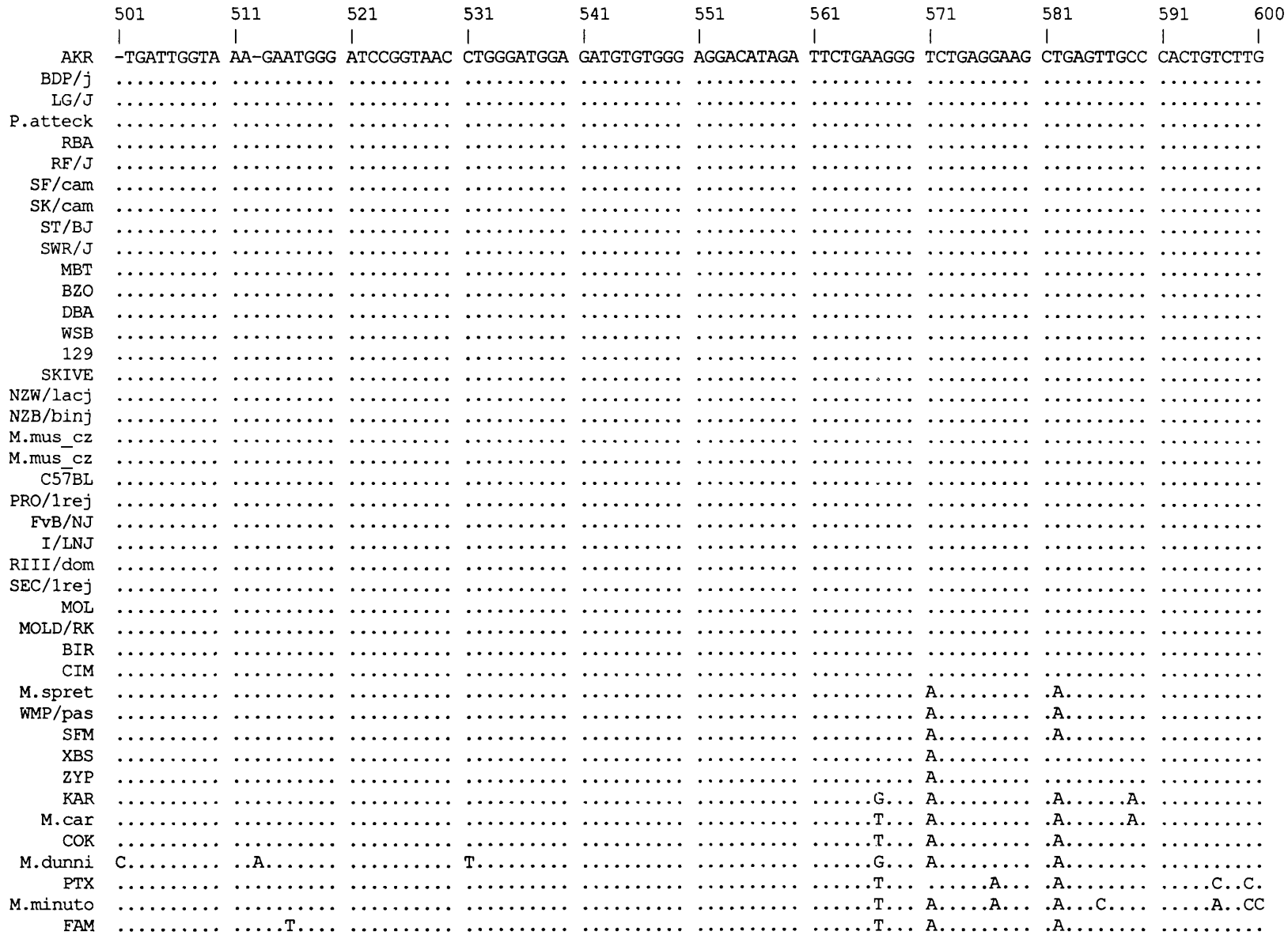

1 11 21 31 41 51 61 71 81 91 100
AKR ATGAATTTCC CACGTGCGCT TGCTGGTTTC TCGAGCTGGC TCTTCAAACC TGAACTTGCC GAGGACTCTC CGGATAATGA CTCTCCGGAT AATGACACTG
BDP/j
LG/J
P.atteck
RBA
RF/J
SF/cam
SK/cam
ST/BJ
SWR/J
MBT
BZO
DBA
WSB
129
SKIVE
NZW/lacj
NZB/binj
M.mus_cz ..C.....
M.mus_cz
C57BL
PRO/1rej
FvB/NJ
I/LNJ
RIII/dom
SEC/1rej
MOL
MOLD/RK
BIR
CIM
M.spret
WMP/pas
SFM
XBS
ZYP ---.....
KAR ..A..... .....A... .---------- G.GA.....
M.car ..A..... .....A... .---------- G.GA.....
COK .A..... .---------- G.GA.....
M.dunni .....A.. .---------- G.GA....G.
PTX T..... ...T..... .....GG... .---------- G.GA.....
M.minuto ....C.G... .---------- G.GA.T....
FAM ..C..... .---------- G.GA.....

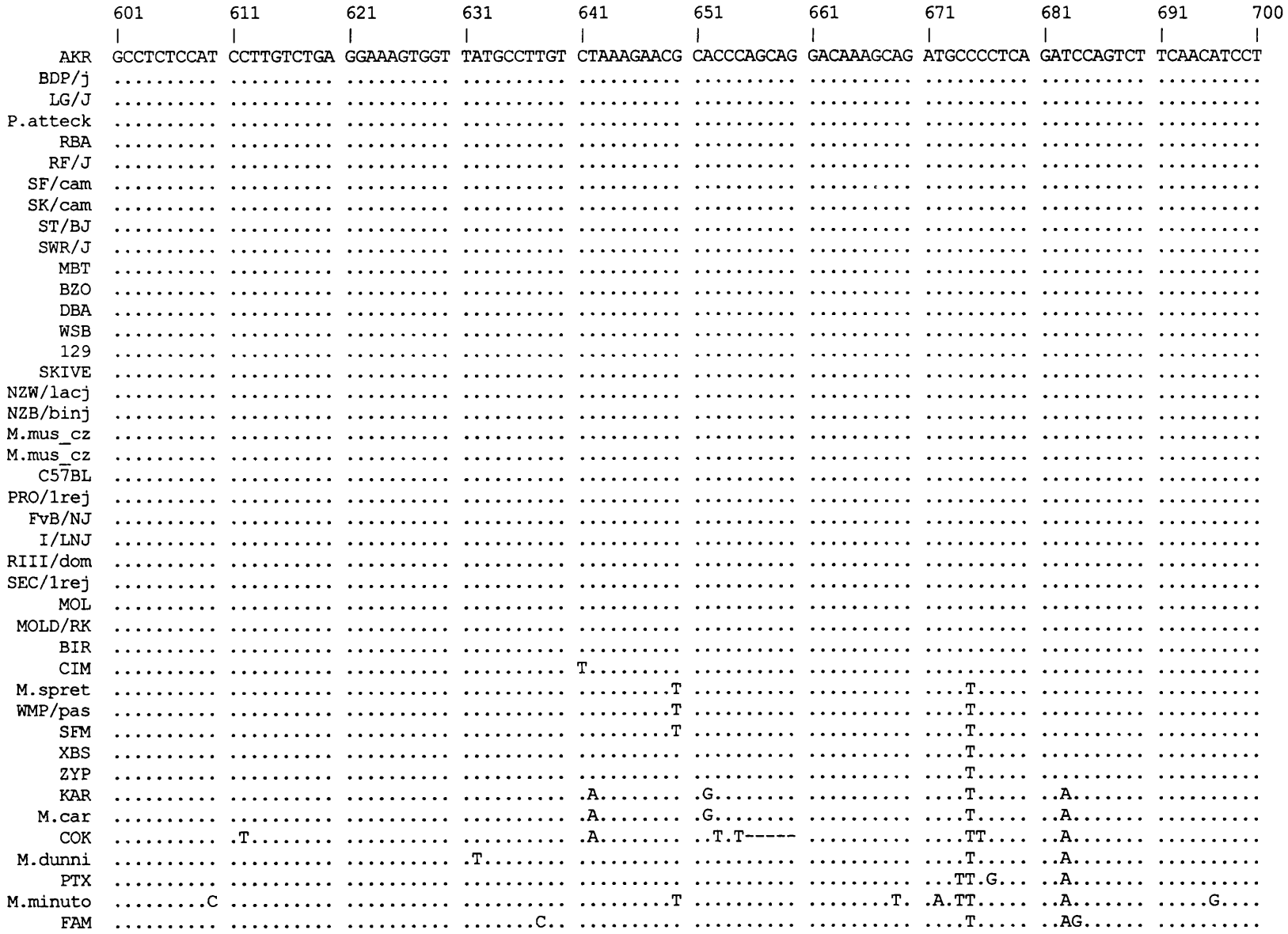
```
          101        111        121        131        141        151        161        171        181        191      200
          |          |          |          |          |          |          |          |          |          |        |
      AKR TTAACCCATG GCGTGAGCTG CTGCAGAAGA TAAATGTGGC CGATCTCCCC GATTCATCCT TTTCGAGCGG TAAGGAACTT AATGACTCTG TGTACCATAC
    BDP/j .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     LG/J .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
P.atteck .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
      RBA .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     RF/J .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   SF/cam .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   SK/cam .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    ST/BJ .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    SWR/J .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
      MBT .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
      BZO .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
      DBA .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
      WSB .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
      129 .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    SKIVE .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
 NZW/lacj .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
 NZB/binj .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
 M.mus_cz .......... .....G.... .......... .......... .......... .......... .......... .......... .......... ..........
 M.mus_cz .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    C57BL .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
 PRO/1rej .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    FvB/NJ .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    I/LNJ .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
 RIII/dom .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
 SEC/1rej .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
      MOL .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
  MOLD/RK .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
      BIR .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
      CIM .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
  M.spret ...GT..... .......... .......... .......... .......... .......... .......... .......... .......... .....G....
  WMP/pas ...GT..... .......... .......... .......... .......... .......... .......... .......... .......... .....G....
      SFM ...GT..... .......... .......... .......... .......... .......... .......... .......... .......... .....G....
      XBS ...G..... .......... .......... .......... .A........ .......... .......... .......... .......... .....G....
      ZYP ...G..... .......... .......... .......... .......... .......... .......... .......... .......... .....GG...
      KAR .......... .......... .......... .......... .......... .......... .......... .......... .......... .....G....
    M.car .......... .......... .......... .......... .......... .......... .......... .......... ....T..... .....G....
      COK .......... .......... .......... .......... .......... .......... .......... .......... .G........ .....G....
  M.dunni .......... .......... .......... .......... ........T. .......... .......... .......... .......... .....G....
      PTX .......G.. .......... .......... ...G..C... .......... .......... .......... .......... .......... ...G.G....
 M.minuto .......... .......... .......... ......CA.. T......... .......... .......T.. ....A..... .......... .....G....
      FAM .......... .......... .......... .......... .......... .......... .......... .......... .......... .....G....
```

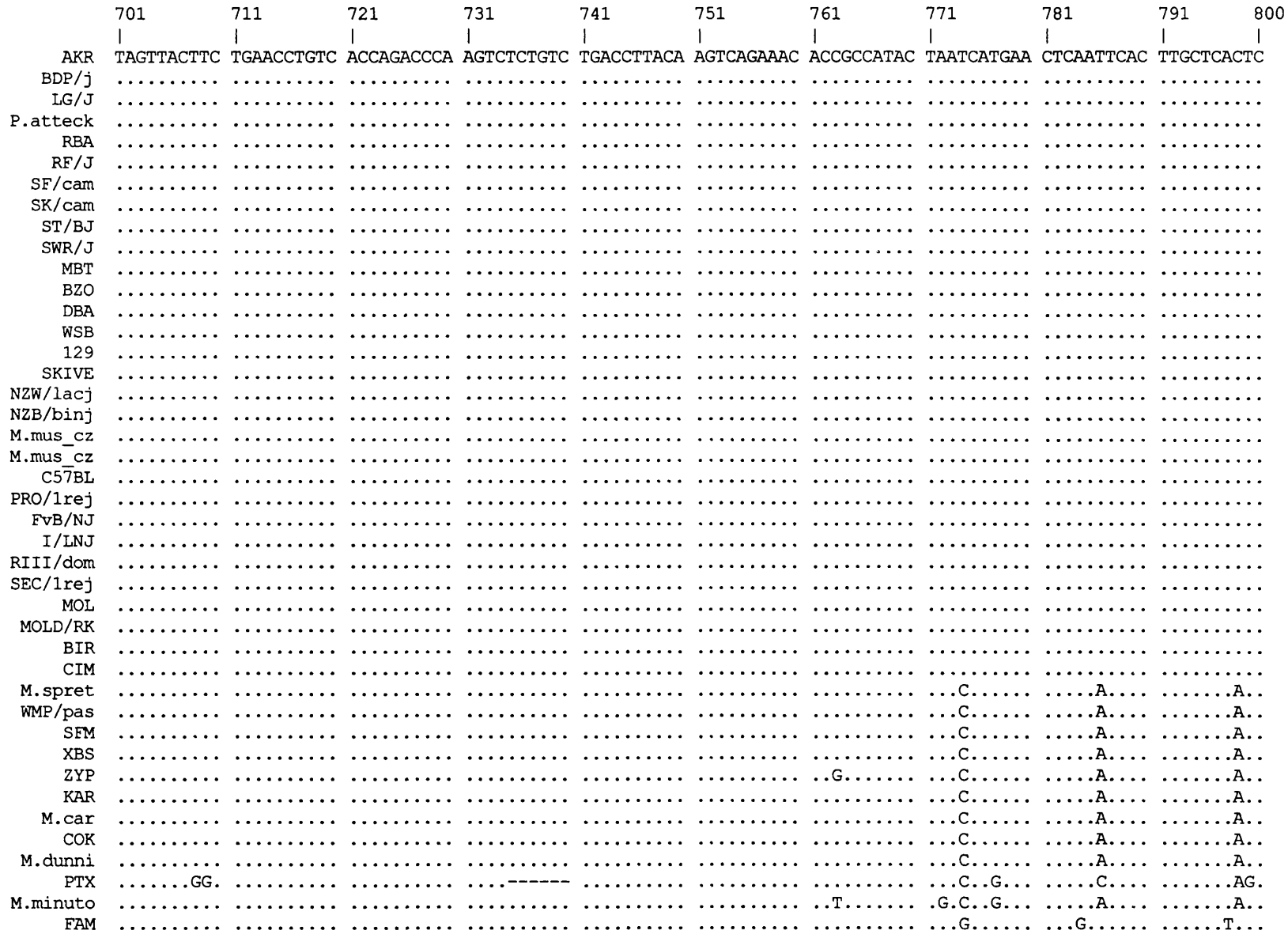

201 211 221 231 241 251 261 271 281 291 300
AKR TTTTGAACAT TTCTGCAAGA TTAGGGACTA TGACGCAGTT GGCGAGCTGC TTCTGGCATT TCTGGATAAA GTAACAAAGG AAAGGGACCA ATTCAGAGAT
BDP/j
LG/J
P.atteck
RBA
RF/J
SF/cam
SK/cam
ST/BJ
SWR/J
MBT
BZO
DBA
WSB
129
SKIVE
NZW/lacj
NZB/binj
M.mus_cz
M.mus_cz
C57BL
PRO/1rej
FvB/NJ
I/LNJ
RIII/dom
SEC/1rej
MOL
MOLD/RK
BIR
CIM
M.spret
WMP/pas
SFM
XBS
ZYP
KAR
M.car
COK
M.dunni ..-G..... ....T.....
PTX
M.minuto .....A.
FAM ...T.....

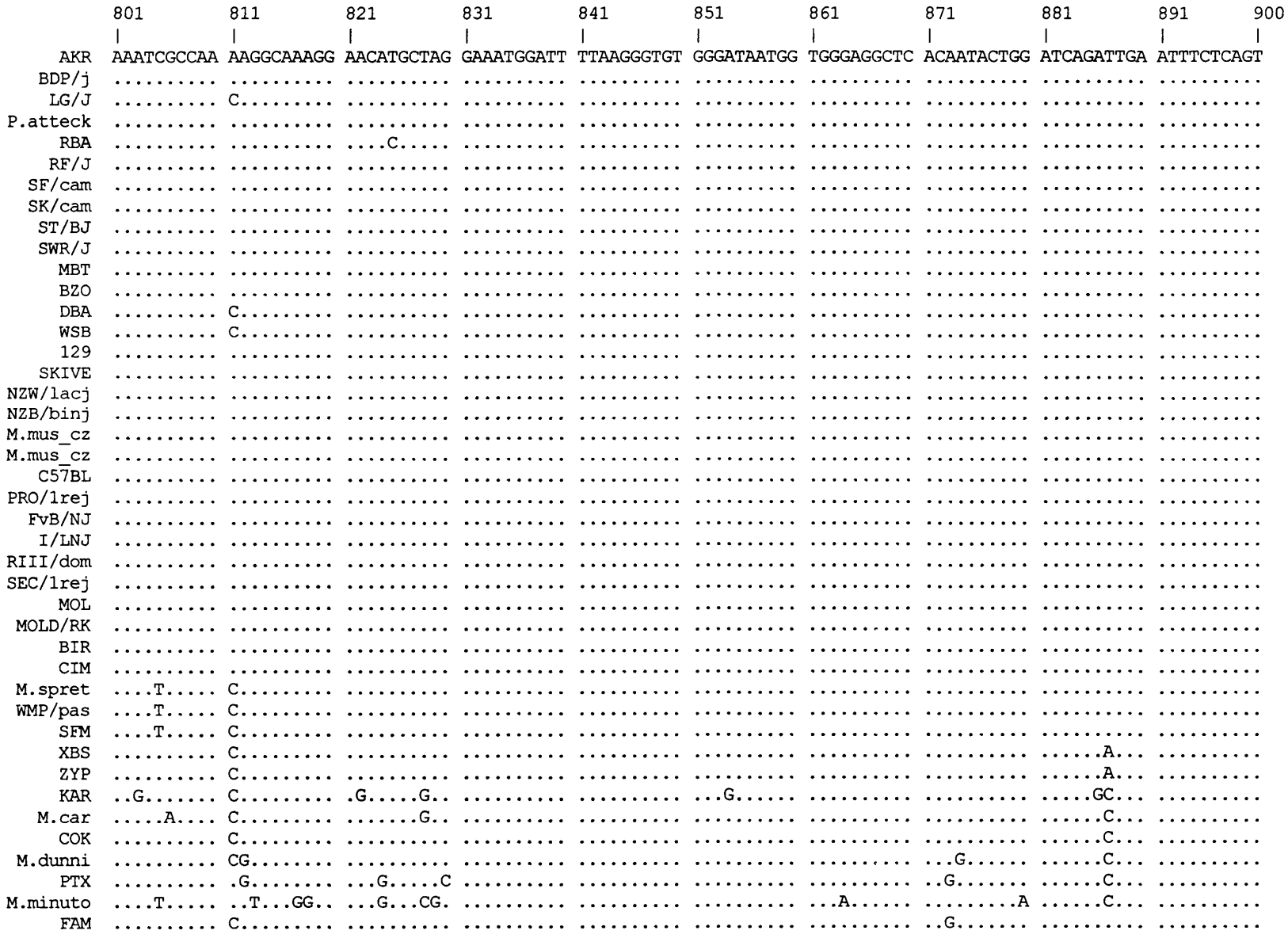

301 311 321 331 341 351 361 371 381 391 400
AKR GAAATTTCCC AGCTCCGAAT GCACATAAAT GATCTAAAGG CTTCTAAGTG TGTCCTGGGG GAGACTCTTC TTTCCTACCG CCACAGGATT GAAGTTGGGG
BDP/j
LG/J
P.atteck
RBA
RF/J
SF/cam
SK/cam
ST/BJ
SWR/J
MBT
BZO
DBA
WSB
129
SKIVE
NZW/lacj
NZB/binj
M.mus_cz
M.mus_cz
C57BL
PRO/1rej
FvB/NJ
I/LNJ
RIII/dom
SEC/1rej
MOL
MOLD/RK
BIR
CIM
M.spret .....C.....A
WMP/pas .....C.....A
SFM .....C.....A
XBS
ZYP
KAR .....C.....
M.car .....C.....
COK .....C.....
M.dunni .....C.....
PTX ....G..... ...T..... .....C.....
M.minuto .....C.....
FAM .....A .....C..T.....

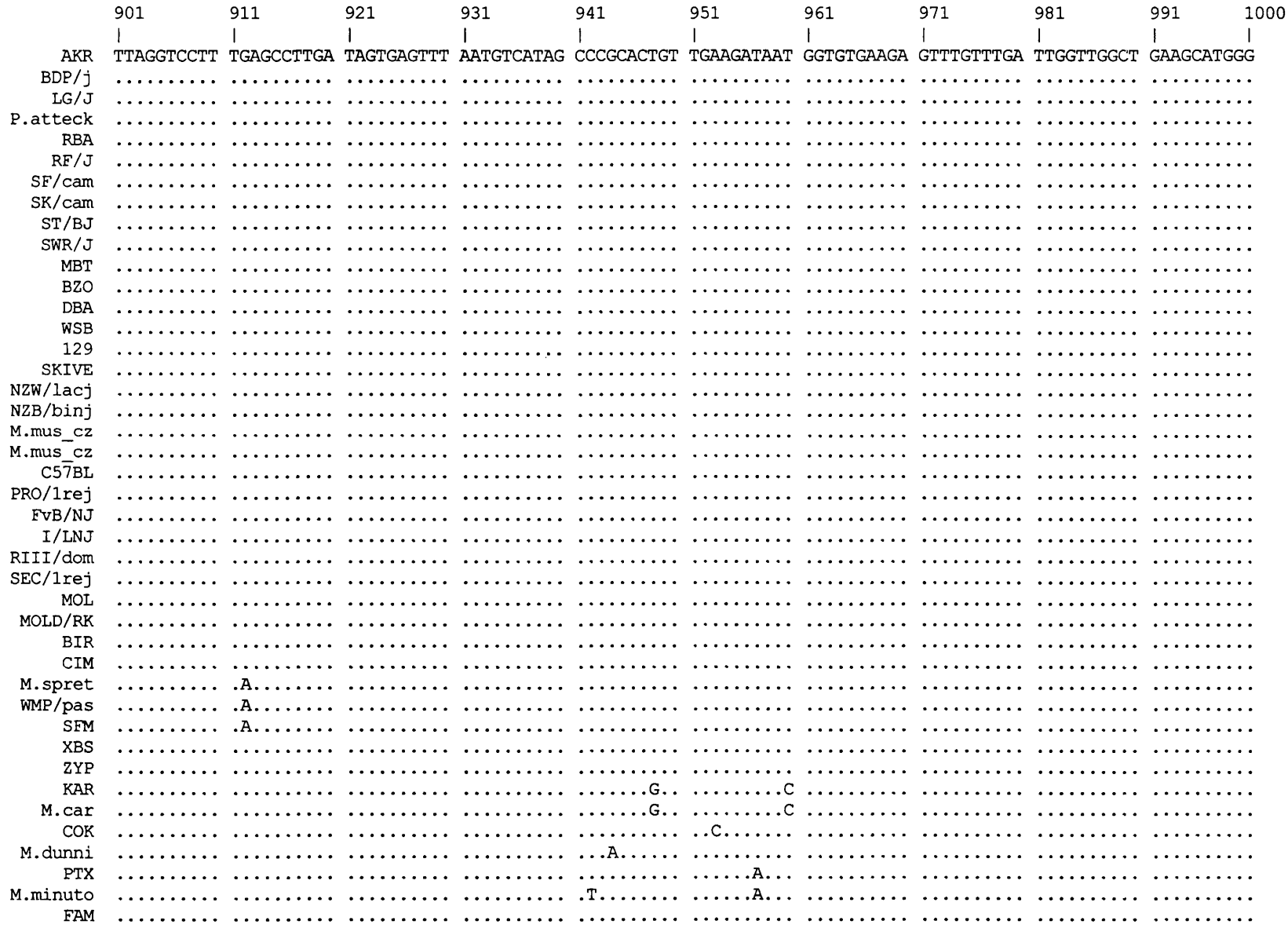

401 411 421 431 441 451 461 471 481 491 500
AKR AAAAACAGAC TGAAGCCCTC ATTGTGAGGT TAGCTGATGT GCAATCCCAG GTCATGTGTC AGCCTGCCCG G-AAAGTGTC TGCAGATAAG GTGAGGGCAC
BDP/j
LG/J
P.atteck
RBA
RF/J
SF/cam
SK/cam
ST/BJ
SWR/J
MBT
BZO
DBA
WSB
129
SKIVE
NZW/lacj
NZB/binj
M.mus_cz
M.mus_cz
C57BL
PRO/1rej
FvB/NJ
I/LNJ
RIII/dom
SEC/1rej
MOL
MOLD/RK
BIR
CIM
M.spret
WMP/pas
SFM
XBS
ZYP
KAR
M.car
COK
M.dunni
PTX
M.minuto
FAM

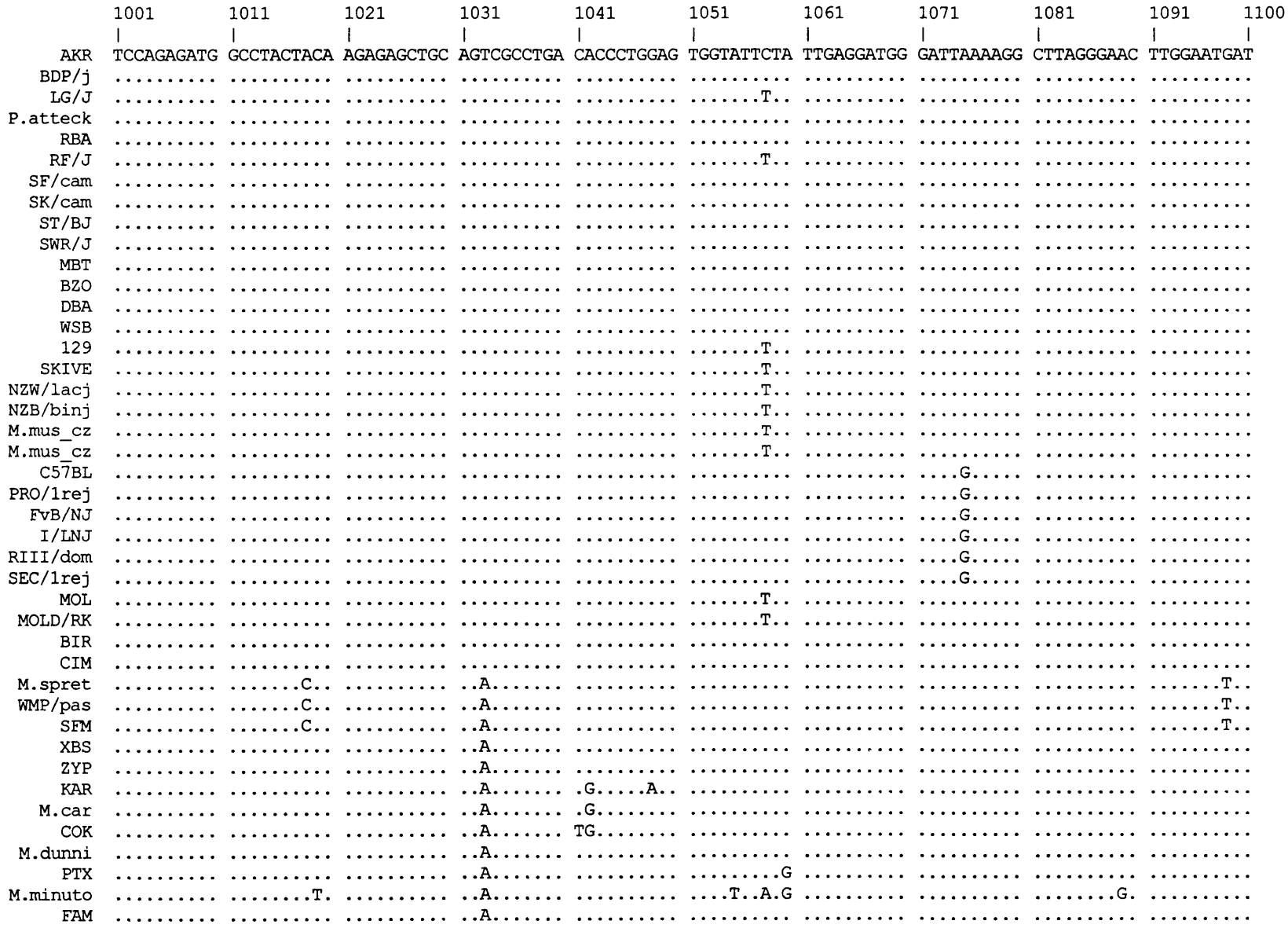

501 511 521 531 541 551 561 571 581 591 600
AKR -TGATTGGTA AA-GAATGGG ATCCGGTAAC CTGGGATGGA GATGTGTGGG AGGACATAGA TTCTGAAGGG TCTGAGGAAG CTGAGTTGCC CACTGTCTTG
BDP/j
LG/J
P.atteck
RBA
RF/J
SF/cam
SK/cam
ST/BJ
SWR/J
MBT
BZO
DBA
WSB
129
SKIVE
NZW/lacj
NZB/binj
M.mus_cz
M.mus_cz
C57BL
PRO/1rej
FvB/NJ
I/LNJ
RIII/dom
SEC/1rej
MOL
MOLD/RK
BIR
CIM
M.spret
WMP/pas
SFM
XBS
ZYP
KAR
M.car
COK
M.dunni
PTX
M.minuto
FAM


601 611 621 631 641 651 661 671 681 691 700
AKR GCCTCTCCAT CCTTGTCTGA GGAAAGTGGT TATGCCTTGT CTAAAGAACG CACCCAGCAG GACAAAGCAG ATGCCCCTCA GATCCAGTCT TCAACATCCT
BDP/j
LG/J
P.atteck
RBA
RF/J
SF/cam
SK/cam
ST/BJ
SWR/J
MBT
BZO
DBA
WSB
129
SKIVE
NZW/lacj
NZB/binj
M.mus_cz
M.mus_cz
C57BL
PRO/1rej
FvB/NJ
I/LNJ
RIII/dom
SEC/1rej
MOL
MOLD/RK
BIR
CIM
M.spret
WMP/pas
SFM
XBS
ZYP
KAR
M.car
COK
M.dunni
PTX
M.minuto
FAM

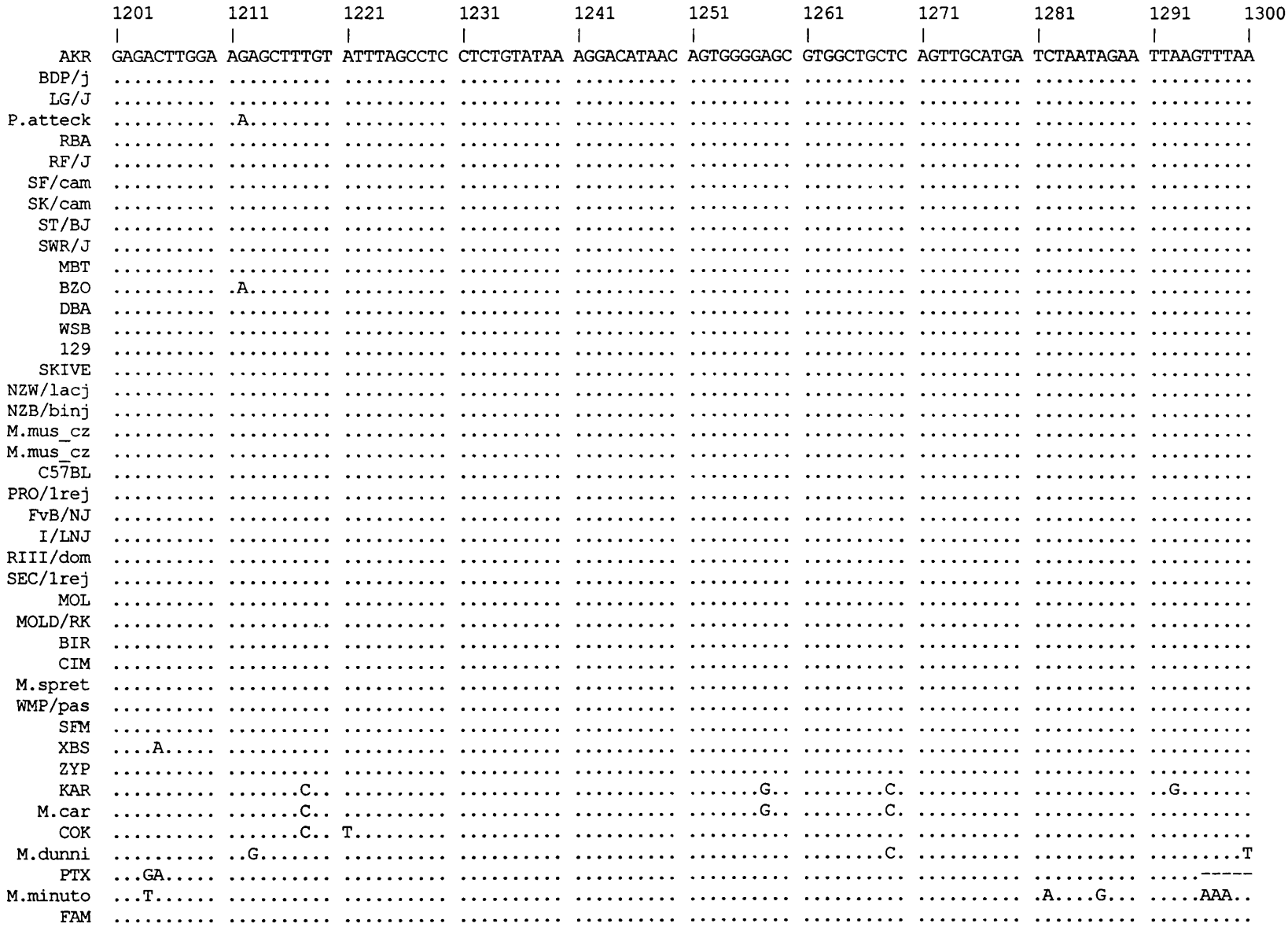

701 711 721 731 741 751 761 771 781 791 800
AKR TAGTTACTTC TGAACCTGTC ACCAGACCCA AGTCTCTGTC TGACCTTACA AGTCAGAAAC ACCGCCATAC TAATCATGAA CTCAATTCAC TTGCTCACTC
BDP/j
LG/J
P.atteck
RBA
RF/J
SF/cam
SK/cam
ST/BJ
SWR/J
MBT
BZO
DBA
WSB
129
SKIVE
NZW/lacj
NZB/binj
M.mus_cz
M.mus_cz
C57BL
PRO/1rej
FvB/NJ
I/LNJ
RIII/dom
SEC/1rej
MOL
MOLD/RK
BIR
CIM
M.spret ...C..... .....A.... ....A..
WMP/pas ...C..... .....A.... ....A..
SFM ...C..... .....A.... ....A..
XBS ...C..... .....A.... ....A..
ZYP ..G..... ...C..... .....A.... ....A..
KAR ...C..... .....A.... ....A..
M.car ...C..... .....A.... ....A..
COK ...C..... .....A.... ....A..
M.dunni ...C..... .....A.... ....A..
PTX .....GG. ....------ ...C..G... .....C.... ....AG.
M.minuto ..T..... .G.C..G... .....A.... ....A..
FAM ...G..... ...G..... ......T...

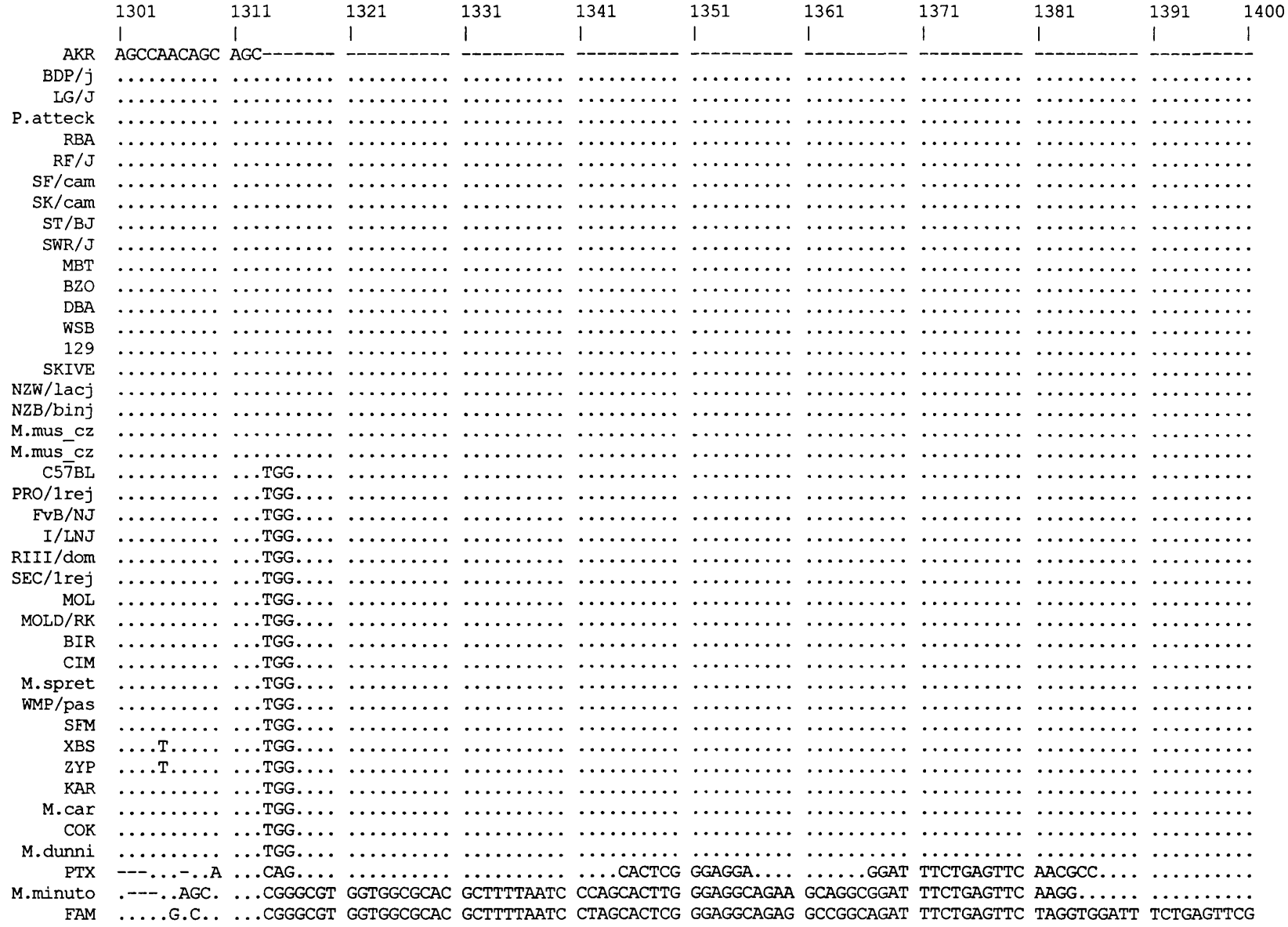

801 811 821 831 841 851 861 871 881 891 900
AKR AAATCGCCAA AAGGCAAAGG AACATGCTAG GAAATGGATT TTAAGGGTGT GGGATAATGG TGGGAGGCTC ACAATACTGG ATCAGATTGA ATTTCTCAGT
BDP/j
LG/J
P.atteck
RBA
RF/J
SF/cam
SK/cam
ST/BJ
SWR/J
MBT
BZO
DBA
WSB
129
SKIVE
NZW/lacj
NZB/binj
M.mus_cz
M.mus_cz
C57BL
PRO/1rej
FvB/NJ
I/LNJ
RIII/dom
SEC/1rej
MOL
MOLD/RK
BIR
CIM
M.spret
WMP/pas
SFM
XBS
ZYP
KAR
M.car
COK
M.dunni
PTX
M.minuto
FAM

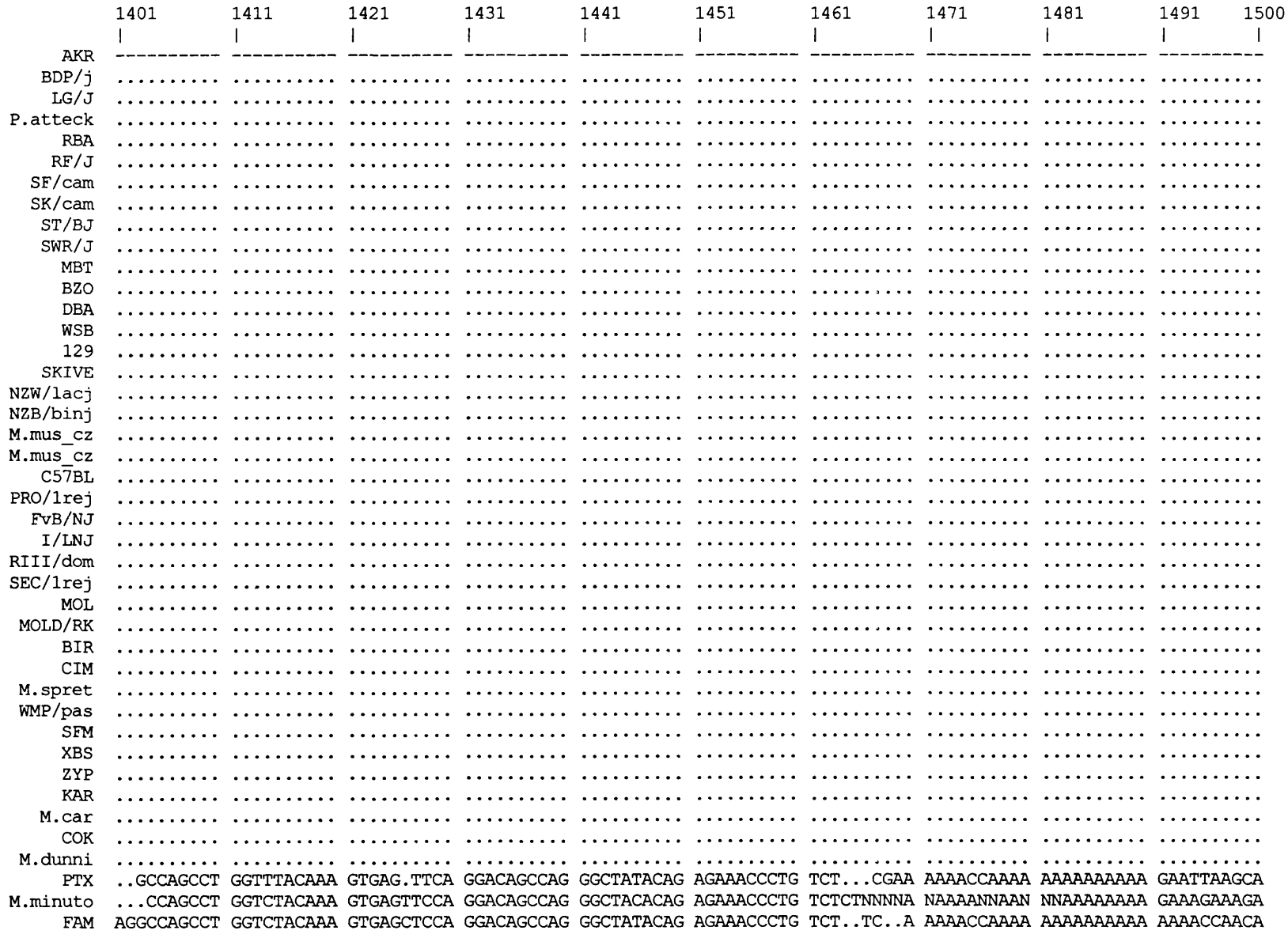

901 911 921 931 941 951 961 971 981 991 1000
AKR TTAGGTCCTT TGAGCCTTGA TAGTGAGTTT AATGTCATAG CCCGCACTGT TGAAGATAAT GGTGTGAAGA GTTTGTTTGA TTGGTTGGCT GAAGCATGGG
BDP/j
LG/J
P.atteck
RBA
RF/J
SF/cam
SK/cam
ST/BJ
SWR/J
MBT
BZO
DBA
WSB
129
SKIVE
NZW/lacj
NZB/binj
M.mus_cz
M.mus_cz
C57BL
PRO/1rej
FvB/NJ
I/LNJ
RIII/dom
SEC/1rej
MOL
MOLD/RK
BIR
CIM
M.spret .A.
WMP/pas .A.
SFM .A.
XBS
ZYP
KAR .....G.. C
M.car .....G.. C
COK ..C.
M.dunni ...A.
PTX .....A...
M.minuto .T..... .....A...
FAM


1001 1011 1021 1031 1041 1051 1061 1071 1081 1091 1100
AKR TCCAGAGATG GCCTACTACA AGAGAGCTGC AGTCGCCTGA CACCCTGGAG TGGTATTCTA TTGAGGATGG GATTAAAAGG CTTAGGGAAC TTGGAATGAT
BDP/j
LG/J
P.atteck
RBA
RF/J
SF/cam
SK/cam
ST/BJ
SWR/J
MBT
BZO
DBA
WSB
129
SKIVE
NZW/lacj
NZB/binj
M.mus_cz
M.mus_cz
C57BL
PRO/1rej
FvB/NJ
I/LNJ
RIII/dom
SEC/1rej
MOL
MOLD/RK
BIR
CIM
M.spret
WMP/pas
SFM
XBS
ZYP
KAR
M.car
COK
M.dunni
PTX
M.minuto
FAM

```
          1101       1111       1121       1131       1141       1151       1161       1171       1181       1191      1200
          |          |          |          |          |          |          |          |          |          |         |
AKR       AGAGTGGCTT TGTGTAAAAG CTACTTGTCC ACAGTGGAGG GGCCCGGAAG ATGTACCCAT CACGAGAGCT ATGAGGATAA CTTTTGTCCG GGAAACTGTA
BDP/j     .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
LG/J      .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
P.atteck  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
RBA       .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
RF/J      .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
SF/cam    .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
SK/cam    .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
ST/BJ     .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
SWR/J     .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
MBT       .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
BZO       .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
DBA       .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
WSB       .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
129       .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
SKIVE     .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
NZW/lacj  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
NZB/binj  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
M.mus_cz  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
M.mus_cz  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
C57BL     .......... .......... .......... .......... .......... .......... .......... .......... .......... .......AG.
PRO/1rej  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......AG.
FvB/NJ    .......... .......... .......... .......... .......... .......... .......... .......... .......... .......AG.
I/LNJ     .......... .......... .......... .......... .......... .......... .......... .......... .......... .......AG.
RIII/dom  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......AG.
SEC/1rej  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......AG.
MOL       .......... .......... .......... .......... .......... .......... .......... .......... .......... .......AG.
MOLD/RK   .......... .......... .......... .......... .......... .......... .......... .......... .......... .......AG.
BIR       .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
CIM       .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
M.spret   .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
WMP/pas   .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
SFM       .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
XBS       .......... .......... .......... .......... .......... ...C...... ........G. .......... .......... ..........
ZYP       .......... .......... .......... .......... .......... ...C...... .......... .......... .......... ..........
KAR       .......... .......... .......... .......... .......... ...C...... .......... .......... .......... ..........
M.car     .......... .......... .......... .......... .......... ...C...... .......... .......... .......... ..........
COK       .......... .......... .......... .......... .......... ...C...... .......... .......... .......... ..........
M.dunni   .......... .......... .......... .......... .......... ...C...... .......... ......C.CC .......... ........A.
PTX       .......... .......... .......... .......... .......... ...C...... ......G... .......... .......... .......A..
M.minuto  .......... .......... ..C....... ..G....... ........T. ...C...... ...T..G... .........G T.......A. ........A.
FAM       .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
```


1201 1211 1221 1231 1241 1251 1261 1271 1281 1291 1300
AKR GAGACTTGGA AGAGCTTTGT ATTTAGCCTC CTCTGTATAA AGGACATAAC AGTGGGGAGC GTGGCTGCTC AGTTGCATGA TCTAATAGAA TTAAGTTTAA
BDP/j
LG/J
P.atteck .A.
RBA
RF/J
SF/cam
SK/cam
ST/BJ
SWR/J
MBT
BZO .A.
DBA
WSB
129
SKIVE
NZW/lacj
NZB/binj
M.mus_cz
M.mus_cz
C57BL
PRO/1rej
FvB/NJ
I/LNJ
RIII/dom
SEC/1rej
MOL
MOLD/RK
BIR
CIM
M.spret
WMP/pas
SFM
XBS ....A.
ZYP
KAR .....C.. .....G.. .....C. ..G.
M.car .....C.. .....G.. .....C.
COK .....C.. T.
M.dunni ..G. .....C. .....T
PTX ...GA..... -----
M.minuto ...T..... .A....G... .....AAA..
FAM

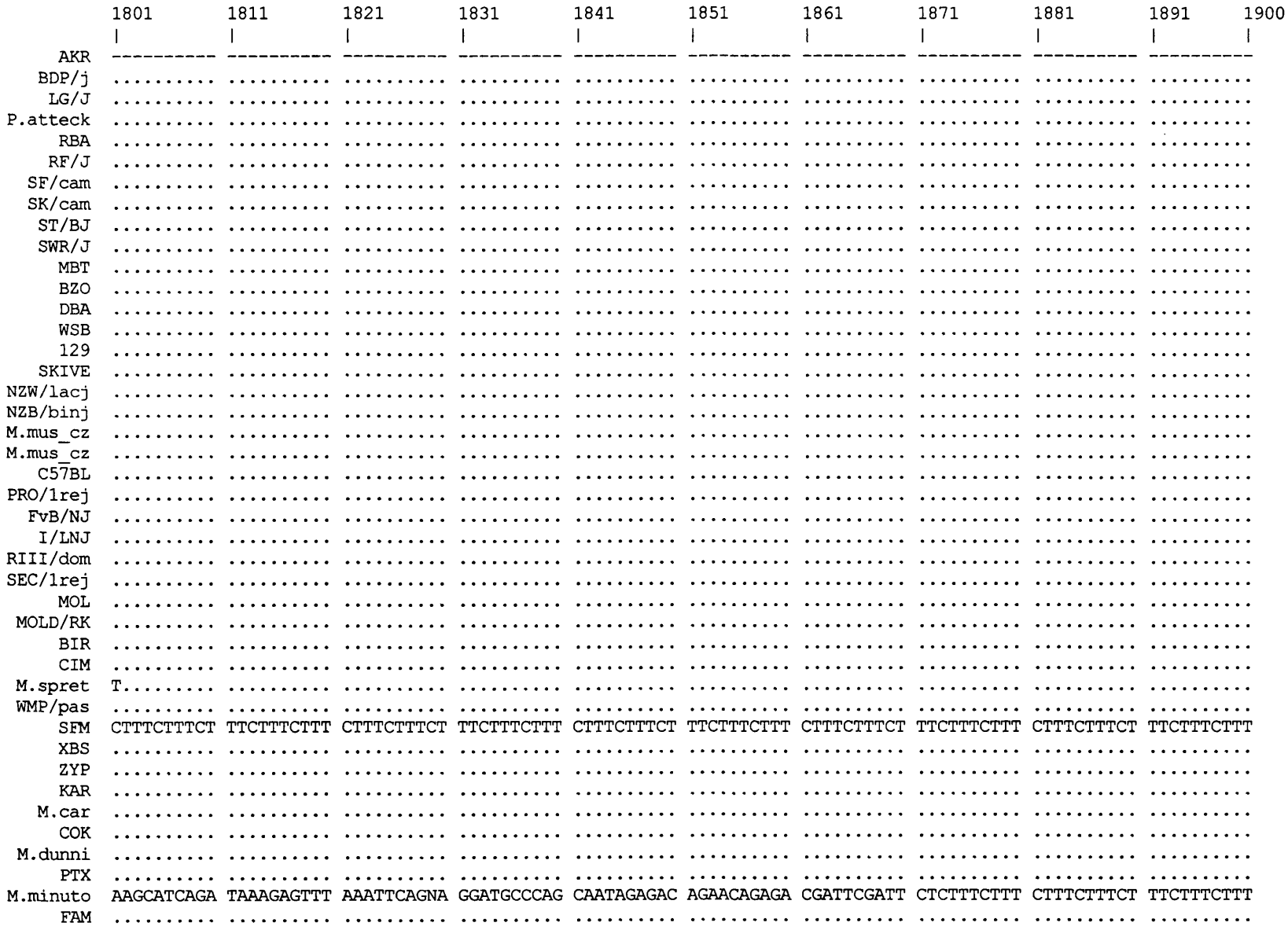

1301 1311 1321 1331 1341 1351 1361 1371 1381 1391 1400
AKR AGCCAACAGC AGC-------
BDP/j
LG/J
P.atteck
RBA
RF/J
SF/cam
SK/cam
ST/BJ
SWR/J
MBT
BZO
DBA
WSB
129
SKIVE
NZW/lacj
NZB/binj
M.mus_cz
M.mus_cz
C57BL ...TGG....
PRO/1rej ...TGG....
FvB/NJ ...TGG....
I/LNJ ...TGG....
RIII/dom ...TGG....
SEC/1rej ...TGG....
MOL ...TGG....
MOLD/RK ...TGG....
BIR ...TGG....
CIM ...TGG....
M.spret ...TGG....
WMP/pas ...TGG....
SFM ...TGG....
XBS ....T..... ...TGG....
ZYP ....T..... ...TGG....
KAR ...TGG....
M.car ...TGG....
COK ...TGG....
M.dunni ...TGG....
PTX ---...-..A ...CAG.... ....CACTCG GGAGGA.... ......GGAT TTCTGAGTTC AACGCC....
M.minuto .---..AGC. ...CGGGCGT GGTGGCGCAC GCTTTTAATC CCAGCACTTG GGAGGCAGAA GCAGGCGGAT TTCTGAGTTC AAGG.....
FAM .....G.C.. ...CGGGCGT GGTGGCGCAC GCTTTTAATC CTAGCACTCG GGAGGCAGAG GCCGGCAGAT TTCTGAGTTC TAGGTGGATT TCTGAGTTCG


1401 1411 1421 1431 1441 1451 1461 1471 1481 1491 1500
AKR ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
BDP/j ....
LG/J ....
P.atteck ....
RBA ....
RF/J ....
SF/cam ....
SK/cam ....
ST/BJ ....
SWR/J ....
MBT ....
BZO ....
DBA ....
WSB ....
129 ....
SKIVE ....
NZW/lacj ....
NZB/binj ....
M.mus_cz ....
M.mus_cz ....
C57BL ....
PRO/1rej ....
FvB/NJ ....
I/LNJ ....
RIII/dom ....
SEC/1rej ....
MOL ....
MOLD/RK ....
BIR ....
CIM ....
M.spret ....
WMP/pas ....
SFM ....
XBS ....
ZYP ....
KAR ....
M.car ....
COK ....
M.dunni ....
PTX ..GCCAGCCT GGTTTACAAA GTGAG.TTCA GGACAGCCAG GGCTATACAG AGAAACCCTG TCT...CGAA AAAACCAAAA AAAAAAAAAA GAATTAAGCA
M.minuto ...CCAGCCT GGTCTACAAA GTGAGTTCCA GGACAGCCAG GGCTACACAG AGAAACCCTG TCTCTNNNNA NAAAANNAAN NNAAAAAAAA GAAAGAAAGA
FAM AGGCCAGCCT GGTCTACAAA GTGAGCTCCA GGACAGCCAG GGCTATACAG AGAAACCCTG TCT..TC..A AAAACCAAAA AAAAAAAAAA AAAACCAACA

```
          1501       1511       1521       1531       1541       1551       1561       1571       1581       1591      1600
          |          |          |          |          |          |          |          |          |          |         |
AKR       ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
BDP/j     .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
LG/J      .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
P.atteck  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
RBA       .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
RF/J      .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
SF/cam    .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
SK/cam    .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
ST/BJ     .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
SWR/J     .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
MBT       .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
BZO       .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
DBA       .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
WSB       .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
129       .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
SKIVE     .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
NZW/lacj  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
NZB/binj  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
M.mus_cz  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
M.mus_cz  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
C57BL     .......... .......... .......... .......CTT GACT...... ........TC TGTAGGCTCT G......... .........T GGGGG.TTCT
PRO/1rej  .......... .......... .......... .......CTT GACT...... ........TC TGTAGGCTCT G......... .........T GGGGG.TTCT
FvB/NJ    .......... .......... .......... .......CTT GACT...... ........TC TGTAGGCTCT G......... .........T GGGGG.TTCT
I/LNJ     .......... .......... .......... .......CTT GACT...... ........TC TGTAGGCTCT G......... .........T GGGGG.TTCT
RIII/dom  .......... .......... .......... .......CTT GACT...... ........TC TGTAGGCTCT G......... .........T GGGGG.TTCT
SEC/1rej  .......... .......... .......... .......CTT GACT...... ........TC TGTAGGCTCT G......... .........T GGGGG.TTCT
MOL       .......... .......... .......... .......CTT GACT...... ........TC TGTAGGCTCT G......... .........T GGGGG.TTCT
MOLD/RK   .......... .......... .......... .......CTT GACT...... ........TC TGTAGGCTCT G......... .........T GGGGG.TTCT
BIR       .......... .......... .......... .......CTT GACT...... ........TC TGTAGGCTCT G......... .........T GGGGG.TTCT
CIM       .......... .......... .......... .......CTT GACT...... ........TC TGTAGGCTCT G......... .........T GGGGG.TTCT
M.spret   .......... .......... .......... .......CTT GGCT...... ........TC TGTAGGCTCT G......... .........T GGGGG.TTCT
WMP/pas   .......... .......... .......... .......CTT GCCTGCTGGT TT..GGCTTC TGTAGGCTCT G......... .........T GGGGG.TTCT
SFM       .......... .......... .......... .......CTT GCCTGCTGGT TT..GGCTTC TGTAGGCTCT G......... .........T GGGGG.TTCT
XBS       .......... .......... .......... .......CTT GACTGCTGGT TT..GGCTTC TGTAGGCTCT G......... .........T GGGGG.TTCT
ZYP       .......... .......... .......... .......CTT GACTGCTGGT TT..GGCTTC TGTAGGCTCT G......... .........T GGGGG.TTCT
KAR       .......... .......... .......... .......CTT GACTGCTGGT TC..GGCTCC TGTGGGCTCG G......... .........T GGCGG.TTCT
M.car     .......... .......... .......... .......CTT GACTGCTGGT TC..AGCTCC TGTGGGCTCC G......... .........T GGCGG.TTCT
COK       .......... .......... .......... .......CTT GACTGCTGGT TT..GGCTCC TGTGGGCTCT G......... .........T GTCGG.TTCT
M.dunni   .......... .......... .......... .......CTT GACTGCTGGT TT..GGCTCC TGTGGGCTCT G......... .........T GGGGG.TTCT
PTX       ACAGCAACTC GCTTGACTG. .......... .......ATG GTTTGTCGGT TT..GGCTCC TGTGGGCCCT G......... .........T GGCGG.....
M.minuto  A.AGAAAGAA AGAAAA..AA AGCCGGGTGT GGTGGCGCTC ACCCCTTTAA TCCCAGCATN TGTGAGGTA. GAGGCAGGTG .GATTTCTGA GTTCGAGGCC
FAM       ACAACAACAA AAAAGTTTAA AGCCAACAGC AGCTG.GCTT GACTGATGGT TT..GGCTCC TGTGGGCTCT G......... .........T GGCGG.TTCT
```

```
          1601       1611       1621       1631       1641       1651       1661       1671       1681       1691      1700
          |          |          |          |          |          |          |          |          |          |         |
     AKR  ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
    BDP/j ..........  .......... .......... .......... .......... .......... .......... .......... .......... ..........
    LG/J  ..........  .......... .......... .......... .......... .......... .......... .......... .......... ..........
P.atteck  ..........  .......... .......... .......... .......... .......... .......... .......... .......... ..........
     RBA  ..........  .......... .......... .......... .......... .......... .......... .......... .......... ..........
    RF/J  ..........  .......... .......... .......... .......... .......... .......... .......... .......... ..........
   SF/cam ..........  .......... .......... .......... .......... .......... .......... .......... .......... ..........
   SK/cam ..........  .......... .......... .......... .......... .......... .......... .......... .......... ..........
    ST/BJ ..........  .......... .......... .......... .......... .......... .......... .......... .......... ..........
    SWR/J ..........  .......... .......... .......... .......... .......... .......... .......... .......... ..........
      MBT ..........  .......... .......... .......... .......... .......... .......... .......... .......... ..........
      BZO ..........  .......... .......... .......... .......... .......... .......... .......... .......... ..........
      DBA ..........  .......... .......... .......... .......... .......... .......... .......... .......... ..........
      WSB ..........  .......... .......... .......... .......... .......... .......... .......... .......... ..........
      129 ..........  .......... .......... .......... .......... .......... .......... .......... .......... ..........
    SKIVE ..........  .......... .......... .......... .......... .......... .......... .......... .......... ..........
NZW/lacj  ..........  .......... .......... .......... .......... .......... .......... .......... .......... ..........
NZB/binj  ..........  .......... .......... .......... .......... .......... .......... .......... .......... ..........
M.mus_cz  ..........  .......... .......... .......... .......... .......... .......... .......... .......... ..........
M.mus_cz  ..........  .......... .......... .......... .......... .......... .......... .......... .......... ..........
    C57BL CTCTCTCTCT CCCTGGA.AA CATCAAAGCA ACAGTTAAAG AGCTTAAAAT CAGGAGGAGG AGCCCCCCCC CCCCCCC..G CAGTGGAGAC AGAGGACAGA
  PRO/1rej CTCTCTCTCT CCCTGGA.AA CATCAAAGCA ACAGTTAAAG AGCTTAAAAT CAGGAGGAG. CCCCCCCCCC CCCCCCC..G CAGTGGAGAC AGAGGACAGA
    FvB/NJ CTCTCTCTCT CCCTGGA.AA CATCAAAGCA ACAGTTAAAG AGCTTAAAAT CAGGAGGAG. CCCCCCCCCC CC.......G CAGTGGAGAC AGAGGACAGA
     I/LNJ CTCTCTCTCT CCCTGGA.AA CATCAAAGCA ACAGTTAAAG AGCTTAAAAT CAGGAGGAG. CCCCCCCCCC CCCCC....G CAGTGGAGAC AGAGGACAGA
  RIII/dom CTCTCTCTCT CCCTGGA.AA CATCAAAGCA ACAGTTAAAG AGCTTAAAAT CAGGAGGAG. CCCCCCCCCC CCCC.....G CAGTGGAGAC AGAGGACAGA
  SEC/1rej CTCTCTCTCT CCCTGGA.AA CATCAAAGCA ACAGTTAAAG AGCTTAAAAT CAGGAGGAG. CCCCCCCCCC CCCC.....G CAGTGGAGAC AGAGGACAGA
      MOL CTCTCTCTCT CCCTGGA.AA CATCAAAGCA ACAGTTAAAG AGCTTAAAAT CAGGAGGAG. CCCCCCCCCC CCCCC....G CAGTGGAGAC AGAGGACAGA
   MOLD/RK CTCTCTCTCT CCCTGGA.AA CATCAAAGCA ACAGTTAAAG AGCTTAAAAT CAGGAGGAG. CCCCCCCCCC CCCCC....G CAGTGGAGAC AGAGGACAGA
      BIR CTCTCTCTCT CCCTGGA.AA CATCAAAGCA ACAGTTAAAG AGCTTAAAAT CAGGAGGAG. CCCCCCCCCC CCCCC....G CAGTGGAGAC AGAGGACAGA
      CIM CTCTCTCTCT CCCTGGA.AA CATCAAAGCA ACAGTTAAAG AGGTTAAAAT CAGGAGGAG. CCCCCCCCCC CCC......G CAGTGGAGAC AGAGGACAGA
   M.spret CTCTCTCTCT CCCTGGA.AA CATTAAAGCA ACAGTTAAAG AGCTTAAAAT CAGGAGGAG. CCCCCCCCCC CCCCCC...G CAGTGGAGAC AGAGGACAGA
   WMP/pas CTCTCTCTCT CCCTGGA.AA CATTAAAGCA ACAGTTAAAG AGCTTAAAAT CAGGAGGAGC CCCCCCCCCC CCCCCCC..G CAGTGGAGAC AGAGGACAGA
      SFM CTCTCTCTCT CCCTGGA.AA CATTAAAGCA ACAGTTAAAG AGCTTAAAAT CAGGAGGAG. CCCCCCCCCC CCCCCC...G CAGTGGAGAC AGAGGACAGA
      XBS CTCTCTCTCT CCCTGGA.AA CATTAAAGCA ACAGTTAAAG AGCTTAAAAT CAGGAGGAG. CCCCCCCCCC CCACCCCC.G CAGTGGAGAC AGAGGACAGA
      ZYP CTCTCTCTCT CCCTGGA.AA CATTAAAGCA ACAGTTAAAG AGCTTAAAAT CAGGAGGAG. CCCCCCCCCA CCCCC....G CAGTGGAGAC AGAGGACAGA
      KAR CTCTCTCTCT CCCTGGA.AA CATCAAAGCA ACAGTTAAAG AGCTTAAAAT CAGGAGGAGG CCCC..... .........G CAGTGGAGAC AGAGGACAGA
    M.car CTCTCTCTCC CCCTGGA.AA CATCAAAGCA ACAGTTAAAG AGCTTAAAAT CAGGAGGAGG TCCC..... .........G CAGTGGAGAC AGAGGACAGA
      COK CTCTCTCTCT CCCTGGA.AA CATCAAAGCA ACAGTTAAAG AGCTTAAAAT CAGGA...GG TCCC..... .........G CAGTGGAGAC AGAGGACAGA
   M.dunni CTCTCTTTCT CCCTGGA.AA CATTAAATCA ATAGTTAAAG AGCTTAAAAT CAGGAGGAGG CCCCCCC... .........G CAGTGGAGAC AGAGGACAGA
      PTX ..TTCTTTCT GCCT.GTGAA CATGAAAGCA TCAGTT..AG AGTTTAAACT CAGGAGGATG CCCCTCCCGT CCCCCAACA. .....GAGAC AGAGGACTGA
  M.minuto AGCCTGGTCT ACAAAGTGAG T.TCCAGG.. ACAGCCA..G GGCTATACA. ...GAGAAAC CCTGTCTC.. ....AAAAAA CTAAAAAAAA AAAACCAAAC
      FAM CTCTCTCTCT CCCTGGA.AA CATCAAAGCA ACAGTTAAAG AGCTTAAAAT CAGGAGGAGG CCCC..... .........G CAGTGGAGAC AGAGGACAGA
```

```
          1701       1711       1721       1731       1741       1751       1761       1771       1781       1791      1800
          |          |          |          |          |          |          |          |          |          |         |
     AKR  ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
   BDP/j  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   LG/J   .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
P.atteck  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     RBA  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    RF/J  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
  SF/cam  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
  SK/cam  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   ST/BJ  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   SWR/J  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     MBT  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     BZO  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     DBA  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     WSB  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     129  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   SKIVE  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
NZW/lacj  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
NZB/binj  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
M.mus_cz  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
M.mus_cz  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   C57BL  GATGATTTTT T.GCT..... .......... .......... .......... .......... .......... .......... .......... ..........
PRO/1rej  GATGATTTTT T.GCT..... .......... .......... .......... .......... .......... .......... .......... ..........
  FvB/NJ  GATGATTTTT T.GCT..... .......... .......... .......... .......... .......... .......... .......... ..........
   I/LNJ  GATGATTTTT T.GCT..... .......... .......... .......... .......... .......... .......... .......... ..........
RIII/dom  GATGATTTTT T.GCT..... .......... .......... .......... .......... .......... .......... .......... ..........
SEC/1rej  GATGATTTTT T.GCT..... .......... .......... .......... .......... .......... .......... .......... ..........
     MOL  GATGATTTTT T.GCT..... .......... .......... .......... .......... .......... .......... .......... ..........
 MOLD/RK  GATGATTTTT T.GCT..... .......... .......... .......... .......... .......... .......... .......... ..........
     BIR  GATGATTTTT T.GCT..... .......... .......... .......... .......... .......... .......... .......... ..........
     CIM  GATGATTTTT T.GCT..... .......... .......... .......... .......... .......... .......... .......... ..........
 M.spret  GATGATATTT .GGCTCTTTC TTTCTTTCTT TCTTTCTTTC TTTCTTTCTT TCTTTCTTTC TTTCTTTCTT TCTTTCTTTC TTTCTTTCTT TCTTTCTTTC
 WMP/pas  GATGATATTT .GGCTCTTTC TTTCTTTCTT TCTTTCTTTC TTTCTTTCTT TCTTTCTTTC TTTCT..... .......... .......... ..........
     SFM  GAGGATATTT .GGCTCTTTC TTTCTTTTTC TTTCTTTCTT TCTTTCTTTC TTTCTTTCTT TCTTTCTTTC TTTCTTTCTT TCTTTCTTTC TTCTTTCTTT
     XBS  GATGATATTT T.GCT..... .......... .......... .......... .......... .......... .......... .......... ..........
     ZYP  GATGATATTT T.GCT..... .......... .......... .......... .......... .......... .......... .......... ..........
     KAR  GATGATTTTT T.GCT..... .......... .......... .......... .......... .......... .......... .......... ..........
   M.car  GATGATTTTT TTGCT..... .......... .......... .......... .......... .......... .......... .......... ..........
     COK  GATGATTTTT T.GCT..... .......... .......... .......... .......... .......... .......... .......... ..........
 M.dunni  GATGATATTT T.GCT..... .......... .......... .......... .......... .......... .......... .......... ..........
     PTX  GACGAGGTTT .GGCT..... .......... .......... .......... .......... .......... .......... .......... ..........
M.minuto  CA.AAACAAA ACAGTTAAAA TCCCACAGTG GCTGGCTACA GATGTTGGCT TGACTCCTGT GGGCTCTGTG GCAGTTATCT CTCTCTGCCT GTGAACATCA
     FAM  GATGATTTTT T.GCT..... .......... .......... .......... .......... .......... .......... .......... ..........
```


1801 1811 1821 1831 1841 1851 1861 1871 1881 1891 1900
AKR
BDP/j
LG/J
P.atteck
RBA
RF/J
SF/cam
SK/cam
ST/BJ
SWR/J
MBT
BZO
DBA
WSB
129
SKIVE
NZW/lacj
NZB/binj
M.mus_cz
M.mus_cz
C57BL
PRO/1rej
FvB/NJ
I/LNJ
RIII/dom
SEC/1rej
MOL
MOLD/RK
BIR
CIM
M.spret T.....
WMP/pas
SFM CTTTCTTTCT TTCTTTCTTT CTTTCTTTCT TTCTTTCTTT CTTTCTTTCT TTCTTTCTTT CTTTCTTTCT TTCTTTCTTT CTTTCTTTCT TTCTTTCTTT
XBS
ZYP
KAR
M.car
COK
M.dunni
PTX
M.minuto AAGCATCAGA TAAAGAGTTT AAATTCAGNA GGATGCCCAG CAATAGAGAC AGAACAGAGA CGATTCGATT CTCTTTCTTT CTTTCTTTCT TTCTTTCTTT
FAM

```
          1901       1911       1921       1931       1941       1951       1961       1971       1981       1991      2000
          |          |          |          |          |          |          |          |          |          |         |
AKR       ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
BDP/j     .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
LG/J      .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
P.atteck  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
RBA       .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
RF/J      .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
SF/cam    .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
SK/cam    .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
ST/BJ     .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
SWR/J     .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
MBT       .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
BZO       .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
DBA       .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
WSB       .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
129       .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
SKIVE     .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
NZW/lacj  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
NZB/binj  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
M.mus_cz  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
M.mus_cz  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
C57BL     .......... .......... ...TTCTTTC AACATAGATT TTAGAGCC.A GAGGTGGGTT ACCTTCCTAT GATACCCCTG ATGCGCCACA AACCCGAGAA
PRO/1rej  .......... .......... ...TTCTTTC AACATAGATT TTAGAGCC.A GAGGTGGGTT ACCTTCCTAT GATACCCCTG ATGCGCCACA AACCCGAGAA
FvB/NJ    .......... .......... ...TTCTTTC AACATAGATT TTAGAGCC.A GAGGTGGGTT ACCTTCCTAT GATACCCCTG ATGCGCCACA AACCCGAGAA
I/LNJ     .......... .......... ...TTCTTTC AACATAGATT TTAGAGCC.A GAGGTGGGTT ACCTTCCTAT GATACCCCTG ATGCGCCACA AACCCGAGAA
RIII/dom  .......... .......... ...TTCTTTC AACATAGATT TTAGAGCC.A GAGGTGGGTT ACCTTCCTAT GATACCCCTG ATGCGCCACA AACCCGAGAA
SEC/1rej  .......... .......... ...TTCTTTC AACATAGATT TTAGAGCC.A GAGGTGGGTT ACCTTCCTAT GATACCCCTG ATGCGCCACA AACCCGAGAA
MOL       .......... .......... ...TTCTTTC AACATAGATT TTAGAGCC.A GAGGTGGGTT ACCTTCCTAT GATACCCCTG ATGCGCCACA AACCCGAGAA
MOLD/RK   .......... .......... ...TTCTTTC AACATAGATT TTAGAGCC.A GAGGTGGGTT ACCTTCCTAT GATACCCCTG ATGCGCCACA AACCCGAGAA
BIR       .......... .......... ...TTCTTTC AACATAGATT TTAGAGCC.A GAGGTGGGTT ACCTTCCTAT GATACCCCTG ATGCGCCACA AACCCGAGAA
CIM       .......... .......... ...TTCTTTC AACATAGATT TTAGAGCC.A GAGGTGGGTT ACCTTCCTAT GATACCCCTG ATGCGCCACA AACCCGAGAA
M.spret   .......... .........C CTTTTCTTTC AACATAGATT TTAGAGCC.A GAGGTGGGTT ACCTTCCTAT GATACTCCTG ATGCGCCACA AACCCGAGAA
WMP/pas   .......... .........C CTTTTCTTTC AACATAGATT TTAGAGCC.A GAGGTGGGTT ACCTTCCTAT GATACTCCTG ATGCGCCACA AACCCGAGAA
SFM       CT........ .........C CTTTTCTTTC AACATAGATT TTAGAGCC.A GAGGTGGGTT ACCTTCCTAT GATACTCCTG ATGCGCCACA AACCCGAGAA
XBS       ......TTCT TTCTTTTTTC CTTTTCTTTC AACATAGATT TTAGAGCC.A GAGGTGGGTT ACCTTCCTAT GATACCCCTG ATGCGCCACA AACCCGAGAA
ZYP       ......TTCT TTCTTTTTTC CTTTTCTTTC AACATAGATT TTAGAGCC.A GAGGTGGGTT ACCTTCCTAT GATACCCCTG ATGCGCCACA AACCCGAGAA
KAR       ..TGCTTTCT TTCTTT.TTC CTTTTCTTTC AACATAGATT TTAGAGCC.A GAGGTGGGCT ACCTTCCTAT GATACCCCTG ATGCGACACA AACCCGAGAA
M.car     ..TGCTTTCT TTCTTT.TTC CTTTTCTTTC AACATAGATT TTAGAGCC.A GAGGTGGGTT ACCTTCCTAT GATACCCCTG ATGCGCCACA AACCCGAGAA
COK       ..TGCTTTCT TTCTTT.TTC CTTTTCTTTC AACATAGATT TTAGAGCC.A GAGGTGGGTT ACCTTCCTAT GATACCCCTG ATACGCCACA AACCCGAAAA
M.dunni   ..TACTTTCT TTCTTT.TTC CTTTCCTTTC AACATAGATT TTAGAGCCCA GAGGTGGGTT ACCTCCCTAT GATATCCCTG ATGCGCCACA AACCTGAGAA
PTX       ....CTTTCT TTCTTTTTTT CTTTTCTTTC AACATAGACT TTAGAGCC.A GAGGTGGGTT ACCTTCCTAT GATATCCCTG ATGCGCCAGA AACCCAAAAA
M.minuto  CTTTCTTTCT TTCTTT.TT. CTTTTCTTTC AACATAGATT TTAGAGCCC. GAGGTGGGTT ACCTTCCTAT GATATCCCTG ATGCACCACA AACCCAAAAA
FAM       ..TTCTTTCT TTCTTT.TTC CTTTTCTTTC AACATAGATT TTAGAGCC.A GAGGTGGGTT GCCTTCCTAT GATACCCCTG ATGTGCCACA AACCCGAGAA
```

```
          2001       2011       2021       2031       2041       2051       2061       2071       2081       2091      2100
          |          |          |          |          |          |          |          |          |          |         |
     AKR  ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
   BDP/j  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   LG/J   .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
P.atteck  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     RBA  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    RF/J  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
  SF/cam  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
  SK/cam  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   ST/BJ  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   SWR/J  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     MBT  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     BZO  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     DBA  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     WSB  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     129  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   SKIVE  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
NZW/lacj  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
NZB/binj  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
M.mus_cz  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
M.mus_cz  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    C57BL .GA.CTATTG CCACTGGTTG GTTGCAAACA GGAGCTAGAT TGTAAGATCC CAACGCTGAC AA...CATCA CAGTGCCTTG AACCTAGGAC GCAGGGGAGT
  PRO/1rej .GA.CTATTG CCACTGGTTG GTTGCAAACA GGAGCTAGAT TGTAAGATCC CAACGCTGAC AA...CATCA CAGTGCCTTG AACCTAGGAC GCAGGGGAGT
   FvB/NJ .GA.CTATTG CCACTGGTTG GTTGCAAACA GGAGCTAGAT TGTAAGATCC CAACGCTGAC AA...CATCA CAGTGCCTTG AACCTAGGAC GCAGGGGAGT
    I/LNJ .GA.CTATTG CCACTGGTTG GTTGCAAACA GGAGCTAGAT TGTAAGATCC CAACGCTGAC AA...CATCA CAGTGCCTTG AACCTAGGAC GCAGGGGAGT
  RIII/dom .GA.CTATTG CCACTGGTTG GTTGCAAACA GGAGCTAGAT TGTAAGATCC CAACGCTGAC AA...CATCA CAGTGCCTTG AACCTAGGAC GCAGGGGAGT
  SEC/1rej .GA.CTATTG CCACTGGTTG GTTGCAAACA GGAGCTAGAT TGTAAGATCC CAACGCTGAC AA...CATCA CAGTGCCTTG AACCTAGGAC GCAGGGGAGT
     MOL  .GA.CTATTG CCACTGGTTG GTTGCAAACA GGAGCTAGAT TGTAAGATCC CAACGCTGAC AA...CATCA CAGTGCCTTG AACCTAGGAC GCAGGGGAGT
   MOLD/RK .GA.CTATTG CCACTGGTTG GTTGCAAACA GGAGCTAGAT TGTAAGATCC CAACGCTGAC AA...CATCA CAGTGCCTTG AACCTAGGAC GCAGGGGAGT
     BIR  .GA.CTATTG CCACTGGTTG GTTGCAAACA GGAGCTAGAT TGTAAGATCC CAACGCTGAC AA...CATCA CAGTGCCTTG AACCTAGGAC GCAGGGGAGT
     CIM  .GA.CTATTG CCACTGGTTG GTTGCAAACA GGAGCTAGAT TGTAAGATCC CAACGCTGAC AA...CATCA CAGTGCCTTG AACCTAGGAC GCAGGGGAGT
   M.spret .GA.CTATTG CCACTGGTTG GTTGCAAACA GGAGCTAGAT TGTAAGATTC CAACGCTGAC GATGACATCG CAGTGCCTTG AACCTAGGAC GCAGGGGAGT
   WMP/pas .GA.CTATTG CCACTGGTTG GTTGCAAACA GGAGCTAGAT TGTAAGATTC CAACGCTGAC GATGACATCG CAGTGCCTTG AACCTAGGAC GCAGGGGAGT
     SFM  .GA.CTATTG CCACTGGTTG GTTGCAAACA GGAGCTAGAT TGTAAGATTC CAACGCTGAC GATGACATCG CAGTGCCTTG AACCTAGGAC GCAGGGGAGT
     XBS  .GA.CTATTG CCACTGGTTG GTTGCAAACA GGAGCTAGAT TGTAAGATTC CAACGCTGAC GA...CATCA CAGTGCCTTG AACCTAGGAC GCAGGGGAGT
     ZYP  .GA.CTATTG CCACTGGTTG GTTGCAAACA GGAGCTAGAT TGTAAGATTC CAACGCTGAC GA...CATCA CAGTGCCTTG AACCTAGGAC GCAGGGGAGT
     KAR  .GA.TTATNG CCACTGGTTG GTTGCAAACA GGAACTAGAT TATAAGATCC CAACGCTGAC GG...CATCA CAGTGCCTTG AACCTAGGAC GCAGGGGAGT
    M.car .GA.TTATTG CCACTGGTTG GTTGCAAACA GGAACTAGAC TATAAGATCC CAACGCTGAC GA...CATCA CAGTGCCTTG AACCTAGGAC GCAGGGGAGT
     COK  .GA.TTATTG CCACTGGTTG GTTGCAAACA GGAACTAGAC TATAAGATCC CAATGCTGAC GA...CATCA CAGTGCCTTG AACCTAGGAC ACAGGGGAGT
   M.dunni .GAATTATTG CCACTGGTTG GTTGCAAACA GGAACTAGAT TATAAGATCC CAACGCTGAC GA...CATCA CAGTGCCTTG AACCTAGGAC GCAGGGGAAT
     PTX  AGA.TTATTG CCACTGGTCG GTTGCAAACA GGAACTAGAT GATAAGATCC TAACACTGAC GC...CATC. ..GTGCCTTG AAAGTA.... ..AGGGGAGT
   M.minuto .GA.TTATTG CCACTGGTTG GTTGCAAACA GGAACTAGAT TATAAGAGCC TAACGCTGAC AA...CATCA C.GTGCCTTG AACCTAGGAC ATAGGGGAGT
     FAM  .GA.TTACTG CCACTGGTTG GTTGCAAACA GGAGCTAGAT TATAAGATCC CAACGCTGAC GA...CATCA CAGTGCCTTG AACCTAGGAC GCAGGGGAGT
```

```
          2101       2111       2121       2131       2141       2151       2161       2171       2181       2191      2200
          |          |          |          |          |          |          |          |          |          |         |
     AKR  ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
   BDP/j  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    LG/J  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
P.atteck  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     RBA  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    RF/J  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   SF/cam  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   SK/cam  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    ST/BJ  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    SWR/J  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     MBT  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     BZO  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     DBA  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     WSB  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     129  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   SKIVE  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
NZW/lacj  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
NZB/binj  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
M.mus_cz  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
M.mus_cz  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   C57BL  A...AGGCAT GGACTGTGCT CAATGAAGGC TTCCT....G CTGGCTGGC. ATTCCTAGTG C.TAGAAAGT GCTGTGCA.G ATTACTGGTG TCCTTGTTTC
PRO/1rej  A...AGGCAT GGACTGTGCT CAATGAAGGC TTCCT....G CTGGCTGGC. ATTCCTAGTG C.TAGAAAGT GCTGTGCA.G ATTACTGGTG TCCTTGTTTC
  FvB/NJ  A...AGGCAT GGACTGTGCT CAATGAAGGC TTCCT....G CTGGCTGGC. ATTCCTAGTG C.TAGAAAGT GCTGTGCA.G ATTACTGGTG TCCTTGTTTC
   I/LNJ  A...AGGCAT GGACTGTGCT CAATGAAGGC TTCCT....G CTGGCTGGC. ATTCCTAGTG C.TAGAAAGT GCTGTGCA.G ATTACTGGTG TCCTTGTTTC
RIII/dom  A...AGGCAT GGACTGTGCT CAATGAAGGC TTCCT....G CTGGCTGGC. ATTCCTAGTG C.TAGAAAGT GCTGTGCA.G ATTACTGGTG TCCTTGTTTC
SEC/1rej  A...AGGCAT GGACTGTGCT CAATGAAGGC TTCCT....G CTGGCTGGC. ATTCCTAGTG C.TAGAAAGT GCTGTGCA.G ATTACTGGTG TCCTTGTTTC
     MOL  A...AGGCAT GGACTGTGCT CAATGAAGGC TTCCT....G CTGGCTGGC. ATTCCTAGTG C.TAGAAAGT GCTGTGCA.G ATTACTGGTG TCCTTGTTTC
 MOLD/RK  A...AGGCAT GGACTGTGCT CAATGAAGGC TTCCT....G CTGGCTGGC. ATTCCTAGTG C.TAGAAAGT GCTGTGCA.G ATTACTGGTG TCCTTGTTTC
     BIR  A...AGGCAT GGACTGTGCT CAATGAAGGC TTCCT....G CTGGCTGGC. ATTCCTAGTG C.TAGAAAGT GCTGTGCA.G ATTACTGGTG TCCTTGTTTC
     CIM  A...AGGCAT GGACTGTGCT CAATGAAGGC TTCCT....G CTGGCTGGC. ATTCCTAGTG C.TAGAAAGT GCTGTGCA.G ATTACTGGTG TCCTTGTTTC
 M.spret  A...AGGCGT GGACCGTGCT GAGTGAAGGC TTCCT....G CTGGCTGGC. ATTCCT.GTG C.TAGAAAGT GCTGTGCA.C ACTACTGGTG TCCTTGTTTC
 WMP/pas  A...AGGCGT GGACCGTGCT GAGTGAAGGC TTCCT....G CTGGCTGGC. ATTCCT.GTG C.TAGAAAGT GCTGTGCA.C ACTACTGGTG TCCTTGTTTC
     SFM  A...AGGCGT GGACCGTGCT GAGTGAAGGC TTCCT....G CTGGCTGGC. ATTCCT.GTG C.TAGAAAGT GCTGTGCA.C ACTACTGGTG TCCTTGTTTC
     XBS  A...AGGCGT GGACTGTGCT GAATGAAGGC TTCCT....G CTGGCTGGC. ATTCCTAGTG C.TAGAAAGT GCTGTGCA.C ACTACTGGTG TCCTTGTTTC
     ZYP  A...AGGCGT GGACTGTGCT GAATGAAGGC TTCCT....G CTGGCTGGC. ATTCCTAGTG C.TAGAAAGT GCTGTGCA.C ACTACTGGTG TCCTTGTTTC
     KAR  A...AGGCAT GGACTGTGCT GAATGAAGGC TTCCT....G CTGGCTGGC. ATTCCTAGTG C.TAGAAAGT GCCGTGCA.G ACTACTGCTG TCCTTGTTTC
   M.car  G...AGGCAT GGACTGTGCT GAATGAAGGC TTCCT....G CTGGCTGGC. ATTCCTAGTG C.TAGAAAGT GCCATGCA.G ACTACTGCTG TCCTTGTTTC
     COK  C...AGGCAT GGACTGTGCT GAAAGAAGGC TTCCT....G CTGGCTGGC. ATTCCTAGTG C.TAGAAAGT GCTGTGCA.G ACTACTGTTG TCCTTGTTTC
 M.dunni  A...AGAC.T GGACTGTGCT GAATGAAGGC TTCCT....G CTGGCTGGCC ATTCCTAATG CCTAGAAAGT GCTGTGCAAG .CCACTGTTG TCTTTGTCTC
     PTX  A...ATGCAC GGACTGTGCT AAATGAAGGC TTCCTCCCTG CTGGCTGGC. ATTCCTAGTG C.TAGAAAGT GCTGTGCA.G ACTGCTGTTG TCCTTGTTCC
M.minuto  AGTAACGCAT GGACTGTGCT GAGTGAAGGC TTCCTCCCTG CTGGCTGGC. ATTCCTAGTG C.TAGAGAGT GCTGTGCAGG .CTGCTGTTG TCCCTGTTCC
     FAM  A...AGGCAT GGACTGTGCT GAATGAAGGC TTCCT....G CTGGCTGGC. ATTCCTAGTG C.TAGAAAGT GCTGTGCA.G ACTACTGTTG TCCTGGTTTC
```

```
          2201       2211       2221       2231       2241       2251       2261       2271       2281       2291      2300
          |          |          |          |          |          |          |          |          |          |         |
      AKR ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
    BDP/j ..........  .........  .........  .........  .........  .........  .........  .........  .........  .........
     LG/J ..........  .........  .........  .........  .........  .........  .........  .........  .........  .........
P.atteck ..........  .........  .........  .........  .........  .........  .........  .........  .........  .........
      RBA ..........  .........  .........  .........  .........  .........  .........  .........  .........  .........
     RF/J ..........  .........  .........  .........  .........  .........  .........  .........  .........  .........
   SF/cam ..........  .........  .........  .........  .........  .........  .........  .........  .........  .........
   SK/cam ..........  .........  .........  .........  .........  .........  .........  .........  .........  .........
     ST/BJ ..........  .........  .........  .........  .........  .........  .........  .........  .........  .........
     SWR/J ..........  .........  .........  .........  .........  .........  .........  .........  .........  .........
      MBT ..........  .........  .........  .........  .........  .........  .........  .........  .........  .........
      BZO ..........  .........  .........  .........  .........  .........  .........  .........  .........  .........
      DBA ..........  .........  .........  .........  .........  .........  .........  .........  .........  .........
      WSB ..........  .........  .........  .........  .........  .........  .........  .........  .........  .........
      129 ..........  .........  .........  .........  .........  .........  .........  .........  .........  .........
    SKIVE ..........  .........  .........  .........  .........  .........  .........  .........  .........  .........
 NZW/lacj ..........  .........  .........  .........  .........  .........  .........  .........  .........  .........
 NZB/binj ..........  .........  .........  .........  .........  .........  .........  .........  .........  .........
 M.mus_cz ..........  .........  .........  .........  .........  .........  .........  .........  .........  .........
 M.mus_cz ..........  .........  .........  .........  .........  .........  .........  .........  .........  .........
    C57BL CACT.CAG.C TATGGACTCT AAATGTTAAT GTCATCAGCA CTA.CAGG.C AGCTAGCTG. .........  .........  .........  .........T
 PRO/1rej CACT.CAG.C TATGGACTCT AAATGTTAAT GTCATCAGCA CTA.CAGG.C AGCTAGCTG. .........  .........  .........  .........T
   FvB/NJ CACN.CAG.C TATGGACTCT AAATGTTAAT GTCATCAGCA CTA.CAGG.C AGCTAGCTG. .........  .........  .........  .........T
    I/LNJ CACT.CAG.C TATGGACTCT AAATGTTAAT GTCATCAGCA CTA.CAGG.C AGCTAGCTG. .........  .........  .........  .........T
 RIII/dom CACT.CAG.C TATGGACTCT AAATGTTAAT GTCATCAGCA CTA.CAGG.C AGCTAGCTG. .........  .........  .........  .........T
 SEC/1rej CACT.CAG.C TATGGACTCT AAATGTTAAT GTCATCAGCA CTA.CAGG.C AGCTAGCTG. .........  .........  .........  .........T
      MOL CACT.CAG.C TATGGACTCT AAATGTTAAT GTCATCAGCA CTA.CAGG.C AGCTAGCTG. .........  .........  .........  .........T
  MOLD/RK CACT.CAG.C TATGGACTCT AAATGTTAAT GTCATCAGCA CTA.CAGG.C AGCTAGCTG. .........  .........  .........  .........T
      BIR CACT.CAG.C TATGGACTCT AAATGTTAAT GTCATCAGCA CTA.CAGG.C AGCTAGCTG. .........  .........  .........  .........T
      CIM CACT.CAG.C TATGGACTCT AAATGTTAAT GTCATCAGCA CTA.CAGG.C AGCTAGCTG. .........  .........  .........  .........T
  M.spret CACT.CAG.C TATGGACTCT AAATGTTAAT GTCATCAGCA CTA.CAGG.C AGCTAGCTG. .........  .........  .........  .........T
  WMP/pas CACT.CAGA. TATGGACTCT AAATGTTAAT GTCATCAGCA CTA.CAGG.C AGCTAGCTG. .........  .........  .........  .........T
      SFM CACT.CAG.C TATGGACTCT AAATGTTAAT GTCATCAGCA CTA.CAGG.C AGCTAGCTG. .........  .........  .........  .........T
      XBS CACT.CAG.C TATGGACTCT AAATGTTAAC GTCATCAGCA CTA.CAGG.C AGCTAGCTG. .........  .........  .........  .........T
      ZYP CACT.CAG.C TATGGACTCT AAATGTTAAC GTCATCAGCA CTA.CAGG.C AGCTAGCTG. .........  .........  .........  .........T
      KAR CACT.CAG.C TATGGACTCT AAATGTTAAT GTCATCAGCA CTA.CAGG.C AGCT.G.... .........  .........  .........  .........T
    M.car CACT.CAG.C TATGGACTCT AAATGTTAAT GTCATCAGCA CTA.CAGG.C AGCT.G.... .........  .........  .........  .........T
      COK CACT.CAG.C TATGGACTCT AAATGTTAAT GTCATCAGCA CTA.CAGG.C AGCT.G.... .........  .........  .........  .........T
  M.dunni CACTTCAGGC AATGGACTGT AAAAGTTAAA GTTTTCAGCA ATAACAGGGC AGCT.G.... .........  .........  .........  .........T
      PTX CACT.CAG.C TATGGACTCT AAATGTTAAT GTCATCAGTA CTA.CAGG.C TGCT.GCTGC TGATGATGAT AATGATGATA ATGATGATGA CGATGATGAT
 M.minuto CACTCCA..C TATGGACTCT AAATGTTAAT GTCATCAGCA CCA.CAGG.C AGCT.G.... .........  .........  .........  .........T
      FAM CACT.CAG.C TATGGACTCT AAATGTTAAT GTCATCAGCA CTA.CAGG.C AGCT.G.... .........  .........  .........  .........T
```

```
          2301       2311       2321       2331       2341       2351       2361       2371       2381       2391      2400
          |          |          |          |          |          |          |          |          |          |         |
     AKR  ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
   BDP/j  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    LG/J  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
P.atteck  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     RBA  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    RF/J  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
  SF/cam  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
  SK/cam  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   ST/BJ  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   SWR/J  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     MBT  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     BZO  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     DBA  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     WSB  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     129  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   SKIVE  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
NZW/lacj  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
NZB/binj  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
M.mus_cz  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
M.mus_cz  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    C57BL AGGGGGTT.G GG.T.TGGTC TGGTGCTG.. ..CTTG.TAG TCAAAAGCTA AATTCTGTTT CCCAAGATCC GATTGCTCCA AGGTGAGCCG ATCCTCACGT
  PRO/1rej AGGGGGTT.G GG.T.TGGTC TGGTGCTG.. ..CTTG.TAG TCAAAAGCTA AATTCTGTTT CCCAAGATCC GATTGCTCCA AGGTGAGCCG ATCCTCACGT
    FvB/NJ AGGGGGTT.G GG.T.TGGTC TGGTGCTG.. ..CTTG.TAG TCAAAAGCTA AATTCTGTTT CCCAAGATCC GATTGCTCCA AGGTGAGCCG ATCCTCACGT
     I/LNJ AGGGGGTT.G GG.T.TGGTC TGGTGCTG.. ..CTTG.TAG TCAAAAGCTA AATTCTGTTT CCCAAGATCC GATTGCTCCA AGGTGAGCCG ATCCTCACGT
  RIII/dom AGGGGGTT.G GG.T.TGGTC TGGTGCTG.. ..CTTG.TAG TCAAAAGCTA AATTCTGTTT CCCAAGATCC GATTGCTCCA AGGTGAGCCG ATCCTCACGT
  SEC/1rej AGGGGGTT.G GG.T.TGGTC TGGTGCTG.. ..CTTG.TAG TCAAAAGCTA AATTCTGTTT CCCAAGATCC GATTGCTCCA AGGTGAGCCG ATCCTCACGT
     MOL AGGGGGTT.G GG.T.TGGTC TGGTGCTG.. ..CTTG.TAG TCAAAAGCTA AATTCTGTTT CCCAAGATCC GATTGCTCCA AGGTGAGCCG ATCCTCACGT
   MOLD/RK AGGGGGTT.G GG.T.TGGTC TGGTGCTG.. ..CTTG.TAG TCAAAAGCTA AATTCTGTTT CCCAAGATCC GATTGCTCCA AGGTGAGCCG ATCCTCACGT
     BIR AGGGGGTT.G GG.T.TGGTC TGGTGCTG.. ..CTTG.TAG TCAAAAGCTA AATTCTGTTT CCCAAGATCC GATTGCTCCA AGGTGAGCCG ATCCTCACGT
     CIM AGGGGGTT.G GG.T.TGGTC TGGTGCTG.. ..CTTG.TAG TCAAAAGCTA AATTCTGTTT CCCAAGATCC GATTGCTCCA AGGTGAGCCG ATCCTCACGT
  M.spret AGGGGGTTTG GG.T.TGGTC TGGTGCTG.. ..CTTG.TAG TCAAAAGCAA AATTCTGTTT CCCAAGATCC GATTGCTCCA AGGTGAGCCA ATCCTCACGT
  WMP/pas AGGGGGTT.. GGGT.TGGTC TGGTGCTG.. ..CTTG.TAG TCAAAAGCAA AATTCTGTTT CCCAACATCC GATTGCTCCA AGGTGAGCCA ATCCTCACGT
     SFM AGGGGGTTTG GGGTCCGGTC TGGTGCTG.. ..CTTG.TAT TCAAAAGCAA AATTCTGTTT CCCAAGTTCC GATTGCTCCA AGGTGAGCCA ATCCTCACGT
     XBS AGGGGGTT.G GG.T.TGGTC TGGTGCTG.. ..CTTG.TAG TCAAAAGCAA AATTCTGTTT CCCAAGATCC GATTGCTCCA AGGTGAGCCG ATCCTCACGT
     ZYP AGGGGGTT.G GG.C.TGGTC TGGTGCTG.. ..CTTG.TAG TCAAAAGCAA AATTCTGTTT CCCAAGATCC GATTGCTCCA AGGTGAGCCG ATCCTCACGT
     KAR AGGGGGGT.G GG.TN.GGTC TGGTGCTG.. ..CTTG.TAG TCAAAAGCTA AATTCTATTT CCCAAGATCC GATTGCTCCA AGGTGAGCCG ATCCTTACGT
   M.car AGGGGGGT.G GG.T.CGGTC TGGTGCTG.. ..CTTG.TAG TCAAAAGCTA AATTCTATTT CCCAAGATCC GATTGCTCCA AGGTGAGCCG ATCCTTACGT
     COK AGGGGGTT.G GG.T.TGGTC TGGTGCTG.. ..CTTG.TAG TCAAAAGCTA AATTCTATTT CCCAAGATCC GATTGCTCCA AGGTGAGCCG ATCCTCACGT
  M.dunni AGGGGGTT.G GG.T.TGGTT CAGTGCTGGC TGGCTGCTAG TCAAAAGCTG AATTCTGTTT CCCAAGATCC GACTGCTCCA AGGTGAGCCG ATCCTCACGT
     PTX AGGGGGTT.G GG.T.TGGTC TCGTGCTG.. ..CTTG.TAG TCAAATGCTA AATTCTATTT CCCAAGATGT GGTTGCTCCA AGGTGAGCCG ATCCTCACGT
  M.minuto AGGGGGTT.. GGGT.TGGTC TGGTGCTG.. ..CTTGC.AG TCAAATGCTA GATTCTGTTT CCCAAGATCT GGTTGCAACA AGGTGAGCAG ATCCTCACA.
     FAM AGGGGGTT.G GG.T.TGGTC TGGTGCTG.. ..CTTG.TAG TCAAAAGCTA AATTCTGTTT CCCAAGATCC GATTGCTCCA AGGTGAGCGG ATCCTCACGT
```

```
          2401       2411       2421       2431       2441       2451       2461       2471       2481       2491      2500
          |          |          |          |          |          |          |          |          |          |         |
     AKR  ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
   BDP/j  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    LG/J  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
P.atteck  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     RBA  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    RF/J  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
  SF/cam  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
  SK/cam  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   ST/BJ  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   SWR/J  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     MBT  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     BZO  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     DBA  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     WSB  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     129  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   SKIVE  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
NZW/lacj  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
NZB/binj  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
M.mus_cz  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
M.mus_cz  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   C57BL  ACTCCAGCCT T.CTTATAAA CCTTA..CTC CCCAGTTTGA AACTATCGGT TAAATAAAAC TGGTGATAAC CAATTACTGG CGGGGGGGGG GGGGGG....
PRO/1rej  ACTCCAGCCT TTCTTATAAA CCTTA..CTC CCCAGTTTGA AACTATCGGT TAAATAAAAC TGGTGATAAC CAATTACTGG CNGGGGGGGG GGGGGGG...
  FvB/NJ  ACTCCAGCCT TTCTTATAAA CCTTA..CTC CCCAGTTTGA AACTATCGGT TAAATAAAAC TGGTGATAAC CAATTACTGG CGGGGGGGGG GGGGGGG...
   I/LNJ  ACTCCAGCCT TTCTTATAAA CCTTA..CTC CCCAGTTTGA AACTATCGGT TAAATAAAAC TGGTGATAAC CAATTACTGG CGGGGGGGGG GGGGGGG...
RIII/dom  ACTCCAGCCT TTCTTATAAA CCTTA..CTC CCCAGTTTGA AACTATCGGT TAAATAAAAC TGGTGATAAC CAATTACTGG CGGGGGGGGG GGGGGGG...
SEC/1rej  ACTCCAGCCT TTCTTATAAA CCTTA..CTC CCCAGTTTGA AACTATCGGT TAAATAAAAC TGGTGATAAC CAATTACTGG CGGGGGGGGG GGGGGGG...
     MOL  ACTCCAGCCT TTCTTATAAA CCTTA..CTC CCCAGTTTGA AACTATCGGT TAAATAAAAC TGGTGATAAC CAATTACTGG CGGGGGGGGG GGGGG.....
 MOLD/RK  ACTCCAGCCT TTCTTATAAA CCTTA..CTC CCCAGTTTGA AACTATCGGT TAAATAAAAC TGGTGATAAC CAATTACTGG CGGGGGGGGG GGGGG.....
     BIR  ACTCCAGCCT TTCTTATAAA CCTTA..CTC CCCAGTTTGA AACTATCGGT TAAATAAAAC TGGTGATAAC CAATTACTGG CGGGGGGGGG GGG.......
     CIM  ACTCCAGCCT TTCTTATAAA CCTTA..CTC CCCAGTTTGA AACTATCGGT TAAATAAAAC TGGTGATAAC CAATTACTGG CGGGGGGGGG GGGGG.....
 M.spret  ACTCCAGCCT TTCTTATAAA CCTTA..CTC CCCAGTTTGA AACTATCGGT TAAATAAAAC TGGTGATAAC CAATTACTGG CAGGGGGGGG GGG..AAGG.
 WMP/pas  ACTCCAGCCT TTCTTATAAA CCTTA..CTC CCCAGTTTGA AACTATCGGT TAAATAAAAC TGGTGATAAC CAATTACTGG CAGGGGGGGG GGGGGAAGG.
     SFM  ACTCCAGCCT TTCTTATAAA CCTTA..CTC CCCAGTTTGA AACTATCGGT TAAATAAAAC TGGTGATAAC CAATTACTGG CAGGGGGGGG GGGGGAAGG.
     XBS  ACTCCAGCCT TTCTTATAAA CCTTA..CTC CCCAGTTTGA AACTATCGGT TAAATAAAAC TGGTGATAAC TAATTACTGG CGGGGGCGCG GGG.......
     ZYP  ACTCCAGCCT TTCTTATAAA CCTTA..CTC CCCAGTTTGA AACTATCGGT TAAATAAAAC TGGTGATAAC TAATTACTGG CGCGGGCGCG GGG.......
     KAR  ACTCCAGCCT TTCTTATAAA CCTTA..CTC CCCAGTTTGA AACTACCAGT TAAATAAAAC TGGTGATAAC CAATTACTGG CGGGGGGCGG GTGGGG.NAA
   M.car  ACTCCAGCCT TTCTTATAAA CCTTA..CTC CCCAGTTTGA AACTACCAGT TAAATAAAAC TGGTGATAAC CAATTACTGG CGGGGGGCGG GTGGGGG.AG
     COK  ACTCCAGCCT TTCTTATAAA CCTTA..CTC CCCAGTTTGA AACTATCGGT TAAATAAAAC TGGTGATAAC CAATTACTGG CAGGGGGTGG GGAGGGGTAG
 M.dunni  ACTCCAGCCT TTCTTATAAA CCTTA..CTC CCCAGTTTGA AACTATCGGT TAAATAAAAC TGGTGATAAC CAATTACTGG CGGGGGGGGG GGGGGGGGG.
     PTX  ACTCCAGCTT TTGTTATAAA CCTTACTCTC CCCAGTTTGA AACTATCGGT TAAATAGAAT TGGT...... .......... .......... ..........
M.minuto  ...CCAGCCT TTGTT..... .CATA..... ...AGT.... ..CGATCAGA TAAT...... .....ATAA. .......... .......... ..........
     FAM  ACTCCAGCCT TTCTTATAAA CCTTA..CTC CCCAGTTTGA AACTATGGGT TAAATAAAAC TGGTGATAAC CAATTACTGG CGGGGG.TGG G.GGGGG.AG
```

```
          2501       2511       2521       2531       2541       2551       2561       2571       2581       2591      2600
          |          |          |          |          |          |          |          |          |          |         |
     AKR  ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
   BDP/j  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    LG/J  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
P.atteck  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     RBA  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    RF/J  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   SF/cam  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   SK/cam  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    ST/BJ  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    SWR/J  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
      MBT  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
      BZO  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
      DBA  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
      WSB  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
      129  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    SKIVE  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
NZW/lacj  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
NZB/binj  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
M.mus_cz  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
M.mus_cz  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    C57BL  ......AGTT TGAGTTACCG GACTGGTCGA ACTAAGTCGA A....CTGAT AATATAAAGC AACATTTGAT TGTTCACATA TGCTATGAGT AGATAGATGG
PRO/1rej  ....GGAGTT TGAGTTACCG GACTGGTCGA ACTAAGTCGA A....CTGAT AATATAAAGC AACATTTGAT TGTTCACATA TGCTATGAGT AGATAGATGG
  FvB/NJ  .....GAGTT TGAGTTACCG GACTGGTCGA ACTAAGTCGA A....CTGAT AATATAAAGC AACATTTGAT TGTTCACATA TGCTATGAGT AGATAGATGG
   I/LNJ  ....GGAGTT TGAGTTACCG GACTGGTCGA ACTAAGTCGA A....CTGAT AATATAAAGC AACATTTGAT TGTTCACATA TGCTATGAGT AGATAGATGG
RIII/dom  ....GGAGTT TGAGTTACCG GACTGGTCGA ACTAAGTCGA A....CTGAT AATATAAAGC AACATTTGAT TGTTCACATA TGCTATGAGT AGATAGATGG
SEC/1rej  ......AGTT TGAGTTACCG GACTGGTCGA ACTAAGTCGA A....CTGAT AATATAAAGC AACATTTGAT TGTTCACATA TGCTATGAGT AGATAGATGG
     MOL  ......AGTT TGAGTTACCG GACTGGTCGA ACTAAGTCGA A....CTGAT AATATAAAGC AACATTTGAT TGTTCACATA TGCTATGAGT AGATAGATGG
 MOLD/RK  ......AGTT TGAGTTACCG GACTGGTCGA ACTAAGTCGA A....CTGAT AATATAAAGC AACATTTGAT TGTTCACATA TGCTATGAGT AGATAGATGG
     BIR  ......AGTT TGAGTTACCG GACTGGTCGA ACTAAGTCGA A....CTGAT AATATAAAGC AACATTTGAT TGTTCACATA TGCTATGAGT AGATAGATGG
     CIM  .CGGGGAGTT TGAGTTACCG GACTGGTCGA ACTAAGTCGA A....CTGAT AATATAAAGC AACATTTGAT TGTTCACATA TGCTATGAGT AGATAGATGG
 M.spret  .CNGGGAATT TGAATTACCG GACTGGTCGA ACTAAGTCGA A....CTGAT AATATAAAGC AACATTTGAT TGTTCACATA TGCTATGAGT AGATAGATGG
 WMP/pas  .CGGGGAATT TGAATTACCG GACTGGTCGA ACTAAGTCGA A....CTGAT AATATAAAGC AACATTTGAT TGTTCACATA TGCTATGAGT AGATAGATGG
     SFM  .CGGGGAATT TGAGTTACCG GACTGGTCGA ACTAAGTCGA A....CTGAT AATATAAAGC AACATTTGAT TGTTCACATA TGCTATGAGT AGATAGATGG
     XBS  ......AGTT TGAGTTACCG GACTGGTCGA ACTAAGTCAA A....CTGAT AATATAAAGC AACATTTGAT TGTTCACATA TGCTATGAGT AGATAGATGG
     ZYP  ......AGTT TGAGTTACCG GACTGGTCGA ACTAAGTCAA A....CTGAT AATATAAAGC AACATTTGAT TGTTCACATA TGCTATGAGT AGATAGATGG
     KAR  TGGGGGAGTT TGAGTTACAG GACTGGTCGA ACTAAGTCGA A....CTGAT AATATAAAGC AACATTTGAT TGTTCACATA TGCTATGAGT AGATAGCTGG
   M.car  TGGGGGAGTT TGAGTTACCG GACTGGTCGA ACTAAGTCGA A....CTGAT AATATAAAGC AACATTTGAT TGTTCACATA TGCTATGAGT AGATAGCTGG
     COK  TGGGGGAGTT TGAGTTACCG GACTGGTCGA ACTAAGTCGA A....CTGAT AATATAAAGC AACATTTGAT TGTTCACATA TGCTATGAGT AGATAGATGG
 M.dunni  ......AGTT TGAGTTACCG GACTGGTCGA ACTAAGTCAA A....CTGAT AATATAAAGC AACATTTGAT TGTTCACATA TGCTATGAGT AGATAGATGG
     PTX  .......GTT TGAGTTACCG GACTGGTCGA ACTAAGTCGA ATTTCCTGAT AATATAAAGC AACATTTGGT TGTTCACATA TAC..TGAGT AGATAGATGG
M.minuto  .......... .......... GACTAG..AA AC........ .......... ......AAGC AACATTTGGT TGTTCACATA TGCTATGAGT AGATAGATGG
     FAM  TGGGGGAGTT TGAGTTACCG GACTGGTCGA ACT....... .......GAT AATATAAAGC AACATTTGAT TGTTCACATA TGCTATGAGT AGATAGATGG
```

```
          2601       2611       2621       2631       2641       2651       2661       2671       2681       2691      2700
          |          |          |          |          |          |          |          |          |          |         |
     AKR  ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
   BDP/j  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   LG/J   .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
P.atteck  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     RBA  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    RF/J  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
  SF/cam  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
  SK/cam  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   ST/BJ  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   SWR/J  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     MBT  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     BZO  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     DBA  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     WSB  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     129  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   SKIVE  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
NZW/lacj  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
NZB/binj  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
M.mus_cz  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
M.mus_cz  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   C57BL  TCTTCAAAAA CCTCGGAGGC CACAGAAAAT GGCATTTAAA GGTGTTTTT. ATTAATTTAA GGCCTTTTTG ACATTGAGAC ATGTCAGCTC CTGGCAGCAC
PRO/1rej  TCTTCAAAAA CCTCGGAGGC CACAGAAAAT GGCATTTAAA GGTGTTTTT. ATTAATTTAA GGCCTTTTTG ACATTGAGAC ATGTCAGCTC CTGGCAGCAC
  FvB/NJ  TCTTCAAAAA CCTCGGAGGC CACAGAAAAT GGCATTTAAA GGTGTTTTT. ATTAATTTAA GGCCTTTTTG GCATTGAGAC ATGTCAGCTC CTGGCAGCAC
   I/LNJ  TCTTCAAAAA CCTCGGAGGC CACAGAAAAT GGCATTTAAA GGTGTTTTT. ATTAATTTAA GGCCTTTTTG ACATTGAGAC ATGTCAGCTC CTGGCAGCAC
RIII/dom  TCTTCAAAAA CCTCGGAGGC CACAGAAAAT GGCATTTAAA GGTGTTTTT. ATTAATTTAA GGCCTTTTTG ACATTGAGAC ATGTCAGCTC CTGGCAGCAC
SEC/1rej  TCTTCAAAAA CCTCGGAGGC CACAGAAAAT GGCATTTAAA GGTGTTTTT. ATTAATTTAA GGCCTTTTTG ACATTGAGAC ATGTCAGCTC CTGGCAGCAC
     MOL  TCTTCAAAAA CCTCGGAGGC CACAGAAAAT GGCATTTAAA GGTGTTTTT. ATTAATTTAA GGCCTTTTTG ACATTGAGAC ATGTCAGCTC CTGGCAGCAC
 MOLD/RK  TCTTCAAAAA CCTCGGAGGC CACAGAAAAT GGCATTTAAA GGTGTTTTT. ATTAATTTAA GGCCTTTTTG ACATTGAGAC ATGTCAGCTC CTGGCAGCAC
     BIR  TCTTCAAAAA CCTCGGAGGC CACAGAAAAT GGCATTTAAA GGTGTTTTT. ATTAATTTAA GGCCTTTTTG ACATTGAGAC ATGTCAGCTC CTGGCAGCAC
     CIM  TCTTCAAAAA CCTCGGAGGC CACAGAAAAT GGCATTTAAA GGTGTTTTT. ATTAATTTAA GGCCTTTTTG ACATTGAGAC ATGTCAGCTC CTGGCAGCAC
 M.spret  TCTTCAAAAA CCTCAGAGGC CACAGAAAAT GGC.TTTAAA GGTGTTCTT. ATTAATTTAA GGCCTTTTTG ACATTGAGAC ATGTCAGCTC CTGGCAGCGC
 WMP/pas  TCTTCAAAAA CCTCAGAGGC CACAGAAAAT GGC.TTTAAA GGTGTTCTT. ATTAATTTAA GGCCTTTTTG ACATTGAAAC ATGTCAGCTC CTGGCAGCGC
     SFM  TCTTCAAAAA CCTCAGAGGC CACAGAAAAT GGC.TTTAAA GGTGTTCTT. ATTAATTTAA GGCCTTTTTG ACATTGAGAC ATGTCAGCTC CTGGCAGCGC
     XBS  TCTCCAAAAA CCTCAGAGGC CACAGAAAAT GGCATTTAAA GATGTTCTT. ATTAATTTAA GGCCTTTTTG ACATTGAGAC ATGTCAGCTC CTGGCAGCGC
     ZYP  TCTCCAAAAA CCTCAGAGGC CACAGAAAAT GGCATTTAAA GATGTTCTT. ATTAATTTAA GGCCTTTTTG ACATTGAGAC ATGTCAGCTC CTGGCAGCGC
     KAR  TCTTCAAAAA CCTCAGAGGC CACAGAAAAT GGCATTTAAA GATGTTTTT. ATTAATTTAA GGCCTTTTTG ACATTGAGAC ATGTCAGCTC CTGGCAGCAC
   M.car  TCTTCAAAAA CCTCAGAGGC CACAGAAAAT GGCATTTAAA GATGTTTTT. ATTAATTTAA GGCCTTTTTG ACATTGAGAC ATGTCAGCTC CTGGCAGCAC
     COK  TCTTCAAAAA CCTCAGAGGC CACAGAAAAT GGCATTTAAA GATGTTTTTT ATTAATTTAA GGCCTTTTTG ACATTAAGAC ATGTCAACTC CCGGCAGCAC
 M.dunni  TCTTCAAAAA CCTCAGAGAC CACAGAAAAT GGCATTTAAA GATGTTTTT. ATTAATTTAA GGCCTTTTTG ACATTGAGAC ATGTCAGCTC CT........
     PTX  TCTTCAAAAA CCTCAGAGGC CACAGAAAAT GGCATTTAAA GATGGTTTT. ATTAGTTTAA GGCCTTTTTG ACATTGAGAC ATGTCAGCTC CTGGCAGCAC
M.minuto  TCTTCAAAAA CCTCAGAGGC CACAGAAAAT GGCATTTAAA GAT.TTTTTT ATTAACTTAA GGCCTTTTTG ACATTGAGGC ATGTCAGCTC CTGGCAGCAC
     FAM  TCTTCAAAAA CCTCAGAGGC CACAGAAAAT GGCATTTAAA GATGTTTTT. ATTAATTTAA GGCCTTTTTG ACATTGAGAC ATGTCAGCTC CTGGCAGCAC
```

```
          2701       2711       2721       2731       2741       2751       2761       2771       2781       2791       2800
          |          |          |          |          |          |          |          |          |          |          |
     AKR  ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
   BDP/j  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    LG/J  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
P.atteck  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     RBA  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    RF/J  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
  SF/cam  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
  SK/cam  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   ST/BJ  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   SWR/J  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     MBT  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     BZO  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     DBA  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     WSB  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     129  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   SKIVE  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
NZW/lacj  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
NZB/binj  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
M.mus_cz  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
M.mus_cz  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    C57BL CCC.ATTCGA CTTGGAAAAG GAAGGATAAG CATTGAAGAA TCTCCGTTAG GAGAT..... GGATTCTAAT GTAGCAAGTT GCCACTGGGC AAAGAAAATG
 PRO/1rej CCC.ATTCGA CTTGGAAAAG GAAGGATAAG CATTGAAGAA TCTCCGTTAG GAGAT..... GGATTCTAAT GTAGCAAGTT GCCACTGGGC AAAGAAAATG
   FvB/NJ CCC.ATTCGA CTTGGAAAAG GAAGGATAAG CATTGAAGAA TCTCCGTTAG GAGAT..... GGATTCTAAT GTAGCAAGTT GCCACTGGGC AAAGAAAATG
    I/LNJ CCC.ATTCGA CTTGGAAAAG GAAGGATAAG CATTGAAGAA TCTCCGTTAG GAGAT..... GGATTCTAAT GTAGCAAGTT GCCACTGGGC AAAGAAAATG
 RIII/dom CCC.ATTCGA CTTGGAAAAG GAAGGATAAG CATTGAAGAA TCTCCGTTAG GAGAT..... GGATTCTAAT GTAGCAAGTT GCCACTGGGC AAAGAAAATG
 SEC/1rej CCC.ATTCGA CTTGGAAAAG GAAGGATAAG CATTGAAGAA TCTCCGTTAG GAGAT..... GGATTCTAAT GTAGCAAGTT GCCACTGGGC AAAGAAAATG
      MOL CCC.ATTCGA CTTGGAAAAG GAAGGATAAG CATTGAAGAA TCTCCGTTAG GAGAT..... GGATTCTAAT GTAGCAAGTT GCCACTGGGC AAAGAAAATG
 MOLD/RK CCC.ATTCGA CTTGGAAAAG GAAGGATAAG CATTGAAGAA TCTCCGTTAG GAGAT..... GGATTCTAAT GTAGCAAGTT GCCACTGGGC AAAGAAAATG
      BIR CCC.ATTCGA CTTGGAAAAG GAAGGATAAG CATTGAAGAA TCTCCGTTAG GAGAT..... GGATTCTAAT GTAGCAAGTT GCCACTGGGC AAAGAAAATG
      CIM CCC.ATTCGA CTTGGAAAAG GAAGGATAAG CATTGAAGAA TCTCCGTTAG GAGAT..... GGATTCTAAT GTAGCAAGTT GCCACTGGGC AAAGAAAATG
  M.spret CTC.ATTCGG CTTGGAAAAG TAAGGATGAG CATTGAAGAA TCTCCGTTAG GAGAT..... GGATTCTAAT GTAGCAAGTT GCCACTGGGC AAAGAAAATG
  WMP/pas CTC.ATTCGG CTTGGAAAAG TAAGGATGAG CATTGAAGAA TCTCCGTTAG GAGAT..... GGATTCTAAT GTAGCAAGTT GCCACTGGGC AAAGAAAATG
      SFM CTC.ATTCGG CTTGGAAAAG TAAGGATGAG CATTGAAGAA TCTCCGTTAG GAGAT..... GGATTCTAAT GTAGCAAGTT GCCACTGGGC AAAGAAAATG
      XBS CTC.ATTCGG CTTGGAAAAG TAAGGATGAG CATTGAAGAA TCTCCGTTAG GAGAT..... GGATTCTAAT GTAGCAAGTT GCCACTGGGC AAAGAAAATG
      ZYP CTC.ATTCGG CTTGGAAAAG TAAGGATGAG CATTGAAGAA TCTCCGTTAG GAGAT..... GGATTCTAAT GTAGCAAGTT GCCACTGGGC AAAGAAAATG
      KAR CCC.ATTCGA CTTGGAAAAG GAAGGATGAG CATTGAAGAA TCTCCATTAG GAGAT..... GGATTCTAAT GTAGCAAGTT GCCACTGGGC AAAGAAAATG
    M.car CCC.ATTCGA CTTGGAAAAG GAAGGATGAG CATTGAAGAA TCTCCATTAG GAGAT..... GGATTCTAAT GTAGCAAGTT GCCACTGGGC AAAGAAAATG
      COK CCC.ATTCGA CTTGGAAAAG GAAGGATGAG CATTGAAGAA TCTCCGTTAG GAGATTAGAT GGATTCTAAT GTAG...... .........C AAAGAAAACG
 M.dunni ..C.ATTCGA CTTGG..... .....ATGAG CATTGAAGAA TCTCCGTTAG GAGAT..... GGATTCTAAT GTAGCAAGTT GCCACTGGGC AAAGAAAATG
      PTX CCC.ATTCGA CTTGGAA... .AAGGATGAG CATTGAAGAA TCACCATTAG GAGAT..... GGAGTCTAAT GTAGCAAGCT GCTACCGGGC AAAGAAAATG
M.minuto .TCCATTCGA CTTGGAAAAG GAAGGATAAG CATTGAAGAA TCTCCGTTAG GAGAT..... GGATTCTAAT GTAGCAAGTT GCCACTGGGC AAAGAAAATG
      FAM CCCCATTCGA CTTGGAAAAG GAAGGATGAG CATTGAAGAA TCTCCGTTAG GAGAT..... GGATTCTAAT GTAGCAAGTT GCCACTGGGC AAAGAAAATG
```

```
            2801       2811       2821       2831       2841       2851       2861       2871       2881       2891      2900
              |          |          |          |          |          |          |          |          |          |         |
       AKR ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
     BDP/j ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
     LG/J ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
  P.atteck ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
       RBA ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
      RF/J ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
    SF/cam ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
    SK/cam ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
     ST/BJ ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
     SWR/J ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
       MBT ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
       BZO ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
       DBA ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
       WSB ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
       129 ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
     SKIVE ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
   NZW/lacj ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
   NZB/binj ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
   M.mus_cz ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
   M.mus_cz ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
     C57BL CCTTATTCTC TCTGCAGATA GGATGCTGTC CAAAATGGGC GAATTGAAAT GCAGGCGAAG TCGACTGCCT AACTCTGCCA AGACAGGGTA CACAAGTCCT
   PRO/1rej CCTTATTCTC TCTGCAGATA GGATGCTGTC CAAAATGGGC GAATTGAAAT GCAGGCGAAG TCGAC..... ..........  ..........  ..........
     FvB/NJ CCTTATTCTC TCTGCAGATA GGATGCTGTC CAAAATGGGC GAATTGAAAT GCAGGCGAAG TCGAC..... ..........  ..........  ..........
     I/LNJ CCTTATTCTC TCTGCAGATA GGATGCTGTC CAAAATGGGC GAATTGAAAT GCAGGCGAAG TCGAC..... ..........  ..........  ..........
   RIII/dom CCTTATTCTC TCTGCAGATA GGATGCTGTC CAAAATGGGC GAATTGAAAT GCAGGCGAAG TCGAC..... ..........  ..........  ..........
   SEC/1rej CCTTATTCTC TCTGCAGATA GGATGCTGTC CAAAATGGGC GAATTGAAAT GCAGGCGAAG TCGAC..... ..........  ..........  ..........
       MOL CCTTATTCTC TCTGCAGATA GGATGCTGTC CAAAATGGGC GAATTGAAAT GCAGGCGAAG TCGAC..... ..........  ..........  ..........
    MOLD/RK CCTTATTCTC TCTGCAGATA GGATGCTGTC CAAAATGGGC GAATTGAAAT GCAGGCGAAG TCGAC..... ..........  ..........  ..........
       BIR CCTTATTCTC TCTGCAGATA GGATGCTGTC CAAAATGGGC GAATTGAAAT GCAGGCGAAG TCGAC..... ..........  ..........  ..........
       CIM CCTTATTCTC TCTGCAGATA GGATGCTGTC CAAAATGGGC GAATTGAAAT GCAGGCGAAG TCGAC..... ..........  ..........  ..........
    M.spret CCTTATTCTC TCTGCAGACA GGATGCTGTC CAAAATGGGC AAATTGAAAT GCAGGCAAAG TCCACTGCCT AACTCTGCCA AGACAGGGTA CACAAGTCCT
   WMP/pas CCTTATTCTC TCTGCAGACA GGATGCTGTC CAAAATGGGC AAATTGAAAT GCAGGCAAAG TCCACTGCCT AACTCTGCCA AGACAGGGTA CACAAGTCCT
       SFM CCTTATTCTC TCTGCAGACA GGATGCTGTC CAAAATGGGC AAATTGAAAT GCAGGCAAAG TCCACTGCCT AACTCTGCCA AGACAGGGTA CACAAGTCCT
       XBS CCTTATTCTC TCTGCAGACA GGATGCTGTC CAAAATGGGC AAATTGAAAT GCAGGCGAAG ACCACTGCCT AACTCTGCCA AGACAGGGTA CACAAGTCCT
       ZYP CCTTATTCTC TCTGCAGACA GGATGCTGTC CAAAATGGGC AAATTGAAAT GCAGGCGAAG ACCACTGCCT AACTCTGCCA AGACAGGGTA CACAAGTCCT
       KAR CCTTATTCTC TCCACAGACA GGATGCTGTC CAAAATGGGC AAATTGAAAT GCAGGTGAAG TCGAC..... ..........  ..........  ..........
     M.car CCTTATTCTC TCTGCAGACA GGATGCTGTC CAAAATGGGC AAATTGAAAT GCAGGCAAAG TCGAC..... ..........  ..........  ..........
       COK CCTTATTCTC TCTGCAGACA GGATGTTGTC CAAAATGGGC AAATTGAAAT GCAGGCGAAG TCGAC..... ..........  ..........  ..........
    M.dunni CCTTATTCTC TCTGCAGACA GGATGCTGTC CAAAATGGGC AAATTGAAA. ..........  ..........  ..........  ..........  ..........
       PTX CCTTATTCTC TCTGCAGACA GGATGCTGTC CAAAATGGGC AAATTGAAAT GCAGGCAAAG TCAACTGCCT AACTCTGCCA AGACAGGGTA TGCAAGTCCT
   M.minuto CCTTATTCTC TCTGCAGATA GGATGCTGTC CAAAATGGGC GAATTGAAAT GCAGGCGAAG TCGACTGCCT AACTCTGCCA AGACAGGGTA CACAAGTCCT
       FAM CCTTATTCTC TCTGCAGACA GGATGCTGTC CAAAATGGGC AAATTGAAAT GCAGGCGAAG TCGAC..... ..........  ..........  ..........
```

```
          2901       2911       2921       2931       2941       2951       2961       2971       2981       2991       3000
          |          |          |          |          |          |          |          |          |          |          |
     AKR  ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
   BDP/j  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    LG/J  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
P.atteck  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     RBA  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    RF/J  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
  SF/cam  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
  SK/cam  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   ST/BJ  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   SWR/J  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     MBT  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     BZO  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     DBA  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     WSB  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     129  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   SKIVE  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
NZW/lacj  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
NZB/binj  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
M.mus_cz  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
M.mus_cz  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   C57BL  T.ATATGCCT TGCTTCACAA ATATGCCTGT CAGGTA.... TACTGCCAGA AGGTTGAAGA TGGATGCTCC AACATTATCT AGACAATT.C TGGGGAACTG
PRO/1rej  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
  FvB/NJ  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   I/LNJ  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
RIII/dom  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
SEC/1rej  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     MOL  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
 MOLD/RK  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     BIR  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     CIM  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
 M.spret  TCATATGCCT TGCTTCACAA ATATGCCTGT CAGGTA.... TACTGCCAGA AGGTTGAAGA TGGATGCTTC AACATTATCT AGACAATT.C TGGGGAACTG
 WMP/pas  TCATATGCCT TGCTTCACAA ATATGCCTGT CAGGTA.... TACTGCCAGA GGGTTGAAGA TGGATGCTTC AACATTATCT AGACAATTTC TGGGGAACTG
     SFM  TCATATGCCT TGCTTCACAA ATATGCCTGT CAGGTA.... TACTGCCAGA AGGTTGAAGA TGGATGCTTC AACATTATCT AGACAATT.C TGGGGAACTG
     XBS  TCATATGCCT TGCTTCACAA ATATGCCTGT CAGGTA.... TACTGCCAGA AGGTTGAAGA TGGATGCTCC AACATTATCT AGACAATT.C TGGGGAACTG
     ZYP  TCATATGCCT TGCTTCACAA ATATGCCTGT CAGGTA.... TACTGCCAGA AGGTTGAAGA TGGATGCTCC AACATTATCT AGACAATT.C TGGGGAACTG
     KAR  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   M.car  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     COK  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
 M.dunni  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     PTX  TCATATGTCT TGCTTCACAA ATATGCCTGT CAGGTAGGTA TACTGCCAGA AGGCTGAAGA TGGATGCTCC AACATTATCG AGACAATT.C TGGGGAACTG
M.minuto  TCATATGCCT TGCTTCACAA ATATGCCTGT CAGGTA.... TACTGCCAGA AGGTTGAAGA TGGATGCTCC AACATTATCT AGACAATT.C TGGGGAACTG
     FAM  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
```

```
          3001       3011       3021       3031       3041       3051       3061       3071       3081       3091       3100
          |          |          |          |          |          |          |          |          |          |          |
     AKR  ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
   BDP/j  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    LG/J  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
P.atteck  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     RBA  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    RF/J  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
  SF/cam  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
  SK/cam  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   ST/BJ  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   SWR/J  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     MBT  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     BZO  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     DBA  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     WSB  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     129  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   SKIVE  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
NZW/lacj  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
NZB/binj  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
M.mus_cz  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
M.mus_cz  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   C57BL  TCAGACAGCT ATTT....GT CTTCTCTGTA GCCTTGGAAG ATGCTTGCCT GTGCTTCCTG CCAACTCAGG CAATATTTAT CCTTCTGATG TCT.CTGATG
PRO/1rej  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
  FvB/NJ  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   I/LNJ  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
RIII/dom  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
SEC/1rej  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     MOL  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
 MOLD/RK  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     BIR  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     CIM  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
 M.spret  TCAGACAGCT ATTT....GT CTTCTCTGTA GCCTTGGAAG ATGCTTGCCT GTGCTTCCTG CACACTCAGG CAATATTTAT CCTTCTGATG TCT.CTGATG
 WMP/pas  TCAGACAGCT ATTT....GT CTTCTCTGTA GCCTTGGAAG ATGCTTGCCT GTGCTTCCTG CACACTCAGG CAATATTTAT CCTTCTGATG TCTTCTGATG
     SFM  TCAGACAGCT ATTT....GT CTTCTCTGTA GCCTTGGAAG ATGCTTGCCT GTGCTTCCTG CACACTCAGG CAATATTTAT CCTTCTGATG TCT.CTGATG
     XBS  TCAGACAGCT ATTT....GT CTTCTCTGTA GCCTTGGAAG ATGCTTGCCT GTGCTTCCTG CACACTCAGG CAATATTTAT CCTTCTGATG TCT.CTGATG
     ZYP  TCAGACAGCT ATTT....GT CTTCTCTGTA GCCTTGGAAG ATGCTTGCCT GTGCTTCCTG CACACTCAGG CAATATTTAT CCTTCTGATG TCT.CTGATG
     KAR  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   M.car  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     COK  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
 M.dunni  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     PTX  TCAGACAGCT ATTTCTCTGT CTTTTCTGTA GCCTTGGAAG CTGCTTGCCT GTGCTTCCTG CACACTCAGG CAATATTTAT CCTTCTGATG TCT.CTGAGG
M.minuto  TCAGACAGCT ATTT....GT CTTCTCTGTA GCCTTGGAAG ATGCTTGCCT GTGCTTCCTG CACACTCAGG CAATATTTAT CCTTCTGATG TCT.CTGATG
     FAM  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
```

```
          3101       3111       3121       3131       3141       3151       3161       3171       3181       3191      3200
          |          |          |          |          |          |          |          |          |          |         |
     AKR  ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
   BDP/j  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    LG/J  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
P.atteck  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     RBA  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    RF/J  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   SF/cam  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   SK/cam  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    ST/BJ  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    SWR/J  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
      MBT  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
      BZO  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
      DBA  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
      WSB  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
      129  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    SKIVE  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
NZW/lacj  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
NZB/binj  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
M.mus_cz  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
M.mus_cz  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    C57BL  GAATTGAAGA CTAGAT.AGT TATAGTCTCA CACTTAAGCT TAAGTTGTTT TGGATTTAAG A.GGTTTTTA GGTCTAAGAA AATGTTTTTA GGTTAATGCA
PRO/1rej  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   FvB/NJ  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    I/LNJ  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
RIII/dom  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
SEC/1rej  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     MOL  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
 MOLD/RK  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     BIR  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     CIM  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
 M.spret  GAATTGAAGA CTAGAT.AGT TATAGTCTCA CACTTAAGCT TAAGTTGTTT TGGATTTAAG A.GGTTTTTA GGTCTAAGAA AATGTTTTTA GGTTAATGCA
 WMP/pas  GAATTGAAGA CTAGATTAGT TGTTGGGAGC CGC....GCC CACATTCGCC ..GTCACAAG ATGGCGCTGA CAGCGGCCGC TCTAGCCCGG GCGGAACCCC
     SFM  GAATTGAAGA CTAGAT.AGT TATAGTCTCA CACTTAAGCT TAAGTTGTTT TGGATTTAAG A.GG
     XBS  GAATTGAAGA CTAGAT.AGT TATAGTCTCA CACTTAAGCT TAAGTTGTTT TGGATTTAAG A.GGTTTTTA GGTCTAAGAA AATGTTTTTA GGTTAATGCA
     ZYP  GAATTGAAGA CTAGAT.AGT TATAGTCTCA CACTTAAGCT TAAGTTGTTT TGGATTTAAG A.GGTTTTTA GGTCTAAGAA AATGTTTTTA GGTTAATGCA
     KAR  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   M.car  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     COK  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
 M.dunni  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     PTX  GAATTGAAGA CTAGAT.AGT TATAGTCTCA CACTTAAGCT TAAGTTGTTT TGGATTTAAG A.GGTTTTTA GGTCTAAGAA AATGTTTTGG GGTTAACGTA
M.minuto  GAATTGAAGA CTAGAT.AGT TATAGTCTCA CACTTAAGCT TAAGTTGTTT TGGATTTAAG A.GGTTTTTA GGTCTAAGAA AATGTTTTTA GGTTAATGCA
     FAM  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
```

```
          3201       3211       3221       3231       3241       3251
          |          |          |          |          |          |
     AKR  ---------- ---------- ----TACAAA ACTCTGAACT CACCAAGATA GTATAGATG
   BDP/j  .......... .......... .......... .......... .......... .........
    LG/J  .......... .......... .......... .......... .......... .........
P.atteck  .......... .......... .......... .......... .......... .........
     RBA  .......... .......... .......... .......... .......... .........
    RF/J  .......... .......... .......... .......... .......... .........
  SF/cam  .......... .......... .......... .......... .......... .........
  SK/cam  .......... .......... .......... .......... .......... .........
   ST/BJ  .......... .......... .......... .......... .......... .........
   SWR/J  .......... .......... .......... .......... .......... .........
     MBT  .......... .......... .......... .......... .......... .........
     BZO  .......... .......... .......... .......... .......... .........
     DBA  .......... .......... .......... .......
     WSB  .......... .......... .......... .......... .......... .........
     129  .......... .......... .......... .......
   SKIVE  .......... .......... .......... .......... .......... .........
NZW/lacj  .......... .......... .......... .......... .......... .........
NZB/binj  .......... .......... .......... .......... .......... .........
M.mus_cz  .......... .......... .......... .......... .......... .........
M.mus_cz  .......... .......... .......... .......... .......... .........
   C57BL  TGTTATCACA AAAAATTATT TAGA...... .......... .......... .........
PRO/1rej  .......... ..........
  FvB/NJ  .......... ..........
   I/LNJ  .......... ..........
RIII/dom  .......... ..........
SEC/1rej  .......... ..........
     MOL  .......... ..........
 MOLD/RK  .......... ..........
     CIM  .......... ..........
     BIR  ..........
 M.spret  TGTTATCACA AAAAATTATT TAGA...... .......... .......... .........
 WMP/pas  CGGGCTG.CA GGAA.....T TCGA..TC.. G..TATCGA. -...--.TCG ACC.C..G
     XBS  TGTTATCACA AAAAATTATT TGGA...... .......... .......... .........T
     ZYP  TGTTATCACA AAAAATTATT TGGA...... .......... .......... .........
     KAR  .......... ..........
   M.car  .......... ..........
     COK  .......... ..........
 M.dunni  .......... ..........
     PTX  TGTTATCACA AAAAGTTATT TAGA...... ...T...... .......... .........
M.minuto  TGTTATCACC AAAAATTATT TAGA...... .......... .......... T........
     FAM  .......... ..........
```

Appendix 4 Multiple protein sequence alignments using the *Fv1* sequence from the mice used in the sequencing project

```
      1         11        21        31        41        51        61        71        81        91       100
      |         |         |         |         |         |         |         |         |         |         |
AKR   MNFPRALAGF SSWLFKPELA EDSPDNDSPD NDTVNPWREL LQKINVADLP DSSFSSGKEL NDSVYHTFEH FCKIRDYDAV GELLLAFLDK VTKERDQFRD
BDP/j ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
LG/J  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
P.atteck ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
RBA   ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
RF/J  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
SF/cam ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
SK/cam ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
ST/BJ ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
SWR/J ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
MBT   ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
BZO   ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
DBA   ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
WSB   ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
129   ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
SKIVE ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
NZW/lacj ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
NZB/binj ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
M.mus_cz ..........  .......A..  ..........  ........G.  ..........  ..........  ..........  ..........  ..........  ..........
M.mus_cz ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
C57BL ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
PRO/1rej ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
FvB/NJ ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
I/LNJ ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
RIII/dom ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
SEC/1rej ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
MOL   ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
MOLD/RK ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
BIR   ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
CIM   ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........  ..........
M.spret ..........  ..........  ..........  ....S.....  ..........  ..........  .....D....  ..........  ..........  ..........
WMP/pas ..........  ..........  ..........  ....S.....  ..........  ..........  .....D....  ..........  ..........  ..........
SFM   ..........  ..........  ..........  ....S.....  ..........  ..........  .....D....  ..........  ..........  ..........
XBS   ..........  ..........  ..........  ....S.....  ........N..  ..........  .....D....  ..........  ..........  ..........
ZYP   ..........  ..........  ..........  -...S.....  ..........  ..........  .....G....  ..........  ..........  ..........
KAR   ..........  ..........  ..T....---  EN........  ..........  ..........  .....D....  ..........  ..........  ..........
M.car ..........  ..........  ..T....---  EN........  ..........  ..........  .V...D....  ..........  ..........  ..........
COK   ..........  ..........  .......---  EN........  ..........  ..........  S....D....  ..........  ..........  ..........
M.dunni ..........  ..........  .....K.---  EN........  .........L  ..........  .....D....  .......GQL ASCFWH.WI. *
PTX   ...L......  .......D..  .....G.---  EN........  ....SA....  ..........  ....CD....  ..........  ..........  ..........
M.minuto ..........  ..........  .....S.---  EN........  .....A....  ........K.  .....D....  ......N...  ..........  ..........
FAM   ..........  ..........  .......---  EN........  ..........  ..........  .....D....  ..........  ..........  ..........
```

```
            101         111         121         131         141         151         161         171         181         191      200
            |           |           |           |           |           |           |           |           |           |        |
       AKR  EISQLRMHIN  DLKASKCVLG  ETLLSYRHRI  EVGEKQTEAL  IVRLADVQSQ  VMCQPARKVS  ADKVRALIGK  EWDPVTWDGD  VWEDIDSEGS  EEAELPTVLA
     BDP/j  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
      LG/J  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
 P.atteck  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
       RBA  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
      RF/J  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     SF/cam .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     SK/cam .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     ST/BJ  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     SWR/J  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
       MBT  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
       BZO  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
       DBA  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
       WSB  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
       129  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     SKIVE  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
  NZW/lacj  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
  NZB/binj  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
  M.mus_cz  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
  M.mus_cz  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     C57BL  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
  PRO/1rej  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     FvB/NJ .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
      I/LNJ  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
  RIII/dom  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
  SEC/1rej  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
       MOL  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    MOLD/RK .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
       BIR  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
       CIM  .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
   M.spret  .......... .......... .......... ...K...... .......... .......... .......... .......... .........T ..........
   WMP/pas  .......... .......... .......... ...K...... .......... .......... .......... .......... .........T ..........
       SFM  .......... .......... .......... ...K...... .......... .......... .......... .......... .........T ..........
       XBS  .......... .......... .......... .......... .......... .......... .......... .......... .........T ..........
       ZYP  .......... .......... .......... .......... .......... .......... .......... .......... .........T ..........
       KAR  .......... .......... .......... .......... .......... .......... .......... .......... .........T .....T....
     M.car  .......... .......... .......... .......... ......L... .......... .......... .......... .......D.T .....T....
       COK  .......... .......... .......... .......... ......L... ......Q... .......... .......... .......D.T ..........
       PTX  ........V. .......... .......... .......... ......L... IT.......F .......... .......... .......D.. .K.....AS.
  1.minuto  .......... .......... .......... .......... ......L... .T........ .......... .......... .......D.T .K.....DS.
       FAM  .......... .......... .......... .......... ......L... .......... .......... D......... .......D.T ..........
```

```
                201         211         221         231         241         251         261         271         281         291         300
                |           |           |           |           |           |           |           |           |           |           |
          AKR   SPSLSEESGY ALSKERTQQD KADAPQIQSS TSLVTSEPVT RPKSLSDLTS QKHRHTNHEL NSLAHSNRQKAKREHARKWIL RVWDNGGRLT ILDQIEFLSL
        BDP/j   .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
         LG/J   .......... .......... .......... .......... .......... .......... .........Q .......... .......... .......... ..........
    P.atteck    .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
         RBA    .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
         RF/J   .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
       SF/cam   .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
       SK/cam   .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
        ST/BJ   .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
        SWR/J   .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
          MBT   .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
          BZO   .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
          DBA   .......... .......... .......... .......... .......... .......... .........Q .......... .......... .......... ..........
          WSB   .......... .......... .......... .......... .......... .......... .........Q .......... .......... .......... ..........
          129   .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
        SKIVE   .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    NZW/lacj    .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    NZB/binj    .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    M.mus_cz    .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    M.mus_cz    .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
        C57BL   .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    PRO/1rej    .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
        FvB/NJ  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
         I/LNJ  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    RIII/dom    .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
    SEC/1rej    .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
          MOL   .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
       MOLD/RK  .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
          BIR   .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
          CIM   .......... ..F....... .......... .......... .......... .......... .......... .......... .......... .......... ..........
      M.spret   .......... .....L.... .......... .......... .......... .......... K...Q..C.Q .......... .......... .......... ..........
      WMP/pas   .......... .....L.... .......... .......... .......... .......... K...Q..C.Q .......... .......... .......... ..........
          SFM   .......... .....L.... .......... .......... .......... .......... K...Q..C.Q .......... .......... .......... ..........
          XBS   .......... .......... .......... .......... .......... .......... K...Q....Q .......... .......... ....N..... ..........
          ZYP   .......... .......... .......... .......... .......... ...G...... K...Q....Q .......... .......... ....N..... ..........
          KAR   .......... ......A... ......N... .......... .......... .......... K...Q.S..Q .......... ...G...... ....A..... ..........
        M.car   .......... ......A... ......N... .......... .......... .......... K...Q..H.Q .......... .......... ....T..... ..........
          COK   .......... ......I*
          PTX   .......... .......... ...VR.N... .....A.... ....--.... .......Q.. ....QA...R ...R.T.... .......... ....T..... ..........
    M.minuto    ..P....... .....L.... ..EV..N... .......... .......... ...C..DQ.. K...Q..C.N .G.R.G.... .......... ..N.T..... ..........
          FAM   .......... .S........ ......K... .......... .......... ......K... D...L....Q .......... .......... .......... ..........
```

```
          301        311        321        331        341        351        361        371        381        391      400
          |          |          |          |          |          |          |          |          |          |        |
     AKR  GPLSLDSEFN VIARTVEDNG VKSLFDWLAE AWVQRWPTTR ELQSPDTLEW YSIEDGIKRL RELGMIEWLC VKATCPQWRG PEDVPITRAM RITFVRETVE
     BDP/j .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
      LG/J .......... .......... .......... .......... .......... .F........ .......... .......... .......... .......... ..........
P.atteck .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
      RBA .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
      RF/J .......... .......... .......... .......... .......... .F........ .......... .......... .......... .......... ..........
     SF/cam .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     SK/cam .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
      ST/BJ .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
      SWR/J .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
        MBT .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
        BZO .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
        DBA .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
        WSB .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
        129 .......... .......... .......... .......... .......... .F........ .......... .......... .......... .......... ..........
      SKIVE .......... .......... .......... .......... .......... .F........ .......... .......... .......... .......... ..........
NZW/lacj .......... .......... .......... .......... .......... .F........ .......... .......... .......... .......... ..........
NZB/binj .......... .......... .......... .......... .......... .F........ .......... .......... .......... .......... ..........
M.mus_cz .......... .......... .......... .......... .......... .F........ .......... .......... .......... .......... ..........
M.mus_cz .......... .......... .......... .......... .......... .F........ .......... .......... .......... .......... ..........
     C57BL .......... .......... .......... .......... .......... .......E.. .......... .......... .......... .......... .........R.
PRO/1rej .......... .......... .......... .......... .......... .......E.. .......... .......... .......... .......... .........R.
     FvB/NJ .......... .......... .......... .......... .......... .......E.. .......... .......... .......... .......... .........R.
      I/LNJ .......... .......... .......... .......... .......... .......E.. .......... .......... .......... .......... .........R.
RIII/dom .......... .......... .......... .......... .......... .......E.. .......... .......... .......... .......... .........R.
SEC/1rej .......... .......... .......... .......... .......... .......E.. .......... .......... .......... .......... .........R.
      MOL .......... .......... .......... .......... .......... .F........ .......... .......... .......... .......... .........R.
     MOLD/RK .......... .......... .......... .......... .......... .F........ .......... .......... .......... .......... .........R.
      BIR .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
      CIM .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
     M.spret .......... .......... .......... ........P. ...T...... .......... ....I..... .......... .......... .......... ..........
     WMP/pas .......... .......... .......... ........P. ...T...... .......... ....I..... .......... .......... .......... ..........
      SFM .......... .......... .......... ........P. ...T...... .......... ....I..... .......... .......... .......... ..........
      XBS .......... .......... .......... .......... ...T...... .......... .......... .......... ...A....G. .......... ..........
      ZYP .......... .......... .......... .......... ...T...... .......... .......... .......... ...A...... .......... ..........
      KAR .......... .......... .......... .......... ...T..A.K. .......... .......... .......... ...A...... .......... ..........
     M.car .......... .......... .......... .......... ...T..A... .......... .......... .......... ...A...... .......... ..........
      PTX .......... .......E.. .......... .......... ...T...... ..V....... .......... .......... ...A...... .........I.
M.minuto .......... .......E.. .......... ........I. ...T...... FYV....... .......... ...P..R... .D.A...... ..V.....E.
      FAM .......... .......... .......... .......... ...T...... .......... .......... .......... .......... .......... ..........
```

```
          401        411        421        431        441        451        461        471
          |          |          |          |          |          |          |          |
      AKR TWKSFVFSLL CIKDITVGSV AAQLHDLIEL SLKPTAATKL *
    BDP/j .......... .......... .......... .......... .
     LG/J .......... .......... .......... .......... .
P.atteck .......... .......... .......... .......... .
      RBA .......... .......... .......... .......... .
     RF/J .......... .......... .......... .......... .
    SF/cam .......... .......... .......... .......... .
    SK/cam .......... .......... .......... .......... .
     ST/BJ .......... .......... .......... .......... .
     SWR/J .......... .......... .......... .......... .
       MBT .......... .......... .......... .......... .
       BZO .......... .......... .......... .......... .
       DBA .......... .......... .......... .......... .
       WSB .......... .......... .......... .......... .
       129 .......... .......... .......... .......... .
     SKIVE .......... .......... .......... .......... .
NZW/lacj .......... .......... .......... .......... .
NZB/binj .......... .......... .......... .......... .
M.mus_cz .......... .......... .......... .......... .
M.mus_cz .......... .......... .......... .......... .
    C57BL .......... .......... .......... .......GLT SVGSVGVLSL SPWKHQSNS*
PRO/1rej .......... .......... .......... .......GLT SVGSVGVLSL SPWKHQSNS*
   FvB/NJ .......... .......... .......... .......GLT SVGSVGVLSL SPWKHQSNS*
    I/LNJ .......... .......... .......... .......GLT SVGSVGVLSL SPWKHQSNS*
RIII/dom .......... .......... .......... .......GLT SVGSVGVLSL SPWKHQSNS*
SEC/1rej .......... .......... .......... .......GLT SVGSVGVLSL SPWKHQSNS*
      MOL .......... .......... .......... .......GLT SVGSVGVLSL SPWKHQSNS*
  MOLD/RK .......... .......... .......... .......GLT SVGSVGVLSL SPWKHQSNS*
      BIR .......... .......... .......... .......GLT SVGSVGVLSL SPWKHQSNS*
      CIM .......... .......... .......... .......GLT SVGSVGVLSL SPWKHQSNS*
  M.spret .......... .......... .......... .......GLA SVGSVGVLSL SPWKH*
  WMP/pas .......... .......... .......... .......GLP AGLASVGSVG VLSLSPWKH*
      SFM .......... .......... .......... .......GLP AGLASVGSVG VLSLSPWKH*
      XBS N......... .......... .......... .......GLT AGLASVGSVG VLSLSPWKH*
      ZYP .......... .......... .......... .......GLT AGLASVGSVG VLSLSPWKH*
      KAR .......... .........G. .......... .......GLT AGSAPVGSVA VLSLSPWKHQ SNS*
    M.car .......... .........G. .......... .......GLT AGSAPVGSVA VLSLSPWKHQ SNS*
      PTX D......... .......... ..........KEASTREED F*
M.minuto S......... .......... ......IM.. RK.KP.GRGG ARF*
      FAM .......... .......... .......... ....A..GRG GARF*
```

Appendix 5 Insertion upstream of PTX *Fv1*

TCCCCAACTTCAATCCTGTGGCGGTAGGAGAGAAGAGTCTCCCCCAGGAC
ACACTTAGAAGCCTTTAAATCATTTACGTGCATTCGGAGCTGGGAAATTTC
ATCTCTGAATTGGTCCCTTTCCTTTGTTACTTTATCCAGAAATGCCAGAAGC
AGCTCGCCAACTGCGTCATAGTCCCTAATCTTGCATAAATGTTCAAAAGTA
TCGCACACAGAGTCATTAAGTTCCTTACCGCTCGAAAAGGATGAATCGGG
GAGATCGGCCGCACTTATCTTCTGCAGCAGCTCACGCCACGGGTTAACAG
TGTTCTCGTCACCATCCGGAGTGTCCTCGGCAAGATCAGGTTTGAAGAGCC
AGCTCGAGAAACCAGCAAGCGCACGTAGGAAATTCATCTTCAGACTTTTG
TTTCCCTAGAACCCGGCTGCTGCGTCCTACAAGCCTCGGATCCAAATTCC
AGTTCGTCCCTGGCAAAAGCACTCCGGGGAGTTGCTAGTGTGGGGTCCGC
GTGGAAAGGTCAAATGGTCGGAGACGACGATAGCGGCAGTGGCAAAGCA
GTTTTCTCCTCTCAATTTGTTTTTTCTGCCAGTTGGAAGGTGCCGCCCACAA
TGAGGGTGGGTCTTCCTCCTGTCAGTCATACTGCCCCACAGGCCAATCCC
CTCCTCTCTTCACCCGCCCCCAGACAGTGACGATTCTCCAGACAACTTGA
CACCAGAGCGGTTTTTTTTCCCCCTCTTCTTTACCTTTCCTCCCTTTTTATGA
CCTCCTCCCCCCCCACGTGTGAAATATCTGTTGTTAATATCTGTTTTAAGTGT
TGGAACACTTAAATTTTTTTTTTTTTTTGGTTTTTTTTGGTTTTTCGAGGCAGGG
TTTCTCTGTATAGCCCCGGCTGTCCTGGAACTTACTTTGTAGACCAGGCTG
GCCTCGAACTCAGAAATCCGCCTGCCTCTGCCTCCCGAGTGCTGGGATTA
AAGGTGTGCGCCACCACGCCTGGCCACTTAAAAATTTTTGAGTCGTGGGG
CTACTGAGGTAGCTCAGCAAGTGAAGACATTTGCCTCCAAGCCTAAATAC
CTGCTTTTGAGCCCCATAGGTGGGAAAGAACGGTTTGGTTTCGTTTTTGCTT
TAAGTCAGTGTCTCCCTTTGTAAACCCGGCTGGCCTGGAAACTGTTTATAA
TAATACAGGCKGGCTTTGAACGCACAGAGATCTGATAAGTTATTTCTCCGC
TGGTGAAATGAGCCAATTCAAACCAGATTAAAATATATAAACACAGGTGT
ATTGGGAAGCTGGTCTAGGGCAGGTGAGTTCACTGGCCTCCTTCCTTGGGG
GAAGTCGCCATAGGAAAGGGGGTGGGGAGAAAGCACATGCGCAGAGAAA
AGCAAAGCAGGGGGAGGGGAGGGCGAGGGCGGGAGAAACTAAAATGTAT
GGGTTATATAGCAAAGAGCCTCAGGGCAGTCAAGCCCCCGAGCTGGAAA
GCTGCAAGACCAGGAGGCCAGATCTGTTTTGGTAAAATATGCACCCCAGC
CCCTCGCCCCGGGGTCTGTAACCCAATAAACGCTCTTTGAACGCTAGGAT
TAAAGGTGTGTGTTAGCACCCCTGAAGTTTGTTTTCTGTTAGCCTCAAGTTC
AGATCCTCTTGCCTCTCTACCTTCCTGTCCCGTCCTGTACCACCACACCCA
GATAATTCATTAGACCTTTTCAAGTGCTGCAAGACCCCAAGAGAGAAGAT
CCTGACGCTTCCCACCTCCGAACACTAAGAAAGAGGGGTCYATGCTTCAA
GAGCTTTTTCTGTCTACTCGGAGTATTCTGCTTTGACTCCTTGTCTGCTAAGG
AGACAAATGGTATTGTTTGTGTTTTGTTTCACACTGGAACAGGGGACTAAG
GTGCATATTTTACATATCAAAGCAGACCTGGCCTTCTGGTTCTACAACnAC
CCTCAGTCCCTACCTGGCATACCCTGCCCCCAACCCAGAATTTTCCAGAC
CAGGGGCTGCTTTTCCCCCAGAGGCTCCCCCCTATATAATCCAGACTTTTT
TTTGGGGGGGGGGTCGTCTTTCTATTTCTCTCTCTCCCTCTCCCTCTCTCCTC
CCTATATAATCTAGACATTTTGCTTTCTCTCTCCTTCCTCTCTCCCTCCCCTT
TCTCCCTGTACTTCTCTTTGTACATTGTACCCCCACCCCCATCCCATGGTG
ACTTCCCTGGCCTCCTTCCTTGGGGCCAGTGAATTCAATAAACCTGCCTTT
AATATATTCTAATCTGGTTTGAATTGCTTTACTTCACCGGTGGGAnAAATAA
CCTATCAnTTTGCGTTTTGAAAAAnCATCTC

Appendix 6 Sequences of lambda subclone

clone 2dk1

(a)AGCTTGGGGATTGTCTAGGGGCAAACCTTCAnnnnGAGAAACGCTCTAA
GTCGGGGTCTCATACTATCTGGAGCTGGCACTGAAACCAGAACTGGAAGC
ATTAAGCCTTTTTCTCTGCCACAAGTCTTCTCCTGGTCAATTCATTTCTTTGC
TTTTCTTTCTTTCTTCCTTTCTTTAATTTGCATGTATATCCCAAGGTCACACA
GCTACTAAATGTCAAAAGCAGGATTTGAACCAAAATGTAAGTTCCTGCTTG
TTGTGCAGTGGCTACTTCCCCTGACTTTCAAACTTCTCCCTGATGGTGGGA
GGAGAGAATCTCTTCAGAAGGCTGCAAATTCCTAGAGCTCCCTCCACAAG
GTTGATTTAAACTATGTCCAGGGGAGAGCATTCTCAGATGGAGCTGCCTGA
AAGCTGCGCGGGGGGTGGATCAGAACCCGTGCTCGGTGGGACTGTCAGTG
ATGTAGCTGCCTTTGGGTCATGACCCCAATACATACACTCCTTGGTTCACC
CAGCCACACTTGGGTGGAATCGTTTCTCTGTTCTGTTGGCTGTGCTCTGGGG
CAAATAAACCTTTGTTGACACATGACTCCCTATGAAGGGTCTGCACGGAG
GAAAGCATGGCCCTGACTGGGACACAAACCCCACCCCCACCCCTTCTCC
GGCCAGCTTTGGTCCTCAAAAATTATACTCAGGGGCTCTCTCTCTCTCTCT
GTCTCTCTCTCTCTCTCTCTCTCTCTCTA**********
****
**********ATTCTTC
CAGCGGATTAGACCCCCTATCCTTCCCAGCCCTGCCACCTAACAAGCCC
CTGGAGAATTGGGTTTCGGTGCAGGCTGCAGGGGCGAGCAGAGAGACCC
CAGCATCTCACTCACGCGCATTCAGGCAGGGTGCACACCGGGCCTAGGC
GCTTGCCCCTTCTCTACCGTCTAGCGGGACGAAGGCACGCCTCTGTGTTCC
CGAGATCAGCGGCTGCCATGGATAAAAGGACGCCAACGCTCTCCGCCCG
AAAGCCCAGATTCCCAGCACCCGGGGATCCGCCTCCAGCAGGGGCGGAC
CTCCAGTGGTTGTCCTAATTGGCTAAGGCTGGGGCGGGACTCCGCCGCTGT
CCTGATCGGCCAATTGACAGTGCCAGGACGCCTTATATTTGTTTCTTCGCA
TCCAGAAGCGGCAGTTTCAATATGGAAAGGAGGGGCTAATGGGACCTCAT
TGCTGATTGGTTAACTGTTAATGCTAATTCTACCAATTCAAATTGAGGACAG
GCTCTGGTTGGTCAGTAGTACTCGGTTACGCAATTTCCGGATGTAAAGTCT
CTAATGGCAGTGGATAGGTGGGGCTAGAGACTCCGGCAACTTTGACCTTTT
CACGCAGACCCCACACTAGCAACTCCCCAGAGTGCTTTTGCCAGGGACG
AACTGGAATTTGGATCCGAGGCTT**********
**********GCTGGTTTCTCGATCTGGCTCTTCAAACCTGAACTTGCCGA
GGACTCTCCGGATAATGACTCTCCGGATAATGACACTGTTAACCCATGGC
GTGAGCTGCTGCAGAAGATAAATGTGGCCGATCTCCCCGATTCATCCTTTT
CGAGCGGTAAGGAACTTAATGACTCTGTGTACCATACTTTTGAACATTTCT
GCAAAATTAGGGACTATGACGCAGTTGGCGAGCTGCTTCTGGCATTTCTGG
ATAAAGTAACAAAGGAAGGGGACCAATTCAGAGATGAAATTTCCCAGCTC
CGAATGCACATAAATGATCTAAGGGCTTCTAAGTGTGTCCTGGGGGAGACT
CTTCTTTCCTACCGCCACAGGATTGAAGTTGGGGAAAAACAGACTGAAGC
CCTCATTGTGAGGTTAGCTGATGTGCAATCCCAGGTCATGTGTCAGCCTGC
CCGGAAAGTGTCTGCAGATAAGGTGAGGGCACTGATTGGTAAAGAATGGG
ATCCGGTAACCTGGGATGGAGATGTGTGGGAGGACATAGATTCTGAAGGG
TCTGAGGAAGCTGAGTTGCCCACTGTCTTGGCCTCTCCATCCTTGTCTGAG
GAAAGTGGTTATGCCTTGTCTAAAGAACGCACCCAGCAGGACAAAGCAG
ATGCCCCTCAGATCCAGTCTTCAACATCCTTAGTTACTTCTGAACCTGTCA
CCAGACCCAAGTCTCTGTCTGACCTTACAAGTCAGAAACACCGCCATACT
AATCATGAACTCTATTCTCTTGCTCACTCAAATCGCCAAAAGGCAAAGGA

* denotes unsequenced base, as determined by comparison to published *Fv1*^{n} sequence (X97720)
Cloning sites are underlined when present

ACATGCTAGGAAATGGATTTTAAGGGTGTGGGATAATGGTGGGAGGCTCAC
AATACTGGATCAGATTGAATTTCTCAGTTTAGGTCCTTTGAGCCTTGATAGT
GAGTTTAATGTCATAGCCCGCACTGTTGAAGATAATGGTGTGAAGAGTTTG
TTTGATTGGTTGGCTGAAGCATGGGTCCAGAGATGGCCTACTACAAGAGAG
CTGCAGTCGCCTGACACCCTGGAGTGGTATTTTATTGAGGATGGGATTAAA
AGGCTTAGGGAACTTGGAATGATAGAGTGGCTTTGTGTAAAAGCTACTTGT
CCACAGTGGAGGGGCCCGGAAGATGTACCCATCACGAGAGCTATGAGGA
TAACTTTTGTCCGGGAAACTGTAGAGACTTGGAAGAGCTTTGTATTTAGCCT
CCTCTGTATAAAGGACATAACAGTGGGGAGCGTGGCTGCTCAGTTGCATG
ATCTAATAAAATTAAGTTTAAAGCCAACAGCAGCTACAAAACTCTGAACT
CACCAAGATAGTATAGATGATACTTATCCAGAGGTTGCCAAATACAAACA
AACTGGACTTTTTAAATGTCTTACCTAATCATTTTCATAATTGTTCCTACTTT
ATGTTGTTTATTGCTGTAAAAGAATGACTTTTTATGTACACAAAAAGGAGGA
AATGTATAGGGTTGGTCTGGTGCTGCTTGTATTCAAATGCTAAATCCTGTCT
GGGTTGCTCCAAGATGAGCCGTTCGATCTTCATGTATACCAGCCTTTGTTGT
ATAAACCTTGCTCCCCAATTTAAAACTATTGATTAGGGGCTGGAGAGTGGC
TTAGCGGTTAAGAGCACTGACTGCTCTCTGAAGATCTGAGTTCAAATCCCA
GCAACCACATGGTGATTCACTGTTGGGAGCGGCGCCAACTTTCGCCGTAA
CAAGATGGCGCTGACAGCTGTGTTCTAAGTGGTAAACAAATAATCTGCGC
ATGTGCCAAGGTATCTTAGGAAACTTGTGTTCTGCTTTCCCGTGACGTCAAT
TCGGCCGATGGGCTGCAGCCAATCAGGGAGTGACACGTCCGAGGCGAAG
GAGAATGCTCCTTAAGAGGGACGGGGTTTCGTTTTCTCTCTCTCTTGCTTCT
TGCTCTCTTGCTTCTTACACGCTTGCTCCTGAAGATGTAAGAAATAAAGCTT
TGCCGCAGAAGATTCTGGTCTGTGGTGTTCTTCCTGGCCGGTCGTGAGAAC
GCGTCTAATAACAATTGGTGCCGAATTCCGGGACGAGAAAAAACTCGGGA
CTGGCGCAAGGAAGATCCCTCATTCCAGAACCAGAACTGCGGGTCGCGG
TAATAAGGATTCCCGTAAAGCAGACT

clone 6L10E/H

GAATTCCTAGAGCTCCCTTCCAACAAGGTTGATTTAAACTATGTCCAGGGG
AGAGCATTCTCAGATGGAGCTGCCTGAAAGCTGCGCGGGGGGTGGATCAG
AACCCGTGCTCGGTGGGACTGTCAGTGATGTAGCTGCCTTTGGGTCATGAC
CCCAATACATACACTCCTTGGTTCACCCAGCCACACTTGGGTGGAATCGTT
TCTCTGTTCTGTTGGCTGTGCTCTGGGGCAAATAAACCTTTGTTGACACATG
ACTCCCTATGAAGGGTCTGCACGGAGGAAAGCATGGCCCTGACTGGGAC
ACAAACCCCACCCCCACCCCTTCTCCGGCCAGCTTTGGTCCTCAAAAATT
ATACTCAGGGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
CTCTCAAAGATTATTTATTTGTTCGGACACACCAGAAGAGGCGTAGGATTC
CATTACTGATGGTTGTGAGCCACCATGTGGTTGCTGGGAATTGAACTCAGT
GGTTAAGAGCACTGAACCGTCTCTCCAGACACTTTCTTTTGCTTCTGGAAG
AGATGTCTGCCATATAAAGAACGCTTCCCACTTCTCTTACAACCCAGCCC
AGACTCCTCCAGCCGAATCAGACCCCCTATCCTTCCCAGCCCTGCCACC
CTAACAGCCCCTGGAGACTTGGGTTTCCGGTGCAGGCTGCAGGGGGCGAG
CAGAGAAGACCCCAGCATCTCACTCACGCGCACTCAGGCAGGGTGCACA
CCGGGCCTAGGCGCTTGCCCCTTCTCTACCGGTCTAGCGGGACGAAGGCA
CGCCTCTGTGGTCCCGAGATCAGCGGCTGCCATGGATAAAAGGACGCCA
ACGCTCTCCGCCCGAAAGCCCAGATTCCCAGCACCCGGGGAATCCGCCT
CCAGCAGGGGCGGAACCTCCAGTGGTTGTCCTAATTGGCTAAGGCTGGGG
CGGGGACTCCGCCGCTGTCCTGATCGGCCAATTGACAGTGCCAGGACGC
CTTATATTTGTTTCTTCGCATCCGGAAGCGGCAGTTTCAATATGGAAAGGA
GGGGCTAATGGGACCTCATTGCTGATTGGTTAACTGTTAATGCTAATTCTAC

* denotes unsequenced base, as determined by comparison to published *Fv1*^{n} sequence (X97720)

Cloning sites are underlined when present

The *Fv1*^{nr}-associated base change is highlighted in bold and underlined

CAATTCAAATTGAGGACAGGCTCTGGTTGGTCAGTAGTACTCGGTTACGCA
ATTTCCGGATGTAAAGTCTCTAATGGCAGTGGATAGGTGGGGCTAGAGACT
CCGGCAACTTTGACCTTTTCACGCAGACCCCACACTAGCAACTCCCCAGA
GTGCTTTTGCCAGGGACGAACTGGAATTTGGATCCGAGGCTTGTAGGACGC
AGCAGCCGAGTTCTAGGGAAACAAAAGTCTGAAGAGAA**********GC
TGGTTTCTCGAGCTGGCTCTTCAAACCTGAACTTGCCGAGGTACTCTCCGG
ATAATGACTCTCCGGATAATGACACTGTTAACCCATGGCGTGAGCTGCTGC
AGAAGATAAATGTGGCCGATCTCCCCGATTCATCCTTTTCGAGCGGTAAGG
AACTTAATGACTCTGTGTACCATACTTTTGAACATTTCTGCAAAATTAGGGA
CTATGACGCAGTTGGCGAGCTGCTTCTGGCATTTCTGGATAAAGTACCAAA
GGAAGGGGACCAGTTCAGAGATGAAATTTCCCAGCTCCGAATGCACATAA
ATGATCTAAGGGCTCCTAATTGTGTCCTGGGGGAGACTCTTCTTCCCACCG
GCCACGGGATTGAAAGTTGGGGAAAAACAGACTGAAGCCCTCATTGTGAG
GTTAGCTGATGTGCAATCCCAGGTCATGTGTCACCCTGCCCGGAAAGTGTC
TGCAGATAAGGTGAGGGCACTGATTGGTAAAGAATGGGATCCGGTAACCT
GGGATGGAGATGTGTGGGAGGACATAGATTCTGAAGGGTCTGAGGAAGCT
GAGTTGCCCACTGTCTTGGCCTCTCCATCCTTGTCTGAGGAAAGTGGTTAT
GCCTTGTCTAAAGAACGCACCCAGCAGGACAAAGCAGATGCCCCTCAGA
TCCAGTCTTCAACATCCTTAGTTACTTCTGAACCTGTCACCAGACCCAAGT
CTCTGTCTGACCTTACAAGTCAGAAACACCGCCATACTAATCATGAACTC
AATTCACTTGCTCACTCAAATCGCCAAAAGGCAAAGGAACATGCTAGGAA
ATGGATTTTAAGGGTGTGGGATAATGGTGGGAGGCTCACAATACTGGATCA
GATTGAATTTCTCAGTTTAGGTCCTTTGAGCCTTGATAGTGAGTTTAATGTCA
TAGCCCGCACTGTTGAAGATAATGGTGTGAAGAGTTTGTTTGATTGGTTGGC
TGAAGCATGGGTCCAGAGATGGCCTACTACAAGAGAGCTGCAGTCGCCTG
ACACCCTGGAGTGGTATTTTATTGAGGATGGGATTAAAAGGCTTAGGGAAC
TTGGAATGATAGAGTGGCTTTGTGTAAAAGCTACTTGTCCACAGTGGAGGG
GCCCGGAAGATGTACCCATCACGAGAGCTATGAGGATAACTTTTGTCCGG
GAAACTGTAGAGACTTGGAAGAGCTTTGTATTTAGCCTCCTCTGTATAAAG
GACATAACAGTGGGGAGCGTGGCTGCTCAGTTGCATGATCTAATAGAATTA
AGTTTAAAGCCAACAGCAGCTACAAAACTCTGAACTCACCAAGATAGTAT
AGATGATACTTATCCAAAGGTTGCCAAATACAAACAAACTGGACTTTTTAA
ATGTCTTACCTAATCATTTTCATAATTGTTCCTACTTTATGTTGTTTATTGCTG
TAAGAGAATGACTTTTTATGTACACAAAAAGGAGGAAATGTATAGGGTTGG
TCTGGTGCTGCTTGTATTCAAATGCTAAATCCTGTCTGGGTTGCTCCAAGAT
GAGCCGTTCGATCTTCATGTATACCAGCCTTTGTTGTATAAACCTTGCTCCC
CAATTTAAAACTATTGATTAGGGGCTGGAGAGATGGCTTAGCGGTTAAGAG
CACTGACTGCTCTTCTGAAGATCCTGAGTTCAAATCCCAGCAACCACATG
GTGACTCACTGTTGGGAGCCGCGCCCACATTCGCCGTTACAAGATGGCGC
TGACAGCTGTGTTCTAAGTGGTAAACAAATAATCTGCGCATGTGCCAAGGG
TATCTTATGACTACTTGTGCTCTGCCTTCCCCGTGACGTCAACTCGGCCGA
TGGGCTGCAGCCAATCAGGGAGTGACACGTCCGAGGCGAAGGAGAATGC
TCCTTAAGAGGGACGGGGTTTCGTTTTCTCTCTCTCTTGCTTCTTGCTCTCTT
GCTTCTTACACGCTTGCTCCTGAAGATGTAAGAAATAAAGCTT

clone 6L11Bam

(ggat)CCTGGGATGAGTGTCAACNNNNNAACCAGTATCCAGCAGACCCCG
AAAAGTCTGATTATTTCCTTTTCCCCAGTGTACAGTTACCCTTGTAAAAGGC
CGTAGGTCCCTCTGGGGAAGAACTGGAGAAAGGGTAACAGCAAAACCTTT
GAGTGTCTTATCAAGATCCTTCCTCAGCGGAACCTGGCCACCCCTTCATTC
AAGGGGTTCTGGATCTGCAAACTGTCTCAAGTCTGGAAATTGATTCACTGG

* denotes unsequenced base, as determined by comparison to published *Fv1*" sequence (X97720)

Cloning sites are underlined when present

CCGAGATTGCTGTTTACCACGATCTAATGTAGCCTTTCTTTCATTTGGCTGTT
TACCACAATCTAATGTAGCCTTTCTTTCATTTGTTTGAGAATTTTTCTGCTTA
TACAGATCAAACAAATATGCAGTAGGCTTCCTAGTATTTCATTCCTGGAAA
CACCATGATTGGTTAGCCAGTACCAAAGGTCCAACCGAGTCATGCCATTA
TAAATTTCACCTCTCCTGTGCTGACCATTACTGGGTATGTTATTATAAACAT
TCTTTTGTCTACGCTGTCCATTATAATAACTAGAATCACCTTGTCTCCGGTG
AATCAATGCTGCCACCTGGCCCCTGTTACCTTGGAATCCAACTAAACCCA
TGAATTAAATCATCTAATTGAACAGAACATCTCCAATGCTAAGATCTGGCA
CAAGGAAAAGGGAAGAACAAAACCCTCCAATGTGCTGGTGCCCTCTCAC
CAATTTGCGTCTTATAGAGCTGGTGAAAGGCATATCTTCTGACCTTCCCATT
GTGGACAATTATGCTTAACACAATATATCCACTCTAGCATTGCAATTTCCC
TAAGTCTTAAATTCCCTTCATCAACACTAAGCCAGGGAATTTCAGGCATCT
CCAAGTCATTTCCAGTAGGCCATCTTTTGATAAACACTTCAGCCAACCATT
CAAACAAACTTTTGACACCTTTTTTAACTATGCGAGCTTCTGTATTAAACCT
AGATCTCTATTCAGAGGACCCATGTCAATAAATTCAGCCTGTTCTAGTTTTA
TGTTCCTTCCACCCTTATCCCACACCCTTAAAATCCTTCCCACACATATTC
ACCAGGTTTCTGCTTGAATGAATTAGCAAACTCATTAAGCTCCTTAGTAGT
GTAGCGAATTTCCTCATGGACTACACTTTCTACCTCCCCTCTAGGAGCCTG
TTTTGCTTTGAGTCTGGTTACAGGTCTAGAAGAAACTATTGGTGGGCCTTGA
GAAACATCAGTAGTGAAAGTCATTGCTGGTTTATCAGACTCTGCAAAATTA
ATTTACTCATGTGGGGAAGGCATTATTTCAAGGGGTGGGGCTGAGGGTACT
ACTTCCTCAGGTGGGGCAAACCCTTGAGAATCTGAAGATTCTTTATTCTCA
GCTTAGACATGGTCTTCCCACACATCCCCGTCCCATGTTGTN(gga)TCC

clone 6L1E/H
rev:
TTTACTGGCTGGCTGCTCTTAGTATCTGTGGANTTTATATGAATGAAAAGAA
GGAGTTGTGTGATAAAATCTAAAGGCTCCAGACACAAGTAAACGATCTAA
AAGTTGCTAAGTGTGTCCTTGAGGAGAATCTTCTCTCTTGTAGCAATAGAG
CTCAAGTTGCAGAAAATCAAACAGAAACTCTCATTGTAAGGTTGGCTGAA
CTACAGCGAAAATTCAAGTCTCAGCCTCAGAGTGTGTCCACAGTTAAAGT
AAGGGCTCTAATTGGCAAAGAATGGGATCCTACAACATGGGATGGGGATG
TGTGGGAAGACCAGGTTGAAGCTGAGAATTTTGAATCCTCAGATTCTCAAG
GGTTGCCCCACCTGAAGAAGTAGTACCTCAGCCCCACCTCTTGAAATAAT
GCCTTCCCCACATGAAGAAATTAATTTGCAGAATCTGTTCACGGCCCACC
AATAGTTTCTTCTAGAAGTGTAACCAGACTCAAAGCAAAACAGGCTCCTA
GAAGGGAAGTAGAAAAGTGTAGTT

-21:
AAGCTTTCTTGGTGCATGGGGAAAAAACNAAAACCAATGGGGAAATCAN
CACAGGATGTGTCCGGGGGACCTATTGGACCTACTGTGAGTCGGACATCA
GTCAAAACTCCATTAATCACCTGCCCTCCATAAGCCCCTACTTTAACTGGA
GGGCCACAATGTTTCTTGGGATCCCCTGGGATCAGTGTCAACTCAGAACC
AGTATCCAGCAGACCCGAAAAGTCTGATTATTTCCTTTCCCCAGTGTACAG
TTACCCTTGTAAAAGGGCGTAGGTCCCTCTGGGGAAGAACTGGAGAAAGG
GTAACAGCAAAACCTTTGAGTGTCTTATCAAGATCCNNCCNCAGTGGAAC
NGGCCACCCTTCATTCAAGGGGTTCTGGATCTGCAAACTGTCCAAGTCTG
GAAATTGATTCACTGGCCGAGAATGCTGTTTACCACGATCTAATGTACCTT
TCTTTCATTTGGGTGTTTAACACGAACTAATGTAGCCTTTCCTTCATTTGTTT
GAGAATTTTTCTGCTTATACAGATCAACCAAATATGCAGTAGGCTTCCTAT
GTTATTCATTCCTGGNAAACACATGAATGGGTNACCACTACAAAGGTCAA
CCGAATCATGCCATT

Cloning sites are underlined when present

clone 14
-21:
(gaatt)CTGCCAATTTCTCTGTATATCTATCCCAATTATACATTCTGGAACTGG
GGAAATAACCACAGGATGTGTTCTGGGACCTACTGGACCTACTGTGAGTC
AGACATCAGCCAAAACACTATTAATCACCTGTCCTCCATAAGCCCCTACT
TTAACTGGAGGGCCACAATCTTTCTTGGGATCTCCTGGGATCAGTGTCAAT
TCAGAACCCCCTGGTGTACAGTCCTCTACCCCAGTGTACAGTTACCCATGT
AAAAGGCCATAGCTCCCTCTGGGAAAGAATCGGAGAAAGGCTAACAGCA
AAACTTTTAGGTGTCTTATCAATATCCTTCCTCTCAGGGGAACCTGGCCAC
CCCTTCATTCAAGGGGTTCTGGATCTGCAAACTGGCTCAAGTCTGGAAGTT
GATTCACTGGCCCAGATTGTCTTTTGCCATGATCCAATGTAGCCTTTCTTTA
ATTTGAGAGTTTTTCTGCTTATACAGATCAAACAGATATGCAAATAAGTAG
GCTTCCTATCTATTTCATGCTTGGAAACACCATGATTGATTAGCCAGTACC
AAAGGTCCAAACGAGTCATGCCATTATAAATTTCACCTCTCCCGTGATGAA
TAATAAGGGGTATGTGAATAATAAACAATGTTCACACTGTCCATTAATAAC
AACTATGAACACCTTGTCTGGCAATTCAATTGCTGCTACCTGGCCCCTTTTT

rev:
AAGCTAATGCCTTTAATTATCTTGATGAATTAGGTGATTTGGTGTACAAAAC
TTTTTACACGCCGGGGGGAAAAATCAGAAAATGATTTTGCTGGCTGGTTGT
TCTTAGCATCTCAGAAAAAAAATGATGAAGGAAAAGGAGTTGTATGATAA
AACTGAATGGCTCCAGACGCAAGTAAACAATCTAAAAGTTGCTAAACGTA
TCCTTGAATAAAATCTTCTCTACAGCAACCATAGAGCTCAAGTTGCAGAAA
ATCAAACTGAAGCCCTCATAAGGTAAGCTGAATTACAACAAAATTTCAAG
TCTCAGTCCCAGAGGGTGTCAGCAGTTAAAGTAAGGGCATTACTTGGCAA
ATGATGGTGGGATCCTATAACTTGGGATGGGGATGAGTGAGAAGACCCTGT
TGAAGCTAAGAATTTTGGATCTTCAGATTCTCATGGGTTTGCCCCACCTGC
GGAAGTAGTACCCTCAGTCCCACCCCTTGAAATAATGTGTTCCCTACATGA
GAAAATTAATTCCTCAGAGTCTATAAACCAGCAGTGACTTTTGATGAAGAA
AATGCCAGACAAGACAATACTGATGTTTCTCAAGGCCCACCAAGTTTCCTT
CTAAGGCCTATAACCAATCTCAAGGCAAAGCAGGTCCCTAGAAGGGGAT
GT

clone 9
-21:
(gaattctg)CCAATTTCTGGATAGATCTATCCCAATTATACATTCTTGGAACTGG
GGAAATCATCACAGGATGTGTCCGGGGACCTACTGGACCTACTGTGAGTC
GGACATCAGTCAAAACTCCATTAATCACCTGCCCTCCATAAGCCCCTACT
TTAACTGGAGGGCCACAATGTTTCTTGGGATCCCCTGGGATCAGTGTCAAC
TCAGAACCAGTATCCAGCAGACCCCGAAAAGTCTGATTATTTCCTTTTCCC
CAGTGTACAGTTACCCTTGTAAAAGGCCGTAGGTCCCTCTGGGGAAGAAC
TGGAGAAAGGGTAACAGCAAAACCTTTGGGTGTCTTATCAAGATCCTTCCT
CAGCGGAACCTGCCACCCCTTCATTCAAGGGGTTCCAGATCTGCAAACTG
TCTCAAGTCTGGAAATTGATTCACTGGCCGAGATTGCTGTTTACCATGATCT
AATGTACCTTTCTTTCATTTGTTTGAGAATATTTCTGCTTATACAGATCAAAC
AAATATGCAGTAGGCTCCTATGTATTTCATTCCTGGAAACACCGTGAATGA
ATAACCCAGTACCAAAGGTCCAACCGAGTTCATGCCATTATAAATTTCAC
CTCTCCTGTGCTGACCATTACTGGGTATGTTTATTATAAACATTCTTTTGTTC
TACGCTGTCCATTATAATAACCTAGAA

rev:
AAGCTTTCNNCNNGACGGGGAAAAAATCTGAAAATGATTTTACTGGCTGG
CTGCTCTTAGTATCTATGGAAAAAATGATGAATGAAAGGAAGGAGTTGTGT
GATAAAATCGAAAGGCTCCAGACACAAGTAAATGATCTAAAAGTTGCTAA
Cloning sites are underlined when present

GTGTGTCCTTGAGGAGAATCTTCTCTCTTGTAGCAATAGAGCTCAAGTTGC
AAAAAATCAAACAGAAACTCTCATTGTAAGGTTGGCTGAACTACAGCGAA
AATTCAAGTCTCAGCCTCAGAGTGTGTCGACAGTTAAAGTAAGGGCTCTAA
TTGGCAAAGAATGGAATCCTATAACATGGGACGGGGATGTGTGGGAAGAC
CATGTTGAAGCTGAGAATTTTGAATCTTCAGATTCTCAAGGGCTTGCCCCA
TCTGAGGAAGTAGTTTTCTCAGCCCCACCCCTTGAAATAATGCCTTCCCCA
CATGAGGAAATTAATTTTGCAGAGTCTGATAAACCAGCAATGACTTTCACT
ACTGATGTTTCTCAAGGCCCACCAATAGTTTCTTCTAGACCTGTAACCAGA
CTCAAAGCAAAAACAGGCTCCTAGAAGGGGAAGTAGAAAGTGTAGTCCA
TGAAGAAATTCGCTACACTACTAAGGGAGCTTAATGTGTTTGCTAATTCAT
CCAAGCAGAAACCTAATGAATATGTTGTTGGGAATGGAATTTAAGGGT

clone 3
ACATGGGACGGGGATGTGTGGGAAGACCATGTCTAAGCTCAGAATAAAGA
ATCTTCAGATTCTCAAGGGTTTGCCCCACCTGAGGAAGTAGTACCCTCAGC
CCCACCCCTTGAAATAATGCCTTCCCCACATGAGTAAATTAATTTTGCAGA
GTCTGATAAACCAGCAATGACTTTCACTACTGATGTTTCTCAAGGCCCACC
AATAGTTTCTTCTATACCTGTAACCAGACTCAAAGCAAAACAGGCTCCTAG
AGGGGAGGTAGAAAGTGTAGTCCTTGAGGAAATTCGCTACACTACTAAGG
AGCTTAATGAGTTTGCTAATTCATTCAAGCAGAAACCTGGTGAATATGTGT
GGGAAGGATTTTAAGGGTGTGGGATAATTTTGGAAGGAACATAAAACTAGA
ACAGGCTGAATTTATTGACATGGGTCCTCTGAATAGAAATCTAGGTTTAAT
ACAGAACTCGCATAGTTAAAAAGGTGTCAAAAGTTTGTTTGAATGGTTGGC
TGAAGTGTTTATCAAAAGATGGTTCACTGGAAATGACTTGGAGATGCCTGA
AATTCCCTGGCTTA

clone 5c
AAGCTTCCGGACTTAAACCTAGAATTTCCACTCAGATGGACCATGTCAATA
AATTCAGCCTGCTCTAGTTTTATGTTCCTTCCACCCTTATCCCACATCCTTA
AAATCTATTCCCACACATATTTGCCAGGTTCCTGCTTGAATGAATTAGCAA
ACTCTTTAAGTTCCTTAGTAGTGTAGCGCATTTCCTCGTGGACTACACTTTC
TACCTCCCCTCTAGGAGCCTGTTTTGCTTTAAGTCTAGTTACAGGTCTAGAA
GAAACTATTGGTGGGCTTTGAGAAACATCAGTATTGTCTTGTCTGGCATTTT
CTTCAGTGAAAGTCACTGCTGGTTTGTCAGACTGCAGAATTAATTTCCTCAT
GTGGGGAAGGCATTATTTCAAGGGATGGGGCTGAGGGTACTACTTCCTCAG
GTGGGGCAAACCCTTGAGAATCTGAAGATTCAAAATTCTCAGCTTCAACAT
TGTCTTCCCACACATCTCCATCCCAAGTTATAGGATCCTATTCTTTGCCAAT
TAGTGCCCTTACTTTAACTGTTGACACACTCTGAGGCTGAGACTTGAATTTT
CACTGTAGTTCAGCCAACCTTACAATGAGGGTTTCAGTTTGATTTTCTGCAA
CTTTAGCTCTATGGCTGCAAGAGAGAAGATTCTCTTCAAGGACACACTTAA
CAAAATTTAAATCATTTACTTGTGTCTGGGGCCTTTCGATTTTGTCTCTCAAA
TCCTTCCTTTCATTCATCATTTTTCCCCAGATGCTAAGAGAAGCCAGCCAG
TAAAATCATTTTTTCAATTTTTCCCCCATCTTGTAGAAAGCTT

clone 5b
-21:
TATCCCAATTATACATTCTGGAACTGGGGAGATGACCACAGGATGTGTTCA
GGGACCTACTGGACCTACTGTGAGTCAGATATCAGTCAAAACTCCATTAAT
CACCTCCATAAGCCCCTTCATTCAAGGGGTTCTGGGTCTGCAAACTGGCTC
AAGTCTGGAAATTGATTCACCAGCTGAGGTTGCCTTTTGCCACAATCCAAA
TGTAGCCTTTGTTTCATTTGTTTGAGAAGTTTTCTGCTTATACAGATCAACCA
GATATGCAGTAGGCTTCCTATCTATTTCCTGCCTGTAAACACCATGATTGAT
TAGCCAGTAACAAAGGTCCAAACTAGTCATGTCATTATAACTTTCATCTTT
Cloning sites are underlined when present

CCTGTGCTGACCATTATGGGGTATGTCATTATAAACCTTGTTTTCTCTACGC
TGTCCATTATAATAACTATGATCACCTTGTCTCTGGCAATTCAGTGCTGCCC
CCTGGCCCTTGTTACCTCGATTCCCAATTAAACCCATTGAATTTAATTCATC
TAATTGAGCAGCAGCATCTTCAACCCTAAGGTCTGGCACAAGGAAAAGGG
AAAGAACAAAACCTTCAAATATGCTGGTGCCCCTCTCACCATTTGGTGTTC
TTATGGACTGGGTAAAGGGCATAT

rev:
AAGCTTGGCCCAACACGCCGTATAGTAAGCGGTAAGCTGGAGCTGCAGTG
GTGGAGTGCGTGTTTAGCAATCGGCAGCACAGAGGTGGGGCTGGAGAGGA
GGGAAGGAAGAAAACCAGCTCAAACGCCCACTTTCTCATGGAAACCCAG
ATATATGAATTACCCCATTGGCTATAGATTTAAACAATTGTCTAGACTTCCT
TAGAGGAAAAAATATAGTAATTTCCTTAACTCATCACCAAGATGCCACAG
AAATGGTGCTGACATAACGGTAGAGATTATTTGTTCCTAATTCTACTGCAG
AGGGATGCTTTGGGAGGTTCTAGCCAGCAGCTCTTTCCCCCAGTCTCTTAC
ATGCTGATGCTGAGCAAGAACTTAACCTGCAAGACCCAATGTACTCTGTCT
CTCTCTCCTTCGCCTGCTTGCCATGTGGGACGGAGCAACTGCTAGATCCTG
GGACTTCCATTCACAGCTGCTGCTGACCATTGTTAGGAGTTGGACCGCAGA
CTGTAAGTCATCAACAAATTCCTTTACTATATAGAGACTACCCATAAGTTC
TGTGACTCTAGAGAACCCTGACTAATACAGAAGTTGGTACTGGGGAATGGT
TCTNAAGGAGCAGAAGTATAAGAATGAATCTCTTTAAAAAT

clone 1
-21:
CGGACCATTTCTTTATCTCCTCCTCTTGCTATGGTTGCCATCATGTNGGCAA
CAGCAAGTGGAGTGGCTCTCACTTCATGCTGTCCGATCCCTGTCAAGCCTG
CAAAATTACTGTCTTTTTTTGCATCTTCTGACAGAAACACTCGTCCTTTTTCT
TCGTCTTGAAGCTGTTTGAAATCATTAAAATGATAAATATCCCCCTGCCAG
CTCACAGGACCCATTAATGAAAGTTTATCTGCATACTCCTCAAGGAGATCC
GGGTTAATATCTTTTAGCTCCTTAGCAATGGAAGCAAATGTATTATTGCAGC
TTCTGGCAAAGCTATCATTAAAATTCAGCATCCCATGCTGGTATTTAAGAT
AGGCTCCCCGTTTATTTTCCTGCTGCAGTCAAATTGCCTTTTCGGATCGTCC
AGACCTTGATCAATTGCCGCAGCTGCCACCACTGTTTTAAAAACTGATCCC
ATAATTTGCTGTTTAATCATATGAATTACTGTTCCGCTGCTTTCTTCATCAAA
AGGATTATCTTTGTTAATGGATGGACGTGACACCATAGCCAAAACTGAATT
AATCTCTATATCAAGGAAGAACACTCCCCCCTTTTTAATCCCATGCTGAAC
AGCCAATTTTTC

rev:
GAATTCTGCCAATTTCTGAGTATATCTATCCCAATTATACATTCTGGAACTG
GGGAAATAACCACAGGATGTGTTCTGGGACCTACTGGACCTACTGTGAGT
CAGACATCAGCCAAAACACTATTAATCACCTGTCCTCCATAAGCCCCTAC
TTTAACTGGAGGGCCACAATCTTTCTTGGGATCTCCTGGGATCAGTGTCAA
TTCAGAACCCCCTGGTGTACAGTCCTCTACCCCAGTGTACAGTTACCCATG
TAAAAGGCCATAGCTCCCTCTGGGAAAGAATCGGAGAAAGGCTAACAGC
AAAACTTTTAGGTGTCTTATCAATATCCTTCCTCTCAGGGGACCTGGCCAC
CCCTTCATTCAAGGGGTTCTGGATCTGCAAACTGGCTCAAGTCTGGAAGTT
GATTCACTGGCCCAGATTGTCTTTTGCCATGATCCAATGTACCTTTCTTTAA
TTTGAGAGTTTTTCTGCTTATACAGATCAAACAGATATGCAATAAGTAGGCT
CCTATCTATTTCATGCTTGGGAAACACCATGAATGAATAGCCATACCAAAG
GTCCAACGAGTCATGCCATTATAAATTTCACCTCTCCTGTGATGACTATTAT
GGGG

Cloning sites are underlined when present

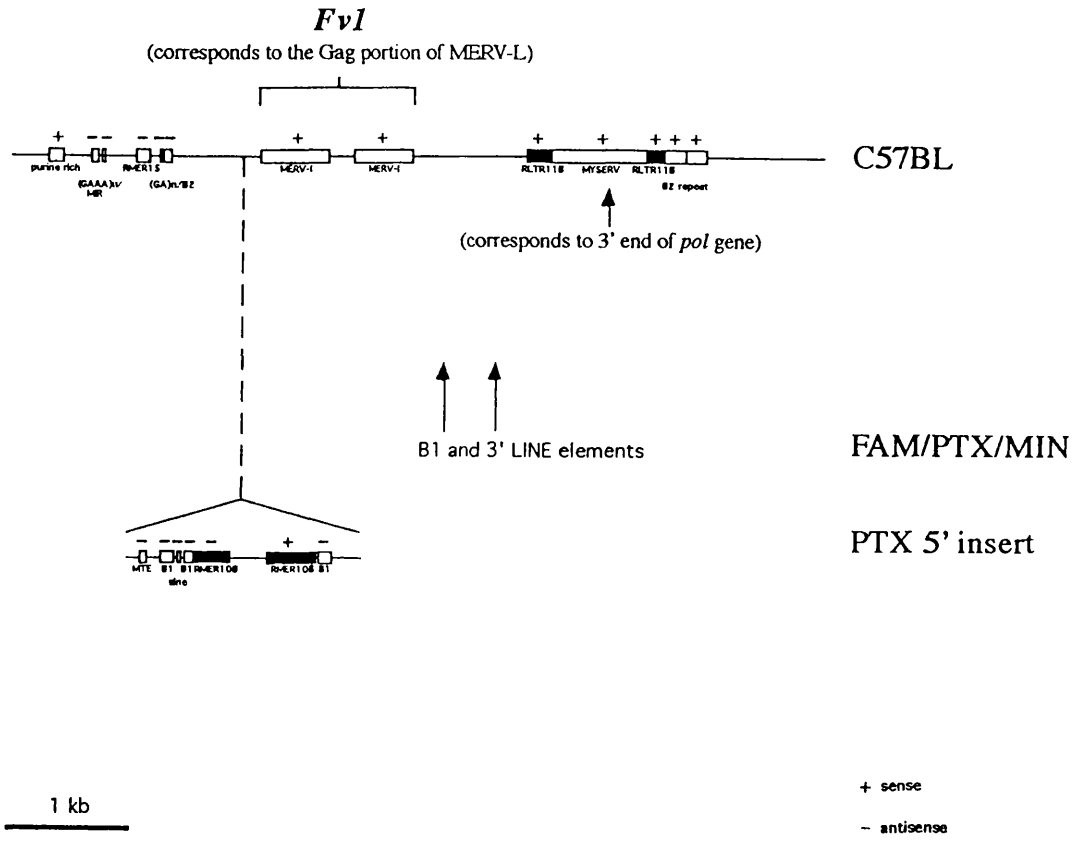
clone 4
CTTTGACCTCCCTGGCTCAGGTGATTCTCTCCCACCTCAGCCTCCAGAGTA
GCTGGGTTCACAGGTGCATGCCACCACACACCTAACTAATTTTTATTTTTTG
TAGAAACAGGGTTTTGCCATGTTGCCTAGGCTGGTCTCAAACTCTGGGCTC
AAGCAATCACGACTAGCTCAAGTGCTGGATTACAGGCATGAGCACCAGA
CCAGCTGA

clone 5
TCCAGCCTGGGCGAAAGAGCGAGACTCCGTCTCAAACAAACAAACAAAC
AAACAAACAAAAAACAGCTATTGTTTGGAATAGTCTATAAACCAGTTTCCT
TGGAGTCTAGAGAGCAGTCAGTTGAGGTGTCTAGATGTTAGACTTGAAAAC
CTCAATTAAGTGGAATAGCAGTGCAATATGAAGATT

clone 22
TGTTAACAATGCAATGTTGTGGGATAAAGTTCTCACTAGTGGGATTGGCCA
TATAACAACTGCTTCCTAAGTGTTGAAGGAACCGATGGAGATAAATCCTAT
CTTATGACAGAAGGATCAGATGAAA

clone23b
GCAGCCCCACCCCTCAAGGTCACATAAGTTGGTGCTGTTGCTTGTAACTAA
TTGGAGATTTTAGTTTTTTCTTTACCCAAAATTAAAACTAAGGACACAATAC
TCTTCAAATACCAATACCTGAGATACATCACTAACAGATCTACATAGGGCT
TTGCTATTAGGTACAGTTAGTAAA

Appendix 7 Repetitive sequences in the *Fv1* region



Repeats in the *Fv1* region were identified by using the repeat filtering program RepeatMasker. The names of the repeats marked on the Figure correspond to sequences submitted to the rodent repetitive element database Repbase. The orientation of these sequences is indicated.
Fv1 shares homology to the *gag* portion of the MERV-L submited sequence. MYSERV corresponds to the *pol* sequence of an endogenous retrovirus. RLTR11B and RMER10B correspond to retroviral LTRs. All other elements are simple repeats.