CHARACTERISING AND MAPPING PORCINE ENDOGENOUS RETROVIRUSES

by

Jun Heon Lee

A thesis submitted to the Faculty of Agricultural science in the fulfilment of the requirements for the degree of Doctor of Philosophy

The University of Sydney
Australia

August 2000

Declaration

The work presented in this thesis is original and was undertaken in the Department of Animal Science while I was enrolled as a PhD student in the Faculty of Agricultural Science, University of Sydney.

I certify that this thesis has not been submitted for any other degree, and that all sources of information and assistance during the experimental work and preparation of the thesis are duly acknowledged.

Jun Heon Lee

August 2000

Dedication

For my beloved people:

My parents, Jil Hyun Lee and Jung Ja Kim

My wife, Myoung Hee Hong

My daughter, Kyong Shik Lee

Acknowledgements

During my three and half years PhD study, a lot of people encouraged me to finish my course in various ways, which I will gratefully acknowledge.

First of all, I am most grateful to my supervisor, Associate Professor Chris Moran, for his support, encouragement and invaluable guidance throughout my PhD candidature. It is a great fortune for me in having had the opportunity to study under his supervision. Chris has been involved at all stages and in all aspects of my studies described in this thesis, making lots of valuable suggestions. Also, he devoted a lot of time and care to critically reading all drafts of this thesis, and gave expert editorial advice for completion of this thesis.

I am also extremely grateful to my associate supervisor, Associate Professor Frank W. Nicholas, who provided stimulating suggestions throughout my Ph.D. course and this thesis. Frank also read the whole thesis and made many constructive suggestions, which are greatly appreciated.
I will always bear my supervisors' suggestions in mind as a scientific researcher for the rest of my life.

My sincere gratitude goes to Dr Graham C. Webb, at University of Adelaide and The Queen Elizabeth Hospital, Adelaide who gave me a lot of help including warm hospitality in his house and experiments on FISH mapping of PERVs.

Thanks to the people at the Westmead Hospital, Sydney, including Dr Phill O'Connell, Jane Burgess, and Tina Patel in relation to valuable discussions and Westran pig DNA samples.

I thank to Associate Professor Mick O'Neill in the Biometry Unit and Dr John James in relation to the confidence limits of my experiment.

I also thank to Dr Yizhou Chen, postdoctoral fellow in my laboratory, who generously shared meaningful insights and discussions on many aspects of my project.

My personal thanks to Seung Soo Lee, a Korean Fellow who has shared most of my time in this laboratory and has shared great knowledge of genetics.

I have much pleasure in acknowledging all of my colleagues in the Animal Genetics section including, Zung Doan, Dr Weiyi Zhang, Dr Mohammad Shariflou, Dr Nauman Maqbool, Nicola Man, Jaime Gongora, and Mark Gear for a very helpful, friendly and entertaining environment.

I also thank the colleagues in the Department of Veterinary Clinical Sciences at Camden (Dr Herman Raadsma, Dr Imke Tammen, Dr Bill Blackhall, Julie Cavanagh, Natatha Ellis, Marilyn Jones, Gina Attard) for their help in relation to the Li-Cor sequencer.

My general thanks also go to many other members of the Department of Animal Science for their kind help in various ways.

I also thank to the people in the Korean Oversea Student Association in the University of Sydney and ROTC colleagues who give me a lot of entertainments and encouragements, especially Jung Won Kim, Wonje Lee, Robin Lee, Hak Hae Kim, Yong Jae Jeon and their families.

My heartfelt thanks go to Myoung Hee Hong, my wife and my colleague, and my daughter Kyong Shik Lee, for their love, encouragement and support during my three years PhD study.

Finally, I would like to gratefully acknowledge all the support of my parents. Without their help, I could not start my PhD either.

This thesis was written using Microsoft Word 97 installed in a personal computer Pentium 200. All result pictures were scanned by Hewlett Packard ScanJet 6200C scanner.

Jun Heon Lee
The University of Sydney
August 2000

Summary

The initial focus of this PhD project was on comparative gene mapping. Comparative gene mapping is facilitated by consensus PCR primers which amplify homologous gene fragments in many species. As a part of an international co-ordinated programme of comparative mapping in pigs, 47 CATS (Comparative Anchor Tagged Sequence) consensus primer pairs for loci located on human chromosomes 9, 10, 20, and 22, were used for amplifying homologous loci in pigs. After optimization of PCR conditions, 23 CATS products have confirmed by comparison with homologous sequences in GenBank. A French somatic cell hybrid panel was used to physically map the 6 porcine CATS products distinguishable from rodent background product, namely *ADRA1A*, *ADRA2A*, *ARSA*, *GNAS1*, *OXT* and *TOP1*. Of these, the map location of *ADRA1A* and *OXT* showed inconsistency with the previously recognised conserved relationship between human and pig. The other four loci mapped to positions consistent with known syntenic relationships. Despite low levels of polymorphism, frequently indistinguishable rodent and porcine products in somatic hybrids and some confusion of identity of gene family members, these CATS primers have made a useful contribution to the porcine-human comparative map.

The focus of the project then changed to genetic and molecular characterisation of endogenous retroviruses in pigs and their relatives. Pigs are regarded as a potentially good source of organs and tissues for transplantation into humans. However, porcine endogenous retroviruses have emerged as a possible problem as they can infect cultured human cells. Two main types of pig retrovirus, determined by envelope protein, PERV-A and PERV-B, are widely distributed in different pig breeds and a third less common type, PERV-C, has also been recognised. Endogenous retroviruses were analyzed from the Westran (Westmead transplantation) inbred line of pig, specially bred for biomedical research. Thirty-one 1.8 kb env PCR product clones were sequenced after preliminary screening with the restriction enzymes *Kpn*I and *Mbo*I. Five recombinant clones between A and B were identified. 55% of clones (17/31) sequenced had stop codons within the envelope protein-encoding region, which would prevent the retrovirus from making full-length envelope protein recognizable by cell-surface receptors of the virus. The endogenous viruses were physically mapped in Westran pigs by FISH (Fluorescence *In Situ* Hybridisation) using PERV-A and PERV-B envelope clones as probes. Preliminary FISH data suggest that there are at least 22 PERVs (13 PERV-A and 9 PERV-B) and the chromosomal locations of these in the Westran strain are quite different from European Large White pigs. The sequences and mapping results of inbred Westran pig suggest that there are relatively few PERV integration sites compared with commercial pigs and further that a large proportion of clones are defective due to premature stop codons in the envelope gene.

To investigate the relationship of endogenous retroviruses in peccaries and pigs, a set of degenerate primers was used to amplify peccary retroviral sequences. The sequences of two putative retroviral clones showed close homology, albeit with a 534 bp deletion, to mouse and pig retroviral sequences. Also, four non-target sequences were amplified from peccary with the degenerate retroviral primers. They are a part of the peccary cofilin gene, a SINE, and a sequence containing a microsatellite. The peccary endogenous retroviral sequences are significant in that they are the first such sequences reported in peccary species and repudiate old claims in the literature that peccaries do not have C-type retroviral sequences.

Abbreviations

^{3}H	=	tritium
5-BrdU	=	5-flourdeoxyuridine
ANGIS	=	Australian genomic information service
APS	=	ammonium persulfate
bp	=	base pair(s)
BSA	=	bovine serum albumin
CATS	=	comparative anchor tagged sequences
cM	=	centi Morgan
cpm	=	counts per minute
dATP	=	deoxyadenosine triphosphate
dCTP	=	deoxycytidine triphosphate
dGTP	=	deoxyguanosine triphosphate
dTTP	=	deoxythymidine triphosphate
DNA	=	deoxyribonucleic acid
dNTPs	=	deoxynucleoside triphosphates (dATP,dCTP,dGTP,dTTP)
EDTA	=	ethylenediamine tetra acetic acid
FISH	=	fluorescence *in situ* hybridisation
FITC-avidin	=	Fluorescein isothiocynate conjugated avidin
g	=	gram
g	=	centrifugation force
kb	=	kilobase pair(s)
LB	=	Luria-Bertani medium
OD	=	optical density
nm	=	nano metre
PBS	=	phosphate-buffered saline
PCR	=	polymerase chain reaction
PERVs	=	porcine endogenous retroviruses
PHA	=	phytohaemagglutinin
PiGMaP	=	European Community initiated international pig gene mapping project
PSF	=	penicillin, streptomycine, and fungizone
PWM	=	pokeweed mitogen
RFLP	=	restriction fragment length polymorphism

RNA	=	ribonucleic acid
Rnase	=	ribonuclease
QTL	=	quantitative trait locus
rpm	=	revolutions per minute
SDS	=	sodium dodecyl sulphate
SNP	=	single nucleotide polymorphism
TAE	=	Tris-acete-EDTA buffer
Taq	=	Taq DNA polymerase
TBE	=	Tris-borate-EDTA buffer
TE	=	Tris-EDTA buffer
TEMED	=	N,N,N',N'-tetra methyl ethylene diamine
Tris	=	Tris(hydroxymethl)aminomethane
µCi	=	micro Curie
µl	=	micro liter
µM	=	micro Molar
UV	=	ultraviolet
V	=	voltage

List of Contents

Chapter One

Introduction

The present thesis has two separate aims. One is developing markers in the pig genome using CATS (comparative anchor tagged sequences) primers and the other aim is characterising and mapping PERVs (Porcine Endogenous Retroviruses). PERVs are a major potential hazard of xenotransplantation.

1.1. Genome mapping

1.1.1. Markers for genome mapping

1.1.1.1. Conventional markers

Throughout human history, animals and plants have been selected based on their superior appearance and performance. Such performance-based selection continues to be vitally important today. However, the availability of genetic markers, initially blood group variants and later biochemical polymorphism (Juneja and Vögeli, 1998), opened new possibilities. Blood group variation is detected as the presence or absence of certain antigens on blood cells. Researchers investigated the relationship between variation in blood types and variation in susceptibility to disease or other traits of interest. Biochemical polymorphism was also investigated based on electrophoretically detectable enzymes and protein variants (also called allozymes). The biochemical polymorphisms studied in the pig focussed mainly on the proteins of blood plasma (serum) and red cells and also to some extent those in milk, semen and other tissues (Juneja and Vögeli, 1998). Although some of these proteins showed relatively high levels of polymorphism, they were replaced by superior DNA markers displaying higher levels of variation or greater availability of markers and detectable using a uniform assay technique.

1.1.1.2. DNA markers

The advent of recombinant DNA technologies, and especially the Polymerase Chain Reaction (PCR) technique, provided ready access to DNA markers for any species. DNA markers can be divided into two categories, namely multilocus markers and single-locus markers. The former includes minisatellites (also called variable number of tandem repeats (VNTR),

although this term also applies to microsatellites), random amplified polymorphic DNA fragments (RAPD) and amplified fragment length polymorphism (AFLP) markers. The latter includes restriction fragment length polymorphisms (RFLPs), microsatellites and single nucleotide polymorphisms (SNPs).

Minisatellites or VNTR markers

Minisatellites were the first developed tandem repeat markers with multiple alleles. Jeffreys *et al*. (1985) discovered human minisatellites, which typically have a core sequence greater than 10 bp, with 3 to 30 copy numbers within a block. Because variability in this marker system is determined by the number of the repeats within the block, minisatellites are also known as VNTRs for Variable Number Tandem Repeats (Nakamura *et al*., 1987). Minisatellites are generally sufficiently informative to reveal a unique genotype in each individual. Thus these markers can provide an individual specific DNA "fingerprint" and can be used for parentage testing and linkage studies (Jeffreys *et al*., 1985). Many minisatellite loci have been found to be extremely variable, with heterozygosity values greater than 90% and mutation rates exceeding 10^{-2} per generation (Bruford and Wayne, 1993). Despite their hyperpolymorphism, minisatellites have fallen into disfavour as genetic markers in recent years for at least three major reasons. Firstly, minisatellites do not uniformly cover the genome in some species, primarily residing in heterochromatic regions near telomeres and centromeres. Secondly, minisatellites are not amenable to PCR, requiring hybridisation technology. Thirdly, the interpretation of complex fingerprint patterns generated by core sequence probes is very difficult and it is difficult or impossible to recognise individual loci (Dodgson *et al*., 1997).

RAPD markers

RAPD markers were the first PCR-based markers to be used (Williams *et al*., 1990). Small primers (8-10 bp) are used to amplify a mixture of random fragments from the genome. Although these markers have the advantage of being very easily generated by PCR and require only a small amount of template DNA (Williams *et al*., 1990), they have severe limitations. They are dominant markers which are intrinsically less informative than codominant markers. Also RAPD markers are not easily reproducible because the appearance of bands is very sensitive to slight changes in PCR conditions. The most serious disadvantage is that a new map must be generated for each new pedigree examined, as there is no locus specificity in the primers used (Crawford *et al*., 2000).

AFLP markers

AFLP markers are also multilocus markers (Vos *et al*., 1995). To produce them, genomic DNA is cut with restriction enzymes and oligonucleotide adaptors are ligated to each end of the restriction fragments. Selective PCR primers are used to amplify subsets of fragments from the mixture of genomic restriction fragments. The selective primers cover the adaptor sequences added onto the end of the restriction fragments and include additional bases to the 3' end of the primers to give additional specificity (Vos *et al*., 1995). These markers have the same advantage as RAPDs, being easily generated by PCR, but are less sensitive to the PCR conditions, giving more reproducible amplification products. However, like RAPDs, these markers have the disadvantage of being diallelic dominant/recessive markers, also requiring a new map for each new resource pedigree being studied (Crawford *et al*., 2000).

RFLP markers

RFLPs are the first reported class of DNA markers. They result from small-scale changes in DNA, usually base substitutions, which create or destroy specific restriction endonuclease cleavage sites. They are detected as size variants on electrophoresis (Botstein *et al*., 1980). Initially, RFLPs were detected by Southern hybridisation analysis (e.g. Archibald *et al*., 1994). Now, most RFLPs used in animal genome analysis are detected by PCR, and are called PCR-RFLPs.

RFLP markers are generally only diallelic, corresponding to the absence or presence of a recognition site for the restriction endonuclease employed, and thus no more than 50% heterozygosity can be expected. However, these markers have advantages of being easy and relatively cheap to generate and being codominant. RFLPs are still used in genome mapping, mainly of type I candidate genes.

Microsatellites

Microsatellites, also called Simple Sequence Repeats (SSR), are tandem repeats of one to four bp, which are interspersed throughout the genomes of mammals, birds, fish, and plants (see review Bruford and Wayne, 1993). Microsatellites are the markers primarily responsible for the recent explosive expansion of genetic linkage maps in many species. They have the advantages of being single-locus, hyperpolymorphic, co-dominant in almost all cases, and uniformly spread throughout the genome. Further they require only a small amount of template DNA and are relatively easy to find and characterise (Crawford *et al*., 2000). Another enormous advantage of microsatellite markers is that they are sequence-tagged sites (STS) and thus their use can be transferred easily between laboratories simply by exchange of primer sequences.

SNPs

SNPs arise primarily from nucleotide substitutions. They are estimated to occur about every kilobase of unique sequence in humans (Cooper *et al*., 1985) and similar frequencies of occurrence are estimated in other mammals. Thus SNPs represent a particularly rich source of genetic variation (Crawford *et al*., 2000). In theory, SNPs can be up to tetra-allelic polymorphisms. However, in practice in humans, tri-allelic and tetra-allelic SNPs are very rare. Generally SNPs are bi-allelic (Brookes, 1999).

Numerous techniques, including PCR-RFLP, are available for detecting some or all SNPs. Orita *et al*. (1989) described a method detecting point mutation as Single Strand Conformation Polymorphism (SSCP). Single-stranded DNA molecules of different sequence display specific sequence-based migration patterns in a non-denaturing polyacrylamide gel, due to alternative three-dimensional conformations that can be formed (Beier, 1993; Spinardi *et al*., 1991).

However, the real importance and potential for application of SNPs lies in their automated detection using DNA chip technology (Chee *et al*., 1996). Although individually not highly informative, SNPs are so numerous that it will be possible to incorporate thousands of SNPs onto single DNA microchips, enabling enormously sensitive parentage testing and rapid and automated genome scanning for QTL (Brookes, 1999).

1.1.2. Gene mapping

A gene map records the location(s) of gene(s) or marker(s) of interest. There are two quite different mapping methods, namely linkage mapping and physical mapping. A linkage map is based on the recombination frequency (r) between genes or markers detectable among the offspring of parents with distinguishable alleles at two or more loci. On the other hand, a physical map does not require variants, but assigns loci to chromosomal positions either by *in situ* hybridisation or by using somatic hybrid panels.

1.1.2.1. Linkage (Genetic) mapping

Loci on different chromosome segregate independently. On the other hand, loci on the same chromosome may show evidence of co-segregation or linkage. The recombination frequency used to estimate the map distance between a pair of loci ranges from 0 for tightly linked loci to 0.5 for loci that are far apart or on different chromosomes. A reference pedigree, also called a mapping resource, can be constructed by crossing genetically divergent populations to increase the informativeness of the parents or can be based on existing families in a population. The construction of a linkage map is a very useful step for the identification of genes

having quantitative effects on traits of interest (so called quantitative trait loci or QTLs) (Archibald and Haley, 1998).

1.1.2.2. Cytogenetic (Physical) mapping

Physical mapping identifies the location of genes or genetic markers on the individual chromosomes or regions of chromosome without analysis of segregation between loci. There are three major physical mapping techniques which allow the identification of the chromosomal location of specific DNA fragments, namely use of i) *in situ* hybridisation, ii) somatic cell hybrids, and iii) radiation hybrids.

ISH (In Situ Hybridisation)

A fragment of DNA of interest can be radioactively or fluorescently labelled and allowed to hybridise to metaphase chromosome spreads. The development of fluorescent *in situ* hybridization (FISH) has almost replaced the use of radiolabelled probes. The sensitivity of detection by FISH is dependent upon the probe size and type of tag used in labelling the probe (Gillois, 1991). The great advantage of FISH compared with radioactive *in situ* hybridization (RISH) is that not only is it safer, but also multicolour analysis is amenable to detection of different loci on the same metaphase spreads using differently coloured dyes (Trask, 1991a; Trask, 1991b).

Somatic cell hybrid mapping

Somatic cell hybrid panels have played a significant role in physical mapping of known genes and markers in humans and livestock species (Payen *et al*., 1995; Rettenberger *et al*., 1994; Robic *et al*., 1996; Williamson *et al*., 1995). The technique basically involves fusion of cell lines originating from different species. Somatic cell hybrids between a donor species (e.g. pig) and a recipient species (normally Chinese hamster or mouse) randomly lose donor chromosomes or fragments of chromosomes so that sublines of cells contain only a few specific donor chromosomes or chromosomal fragments. A panel of these sublines can be tested for presence or absence of a particular DNA sequence and this can then be correlated with presence or absence of a particular chromosome or chromosomal segment to identify the chromosome or chromosomal region where the locus is located (Bosch *et al*., 1993; Iles *et al*., 1992).

In pigs, early attempts at mapping with a somatic cell hybrid panel used Southern blotting techniques (Ruddle, 1981). PCR-based analysis of somatic cell hybrid panels is now routinely applied (Rettenberger *et al*., 1995a; Rettenberger *et al*., 1996; Rettenberger *et al*., 1994; Zijlstra *et al*., 1996).

The INRA laboratories in Toulouse, France, have generated and distributed DNA from a very informative porcine somatic cell hybrid panel which can be used to assign DNA loci to specific chromosomes and chromosomal regions (Yerle *et al*., 1996). This panel consists of 27 pig × rodent somatic cell hybrids in which 127 different subchromosomal regions can be identified as a result of fragmentation of the porcine chromosomes by low level irradiation (see radiation hybrid below). It is an efficient tool for physical mapping and has been used in the study described in this thesis.

Radiation hybrid mapping

Radiation hybrid mapping provides very high-resolution maps (McCarthy, 1996). The technique is identical to construction of a conventional somatic hybrid panel except that the donor (e.g. pig) chromosomes are fragmented by a high dose of X-ray irradiation. The broken fragments are recovered in recipient cells (e.g. rodent), and hybrid clones are analysed for the presence or absence of DNA markers. Whole genome radiation hybrid mapping has become a mainstream method for high-resolution human gene mapping (Gyapay *et al*., 1996; Stewart *et al*., 1997) as well as animal gene mapping (Kwok *et al*., 1998; Murphy *et al*., 1999; Womack *et al*., 1997; Yerle *et al*., 1998). In 1998, a radiation hybrid panel became publicly available for mapping studies in pigs, jointly developed by INRA, France and the University of Minnesota, USA, called the INRA-University of Minnesota porcine Radiation Hybrid (IMpRH) panel. It consists of 118 hybrid clones, and the resolution of the radiation hybrid map is higher than that obtained by linkage analysis from available mapping families, so that unresolved markers in the linkage map can be ordered (Yerle *et al*., 1998).

1.1.3. Pig genome mapping

Pigs have a number of advantages in gene mapping studies. Firstly, pigs produce large litters of ten or more and have a relatively short gestation period of about 114 days. This allows the rapid establishment of experimental pedigrees for mapping studies. Secondly, genetically divergent breeds are available for creating highly informative reference and resource pedigrees. By crossing genetically divergent breeds, the level of heterozygosity in the F1 animals is increased and the pedigree is much more likely to be highly informative in the F2 generation. Finally, pigs have a relatively small chromosome number (2n=38) as compared with other domestic animal species like cattle (2n=60), sheep (2n=54), goat (2n=60), dog (2n=78), horse (2n=64) and chicken (2n=78). Also pig chromosomes are easily distinguishable from each other especially with the help of chromosome banding methods. Therefore physical mapping techniques such as *in situ* hybridisation and chromosome sorting are easier to apply.

In August 1989, there were only 42 markers mapped in the pig, 15 of which were assigned to 7 linkage groups, and 27 markers were mapped in 10 chromosomes (Echard, 1990). Intense efforts were made during the 1990s to develop a comprehensive genome map of the pig, particularly benefiting from the availability of hyperpolymorphic microsatellite markers. Three international research groups have been involved in linkage mapping the porcine genome, namely the Pig Gene Mapping Project (PiGMaP) consortium (Archibald *et al*., 1995), the Scandinavian group (Ellegren *et al*., 1994; Marklund *et al*., 1996) and the United States Department of Agricuture (USDA) group (Rohrer *et al*., 1996; Rohrer *et al*., 1994). The combined efforts of these groups have been placed mare than 1500 polymorphic genetic markers on the porcine linkage map and many genes and markers also have been physically mapped (Yerle *et al*., 1997; PiGBASE, http://www.ri.bbsrc.ac.uk/pigmap/pigbase/pigbase.html).

1.1.3.1. PiGMaP and PiGMaP reference families

The European Pig Gene Mapping Project (PiGMaP) consortium was initiated in 1989 by a group of 10 laboratories within the PiGMaP collaboration (Haley *et al*., 1990). The initial aim of PiGMaP was to map markers at evenly-spaced intervals of approximately 20 cM, covering the whole genome. The group also set out to produce a comprehensive physical map within 3-4 years (Echard *et al*., 1992; Haley *et al*., 1990). Laboratories in Europe, Japan, United States and Australia (Associate Professor C. Moran's laboratory) were included. Echard *et al*. (1992), summarising the early status of PiGMaP, indicated that even by 1992 a total of 84 loci had been assigned to 17 chromosomes. With the arrival of microsatellite markers, a comprehensive linkage map was rapidly achieved. Archibald *et al*. (1995) reported the first PiGMaP linkage map consisting of 245 markers, 81 corresponding to known genes, with the rest being anonymous microsatellite and RFLP markers. Linkage groups were assigned to all 18 autosomal chromosomes and the X chromosome. The number of markers mapped in the pig is still growing. Currently over 2000 loci and more than 3500 markers are described in the linkage database (PiGBASE: http://www.ri.bbsrc.ac.uk/cgi-bin/arkdb/browsers/browser.sh?species=pig).

The PiGMaP reference population consists of six families of a three-generation pedigree based on crosses between genetically divergent breeds. Five centres in Europe contributed animals to the shared PiGMaP mapping pedigrees, from Scottish, Dutch, French, Swedish and German resources. Three divergent pig types were used to establish the PiGMaP reference pedigree, namely i) European commercial breeds, Large White and Pietrain, which have good commercial performance traits; ii) European Wild Boar, whose appearance and physiological properties such as growth rate as well as fat content are very distinct from modern European breeds (Andersson *et al*., 1994a); and iii) Chinese Meishan, which is highly fecund but otherwise

very unproductive. By crossing such phenotypically divergent breeds, it was hoped that it would be easier to map loci controlling monogenic as well as polygenic traits of interest (Andersson *et al*., 1994a).

1.1.4. Application of gene mapping

1.1.4.1. Identification of QTL

Quantitative genetic variation is the major component of intra- and interpopulation differences for many traits of biological, medical and agricultural significance. The development of porcine genome maps allows the identification of specific markers associated with genes influencing economically important traits, which are called QTLs (Quantitative Trait Loci) (Andersson *et al*., 1994b; Archibald *et al*., 1995). QTLs are those loci whose variation jointly contributes to the quantitative trait variation observed between animals in a population. The general principle is simple: if a QTL for a particular trait is closely linked to a marker, the marker alleles will appear to be associated with different levels of performance for the trait. Genome maps provide two advantages for the detection of QTLs. First, they enable choice of the minimum number of markers, evenly spaced in the genome, that will enable a scan of the entire genome. Second, knowing the location of marker linked to a QTL allows the approximate location of the QTL, and this information can be used for selection of candidate gene(s) in the same other species. New statistical and experimental methods are being developed to map QTLs in different types of population structure (Archibald and Haley, 1998; Crawford *et al*., 2000). Once mapped, QTLs have important potential applications in breeding programmes, including marker assisted selection, marker assisted introgression and less directly, positional cloning. Several QTLs have been recorded in pigs (e.g. Andersson *et al*., 1994b).

1.1.4.2. Marker Assisted Selection (MAS)

The progress of genome mapping and the knowledge of co-segregation of a marker allele and a trait may be used to improve efficiency of selection of animals. MAS (Archibald, 1994; Soller, 1994) is particularly relevant for traits which cannot easily be measured in the candidates for selection, such as sex-limited traits, reproductive traits, slaughter traits and processing traits.

1.1.4.3. Identifying the genes underlying QTL

The positional candidate gene approach can help to identify the gene causing variation in a particular trait. Once a QTL has been mapped, candidate gene loci in the vicinity of the QTL can be evaluated for their effects on the trait. This approach has been noticeably successful for the isolation of disease genes (Collins, 1995; Fujii *et al.*, 1991). The success of this approach is greater if all possible candidates have been identified, for example by genome sequencing. For species whose genomes are poorly characterised, such as the pig, cross reference to a well characterised genome, such as human or mouse, will be essential. By placing many type I markers, that is evolutionary conserved coding sequences, on the genetic maps of many mammalian species, one can get a clear picture of syntenic relationships. This augments the number of positional candidate genes in less well studied species by providing large numbers of comparative positional candidates.

1.1.4.4. Comparative gene mapping

Comparative gene mapping, involving comparison of chromosomal locations of homologous genes in different species, is a useful source of information in terms of identifying candidate disease genes and genes determining complex traits. It also facilitates gene mapping in other species and provides insights into genome organisation and evolution (DeBry and Seldin, 1996; Eppig, 1996; Eppig and Nadeau, 1995; Nicholas and Harper, 1996; Womack and Kata, 1995). The first step in the comparative gene mapping approach is the recognition of regions of conserved synteny, that is, groups of linked loci conserved among species. A large effort has been made in recent years to define the conserved synteny between species.

In order to recognise evolutionary breakpoints between species, a set of anchored reference loci suitable for comparative mapping in mammals was first proposed by O'Brien *et al.*, (1993). Subsequently, Lyons *et al.* (1997) designed 410 evolutionary conserved primers to amplify anchor loci in different species, calling them Comparative Anchor Tagged Sequence (CATS) markers. Once the CATS markers have been mapped in each species, it will be easier to detect chromosomal rearrangements and assist in reconstructing mammalian chromosome evolution. However, the CATS approach using consensus primers has not proved to be very efficient, especially in pig, with only a small number of localisation having been obtained (see Chapter three for more detail). Recently, cross-species chromosome painting or Zoo-FISH (Scherthan *et al.*, 1994) has contributed significantly to identification of conserved synteny between species (e.g. Rettenberger *et al.*, 1995b).

1.2. Retroviruses

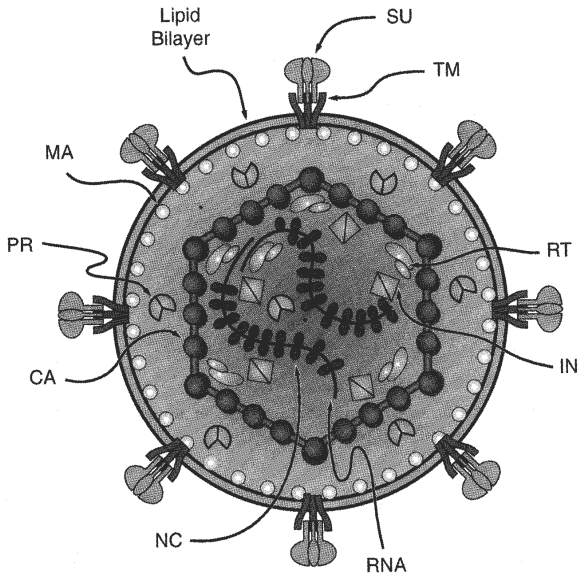
1.2.1. Structure and life cycle of retrovirus

Retroviruses are a group of viruses whose genetic material is comprised of RNA rather than DNA. Retroviruses are associated with a variety of diseases including malignancies, immunodeficiencies, and neurologic disorders (Rosenberg and Jolicoeur, 1997). Normally, the virions are 80 to 100 nm in diameter and their outer envelope is constituted of glycoproteins. Each retrovirus particle contains two copies of 8 to 10 kb genomic RNA held together by hydrogen bonding in a dimeric structure. Broadly, retroviruses are classified into two different groups, namely the simple and the complex, which can be distinguished by the organisation of the RNA genome. Compared to a simple retrovirus genome containing only four major coding regions, *gag*, *pro*, *pol*, and *env,* a complex retroviral genome contains information for regulatory proteins besides the major coding regions (Vogt, 1997).

The *gag* region contains the information for the synthesis of internal virion proteins that form the matrix (MA), the capsid (CA) and the nucleocapsid (NC) protein structure. The *pol* region directs the synthesis of reverse transcriptase (RT) and integrase (IN) enzymes; and the surface (SU) and transmembrane (TM) components of the viral envelope protein are encoded by the *env* coding region. In addition, a small coding domain present in all retroviruses is *pro*, which encodes the virion protease (PR) (Vogt, 1997). Protease is responsible for all the proteolytic cleavages generating the mature *gag* and *pol* proteins during virion maturation (Coffin, 1992) (Fig 1.1). This standardised nomenclature for the proteins common to all retroviruses is on the basis of biological function, enzymatic activity and/or virion location data (Leis *et al*., 1988).

The retroviral life cycle starts from the attachment of the viral surface glycoproteins to specific host plasma membrane receptors, which leads to fusion of virus and cell membranes. The integration of viruses (exogenous viruses) and cell surfaces is highly specific and determines the viral-host range. After penetration into the cell, the RNA genome is transcribed into DNA by reverse transcriptase. This DNA copy is then inserted into the host genome, at which stage it is called an endogenous retrovirus or DNA provirus. Reverse transcription of the RNA genome also generates long terminal repeats (LTRs) found at both ends of the DNA provirus (Figure 1.2).

(a)



(b)

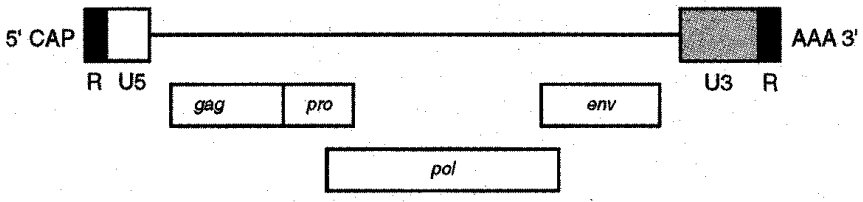


Fig 1.1 (a) Schematic cross section through a retroviral particle. (b) A simple retroviral genome which contains four major coding regions (Source: Vogt, 1997).

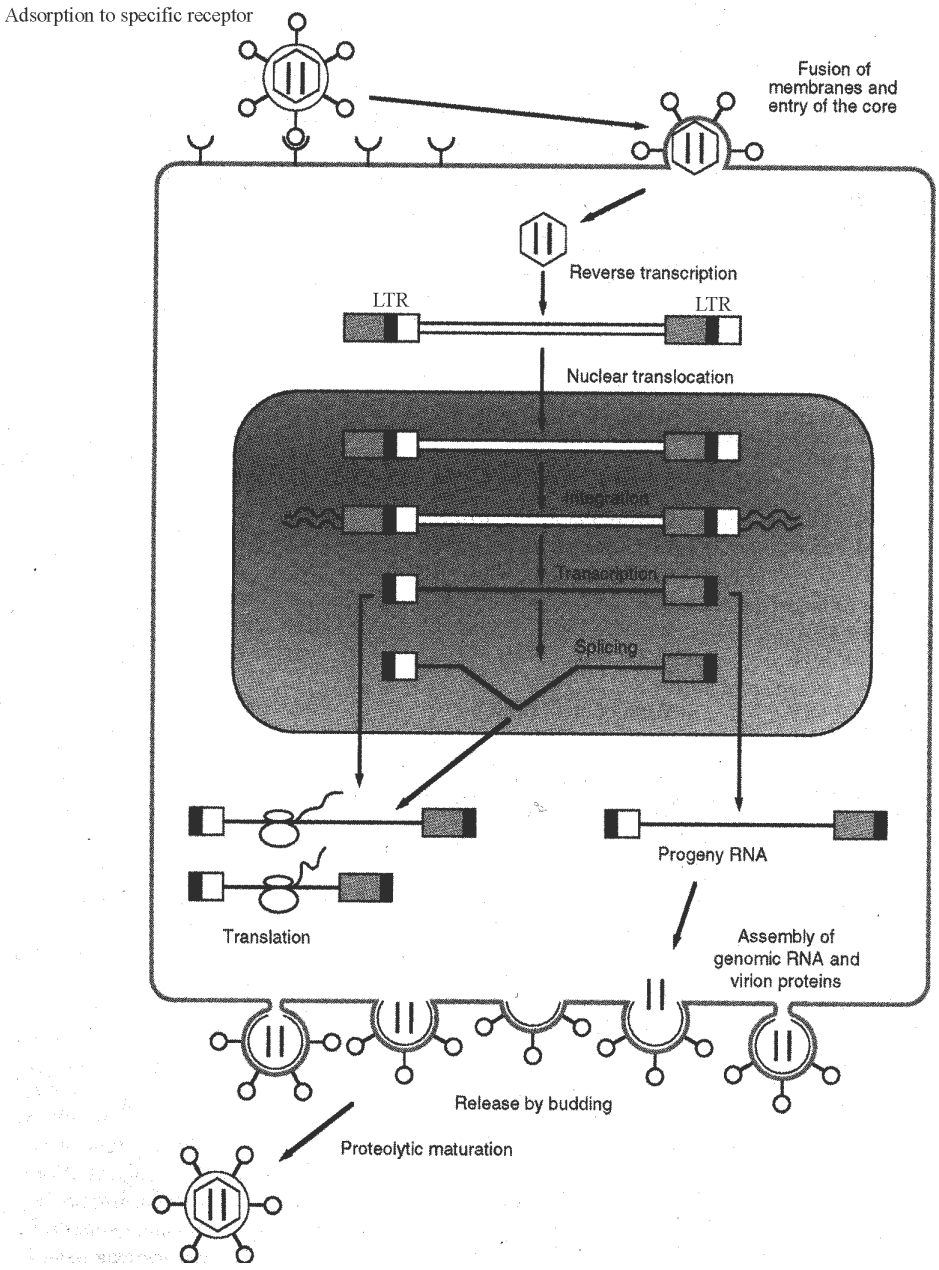


Fig 1.2 The life cycle of a retrovirus (Source: Vogt, 1997).

1.2.2. Classification of retroviruses

The classification and nomenclature for retroviruses presented here is based on Vogt (1997). Retroviruses are subdivided into seven genera (Table 1.1) of which five have oncogenic potential. All oncogenic members except the human T-cell leukaemia virus-bovine leukemia virus (HTLV-BLV) genus are simple retroviruses (Vogt, 1997).

Based on their morphology under the electron microscope, retroviruses can be classified into four categories, namely A, B, C, D type particles. A type particles are 60 – 90 nm in diameter with an electron-lucent center surrounded by a double shell. They occur as intracellular forms only and do not have any infectivity.

B type particles have two morphologically distinctive features. One is that budding particles at the plasma membrane show toroidal (doughnut-shaped) cores about 75 nm in diameter and long spikes are seen at the cell surface. The other feature is that the mature forms contain electron-dense nucleoids about 125 – 130 nm in diameter that are eccentrically located within the enveloped particle.

Table 1.1 Classification of retroviruses.

Genus	Example	Virion morphology[a]	Genome
Avian sarcoma and Leukosis viral group	Rous sarcoma virus	Central, spherical core C particles	Simple
Mammalian B-type Viral group	Mouse mammary tumor virus	Central, spherical core B particles	Simple
Murine leukemia Related viral group	Moloney murine leukemia virus	Central, spherical core C particles	Simple
D-type viral group	Mason-Pfizer monkey virus	Cylindrical core D particles	Simple
Human T-cell Leukemia, bovine Leukemia viral group	Human T-cell leukemia virus	Central, spherical core	Complex
Lentiviruses	Human immuno-deficiency virus	Cone-shaped core	Complex
Spumaviruses	Human foamy virus	Central, spherical core	Complex

[a]Distinctive features seen in transmission electron micrographs. (Source: Vogt, 1997)

The majority of retroviruses isolated to date are classified as C type particles. The first distinct viral structure at the plasma membrane is an electron-dense crescent-shaped form that will be the core of the particle. As virus maturation proceeds, the core is eventually seen as a sphere with an electron-lucent center. The C particles have an immature form with centrally located electron-lucent cores, which matures into a centrally located electron-dense core.

D type is typical of many retroviruses found in primates. They resemble B type particles in assembly, maturation and morphology but have less prominent surface projections and a characteristic cylindrical core (Coffin, 1992; Teich, 1982).

1.2.2.1. Retroviruses in different host species

Avian retroviruses

The discovery of retroviruses arose from investigations of neoplastic disease in chickens revealing the existence of Avian leukemia virus (ALV) (Ellermann and Bang, 1908) and Rous sarcoma virus (RSV) (Rous, 1911). These viruses constitute the C-type virus genus, referred to as avian sarcoma/leukemia viruses (ASLV) (Vogt, 1997). Payne (1992) describes the classification of avian retroviruses into four distinct classes.

(1) Leukosis-sarcoma group viruses (also termed avian sarcoma and leukemia viruses). These are exogenous and endogenous viruses recognised mainly in the domestic fowl and which cause a variety of leukotic disorders, sarcomas, and other tumors.

(2) Reticuloendotheliosis viruses. These are exogenous viruses in several species of domestic poultry and appear to be related to mammalian retroviruses. They cause lymphomas and acute reticulum cell and other tumors.

(3) Lymphoproliferative disease virus of turkeys. This is an exogenous virus of turkeys which causes a lymphoproliferative disease.

(4) Pheasant type C oncoviruses. These are endogenous viruses of golden and Lady Amhurst pheasants and are apparently nonpathogenic. They are unrelated to the endogenous leukosis-sarcoma group viruses that also occur in certain species of pheasants.

Murine retroviruses

The retroviruses of rodent species, especially the mouse, have been widely studied because of the short generation length, large litter size and amenable breeding strategy of the host. The

mouse retroviruses can be classified into four groups based on the differences in virion morphology and sequence homology. There are type C murine leukemia viruses (MuLVs or MLVs), type B mammary tumor viruses (MMTVs), type A intracisternal particles (IAPs), and VL30s (Kozak and Ruscetti, 1992).

The C-type viruses are wide distributed both as exogenous and endogenous forms and can cause a wide variety of malignancies, immunosuppression and neurological disease (Petropoulos, 1997). The C-type viruses are classified as ecotropic (capable of growing in normal mouse cells but not in cells from other species), xenotropic (incapable of growing in normal mouse cells but growing well in cells from other species), polytropic (capable of growing well in mouse cells and cells of other species), and amphotropic (capable of growing well in mouse cells and cells of other species, but distinct in antigenic and interference properties compared with polytropic) (Kozak and Ruscetti, 1992).

Mouse mammary tumor viruses (MMTVs) represent a class of murine retroviruses distinct from the MuLVs by virion morphology, by the absence of sequence homology, and by their sensitivity to induction by glucocorticoid hormones (Kozak and Ruscetti, 1992). MMTVs are associated with mammary carcinoma and T-cell lymphomas and can be inherited vertically in certain inbred strains or transmitted horizontally through the milk from the affected females (Petropoulos, 1997).

IAPs are found only in association with the endoplasmic reticulum and have no extracellular phase (Kozak and Ruscetti, 1992). IAPs are not known to encode replication-competent viruses. However, they are present in mice at approximately 1000 to 2000 copies per cell and are often expressed in plasma-cell tumors at high level (Boeke and Stoye, 1997).

VL30s are endogenous retrovirus-like sequences which are not known to produce any virion structural components, but can be efficiently packaged and transmitted as pseudotypes of type C viruses (Kozak and Ruscetti, 1992).

Feline retroviruses

Based on Hardy (1992), feline species have three major groups of retroviruses, namely Oncovirus, Lentivirus and Spumavirus.

The oncovirus group is now classified as the Murine leukemia virus (MLV)-related retrovirus genus and is present in both endogenous and exogenous forms. The virions exhibit a C type morphology and Feline sarcoma and leukemia viruses (FeSV/FeLV) are included. They can causes malignancies, immunosuppression, and neurological disease similar to C type murine leukemia viruses in mice (Petropoulos, 1997).

FIV (Feline immunodeficiency virus) and FSV (Feline syncytial virus) belong to the Lentivirus and Spumavirus genera, respectively. FIV induces an AIDS-like syndrome in cats. FSV pathogenicity has not been demonstrated (Petropoulos, 1997).

Bovine retroviruses

Two retroviruses in the HTLV/BLV genus have been recognised in bovine species. They are the Bovine leukemia virus (BLV) and Bovine immunodeficiency virus (BIV). BLV is the cause of enzootic bovine leukosis (Kettmann *et al*., 1992). BLV can be transmitted vertically by the transfer of infected lymphocytes in milk from a dam to her calf, or horizontally between cattle by biting insects or via contaminated instruments (Pelzer and Sprecher, 1993). BIV is a horizontally transmitted, nononcogenic, pathogenic retrovirus and member of the Lentivirus genus. There has been a interest in BIV because Human immunodeficiency virus (HIV) is also included in the Lentivirus genus (Gonda, 1992).

Other retrovirus genera also exist in bovine species. These are Bovine pulmonary adenocarcinoma virus, a type D retrovirus genus, and Bovine syncytial virus (BSV) in Spumavirus genus. The bovine pulmonary adenocarcinoma virus is the bovine version of the Jaagsiekte sheep retrovirus (JSRV) (Petropoulos, 1997).

Ovine retroviruses

JSRV (Jaagsiekte sheep retrovirus) is a type D retrovirus, comprising both endogenous and exogenous viruses of sheep. These viruses are associated with pulmonary cancer in sheep (Petropoulos, 1997). Similar endogenous proviruses are detected in goats and other ungulates (Hecht *et al*., 1996).

Maedi and Visna virus (MVV) and Caprine arthritis encephalitis virus (CAEV) are caused by ovine lentiviruses (OvLV) (Petropoulos, 1997). Maedi, an Icelandic word meaning dyspnoea, is used to describe a chronic interstitial pneumonia, and Visna refers to a slow, progressive disease of the central nervous system. Maedi and Visna, therefore, are chronic multisystemic diseases of sheep (DeMartini *et al*., 2000). CAEV also induces diseases of the central nervous system (CNS) (McGuire *et al*., 1990).

Human retroviruses

Three different genera of retroviruses are identified in humans. They are Human T-cell leukemia virus (HTLV) in the HTLV/BLV genus, Human foamy virus (HFV) in the Spumavirus genus and Human immunodeficiency virus (HIV) in the Lentivirus genus (Petropoulos, 1997).

HTLV can be divided into two species (HTLV-1 and HTLV-2) and only exogenous viruses are known. Infection with HTLV is associated with B- and T-cell leukemias, lymphomas

and neurological disease (Petropoulos, 1997). HIV is an exogenous retrovirus and no closely related endogenous retrovirus have been described. Members of the Lentivirus genus, including HIV, are the causative agents of immunodeficiencies, neurological degeneration and arthritis (Petropoulos, 1997).

Porcine retroviruses

Porcine type C viruses are classified in the genus Murine leukemia virus (MLV)-related virus (Petropoulos, 1997). The first description of porcine C type retroviruses was from cultured pig kidney cells (Breese, 1970; Armstrong *et al*., 1971). These porcine retroviruses are associated with lymphosarcomas and tumors (Bostock and Owen, 1973). There has been emerging interest in porcine retroviruses in relation to the potential use of pig organs and tissues for xenotransplantations (Cozzi and White, 1995) and the theoretical hazards to transplant recipients associated with endogenous proviruses (Stoye and Coffin, 1995).

Recently, much research has been carried out on porcine retroviruses. Retroviral infection has been investigated in humans and baboons, both *in vitro* (Martin *et al*., 1998a; Patience *et al*., 1998; Wilson *et al*., 1998) and *in vivo* (Martin *et al*., 1998b; Paradis *et al*., 1999; Pitkin and Mullon, 1999; Switzer *et al*., 1999).

1.2.3. Retroviral variation

1.2.3.1. Mutation and deletion

Retroviruses are prone to errors during replication (Temin *et al*., 1991) and variation among retrovirus populations has been observed since their first discovery (Rous and Murphy, 1913). Three different polymerases are involved in retrovirus replication, namely reverse transcriptase, DNA polymerase, and RNA polymerase II (Pathak and Temin, 1990b). Cellular DNA polymerases have a very low mutation rate because of their proofreading mechanisms, and their contribution to the retroviral mutation rate is negligible (Drake, 1969). The rate of mutation caused by RNA polymerase II has not been measured and its contributions to retroviral mutation rates is unknown (Pathak *et al*., 1990). On the other hand, the mutation rate of avian myeloblastosis virus (AMV) and human immunodeficiency virus (HIV) reverse transcriptase determined by *in vitro* assay has been estimated to be 2.3×10^{-5} and 4.3×10^{-5} substitutions per bp per polymerisation cycle, respectively (Preston *et al*., 1988). Gojobori and Yokoyama (1985) compared the rate of nucleotide substitution for the retroviral *v-mos* oncogene of Moloney murine sarcoma virus (Mo-MuSV) and its cellular counterpart *c-mos*. They estimated the rate of nucleotide substitution for *v-mos* and *c-mos* as 1.31×10^{-3} and 1.71×10^{-9} substitutions per site

per year, respectively. The difference in rate of mutations between *v-mos* and *c-mos* confirmed that DNA polymerase-based mutation was far less frequent than reverse transcriptase-based mutation as indicated by Drake (1969). Another study measured the forward mutation rates *in vivo* for spleen necrosis virus (SNV). Substitutions occurred at a rate of 7 X 10^{-6} per nucleotide per cycle and frameshifts occurred at a rate of 1 X 10^{-6} per nucleotide per cycle (Pathak and Temin, 1990b). Deletions and deletions with insertions occurred at a similar rate of 2 X 10^{-6} per nucleotide per cycle (Pathak and Temin, 1990a).

1.2.3.2. Recombination

Since retroviral particles have two genomic RNAs in their virus particles, they can easily recombine. The study of retrovirus recombination suggests a selective value for recombination in removing genomic damage and in increasing the amount of variation in retrovirus populations (Temin, 1991).

To measure the rate of recombination in a single cycle of retrovirus replication, helper cells were infected with spleen necrosis virus-based splicing vectors containing either the neomycin-resistance gene (*neo*) or the hygromycin B phosphotransferase gene (*hygro*). Progeny proviruses were screened for proviruses expressing both wild type genes and were then characterized by restriction enzyme digestion. The rate of retroviral recombination is about 2 % per 1 kb per replication cycle (Hu and Temin, 1990). Another experiment showed that recombination between an homologous sequence of about 560 bp occurred with a frequency of about 10^{-4} per virus replication cycle (Stuhlmann and Berg, 1992). Recently, recombination between species has been identified. For example, BaEV (baboon endogenous retrovirus) is a recombinant between two primate viruses (Mang *et al*., 1999; van der Kuyl *et al*., 1997) and RD-114 in cat is a recombinant between baboon and primate retroviruses (van der Kuyl *et al*., 1999). Also, Jaagsiekte sheep retrovirus (JSRV) appears to be a type B and D chimera because it contains type D capsid sequences and type B envelope sequences (Hecht *et al*., 1996).

1.2.4. Retroelements

Retroelements are dispersed repetitive elements whose movement is mediated via an RNA intermediate. These elements, also called retroposons or retrotransposons, consist of SINEs (**S**hort **I**nterspersed **N**uclear **E**lements), LINEs (**L**ong **I**nterspersed **N**uclear **E**lements), endogenous retroviruses and processed pseudogenes (Moran, 1998). The common feature of all retroelements is that DNA copies have been made from RNA templates using reverse transcriptase, and then integrated into the host genome.

1.2.4.1. Endogenous retroviruses (ERVs)

ERVs are very common in the mouse genome, comprising up to 5 % of the genome, whereas in humans, they have been estimated to constitute up to 0.1 % of germ-line DNA (Patience *et al*., 1997b). Complete ERVs, like the RNA of simple exogenous retroviruses, consist of the *gag*, *pol*, and *env* genes flanked by LTRs, although many endogenous retroviruses may be defective and undergoing mutational decay. Although many are capable of completing the infectious viral cycle and thus are capable of horizontal transmission, most endogenous retroviruses are stably inherited and are vertically transmitted only (Patience *et al*., 1997b).

1.2.4.2. LINEs

LINEs (Long Interspersed Nuclear Elements) are a class of retroposon distinguishable from endogenous retroviruses by the lack of *env* sequences (Moran, 1998). LINEs are defined as repeated sequences of DNA greater than 5 kb in length. They are usually present at copy numbers of greater than 10^4 per mammalian genome (Hutchison *et al*., 1989; Singer, 1982). Most LINEs are truncated and stranded in their current genomic location, where they are doomed to gradually decay due to unconstrained mutation, even though some LINEs have at least a theoretical potential for further movement (Moran, 1998).

One common LINE family in mammalian species, LINE-1 or L1, contains two open reading frames (ORFs) potentially encoding proteins, although these ORFs are frequently interrupted by mutation (Martin, 1991). All LINEs, including the truncated ones, end with polyadenylation or at least an A rich region, and are bounded by short direct repeats. Some features such as the lack of LTRs, the presence of an internal RNA polymerase II promoter and reverse transcriptase-like sequences are likely to be common to all LINE elements (Martin, 1991).

Homologous sequences of the L1 family have been found in animals, insects, higher plants and even in protozoans suggesting that it may be even the ancestor of retroviruses (Hutchison *et al*., 1989). Only about 10 % of mouse L1 elements are the full length of about 7 kb, with the remainder being shorter truncated elements down to 500 bp in size. The L1 family alone is present at more than 80,000 sites within the mouse genome. Given the universal distribution of L1-like elements in eukaryotes, a similar number of related elements would be expected in the porcine genome (Moran, 1998). Miller (1994) has described the existence of porcine LINEs. Porcine LINEs are uniformly distributed throughout the euchromatic part of porcine genome, with a slight bias towards G-bands (Thomsen and Miller, 1996).

1.3. Xenotransplantation

Xenotransplantation is the transplantation of cells, tissues or organs between different species. The interest in xenotransplantation has been growing during the past decade because of the potential for using pigs as a source for unlimited donor organs and tissues to overcome a shortage of human materials for clinical transplantation. For example, 4000 people in the USA died waiting for a transplant in 1998 and every 16 minutes a new name is added to the waiting list (Melton, 1999). The comparison between waiting lists and transplants is illustrated in Fig 1.3 for the UK, and similar trends are reported throughout the world.

According to White and Nicholson (1999), clinical transplantation can be divided into two eras separated at the middle of the 1960s, prior to and after availability of treatment for immunosuppression. Before the availability of appropriate immunosuppression, the results of transplants were very poor. The use of immunosuppressive chemicals like cyclosporin provided long-term survival of grafted organs but could not eliminate progressive graft rejection (Bailey *et al*., 1985). There have been a few attempts during the past century to transplant animal organs into humans. More recently, molecular and immunological understanding of xenograft rejection and the feasibility of genetic modification of donors has made xenotransplantation more likely (Platt and Lin, 1998; Vanhove *et al*., 1998).

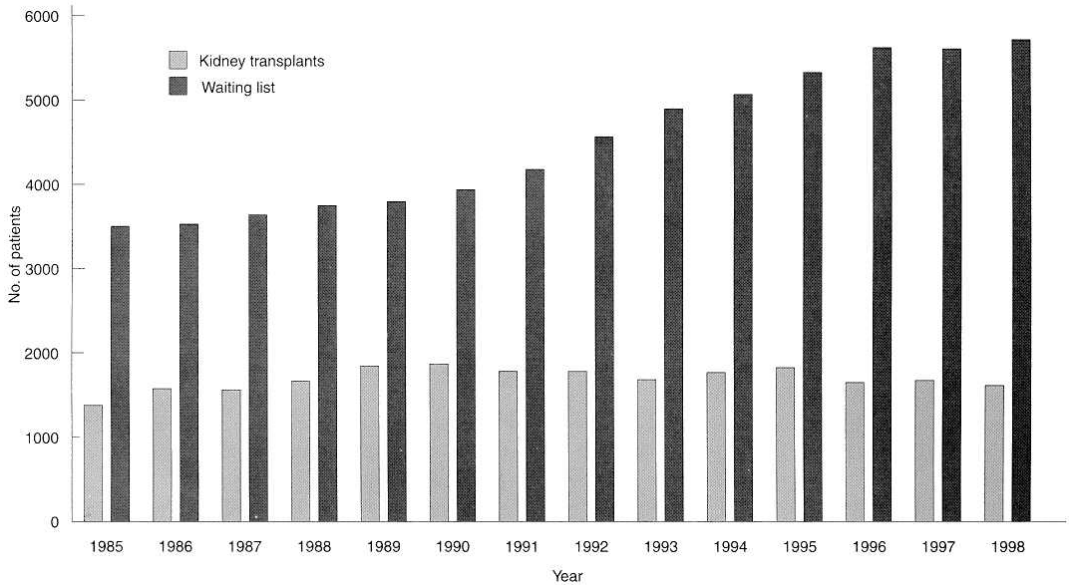


Fig 1.3 Discrepancy between number of patients awaiting renal transplantation and the number of renal transplants performed in UK (United Kingdom Transplant Support Services Association, 1998).

For several reasons, pigs rather than non-human primates are expected to be the best sources of xenograft donors. Firstly, porcine organs are physiologically and anatomically similar to those of humans. Secondly, pigs can be bred economically and quickly in large numbers in a disease-free environment. Finally, there are few ethical qualms about breeding pigs for slaughter for xenotransplantation because pigs are already bred for human consumption (White and Nicholson, 1999).

1.3.1. Barriers to xenotransplantation

1.3.1.1. Hyperacute rejection

Hyperacute rejection is the first major barrier to xenotransplantation and occurs immediately after exposure of the donor organs to a recipients blood flow. A single donor gene and its products are responsible for the hyperacute rejection of organs transplanted from non-primates into primates. A disaccharide sugar, galactose α (1,3) galactose, called α-gal antigen, is present as the terminal residue of glycoproteins and glycolipids on the surface of cells from pigs and other mammals but is absent from higher apes (Mollnes and Fiane, 1999). Naturally occurring antibodies in primates, formed in response to the same antigen in bacteria, recognise galactose α (1,3) galactose. This antigen-antibody system activates the complement cascade causing hyperacute rejection. Hyperacute rejection leads to destruction of the graft within minutes to a few hours (Platt and Bach, 1991). Endothelial cells in blood vessels are the primary target of the host immunity in hyperacute rejection (Platt *et al*., 1990).

Under normal circumstance, the complement cascade is inhibited by a number of proteins in the plasma and on the surface of the cells (Hourcade *et al*., 1989). These protective proteins include Decay accelerating factor (DAF), Membrane cofactor protein (MCP) and CD59. They are also collectively called regulators of complement activation (RCAs) (Cozzi and White, 1995).

1.3.1.2. Cell-mediate immunity

Cell-mediated rejection is becoming of increasing importance as it causes delayed xenograft rejection (White and Nicholson, 1999). Major histocompatibility complex (MHC) antigens play a major role in eliciting this response, even though the mechanism of cell-mediated immunity is not fully understand (White and Nicholson, 1999).

1.3.2. Genetic manipulation to avoid xenograft rejection

Knocking out α (1,3) galactosyl transplantation

The α (1,3) galactose epitope is the major xenoantigen expressed on porcine cells and is responsible for hyperacute rejection. All humans have circulating antibodies, which recognise gal antigen since similar antigens are expressed in bacteria. Exposure of recipients to porcine tissue expressing gal antigen causes activation of the complement cascade and ultimately hyperacute rejection. The enzyme, α (1,3) galactosyl transferase, creates this antigen and it would be desirable to eliminate it from porcine organs to be used for xenotransplantation. This porcine gene was cloned by Dabkowski *et al.* (1994). Homologous recombination provides a theoretical mechanism for knocking out the function of α (1,3) galactosyl transferase but all attempts to develop porcine embryonic stem (ES) cells have failed (Vanhove *et al*., 1998). Very recently, five piglets were cloned by nuclear transfer (Bradbury, 2000) bringing closer the possibility of a knockout of α (1,3) galactosyl transferase and thus ultimately overcoming this major hurdle to xenotransplantation.

Alternatively, Ogawa *et al*. (1999) demonstrated that transfected cells for splicing variants of α (1,3) galactosyl transferase suppress the expression of α-gal antigen. The transgene cDNAs without exon 5 and 8 (Δ58), with insertion of 47 bp between exon 7 and 8 (Full+47) or with two point mutations in exon 9 (Mut 9), reduced α-gal expression in COS7 cells by 27%, 36%, 23%, respectively. The Δ58, Full+47 and Mut 9 showed 9%, 36% and 42% suppression of α-gal expression in PK15 cells.

Over expressing H-transferase

The most widely explored alternative mechanism for preventing expression of galactose α (1,3) galactose on porcine glycoproteins is to overexpress another glycosyl transferase, α (1,2) fucosyl transferase or H-transferase. The overexpression of H-transferase suppresses the appearance of galactose α (1,3) galactose in mouse and pigs (Sharma *et al*., 1996). However, this effect of reducing Gal expression is dependent on cell type in the mouse (Chen *et al*., 1998). Transgenics for the fucosyl transferase gene in the rat have developed carcinoma of the colon, raising fears that overexpression of this enzyme and its product might be pathogenic in pigs (Hallouin *et al*., 1999).

Protective proteins

Protective proteins like DAF (Decay accelerating factor, CD55), MCP (Membrane cofactor protein, CD46) and CD59 can be overexpressed in pigs to inhibit activation of the complement cascade and initiation of hyperacute rejection (Cozzi and White, 1995; Mollnes and Fiane, 1999).

Initial attempts involved microinjection of human protective proteins into porcine zygotes. Levels of transgenic expression of human DAF in porcine donors are very variable, but 67% of transgenic pigs express the transgene in skin, liver, spleen, lung and heart (Langford *et al*., 1994). Primate studies have demonstrated survival of porcine hearts transgenic for human DAF for up to 60 days without evidence of rejection. Unfortunately, high levels of immunosuppression are required and have led to adverse side effects in the recipients (Dabkowski *et al*., 1994). Human DAF transgenic pigs were also used for renal transplantation to primates without hyperacute rejection and a maximum survival of 78 days was achieved (Bhatti *et al*., 1998; Zaidi *et al*., 1998).

Mollnes and Fiane (1999) indicated that there may be possible health concern arising from overexpression of membrane complement regulators, due to their affinity for different ligands. For example, DAF is a high-affinity ligand for the seven-span transmembrane molecule, CD97, which is rapidly expressed upon activation of many leukocytes and natural killer cells (Hamann *et al*., 1996). Thus overexpression of DAF might interfere with control of general cell mediated immunity. In addition, MCP is a receptor for measles virus and DAF is a receptor for echovirus and coxackie B picornavirus and overexpression of these proteins could affect susceptibility to virus infection (Weiss, 1998).

1.3.3. Progress in xenotransplantation

Organ xenotransplantation, including heart, liver and kidney from transgenic pigs, has been tested in non-human primates (Byrne *et al*., 1997; Daggett *et al*., 1997; McCurry *et al*., 1995; Schmoeckel *et al*., 1998; Zaidi *et al*., 1998). Also various porcine cells, including foetal brain cells, pancreatic islet cells and liver cells also have been tested for transplantation (Chari *et al*., 1994; Deacon *et al*., 1997; Groth *et al*., 1994; Heneine *et al*., 1998; Pitkin and Mullon, 1999). Table 1.2 shows the recent progress and potential for xenotransplantation using porcine organs and cells.

Most progress has been made in transplanting porcine foetal brain cells into humans to cure Parkinson's and Huntington's disease. The brain is protected by the blood-brain barrier from circulating immune factors and neural cells do not strongly express MHC class I and class II antigens for the xenograft rejection. Recently, Schumacher *et al*. (2000) reported results of transplantation of porcine embryonic ventral mesencephalic tissue into patients with advanced Parkinson's disease, providing improvements in the total Unified Parkinson's Disease Rating Scale scores of 19%.

Table 1.2 Potential xenotransplantation therapy using porcine cells and organs.

Disease	Defect	Therapeutic effect	Stage	Company
Parkinson's	Death of dopaminergic neurons in substantia nigra	Regulated release of dopamine at synapses	Phase I	Diacrin Inc
Huntington's	Death of GABAergic neurons in corpus striatum	Recognition of neuronal pathways and regulated release of GABA at synapses	Phase I	Diacrin Inc
Focal epilepsy	Inappropreate neuronal firing	Inhibition of epileptic focus	Preclinical	Diacrin Inc
Cognitive disorders	Depletion with aging of cholinergic neurons in neocortex and hippocampus	Regulated release of acetylcholine at synapses	Preclinical	Diacrin Inc
Acute liver failure	Hepatocyte death	Liver function	Preclinical	Diacrin Inc
Familial hyperchlesterolemia	Defect of LDL receptor on liver cells	Enhanced metabolism of circulating cholesterol	Preclinical	Diacrin Inc
Cardiac disease	Diseased or damaged myocardium	Provision of functional cardiac contractile tissue	Preclinical	Diacrin Inc
Diabetes	Death of insulin producing cells in pancreatic islets	Provision of insulin in response to changes in blood sugar	??	??
Cardiac disease	Diseased or damaged heart	Replacement of non-functional heart	Preclinical	Norvatis

(Personal Communication: Chris Moran)

1.3.4. Dangers of porcine xenotransplantation

With the advent of xenotransplantation, zoonoses - the inadvertent transmission of pathogens from animal organs or cells into human recipients - have become a concern. Pathogens specific to baboons and pigs, which can infect humans, have been reviewed by Michaels and Simmons (1994). To reduce the potential risks of transmitting animal infections to the recipients, consideration has been given to raising donor animals under specific pathogen free (SPF) conditions. In this process, any known exogenous pathogens can be excluded (Swindle, 1998).

However, endogenous retroviruses, belonging to the category of retroelements (see section 1.1.4 for more details), form part of the genome of the donor. They are hard to detect and difficult or impossible to eliminate from the donor species, even under SPF conditions, but have the potential to cause serious disease in humans (Stoye *et al.*, 1998).

Three types of porcine endogenous retroviruses (PERVs) differentiated by the envelope (*env*) gene, called PERV-A, PERV-B and PERV-C, have been recognised. They are present at approximately 50 copies in the pig genome (Akiyoshi *et al.*, 1998; Le Tissier *et al.*, 1997). Host range and interference studies in the three classes of PERVs show that each of them recognise different receptors and all of them infect at least one human cell line (Takeuchi *et al.*, 1998).

Recent coculture and infectivity experiments have shown that PERVs released from pig kidney cell lines, from mitogenically activated porcine peripheral blood mononuclear cells (PBMCs), or from porcine endothelial cells, can infect human cells and cell lines *in vitro*, raising concerns about the possibility of cross-species infection after xenotransplantation (Martin *et al.*,

1998a; Patience *et al*., 1997a; Wilson *et al*., 1998). However, there has been no evidence of PERV infections *in vivo* in baboons and humans (Martin *et al*., 1998b; Paradis *et al*., 1999; Pitkin and Mullon, 1999; Switzer *et al*., 1999).

1.4. Aims of this thesis

At the time that the research described in this thesis was commenced, there was little success with CATS primers in the pig genome as comparative anchor markers. Therefore the first part of this thesis describes gene mapping use CATS primers. The rest of the thesis describes characterising and mapping PERVs (porcine endogenous retroviruses). As a first step in defining the real PERV hazard for xenotransplantation, this thesis includes the following aspects: i) Characterising PERVs in Westran pigs. ii) Physical mapping including FISH mapping and somatic cell hybrid mapping of PERVs. iii) Characterising ERVs (Endogenous Retroviruses) in the Collared Peccary in order to obtain insight into PERV evolution by looking at endogenous retroviruses from relatives of the pig.

This thesis is organised into seven chapters. General Materials and Methods are described in Chapter Two. The mapping of a subset of CATS markers in the pig genome is described in Chapter Three. Characterization and mapping of Porcine Endogenous Retroviruses (PERVs) is presented in Chapter Four and Chapter Five, respectively. Chapter Six reports the discovery of endogenous retroviruses in the Collared peccary. Finally, a general discussion and conclusions are presented in Chapter Seven.

Chapter Two

General Materials and Methods

2.1. Introduction

This chapter describes the general and common materials and methods used in more than one chapter in this thesis. Common laboratory techniques, materials, and general procedures for preparation of reagents presented here are based on *Molecular Cloning: a Laboratory Manual* (Sambrook *et al.*, 1989), unless otherwise mentioned. All specific materials and methods are fully described in relevant chapters. All glassware and plasticware were sterilised by either autoclaving at 120°C for 20 minutes or baking at 200°C overnight. Chemical reagents used were Analytical Reagent (AR) or Molecular Biology grades unless otherwise specified.

2.2. Preparation of buffers

2.2.1. Commonly used buffers

2.2.1.1. Tris-HCl

To make 1 M Tris-HCl with a required pH (7.6 or 8.0), 121.1 g of Tris base was initially dissolved in 800 ml distilled water and the pH was adjusted to the desired value by adding concentrated HCl at room temperature. The volume of the solution was then adjusted to 1000 ml by adding distilled water and the mixture was then autoclaved at 120°C for 20 minutes.

2.2.1.2. EDTA

The stock solution of 0.5 M EDTA (pH 8.0) was made by adding 186.1 g of disodium ethylenediaminetetraacetate.$2H_2O$ (EDTA) to an initial volume of 800 ml distilled water. About 20 g sodium hydroxide (NaOH) was added to hasten dissolving of EDTA. The solution was stirred on a magnetic stirrer for at least 3 hours. Finally, the pH was adjusted to 8.0 with further NaOH and the volume was made up to 1000 ml. The solution was then autoclaved at 120°C for 20 minutes.

2.2.1.3. TE

TE (10 mM Tris-HCl and 1 mM EDTA) buffer was prepared in aliquots of 500 ml at a time. To make this volume of TE, 1 ml of 0.5 M EDTA (pH 8.0) and 5 ml of 1 M Tris-HCl (pH 7.6 or 8.0) was mixed and the volume was made up with distilled water. The solution was then autoclaved at 120°C for 20 minutes.

2.2.1.4. TAE

A stock solution (50X) of TAE (Tris-acetate EDTA) was prepared by mixing 242 g of Tris base, 57.1 ml of glacial acetic acid and 100 ml of 0.5 M EDTA (pH 8.0) and adjusting the volume to 1000 ml with distilled water. A working solution (1X) was prepared by dispensing 400 ml of the stock solution and bringing the final volume to 20 litres with distilled water.

2.2.1.5. TBE

A 20X TBE (Tris-borate EDTA) stock was made by mixing 216 g of Tris base, 110 g boric acid and 80 ml of 0.5 M EDTA (pH 8.0) and adjusting to final volume of 1000 ml by adding distilled water.

2.2.1.6. SSC

A 20X SSC stock was prepared by mixing 175.3 g of NaCl and 88.2 g of sodium citrate in 800 ml of distilled water. The solution was then adjusted the pH to 7.0 with few drops of NaOH and adjusting to final volume of 1000 ml by adding distilled water.

2.2.2. Buffers used for DNA extraction

2.2.2.1. Cell lysis buffer

To make 100 ml of cell lysis buffer, 1 ml of Triton X-100 (1% final concentration), 100 µl of 1 M Tris-HCl pH 8.0 (1 mM final concentration) and 500 µl of 1 M $MgCl_2$ (5 mM final concentration) were mixed, and distilled water was added to adjust the total volume to 100 ml. Prior to use, 32 ml 1M sucrose were added to 68 ml of the solution (0.32 M final concentration). This buffer was freshly made and used for lysing cells when genomic DNA was extracted from blood samples.

2.2.2.2. Nuclear lysis buffer

A stock of 100 ml nuclear lysis buffer was prepared by mixing 1 ml of 1 M Tris-HCl, pH 8.0 (10 mM final concentration), 1 ml of 1 M KCl (10 mM final concentration), 400 µl of 0.5 M EDTA, pH 8.0 (2 mM final concentration), 500 µl of SDS (0.5% final concentration) and 50 µl of 1 M NaCl (0.5 mM final concentration). The final volume was adjusted to 100 ml with distilled water. This buffer was used for lysing nuclei when genomic DNA was extracted from blood samples.

2.2.2.3. Solution I

A stock of 100 ml solution I was prepared by mixing 2 ml of 0.5 M EDTA pH 8.0 (10 mM final concentration) and 2.5 ml of 1 M Tris-HCl pH 8.0 (25 mM final concentration). The final volume was adjusted to 100 ml distilled water. This buffer was used for extraction of plasmid DNA from bacterial cells.

2.2.2.4. Solution II

This solution was freshly made immediately before use for extraction of plasmid DNA from bacterial cells. 1 ml stock of 10% SDS was mixed with 2 ml of 1 M NaOH. Distilled water was then added to obtain a 10 ml solution II, containing 1% SDS and 0.2 M NaOH.

2.2.2.5. Solution III

This solution was also required for extraction of plasmid DNA from bacterial cells. To prepare a 100 ml of this solution, 11.5 ml of acetic acid (11.5% final concentration) was added to 60 ml stock of 5 M potassium acetate (3 M final concentration) and the final volume was adjusted with distilled water.

2.2.3. Gel loading buffers

2.2.3.1. Agarose gel-loading buffer

To make 10 ml of 10X stock, 0.025 g Bromophenol blue (0.25% final concentration), 0.025g Xylene Cyanole FF (0.25% final concentration) and 5 ml of 30% Ficoll (15% final concentration) were mixed and distilled water was added to adjust the final volume to 10 ml. The

stock was stored at room temperature and used when loading DNA samples in agarose gels to prevent samples from spilling from wells and to provide tracking dyes for electrophoresis.

2.2.3.2. Sequencing gel-loading buffer

Gel-loading buffer for ABI sequencer

A stock was made by mixing 9.8 ml of deionized Formamide (98% final concentration), 200 µl of 0.5 M EDTA pH 8.0 (10mM final concentration) and 0.0025 g of each Bromophenol blue and Xylene Cyanole (0.025% final concentration). This buffer aided loading samples into wells and maintained DNA in a denatured status during electrophoresis.

Gel-loading buffer for Li-Cor sequencer

A stop/loading buffer stock was made by dissolving 40 mg NaOH (10 mM final concentration) and 2 ml 0.5 M EDTA pH 7.6 (10 mM final concentration) into 100ml deionized Formamide. Stock loading dye solution (21 mg/ml Basic Fuchsin, Aldrich) in the ratio 5:100 was added above stop buffer. These stock solutions were stored in the freezer.

2.3. Preparation of medium

2.3.1.1. LB medium

For preparing 1 litre LB (Luria-Broth) medium, 10 g Bacto-Tryptone (1% final concentration) 5 g Bacto-Yeast extract (0.5% final concentration) and 10 g NaCl (1% final concentration) were dissolved in 800 ml distilled water. The pH of the medium was adjusted to 7.0 by adding NaOH and the final volume was increased to 1 litre by adding distilled water. LB medium was autoclaved at 120°C for 20 minutes and stored at 4°C.

2.3.1.2. LB plate with ampicillin or kanamycin

Into 100 ml LB medium, 2 g agar was added (2% final concentration) and the mixture was microwaved until dissolved. The solution was cooled down to 50°C and then 500 µl of 10 mg/ml ampicillin or kanamycin was added (50 µg/ml final concentration) and mixed. Then, 15 ml of this solution was poured in 6 mm Petri dishes. The plates were left at room temperature for 20 minutes to harden the agar, and were then stored in an upside-down position at 4°C until used.

2.4. DNA extraction

2.4.1. Extracting genomic DNA from blood

Porcine whole blood samples in 10ml EDTA vacutainer tubes were removed from the -20°C freezer and thawed at room temperature. 10 ml of the thawed blood was transferred to a 50 ml polypropylene screw-cap centrifuge tube (Beckman) and approximately 5 ml cell lysis buffer (1% Triton X-100, 0.32 M sucrose, 1 mM Tris-HCl pH 8.0 and 5 mM $MgCl_2$) (see section 2.2.2.1. in detail) was added. This solution lyses blood cells and leaves the nuclei intact. A nuclear pellet was obtained after centrifugation at 4,000 rpm (See Appendix 2.3 for the conversion of rpm to *g*) for 10 minutes in a Beckman J2-21 M/E Centrifuge using a JA-20 rotor. The supernatant was decanted into a beaker containing 20 ml of bleach. If the pellet was red, the above lysis step was repeated. After decanting the supernatant, the pellet was resuspended in 5 ml of nuclear lysis solution (10 mM Tris-HCl pH 8.0, 10 mM KCl, 10 mM $MgCl_2$, 0.5 mM NaCl, 0.5% SDS, 2 mM EDTA) (see section 2.2.2.2. in detail) with 40 µl 20 mg/ml Preteinase K. The contents were mixed and the tube incubated at 55°C with constant shaking at 100 rpm overnight.

On the following day, 20 µl of RNase (10 mg/ml) was added into the solution and the tube was incubated a further 1 hour at 55°C. After finishing incubation, 3 ml of phenol and 3 ml of chloroform were added and the solution was mixed by inversion. After the tube was centrifuged at 10,000 rpm for 10 minutes in the Beckman JA-20 rotor with a suitable rubber adapter supplied by Beckman, the upper aqueous layer was carefully removed to a fresh 10 ml yellow cap graduated polypropylene tube. The purified DNA was then precipitated by the addition of 500 µl of 3 M sodium acetate (NaOAc) and 1 volume of isopropanol and the centrifugation was performed at 10,000 rpm for 5 minutes to pellet the DNA. After carefully pouring off the supernatant, 1 ml of 70% ethanol was added to rinse the DNA which was again centrifuged at 10,000 rpm for 5 minutes. The DNA pellet was dried at room temperature until the residual liquid disappeared. The DNA pellet was resuspended in 500 µl of TE buffer (pH 7.5) and 2 µl of RNase (10 mg/ml) and left at room temperature overnight to dissolve prior to storage at 4°C.

2.4.2. Extracting plasmid DNA from bacterial cells

Mini-preparation of Plasmid DNA

The alkaline minilysis method of Birnboim and Doly (Birnboim and Doly, 1979) was adapted for the preparation of plasmid DNA.

Individual positive plasmid colonies were picked with a sterile tip to inoculate a 10 ml tube containing 5 ml LB medium with 50 μg/ml ampicillin or kanamycin depending on the vector used. The tube was then incubated at 37^0C for 10-14 hours. After finishing incubation, 0.5 ml of each culture solution was used to make a glycerol stock for long-term storage of live bacteria. A bacterial cell pellet was obtained after centrifugation at 5,000 rpm for 5 minutes in a Beckman J2-21 M/E centrifuge using a JA-20 rotor with rubber adapter. The supernatant was discarded into a beaker with 20 ml of bleach.

The residual cell pellet was thoroughly resuspended in 100 μl of ice-cold miniprep solution I (25 mM Tris-HCl pH 8.0, 10 mM EDTA pH 8.0) (see section 2.2.2.3. in detail). After vigorous vortexing, the tube was stood at room temperature for 5 minutes. To lyse the cells, 200 μl of freshly prepared miniprep solution II (section 2.2.2.4) was added to the cell suspension. The tube was mixed gently by inversion to ensure that the entire surface of the tube came in contact with solution II, then placed on ice for 5 minutes. To precipitate proteins and cell debris, 150 μl of ice-cold miniprep solution III (60 ml 5 M potassium acetate, 11.5 ml glacial acetic acid, and 28.5 ml distilled water) (see section 2.2.2.5. in detail) was added and then mixed by vortexing for a few seconds. The tube was chilled in ice for 10 minutes to allow most of the protein, high molecular-weight RNA and chromosomal DNA to precipitate.

The tube was centrifuged at 12,500 rpm for 5 minutes at room temperature to yield a clear supernatant. The supernatant, which contained the plasmid DNA, was transferred to a fresh tube, while the pellet which contained precipitated proteins and cell debris was discarded. An equal volume of phenol (400 μl) was added to the supernatant and the tube was vortexed for one minute, to create an aqueous phenol emulsion. The tube was then spun at 12,500 rpm for 5 minutes to separate the phenol and aqueous layers. Immediately after centrifugation, the aqueous layer (top phase) was transferred to another fresh tube, mixed with an equal volume of chloroform (400 μl) and centrifuged at 12,500 rpm for 5 minutes. Again, the aqueous layer was removed after centrifugation and transferred to a fresh Eppendorf tube. To precipitate the plasmid DNA, 0.05 volume of 4 M NaCl followed by 2 volumes of absolute ethanol were added to the tube. The tube was placed at -80^0C for 20 minutes.

After centrifugation for 10 minutes, the supernatant was discarded and the plasmid DNA pellet was collected. The pellet was rinsed gently with 1 ml of 70% ethanol and dried by placing the tube into a heater block at 37^0C for about 10-20 minutes. Finally, the plasmid DNA pellet was dissolved in 30 μl TE buffer (pH 7.5) containing 10 μg/ml DNase-free pancreatic RNase and stored at 4^0C.

Mini-preparation of Plasmid DNA for sequencing

Pure plasmid DNA was needed for sequencing. A JETquick Plasmid Miniprep Spin Kit (GENOMED Inc.) was employed for this purpose.

Bacterial cells were harvested by the previously described method in a 10 ml tube. The cell pellet was thoroughly resuspended in 210 µl of solution G1 (50 mM Tris-HCl pH 8.0, 10 mM EDTA pH 8.0, 100 µg/ml RNase A). After vigorous vortexing, the homogeneous suspension was transferred to a 1.5 ml Eppendorf tube. To lyse the cells, 210 µl of solution G2 (0.2 M NaOH, 1% SDS) was added to the cell suspension. The tube was mixed gently by inversion to ensure that the entire surface of the tube came in contact with solution G2, then incubated at room temperature for 5 minutes. To precipitate proteins and cell debris, 280 µl of solution G3 (contains acetate and guanidine hydrochloride; details not provided by manufacturer) was added and then mixed by inverting the tube several times until a homogenous phase was obtained. After centrifugation for 10 minutes at 12,500 rpm, the supernatant was loaded into a JETquick spin column in a 2 ml receiver tube. Immediately after centrifugation for 1 minute, the flowthrough in the receiver tube was discarded. The spin column was reinserted into the receiver tube and 500 µl of solution GX (contains acetate, guanidine hydrochloride, EDTA, and ethanol; details not provided by manufacturer) added. After incubation for 1 minute at room temperature, the column was centrifuged again at 12,500 rpm for 1 minute and the flowthrough discarded again. The spin column was reinserted into the receiver tube and 700 µl of solution G4 (contains ethanol, NaCl, EDTA, and Tris-HCl; details not provided by manufacturer) added. After centrifuging twice at 12,500 rpm for 1 minute to get rid of all residual solutions, the JETquick spin column was placed into a new 1.5 ml Eppendorf tube. 50 µl of preheated (65-70^{0}C) TE buffer (10 mM Tris-HCl and 1 mM EDTA) was added directly onto the centre of silica matrix of the spin column to elute the DNA from the column. After centrifuging at 12,500 rpm for 2 minutes, the plasmid DNA eluted was stored at 4^{0}C until use.

Large-scale preparation of plasmid DNA

This is a modification of the alkaline minipreparation method. 100 ml of LB medium in a 500 ml conical flask was inoculated with cells from a single colony of bacteria. Cells were grown overnight with shaking at 37^{0}C. The medium was collected in a 250 ml polypropylene screw-cap centrifuge tube (Beckman) and spun in a JA-14 rotor in a Beckman J2-21M/E centrifuge at 5,000 rpm for 5 minutes. The bacterial pellet was resuspended in 2.5 ml of solution I (25 mM Tris-HCl pH 8.0, 10 mM EDTA pH 8.0) (see section 2.2.2.3. in detail) and transferred to a 50 ml polypropylene screw-cap centrifuge tube (Beckman). 5 ml of freshly prepared solution II (0.2 M NaOH, 1% SDS) (see section 2.2.2.4. in detail) was then added and mixed by gentle inversion,

followed by placing the tube on ice for 5 minutes. Then 3.75 ml of solution III (3 M potassium acetate, 11.5 % acetic acid) (see section 2.2.2.5. in detail) was added and mixed by vortexing for a few seconds before the tube was placed in ice for 10 minutes. The mixture was then centrifuged in a JA-20 rotor at 4^0C for 15 minutes at 15,000 rpm and the supernatant was transferred to a new 50 ml centrifuge tube.

To this supernatant, 50 µl of DNase-free pancreatic RNase (10 mg/ml) was added and the tube was incubated at 37^0C for 30 minutes. This was followed by adding 40 µl of Proteinase K (20 mg/ml) and further incubation at 37^0C for 1 hour. After incubation, 1/2 volume of isopropanol (5.6 ml) was added and the contents of the tube were mixed. The tube was then spun at 4^0C for 15 minutes at 10,000 rpm. After centrifugation, the supernatant was poured off and the DNA pellet was resuspended in 300 µl of 0.3 M sodium acetate. The solution was transferred to a fresh Eppendorf tube after dissolution. 3 volumes of absolute ethanol (900 µl) were added and the tube stored at -80^0C for 20 minutes, followed by centrifugation for 5 minutes. The plasmid DNA was then rinsed with 70% ethanol, briefly dried, and dissolved in 500 µl of TE (pH 7.5).

2.4.3. Isolation of DNA from gel slices

2.4.3.1. BRESAclean™ DNA purification kit

The isolation of pure PCR products is essential for the cloning reaction. For this purpose, a BRESAclean™ DNA purification kit (Bresatec Inc.) was used.

DNA fragments were excised from a TAE gel under UV light, weighed, mixed with 3 volumes of BRESA-SALT™ (bottle no. 1), and incubated at 55°C for 5 minutes. The completely dissolved gel slice was mixed with 5 µl plus 1 µl/µg DNA of BRESA-BIND™ (bottle no. 2) and incubated for 5 minutes at room temperature to bind DNA with regular mixing to keep the BRESA-BIND™ in suspension. The BRESA-BIND™/DNA complex pellet was obtained after centrifugation for 5 seconds with microcentrifuge. The supernatant was poured off and the pellet was washed once with BRESA-WASH™ (bottle no. 3) in a volume equivalent to the amount of BRESA-SALT™ (bottle no. 1) used. After 5 seconds centrifugation again all traces of wash solution were removed. Purified DNA was eluted with 20 µl of TE buffer followed by incubation at 55°C for 5 minutes and by centrifugation 1 minute in microcentrifuge. DNA was stored at –20 °C until use.

2.4.3.2. Low melting point agarose gel

The second method for gel purification of DNA used low-melting point agarose (Wieslander, 1979). Restriction digests or PCR reaction products were run at 100 V in 1% low-melting point agarose gel. Electrophoresis was stopped at an appropriate time, and the DNA band of interest was excised from the gel and placed in an Eppendorf tube containing 100 µl of TE (pH 7.5). The gel was melted using a heater block at 65^0C for 15 minutes, and the insert DNA fragment was then purified by phenol and chloroform extractions. Finally, after ethanol precipitation, the DNA fragment was dissolved in TE (pH 7.5) and stored at 4^0C.

2.5. Measuring the concentration of DNA

2.5.1. Spectrophotometric measurement

The DNA samples to be quantified were diluted in individual 1.5 ml Eppendorf tubes containing 2 µl of extracted genomic or plasmid DNA in 198 µl of TE (pH 7.5). The concentration of DNA was measured by spectrophotometry (Pharmacia Biotech Gene Quant II RNA/DNA Calculator) at wavelength of 260 nm (OD_{260}) and 280 nm (OD_{280}). The OD_{260} was used for determining genomic DNA concentration using the relationship that OD_{260} of 1 corresponds to about 50 µg/ml of double-stranded DNA. The ratio between OD_{260} and OD_{280} indicates the extent to which there has been contamination of the extracted DNA with protein. When the ratio is less than 1.8, it indicates that the sample is contaminated with protein (Sambrook *et al*., 1989).

2.5.2. Agarose gel quantification

To check the concentration of DNA or to determine whether the DNA was sheared or otherwise degraded, 1 µl of extracted DNA was mixed with agarose loading dye and run on a 0.8% agarose gel at 100 voltage for 40 minutes. A range of known concentrations of DNA (λ/*Hind*III 1 µg/ml) was used to estimate the concentration of extracted DNA samples.

2.6. Designing Primers

Primers were designed from the relevant GenBank database sequences using the computer program "Primer Design – Version 2.0" (Scientific and Educational Software, 1991). The criteria specified in the selection of a primer were a 45-55% GC content, no hairpins, no dimerisation within or between primer pairs, and a primer length of 20 nucleotides. Primers for the

comparative mapping research (Comparative Anchor Tagged Sequence (CATS) primers) were kindly supplied by Dr. Leslie Lyons. All other primers were synthesised commercially unless specified. Primers were diluted to 10 pmole/µl with sterile MilliQ water and stored at -20^{0}C. The Primer Design program estimates the melting temperatures for each primer and normally the annealing temperature for PCR reaction was set at 5^{0}C lower than melting temperature.

2.7. PCR

2.7.1. PCR conditions

PCR conditions were optimised by testing four levels of $MgCl_2$ concentration (0.5 mM, 1 mM, 2 mM, 3 mM) and a range of annealing temperatures using identical control genomic DNA before moving to the main experiments. The optimum concentration of $MgCl_2$ and annealing temperature for each primer pair is presented in relevant chapters.

2.7.2. PCR amplification

PCR reactions were performed in a 25 µl reaction volume and the composition is outlined below.

Constituent	Final
Genomic DNA	~100 ng
Forward primer	10 pmol
Reverse primer	10 pmol
$MgCl_2$	0.5 mM – 3 mM
Each dNTP (dATP, dGTP, dCTP, dTTP)	100 µM
10 x *Taq* polymerase buffer	2.5 µl
Taq DNA polymerase	1 unit

Distilled water	up to 25 µl

The mixture was overlaid with one drop of Paraffin oil and PCR was carried out in a PTC-100 Programmable Thermal Controller (MJ Research, Inc.) with the following thermocycling programme (Table 2.1). After the thermocycle reaction, the PCR products were directly analyzed or stored at -20^{0}C until used later.

Table 2.1 Thermocycling profile used for PCR amplification.

Purpose	Temperature	Duration	No. of cycles
Initial denaturation	95 ^{0}C	5 minutes	1
Denaturation	95 ^{0}C	30 seconds	30
Annealing	Variable (50-65 ^{0}C)	30 seconds	
Extension	72 ^{0}C	30 seconds	
Final extension	72 ^{0}C	20 minutes	1

2.7.3. Touchdown PCR

Touchdown PCR is based on progressively lowering the annealing temperatures during PCR (Don *et al*., 1991). The high annealing temperature at the beginning of the PCR reaction ensures that only the correct products are amplified in the beginning, even though less efficiently than in the conventional PCR cycles. Later the decreased annealing temperature provides more efficient amplification when the correct products predominant as template.

2.7.4. Screening of clones using boiled preparation

After a single colony was cultured in 5 ml LB medium overnight, a bacterial cell pellet was obtained by centrifugation at 5,000 rpm for 5 minutes in a Beckman J2-21 M/E centrifuge using a JA-20 rotor with rubber adapter. The supernatant was discarded. 5 µl of the residual cell pellet was mixed thoroughly with 5 µl sterile water and heated at 95 °C for 5 minutes using a PTC-100 Programmable Thermal Controller (MJ Research Inc.) to release DNA from the cell. PCR amplification was undertaken on 1 µl of the boiled extracts.

2.8. Restriction enzyme digestion

All the restriction enzymes and the restriction enzyme buffers were purchased from Promega, and digestions were carried out using the manufacture's recommendations. The following cocktail components were used to digest PCR products or plasmid DNA.

PCR products or plasmid DNA	= 0.5-2µg
10X restriction enzyme buffer	= 3 µl
Restriction enzyme	= 5 units
Distilled water up to	30 µl

After the components above were mixed, the tubes were normally incubated overnight in a hybridisation oven (Hybaid Inc) at an appropriate temperature specified in the Promega catalog. The reaction was stopped by adding 10X gel loading buffer before electophoresis.

2.9. Gels and gel electrophoresis

2.9.1. Agarose gels

Agarose gel electrophoresis was used to separate DNA fragments and was performed in a horizontal gel electrophoresis tank (OWL Scientific Plastics Inc.). The concentration of the gel used was dependent on the size of DNA fragments to be resolved. To make a 1% agarose gel, 1 g of agarose powder (Progen Industries Ltd.) was mixed with 100 ml of 1X TAE. The solution was microwaved with shaking at one minute intervals until the agarose was completely dissolved. After cooling the solution to about 60°C, 2 µl of 10 mg/ml ethidium bromide was added before pouring in a 20 cm horizontal gel electrophoresis tank (Owl Scientific Inc.). A well-forming comb was put in place and the solution was allowed to set for about one hour. Samples of PCR products were loaded with loading buffer along with molecular size markers to estimate the size of the product. Electrophoresis of the gel was performed in 1X TAE at 100 to 120 volts for 20 to 45 minutes. Gels were then removed and the DNA bands detected by ethidium bromide fluorescence using an ultraviolet light transilluminator (UVP Inc.). Gels were photographed using a polaroid camera (Polaroid MP4 Land Camera), using a red filter and 667 Kodak black and white instant film.

2.9.2. Sequencing gels

2.9.2.1. ABI 373 sequencing gels

6% Denaturing polyacrylamide sequencing gels were made to separate DNA fragments differing by as little as one nucleotide in length. A 6% polyacrylamide gel was prepared by dissolving 40 gm of urea (ICN Biomedical Inc.) in 28ml of distilled water with 12 ml of 40% acrylamide/bis solution (19:1, Gradipore Ltd) and 8ml of 10 x TBE. The gel solution was mixed in a beaker and stirred until completely dissolved. Meanwhile, the glass plates from a vertical 25 cm x 42 cm sequencing gel apparatus (Applied Biosystems, Perkin-Elmer) were thoroughly cleaned using glass plate washing solution and distilled water. Then, the gel solution was mixed with 400 µl of 10% APS (ammonium persulfate) and 40 µl of TEMED (N, N, N', N', - tetra methyl ethylene diamine) and immediately the solution was poured between the two glass plates

separated by 0.4 mm spacer. Finally a well-forming comb (36-well sharks-tooth comb) was inserted immediately and the gel left at the room temperature to set for more than 2 hours. After pre-electrophoresis at 2000 volts, 40 watts for 20 minutes in 1X TBE, sequencing samples were loaded and the gel run at 2000 volts for 14 hours.

2.9.2.2. Li-cor sequencing gels

For the purpose of long readable single pass-sequencing, a Li-Cor sequencer (Model 4200, Li-Cor Inc.) was used. There are three sizes of gels (25 cm, 41 cm, 66 cm). A 66 cm gel, which can be read for up to 1,000 bp, was used for most of the experiments in this thesis.

A 4% polyacrylamide gel was prepared by mixing 18.9 gm of urea, 3.6 ml 50% Long Ranger acrylamide solution (BioProducts Com.) and 10.8 ml of 5 x TBE in a beaker and stirred until completely dissolved. Distilled water was added to bring the total volume to 45 ml or 50.625 gm. While the gel solution was mixed, the 66 cm glass plates were thoroughly cleaned using distilled water and isopropanol. When the gel solution was dissolved completely, 200 µl of 10% APS and 20 µl of TEMED were mixed and immediately the solution was poured between the two glass plates separated by a 0.25 mm spacer. Immediately, a well-forming comb (48-well rectangular comb) was inserted. The gel was left at room temperature to set for more than 2 hours.

2.10. Sequencing – ABI/Li-Cor

Cycle sequencing was performed with a dye-labelled (for ABI) or unlabelled (for Li-Cor) terminator, using a method modified from Sanger's dideoxy sequencing method (Sanger *et al*., 1977).

2.10.1. ABI Dye Terminator Cycle Sequencing

Fuorescence-labelled cycle sequencing reactions were performed in 20 µl volumes using an ABI sequencing kit (Perkin Elmer). Each reaction contained 100 ng of plasmid or PCR products template DNA, 8 µl of Terminator Ready Reaction Mix and 3.2 pmole primer. The mixture was overlaid with one drop of paraffin oil, and the reaction was performed in a PTC-100 Programmable Thermal Controller (MJ Research, Inc.). Reaction profiles included 25 cycles of 30 seconds denaturation at 96^0C, 15 seconds annealing at 50^0C, and 4 minutes of extension at 60^0C. After PCR, all 20 µl of the reaction contents were transferred to a 1.5 ml Eppendorf tube which contained 2 µl of 3 M Sodium acetate (pH 4.6) and 50 µl of 95% ethanol. The tube was

vortexed and kept on ice for 10 minutes. Then the tube was spun at 12,500 rpm for 15 minutes to remove the unincorporated terminators and to pellet the DNA. After rinsing with 70% ethanol, the pelleted DNA was dried. The dried DNA sample was resuspended by adding 5 µl of auto-sequencing gel loading buffer (Section 2.2.3.2) and completely dissolved by vortexing. Then, the DNA sample was denatured at 95^0C for 5 minutes, and placed on ice until all of the sample solution was loaded in a sequencing gel. After running the gel, the sequence data were analysed with DNA sequencing analysis software (version 3.3, Perkin Elmer).

2.10.2. Li-Cor cycle sequencing

Li-Cor cycle sequencing reactions were carried out using a SequiTherm EXCEL™ II DNA Sequencing Kits-LC (Epicentre Technologies, Inc.). Compared to the ABI system, the Li-Cor system has only two infra-red dyes available (Wavelength 700 and 800). The cycle sequencing reaction products must be resolved in four lanes.

The 17 µl volume of premix was made in a 0.5 ml microcentrifuge tube by combining 7.2 µl of 3.5X SequiTherm EXCEL II sequencing buffer, 2 µl (1pmole/µl) of IRD-labelled primer, 100 to 250 fmole of DNA template, 1 µl of SequiTherm EXCEL II DNA Polymerase (5 U/µl), and deionized water to 17 µl. This premix was placed on ice until use. For each template, four 0.2 ml PCR tubes were labeled G, A, T, C and placed on ice. Then 2 µl of each SequiTherm EXCEL II-LC termination mix, G, A, T or C, were added to the G, A, C, T tubes, respectively. 4 µl of premix was added to each of the four tubes of termination mix and thoroughly mixed. The mixture was overlaid with one drop of paraffin oil. After briefly centrifuging, the PCR reaction was performed in a PTC-100 Programmable Thermal Controller (MJ Research, Inc.). Reaction profiles included a 5 minute denaturation step at 95^0C followed by 30 cycles, each consisting of 30 seconds denaturation at 95^0C, 15 seconds annealing at 50^0C, and 1 minute of extension at 70^0C. After the PCR, 3 µl of stop/loading buffer was added to each reaction which was stored at –20 ^{0}C until use. After running the Li-Cor sequencing gel, the sequence data were analysed with image analysis (version 4.0, Li-Cor Inc.) and manipulated with image manipulation (version 4.0, Li-Cor Inc).

2.11. Cloning PCR products

Two vector systems were employed: pCR® 2.1 – TOPO vector (Invitrogen, TOPO™ TA Cloning Kit) (see Appendix 2.1.) for products amplified with *Taq* DNA polymerase, and pCR® -

Blunt vector (Invitrogen, Zero Blunt™ PCR Cloning Kit) (see Appendix 2.2.) for products amplified with *Pfu* proof reading DNA polymerase.

2.11.1. TOPO™ TA Cloning Kit

The plasmid vector, pCR® 2.1 – TOPO, is supplied linearised with single 3' thymidine (T) overhangs for cloning PCR products which generally have 3' adenine (A) overhangs. Topoisomerase is covalently bound to the vector to increase the efficiency of ligation at room temperature.

2.11.1.1. Ligation of plasmid and insert DNA

To obtain the optimal ratio of plasmid-to-insert DNA, a 1:1 molar ratio of vector:insert was tried. In general, 0.5 to 2 µl of a typical PCR sample (10 ng/µl) with an average insert length of 400 to 1,000 bp will give the proper insert:vector ratio. In a 5 µl reaction volume, the composition contained:

Fresh PCR product	0.5 to 2 µl
pCR® 2.1 – TOPO vector	1 µl

Deionized water to final volume 5 µl

The ligation reaction was performed for 5 minutes at room temperature (~25°C). The reaction to be was briefly centrifuged and then placed on ice prior to the One Shot™ transformation reaction.

2.11.1.2. Transformation reaction

50 µl of TOP10 One Shot™ competent cells were thawed on ice. 2 µl of 0.5 M β-mercaptoethanol were added and mixed by stirring gently with a pippette tip. 2 µl of the TOPO cloning solution were then added. The tube was incubated on ice for 30 minutes. The cells were heat shocked for 30 seconds at 42°C. Then the tube was transferred into ice and incubated for another 2 minutes. 250 µl of SOC medium (2% Tryptone, 0.5% Yeast Extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM Glucose) was added and shaken gently at 37°C for 30 minutes (for ampicillin selection) or 1 hour (for kanamycin selection) to allow the cells

recover. The cells then be plated directly onto a 1.5% agar LB medium plate containing 50 μg/ml ampicillin or kanamycin and incubated at 37°C overnight.

2.11.2. Zero Blunt™ PCR Cloning Kit

2.11.2.1. Ligation of plasmid and insert DNA

In a 10 μl reaction volume, the reaction mix contained:

Linearized, blunt pCR-Blunt (25ng)	1 μl
Blunt-end PCR product	1 to 5 μl
10X Ligation Buffer	1 μl
T4 DNA Ligase (4U/μl)	1 μl
---	---
Deionized water to final volume	10 μl

The ligation reaction was performed for 1 hour at 16°C. The rest of the procedures are the same as above for the TOPO™ TA Cloning Kit.

2.11.2.2. Transformation reaction

Most of the procedures are the same as for the TOPO™ TA Cloning Kit except the recommended duration of the heat shock is 45 seconds rather than 30 seconds and β-mercaptoethanol is not needed for the blunt-end cloning reaction.

2.12. Mapping resources

To physically map PCR products, a French somatic cell hybrid panel was employed. It consists of 27 somatic cell hybrid lines including 19 pig × Chinese hamster somatic cell hybrids (numbers 1-19) and 8 pig × mouse somatic cell hybrids (numbers 20-27) (Yerle *et al*., 1996). DNA from these was kindly provided by Dr Martine Yerle of INRA, France. Further details are provided Chapter 1.

Chapter Three

Improving the comparative porcine map relative to human chromosomes 9, 10, 20 and 22

3.1. Introduction

Since Haldane's review of coat-colour-determining genes in several species (Haldane, 1927), it has been postulated that genomic organisation is conserved among mammals even though mammals arose and diverged over the past 70 million years (Andersson *et al*., 1996). Mammals are estimated to have 70,000 to 100,000 genes (Bird, 1995) and presumably share most of these genes with even distantly related vertebrates like birds and even fish. Genome mapping allows development of comparative gene maps, showing conserved synteny between species, from which conserved and ultimately ancestral chromosome organisation can be inferred. For example, chicken chromosome 1 exhibits conserved synteny with human chromosome 1 (HSA 1) (Smith *et al*., 1997). The loci on zebrafish (*Danio rerio*) linkage group 9 imply that this region is homologous to the long arm of HSA 2 (Postlethwait *et al*., 1998). On the assumption that relatively large chromosomal segments are conserved among species of mammals, it is possible to deduce the position of a gene in one species by knowledge of its position in another (Eppig and Nadeau, 1995; Gillois, 1991). Because of this, comparative genome mapping has emerged as an important research area (Andersson *et al*., 1996; Wakefield and Graves, 1996). The recognition of conservation of genomes across species has been useful in identifying and mapping new genes, in searching for candidate genes for disease phenotypes, and in analyzing genome organization and evolution (Eppig, 1996; Eppig and Nadeau, 1995; Nadeau and Sankoff, 1997; Nicholas and Harper, 1996).

Characterisation of conserved chromosomal segments among mammalian species will enable animal geneticists to exploit the rapid advances in human and mouse molecular genomics, including the soon to be completed total sequencing of the human genome (Dunham *et al*., 1999; Hattori *et al*., 2000). For instance, using already identified mouse and human gene(s) as references and a comparative map, one can predict the location(s) of the nearby gene(s) in animal genomes and *vice versa*. For this reason, the highly developed genetic maps of humans and mouse are potentially very

valuable resources for genome mappers working in other mammalian species. Conversely the results of QTL studies being performed in domestic and model species can be mined for their biomedical significance. Thus the effort of producing detailed comparative maps between species is justified. Numerous publications have examined comparative mapping (between human and mouse: Cole *et al*., 1998; Thomas *et al*., 1999, between human and pig: Bruch *et al*., 1996; Hu *et al*., 1997; Lahbib-Mansais *et al*., 1996; Lahbib-Mansais *et al*., 1999; Van Poucke *et al*., 1999, between human and cattle: Aleyasin and Barendse, 1999; Gao and Womack, 1997; Lanneluc *et al*., 1996; Schlapfer *et al*., 1998; Yang *et al*., 1998, between human and cat: Murphy *et al*., 1999; O'Brien *et al*., 1997, between human and sheep: Broad *et al*., 1996; Lanneluc *et al*., 1996, between human and horse: Caetano *et al*., 1999, between human and chicken: Smith *et al*., 1997, and between human and dog: Lyons *et al*., 1999).

Heterologous chromosomal painting (or Zoo-FISH), i.e. hybridisation of chromosome-specific DNA libraries from one species onto metaphase spreads of a second species, has broadly established the relationship between all human (*Homo sapiens*, HSA) and all pig (*Sus scrofa*, SSC) chromosomes. For example, HSA 20 is homologous with SSC 17 and HSA 9 with SSC 1. HSA 22 corresponds to two different porcine chromosome segments, on SSC 5p and SSC 14q, and HSA 10 corresponds to SSC 10 and 14 (Fröncke *et al*., 1996; Goureau *et al*., 1996; Johansson *et al*., 1995; Rettenberger *et al*., 1995b). The Zoo-FISH technique, using human chromosome-specific painting probes, has also been applied to other important animal species, such as cattle (Chowdhary *et al*., 1996; Hayes, 1995; Solinas-Toldo *et al*., 1995), muntjac (Yang *et al*., 1997), horse (Raudsepp *et al*., 1996), cat (Rettenberger *et al*., 1995a), mink (Hameister *et al*., 1997) and mice (Scherthan *et al*., 1994). Chromosome specific-painting probes are also available for animals like mice (Rabbitts *et al*., 1995), rat (Hoebee *et al*., 1994), pig (Schmitz *et al*., 1992) and cattle (Schmitz *et al*., 1995). Bi-directional chromosomal painting can more accurately determine the chromosomal rearrangement and syntenic relationships between species. The current comparative status between human and porcine chromosomes, based mainly on the work of Frönicke *et al.* (1996), is shown in Table 3.1.

At the Human Genome Organisation (HUGO) sponsored Comparative Genome Organisation Workshop held on Fraser Island, Australia, in 1995 (Andersson *et al*., 1996), an international collaborative comparative mapping programme for pigs was established between six groups (Leif Andersson: Sweden, Alan Archibald: UK, Chris Moran: Australia, Larry Schook: USA, Chris Tuggle:USA, Martin Yerle: France) to share the mapping of the newly available Comparative Anchor Tagged Sequence (CATS) markers. Although the original marker CATS was not published until 1997 (Lyons *et al*., 1997), Drs Lyons and O'Brien had suggested at this meeting that conserved

sequences in expressed genes (Type I markers) can be used for making consensus PCR primers for use in systematically constructing comparative maps for all mammals, and they produced a set of CATS primers for this purpose. The original intention was to linkage-map the markers, but the French group subsequently agreed to supply a somatic cell hybrid panel (Yerle *et al*., 1996) for physical mapping as well.

Table 3.1 Comparative status of human and porcine chromosomes as identified by gene mapping (both physical and linkage mapping) and Zoo-FISH (Modified from Frönicke *et al*., 1996).

Human chromosome number	Expected homologies on pig chromosomes based on gene mapping	Homologous porcine segments Observed through Zoo-FISH[1]
1	4,6,9,14	4q1.6-qter(S);6q2.3-q2.8(S); 6q3.2-qter(S);9q2.4-q2.6;10p(S)
2	3,12,15	3cen-qter(S);15q1.1-q1.4(S);15q2.1-qter(S)
3	13	13cen-q4.6(S)
4	8	8(S)
5	2,14,16	2q2.1(middle)-qter(S);16(S)
6	1,7	1p(S);7pter-q1.3(MS)
7	9,12,18	9cen-q2.3(S);1.8(S)
8	14	4qter-q1.5(S);14q1.2-q1.3(M);15q1.5(M)
9	1	1q2.4-qter(S);10cent-q1.1(MS)
10	14	10q1.2-qter(S);14q2.3-qter(S)
11	2,9	2pter-p1.1(WM);9pter-p1.1(M)
12	5,14	5p1.4-qter(S);14q1.4(MS)
13	11	11
14	7	1q2.2-q2.3(S);7q1.5-q2.2(S);7q2.4-qter(S)
15	7	1q1.4-q1.8(S);7q1.4(M);7q2.3-q2.4(MS)
16	3	3p(M);6p(S)
17	12,14	12(S)
18	-	1q1.1-q1.3(S);1q1.8(half)-q2.1(S); 6q2.8-q3.1(MS)
19	2,6	2cent-q2.1(MS);6cent-q2.1(MS)
20	-	17(S)
21	7,9,13	13q4.7-qter(S)
22	5,14	5pter-p1.4(S);14q1.6-q2.2(MS)
X	X	X(S)

[1]The intensity of signal on each of the porcine segments painted is presented in brackets as follows: (S) = strong; (M) = medium; (W) =weak. In a few cases, intermediate signal intensities (MS =medium strong, WM =weak medium) were also observed. (?) questions validity of this result because hybridisation was on the peri-centromeric region.

The international CATS project was headed by Drs. O'Brien and Lyons and their colleagues at Frederick, Maryland, USA. They designed and distributed 313 CATS primers and relevant information with the objective of enabling the mapping of CATS markers in any species for which the consensus primers yielded the appropriate PCR product (Lyons *et al*., 1997). The CATS primers were distributed among the six participating groups according to human chromosome numbers. The Australian group was allotted 53 pairs of primers, initially 36 from human chromosome 9 and 10 (Zhang *et al*., 1996) and later 17 primer pairs coming from markers mapping to human chromosome 20 and 22 (Lee *et al*., 1998).

Recently, these anchor markers have been ordered in the human linkage map to provide a framework for comparative gene mapping of representative mammalian species (Chen *et al*., 1999). Other attempts to make cross-species primers have been reported by Venta *et al.* (1996) (Universal Mammalian Sequences-Tagged Sites, UM-STSs) and Jiang *et al.* (1998) (Traced Orthologous Amplified Tags, TOASTs).

Dr Weiyi Zhang initially screened 36 primer pairs coming from human chromosomes 9 and 10 in his PhD study, with six loci being physically mapped (Zhang, 1997). In the study reported here, 17 new pairs of CATS primers from human chromosomes 20 and 22 have been tested and the products of 30 primer pairs, which had not been mapped in Zhang's study (1997), were re-evaluated.

3.2. General strategy

On the assumption that most functional genes (type I markers) are present in all mammalian species, the comparison of the gene position and order in one species will identify homologous regions and possibly conserved synteny in another species. If the gene homologies are correctly identified, then similar gene order identifies conserved synteny. For example, most of the q arm of human chromosome 4 is homologous with most of porcine chromosome 8, although the linear order of the markers is not identical in the two species (Ellegren *et al*., 1993; Johansson *et al*., 1995). By mapping a strategic subset of type I markers in a "map-poor" species like a pig, it becomes possible to mine the genetic maps from "map-rich" species such as human and mouse. These two species maps have extensive information on the localisation of coding sequences, especially arising from the human genome project. Well defined comparative maps can be used not only to predict gene locations in "map-poor" species, but also identify candidate disease genes, characterize the genetic basis for complex traits like QTL, predict genome organisation and reveal chromosome evolution (Eppig and Nadeau, 1995; Nadeau and Sankoff, 1997; O'Brien *et al*., 1988).

The 313 CATS primers are mostly based on the consensus of coding sequences from human and mouse and sometimes other species of mammals. The intention of Lyons *et al.* (1997) was that they would flank introns, but this is frequently not the case (Zhang, 1997). In theory, the PCR products are designed to contain conserved sequences adjacent to the primers for gene identification. The less-conserved intronic sequences would be a source of polymorphisms. Preliminary analysis of the CATS primer has been reported on the domestic cat (Lyons *et al.*, 1997) and dog (Lyons *et al.*, 1999).

The task assigned to this laboratory by the "pig CATS consortium" was to physically and linkage map 53 CATS loci in the pig genome. The basic strategy used is diagrammatically presented in Fig 3.1. The specific objectives are, i) to optimise PCR conditions for the primers, ii) to sequence the PCR products and to verify their identity with the target loci by comparison with the relevant entries in the GenBank database; iii) to physically map the loci using a porcine somatic cell hybrid panel and iv) to genetically map the loci using the PiGMaP (Pig Gene Mapping Project) reference families.

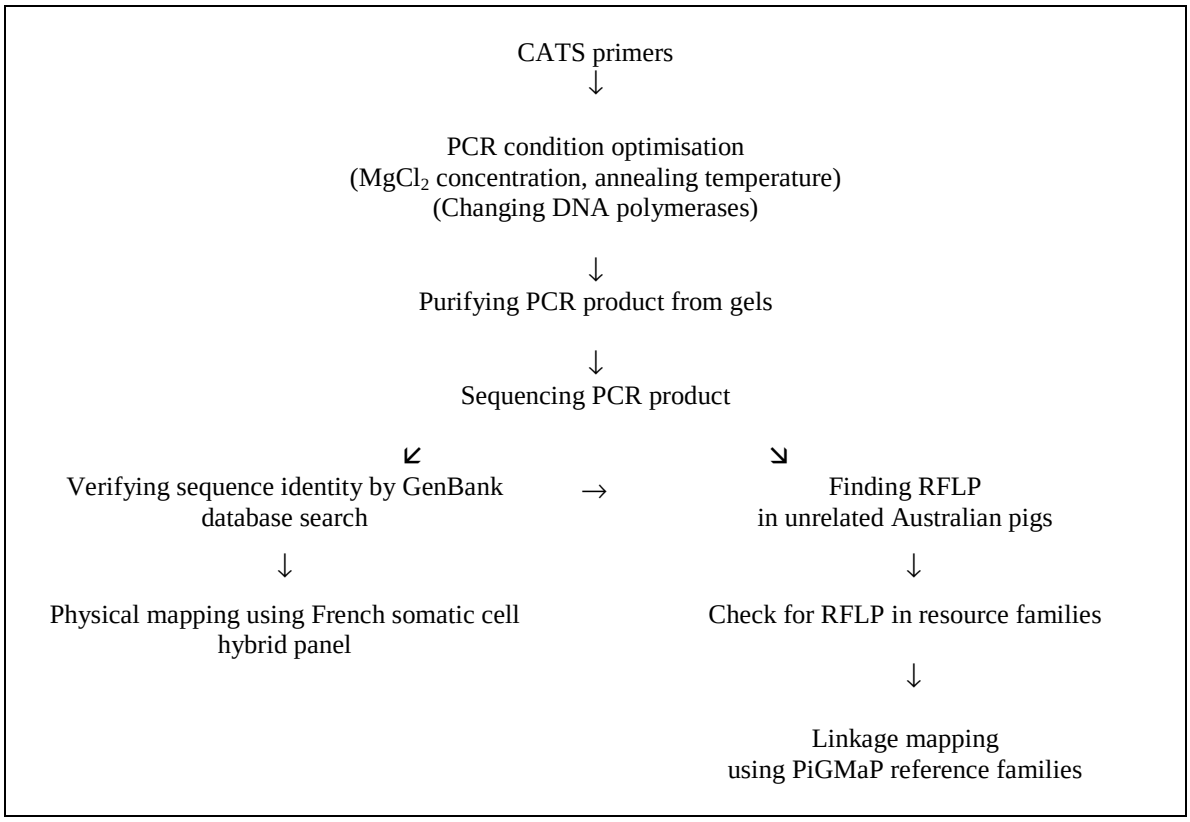


Fig 3.1 Strategy used to map CATS loci onto pig chromosomes.

3.3. Materials

3.3.1. CATS primers

The CATS primers were chosen from loci spaced at 5 to 10 centiMorgan intervals throughout the mouse and human genomes (O'Brien *et al*., 1993).

Of 53 CATs loci allotted to the University of Sydney group, 47 CATS primer pairs, from human chromosome 9, 10, 20 and 22, were used in this study for comparative mapping between human and pigs, excluding 6 loci previously mapped by Zhang (1997). The CATS primer sequences, human location, and species from which the primer consensus sequence was derived are listed in Table 3.2.

Table 3.2 A list of 53 CATS loci, their human location, primer sequences and aligned species (Lyons *et al*., 1997).

Locus	Human Location	Primer sequences(5'→3') (Forward and Reverse)	Species contributing to consensus[1] Forward/Reverse
ABL1 (abelson murine lukemia viral oncogene)	9q34.1	CTGAATGAAGATGGTGGGC TAAGACCCGGAGCTTTTCAC	HM/HM
ALDH1 (aldehyde dehydrogenase-1)	9q21	CGAGGTCTTCTGCAACCAG TGTCCAAATCCACCAGGTAG	HMR/HMR
ALDOB (aldolase B)	9q21.3-q22.2	AACACTGAAGAGAACCGCC GCCACTTCCCAAAGTCAAC	HKS/HKS
AMBP (alpha-1-microglobulin/bikunin)*	9q32-34	AGTGTCTGCAGACCTGCCG AGTAGAACTTGTTGCCGTTGCC	3HJP/3HJP
ANX1 (Annexin I)	9q11-22	CATCACCTCAGACACATCTGG CACATCTGTCCCCTTTCTCC	CHKNR/ CHKNR
ASS (argininosuccinate synthetase)	9q34.1	TCATAGCCTACCTGGCCAAC GCCGTGAGACACATACTTGG	CHKM/CHKM
C5 (complement Component 5)	9q32-34	ATGGGAAATTCAAGGCATTG GGAAGCATTGTGAATGTCACC	HM/HM
CNTFR (ciliary neuro-trophic factor receptor)	9p13	CCAAGGACAATGAGATTGGG AGATCTTCGTGGTAGGTGGG	HR/HR
DBH (dopamine beta-hydroxylase)	9q34	CCCCAGCAGGACTACCAG AACGGCTCCTCCAGGATC	HR/HR
GALT (galactose-1-phosphate uridyltransferase)	9p13	TCTAGCCACTGCACTCCATC CCCATGGAGTAGGGAAAGG	2HM/HM
GGTB2 (glycoprotein 4-beta-galactosyl-transferase-2)	9p13	CACCCTCGTCTATTACCTGG TGAGCAGCGGGGACT	CHM/CHM
GRP78 (glucose related protein 78 kd)*	9q33-34.1	CAACGATCAGGGCAACC TCATTTTAGTGAGAACCATGGC	CHJKM/ CHJKM

GSN (gelsolin)	9q32-q34	CACCCCGAATTCCTAAAGG CCAGCCATTAGTGGAGGTC	HM/HM
HXB (hexabrachion)	9q32-q34	ACTGGCCTTGCTCCTGG TCAGGTTCCCGATGGAGTAC	HMP/HMP
IFN1@ (interferon Alpha) (gene family)	9p22	TTCTCCTGCCTGAAGGACAG GGATCTCATGATTTCTGCTCTGAC	HQ/HQ
IFNB1 (interferon Beta-1)*	9p22	AGAACTGAAAGTGGGAAATTCCTC GTCTCATTCCAGCCAGTGCTAG	CHMQ/CHQ
RLN1 (relaxin)	9pter-q12	CTCCTGGGGAAGAACTGCTC TTCAGCTCCTGTGGCAAATTAG	HP/HP
SPTAN1(spectrin nonerythroid alpha subunit)*	9q33-34	TGCACAGAGTTAAACCAGGC GCTGCTGTCCAAACTGCTC	HKM/KM
TXN (thioredoxin)*	9q31	TGTGAAGTCAAATGCATGCC ATGGTGGCTTCAAGTTTTTCC	HMS/HS
ADRA2A(alpha-2A-adrenergic receptor)	10q23-q25	GCACCTGTGCGCCATC CTTCTCGATGGAGATGAGCG	CHJKMPO/ CHJKMPO
ADRB1 (beta-1-adrenergic receptor)	10q24-26	CCTCTTCATCATGTCCCTGG TGACACACAGGGTCTCGATG	2HM/2HM
ALOX5(arachidonate 5-lipoxygenase)	10q11.2	GGGGACTACATCGAGTTCCC GGGTTCCACTCCATCCATC	HM/HM
APT1 (apoptosis Antigen 1)	10q23	TCTGGACCCTCCTACCTCTG TTGTCTGTGTACTCCTTCCCTTC	HM/HM
BMI1 (oncogene BMI1)	10p13	CAGCTGATGCTGCCAATG CATCACAGTCATTGCTGCTG	HM/HM
CHAT (choline acetyltransferase)	10q11.2	CCATTGTGCAGCAGTTTGG TGGAGTTGACAGGCAGGG	HP/HP
CREM (cyclic-AMP-response- Element modulator)	10p11.2	AGACTAGCACGGGGCAATAC CAAAGCAGTAGTAGGAGCTCGG	HMR/HMR
CYP2E (cytochrome p-450, family II, subfamily E)	10q24.3-qter	GAAGTATCTGAGGCTGATGAGTTTG TCCAGTGACTGATGGTGCTC	HNR/HNR
DNTT (terminal deoxynucleotidyl Transferase)	10q23-q24	TGGAGAAGAAAATGGGAACC CGATCAGCCAGGAGACATC	CH/CH
EGR2 (early growth response-2)*	10q21.1-q22.1	CCCAAAAGACCAAGCAAGAC GCAGATCCGACACTGGAAG	HR/HR
HK1 (hexokinase-1)	10q22	CGGGTCTTCCTTTCGAATTC ATGTTGGCATCATAGTCCCC	CHR/CHR
IL2R (interleukin-2 receptor)	10p14-p15	CACGCCACATTCAAAGCC GATGAGTGGCTTGAGTTTCCTG	HFM/HF
OAT (ornithine aminotransferase)	10q26	TGAGAAAGGAGCTCATGAAGC GGCCAGAAGCCCATTATCTC	HMR/HMR
PLAU (plasminogen activator urinary)	10q24	CCAACAAGTACTTCTCCAACATTC AGTTCCAGGCCAGGCAG	2CHP/2CHMP
PP (pyrophosphatase inorganic)	10q11.1-q24	ACAAGGCGTTCACCAAGG AACGAACCTCTTGCACATTTG	CHNR/CHNPR
PRF1 (perforin 1)	10q22	AGACCCACCAGGACCAGTAC GCTTCTTCTTCTTCTCCTCACAG	HM/HM
VIM (vimentin)	10p13	ATTGAGATTGCCACCTACAGG TGAGTGGGTGTCAACCAGAG	CHJMR/CHJMR

PRNP (Prion Protein)	20pter-p12	TGGTGGCTGGGACAG GGCACTTCCCAGCATGTAG	24CGJLMPRS / 24CGJLMPRS
OXT (Oxitocin, prepro-(neurophysin –I))	20p13	CGCCTGCTACATCCAGAAC CGACGGCAGGTAGTTCTCC	CHMRS / 2HMPRS
SRC (V-Src avian sarcoma viral oncogene homolog)	20q11.2	CCTTCCTCCTCGTTCGTGAGAG TGATGTAATAGCCACCCGTG	FH / FH
TOP1 (Topoisomerase (DNA) I)	20q11.2-13.1	CTGATGCCAAGGTCATGAAG GTCCAGATAATTGAGTTTGGAGG	HM / HM
ADA (Adenosine deaminase)	20q12-q13.11	CCTGGTCCAGCTACCTCAC GCGGGTCATCTGTGTTGAG	HM / HM
CD40 (CD40 antigen)	20q12-q13.2	CTTATCACCCAGATGATTGGG AGTTCAGTAAGGATAAGGACCCC	CH / CH
GNAS1 (Guanine nucleotide binding protein, α stimulating activity polypeptide 1)	20q13.2-q13.3	GGAGGACAACCAGACCAAC CTTCGATTTTCCAGCAAGG	CH / CH
ADRA1A (Adrenergic, α-1A)	20	CAACTATTTCATCGTGAACCTGG TACACGCGGCAGTACATGAC	HR / HMR
ADRBK2 (Adrenergic, beta, recepter kinase 2)	22q11	ACCTTGGGTCTCATTATGGG CTCACCTTCATAGCTTTCAAACC	HMR / HR
GNAZ (Guanine nucleotide binding protein (G protein), alpha-z-polypeptide)	22q11.2	CGCTCACGGGACATGAC ACAGAAGATGATGGCTGTTACG	HR / HR
IGL@ (Immunoglobulin lambda gene cluster)	22q11.2	CCCGCCCTCCTCTGAGG GCCGCGTACTTGTTGTTGCT	2HNQ / 2HNQ
ACO2 (Aconitase 2, mitochondrial)	22q11.2-q13.1	CAGTGGCCAAGCAGGC GCCAGGGCTGTGACAATC	HMP / HMPR
LIF (Leukemia inhibitory factor)	22q12	CCCATTTGAGCATGAACCTC CCCAGCATCTCTAGGCAGTG	CHM / CH
NEFH (Neurofilament)	22q12.1-q13.1	AGCTGCTCGGACAGATCC CCTTGTCTTCTGTGGCCTTC	HM / HR
PDGFB (V-sis platelet-derived growth factor beta polypeptide (Simian sarcoma viral oncogene homolog))	22q12.3-q13.1	TTCAAGAAGGCCACAGTGAC GTCTCCTTCAGTGCCTTCTTG	HFMR / HM
CYP2D@ (Cytochrome P 450)	22q13	ATGGTGACCACCTCGACC CAGCACCGATGACAGGTTG	2CHM / 2CHM
ARSA (Arylsulfatase A)	22q13.3-qter	AGGCTACCTCACAGGAATGG AAGCAGGTCAGGTTCTGGC	HM / HM

[1]Species or order codes for both alignment and comparative data are: A = chicken, B = bat, C = cow, D = dog, E = echidna, F = cat, G = goat, H = human, I = shrew, J = Chinese hamster cell line (E36), K = cavia, L = , M = mouse cell line, (RAG), N = rabbit, O = opossum, P = pig, Q = horse, R = rat, S = sheep, T = deer, U = seal, V = vole, W = wallaby, X = xiphosphorus, Y = giant panda, Z = humpback whale, 1 = leopard cat, 2 = non-human primate, 3 = non-Mus and non-Cricetus rodent, and 4 = other species, including; bulbarus, cavia, callitrix, cervid, desmodus, didelphis, macropus, mustela, odocoileus, tragelaphus, trichosurus, tupaia, vulpes, and Xenopus. Exceptions: J = cricetulus and M = mouse for aligned species column.

*6 CATS loci already mapped by Zhang (1997).

3.3.2. Pig genomic DNA

Porcine genomic DNA from unrelated Australian pigs (Table 3.3) was available for a preliminary PCR test and for sequencing of PCR products. The PiGMaP reference pedigree (see Section 1.3.3.1 for more formation and see Appendix 3.1 for the pedigree) was available for linkage mapping and for evaluating RFLPs.

Table 3.3 Porcine DNA samples from 23 unrelated Australian animals representing nine European breeds (kindly provided by Dr P. Le Tissier).

I. D. number	Pig breed
01	Berkshire (BS)
02	Berkshire (BS)
03	Large White (LW)
04	Large Black (LB)
05	Tamworth (TW)
06	Landrace (LD)
07	Wessex Saddleback (WS)
08	Wessex Saddleback (WS)
09	Wessex Saddleback (WS)
10	Duroc (DC)
11	Duroc (DC)
12	Duroc (DC)
13	Hampshire (HS)
14	Hampshire (HS)
15	Hampshire (HS)
16	Berkshire (BS)
17	Large White (LW)
18	Large White (LW)
19	Landrace (LD)
20	Landrace (LD)
21	Landrace (LD)
22	Large Black (LB)
23	Welsh (WL)

3.3.3. Somatic cell hybrid panel

The description and method of analysis of the somatic cell hybrid panel is described in Section 1.3.2.2 and 3.4.8.

3.4. Methods

3.4.1. Preliminary PCR testing of CATS primers on porcine genomic DNA

Approximately 100 ng porcine genomic DNA was used as template for amplification using conventional *Taq* DNA Polymerase in a 25 µl PCR reaction volume, containing 1 to 3 mM $MgCl_2$ (depending on the primers), 100 µM dNTPs (dATP, dGTP, dCTP, and dTTP), 10 x *Taq* DNA Polymerase buffer and 10 pmol of a pair of CATS primers. This solution was overlaid with paraffin oil and the PCR was carried out initially with two thermocycling programs, namely "touchdown" PCR (more details see Section 2.7.3) and conventional PCR (more details see Section 2.7.2). In spite of optimization of PCR conditions, multi-banded products or smears were sometimes amplified. For these recalcitrant primers, AmpliTaq Gold DNA Polymerase (Perkin-Elmer, 5 units/µl) was tried. This modified *Taq* Polymerase is provided in an inactive state and can be activated in a pre-PCR heat step and/or slowly during thermal cycling (Slow-release PCR).

The annealing temperature and number of PCR reaction cycles was varied depending on primers and template DNA concentrations. PCR products were electrophoresed in 2% agarose gel (Progen) to determine whether single-banded products were being amplified. The PCR product sizes in base pairs were estimated comparing with ϕ X 174 RF DNA/*Hae* III size-standard ladder (Promega).

3.4.2. Purifying PCR product from agarose gel slices

Specific PCR product of similar size to that found in cats (Lyons *et al*., 1997) was cut from a 2% low melting point agarose gel (Progen). The gel slice was put into a 1.5 ml Eppendorf tube and 100 µl of TE (Tris-EDTA, pH = 8.0) was added. This mixture was placed onto a Heat Block (80-85 °C) until it became liquid. 400 µl of phenol (approximately equal to the gel slice volume) was added to this solution and mixed thoroughly. The mixture was put into the heat block again to liquify and was then centrifuged for 10 minutes in a microcentrifuge. The supernatant was collected without disturbing the intermediate (agarose) and lower (phenol) layers. 500 µl of chloroform was added to the supernatant in a new tube and spun for 5 minutes. The aqueous phase was collected again into a new Eppendorf tube. 40 µl (one-tenth of the total volume) 3 M sodium acetate and 700 µl (equal volume) of isopropanol were added and the tube placed at -80 °C for 15 minutes. The solution was spun for 10 minutes to precipitate the DNA. The supernatant was poured off and 500 µl of 70 %

ethanol added to rinse away residual sodium acetate. The ethanol was then poured off and the DNA pellet was briefly dried on a heat block (42 °C) for 5 minutes. 10 to 15 µl of TE was added to the Eppendorf tube to dissolve the DNA.

3.4.3. Direct Sequencing of PCR products using ABI 373 automatic DNA sequencer

When a single-banded PCR product of the correct size was amplified, 100 ng of PCR product was used directly as template for sequencing using an ABI 373 automatic sequencer (Perkin Elmer) (Section 2.10.1). 3.2 pmol of forward or reverse primer was used to enable sequencing from one strand of template DNA.

3.4.4. Cloning and sequencing OXT PCR products from Landrace and Large White breeds

To investigate the difference in sequence of the *OXT* gene between two breeds, the *OXT* PCR products of Landrace and Large White pig breeds were cloned into pCR® 2.1 – TOPO vector (Invitrogen, TOPO™ TA Cloning Kit) (see Section 2.1 and Appendix 2.1 for more details) in order to look for Single Nucleotide polymorphisms (SNPs).

3.4.5. Verification of the sequences

The sequences obtained were screened against the GenBank database using the FASTA program (Pearson and Lipman, 1988) via the ANGIS (Australian National Genomic Information Service; http://www.angis.usyd.edu.au) interface, to investigate homology between PCR product sequences and the target sequences.

3.4.6. Searching for restriction enzyme sites for distinguishing porcine and rodent PCR products from the somatic cell hybrid panel

Because a somatic cell hybrid panel was used as a physical mapping tool, a major problem arose from the fact that the CATS primers were very likely to amplify PCR product from the rodent (mouse or hamster) background present in all clones in the somatic hybrid panel. Since mouse

sequence (Lyons *et al*., 1997) was very frequently included in the consensus, the problem was severe. When all somatic cell hybrid clones generated identical PCR products, restriction enzymes, especially four-base cutters, viz. *Alu*I, *Hha*I, *Hae*III, *Msp*I, *Rsa*I, *Sau3A*I and *Taq*I, were used to attempt to distinguish porcine and rodent PCR products. If there were no differences detected by the above restriction enzymes, mouse sequences were obtained from Genbank and in some cases hamster sequences were generated using an ABI 373 automatic sequencer and then confirmed by FASTA comparison. Restriction enzyme maps were generated for pig, mouse and hamster, using the GCG MAP program, to identify an enzyme able to distinguish the products from these species.

3.4.7. Screening for porcine RFLPs

PCR products, verified by comparison with GenBank, were tested for porcine RFLPs using four-base recognition sequence restriction enzymes (*Alu*I, *Hha*I, *Hae*III, *Msp*I, *Rsa*I, *Sau3A*I and *Taq*I) on products from 23 unrelated Australian pigs (see Table 3.3 for the list of pig breeds). The protocol used for restriction enzyme digestion has been described in Section 2.8. The restriction enzyme digestion reaction was stopped by adding 5 µl of 10 x agarose gel loading buffer and the digested PCR products were loaded onto a 2 % agarose gel containing 0.5 µg/ml ethidium bromide. Electrophoresis was run at approximately 100 V (10 V/cm) so that optimal separation of the fragments of interest occurred within an hour. Finally, the gel was viewed and photographed on an ultra-violet light transilluminator.

3.4.8. Physical mapping - somatic cell hybrids

DNA from the hybrid clones was analyzed by PCR for all CATS primer pairs confirmed by sequencing to amplify the correct porcine genes. Each primer pair was used with its optimal $MgCl_2$ concentration and annealing temperature to allow the visualization of the porcine-specific product on an agarose gel. Diagnostic PCRs were carried out in 25 µl reaction volume of 1 × PCR buffer (Promega), consisting of 20 ng DNA from 27 hybrid lines, pig and rodent parental lines (mouse LMTK and Chinese hamster Wg3hcl2), 10 pmol of each primer, 100 µM of each dNTPs, 0.5 to 3 mM $MgCl_2$ and 1 U *Taq* DNA polymerase. Reactions were performed in a PTC-100 Programmable Thermal Controller (MJ Research, Inc.) using the PCR programme described previously except the number of cycles was increased to 45 cycles (see Table 2.1 for more details).

Chromosomal and regional assignments were achieved through observing the concordant segregation of porcine PCR products with chromosome fragments retained in the hybrid cells. The pattern of amplification of porcine PCR products depends on the presence/absence of the relevant porcine chromosomal regions in each of the 27 somatic hybrid clones developed by Yerle *et al.* (1996).

A World Wide Web (WWW) site at the INRA laboratory, Toulouse, France, was accessed to obtain all relevant somatic cell hybrid panel information and to interpret the PCR data. The address is http://bio.toulouse.inra.fr/lgc/pig/hybrid.htm. This site enables entry of the results from the somatic cell hybrid panel and provides an immediate assignment to a chromosomal region using a programme to interpret the new PCR data in relation to all previously lodged data, and also provides statistical output on the reliability of the assignment (Chevalet *et al.*, 1997). Loci can be assigned to 127 different non-overlapping subchromosomal regions. The programme also calculates correlation coefficients, which are measures of association (co-segregation) between markers, and which have values of 1 or close to 1 when two loci are syntenic. However, provided the correlation coefficient is greater than 0.8, a locus can be reasonably assigned to a specific chromosomal region (Chevalet *et al.*, 1997; Robic *et al.*, 1996; Yerle *et al.*, 1996).

3.5. Results

3.5.1. Optimum conditions for PCR

Factors affecting PCR yield are mainly annealing temperature and $MgCl_2$ concentration. Because some CATS primers may not perfectly match the pig template, adjustment of these factors may improve PCR results. For the CATS loci located on human chromosome 9 and 10, PCR conditions were initially evaluated by Zhang (1997). Of the 47 primer pairs tested, 21 gave single-banded PCR products using conventional *Taq* DNA polymerase with minor adjustment of reaction conditions. With use of AmpliTaq Gold Polymerase, 14 more primer pairs produced better results with a single or clear dominant band. 12 primer pairs continued to give double, triple or multiple bands even with variation of PCR conditions (Table 3.4).

Table 3.4 Optimum PCR conditions and porcine PCR product size for CATS markers mapping to human chromosome 9, 10, 20, and 22. The feline PCR product sizes (Lyons *et al.*, 1997) are also presented for comparison.

Locus symbol[1]	PCR product size (bp)[2]	$MgCl_2$ (mM)	Annealing temperature (°C)[3]	PCR product size in cat (bp)
ABL1	1,200*#	2	50	1,100
ALDH1	900*	2	59	1,000
ALDOB	300	2	54	300
*AMBP***$**	450	2	58	450
ANX1	150	2	53	600
ASS	700, 400	3	50	1,300
C5	200*#	3	50	1,000
CNTFR	270, 600	2	50, 57	600
DBH	210, 500	2	55, T	500
GALT	500	2	57	475
GGTB2	320	2	55	1,100
*GRP78***$**	800	2	52	400
GSN	150	2	52	194
HXB	150	2	56	600
IFN1@	400	2	54	400
*IFNB1***$**	450	2	58	450
RLN1	300	2	56	500
*SPTAN1***$**	180	2	54	700
*TXN***$**	800*	2	55	600
ADRA2A	170	1	53	200
ADRB1	200	1.5	54	190
ALOX5	1,500#	2	52	500
APT1	270	2	56	400
BMI1	750	1	64	200
CHAT	Multiple bands	2	55	861
CREM	320	1	51	1,000
CYP2E	Multiple bands	2	58	900
DNTT	Multiple bands	2	53	900
*EGR2***$**	150	1	51	800
HK1	150, 1,200#	2	50	800
IL2RA	180	2	55	100
OAT	1,500*	2	58	2,000
PLAU	600*	2	52	500
PP	Multiple bands	2	50	600
PRF1	120	2	53	1,000
VIM	200#	2	55	750
PRNP	250*	2	51	250
OXT	600	1	58	900
SRC	120	1.5	60	300

TOP1	400*	2	54	500
ADA	120	1	60	1,000
CD40	130*#	2	52	2,000
GNAS1	320*	2	55	500
ADRA1A	450	2	65	453
ADRBK2	Multiple bands	1.5	51	1,000
GNAZ	150*	2	50	150
IGL@	190*	2	60	220
ACO2	280, 600	2	55, T	1,200
LIF	100, 250, 450	1, 1, 2	51, 60, T	900
NEFH	400, 600*	2, 1.5	60, 54	375
PDGFB	200*, 280	1.5, 2	54, 51	603
CYP2D@	500, 450	2	53, T	700
ARSA	230*	2	60	600

[1]**$**: Mapped by Zhang (1997)
[2]: *PCR products amplified by AmpliTaq Gold DNA Polymerase (Perkin-Elmer); # PCR fragments with very weak bands.
[3]T: 'Touchdown' PCR

3.5.2. Sequences of PCR products

For the primers giving strong major bands, one of the primers was used to directly sequence the PCR product (more details see Section 2.10.1). For the loci from human chromosome 9 and 10, if the amplified PCR product size was approximately the same as obtained by Zhang (1997), a sequencing reaction was not carried out because this would have simply duplicated Zhang's results. The PCR products from the rest CATS primers were sequenced partially or fully except for five primer pairs (*CHAT, CYP2E, DNTT, PP, ADRBK2*) which gave multiple bands.

3.5.3. Comparison of the sequences with the GenBank database

The putative homologues of the submitted sequences were identified using FASTA searches of the GenBank database. 23 CATS products, including 13 analysed by Zhang (1997), were confirmed to match the target loci because they aligned with relevant genes of human or other mammals (Table 3.5 and Appendix 3.1). Zhang (1997) also generated a product for *BMI1*, but this did not match with anything in the GenBank database. In the second trial, a longer and more accurate *BMI1* CATS sequence was obtained, confirming that the appropriate product had been amplified. However 25 products still did not match the CATS target loci.

Table 3.5 Alignment of sequences from 23 CATS PCR products with genes in the database.

Gene symbol	Matching Gene in GenBank database	Species in the database	Expectation for the best correct match[1]
ADRA2A	PIGA2AR	Pig	$1.1 \times e^{-30}$
*ADRB1**	MMB1AR	Monkey	$2 \times e^{-25}$
*ALDOB**	HUMALDB1	Human	1.8
*AMBP**	SSAMGBIK	Pig	$2 \times e^{-6}$
*APT1**	MUSFASANT	Mouse	4.5
BMI1	HUMBMI1X	Human	$4.5 \times e^{-15}$
*EGR2**	HUMEGR2A	Human	$1.8 \times e^{-11}$
*GRP78**	SSGRP78	Pig	$8.1 \times e^{-7}$
*GSN**	SSGELSOL	Pig	$1.3 \times e^{-30}$
*IFN1@**	SSPOIII5	Pig	$4.2 \times e^{-10}$
*IFNB1**	PIGINFB	Pig	0
*OAT**	HUMOAT10	Human	$9.4 \times e^{-5}$
*PLAU**	SSUPAG	Pig	$3.3 \times e^{-26}$
*SPTAN1**	HUMASPX	Human	0.0029
*TXN**	HSTHDC	Human	$8 \times e^{-6}$
OXT	HUMOTNP1	Human	$2.1 \times e^{-10}$
TOP1	HUMTOP18	Human	$1.1 \times e^{-24}$
CD40	HUMCD40L1	Human	$3.4 \times e^{-6}$
GNAS1	HUMGNAS6	Human	$5.4 \times e^{-31}$
ADRA1A	HUMA1AADR	Human	0
GNAZ	HUMGNAZ	Human	$6.4 \times e^{-27}$
IGL@	PIGIGLVJC	Pig	$1.5 \times e^{-29}$
ARSA	HSARYLA	Human	$3.5 \times e^{-7}$

*Data were generated by Zhang (1997).

[1]Statistical expectation for the sequence. The number of sequences with this identity expected in a database of this size by chance alone. An value less than 0.05 is considered statistically significant.

There are several possible explanations for the unmatched CATS sequences. Firstly, these sequences may be predominantly from nonconserved introns and lack sufficient length of conserved exonic sequence to allow alignment with GenBank sequences. For example, the 160 bp *ARSA* CATS sequence did not match significantly with any sequence in the GenBank database search. However, 40 bp at the beginning of the sequence were well matched with *ARSA* human sequences. Because most of the *ARSA* CATS sequence was from an unconserved intron, the FASTA alignment could not match it with homologous sequences from other species. Secondly, since the CATS primers were designed predominantly from human and mouse sequences, it is possible that the primers have amplified artefactual products in the pig genome. Thus, it is not clear whether the sequences from these 25 CATS PCR products are primarily intronic or artefactual. Thirdly, it is possible to have

errors in primer design or recording. Zhang (1997) found that the primer pair for *RLN1* flanks a fragment of about 5.6 kb in the human genome, which is impossible to amplify reliably from a genomic template under standard PCR conditions. Also, he found that the reverse primer of *ANX1* was not found in any relevant gene sequences when used in FASTA to query the GenBank database. Therefore it is not surprising that the PCR product amplified by the *ANX1* primer pairs could be an artefact. Finally, direct sequencing of the CATS PCR products, which were sometimes of poor quality, generated many unreadable nucleotides, $(N)_n$. In these cases, finding a match to the target locus in the GenBank data base was made more difficult.

3.5.4. Screening for PCR-RFLP in CATS products

Products from the following six loci were tested for RFLP: *ADRA1A, BMI1, GNAS1, GNAZ, OXT* and *TOP1*. After cutting with four-base pair recognition restriction enzymes, including *Alu*I, *Hha*I, *Hae*III, *Msp*I, *Rsa*I, *Sau3A*I and *Taq*I, an RFLP was obtained for only one of the five loci, namely *OXT* (Fig 3.2 and Fig 3.3).

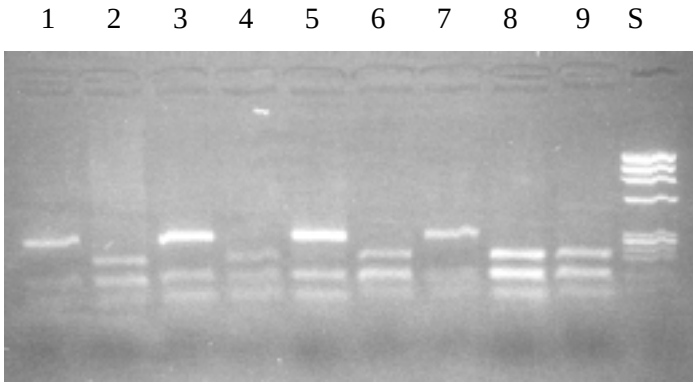


Fig 3.2 The *OXT Msp*I RFLP genotypes seen in unrelated Australian pigs, S: size standards, ϕX174/*Hae*III DNA ladder (Promega). Lane 1 and 7: Large White, lane 2 and 8: Landrace, lane 3: Wessex Saddleback, Lane 4: Duroc, Lane 5: Hampshire, Lane 6: Berkshire, Lane 9: Welsh. Lanes 1, 3, 5 and 7 are AA genotype and lanes 2, 4, 6, 8 and 9 are aa genotype.

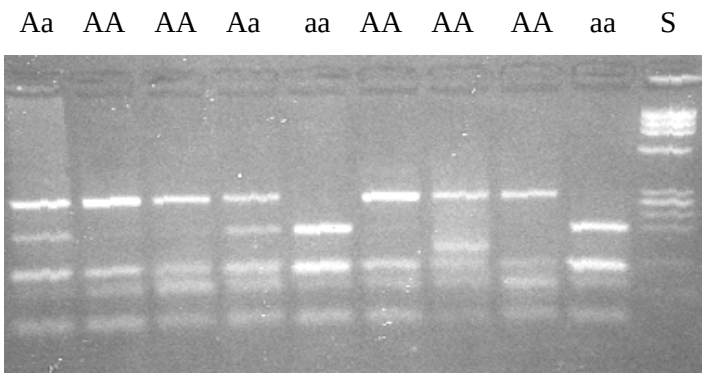


Fig 3.3 The *OXT Msp*I RFLP genotypes seen in the PiGMaP reference families, showing three apparent genotypes (see lane headings). S: size standards, ϕX174/*Hae*III DNA ladder (Promega).

3.5.4.1. An *Msp*I RFLP for *OXT*

A *Msp*I restriction site (C↓CGG) polymorphism was found for the *OXT* PCR product. When the PCR product was digested using *Msp*I, two different genotypes (AA and aa) were observed from several unrelated Australian pig samples, and all three genotypes (AA, Aa and aa) were observed in the PiGMaP reference families. This *Msp*I RFLP was genotyped in the PiGMaP pedigree. Among the six PiGMaP families, the Edinburgh-1, German and Swedish families were informative, but no variation was found in the Edinburgh-2, French and Dutch families (see Appendix 3.2 for detailed information). Interestingly, a null allele also was detected in the sample of unrelated Australian pigs and apparently in the PiGMaP reference families. *OXT* genotypes did not follow simple Mendelian codominant inheritance patterns, apparently also because of the presence of a null allele, and thus the locus could not be linkage mapped.

3.5.4.2. Sequence comparison between Large White and Landrace breeds for *OXT*

Different *Msp*I variants were found in the Large White and Landrace samples from our collection of DNAs from unrelated Australian pig. Sequencing showed that the polymorphic *Msp*I restriction site is located in an intron. The exon sequences for the Large White and Landrace animals were identical but there were numerous other intronic variants in addition to the *Msp*I site (Fig 3.4). These results for porcine *OXT* confirm the validity of Lyons *et al*. (1997) general approach since the conserved exons enabled amplification and the variable introns have provided polymorphism.

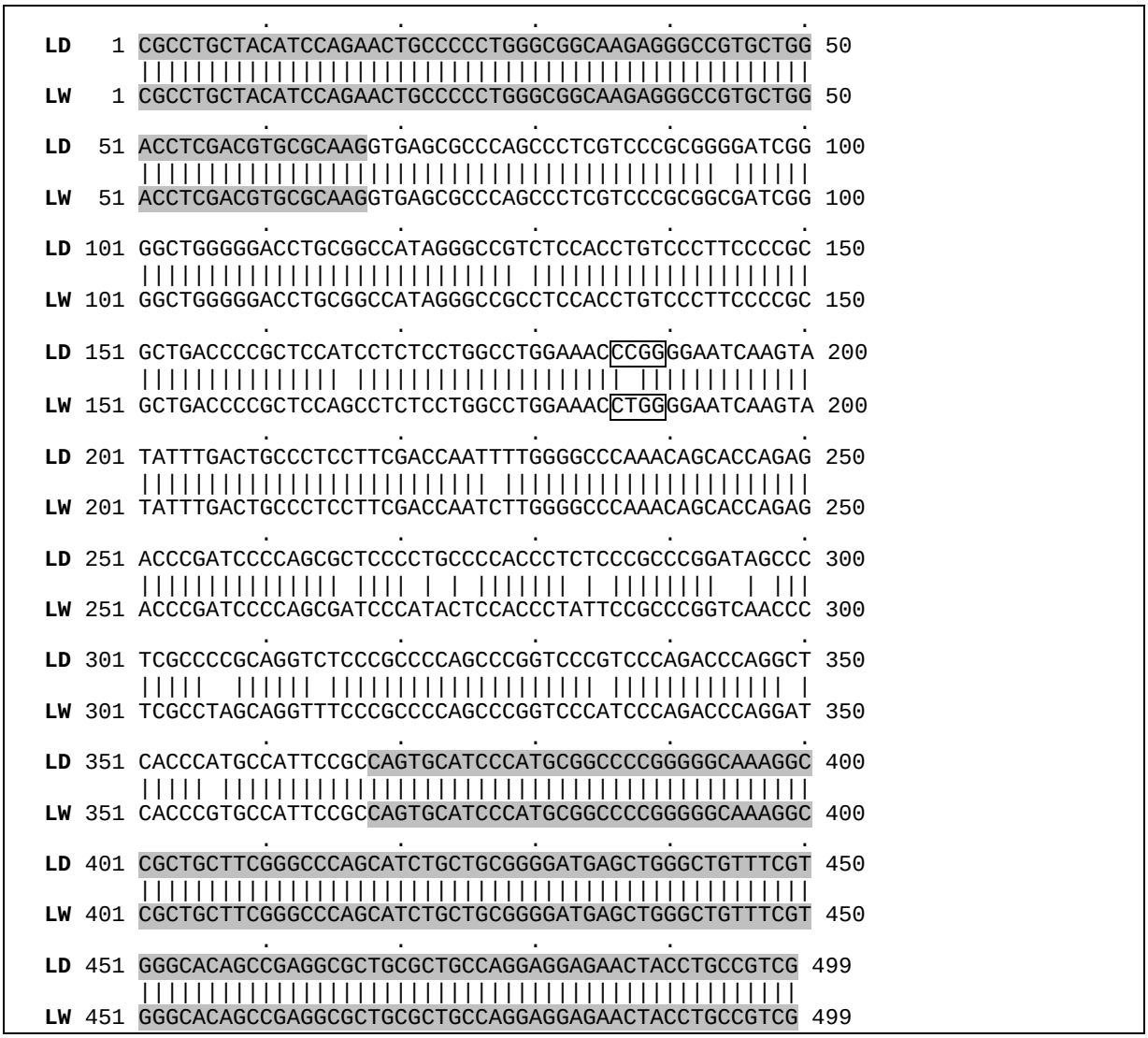


Fig 3.4 Alignment of OXT sequences from CATS PCR products from Landrace and Large White pigs. Exon sequences are shaded (…). The *Msp*I variant restriction site is boxed (…).

3.5.5. Restriction enzyme digestion of mouse, hamster, and pig template for analysis of the somatic cell hybrid panel

For seven loci, namely *ADRA1A, ADRA2A, ADRB1, GNAS1, GNAZ, GSN* and *IGL@*, an attempt was made to discriminate between porcine, mouse and hamster PCR products using restriction enzymes. The choice of restriction enzyme was guided by alignment of pig and rodent sequences. The restriction enzymes evaluated are summarised in Table 3.6. An example, for

ADRA2A, where the porcine product can be distinguished from the rodent PCR products, is shown in Fig 3.5.

Table 3.6 Restriction enzymes used to attempt to distinguish the porcine, mouse and hamster PCR products amplified by CATS primers.

Locus	Restriction enzymes evaluated	Enzymes providing discrimination
ADRA1A	*Rsa*I, *Alu*I, *Sac*II, *Apa*I, *Nco*I	*Rsa*I
ADRA2A	*Alu*I, *Taq*I	*Alu*I, *Taq*I
ADRB1	*Mbo*I, *Rsa*I	*Rsa*I
GNAS1	*Cfo*I	*Cfo*I
GNAZ	*Cfo*I, *Taq*I, *Sac*I, *Ban*II	*Taq*I, *Sac*I, *Ban*II
GSN	*Hae*III, *Hinf*I, *Taq*I, *Ban*I	*Hae*III
IGL@	*Sau3A*I, *Rsa*I, *Kpn*I, *Taq*I, *Ban*I, *BstO*I, *Nco*I	*Ban*I, *BstO*I

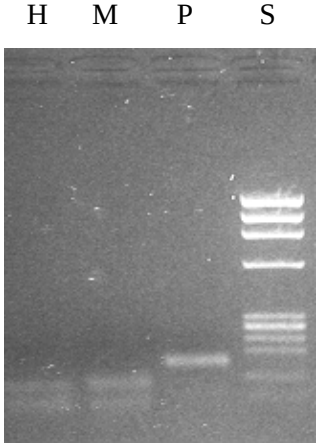


Fig 3.5 Interspecific *Taq*I restriction variants for the *ADRA2A* products. H: Chinese hamster, M: Mouse, P: Pig, S: size standards (ϕX174/*Hae*III DNA ladder, Promega).

3.5.6. Somatic cell genetics of ADRA1A, ADRA2A, ARSA, GNAS1, OXT, TOP1

The distributions of positive PCR results from the 27 somatic hybrid clones for *ADRA1A, ADRA2A, ARSA, GNAS1*, *OXT* and *TOP1* are presented in Table 3.7 and the definition of cytogenetic regions detected are presented in full in Appendix 3.3. A summary of the mapping results, including chromosomal and regional probabilities, is presented in Table 3.8. For illustration, the CATS amplification products for *ARSA* from the somatic hybrid panel are shown in Fig 3.6(a) and their

interpretation in Fig 3.6(b). PCR screening of the somatic cell hybrid panel assigned *GNAS1* and *TOP1* to porcine chromosome 17, *ARSA* to porcine chromosome 5, *ADRA2A* to porcine chromosome 6 or 14, *ADRA1A* to porcine chromosome 16, and *OXT* to porcine chromosome 14.

Unfortunately, 11 CATS products, namely *ADRB1*, *ALDOB*, *APT1*, *BMI1*, *CD40*, *GSN*, *GNAZ*, *IFNB1@*, *IGL@*, *OAT* and *PLAU* could not be mapped using the somatic cell hybrid panel, mainly due to strong background amplification from every clone in the somatic cell hybrid panel. Even though restriction enzymes were found that could distinguish porcine from rodent PCR products from control DNAs for these species, the restriction digestion test could not be reliably applied to products amplified from the hybrid panel.

Table 3.7 The distribution of PCR results in somatic cell hybrid panel

Locus	Somatic cell hybrid clones[1]																									
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
ADRA1A	-	+	-	+	-	+	+	-	+	+	+	+	-	-	-	+	-	-	+	-	-	-	-	-	-	-	-
ADRA2A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	+	-
ARSA	-	+	+	+	-	+	+	-	+	+	+	-	+	+	-	-	+	+	+	-	-	+	-	-	-	-	+
GNAS1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-	-	-	-
OXT	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	+	-
TOP1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-	-	-	-

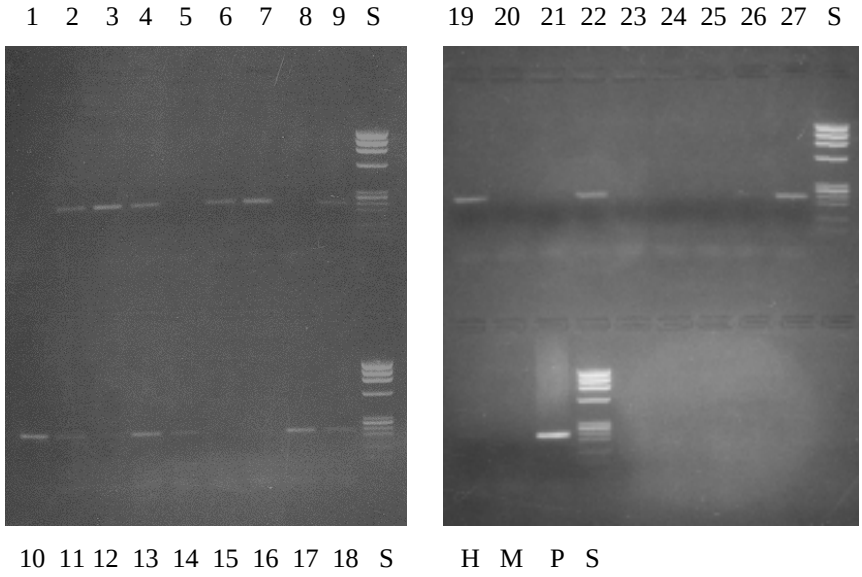
[1]+ and - indicate the positive and negative PCR products

Table 3.8 Cytogenetic localisation, correlation coefficients and probability values of *ADRA1A*, *ADRA2A*, *ARSA*, *GNAS1*, *OXT* and *TOP1* CATS products on the somatic cell hybrid panel.

Locus symbol	Human location	Porcine Location	Error risk[1]	Chromosomal probability[2]	Regional probability[3]	Comment[4]
ADRA1A	20	16q1.4/ 16q2.2-2.3	0.1 %	1	0.47/0.47	*Rsa*I (all clones)
ADRA2A	10q23-q25	6(1/2q2.1)/ 14q2.5-2.9	5 %	0.52/0.48	0.42/0.42	*Taq*I (all clones)
ARSA	22q13.31-qter	5p1.1-p1.5	0.1 %	1	1.00	
GNAS1	20q13.2-q13.3	17(1/2q2.1)-2.3	0.5 %	0.98	0.88	*Cfo*I (mouse background clones only)
OXT	20p13	14q2.5-2.9	5 %	0.79	0.69	
TOP1	20q11.2-q13.1	17(1/2q2.1)-2.3	0.5 %	0.98	0.88	

[1]The probability that the proposed chromosome and region are incorrect. [2]The posterior probability that the locus is located on the nominated chromosome. [3]The posterior probability that the locus is located in the nominated region. [4]Restriction digestion required distinguishing porcine product from hamster or mouse background PCR product.

(a)



(b)

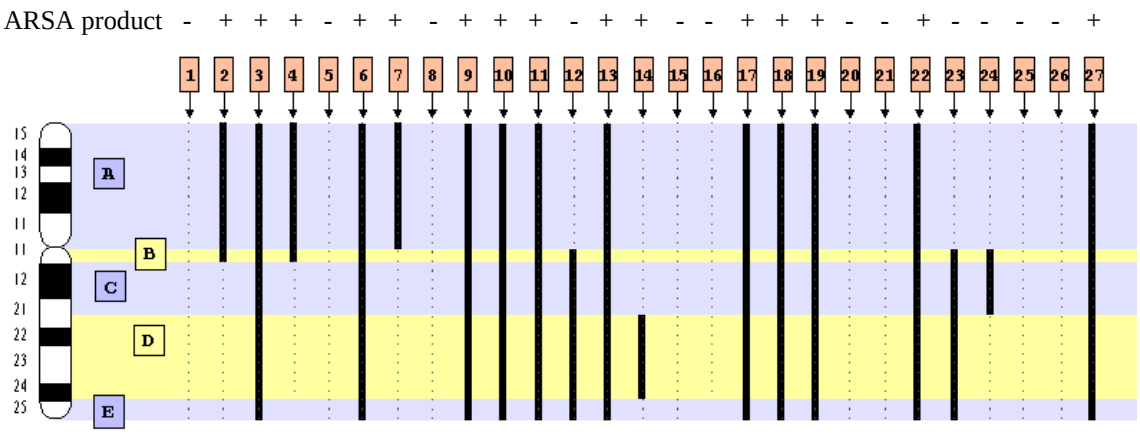


Fig 3.6 Example of physical mapping of a CATS product to SSC 5 using the somatic cell hybrid panel. (a) PCR amplification of *ARSA* product using CATS primers on the 27-clone somatic cell hybrid panel. In this case, a rodent background product did not amplify. H: Chinese hamster, M: Mouse, P: Pig, S: size standards (ϕX174/*Hae*III DNA ladder, Promega) (b) Proportion of pig chromosome 5 retained in each of the 27 somatic hybrid clones. Clones positive for the porcine *ARSA* products are indicated by a +. The presence of product for clone 14 may be an artefact or may indicate that the description of this clone 14 is incorrect. A, B, C, D, E indicate regions of SSC 5 distinguishable by the panel.

3.5.7. Comparative map

Six loci only were physically mapped in this study from 47 pairs of CATS primers. Of these, five loci (*ADRA1A, ARSA, GNAS1, OXT, TOP1*) came from human chromosome 20 and 22. Only one additional locus (*ADRA2A*) from human chromosomes 9 and 10 was mapped from the 30 CATS primer pairs for which Zhang (1997) was previously unsuccessful. There is no significant difference between the success rates for mapping loci from human chromosomes 9 and 10 versus chromosomes 20 and 22 based on the Chi-square homogeneity test ($\chi_1^2 = 0.655$, p = 0.419). A comparison of the location of these mapped CATS loci between human and pig confirms recognised conserved synteny between these two species and agrees well with previously recognised conserved relationships between pig and human genomes (Johansson *et al*., 1995; Rettenberger *et al*., 1995b; Fröncke *et al*., 1996) except for the locations of *OXT* and *ADRA1A*. As discussed below, the map location of *OXT* is less certain than other loci and the porcine location of *ADRA1A* shows a potential confusion of gene family members. Fig 3.7 shows the map locations in pigs compared with the human physical map location, including 6 CATS loci mapped by Zhang (1997).

3.6. Discussion and conclusion

In this study, 47 consensus primer pairs, mainly based on the human and mouse sequences, were tested in pigs for identification of conserved chromosomal regions between species. However, the sequence divergence between human/mouse and pig may lead to mismatches in the primer target region and/or the creation of spurious complementarity in non-target regions. Therefore, it is highly desirable to eliminate artefact bands or at least identify target bands to enable appropriate interpretation of the CATS PCR products for linkage and physical mapping.

AmpliTaq Gold DNA Polymerase used in the slow-release protocol helped in eliminating some confusing multiple banding patterns, but did not eliminate the problem completely. Ultimately, 12 primer pairs (*ACO2, ADRBK2, ASS, CHAT, CNTFR, CYP2D@, CYP2E, DBH, DNTT, HK1, LIF, PP*) failed to yield interpretable PCR products for sequencing because of the confusing pattern of multiple smearing bands with different trials of PCR conditions. Including Zhang's (1997) results, 12 loci have now been physically mapped and 23 CATS primers have amplified PCR products that have been confirmed to match their target loci. The success rate for amplification of target PCR product was 43% (23/53) based on short single-pass sequences obtained directly from the CATS PCR products. It is possible that some of these sequences were intronic only and therefore were unlikely to find a match in the GenBank database, so the true success rate could be higher.

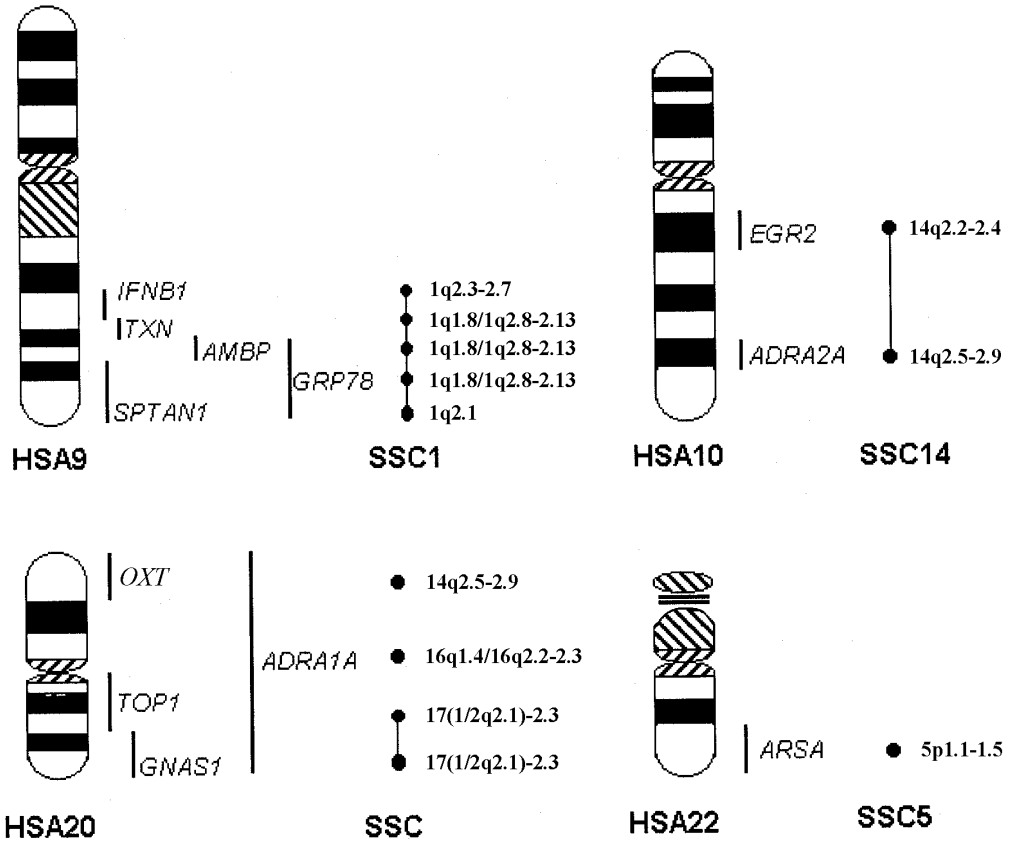


Figure 3.7 Physical localisation of CATS loci in pig, shown relative to their human locations. The source of human map locations is based on the human genome database (GBD, http://gdbwww.gdb.org/)

A disappointingly high proportion of CATS primers were found to amplify only coding sequences contrary to their stated design principles. The exon sequences are unlikely to be informative for linkage mapping and as a result only one locus could be linkage mapped by Zhang (1997) with a reasonable amount of effort in detecting polymorphism. More disappointingly, the nature of the CATS consensus sequence design meant that for about 50% (11/23) loci, rodent PCR products coamplified and could not be distinguished from porcine products in the somatic cell hybrid panel with the range of enzymes employed. Almost every consensus primer is based on mouse (rodent) sequence contributing to consensus. Therefore on average CATS primers will work much better with mice (rodents) than other species.

An inconsistency between Zoo-FISH and somatic cell hybrid panel results was found for the *OXT* and *ADRA1A* loci. *OXT* maps to SSC 14. This position is not consistent with HSA 20 but rather

with HSA 10. However, the syntenic relationship between SSC 14 and human chromosomes has not been well defined with Zoo-FISH (cytogenetic map of the pig, http://www.toulouse.inra.fr/lgc/pig/cyto/cyto.htm; August 2000). Also it was hard to tell whether this *OXT* map position was correct or the results of artefactual PCR bands from the somatic cell hybrid panel because the 5 % error rate in the analysis is the borderline value, above which no decisions can be made on location (see Appendix 3.3) and the low number of positive clones in the somatic cell hybrid panel. If the location on SSC 14 is correct, this result indicates that there is an unidentified homologous chromosomal segment between HSA 20 and SSC 14. Also it is possible that small chromosomal segments could have been rearranged in the somatic hybrids, generating spurious mapping results (Dr. Bhanu Choudhary: Personal communication). Fortunately, *OXT* sequences for two pig breeds (Large White and Landrace) have been generated in this study and will provide resources to linkage and physically map this locus in the future.

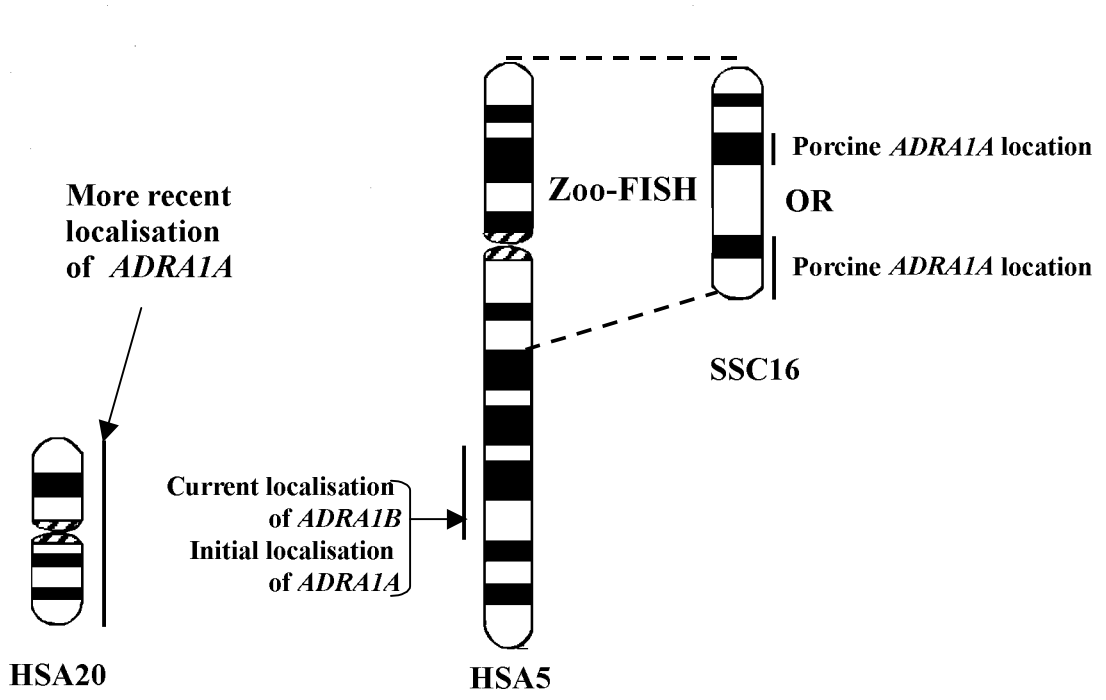


Fig 3.8 *ADRA1A* in humans was assigned first to HSA5q and then reassigned to HSA20. SSC16 is homologous with HSA5p extending a small way into the q arm. *ADRA1B* (a member of the same gene family as *ADRA1A*) maps on HSA5q but again outside of the region of recognised homology between HSA5 and SSC16.

The porcine *ADRA1A* CATS product maps to SSC 16. Human *ADRA1A* was initially assigned to HSA5 (Yang-Feng *et al*., 1990; Lomasney *et al*., 1991). All of SSC16 is syntenic with HSA5, but the region of synteny does not include the site of this original human assignment nor the current site of localisation of *ADRA1B* to which the porcine CATS product also shows high sequence similarity (Fig 3.8). The porcine localisation of *ADRA1A* is inconsistent with the current, reassigned human position on HSA20 (Loftus *et al*., 1994). The results suggest that a careful revision of the human localisations of *ADRA1A* and *ADRA1B* is required. They also suggest that more complex small rearrangements may have occurred between human and pig genomes that affect the segment containing *ADRA1A*.

ADRA2A was assigned with equal probability to two different physical map locations, 6(1/2q2.1) and 14q2.5-2.9. However, Rettenberger *et al*. (1996) have previously shown that *ADRA2A* maps to SSC 14, not SSC 6, consistent with this assignment to SSC 14q2.5-2.9.

In August 1998, a prelimnary comparative mapping report was presented at the 26th ISAG (International Society for Animal Genetics) Conference, held in Auckland, New Zealand. Results collected from several pig mapping groups showed that over 131 CATS primer pairs had been tried, with 35 CATS PCR products confirmed by sequencing, and 22 subsequently mapped, including the results presented in this Chapter.

Due to the low levels of polymorphism, frequently indistinguishable rodent and porcine products in somatic hybrids and some confusion of identity of gene family members, the CATS primers were not a very powerful tool for cross-species identification and mapping of type I markers. However these CATS primers did provide some useful information. Also if better designed, for example to ensure presence of introns in amplified product and to avoid over reliance on mouse sequences, they might more usefully contribute to the porcine-human comparative map. The comparative map, which is based on Type I markers, will permit the identification of intrachromosomal rearrangements which are not visible to Zoo-FISH, allow a more precise recognition of evolutionary breakpoints, and thus contribute to the understanding of the evolutionary relationships between mammalian genomes and the reconstruction of the evolutionary history of chromosomes within mammals.

Chapter Four

Characterising Porcine Endogenous Retroviruses (PERVs) in Westran pigs

4.1. Introduction

Porcine endogenous retroviruses (PERVs) are proviral forms of retrovirus and are inherited in a stable Mendelian fashion (Patience *et al*., 1997), as well as being acquired through infection. The discovery of porcine retroviruses dates back to 1970. Type-C virus particles released in two cell lines from pig kidney were first described by Breese (1970), with particles being morphologically similar to the mouse type-C leukaemia viruses. Soon after, C-type viral particles were reported in five different pig leukemia cell lines (Armstrong *et al*., 1971). Todaro *et al*. (1974) showed that porcine retroviruses were present in multiple copies in the porcine genome in DNA from different tissues and cells. Lieber *et al*. (1975) described the biological and immunological properties of porcine type-C viruses. Cell lines from most mammalian species were found resistant to infection with the viruses but they are otherwise very similar to other mammalian C-type viruses by morphological, biochemical and immunological criteria. Benveniste and Todaro (1975) showed that related endogenous retroviral sequences were found in close wild relatives within Suidae, like bush pig and wart hog, and were absent from the peccary and other artiodactyls, like cattle. They concluded that C-type viruses were introduced into the Suidae lineage as a result of trans-species infection by an ancestral xenotropic murine virus.

Benveniste and Todaro (1975) could not find evidence for expression of porcine endogenous retroviruses in normal non-leukomogenic tissues, but they are expressed in most cell lines and in lymphosarcomas. Lymphosarcomas occur at a frequency of 3-50 x 10^{-6} among slaughtered animals and account for more than 25% of porcine neoplasms detected in abattoirs (Bostock and Owen, 1973). Suzuka *et al*. (1985) reported the isolation of a swine C-type retrovirus from malignant lymphomas. This C-type retrovirus was cloned into a λ phage vector and characterised by restriction digestion of the entire 8.8 kb virus (Suzuka *et al*., 1986).

Due to the shortage of human organs and tissues for transplantation, the use of non-human species is considered a possible solution. Success in modulating immunological rejection by

transgenic modifications to animals has raised the possibility of clinical introduction of xenotransplantation (Cozzi and White, 1995; Sharma *et al*., 1996; Fodor *et al*., 1994). Pigs are regarded as a better source than primates for xenotransplantation into humans for safety, financial, ethical, and practical reasons. Clinical trials have been carried out with pig xenografts, including perfusion of human blood through pig livers or using porcine hepatocytes as a bridging strategy for hepatic failure, the use of pancreatic islet cells as a treatment for chronic diabetes, and the implantation of fetal neuronal tissue as a therapy for Parkinson's disease (Chari *et al*., 1994; Deacon *et al*., 1997; Groth *et al*., 1994).

Recently porcine retroviruses have become a focus of concern, as they infect human cells *in vitro* (Martin *et al*., 1998a; Patience *et al*., 1998; Wilson *et al*., 1998), although there is no evidence that this occurs *in vivo* in baboons (Martin *et al*., 1998b) or in humans (Paradis *et al*., 1999; Pitkin and Mullon, 1999; Switzer *et al*., 1999). Akiyoshi *et al*. (1998) suggested that the risk of viral infection would be increased in xenotransplantation by the presence of factors commonly associated with viral infection e.g. immune suppression, graft-versus-host disease, graft rejection, viral co-infection, and cytotoxic therapies. Very recently, the transplantation of porcine pancreatic islets into SCID (severe combined immunodeficiency) mice led to *in vivo* expression of PERVs, reinforcing fear about the risk of PERV infection in immunosuppressed human patients (van der Laan *et al*., 2000).

The viral envelope is the major determinant of host range and is essential for infection. Two main types of pig retrovirus, PERV-A and PERV-B, which differ by 507 bases in their envelope (*env*) gene, are widely distributed in different pig breeds, with PERV-A predominant in European pigs (~70%) and PERV-B predominant in Asian (~60%) pigs by Southern hybridisation (Le Tissier *et al*., 1997). PERVs are present at approximately 50 copies in different breeds of pig (Akiyoshi *et al*., 1998; Le Tissier *et al*., 1997).

Host range analysis initially showed that PERVs are restricted in their species tropism, infecting only porcine cells and not cell lines derived from a range of species including chimpanzee, rhesus monkey, horse, mink, bat, rabbit, cow, cat, dog, and mouse (Todaro *et al*., 1974). The receptor specificities of PERV-A, PERV-B, and PERV-C were investigated with LacZ pseudotype vectors (Takeuchi *et al*., 1998). The results showed no cross-interference, implying that these Env proteins recognise different cell surface receptors. Also, host range analyses by the vector transduction assay showed that PERV-A and PERV-B viruses have wider host ranges, including several human cell lines, compared with PERV-C viruses which infected only two pig cell lines and one human cell line (Takeuchi *et al*., 1998). Recently, the *in vitro* host range of PERV was studied in human primary cells and cell lines, as well as in cell lines from non-human primates and other species. The analysis revealed that there were three distinct patterns of susceptibility to infection among these host cells. Firstly, some cells are resistant to

infection in the assay. Secondly, other cells are infected by virus but the cells are not permissive to productive replication and spread. Thirdly, the final category of cells is permissive for productive infection and spread (Wilson *et al*., 2000).

In March 1994, the transplantation research group at Westmead hospital in Sydney, Australia initiated an inbreeding program in a particular stock of feral pigs, as tools for transplantation research and in the long term as potential donors for xenotransplantation. The line is called Westran (Westmead transplantation) and pigs from it are the subject of the research reported here. Specifically, this Chapter describes a detailed study of PERV *env* sequences in the Westran pig line in relation to the potential hazards of these viruses for xenotransplantation.

4.2. Materials

4.2.1. Animal resources

The Westran line are believed to be descended from a pair of pigs released on Kangaroo Island, off the coast of South Australia, in 1803 by a French navigator and explorer, Captain Nicholas Baudin (Cooper, 1954). Captured feral pigs from Kangaroo Island were transferred to Adelaide University for biomedical research in 1976 (McIntosh and Pointon, 1981). After being maintained as a very small colony for about 15 years, a pair of full sibs was transferred to Westmead Hospital in Sydney for transplantation research. Since then, the core breeding line has been maintained by deliberate full-sib mating up to the current 8th generation. Assessment of genetic and immunological composition has been performed by a Westmead/Animal Science research team using ABO blood grouping and mixed lymphoctye reaction (MLR), lymphocytotoxity assay (LCA) and monitoring of hyperpolymorphic microsatellite markers. Selection during inbreeding has been based on body size, skin colour and ABO blood group. The surviving Westran pigs are smaller in size and are blood group O animals that are unreactive in MLR, implying SLA (Swine Leukocyte Antigen) class II identity, and have no reactivity to each other on LCA. Their highly inbred status has been confirmed by very high levels of microsatellite homozygosity (94% (47/50) of microsatellites are monomorphic at the 8th generation of full sib mating) (Hawthorne *et al*., 1999). Cross-over skin grafts also have been performed in Westran pigs. Littermate male-to-male, female-to-female and female-to-male skin grafts are accepted longterm without evidence of rejection. The male-to-female grafts are rejected by 7 days suggesting the presence of a Y-chromosome linked minor histocompatibility antigen (Dr Phillip J. O'Connell: Personal Communication). Figure 4.1 shows typical Westran pigs with black spots on white background and solid white coat colour. A solid white coat colour is more common than spotted in Westran pigs and is fixed in the core breeding line.

The origin of Westran pig ancestors was investigated with phylogenetic analysis of mitochondrial DNA D-loop sequence (Personal commonication: Kyu-Il Kim *et al*). Two Westran pigs from generation two (Animal no. 1) and three (Animal no. 11) of inbreeding showed 100% homology in the D-loop sequence. Surprisingly they clustered with Asian pigs (Fig. 4.2), suggesting an Asian origin.

Endogenous retroviruses have been analyzed in this Chapter from a generation-six boar (Animal no. 115).

(a)

(b)

Fig 4.1 Typical Westran pigs with black spots on white background (a) and solid white coat colour (b).

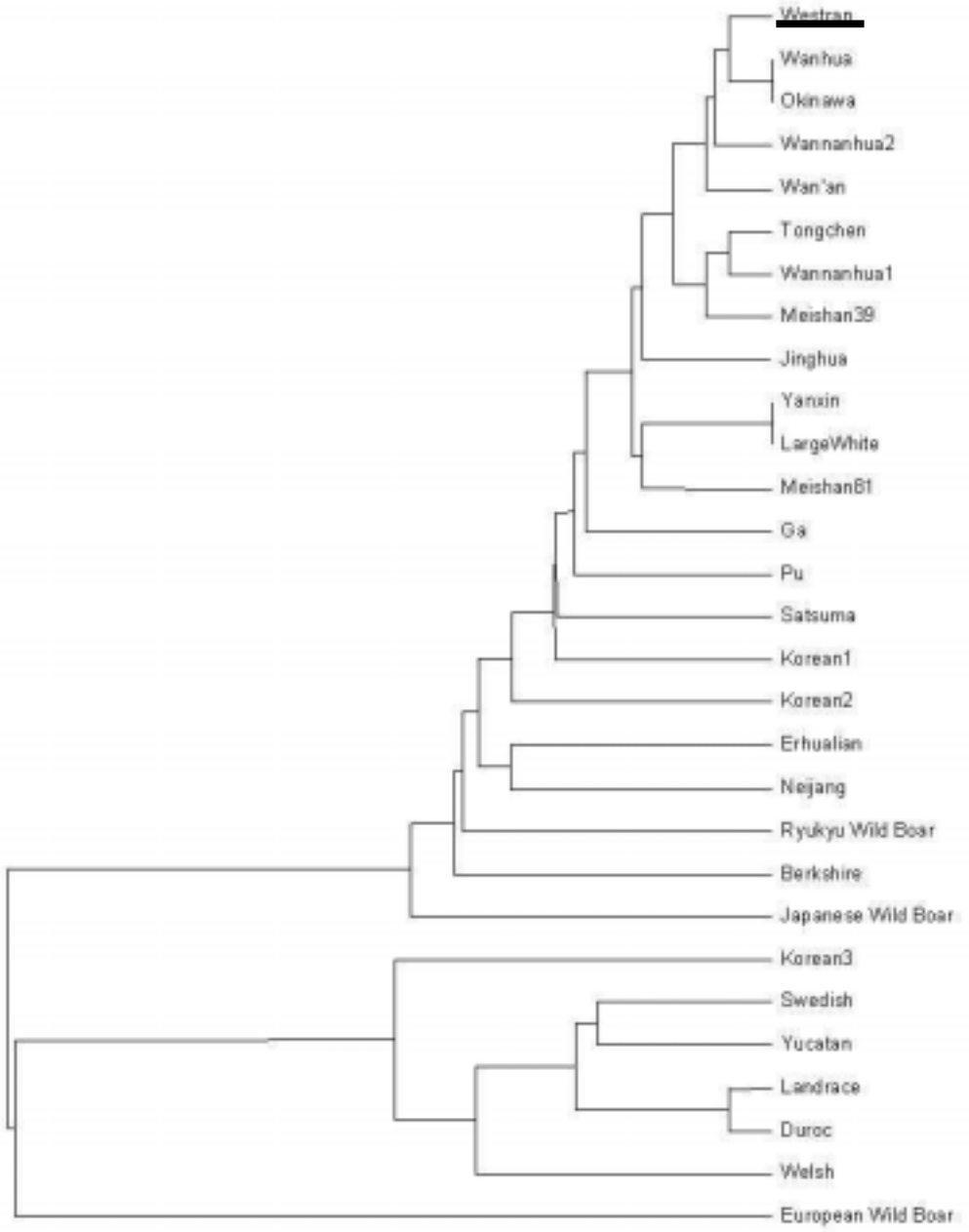


Fig 4.2 mtDNA phylogeny showing Westran sequence clustered with predominantly Asian pig breeds.

4.2.2. Primers

4.2.2.1. PERV PCR amplification

The primers used in this work were designed using the Primer Design program (Version 2.0, Scientific and Educational Software, 1991) as described in section 2.6. The primers are based on the sequences of Akiyoshi *et al*. (1998) for PERV-C and Le Tissier *et al*. (1997) for PERV-A and PERV-B. Primers were synthesised by Life Technologies Pty Limited, Victoria, Australia.

4.2.2.2. Primers for sequencing

A pair of M13 vector primers labelled with different infrared-sensitive dyes suitable for the Li-Cor sequencing system were used for cycle sequencing. Two M13 forward dye-labeled primers called –29 and –38 are available commercially from Li-Cor Inc. After testing, the –38 M13 forward primer proved more suitable. The primer sequences and annealing temperatures for the chosen pair of primers are presented in Table 4.1 (DNA sequencing manual, Li-Cor Inc).

Table 4.1 The pair of vector primers used for sequencing.

Name of primer	Primer sequences	Annealing Temp.
M13 Forward (-38)/ IRD800 Dye-labelled primer	5'-TTTCCCAGTCACGACGTTG-3'	50 °C
M13 Reverse/ IRD700 Dye-labelled primer	5'-GGATAACAATTTCACACAGG-3'	50 °C

4.3. Methods

4.3.1. PCR conditions

4.3.1.1. PCR using *Taq* DNA polymerase

PCR was carried out in a 50 µl volume with approximately 100 ng of template genomic DNA, 0.2 mM of each primer, 600 µM of each dNTP, 1 or 2 mM $MgCl_2$, 5 µl 10x PCR buffer (Invitrogen), and 2.5 units of *Taq* polymerase. This mixture was overlaid with one drop of mineral oil. Amplification was performed in a PTC-100 Programmable Thermal Controller (MJ Research Inc.) for 45 cycles with denaturation for 1 minute at 94 °C, annealing for 1 minute at 65 °C (PERV-A and -B) or 58 °C (PERV-C) and extension for 1 minute 72 °C. The final extension was for 20 minutes at 72 °C.

4.3.1.2. PCR using *Pfu* proof-reading DNA polymerase

The *Pfu* proof-reading DNA polymerase was chosen for the second set of PERV PCR reactions because its known lower error rate compared with *Taq* DNA polymerase. PCR using *Pfu* proof-reading DNA polymerase was carried out in a 50 µl volume with approximately 100 ng of template genomic DNA, 0.2 mM of each primer, 0.5 mM of each dNTP, 10 × PCR buffer (Stratagene), and 2.5 units of cloned *Pfu* polymerase (Stratagene). This mixture was overlaid with one drop of mineral oil. Amplification was performed in a MJ thermocycler with initial denaturation for 3 minutes at 95 °C, followed by 35 cycles of denaturation for 1 minute at 95°C, annealing for 1 minute at 65°C and extension for 5 minutes at 72°C. The final extension was for 10 minutes at 72°C.

4.3.1.3. Cycle sequencing reaction

Cycle sequencing was performed using a PTC-100 Programmable Thermal Controller (MJ Research Inc.) with the pair of M13 vector primers shown in Table 4.1, and then a Li-Cor sequencer, as described in Section 2.10.2.

4.3.2. Cloning of PERV PCR products

The *env* PCR products were cloned into pCR®-Blunt plasmid vector (Invitrogen) for *Pfu* polymerase generated PCR products and pCR®2.1-TOPO plasmid vector (Invitrogen) for *Taq* polymerase generated PCR products. The detailed information for cloning is described in Section 2.11. Detailed maps of the cloning vectors are presented in Appendices 2.1 and 2.2.

4.3.3. Checking clone inserts by PCR and restriction digestion

To check that plasmid clones contained a correct insert, boiled preparation was tested as described in section 2.7.3, followed by restriction digestion. PCR was performed with *Taq* DNA polymerase as described in Section 4.3.1.1. Two restriction enzymes, *Kpn*I and *Mbo*I, were used for preliminary screening of the clones for characteristic features of PERV-A and -B (Le Tissier *et al*. 1997). The detailed information for restriction enzyme digestion is described in Section 2.8.

4.3.4. Extracting plasmid DNA from bacterial cells

The plasmid DNA was extracted using a JETquick Miniprep Spin Kit as described in Section 2.4.2.

4.3.5. Sequencing of PERV clones

A *SequiTherm* EXCEL™ Long-Read™ DNA sequencing kit (Epicentre) and Li-Cor sequencer (Model 4200, Li-Cor Inc.) were used to read approximately 1.8kb of full-length insert sequences, using the pair of vector primers (Li-Cor Inc.) shown in Table 4.1. Sequences were analyzed using Base ImageIR™ software version 4.1 (Li-Cor Inc.). The detailed protocols for sequencing using the Li-Cor sequencer are described in Section 2.10.2. Figure 4.3 shows the schematic diagram of the sequencing.

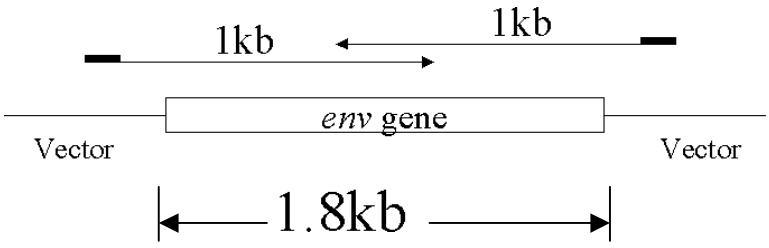


Fig 4.3 Schematic representation of single-pass sequencing with a pair of vector primers (Solid bars (—)).

4.3.6. Orientation and amalgamation of sequences

SeqEd software version 1.0.3 (Perkin-Elmer, Applied Biosystems) was used to amalgamate and orient the sequences with respect to published PERV sequences. Full sequences of the PERV *env* PCR products were assembled by overlapping forward and reverse sequencing products. The overlapping sequences at each end of the long reads may compensate for the less accurate reads at the ends.

4.3.7. Sequence alignment and translation

Alignment of the PERV *env* sequences was performed using the Clustalw and Pileup programs of the multiple sequence alignment option in GCG using the Web ANGIS interface

(http://www.angis.org.au/). The putative amino acid sequences were determined using the Translate program in GCG.

4.3.8. Phylogenetic analysis

Phylogenetic analysis of PERV *env* sequences was performed via PAUP software version 4 beta 3a for PPC (Swofford, 1999), using the UPGMA (unweighted pair group method using arithmetic average) method. The UPGMA method assumes that the rate of substitution is constant and distances are linear with time.

4.4. Results

4.4.1. Designing primers

The published PERV-A and -B *env* sequences (Le Tissier *et al*. 1997) were aligned using the GAP program available in WebANGIS (Appendix 4.1). The *env* genes of PERV-A and PERV-B are highly conserved at both the 5' and 3' ends, which flank a distinct region which forms the basis for recognising the A and B variants. The primers were chosen from the highly conserved regions in order to search for novel variants in the less conserved region and to amplify all possible PERV inserts in the pig genome (Fig 4.4). The primer information is presented in Table 4.2 and the sequence alignment between PERV-A and PERV-B is presented in Appendix 4.1.

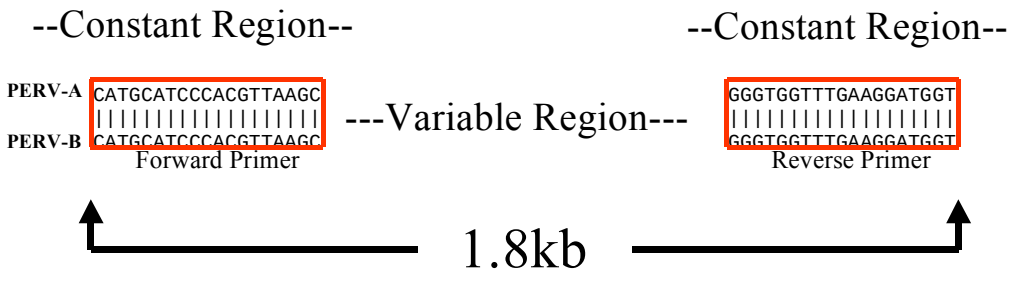


Figure 4.4 Strategy for PERV PCR amplification using consensus PERV-A and PERV-B primers from the constant regions located at the 5' and 3' ends of the *env* gene.

Table 4.2 The primer sequences, PCR conditions ($MgCl_2$ concentration and annealing temperature) and expected product size for amplifying the PERV *env* sequences.

Primer sequences[1]	$MgCl_2$ Conc.	Annealing Temp.	Expected size (bp)
F. 5'-CATGCATCCCACGTTAAGC-3' R. 5'-ACCATCCTTCAAACCACCC-3'	2 mM	65 °C	1784 bp (PERV-A) 1775 bp (PERV-B)
F. 5'-ATTCCTAGTCTGACCTCACC-3' R. 5'-TGTGAGTAACAGGAGGAGGA-3'	1 mM	58 °C	1515 bp (PERV-C)

[1]F: forward primer; R: reverse primer

4.4.2. PCR for PERV-C amplification

It was initially hoped that the PERV-A and -B consensus primers would also amplify PERV-C inserts and possibly even new categories of PERV. However analysis of the *env* sequences showed that no PERV-C insert amplified. Therefore a new set of PERV-C specific primers (Table 4.2) were designed based on the sequence of Akiyoshi *et al.* (1998). The PCR results for PERV-C in a Westran pig are presented in Fig 4.5, clearly showing that PERV-C is present in the Westran line of pigs but not in the Large White control DNA sample used.

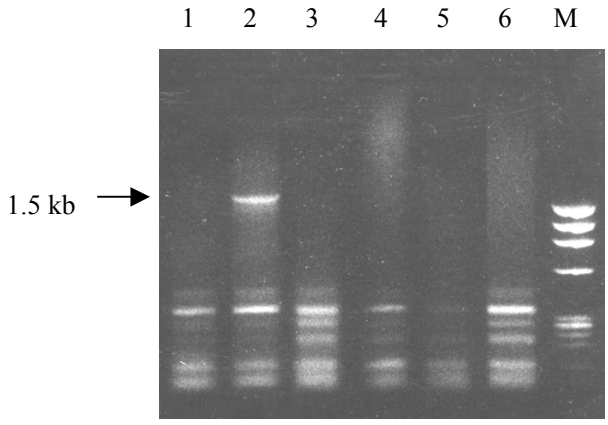


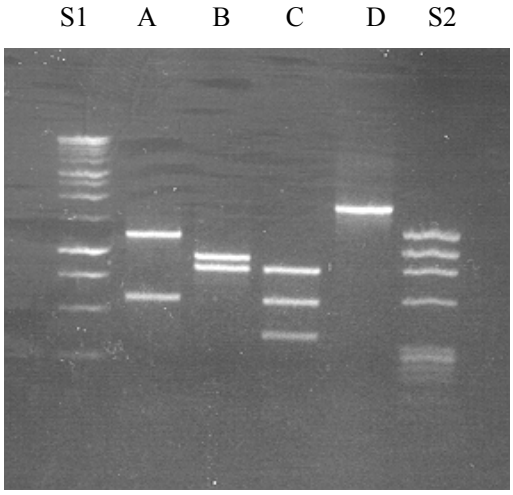
Figure 4.5 1.5 kb PERV-C PCR product in a Westran pig (indicated by an arrow). Lanes 1, 3 and 5 have Large White genomic DNA as template. Lanes 2, 4 and 6 have genomic DNA from Westran pig No 115 as template. Different $MgCl_2$ concentrations were tested (lanes 1 and 2: 1 mM, lane 3 and 4: 2 mM, lane 5 and 6: 3 mM). M is ϕX174/*Hae*III Marker (Promega) size standards. The only combination to give the expected size product is Westran genomic template with 1 mM $MgCl_2$. PERV-C is clearly present in the Westran inbred line but amplification of PERV-C envelope sequence is sensitive to $MgCl_2$ concentration.

4.4.3. Restriction enzyme digestion for screening PERV clones

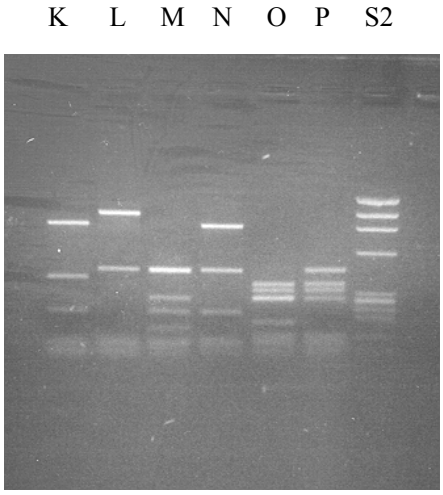
Cloned PERV-A and -B *env* PCR products, amplified by *Taq* DNA Polymerase (64 clones) and *Pfu* proof-reading DNA polymerase (346 clones), were screened by *Kpn*I and *Mbo*I restriction enzymes. Four patterns of *Kpn*I digestion (Fig 4.6(a)) and six patterns of *Mbo*I digestion (Fig 4.6(b)) were identified in PCR product amplified by *Taq* DNA polymerase. Each of the *Kpn*I/*Pfu* patterns corresponded to a *Kpn*I/*Taq* pattern (A, B and D). Four of the *Mbo*I/*Pfu* patterns corresponded to *Mbo*I/*Taq* patterns (L, M, N and P). Three patterns of *Kpn*I digestion (Fig 4.6(c)) and twelve patterns of *Mbo*I digestion (Fig 4.6(d)) were identified in PCR product amplified by the *Pfu* proof-reading DNA polymerase. Combining the results for the two restriction enzymes, nine types of clone were recognizable for *Taq* DNA polymerase amplified clones and thirteen types were recognizable for *Pfu* proof-reading DNA polymerase amplified clones (Table 4.3).

The *Kpn*I restriction digestion pattern A is characteristic of PERV-A and pattern B of PERV-B, based on the published PERV-A and PERV-B sequences (Le Tissier *et al*., 1997). Digestion patterns AM, AN, AP and BL were found for clones generated with both *Taq* and *Pfu* polymerases. On the other hand, restriction digestion patterns AK, AO, BN, CL, DL were found only in the clones generated by *Taq* DNA polymerase and restriction enzyme pattern AV, AX, AQ, AR, AS, AT, AU, AW, DU showed only in the clones generated using *Pfu* DNA polymerase. These unique clones constitute 8 % (5/64) for the *Taq* DNA polymerase clones and 5 % (18/346) for the *Pfu* DNA polymerase amplified clones. Among both *Taq* and *Pfu* amplified clones, the BL type is predominant, constituting 64 % (41/64) of the *Taq* polymerase amplified clones and 72 % (248/346) of the *Pfu* amplified clones (Table 4.3). All restriction enzyme patterns of all clones are presented in Appendix 4.2.

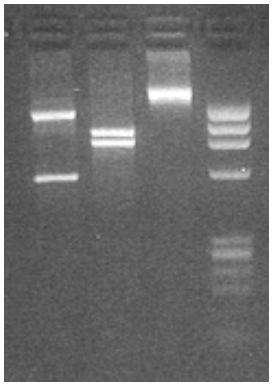
(a) *Kpn*I, *Taq* DNA polymerase



(b) *Mbo*I, *Taq* DNA polymerase



(c) *Kpn*I, *Pfu* DNA polymerase



(d) *Mbo*I, *Pfu* DNA polymerase

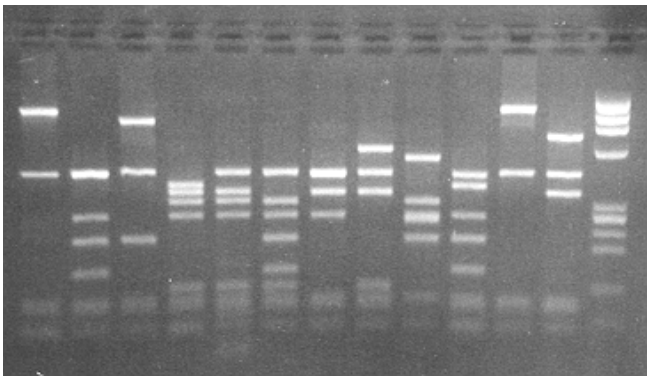


Fig 4.6 Restriction digestion patterns of PERV clones. (a), (b): Four patterns (A, B, C, D) of *Kpn*I digestion and six patterns (K, L, M, N, O, P) of *Mbo*I digestion were identified in PCR product amplified by *Taq* DNA polymerase. (c), (d): Three patterns (A, B, D) of *Kpn*I digestion and twelve patterns (L, M, N, X, P, Q, R, S, T, U, V, W) of *Mbo*I digestion were identified in PCR product amplified by *Pfu* proof-reading DNA polymerase. S1 and S2 are the 1kb ladder and ϕX174/*Hae*III Marker (Promega) size standards, respectively.

Table 4.3 Restriction enzyme digestion and sequence analysis of PERV clones.

Polymerase used	Digestion pattern	Number*	No of clones sequenced (IDs)	Most likely type from sequence comparison	Premature stop codon (IDs)
Taq	AK	1	1 (1)	A+B	Yes (1)
	AM	10	3 (6, 10, 11)	A	Yes (6, 10, 11)
	AN	4	2 (8, 82)	A	Yes (8, 82)
	AO	1	1 (17)	A	Yes (17)
	AP	4	2 (54, 56)	A	Yes (54)
	BL	41	6 (3, 4, 9, 12, 14, 20)	B	Yes (3, 12)
	BN	1	1 (24)	B	No
	CL	1	1 (21)	A+B	No
	DL	1	1 (29)	B	Yes (29)
Pfu	AV	2	1 (251)	A	No
	AM	15	1 (6)	A	No
	AN	17	1 (11)	A	No
	AX	1	1 (260)	A+B	Yes (260)
	AP	48	1 (3)	A	No
	AQ	5	1 (56)	A	Yes (56)
	AR	2	1 (62)	A	No
	AS	4	1 (112)	A	Yes (112)
	AT	1	1 (115)	A	Yes (115)
	AU	1	1 (232)	A+B	Yes (232)
	AW	1	1 (295)	A	Yes (295)
	BL	248	1 (1)	B	No
	DU	1	1 (345)	A+B	No

*number of clones with this digestion pattern

4.4.4. Sequences of PERV clones

4.4.4.1. Sequence similarities of PERV clones

After consideration of the restriction digestion patterns and to ensure the widest possible range of clone types were included, 18 *Taq* amplified clones and 13 *Pfu* proof-reading amplified clones were sequenced (Table 4.3). Table 4.4 summarises the sequence differences among PERV clones and published PERV-A (EMBL Nucleotide Sequence Database accession number Y12238), PERV-B (EMBL Nucleotide Sequence Database accession number Y12239) and

PERV-C (GenBank accession number AF038600) published sequences (Le Tissier *et al*., 1997; Akiyoshi *et al*., 1998).

As noted previously, the PERV-A and PERV-B consensus envelope primers amplified 9 different types of *Taq* polymerase generated PERV clones and 13 types of *Pfu* polymerase generated clones recognisable by screening with two restriction enzymes. Of course any PERVs differing by substantial substitution or deletion in the primer complementary region could not be amplified by these primers. Thus there is a limit to the variety of novel PERV insertions which could be characterised in this way.

To determine the envelope types, each clone was initially aligned with the published PERV-A and PERV-B sequences. Seventeen clones were very similar to the PERV-A sequence, differing from it by only 44 to 54 bases. They are designated as PERV-A clones. Nine clones were very similar to the PERV-B sequences, differing from it by only 1 to 15 bases. These are designated as PERV-B clones. The remaining five clones differed from both PERV-A and PERV-B by at least 94 bases. Sequence comparison (Appendix 4.3) showed that these five clones are actually recombinants between PERV-A and PERV-B. Comparisons amongst all sequences are colour-coded in Table 4.4. Excluding recombinant clones, the absolute number of nucleotide differences among PERV-A clones is between 5 to 54 base pairs in 1,785 base pairs. For PERV-B clones, there were between 1 to 24 base pair differences in 1,776 base pairs. There are about 400 base pairs different between PERV-A group and PERV-B groups (Table 4.4).

Table 4.4 Absolute difference matrix of 31 PERV sequences and PERV-A, PERV-B, and PERV-C published sequences (number of nucleotides by which pairs of sequences differ). The ID of unpublished sequences is expressed as *polymerase*-x, where *polymerase* is either *Taq* or *Pfu* and x is the number of the clone generated by that polymerase. The comparisons among PERV-A clones (green), PERV-B clones (blue), between PERV-A and PERV-B clones (black), and between recombinants and all other clones (red) are indicated by different font colours.

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
		PERV-B	Taq-14	Taq-4	Pfu-1	Taq-20	Taq-12	Taq-24	Taq-3	Taq-9	Taq-29	Taq-21	Taq-11	Taq-6	Pfu-6	Taq-10	Taq-8	Taq-82	Pfu-56	Pfu-11	Pfu-251	Pfu-112	Pfu-295	Pfu-3	Taq-56	Taq-54	Pfu-62	Taq-17	Pfu-115	PERV-A	Pfu-232	Taq-1	Pfu-260	Pfu-345	PERV-C
1	PERV-B	-																																	
2	Taq-14	1	-																																
3	Taq-4	2	3	-																															
4	Pfu-1	3	4	5	-																														
5	Taq-20	3	4	5	6	-																													
6	Taq-12	5	6	7	6	8	-																												
7	Taq-24	7	8	7	10	10	12	-																											
8	Taq-3	15	16	17	18	18	20	22	-																										
9	Taq-9	10	11	10	13	13	15	13	11	-																									
10	Taq-29	9	10	11	12	10	14	16	24	19	-																								
11	Taq-21	97	98	99	100	100	102	102	108	103	106	-																							
12	Taq-11	395	396	395	397	394	400	398	399	399	400	348	-																						
13	Taq-6	396	397	396	398	395	401	399	400	400	401	349	6	-																					
14	Pfu-6	390	391	390	392	389	395	393	394	394	395	345	11	13	-																				
15	Taq-10	393	394	393	395	392	398	396	397	397	398	347	18	20	11	-																			
16	Taq-8	400	401	400	402	399	405	404	404	404	403	352	28	30	29	36	-																		
17	Taq-82	403	404	403	405	402	408	407	407	407	406	355	29	31	34	41	5	-																	
18	Pfu-56	400	401	400	402	399	405	404	404	404	405	352	30	32	35	42	22	23	-																
19	Pfu-11	397	398	397	399	396	402	401	401	401	400	350	34	36	27	24	12	17	34	-															
20	Pfu-251	396	397	396	398	395	401	400	400	400	399	349	39	40	34	29	22	25	42	10	-														
21	Pfu-112	393	394	393	395	392	398	397	397	397	396	346	41	42	34	29	32	37	32	20	12	-													
22	Pfu-295	389	390	389	391	388	394	392	393	393	392	342	29	30	24	21	32	35	40	22	12	12	-												
23	Pfu-3	393	394	393	395	392	398	397	397	397	398	349	44	46	38	32	41	46	29	30	22	9	21	-											
24	Taq-56	396	397	396	398	395	401	400	402	400	401	349	47	48	40	35	45	50	29	34	26	14	26	12	-										
25	Taq-54	394	395	394	396	393	399	398	398	398	399	347	47	48	40	35	46	51	30	34	26	14	26	11	10	-									
26	Pfu-62	390	391	390	392	389	395	393	394	394	395	343	34	35	27	22	41	46	35	29	21	17	15	17	15	15	-								
27	Taq-17	401	402	401	403	400	406	405	405	405	404	356	36	37	39	44	35	36	21	41	31	21	29	23	23	23	26	-							
28	Pfu-115	401	402	401	403	400	406	405	405	405	404	354	33	36	25	31	31	36	27	28	35	28	36	29	31	30	34	30	-						
29	PERV-A	394	395	394	396	393	399	398	400	398	399	341	49	51	45	47	48	53	50	45	47	44	46	48	49	48	48	54	49	-					
30	Pfu-232	337	338	337	339	336	342	340	344	341	342	340	62	65	60	63	76	81	82	77	84	83	73	86	90	91	77	91	81	94	-				
31	Taq-1	344	345	344	346	343	349	348	351	348	347	349	88	91	82	85	64	69	86	65	75	84	84	93	96	98	90	96	87	103	26	-			
32	Pfu-260	339	340	339	341	338	344	343	346	343	344	344	95	97	89	88	94	99	76	89	81	68	78	65	63	64	68	70	80	99	35	42	-		
33	Pfu-345	287	288	287	289	286	292	290	296	293	294	380	116	119	110	113	124	129	132	123	133	134	124	134	138	140	126	142	133	143	59	71	85	-	
34	PERV-C	441	442	443	443	442	446	445	440	443	448	346	245	247	241	243	249	254	248	247	248	246	244	246	248	244	240	253	247	222	274	287	280	318	-

4.4.4.2. Empirical cycle-sequencing error rate, using the Li-Cor sequencer

Because single-pass sequencing reaction was applied in this experiment except for the middle of the sequence where there was a short overlap, sequencing errors are to be expected. After 600 to 800 base pairs of a single long sequence read, the accuracy is expected to decline. To determine the error rate of cycle sequencing using the Li-Cor sequencer, a sample of four clones was sequenced twice under exactly the same experimental conditions and editing procedures. That is, for each of four clones, two forward and two reverse products were generated and sequenced. Table 4.5 shows the number of nucleotides by which the sequence of the two products of each primer differed. For the four clones sequenced twice, there was an error rate of 0.4% (28/7020). This error rate is based on comparison of two sequences, each of which are equally likely to have errors. Therefore the true error rate is 0.2% (0.4/2). This sequencing accuracy is better than the 99% rate indicated by the company (IR^2 system specification, Li-Cor Inc.).

Table 4.5 The empirical error rate of cycle sequencing using the Li-Cor sequencer estimated from duplicate sequencing of the same clones

Clone	Vector primer used	Differences
Taq-11	M13 Forward	5/900
	M13 Reverse	7/900
Taq-54	M13 Forward	4/900
	M13 Reverse	3/650
Taq-56	M13 Forward	3/1000
	M13 Reverse	4/900
Taq-6	M13 Forward	1/900
	M13 Reverse	1/870
Total		28/7020

4.4.4.3. Frame shift mutations and premature stop codons

The nucleotide sequence alignments of the 17 PERV-A clones, 9 PERV-B clones, and 5 recombinant clones are presented in Appendix 4.3. The alignment of deduced amino acid sequences of the PERV clones is presented in Appendix 4.4 together with PERV-A, PERV-B, and PERV-C published sequences. 55% of the clones (17/31) sequenced have premature stop codons within the envelope protein-encoding region, which would prevent the retrovirus from making full length envelope protein recognizable by cell surface receptor for the virus. The positions of premature stop codons and their causes are listed in Table 4.6. Fourteen stops were caused by frame-shift mutation and only three were caused by base substitutions. A hot spot for frame-shift mutations was found at nucleotide position 1134, with ten of fourteen found at this

position. Three clones, namely *Taq*-10, *Taq*-11 and *Pfu*-232, have two frame-shift mutations. An example of a premature stop codon caused by a deletion is illustrated for clone *Taq*-29 in Fig 4.7.

Table 4.6 Sequence analyses of 17 clones with premature stop codons.

Clone ID	Stop codon position[1]	Causes of stop codon[2]
Taq-1	416	ΔG nt 1134 (Frame shift mutation)
Taq-3	578	Base substitution C1733T
Taq-6	467	ΔA nt 1250 (Frame shift mutation)
Taq-8	416	ΔG nt 1134 (Frame shift mutation)
Taq-10	330	ΔC nt 815 (Frame shift mutation), ΔG nt 1134 (Frame shift mutation)
Taq-11	428	ΔG nt 1134, ΔA nt 1250 (Frame shift mutation)
Taq-12	242	Base substitution A725T
Taq-17	394	Base substitution C1184T
Taq-29	373	ΔA nt 1042 (Frame shift mutation)
Taq-54	416	ΔG nt 1134 (Frame shift mutation)
Taq-82	416	ΔG nt 1134 (Frame shift mutation)
Pfu-56	416	ΔG nt 1134 (Frame shift mutation)
Pfu-112	416	ΔG nt 1134 (Frame shift mutation)
Pfu-115	511	Base substitution A1532T
Pfu-232	438	ΔG nt 1134, ΔA nt 1250 (Frame shift mutation)
Pfu-260	416	ΔG nt 1134 (Frame shift mutation)
Pfu-295	416	ΔG nt 1134 (Frame shift mutation)

[1]The position of stop codon is in amino acid sequence
[2]The position of mutation in nucleotide sequence. The number of nucleotide starts from the 5' forward primer sequence as nt number 1

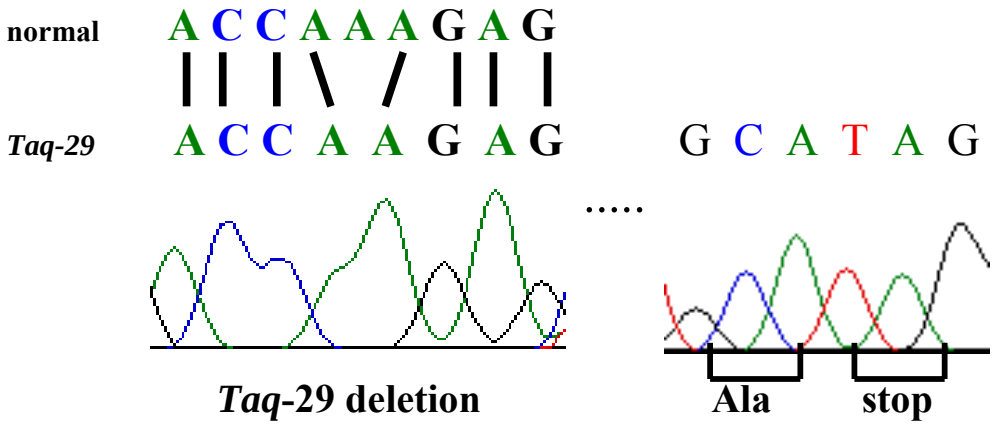


Fig 4.7 Example of premature stop codon in clone *Taq*–29 caused by an upstream (adenine) deletion giving a frame shift mutation.

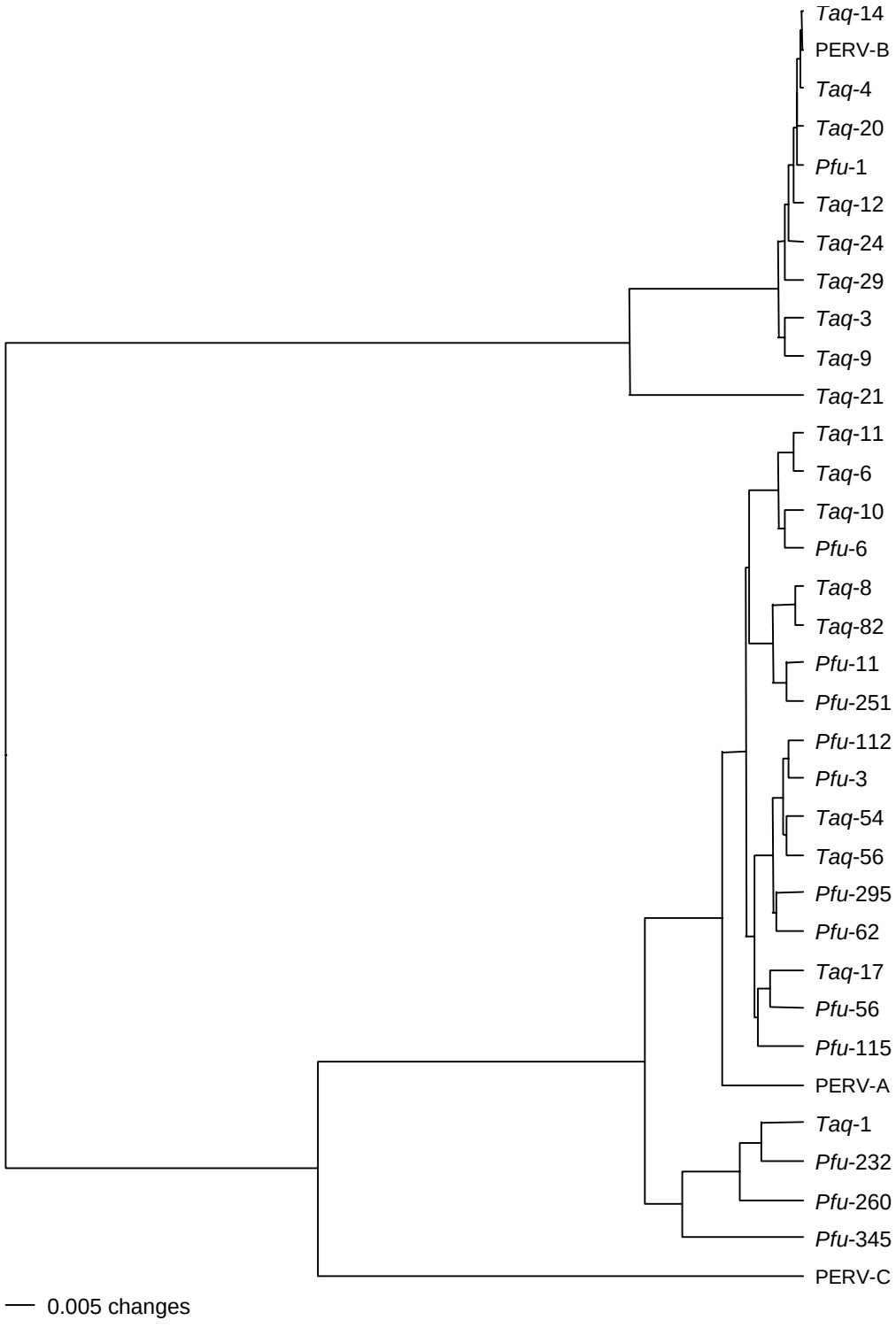


Fig 4.8 Phylogenetic tree of 18 *Taq* and 13 *Pfu* generated clones as well as PERV-A, PERV-B and PERV-C published sequences.

4.4.5. Phylogenetic analysis

A dendrogram summarising the results of phylogenetic analysis of all 34 PERV *env* sequences considered in this study, including the PERV-A, -B, -C published sequences, is presented in Fig 4.8. 21 sequences cluster with PERV-A and 10 with PERV-B. PERV-C is more closely related to PERV-A than PERV-B. The absolute number of nucleotide sequence differences between PERV-A and PERV-C is 222 base pairs and between PERV-B and PERV-C is 441 base pairs (Table 4.4). The recombinant five clones, namely *Taq*-1, *Taq*-21, *Pfu*-232, *Pfu*-260, *Pfu*-345, do not show the same pattern of clustering as the other clones, having more distant nodes relative to the published sequences.

4.4.6. Recombinant clones between PERV-A and PERV-B

The recombinant clones were classified into three groups based on the patterns of break points between PERV-A and –B sequences (Fig 4.9). In the case of *Taq*-1, *Pfu*-232 and *Pfu*-260, two different polymerases gave the same break points in these recombinant clones suggesting that at least one of these recombinant sequences reflect genuine PERV sequence in the pig genome and are not artefacts of the PCR process. The clone *Taq*-21 shows double recombination between PERV-A and PERV-B. The sequence alignments and position of recombination are presented in Appendix 4.3. (c).

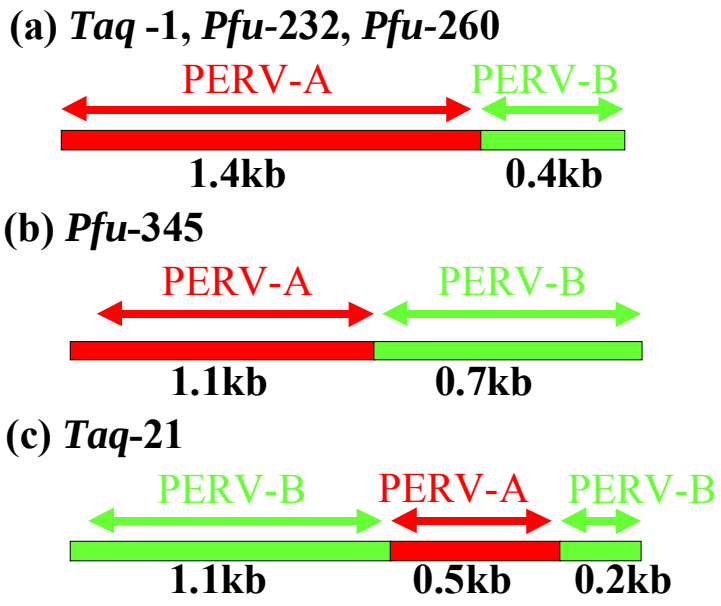


Fig 4.9 Three types of recombinant envelope gene sequence between PERV-A and PERV-B.

4.5. Discussion

This chapter reports studies of PERV *env* sequences obtained from an inbred pig from the Westran line, which is being investigated as a source of tissue for transplantation into humans. A total of thirty-one 1.8 kb PERV *env* clones, generated by either *Taq* and *Pfu* DNA polymerase, were fully sequenced.

About 30% of the 31 envelope sequences described in the Chapter are classified as PERV-A, whether from *Taq* or *Pfu* Polymerase amplified clones. To check whether the number of clones generated in each class is significantly different with *Taq* and *Pfu* polymerase, chi-squire contingency analysis was performed ($\chi^2 = 2.508$, df=2, not significant; for details see Appendix 4.5). The results indicated that the frequency of A, B and recombinant clones is not significantly different between the clones generated by *Taq* and *Pfu*. However, the clones sequenced in this study are a non-random sample because selection of clones for sequencing was based on restriction enzyme digestion results. The restriction digestion results for 410 randomly chosen clones (Table 4.3) show that 27.8 % (114/410) are PERV-A, 71 % (291/410) are PERV-B and 1.2 % (3/410) can be classified as recombinant clones between PERV-A and PERV-B, providing a more accurate representation of the relative occurrence of these three classes. In any case, it is important to realise that the sequences presented here do not necessarily reflect the relative number of PERV-A and –B inserts in Westran pigs. In the next Chapter of this thesis, FISH provides a much more accurate idea of the number of PERV-A and –B inserts in Westran pigs, subject to some possible hybridisation of PERV-A and PERV-B probes with each other or with PERV-C inserts.

Initially *Taq* DNA polymerase was used to generate PCR products for cloning. However due to a substantial level of minor sequence differences between clones, which might have been artefacts of the inability of *Taq* to correct errors during DNA replication, a proof-reading polymerase, *Pfu,* was then tried. *Taq* DNA polymerase has no 3'-proof reading exonuclease activity and a relatively high misincorporation of bases is expected. The usually quoted figure for rate of misincorporation using *Taq* polymerase is about $2x\ 10^{-4}$ per nucleotide (Watson, 1992; Lewin, 1994). However Cantor and Smith (1999) have shown that the overall misincorporation depends not only on the rate of misincorporation of the *Taq* polymerase but also on the PCR reaction conditions and the length of DNA amplified. The number of misincorporations observed in a particular product in a particular set of PCR conditions is predicted by:

Number of misincorporations = LnX_{mis}……..(equation 1)

where L is the length of DNA amplified, n is the number of the PCR cycles, and X_{mis} is the fractional misincorporation rate per nucleotide per PCR cycle (Cantor and Smith, 1999). *Taq* DNA polymerease fidelity has been measured using a PCR based forward mutation assay based on the *lacI* target gene. This provided an estimated fractional misincorporation rate of 8.0×10^{-6} per nucleotide per duplication (Instruction manual for *Pfu* DNA polymerase, Stratagene). *Pfu* has a lower fractional misincorporation rate of 1.3×10^{-6} per nucleotide per duplication, based on the same *lacI* based fidelity assay (Instruction manual for *Pfu* DNA polymerase, Stratagene). Applying these fractional misincorporation rates and the equation to the PERV amplifications (1800 bp product, 45 cycles (*Taq*) or 35 cycles (*Pfu*) of PCR), equation 1 predicts the misincorporation for *Taq* DNA polymerase to be 0.65 nucleotide and for *Pfu* DNA polymerase to be 0.08 nucleotide. Thus two *Taq* clones on average would be expected to differ by 1.3 nucleotides, a *Taq* clone sequence versus a *Pfu* clone sequence by 0.73 nucleotides and a pair of *Pfu* clones by 0.16 nucleotides if misincorporation during PCR was the only source of difference between them. We can conclude that comparisons of *Pfu* clone sequences with other *Pfu* sequences or *Taq* clone sequences are very unlikely to show artefactual differences due to PCR. Even the comparison of *Taq* sequences with other *Taq* sequences are unlikely to be seriously affected by amplification artefacts, as only about one nucleotide difference is expected on average.

Sequencing errors are an additional more important source of artefactual sequence differences between clones. Financial constraints meant that only single-pass sequencing using Li-Cor long-range sequencing was possible. Double pass sequencing would have resolved many ambiguities but it was beyond the resources available to this project to do so. An empirical single-pass cycle-sequence error rate for the Li-Cor sequencer was also calculated in this study by cycle-sequencing four PERV clones in duplicate. Each clone was sequenced bidirectionally providing 8 duplicate sequence comparisons in size from 650 to 1000 bp. The number of discrepancies ranged from a low of 1/900 to a high of 7/900. The data were tested for heterogeneity (Appendix 4.6) by Chi-square analysis (χ^2 = 7.855, df=7, p=0.347). The non-significant result justifies pooling of the data to obtain an estimate of the sequence comparison error rate, p of 0.00398860 (28/7,020). Using this observed estimate of p, the cumulative frequency distribution of errors for a comparison of two sequences 1800 bp long (n=1800) was calculated in Minitab (Minitab Inc., Version 10 Xtra) using the exact Binomial distribution (Table 4.7). This table enables interpretation of the results of comparison of all clones shown in Table 4.4. The probability of getting a comparison differing by more than 12 nucleotides due to sequencing errors alone is 3.18% (ie less than the 5% significance threshold). If the difference is more than 16 nucleotides the probability is 0.12%. By choosing an appropriate threshold stringency, it is possible to reliably distinguish between those clones in Table 4.4 showing

genuine genetic differences and those where genetic and artefactual differences cannot be reliably distinguished. Applying this threshold, at least 12 PERV-A, 5 PERV-B and 5 recombinant clones are distinct sequences due to biological differences or PCR introduced errors.

Table 4.7 The Binomial cumulative frequency distribution for comparison error in cycle sequencing for a 1800 bp sequence comparison (p=0.00398860, n=1800).

Error number (x)	Probability($X \leq x$)
11	0.9385
12	0.9682
13	0.9846
14	0.9930
15	0.9970
16	0.9988
17	0.9995
18	0.9998
19	0.9999

Toward the end of this study, PERV-C specific primers were tested in Westran pigs and amplified the expected size PCR product in several trials. However, these PERV-C PCR amplifications were unreliable, suggesting that there may be mismatch(es) between the PERV-C primers and template. Therefore a new set of PERV-C primers may be needed for efficient PCR amplification and further analysis of this class of PERV. Nevertheless the results indicate that the Westran line is PERV-C positive. The copy number of PERV-C has been estimated as 8 to 15 copies per genome for specific strains of inbred and outbred miniature pig (Akiyoshi *et al*., 1998). There is evidence of variable PERV-C copy number in other pigs. Some pigs have even been found to be negative for PERV-C (Herring *et al*., 1999; Bösch *et al*., 1999). Even though PERV-C has a narrower host range compared to PERV-A and PERV-B (Takeuchi *et al*., 1998), it will be important to further characterise PERV-C in Westran pigs if they are ever to be used for xenotransplantation.

The sequencing results suggest that a large proportion of PERVs (17/31) are defective due to premature stop codons in the envelope gene. Indeed, eleven clones have a frame-shift mutation caused by a deletion of the same nucleotide at position 1135. While some of these clones may simply reflect resampling of the same PERV insertion from the PCR product, it is unlikely that all 11 are. Thus this site may be a hot spot for deletion or reflect independent insertions of PERVs with the same deletion during the evolutionary history of the pig. Bebenek *et al*. (1989) indicated that the reverse transcriptase of HIV-1 is relatively error prone and that errors are non-randomly distributed. They found base substitution and one-base frame shift mutational hot spots

mainly due to the template-primer slippage. Subject to further verification, these results are encouraging as they indicate that a substantial proportion of the PERVs in the Westran line constitute little potential hazard in xenotransplantation as they are non-functional.

There are a few strategies to identify all potentially active PERVs in the porcine genome. First, screening a genomic library for intact PERVs could chive this. Secondly, a cDNA approach could be applied to identify actively transcribed PERVs. Thirdly, an immunological approach can be used to identify expressed viruses. To eliminate these active PERVs for xenotransplantation, either knockouts via homologous recombination or breeding strategies could be considered. Insertional mutation by homologous recombination would knockout the gene. However it would be difficult or impossible to knockout all active PERVs because there are too many integration sites over the chromosomes. Further insertionally inactivated PERVs could regain infectivity through recombination. A breeding strategy also has obstacles. Due to the large number of integration sites, very (impossibly) large numbers of animals would have to be bred to identity the small number of animals lacking PERV sites. At the end of the breeding program, there would still be some constant PERV sites would present and the constant possibility of re-infection.

Retroviruses package two complete viral genomic RNAs in each virion and this specific configuration facilitates recombination. Recombination between retrovirus genomes has been demonstrated during mixed infection with genetically marked avian tumor viruses (Blair *et al*., 1976; Kawai and Hanafusa, 1972; Vogt, 1971; Wyke and Beamand, 1979; Wyke *et al*., 1975), murine leukaemia viruses (Faller and Hopkins, 1978; Wong and McCarter, 1973) and human retroviruses (Clavel *et al*., 1989). Also, exogenous viruses can recombine with endogenous retroviral sequences (Elder *et al*., 1977; Stephenson *et al*., 1974; Weiss *et al*., 1973). Recently, new recombinants between species have been discovered. For example, BaEV (baboon endogenous retrovirus) is a recombinant retrovirus containing type C *gag-pol* genes and type C *env* gene which probably arose by recombination of two primate viruses (Mang *et al*., 1999; van der Kuyl *et al*., 1997). Similarly, RD-114 of cats is a recombinant between *env* of BaEV and *gag-pol* of PcEV (*Papio cynocephalus* endogenous retrovirus) (van der Kuyl *et al*., 1999). The three types of recombinant PERVs reported here thus have ample precedent among retroviruses in other species.

However, the possibility that the recombinant clones are PCR artefacts can not be excluded. Zylstra *et al*. (1998) reported artefactual recombinant sequences for the murine immunoglobulin germline V gene generated by PCR and detected by sequencing about 1 kb of cloned PCR products. Interestingly they found that *Pfu* polymerase generated a higher percentage of recombinant artefacts than *Taq*. The number of artefacts increased with the number of amplification cycles, and pseudo-recombination events were located throughout the whole PCR

product with no preferred sites of template exchange. In the present study, at least one of the apparent recombinant PERVs is more likely to be a genuine recombinant because three clones showing the same recombination pattern with two DNA polymerase were found (Fig 4.9). However, conformation of the reality of the recombinant clones will require their amplification directly from genomic DNA template using recombinant-specific primers. This remains an important task for the immediate future.

Recombinant sequences create a problem for phylogenetic analysis. Most phylogenetic analyses assume that sequences evolve independently. Therefore more different sequences separated further in the past. In other words, more similar sequences shared a common ancestor at a less distant time. Recombination draws some odd conclusions in the phylogenetic analysis because the relationship of the sequences does not reflect independent evolutionary history but a genetic exchange. To solve this problem, a method called "sister-scanning" was recently developed to detect the occurrence of recombination, allow for it in the phylogenetic analysis, and provides some systematic way of representing its occurrence (Gibbs *et al.*, 2000). Unfortunately, sister-scanning analyses of the PERV sequences reported in this present study could not be performed in time for inclusion in this thesis.

The occurrence of recombinant PERVs also provides a salutary reminder of the potential for recombinational repair of defective retroviruses and for recombination of human and porcine retroviruses during xenotransplantation. Defective PERVs could possibly regain infectious potential through recombination. Furthermore PERVs could recombine with human endogenous retroviruses to generate totally novel retrovirus. Thus these results corroborate the recombinogenic potential of retroviruses and highlight the potential danger of intra and/or interspecies recombination of PERVs in xenotransplantation.

Chapter Five

Physically Mapping PERVs

5.1. Introduction

Pigs (*Sus Scrofa*) have a diploid chromosome number of 38 (18 pair of autosomes and XY sex chromosomes). The standard G-banding and R-banding karyotypes have been reported in pig, dividing the chromosomes into cytogenetically defined regions by systematically described landmarks (Gustavsson, 1988). The standard R-banded karyotype has been used for mapping PERVs. Fig 5.1 shows the pig R band karyogram and its ideogram.

PERVs, as described in Chapter 4, have a potential infectious risk in relation to xenotransplantation. Le Tissier *et al*. (1997) and Akiyoshi *et al*. (1998) reported about 50 copies of PERV in the pig genome, of which 10 to 20 copies could correspond to full-length proviruses (Akiyoshi *et al*., 1998).

Rogel-Gaillard *et al*. (1999) reported 62 PERV-positive clones found by screening a Large White porcine BAC (bacterial artificial chromosome) library using specific primers for the protease gene of type-C retrovirus. The PERV clones were analysed using PERV-A and -B specific envelope primers, showing that 28 clones contained PERV-B, 12 clones contained PERV-A envelope, 6 clones were positive for both types and 16 clones were negative for both A and B envelope types. Ultimately, 33 BAC clones containing PERV sequences were mapped by FISH to 22 distinct locations on 14 chromosomes including X and Y. The pig used for making the BAC clones for FISH mapping of PERV in this study was an outbred commercial Large White animal (Rogel-Gaillard *et al*., 1999). Of course such outbred animals may not be the most suitable for xenografts. In this study, the main focus is on the inbred Westran line of pigs where there is expected to be no or little variation between animals in the sites of hybridisation. The locations found in two Westran animals will be compared with those reported in Large White.

The FISH experiments presented here were carried out at both The Queen Elizabeth Hospital and the Waite Campus, University of Adelaide, Australia with the help of Dr Graham C. Webb, a molecular cytogenetist. I generated all the clones, made all the probes and assisted in the hybridisation. Dr Webb scored all PERV signals on the Westran pig chromosomes.

In addition to the FISH experiment, PERVs were physically mapped using a technique developed by Dr Yi-Mo Deng from the Department of Microbiology, Prince of Wales Hospital,

NSW, Australia in conjunction with A/Prof Chris Moran, to amplify by PCR one of the two unique flanks of a repetitive locus (in this, a PERV) and then use this to map the insert physically. This method yields a unique sequence tag for the site and this enables physical mapping on, for example, a somatic cell hybrid panel. The overall process for this method is illustrated in Fig 5.2. The porcine genomic sequences were first amplified by PCR using a PERV-specific primer and a porcine SINE-specific primer. PCR products were cloned and those sequences that contained PERV-flanking regions were selected using a second-round PCR and cloning. This method obviates the need for individual library construction or linker/adaptor ligation, and can be used to locate individual sites of moderately repeated, dispersed DNA sequences in the genome (Deng *et al.*, submitted). To evaluate this method, a PCR product comprising a sequence flanking a PERV was mapped using a porcine-rodent somatic cell hybrid panel.

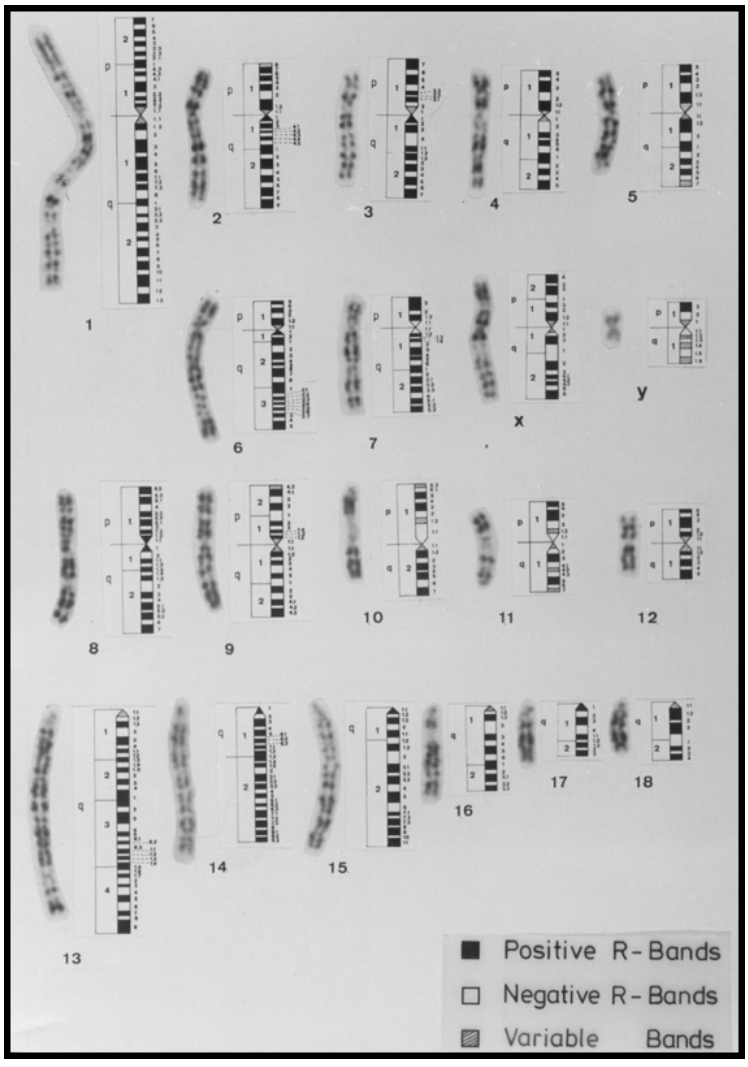


Fig 5.1 Pig R-band karyotype and ideogram (Source: Rønne *et al.*, 1987).

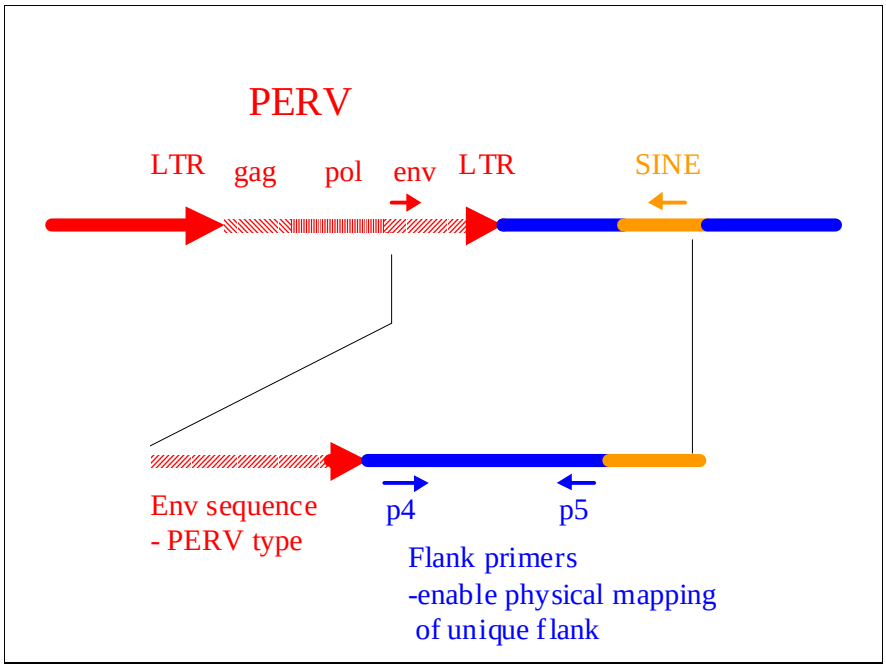


Fig 5.2 Strategy to amplify a PERV junction fragment. A PERV flanking sequence was initially amplified with a PERV-specific envelope primer and a SINE primer. New flanking primers (p4 and p5) were designed based on the flanking sequence information and used to physically map the site using the French somatic cell hybrid panel (Figure courtesy of Dr Yi-Mo Deng and A/Prof Chris Moran).

Thus two separate mapping experiments are presented in this Chapter; large scale, comprehensive FISH mapping of PERVs in two inbred Westran pigs and evaluation of a new mapping method for dispersed repetitive loci using semi-specific PCR cloning and a somatic cell hybrid panel.

5.2. Materials

5.2.1. Animals

White blood cells were cultured from two male Westran animals (No 115 and No 167). No 115 is the most inbred animal from the 5th generation of full-sib mating. No 167 is the son of No 115 from a sib mating.

5.2.2. Probes for FISH

Two randomly chosen PERV clones, *Taq*-82 and *Taq*-9, were used to as probes for PERV-A and PERV-B, respectively as previously described (Section 4.4). To improve the efficiency and specificity, the PERV inserts were excised from the vector by *Eco*RI restriction digestion (Section 2.8).

5.2.3. Primers for precise mapping of a PERV junction

A pair of primers, p4 and p5, was designed by Dr Yi-Mo Deng based on a PERV junction fragment sequence (Table 5.1 and Fig 5.2) obtained from a commercial Large White x Landrace crossbred animal.

Table 5.1 The primer sequences, PCR conditions ($MgCl_2$ concentration and annealing temperature) and expected product size for amplifying the genomic junction fragment flanking a PERV-B.

Primer sequences[1]	$MgCl_2$ concentration	Annealing Temperature	Expected size (bp)
p4. 5'-TCACACCACCTGCTACCTTTCC-3' p5. 5'-TCTGATGTGCCAACTGTGATTA-3'	2 mM	60 °C	115 bp

[1]p4: forward primer; p5: reverse primer

5.2.4. Somatic cell hybrid panel

The somatic cell hybrid panel and the methods for its analysis are as previously described (Sections 1.3.2.2 and 3.4.8).

5.3. Methods

5.3.1. FISH

5.3.1.1. Lymphocyte cell culture

Westran pig blood was collected in 10ml vacutainers containing lithium heparin by Dr Wayne Hawthorne and Dr Richard Allen at the Westran breeding facility at Karuah, NSW. The vacutainers were inverted to mix the lithium heparin to prevent coagulation and were transported by air to Adelaide within 12 hours of collection. 5ml cultures are prepared under aseptic

conditions and cultured at 37 °C for 72 hours. The cultures were inverted once or twice in this time to optimise growth of the cells by preventing the cells from sedimenting out of the solution. The composition of the culture media is outlined below.

Aminomax Basal + supplement medium (Gibco-BRL)	100 ml
Phytohaemagglutinin (PHA, M-form, Gibco-BRL)*	2 ml
Pokeweed Mitogen (PWM, 1 mg/ml, Sigma)	500 µl
Penicillin (10,000 Units/ ml), Streptomycin (10,000 Units/ ml), and Fungizone (25 µg/ml) (PSF, Commonwealth Serum Laboratories)	200 µl
Glutamine (0.2 M)	1 ml
Heparin	200 µl

*Given the unidentified nature of this extract, it is not possible to assign weight/volume values on the active material in the final product (Gibco BRL product and reference guide, Life Technologies)

300 µl blood was added to each 5ml of culture medium and incubated at 37 °C. To synchronise cells in mid-S phase, 300 µg/ml of thymidine (Sigma) was added 24 hours before finishing cell culture. At day three, the cells were rinsed three times with Dulbecco's phosphate buffer saline (PBS, Commonwealth serum labs), each time centrifuging at 1,000 rpm for 10 minutes and then discarding the supernatant. They were suspended once again in the culture medium to which 5-bromodeoxyuridine (5-BrdU, Sigma) was added to a final concentration of 20 µg/ml and FDU (5-flourodeoxyuridine, Sigma) to a final concentration of 0.5 µg/ml. After incubating for six and half hours, Colchicine (Sigma) (2 drops of 100 µg/ml from a 1 ml syringe per culture) was added 10 to 20 minutes before finishing incubation in order to disrupt the mitotic spindle and accumulate cells at metaphase.

To harvest the cells, they were first centrifuged at 1300 rpm (695 x *g*) for 10 minutes. The supernatant was removed and 12 ml hypotonic 0.075 M KCl solution (pre-warmed to 37°C) was added. The cells were incubated at 37 °C for 20 minutes in the hypotonic. 3 ml fixative (3:1 methanol-acetic acid) was then added to the cells in the hypotonic solution. This was centrifuged at 1400 rpm (745 x *g*) for 10 minutes. The supernatant was removed. 6 ml of 3:1 fixative was then added directly to the cells which were resuspended and then centrifuged at 1600 rpm (850 x *g*) for 7 minutes. This fixation and centrifugation step was repeated. Finally 1 ml fixative was added to the pellet and the cell was resuspended. A drop of the suspension was dropped onto a slide to check cell density under a light microscope. If it was necessary to alter cell density, either additional fixative was added or the cells were spun down in 6 ml fixative by centrifugation at 1600 rpm (850 x *g*) for 7 minutes and then resuspended in a smaller volume. Finally slides were

prepared by dropping 3 drops of the suspension from a Pasteur pipette onto each slide. slides, cleaned at manufacture, were further cleaned by soaking in 5 % Decon 90 (Selby) and rinsing with running tap water, each for approximately 2 hours. The slides were then rinsed in distilled water and three changes of 100 % ethanol. They were then dried vertically.

5.3.1.2. Making PERV FISH probes

To label PERV probes with Biotin, a BioNick™ Labeling System (Life Technologies) was employed for nick translation. To check the efficiency of incorporation of the biotin label, tritiated label was incorporated at the same time. 1 µCi of ^{3}H dATP (Amersham) was dried in a 1.5 ml Eppendorf tube. The following were then added:

10 x dNTP mix containing biotin-14-dATP (kit)	5 µl
1 µg of probe DNA	1 µl
distilled H_2O bring total volume up to	45 µl
10x enzyme mix containing DNA polymerase I and DNase I (kit)	5 µl
Total	50 µl

The reaction tube was briefly centrifuged and the labeling reaction was incubated for 1 hour at 16 °C. 5 µl of Stop Buffer (BioNick™ Labeling System, Life Technologies) was then added. Unincorporated nucleotides were separated from the labeled DNA probe using a 5 cm Sephadex G 50 (Pharmacia) column in a short Pasteur pipette plugged with sterile, non-absorbent cotton wool. The column was washed twice with TE (pH 7.6). Then 55 µl of reaction mixture (50 µl of reaction and 5 µl of Stop Buffer) was carefully layered onto the column. For the first fraction, 445 µl of TE was carefully layered and the first fraction was collected. After that, the following fractions were collected by adding the same amount of TE in Eppendorf tubes:

Fraction number	Volume
1	500 µl
2 – 10	100 µl
11 – 14	500 µl

To determine the success of the nick translation reaction, a Liquid Scintillation Counter (Beckman) was used to measure incorporation of the tritiated tracer in a small aliquot from each

fraction. 2 µl samples of each fraction were mixed with 150 µl distilled water and 1.5 ml BCS scintillation fluid (Amersham) in a numbered scintillation tube. The tubes were dipped in ethanol and wiped carefully to remove any static electricity before counting. The scintillation counts for each fraction showed that there are two major peaks, indicating the incorporated tritium (^3H) for the first peak and unincorporated tritium for the second peak. The fractions representing the first peak were pooled into a single tube and the incorporation percent was calculated. Over 10% is considered adequate incorporation. 50 µg sheared salmon sperm DNA (Promega) was added per 100 ng of probe to assist in precipitating the probe and to compete with the probe for non-specific binding to the chromosomes, thus reducing background labelling. $1/10^{th}$ volume of 3 M sodium acetate (pH 5.2) and 2.5 volume of absolute alcohol was added to precipitate the probe, which was mixed well and left at –80 °C for 30 minutes. After centrifugation for 10 minutes at 14,000 rpm, the supernatant was discarded and 200 µl of 70 % ice-cold ethanol was added to the tube. Again, the supernatant was removed and the pellet was air dried at 37 °C for 10 minutes. 40 µl of probe mix, containing 12 µl of probe in water, 8 µl of 10 x SSCP (1.2 M NaCl, 0.15 M NaCitrate, 0.1 M Na_2HPO_4, 0.1 M NaH_2PO_4, pH 6.0) and 20 µl of 20 % dextran sulfate in formamide, were then combined.

5.3.1.3. *In situ* hybridisation of PERV probes to chromosomes

Slides were immersed for 1 hour at 37 °C in 100 µg/ml RNase solution in 2 x SSC. The slides were rinsed 4 times with 2 x SSC for 2 minutes each and dehydrated through 35 %, 70 %, 95 % and 100 % ethanol for 2 minutes, respectively. The slides were then air-dried. Two Coplin jars, one containing slide denaturation mix (70 % deionised formimide, 2 x SSC, pH 7.0) and another empty jar were pre-heated at 70 °C. Slides were pre-heated in the empty jar and transferred to the denaturation mix for 2 minutes with gentle agitation. The denatured slides were immediately put into the cold (about –10 °C) 70 % ethanol and agitated for 2 minutes. This step allows the chromosomes to stay denatured. The slides were then dehydrated through an ethanol series (70 %, 80 %, 95 %, 100 %) for 2 minutes each and drained with the frosted end down.

Meanwhile, 22 X 50 mm coverslips were dipped into 70 % ethanol, dried and polished. The pre-made probe mix was briefly heated to 75 °C and mixed thoroughly. After brief centrifugation, the probe mix were denatured at 75 °C for 10 minutes and placed on ice for at least 2 minutes. 40 µl probe mix then were loaded onto each slide in a line of droplets. The slides were covered with coverslip and sealed with rubber cement. The slides were hybridised overnight at 37 °C in a sealed box. After hybridisation, they were stored at 4 °C.

5.3.1.4. Stringency rinses

To remove any non-specifically bound probe, high stringency rinsing solution (50% formamide, 2 X SSC) was dispensed into three Coplin jars and two of them heated to 39 °C in a water bath. To remove the hybridisation mix, the slides were quickly rinsed in the rinse solution at room temperature. The slides were then washed twice in the high stringency rinse solution for 10 minutes at 39 °C with shaking to remove any non-specifically bound probe. The slides were then rinsed twice in 2 X SSC solution for 5 minutes at 37 °C with shaking and rinsed in 1 X SSC solution for 10 minutes at room temperature with shaking. The final washing was carried out in a solution containing 4 X SSC and 0.05 % Tween 20 (Promega) for 3 minutes at room temperature with shaking.

5.3.1.5. Immunochemical detection of biotin label

When all rinsing steps were finished, excess moisture was removed from the slides by tapping the edge of slides on paper towel. 100 µl of 4 X SSC/1.0 % w/v bovine serum albumin (BSA) was dropped onto each slide. The slides were then covered with coverslip and incubated in a moist chamber for 10 minutes. 1.0 µl of Avidin-FITC (Vector Labs) per 200 µl of 4 X SSC/1 % BSA was thoroughly mixed. The coverslips were removed and the edge of each slide was tapped on a paper towel to remove excess moisture. 70 µl Avidin-FITC mix was placed onto each slide. The slides were covered with coverslips and incubated in the moist chamber for a minimum 40 minutes at 37 °C. The coverslips were again removed and rinsed twice with 4 X SSC/0.05 % Tween 20 for 5 minutes each at 37 °C with agitation. 1.0 µl of biotinylated goat anti-Avidin (Vector Labs) per 100 µl of 4 X SSC/1 % BSA was thoroughly mixed. Then 70 µl biotinylated goat anti-Avidin mix was placed onto each slide. The slides were covered with coverslips and incubated in the moist chamber for at least 40 minutes at 37 °C. The coverslips were then removed and the slides rinsed twice with 4 X SSC/0.05 % Tween 20 for 5 minutes each at 37 °C with agitation. Again, 70 µl Avidin-FITC mix was placed onto each slide. The slides were covered with a coverslip and incubated in the moist chamber for a minimum of 40 minutes at 37 °C.

After finishing the above amplification step, the final rinsing was carried out. The coverslips were removed and rinsed twice with 4 X SSC/0.05 % Tween 20 for 5 minutes each at 37 °C with agitation. The slides were rinsed once with 2 X SSC for 2 minutes and then rinsed twice with PBS for 2 minutes respectively.

5.3.1.6. Staining and viewing of chromosomes

The chromosomes were stained in 5 μg/ml of propidium iodide in PBS for 5 minutes with agitation. The slides were then rinsed three times in PBS for 20 seconds each with agitation. The excess PBS on the slides was removed by tapping on paper towel and 80 μl of PPD11 (p-phenylenediamine dihydrochloride, pH 11) (Lemieux *et al.*, 1992) mountant was placed on the slide. A 24 X 60 mm coverslip were placed onto the slide and the excess mountant was removed. The slides were dried and then sealed with nail varnish.

The slides were examined under blue excitation to give yellow FISH signals and orange stained chromosomes. Photos were then taken of well spread metaphases using colour-positive slide film (ASA 400, Fujichrome).

5.3.1.7. Scoring of FISH signals

FISH signal, appearing as yellow grains on R-banded chromosomes, was scored and the data plotted onto pig standard R-band ideograms of about 300 bands (Gustavsson, 1988). 20 cells were counted in each animal (No 167 and No 115) for each of the PERV-A and PERV-B probes.

5.3.1.8. Statistical analysis of hybridisation signals

The Z_{max} test (Ewens *et al.*, 1992) was used in this study to analyse the cumulative FISH data from 20 metaphase cells to determine the significance of each hybridisation location. This test was originally designed for analysing grain counts from radioactive *in situ* hybridisation but is ideally suited for the present situation where there are multiple sites of hybridisation which must be distinguished from background labelling. The null hypothesis is that there is no preferred site of hybridisation. This hypothesis is equivalent to the assumption that any given hybridisation event will occur on segment *i* with probability p_i (p_i = length of segment *i*, relative to the length of all chromosomes combined) and is to be tested against the alternative hypothesis that there is some (unknown) segment, called segment *j*, containing a hybridisation site at which there is an increased probability of hybridisation. In statistical terms, if Prob(*i*) is the probability of hybridisation to segment *i*, the null hypothesis and alternative hypothesis become

Ho: Prob(i) = p_i,

H_1: Prob(i) = $\theta\, p_i$, $i \neq j$,

Prob(j) = $1 - \theta + \theta\, p_i$

where θ is an unknown parameter ($0 \leq \theta \leq 1$),

j is an unknown number of set (1, 2, …, k)

(Ewens *et al*., 1992).

To calculate Z_{max}, the test statistic Z_i is calculated for each segment i using the equation.

$$Z_i = (n_i - np_i)/\sqrt{np_i(1-p_i)}$$

where n_i = the number of grains counted in i^{th} chromosomal segment

n = total number of grains

p_i = length of segment i, relative to the length of all chromosomes combined

(Ewens *et al*., 1992).

The site(s) giving the highest Z_i (ie Z_{max}) is identified. If this value is significant as judged by comparison with threshold values provided by Ewens *et al.* (1992) (see Table 5.2), the site(s) is eliminated from the analysis. A new set of Z_i are calculated excluding the significant site, and the next Z_{max} is identified and tested for significance. The process is applied recursively until Z_{max} is no longer significant.

Table 5.2 5%, 1%, and 0.1% significance points for Z_{max} (taken from Ewens *et al.*, 1992).

k[1]	5 %	1 %	0.1 %
70	3.19	3.63	4.18
80	3.23	3.66	4.21
90	3.26	3.69	4.24
100	3.29	3.72	4.26
110	3.32	3.74	4.29
120	3.34	3.77	4.31
130	3.36	3.79	4.32
140	3.39	3.80	4.34
150	3.40	3.82	4.35

[1]number of segments being tested.

5.3.2. PCR conditions for somatic cell hybrid mapping

PCR amplication was performed using 20 ng of somatic cell hybrid panel DNAs, 1 X PCR buffer (Perkin Elmer), 2 mM $MgCl_2$, 40 µM each of dNTPs, 0.5 unit of *AmpliTaq* Gold DNA polymerase (Perkin Elmer) and 400 nM of each primer (Table 5.1) in a 25 µl reaction volume. The PCR cycling conditions included an initial denaturation of 1 min at 95 ºC followed by 35

cycles of 30 seconds at 95 ºC, 30 seconds at 60 ºC, and 30 seconds at 72 ºC, with a final 20 minutes extension at 72 ºC using a PTC-100 Programmable Thermal Controller (MJ Research, Inc).

5.4. Results

5.4.1. Incorporation of biotin

The incorporation of biotin was inferred from the incorporation of tritium tracer in the BioNick labelling process. Table 5.3 shows incorporation % for the tritium label estimated by comparing the counts incorporated into the probe with the total count for the PERV-A and PERV-B labelling reactions. Noting that 5 µl of 0.1 mM biotin were used in the BioNick process, The biotin incorporation was estimated from these percentages (Table 5.3). It is evident that the efficiency of biotin incorporation into the PERV-A and PERV-B probes is very similar, indicating that there should be no difference in FISH signals due to properties of the probes.

Table 5.3 Calculated incorporation rate of tritium (^{3}H) and the amount of biotin incorporation for each PERV probe labelling reaction.

PERV-A			
Probe counts (cpm)	Total Counts (cpm)	^{3}H incorporation	Biotin incorporation
3863	38903	9.93 %	49.65 pM
PERV-B			
Probe counts (cpm)	Total Counts (cpm)	^{3}H incorporation	Biotin incorporation
4080	43293	9.42 %	47.1 pM

5.4.2. Chromosomal distributions of PERVs

For animal 115 with the PERV-A probe, 478 grains were found distributed over 106 chromosomal segments (Fig 5.3), giving an average density of 4.51 grains per segment. 745 grains were found over 131 positions from animal 167 with the same PERV-A probe (Fig 5.4), giving a higher mean grain density of 5.69. The pooled PERV-A FISH signals across these two animals (115 + 167) are presented on a secondary plot with mean density of 7.84 (1223 grains/156 chromosomal positions) (Fig 5.5).

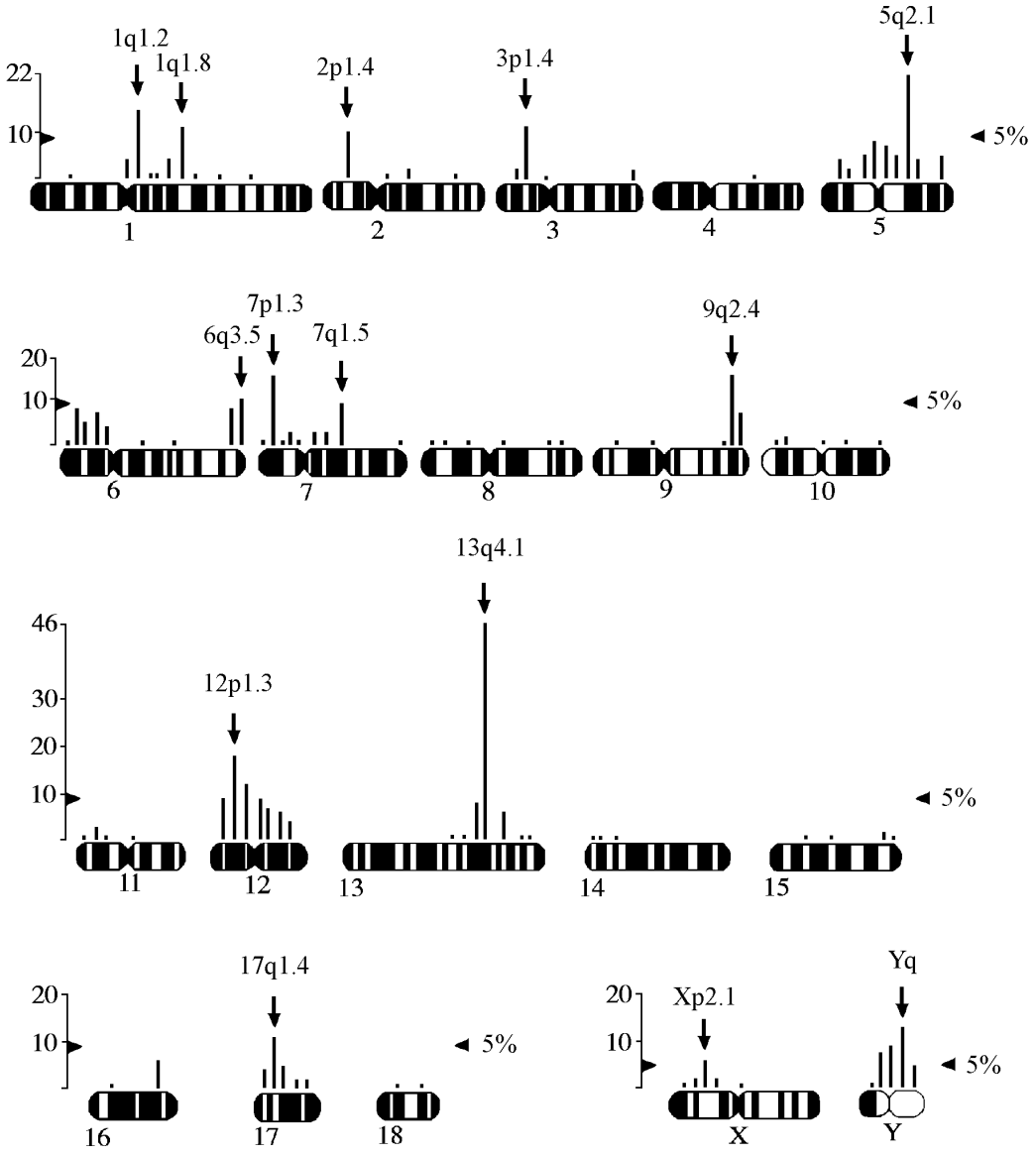


Fig 5.3 The locations of PERV-A in animal 115 detected by FISH. The vertical scale is number of grains. The 5% significance threshold for the Z_{max} test is indicated. Arrows indicate most likely band location of PERVs.

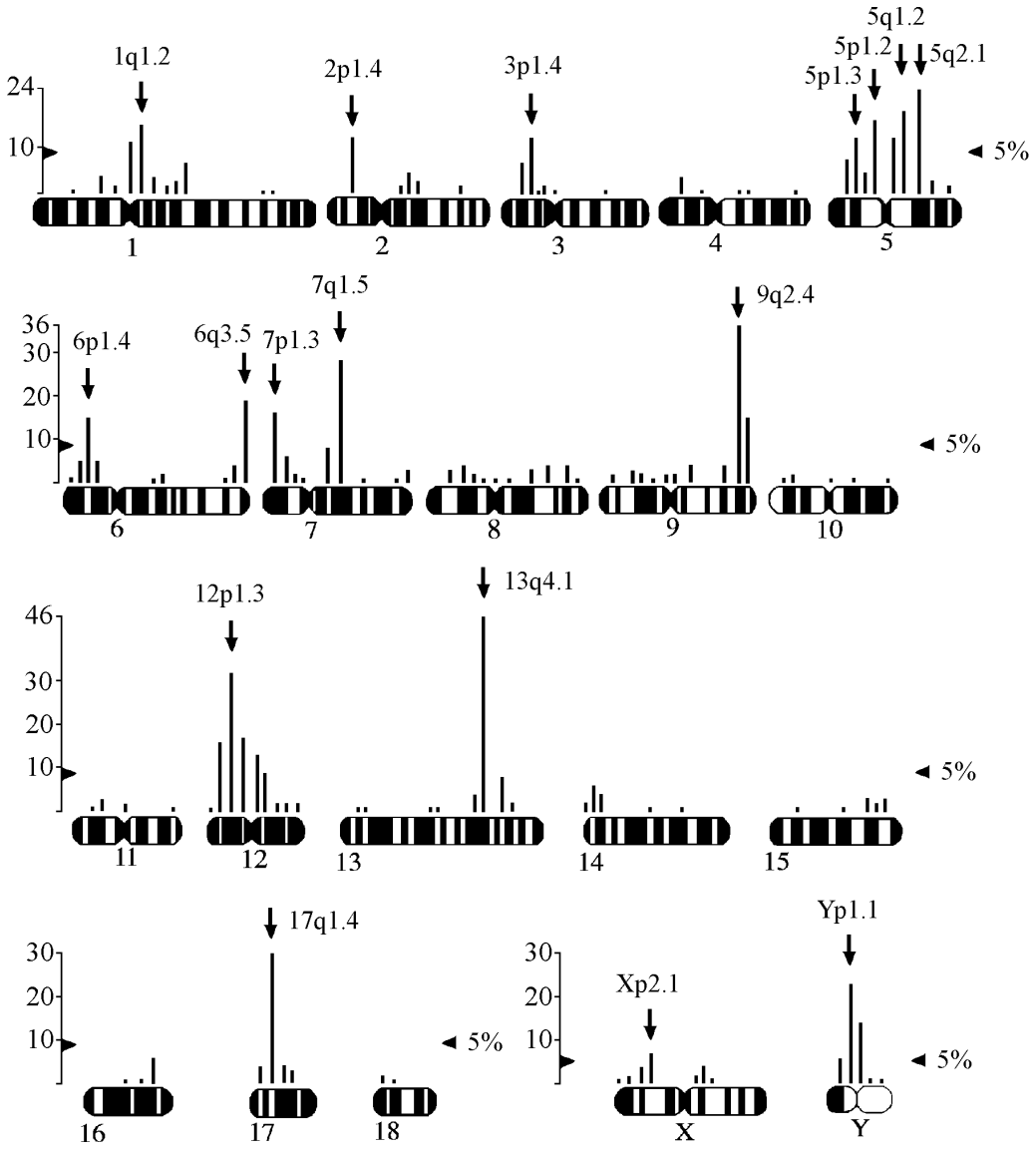


Fig 5.4 The locations of PERV-A in animal 167 detected by FISH. The vertical scale is number of grains. The 5% significance threshold for the Z_{max} test is indicated. Arrows indicate most likely band location of PERVs.

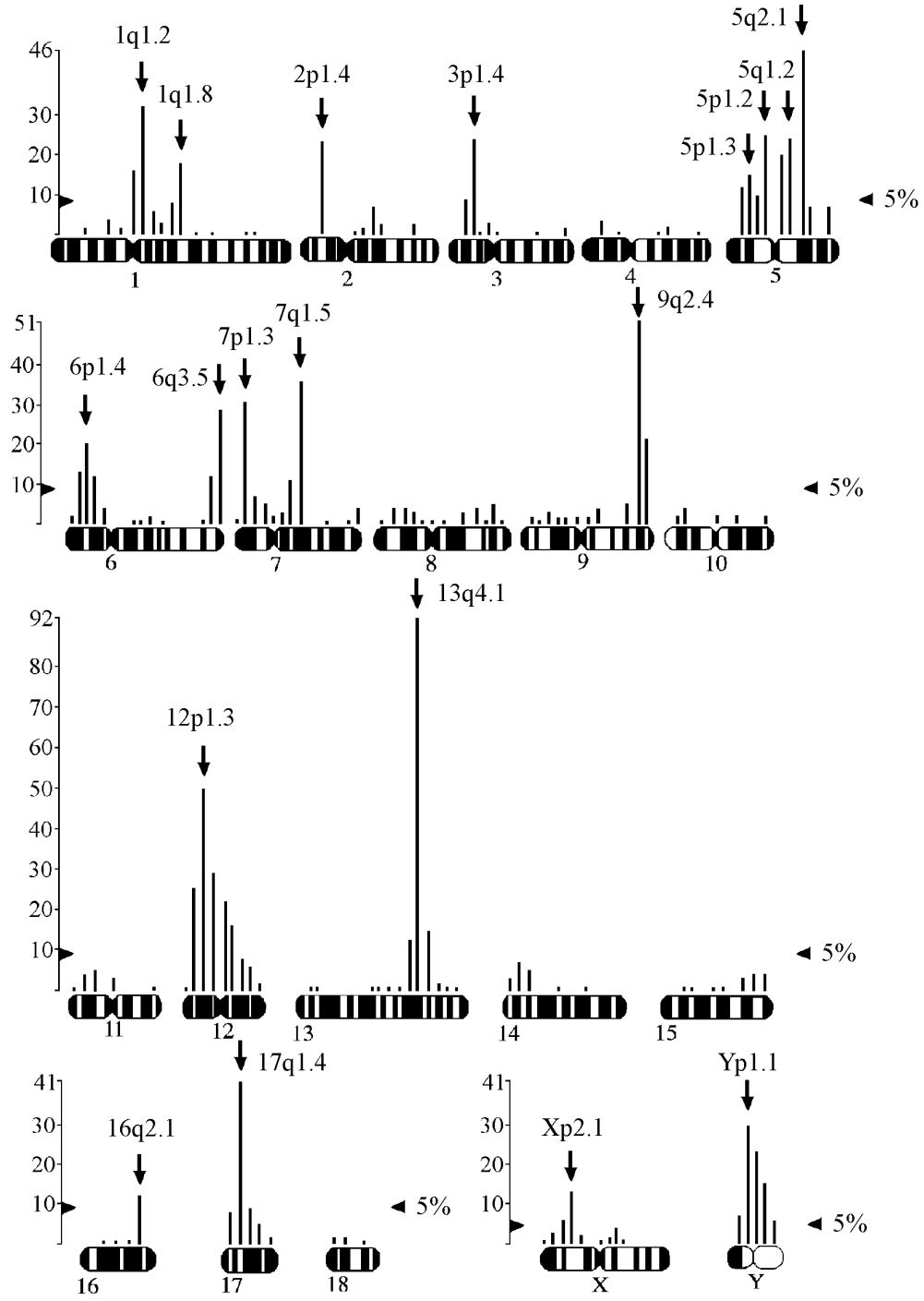


Fig 5.5 The results of PERV-A hybridisation pooled across two animals (115 + 167). The vertical scale is number of grains. The 5% significance threshold for the Z_{max} test is indicated. Arrows indicate most likely band location of PERVs.

Similarly, hybridisation of the PERV-B probe to metaphases from animal 115 and animal 167 were scored (Fig 5.6 and Fig 5.7). Pooled PERV-B FISH results were also obtained (Fig 5.8). The mean grain density with this probe is 3.6 (342 grains/95 positions) for animal 115. Again animal 167 had a higher mean density of 4.44 (560 grains/126 positions). The pooled mean density is 6.05 (902 grains/149 positions) from the results across two animals (115+167). The hybridisation data for animals 115, 167, and the pooled results across the two animals for the PERV-A and PERV-B probes at each chromosomal locations are presented in Appendix 5.1.

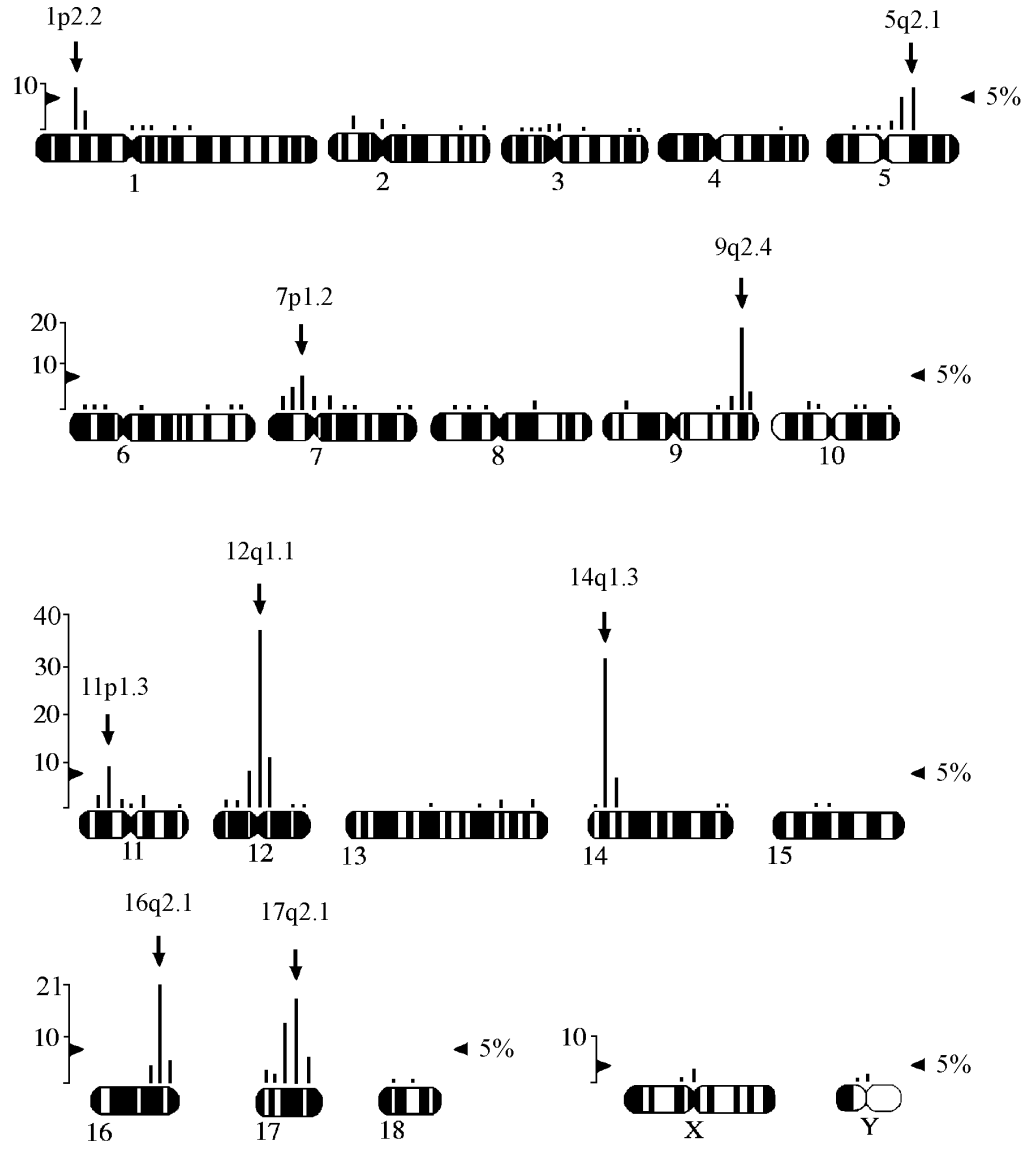


Fig 5.6 The locations of PERV-B in animal 115 detected by FISH. The vertical scale is number of grains. The 5% significance threshold for the Z_{max} test is indicated. Arrows indicate most likely band location of PERVs.

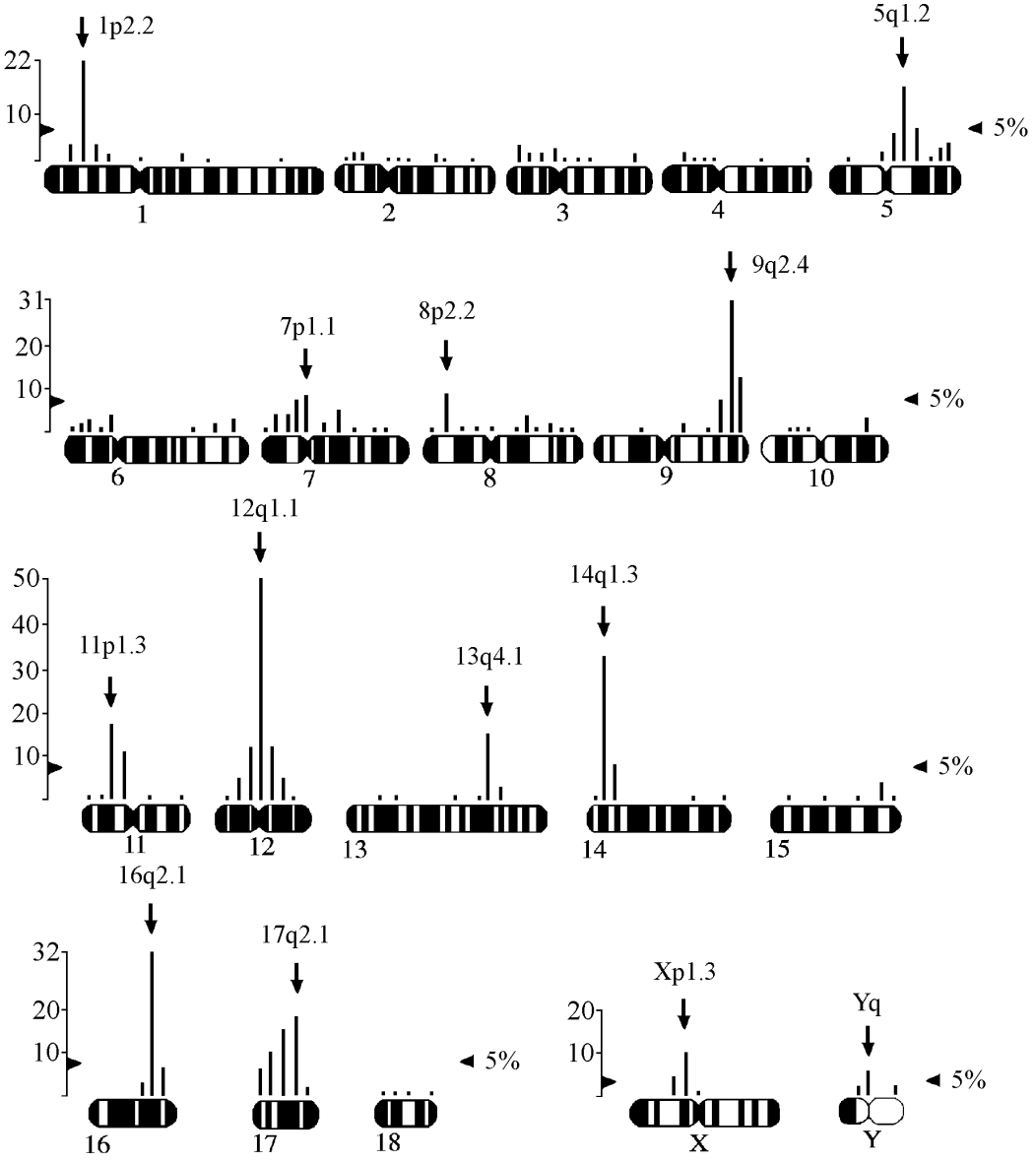


Fig 5.7 The locations of PERV-B in animal 167 detected by FISH. The vertical scale is number of grains. The 5% significance threshold for the Z_{max} test is indicated. Arrows indicate most likely band location of PERVs.

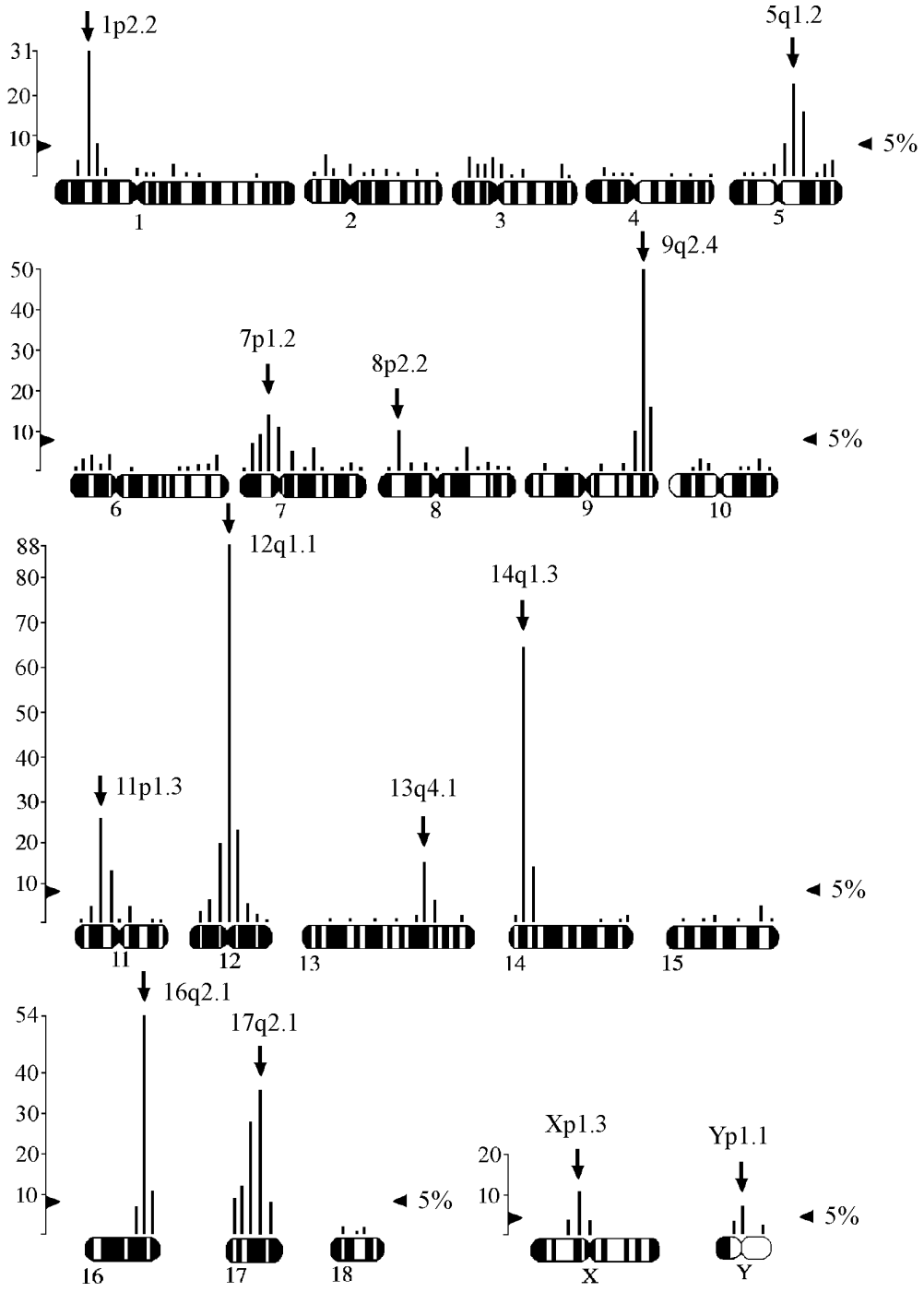


Fig 5.8 The results of PERV-B hybidisation pooled across two animals (115 + 167). The vertical scale is number of grains. The 5% significance threshold for the Z_{max} test is indicated. Arrows indicate most likely band location of PERVs.

The results just quoted (and summarised in Table 5.4) show that the PERV-A probes generated stronger signal (more grains) that the PERV-B. Also it is evident that the overall signal intensity is stronger in animal 167 than animal 115. A difference in the quality of the cultures, where cells from animal 167 grew better than cells from 115, which had to be treated with Fungizone, is the likely explanation for this difference.

Table 5.4 Chi-square contingency analysis of number of grains for PERV-A and PERV-B probes in animals 115 and 167. The expected values based on the null hypothesis of equal total numbers of grains for each probe and each animal are shown in parentheses.

Animal	Probe		
	PERV-A	PERV-B	Total
115	478 (471.93)	342 (348.07)	820
167	745 (751.07)	560 (553.93)	1305
Total	1223	902	2125

X_1^2 = 0.299, p = 0.585

Chi-square homogeneity analysis (Mead *et al.*, 1993) shows that there is no interaction between animal and probe (Table 5.4). This means that it is valid to test the pooled marginal totals separately for the effects of animal and probe against the null hypothesis of equal number of grains (signal intensity). For the comparison of PERV-A versus PERV-B signal intensity, the result is X^2 = 48.79 with one degree of freedom, indicating a very highly significant deviation from expectation: the PERV-A probe hybridised much more strongly than PERV-B, presumably due to a larger number of PERV-A inserts. Another chi-square test was performed to compare the results of the two animals. The result provides even more significant evidence (X_1^2 = 110.69) for deviation from the expected equal signal intensity. The most likely explanation for this difference is the superior quality of the chromosome cultures and spreads from animal 167.

Z_{max} test statistics have been calculated for the hybridisation data from animals 115 and 167, and the pooled results across the two animals (Tables 5.5 to 5.10). In this analysis, each site of hybridisation was taken as a different segment. It was assumed as a first approximation that each segment, is the same size. It should be noted that since both animals were males, segments on the sex chromosomes (X and Y chromosomes) occur only half as often as autosomal segments in these two animals, and this has been taken into account in the Z_{max} statistic.

Table 5.5 Z_{max} test for PERV-A hybridisation in animal 115 (20 cells scored).

Chromosomal location	Number of grains	Z_{max}	Significance level
13q4.1	46	19.064	0.1 %
Yq	14	9.049	
5q2.1	22	8.549	
12p1.3	18	6.845	
Yq	9	6.117	
1q1.2	16	6.031	
Yp1.1	8	5.714	
7p1.3	15	5.679	
9q2.4	15	5.679	
Xp2.1	6	4.797	
12p1.2	12	4.552	
Yq	5	4.163	1 %
1q1.8	11	4.118	
3p1.4	11	4.118	
17q1.4	11	4.118	
2p1.4	10	3.856	
6q3.5	10	3.856	
7q1.5	9	3.464	5 %
12p1.4	9	3.464	
12q1.1	9	3.464	
5p1.1	8	3.223	N.S.
6p1.5	8	3.223	
6q3.4	8	3.223	
13q4.1	8	3.223	
Xp2.2	4	2.920	
Xq1.3	4	2.920	

Table 5.6 Z_{max} test for PERV-A hybridisation in animal 167 (20 cells scored).

Chromosomal location	Number of grains	Z_{max}	Significance level
13q4.1	46	16.467	0.1 %
Yp1.1	22	12.979	
9q2.4	36	12.840	
12p1.3	32	11.508	
17q1.4	30	10.962	
7q1.5	27	9.927	
Yq	14	9.364	
5q2.1	24	8.823	
5q1.2	19	6.920	
6q3.5	19	6.920	
5p1.2	17	6.265	
12p1.2	17	6.265	
1q1.2	16	6.048	
7p1.3	16	6.048	
12p1.4	16	6.048	
6p1.4	15	5.955	
9q2.5	15	5.955	
1q1.1	12	5.408	
2p1.4	13	5.311	
3p1.4	13	5.311	
5p1.3	13	5.311	
5q1.1	13	5.311	
12q1.1	13	5.311	
Xp2.1	7	5.061	
Yp1.3	6	4.440	
12q1.2	9	3.741	1 %
5p1.4	8	3.211	N.S.
7q1.4	8	3.211	
13q4.3	8	3.211	

Table 5.7 Z_{max} test for PERV-A hybridisation data pooled across two animals (115+167).

Chromosomal location	Number of grains	Z_{max}	Significance level
13q4.1	92	29.346	0.1 %
12p1.3	50	15.790	
9q2.4	51	15.738	
Yp1.1	30	15.639	
5q2.1	46	14.713	
17q1.4	41	13.179	
Yq	23	13.166	
7q1.5	36	11.549	
1q1.2	32	10.224	
7p1.3	31	10.041	
6q3.5	29	9.676	
12p1.2	29	9.676	
Yq	15	9.533	
Xp2.1	13	8.499	
5p1.2	25	8.395	
12p1.4	25	8.395	
3p1.4	24	8.309	
5q1.2	24	8.309	
2p1.4	23	8.217	
9q2.5	22	8.141	
12q1.1	22	8.141	
5q1.1	20	7.551	
6p1.4	20	7.551	
1q1.8	18	6.901	
1p1.1	16	6.043	
12q1.2	16	6.043	
5p1.3	15	5.777	
13q4.3	14	5.506	
Yp1.3	7	5.169	
6p1.5	13	5.081	
5p1.4	12	4.753	
6p1.3	12	4.753	
6q3.4	12	4.753	
13q4.1	12	4.753	
16q2.1	12	4.753	
7q1.4	11	4.674	
Xp2.2	6	4.524	
Yq	6	4.524	
5p1.2	10	4.181	1 %
3p1.5	9	3.667	5 %
17q2.1	9	3.667	
1q1.7	8	3.198	N.S.
2q2.1	8	3.198	
12q1.3	8	3.198	
14q1.3	8	3.198	
17q1.2	8	3.198	

Table 5.8 Z_{max} test for PERV-B hybridisation in animal 115 (20 cells scored).

Chromosomal location	Number of grains	Z_{max}	Significance level
12q1.1	38	17.995	0.1 %
14q1.3	32	15.873	
16q2.1	21	10.475	
9q2.4	19	9.762	
17q2.1	18	9.588	
17q2.1	13	6.816	
12q1.2	11	5.750	
1p2.2	9	4.595	
5q2.1	9	4.595	
11p1.3	9	4.595	
12p1.2	8	4.333	
5q1.2	7	3.739	0.5 %
7p1.2	7	3.739	
14q1.4	6	3.212	N.S.
17q2.2	6	3.212	

Table 5.9 Z_{max} test for PERV-B hybridisation in animal 167 (20 cells scored).

Chromosomal location	Number of grains	Z_{max}	Significance level
12q1.1	50	21.386	0.1 %
9q2.4	31	14.252	
14q1.3	33	14.154	
16q2.1	33	14.154	
1p2.2	22	9.988	
17q2.1	18	8.102	
Xp1.3	10	8.024	
11p1.3	17	7.769	
5q1.2	16	7.413	
13q4.1	15	7.035	
17q2.1	15	7.035	
9q2.5	12	5.614	
12p1.2	12	5.614	
12q1.2	12	5.614	
11p1.1	11	5.416	
17q1.3	10	4.907	
8p2.2	9	4.504	
Yq	5	4.303	
7p1.1	8	3.933	1 %
14q1.5	8	3.933	
Xp2.1	4	3.467	5 %
5q2.1	7	3.417	
7p1.2	7	3.417	
9q2.3	7	3.417	
5q1.1	6	2.928	N.S
16q2.2	6	2.928	
17q1.2	6	2.928	

Table 5.10 Z_{max} test for PERV-B hybridisation in pooled across two animals (115+167).

Chromosomal location	Number of grains	Z_{max}	Significance level
12q1.1	88	33.032	0.1 %
14q1.3	65	25.151	
16q2.1	54	21.471	
9q2.4	50	20.548	
17q2.1	36	14.812	
1p2.2	31	12.884	
17q2.1	28	11.790	
11p1.3	26	11.121	
5q1.2	23	9.891	
12q1.2	23	9.891	
12p1.2	20	8.826	
Xp1.3	11	8.165	
13q4.1	17	7.415	
5q2.1	16	7.032	
9q2.5	16	7.032	
7p1.2	14	6.217	
14q1.5	14	6.217	
11p1.1	13	5.904	
Yp1.1	7	5.633	
17q1.4	12	5.437	
7p1.1	11	4.948	
16q2.2	11	4.948	
8p2.2	10	4.648	
9q2.3	10	4.648	
7p1.2	9	4.197	0.5 %
17q1.2	9	4.197	
1p2.1	8	3.707	5 %
5q1.1	8	3.707	
17q2.3	8	3.707	
Xp2.1	4	3.254	N.S.
Xp1.1	4	3.254	
7p1.3	7	3.246	
12p1.3	7	3.246	
16q2.1	7	3.246	

5.4.2.1. PERV-A locations

There are 20 significant sites in animal 115 and 26 in animal 167 (Table 5.5 and Table 5.6). For the pooled data, there are 41 significant sites (Table 5.7). However, it is highly unlikely that there are so many sites of PERV-A insertion. In many cases, significant sites are in adjacent or contiguous bands and are almost certainly the results of “spillover” of signal between bands. To choose one example from many, the site at 9q2.5 in Fig 5.5 is almost certainly spillover from the major peak at 9q2.4.

After making allowances for adjacent/contiguous significant sites, there are 14 major PERV-A peaks identified in animal 115 and 17 major PERV-A peaks identified in animal 167 (indicated with vertical arrows in Fig 5.3 and Fig 5.4). The differences in PERV-A major peaks between 167 and 115 are on chromosomes 1, 5 and 6. In chromosome 1, there are two significant peaks (1q1.2 and 1q1.8) observed in animal 115 but only one location (1q1.2) was significant in animal 167. The peak in the position 1q1.8 is significant in the pooled results across two animals (Fig 5.5). Animal 167 has a peak at the same location but it falls below the significant threshold in the analysis of 167 data alone. On the basis of pooled results it is assumed that the location 1q1.8 is a genuine PERV-A location in Westran pigs. There are four significant locations (5p1.3, 5p1.2, 5q1.2 and 5q2.1) on chromosome 5 in animal 167. Only one location (5q2.1) was observed in animal 115. When the pooled results across two animals were analysed, all four locations are significant. It is quite clear from the pattern of peaks and troughs that chromosome 5 has at least 3 PERVs sites. These sites are apparent in animal 115 but due to the lower overall hybridisation intensity are not significant in this animal. Similarly, the site on 6p1.4 is significant in animal 167 but is not significant in animal 115, lying below the significance threshold in the analysis of 115 data alone. The significant pooled results indicate that the location 6p1.4 is also a genuine PERV-A location in Westran pigs.

Thirteen of the significant PERV-A major peaks identified above (1q1.2, 2p1.4, 3p1.4, 5q2.1, 6q3.8, 7p1.3, 7q1.5, 9q2.4, 12p1.3, 13q4.1, 17q1.4, Xp2.1, Yq) are identical in animals 115 and 167, confirming that these locations have genuine PERV-A insertions. There are another five strongly suggestive PERV-A locations (1q1.8, 5p1.3, 5p1.2, 5q1.2, 6p1.4), which are significant in one animal but are present below the significance threshold in the other animal. The site at 16q2.1 is also a suggestive PERV-A location: although it is significant only in pooled results, the pattern of peaks is identical in both animals. Therefore there are 19 PERV-A locations, including suggestive locations, identified in the Westran line. The results are quite consistent between the two Westran animals, as expected. However, the Westran PERV-A locations are quite different from those of Large White pig (Table 5.11).

Table 5.11 Sites of significant major PERV-A peaks in Westran animals 115, 167 and pooled results for these two animals. The published PERV-A locations in Large White pig are also presented.

Animals			Published PERV-A locations[1]
115 (20 cells)	167 (20 cells)	115+167 (40 cells)	
1q1.2 1q1.8	1q1.2	1q1.2 1q1.8	 1q2.1 1q2.3 1q2.4
2p1.4	2p1.4	2p1.4	
3p1.4	3p1.4	3p1.4	
   5q2.1	5p1.3 5p1.2 5q1.2 5q2.1	5p1.3 5p1.2 5q1.2 5q2.1	
 6q3.5	6p1.4 6q3.5	6p1.4 6q3.5	
7p1.3 7q1.5	7p1.3 7q1.5	7p1.3 7q1.5	
			8p1.2
9q2.4	9q2.4	9q2.4	
12p1.3	12p1.3	12p1.3	
13q4.1	13q4.1	13q4.1	 13q4.2 13q4.3 13q4.9
		16q2.1	
17q1.4	17q1.4	17q1.4	
Xp2.1	Xp2.1	Xp2.1	
Yq	Yp1.1	Yp1.1	Yp1.2

[1]Based on Rogel-Gaillard *et al.* (1999).

5.4.2.2. PERV-B locations

There are 10 significant sites of hybridisation in animal 115 and 17 significant sites in animal 167 (Table 5.8 and Table 5.9). However, after allowing for the adjacent and contiguous sites apparently affected by "spillover" of hybridisation signal (Fig 5.6 and Fig 5.7), it is reasonable to conclude that there are 9 significant major peaks in animal 115 and 13 in animal 167. The differences between animals 115 and 167 are the peaks on chromosomes 8, 13, X, and Y. All these different PERV-B locations are significant in animal 167 but are not above background in animal 115 (Fig 5.6 and Fig 5.7). However, they are significant in the pooled results across the two animals (Fig 5.8).

Nine significant PERV-B major peaks (1p2.2, 5q2.1, 7p1.2, 9q2.4, 11p1.3, 12q1.1, 14q1.3, 16q2.1, 17q2.1) are identical in animals 115 and 167, confirming these 9 as unequivocal PERV-B locations in the Westran line. There are 4 more suggestive locations (8p2.2, 13q4.1, Xp1.3, Yq) in Westran pigs. When these PERV-B locations are compared with the published PERV-B locations in Large White (Gogel-Galliard *et al.*, 1999), 4 PERV-B sites, located on chromosomes 7, 9, 13, and 17, are identical or very close, indicating that these are probably common PERV-B insertions in these two breeds (Table 5.12).

Table 5.12 Significant PERV-B location in animals 115, 167 and pooled results for these two animals. The published PERV-B locations in a Large White pig are also presented.

Animals			Published PERV-B locations[1]
115 (20 cells)	167 (20 cells)	115+167 (40 cells)	
1p2.2	1p2.2	1p2.2	
			4p1.1
5q2.1	5q1.2	5q1.2	
7p1.2	7p1.1	7p1.2	7p1.2 → p1.1
	8p2.2	8p2.2	
9q2.4	9q2.4	9q2.4	9q2.6
			10p1.2
11p1.3	11p1.3	11p1.3	11q1.4
12q1.1	12q1.1	12q1.1	
	13q4.1	13q4.1	13q4.2
14q1.3	14q1.3	14q1.3	14q2.8
16q2.1	16q2.1	16q2.1	
17q2.1	17q2.1	17q2.1	17q2.1
	Xp1.3	Xp1.3	
	Yq	Yq	

[1]Based on Rogel-Gaillard *et al.* (1999)

5.4.3. A physical location for a PERV junction fragment

To evaluate the method presented in Fig 5.2, the PERV junction fragment cloned by Dr Y-Mo Deng was mapped on the somatic cell hybrid panel. Analysis of the amplification pattern in the 27 porcine-rodent somatic cell hybrid clones (Yerle *et al.*, 1996) allowed regional assignment of the PERV junction PCR product to porcine chromosome 17 (1/2 q2.1) - q2.3 with 98% concordance (Chevalet *et al.*, 1997) (Fig 5.9, Fig 5.10 and Appendix 5.2).

Rogel-Gaillard *et al*. (1999) have reported that two BAC clones containing PERV-B envelope sequences map to porcine chromosome 17q2.1 in a Large White pig. This suggests that the PERV junction reported here is the same as the PERV-B site mapped by them. Also, analysis of PERV-B FISH data in the inbred Westran pig shows strong positive signals at this same location on pig chromosome 17 with a PERV-B envelope sequence probe (See Fig 5.6, 5.7 and 5.8).

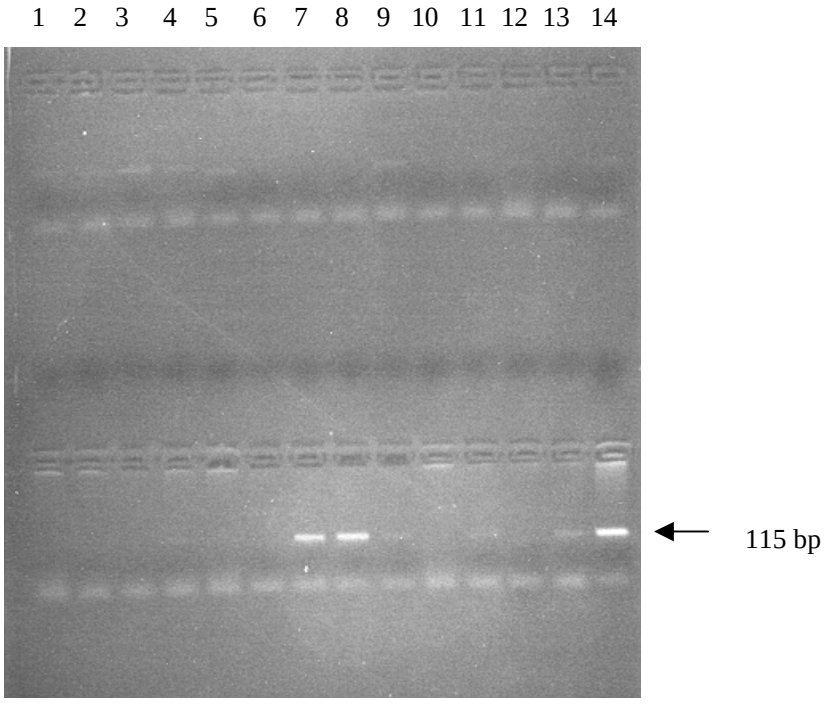
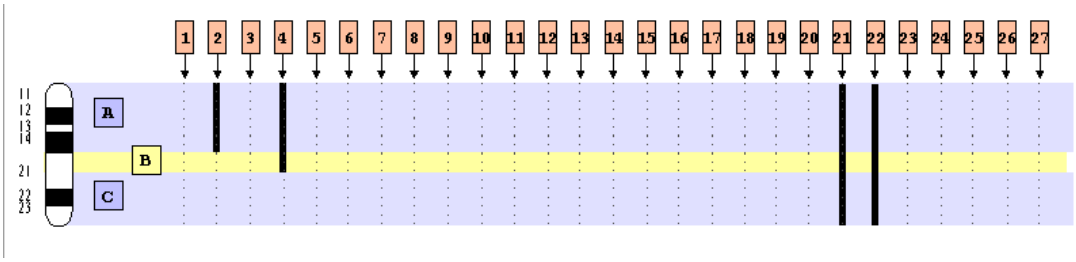


Fig 5.9 PCR results of the PERV-B junction fragment in the 27 somatic cell hybrid clones from the French panel, separated in a 2% agarose gel. Clones 1-19, hybrid cells with Chinese hamster background; Clones 20-27, hybrids with mouse background. Note that the hybrid clones 21 and 22 are positive for the 115 bp fragment, as is the pig control DNA template (P). Mouse and hamster controls were run on a separate gel and no PCR products were obtained under the same PCR conditions.



Pig chromosome 17

Fig 5.10 Diagram representing the presence of fragments of porcine chromosome 17 in each hybrid clone deduced from cytogenetic and genetic analysis of the hybrid panel (Yerle *et al.*, 1996). Note that amplification in clones 21 and 22 is consistent with the site being located in region C (ie 17q ½2.1 - q2.3).

5.5. Discussion

To investigate the distributions of PERVs in the Westran line of pigs, PERV-A and PERV-B envelope probes were hybridised onto the metaphase chromosome spreads of two highly inbred and closely related animals. The FISH results show that there are at least 13 PERV-A and 9 PERV-B definite locations and a further 6 PERV-A and 4 PERV-B possible locations. The higher number of locations for PERV-A than PERV-B is more like the pattern that Le Tissier *et al.* (1997) found for European pigs than for Asian pigs and is contrary to the expectation from analysis of Westran mitochondrial DNA sequence. However the European versus Asian distribution of PERV-A and PERV-B was not accurately estimated by Le Tissier *et al.* (1997) and some Asian breeds could possibly have more PERV-A than PERV-B. Alternatively the Asian mitochondrial DNA found in Westran pigs may be the result of introgression into a predominantly European nuclear genetic background. Le Tissier *et al.* (1997) showed that PERV-A proviruses are present at between 10 to 23 copies and PERV-B between 7 to 12 copies in different pig breeds, providing without any information of PERV locations. On average, inbreeding is as likely to cause loss of a PERV site as it will fixation. Thus an inbred line might be expected to have fewer sites compared with an outbred animal. However, there is a possibility that some PERV locations could not be detected by the method used in the present studies because of mismatch between the probe and target causing weak hybridisation signals.

FISH photos are not presented because it is impossible to illustrate representative hybridisation. The average number of grains per site from 20 metaphase cells is about 17 for PERV-A in animal 167 and the highest number of grains per site is 46 on 1q4.1. If homozygous

hybridisation signals are present for all 20 cells, then the maximum possible number of grains per site is 80 (20 cells X 2 chromosomes X 2 chromatids). Therefore in most cells, particular specific sites are not represented.

Comparison of the FISH results between animal 115 and 167 indicates consistent hybridisation patterns except for four locations detected with the PERV-B probe (8p2.2, 13q4.1, Xp1.3, Yq) (Fig 5.6 and Fig 5.7). Based on the high level of inbreeding and close relationship of these animals, these differences are unlikely to be due to differences in the presence of hybridisation targets. In each case, the sites are present in animal 167 which has an overall higher signal intensity and absent from 115. Thus it would appear that they slipped below the threshold for detection in animal 115, due to the lower efficiency of hybridisation and/or signal detection in this animal.

The peaks over 12p1.3 and Yp1.1 with the PERV-A probe (Fig 5.5) and 17q2.1 with the PERV-B probe (Fig 5.8) are quite broad compared other narrow peaks such as the PERV-A peak at 13q4.1. These locations may result from more than one PERV insert, even though they have been counted as one single PERV location at the point of the highest (major) peak. Clusters of retroviral integration sites have been reported in three different chromosomal locations, 1q2.3, 3p1.1 and 7p1.1, in pig (Rogel-Gaillard *et al.*, 1999). Similar retroviral clusters have been reported in mouse (Frankel *et al*., 1990; Tereba, 1983) and chicken (Tereba, 1983).

A PERV-B site has been mapped to 7p1.2 (Fig 5.8), very close to the known location of the swine major histocompatibility complex (SLA complex) class I region. Two distinct PERV-B integration loci, position 7p1.1 and 7p1.2-1.1, were identified very close to the SLA complex using BAC clones as FISH probes. Therefore it is very likely that this is one of the SLA-close PERV-B locations mapped by Rogel-Gaillard *et al*. (1999) (Table 5.12).

Hybridisation of PERV-A and PERV-B probes to the same site is also possible. There are five possible sites (5q2.1, 9q2.4, 13q4.1, 16q2.1, Yp1.1) showing hybridisation peaks in the same chromosomal locations with PERV-A and PERV-B probes (Fig 5.5 and Fig 5.8). There are three possible explanations. First, PERV-A and PERV-B insertions may lie close to each other. Second, there are small regions of highly conserved sequence between PERV-A and PERV-B probes where the forward and reverse primers are located (See Appendix 4.1 for the sequence alignment between PERV-A and PERV-B *env* gene). This could obviously contribute to some cross-hybridisation although the cross signal would be expected to be very small. Third, recombinant PERVs (Fig 4.8) will also generate hybridisation signals with both probes, likely to be more equal in intensity. Sequence analysis of PERV clones (Chapter 4) has shown the existence of three possible types of recombinant PERV clones. These three sites might correspond to these three different recombinant PERVs.

When the PERV FISH signals were counted, there was some confusion over the identity of chromosome 5 and 10. These chromosomes are of similar size and their R-banding pattern is very similar. This may have caused some mis-allocation of the hybridisation signal between the chromosomes. Similarly, the identification of arms was difficult for chromosome 12. The sizes of the p-arm and q-arm are very similar for SSC12 and the R-banding pattern is almost identical. However, for all other chromosomes, there was no confusion of chromosome or arm identification (Webb, pers. comm.).

The PERV-A locations in Westran pigs, compared with Large White pig, are quite different (Table 5.11). Rogel-Gaillard *et al.* (1999) reported 8 PERV-A locations on four different chromosomes (1, 8, 13, and Y) of their Large White pig. On the other hand, consistent significant PERV-A sites are observed on 13 different chromosomes in Westran pigs. The PERV-A locations on 10 chromosomes (2, 3, 5, 6, 7, 9, 12, 16, 17, and X) are so far unique to Westran pigs. However, a PERV-A site on the Y chromosome seems to be the same in Large Whites and the Westran line. On chromosome 1, there are two significant PERV-A sites (1q1.2 and 1q1.8) in Westran pig and 3 sites (1q2.1, 1q2.3, 1q2.4) in the Large White pig. Of these, only the site at 1q1.8 in Westran and 1q2.1 in the Large White pig could possibly be the same and mis-allocation is highly unlikely given that the sites are far away. It is possible that the significant PERV-A site on 13q4.1 in Westran pigs could be the same as that in the Large White pig allocated to the adjacent location at 13q4.2.

About half of the PERV-B locations appear to be located in the same or adjacent chromosomal bands in Westran and Large White (Table 5.12). They are the sites on chromosomes 7, 9, 13, and 17. Seven different chromosomes (chromosomes 1, 5, 8, 12, 16, X and Y) have PERV-B sites in Westran pigs which are absent in Large White. Two chromosomes (chromosomes 4 and 10) have PERV-B inserts in Large White which are absent in Westran pig. For chromosomes 11 and 14, a single PERV-B site has been identified in Westran and Large White. However, the locations are so far apart (14q1.3 in Westran and 14q2.8 in Large White), being located on different arms for chromosome 11 (11p1.3 in Westran and 11q1.4 in Large White), that they clearly represent different insertions in Large White and Westran pigs. Thus as expected, different pig breeds have some identical PERV insertions in their genome but also have other unique locations.

Laboratory was not set up for Southern hybridisation at the time these experiments were performed. Further it was hoped that sequence and FISH analysis would give more definitive results than Southern hybridisation. It would therefore be desirable in future to use Southern hybridisation to attempt an independent estimate of PERV copy number in Westran pigs.

Finally it has been demonstrated that a strategy for PCR amplification of PERV insertion junctions without cloning, employing PERV and SINE primers, actually works. A physical

location of a genuine junction PERV fragment was obtained using a somatic cell hybrid panel. The mapping data for this first junction fragment obtained in this way is consistent with an identified PERV-B site on porcine chromosome 17. In the very near future, this method will be extended to all PERV sites in the genome so that ultimately simple PCR genotyping tests can be devised using the unique sequence tagged site and PERV sequence to test for the presence of all PERVs in the genome.

Chapter Six

Discovery of Endogenous Retrovirus in Collared Peccaries (*Tayassu tajacu*)

6.1. Introduction

Peccaries belong to the order Artiodactyla (even-toed ungulates) and to the suborder Suiformes, in which three families are listed, namely Hippopotamidae (hippos), Tayassuidae (peccaries) and Suidae (pigs) (Ruvinsky and Rothschild, 1998; Miyamoto *et al*., 1993). They look like small hairy pigs, have a snout disc, long hair, thin legs, small eyes, and small rounded ears (Fig 6.1). Fossil records of Tayassuidae have been found in Eurasia and even in Africa suggesting that they were separated from Suidae not later than in the Oligocene era (30 million years before present, MYBP). However, modern peccary species live only in the Americas (Ducrocq, 1994; Hendey, 1976). The family Tayassuidae seems to have originated in South East Asia and then migrated into the New World, because the most morphologically primitive fossils are found in Thailand (Ducrocq, 1994).

According to the most recent classification, the Tayassuidae family includes three extant species; Collared peccary (*Tayassu tajacu* or *Pecari tajacu*), White-lipped peccary (*Tayassu pecari*) and Chacoan peccary (*Catagonus wagneri*) (Ruvinsky and Rothschild, 1998). The diploid chromosome numbers of Collared peccary, White-lipped peccary and Chacoan peccary are 30, 26 and 20 respectively (Benirschke *et al*., 1985; Benirschke and Kumamoto, 1989; Hufty *et al*., 1973). Controversy remains about the chromosome evolution of peccary species. Anatomical evidence suggests that the Chacoan peccary is the most primitive species compared to the Collared peccary and White-lipped Peccary. This suggests that the chromosome number of 26 for the White-lipped Peccary and 30 for the Collared peccary have evolved from the lower Chacoan peccary number through chromosomal fission (Todd, 1985). On the other hand, others have proposed that chromosomal fusion has reduced the large chromosome number found in the Collared peccary to smaller numbers in the White-lipped and Chacoan peccaries (Benirschke and Kumamoto, 1989; Hufty *et al*., 1973).

The phylogenetic relationships among species within the Suiformes have been examined using sequence data from the gene for 12S rRNA. The results show monophyly of pigs and peccaries within the order Artiodactyla (Douzery and Catzeflis, 1995). Mitochondrial DNA sequence has also been used. The estimated divergence time between Suidae and Tayassuidae, based on the analysis of mitochondrial cytochrome b sequences, is 33 to 37 MYBP (Theimer and Keim, 1998) and 31 to 33 MYBP (Randi *et al*., 1996). The analysis of SINE (Short Interspersed Nuclear Element) and LINE (Long Interspersed Nuclear Element) elements has confirmed the 12S rRNA results, namely that pigs and peccaries form a monophyletic group to the exclusion of hippopotamuses (Nikaido *et al*., 1999). Analysis of SINE sequences has suggested that the separation of Tayassuidae and Suidae occurred about 43.2 MYBP (Yasue and Wada, 1996). However, by contrast, recent analysis of 13 swine PRE-1 (a swine SINE) loci in peccaries suggested the divergence time between the two families was about 16.8 MYBP (Sulandari *et al*., 1997), though this divergence time is not consistent with the fossil records described above.

The C-type retroviruses are widespread as both endogenous and exogenous agents within mammalian species including humans (Tristem *et al*., 1996b). Tristem *et al*. (1996b) designed degenerate primers based on conserved motifs in the protease and reverse transcriptase gene of Murine leukemia viruses (MLVs) for amplifying retroviral sequences in a variety of mammalian species. Other degenerate primers were also designed based on conserved motifs of the retroviral protease and reverse transcriptase protein for the same purpose (Tristem *et al*., 1996b). These primers are capable of amplifying retroviruses and long terminal repeats (LTR) retrotransposons in wide range of vertebrate taxa including mammals, reptiles, amphibians and fish (Herniou *et al*., 1998; Martin *et al*., 1997; Tristem, 1996; Tristem *et al*., 1996a).

Based on very early work of DNA hybridisation, which has limited sensitivity when the probe and target DNA have diverged, it has been claimed that peccaries do not contain virogene sequences related to the pig and mouse type C-virus (Benveniste and Todaro, 1975). Related viruses were detected by this method in members of the Suidae and in the mouse and this was interpreted as evidence for the horizontal transfer of retroviruses from mouse to pig before the divergence of wart hogs, bush pigs and domestic pigs from a common ancestor, but after divergence of the Suidae and Tayassuidae lineages. It was claimed that this transfer from mouse to the pig lineage occurred about 5 to 10 million years ago. A specific objective of this study was to re-examine peccaries for evidence of C type retroviruses.

Mr. Jaime Góngora visited the University of Sydney from Colombia for three months in 1999 to pursue his interest in the cytogenetics and molecular genetics of peccaries. He brought Collared Peccary (*Tayassu Tajacu*) genomic DNAs for his research and provided access to these

samples for this research. This provided an opportunity to re-evaluate the older studies on retroviruses in peccaries using modern molecular methods.

The aim of the work reported here is to investigate the existence and nature of endogenous retroviral sequences in the Collared peccary using the degenerate primers of Tristem *et al*. (1996b).

Fig 6.1 A Collared Peccary (*Tayassu tajacu*) (Courtesy of Jaime Góngora).

6.2. Materials

6.2.1. The Peccary genomic DNA samples

Five Collared peccary genomic DNAs were supplied by Mr. Jaime Góngora who was a visiting scholar in University of Sydney from Colombia for 3 months in 1999. They were collected in 1999 at Santa Cruz Zoo which is located near Bogota, Colombia.

6.2.2. Primers

6.2.2.1. Degenerate primers

The degenerate primers are based on two motifs conserved in a representative set of C-type retroviruses, especially in MLVs. Retroviral sequences have been selectively and specifically amplified in many different mammalian species using these primers (Tristem, 1996; Tristem *et al*., 1996b). The primer sequences are presented in Table 6.1. The primers were synthesised by Life Technologies Pty Limited, Victoria, Australia.

Table 6.1 The primer sequences, PCR conditions ($MgCl_2$ concentration and annealing temperature) and two conserved protein motifs from MLVs. The motif LVDTGA is in the protease protein and the motif QYVDDL is in the reverse transcriptase.

Primer sequences[1]	$MgCl_2$ concentration	Annealing temp.	Peptide motifs
F. 5'-(C/T)TI(T/G)TIGA(T/C)ACIGGIGCI(G/C)A-3' R. 5'-AGIAGGTC(A/G)TCIAC(A/G)TA(C/G)TG-3'	1 mM	47 °C	LVDTGA QYVDDL

[1]F: forward primer; R: reverse primer; I in the primer sequences stand for inosine

6.2.2.2. Internal primers

Based on the peccary retroviral sequence information obtained during the present study, internal primers were designed to amplify peccary-specific MLV-homologous PCR products. The primer information is presented in Table 6.2 and the locations of each primer are shown in Fig 6.4.

Table 6.2 The homologous primer sequences, PCR conditions ($MgCl_2$ concentration and annealing temperature) of the internal primers.

Primer sequences[1]	$MgCl_2$ concentration	Annealing Temperature
F. 5'-CTACCGGAGTCAGACGTTAC-3' R. 5'-TTCGATACTCACTGAGGTCC-3'	2 mM	60 °C

[1]F: forward primer; R: reverse primer

6.3. Methods

6.3.1. PCR conditions

6.3.1.1. PCR conditions for amplifying MLV homologues using degenerate primers

PCR was carried out in a 25 µl volume with approximately 100 ng of template genomic DNA, 0.4 mM of each primer, 600 µM of each dNTP, 1 mM $MgCl_2$, 10 × PCR buffer, and 2.5 units of *Taq* polymerase. This mixture was overlaid with one drop of mineral oil. Amplification was performed in a PTC-100 Programmable Thermal Controller (MJ Research Inc.) for 45 cycles with denaturation for 1 minute at 95°C, annealing for 30 seconds at 47°C and extension for 1 minute at 72°C. The final extension was for 20 minutes at 72°C.

6.3.1.2. PCR conditions for internal primers

AmpliTaq Gold *Taq* polymerase (Perkin Elmer) was employed. Amplification was performed with initial denaturation for 5 minutes at 95°C, followed by 35 cycles with denaturation for 30 seconds at 95°C, annealing for 30 seconds at 60°C and extension for 1 minute at 72°C. The final extension was for 20 minutes at 72°C.

6.3.1.3. Cycle sequencing reaction

DNA fragments for cycle sequencing were generated using a PTC-100 Programmable Thermal Controller (MJ Research Inc.) with two M13 dye-labelled vector primers (Table 4.2) as described in Section 2.10.2.

6.3.2. Purifying PCR products from agarose gels

To get clean PCR products for cloning, PCR bands were isolated from the agarose gel slice using a BRESAclean™ DNA purification kit (Bresatec Inc.) as described in Section 2.4.3.1.

6.3.3. Cloning of PCR products

The peccary retroviral PCR products were cloned into the pCR®2.1-TOPO plasmid vector (Invitrogen). The detailed information for cloning has been described in Section 2.11 and a detailed map of the cloning vector is presented in Appendix 2.1.

6.3.4. Quick PCR method to check clones for correct inserts

To check whether the plasmid clones had an appropriate insert, preparations of DNA were made by boiling as described in Section 2.7.3 and PCR was carried out with *Taq* DNA polymerase as in section 4.3.1.1.

6.3.5. Extraction of plasmid DNA from bacterial cells

After identifying clones with the insert, high quality plasmid DNA was extracted using the alkaline minilysis method or a JETquick Plasmid Miniprep Spin Kit (GENOMED Inc.) as described in Section 2.4.2.

6.3.6. Restriction enzyme digestion

The plasmid DNA with insert was further verified with *Eco*RI restriction enzyme digestion as described in Section 2.8.

6.3.7. Sequencing with Li-Cor sequencer

A *SequiTherm* EXCEL™ Long-Read™ DNA sequencing kit (Epicentre) and Li-Cor sequencer (Model 4200, Li-Cor Inc.) were used to read peccary insert sequences, using two vector primers (Li-Cor Inc.) labeled with different dyes (Table 4.2). Sequences were analyzed using Base ImageIR™ software version 4.1 (Li-Cor Inc.). The detailed protocols for sequencing using a Li-Cor sequencer are described in Section 2.10.2.

6.3.8. Analysis of sequence data

DNA sequences were stored and analysed using the programs provided by the Australian Genomic Information Service (ANGIS). The most similar sequences in GenBank were identified

by FASTA database searching (Pearson and Lipman, 1988). The GAP program (Needleman and Wunsch, 1970) was used to compare peccary sequences in detail with the most similar sequences identified by FASTA searching. The GAP program considers all possible alignments and gap positions and creates the alignment with the largest number of matched bases and the fewest gaps.

6.4. Results

6.4.1. PCR amplification for retroviral sequences in Collared peccary

The degenerate primer pairs (Table 6.1) were tested in five different collared peccary genomic DNAs. The agarose gel shows that the PCR reactions produced two bands in some peccary samples (Fig 6.2). The 434 bp product was common to all templates but the 840 bp products amplified in only two peccary DNA samples. The 840 bp PCR product was found to amplify variably depending upon the PCR conditions (data not shown). Most mammalian species, including pig, gave only a 950 bp product with these degenerate primers (data not shown) as described by Tristem *et al*. (1996b). Therefore, neither the 434 nor 840 bp PCR products were expected as correctly amplified retroviral fragments.

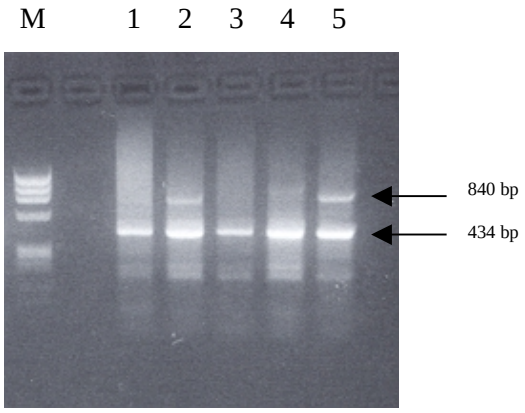


Fig 6.2 PCR products of five Collared Peccaries amplified by degenerate primers (Table 6.1). A 434 bp product is common to all amplifications but the 840 bp product occurs in only two samples (lane two and five). Marker (M) is ϕX174/*Hae*III Marker (Promega) size standard.

6.4.2. *Checking the insert by* Eco*RI Restriction digestion*

After cloning, *Eco*RI digestion was used to release the insert from the vector to confirm its size. The recognition site of this enzyme is located in the multi-cloning site and can be used to excise the DNA insert (The map of vector is shown in Appendix 2.1). Fig 6.3 shows the *Eco*RI restriction enzyme digestion pattern of a series of peccary clones derived from the PCR product of lane number 5 in Fig 6.2. The attempt to clone the "434" bp PCR products gave inserts of varying size (Fig 6.3 lane 1 to 10). Only one clone with the 840 bp insert was analysed (Fig 6.3 lane 11). The clones from the putatively 434 bp PCR product of lane 5 Fig 6.2 are designated as clones S1 to S10 based on their order in Fig 6.3. The 840 bp insert clone (Fig 6.3 lane 11) is designated as clone L1.

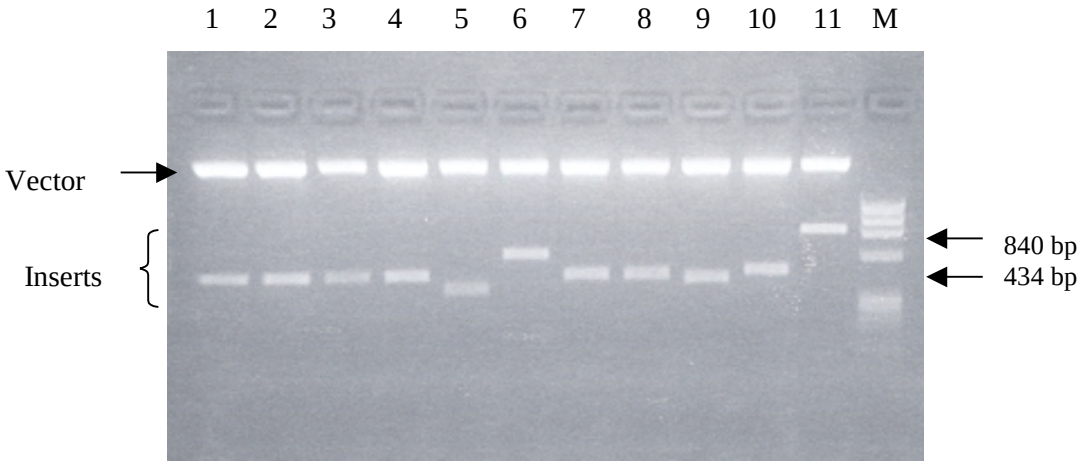


Fig 6.3 *Eco*RI restriction digestion of peccary clones. Lanes 1 to 10 (designated as clones S1 to S10) are the putatively 434 bp insert clones. Lane 11 (designated as clone L1) is the putative 840 bp insert clone. M is ϕX174/*Hae*III Marker (Promega) size standards.

6.4.3. *Sequences of peccary retroviral clones*

Six peccary putative 434 bp clones (clone S1, S2, S5, S6, S9, S10) which were selected to cover all insert sizes and the one 840 bp clone (clone L1) were fully sequenced. The FASTA sequence search results showed that clones S1 and S2 were the only clones with strong matches to retroviral sequences. They were both well matched with known murine leukemia virus sequences in GenBank (Table 6.3), demonstrating that at least some of the putative 434 bp PCR products are from endogenous retroviral sequences in Collared peccary. Interestingly, the FASTA

results show that the peccary retroviral sequences are better matched with mouse sequences than with pig retroviruses, although the match with pig retroviral sequence is also excellent.

The alignments of the two peccary retroviral sequences are presented in Fig 6.4, showing that there are eight base-pair differences between these two clones. The 0.4 % error rate for the Li-Cor sequencer has been calculated previously (see Section 4.4.3.2) from analysis of very long (~1.8 kb) sequence reads from porcine endogenous retrovirus clones. However, most of these sequencing errors are located at the end of the sequences, after 600 – 800 bp. These shorter peccary retroviral sequences are fully overlapped in both forward and reverse directions and thus are confirmed double pass sequences. Thus it can be assumed that these peccary retroviral sequences are truly different and represent multiple copies, as in other species.

Alignment of one of the peccary endogenous retroviral sequences with the GenBank nucleotide sequences shows that it is not continuously matched with the mouse C-type retroviral sequences (Fig 6.5). The sequence from the beginning to 319 bp is matched closely with mouse retroviral sequence but the match disappears abruptly at this point (Fig 6.5a). When the peccary sequence after 319 bp is separated and used to screen GenBank, this remaining sequence also matches best with the same mouse C-type retroviral sequence (Fig 6.5b) but further along the sequence. There is a large deletion of the peccary endogenous retroviral sequence compared to the mouse and other species retroviruses. In fact, these two peccary retroviral sequences have a 534 bp deletion from nucleotide number 2627 to 3160 of the total 8282 nucleotide gag-*pol*-*env* gene compared with published murine leukemia retrovirus sequences (Fig 6.5a, b).

The sequence search results for the 840 bp insert clone (clone L1) show that this PCR product is not derived from an endogenous retrovirus. The only match in GenBank is with an unknown sequence in human. The sequence information and sequence search results for this clone are presented in Appendix 6.1. It is always possible that non-target sequences will be amplified when degenerate primers are used. Similarly the four other clones containing the putative 434 bp insert, namely S5, S6, S9 and S10, failed to match with any retroviral sequences in FASTA sequence searches of GenBank. The information on these non-retroviral inserts is fully described in Section 6.4.5.

Table 6.3 Comparison of the two peccary putative retroviral sequences with the nucleotide sequences in the Genbank database (both sequences gave exactly the same output).

The best scores are:		initn	init1	opt	z-sc	E(1073989)
MLVGPN Murine leukemia virus N-tropic st	(1390)	437	437	613	672.9	1e-30
MLVGPB Murine leukemia virus B-tropic st	(1390)	437	437	613	672.9	1e-30
AF034782 Synthetic helper virus genomic	(5798)	596	459	604	655.1	2.5e-30
A47081 Sequence 1 from Patent WO9527063	(8889)	591	428	599	647.1	4.4e-30
A47083 Sequence 3 from Patent WO9527063	(8889)	591	428	599	647.1	4.4e-30
MLOCG AKV murine leukemia virus, complet	(8374)	565	428	599	647.5	4.5e-30
A47085 Sequence 5 from Patent WO9527063	(8352)	556	428	599	647.5	4.5e-30
MLVPOGAEN Murine Leukemia virus gag gene	(8259)	565	428	599	647.5	4.5e-30
AF136491 Murine leukemia virus erv1 poly	(3539)	565	428	599	652.2	5.8e-30
AF169256 SL3-3 murine leukemia virus, co	(8377)	547	419	595	643.0	8e-30
MMU63133 Mus musculus C-type ecotropic e	(8274)	547	419	595	643.1	8e-30
MLU13766 Murine leukemia virus MCF1233,	(8196)	573	445	559	603.1	1.4e-27
RETCG Retroviridae complete genome (muri	(8135)	490	362	559	603.2	1.4e-27
MLVRT12 Murine leukemia related virus ge	(945)	746	502	551	606.1	8e-27
AF038600 Sus scrofa porcine endogenous r	(8132)	440	370	544	586.5	1.2e-26
A42090 Sequence 2 from Patent WO9501447	(8323)	576	403	542	584.1	1.5e-26
REFMLVCGD Friend murine leukemia virus F	(8323)	576	403	542	584.1	1.5e-26
I50881 Sequence 8 from patent US 5643756	(8323)	576	403	542	584.1	1.5e-26
MULV13893 Murine leukemia virus RNA for	(8282)	567	403	542	584.2	1.5e-26
MLFCG Friend murine leukemia virus, comp	(8282)	567	403	542	584.2	1.5e-26
MLMPOLA Moloney murine leukemia virus po	(1680)	468	403	542	592.9	2.4e-26
A66553 Sequence 3 from Patent WO9740167	(8209)	480	357	537	578.6	3.1e-26
A66552 Sequence 2 from Patent WO9740167	(8196)	480	357	537	578.7	3.1e-26
AF038601 Sus scrofa porcine endogenous r	(7333)	480	357	537	579.3	3.2e-26
D88386 Friend murine leukemia virus comp	(8358)	567	403	533	574.1	5.5e-26
RMU94692 Rauscher murine leukemia virus,	(8282)	567	394	533	574.2	5.5e-26
PERY17013 Porcine endogenous retrovirus	(7808)	471	348	528	568.9	1.1e-25
MLVENVR MuLV (strain RadLV/VL3(T+L+)) RN	(8394)	485	366	524	564.1	2e-25
MUSGAGPOEN Mouse DNA with endogenous mur	(4217)	376	376	524	567.9	2.4e-25
REFMLVCG Friend murine leukemia virus (F	(8359)	558	385	515	554.1	7.2e-25
AF019230 Murine leukemia virus strain SR	(8256)	413	368	515	554.2	7.2e-25
RSVTRANSA Rat sarcoma virus transduction	(4480)	306	306	514	556.4	9.9e-25
RSVVRASX Rat sarcoma virus V-ras oncogen	(4480)	306	306	514	556.4	9.9e-25
AF038599 Sus scrofa porcine endogenous r	(8132)	274	138	511	549.8	1.3e-24
MLVRT10 Murine leukemia related virus ge	(942)	548	393	510	560.6	2.8e-24
MLVGAG Rat leukemia virus gag (gag), pol	(8107)	235	205	505	543.1	3e-24
CASBREML CAS-BR-E murine leukemia virus,	(8231)	476	357	490	526.4	2.5e-23
SSU77600 Sus scrofa domestica polyprotei	(927)	473	370	494	542.9	2.7e-23
MLVRT7 Murine leukemia related virus gen	(948)	459	288	487	535.0	7.4e-23
SSU77599 Sus scrofa polyprotein gene, pr	(927)	480	357	487	535.1	7.4e-23
AF010170 Plasmid pAMS with hybrid amphot	(11328)	511	339	479	512.4	1.1e-22
AF033811 Moloney murine leukemia virus,	(8332)	476	339	479	514.1	1.2e-22
MLMCG Moloney murine leukemia virus, com	(8332)	476	339	479	514.1	1.2e-22
A60206 Sequence 2 from Patent WO9708330	(7616)	511	339	479	514.6	1.2e-22
A60207 Sequence 3 from Patent WO9708330	(7308)	476	339	479	514.8	1.3e-22
A60208 Sequence 4 from Patent WO9708330	(7308)	476	339	479	514.8	1.3e-22
SSAJ5410 Sus scrofa DNA for endogenous r	(276)	388	342	471	524.0	1e-21
SSY18746 Sus scrofa DNA for endogenous t	(276)	342	342	470	522.8	1.2e-21
SSAJ5403 Sus scrofa DNA for endogenous r	(276)	333	333	470	522.8	1.2e-21
FCVGP Feline leukemia virus (FeLV-B) gag	(2565)	282	232	463	502.8	1.7e-21

Sequences sorted according to their “opt” score.
Opt score = alignment score between the query sequence and the database sequence
Initn, init1 = intermediate alignment scores
z-sc = z opt score
E = statistical expectation for this sequence. The number of sequences with this identity expected in a database of this size by chance alone. An E value less than 0.05 is considered statistically significant.
Reference: ANGIS: http://www.angis.org.au

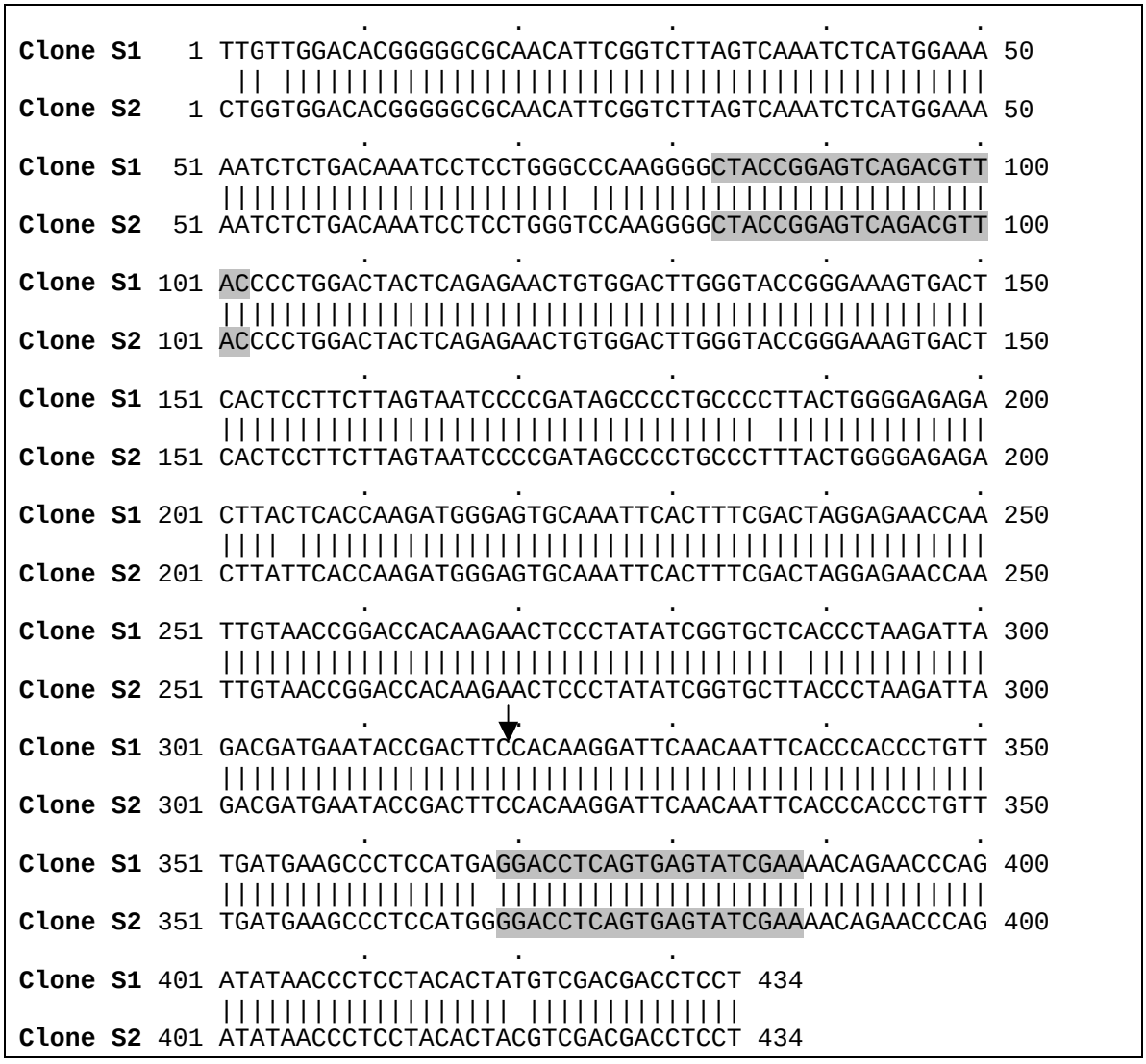


Fig 6.4 Alignment of two peccary retroviral sequences. The clone S1 sequence is shown above and the clone S2 sequence below. They are the identical in length (434 bp) and show over 98% (426/434) sequence similarity. The positions of the internal primers (see sections 6.2.2.2 and 6.4.4) are shaded (…). The position of the 534 bp deletion in the peccary retroviral sequences relative to mouse sequences (see Fig 6.5) is shown with an arrow (↓).

Fig 6.5 Separate alignments of the first 319 bp and the final 115 bp of one of the 434 bp peccary endogenous retroviral sequences (clone S2) with the same murine leukemia virus sequence (GenBank accession number MULV13893). After 319 bp, there is complete loss of similarity in the sequence alignment (a). The final 115 bp from nt 320 to 434 re-establishes a good match with this same MuLV sequence further downstream (b). Relative to the MuLV sequence, this peccary endogenous retrovirus has a 534 bp deletion. The final 115 bp sequence is shown shaded (…).

(a)

```
MULV13893 Murine leukemia virus RNA for gag-pol-env pol (8282 nt)

 initn: 567 init1: 403 opt: 542 Z-score: 584.2 expect() 1.5e-26
 63.975% identity in 322 nt overlap

                                                 10        20        30
Clone S2                                  CTGGTGGACACGGGGGCGCAACATTCGGTC
                                          :: ::::: :: ::::: ::::: :: ::
  MULV13 CTCAAAGTCGGGGGGCAACCCGTCACCTTCCTAGTGGATACTGGGGCCCAACACTCCGTG
       2280      2290      2300      2310      2320      2330

                40        50        60        70        80        90
Clone S2 TTAGTCAAATCTCATGGAAAAATCTCTGACAAATCCTCCTGGGTCCAAGGGGCTACCGGA
          :   : ::  :: ::::    :   :::::: ::  :::::::::::::::::: :::
  MULV13 CTGACCCAAAATCCTGGACCCCTAAGTGACAAGTCTGCCTGGGTCCAAGGGGCTACTGGA
       2340      2350      2360      2370      2380      2390

               100       110       120       130       140       150
Clone S2 GTCAGACGTTACCCCTGGACTACTCAGAGAACTGTGGACTTGGGTACCGGGAAAGTGACT
         :  :  :: :: : :::::: ::  :  :    ::: :: : :  ::::: :: :: ::
  MULV13 GGAAAGCGATATCGCTGGACCACGGATCGCCGAGTGCACCTAGCCACCGGTAAGGTCACC
       2400      2410      2420      2430      2440      2450

               160       170       180       190       200       210
Clone S2 CACTCCTTCTTAGTAATCCCCGATAGCCCCTGCCCTTTACTGGGGAGAGACTTATTCACC
         ::::: ::: :     : :: ::  ::::::  ::: : :: :: ::::: ::  : ::
  MULV13 CACTCTTTCCTCCATGTACCAGACTGCCCCTATCCTCTGCTAGGAAGAGATTTGCTGACT
       2460      2470      2480      2490      2500      2510

               220       230          240       250       260
Clone S2 AAGATGGGAGTGCAAATTCACTTT---CGACTAGGAGAACCAATTGTAACCGGACCACAA
         ::  :   ::  ::::::::::::    ::  :::::  :   ::::    ::::::
  MULV13 AAATTAAAAGCCCAAATTCACTTTGAGGGATCAGGAGCTCAGGTTGT---GGGACCAATG
       2520      2530      2540      2550      2560         2570

         270       280       290       300       310       320
Clone S2 GAACTCCCTATATCGGTGCTTACCCTAAGATTAGACGATGAATACCGACTTCCACAAGGA
         : ::  ::  :    ::::: :::::::   :::: ::::: :: :: :: :
  MULV13 GGACAGCCCCTGCAAGTGCTGACCCTAAACATAGAAGATGAGTATCGGCTACATGAGACC
           2580      2590      2600      2610      2620      2630

         330       340       350       360       370       380
Clone S2 TTCAACAATTCACCCACCCTGTTTGATGAAGCCCTCCATGGGGACCTCAGTGAGTATCGA

  MULV13 TCTAAAGGGCCAGATGTGCCTCTAGGGTCCACATGGCTCTCTGATTTTCCCCAGGCCTGG
           2640      2650      2660      2670      2680      2690
```

(b)

```
MULV13893 Murine leukemia virus RNA for gag-pol-env pol (8282 nt)
 initn: 184 init1: 184 opt: 278 Z-score: 325.0 expect() 4.1e-12
 71.304% identity in 115 nt overlap

                                              330       340       350
Clone S2                                CACAAGGATTCAACAATTCACCCACCCTGT
                                        :::: :: ::::: ::    ::::::::::
  MULV13 TCTCAGGACAATTAACCTGGACCAGACTCCCACAGGGTTTCAAAAACAGTCCCACCCTGT
               3140      3150      3160      3170      3180      3190

                360       370       380       390       400       410
Clone S2 TTGATGAAGCCCTCCATGGGGACCTCAGTGAGTATCGAAAACAGAACCCAGATATAACCC
         ::::::: ::::: ::  :::::::::   :: :  :: :  ::: :::::::  : : ::
  MULV13 TTGATGAGGCCCTGCACAGGGACCTCGCAGACTTCCGGATCCAGCACCCAGACCTGATCC
               3200      3210      3220      3230      3240      3250

                420       430
Clone S2 TCCTACACTACGTCGACGACCTCCT
         : :: :: :: :: :: ::: : ::
  MULV13 TGCTCCAGTATGTAGATGACTTACTGCTGGCCGCCACCTCTGAGCTTGACTGTCAACAAG
               3260      3270      3280      3290      3300      3310
```

6.4.4. PCR with internal primers

To confirm the validity of the PCR product generated with degenerate primers and the existence of the 534 bp deletion in Collared peccary retroviral sequences, a new set of internal primers was designed (Table 6.2) from the new sequence. The internal primers are located either side of the 534 bp deletion break point. Only the expected 306 bp product has been amplified from four peccary DNAs. Interestingly it has been amplified (weakly) from pig control DNA but is absent in mouse DNA (Fig 6.6a, b). This deleted retrovirus is thus not closely related to any full length retroviruses in the peccary genome, but surprisingly appears to be detecting a similarly deleted endogenous virus in the pig. Given the sequence alignments shown in Fig 6.5, it is not surprising that these primers do not amplify a product from mouse template.

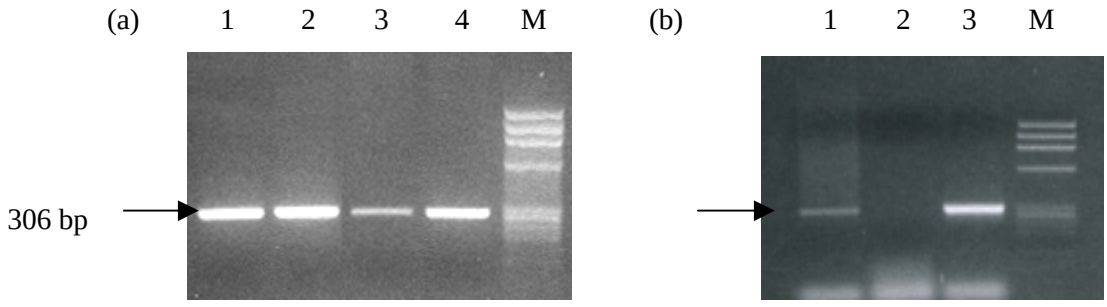


Fig 6.6 (a) The PCR products of four different Collared peccaries (from lane one to four) amplified by internal primers (Table 6.2). (b) The PCR product from pig (lane 1), mouse (lane 2) and one of the peccary DNAs (lane 3) amplified by internal primers. M is ϕX174/*Hae*III Marker (Promega) size standards.

6.4.5. The sequences of non-retroviral inserts

Four clones (clone S5, S6, S9, S10) contained sequences of non-retroviral origin. Clones S6 and S9 do not closely match any known genes in GenBank, but do match sequences of unknown functions in human (see Appendix 6.2 for sequence alignments). An (AG)n microsatellite sequence was found in clone S6 (Appendix 6.2a).

The clone S10 sequence aligns best with the porcine cofilin gene, but also shows excellent alignment with human, mouse and rat cofilin sequences, implying that clone S10 contains part of the Collared peccary cofilin gene (Fig 6.7). The putative peccary cofilin sequence and the pig cofilin gene show 93% identity in nucleotide sequences after 190 bp. The sequence alignment with the pig cofilin cDNA sequence commences just after the initial methionine (ATG), indicating that the preceding porcine sequence is the 5' UTR. The corresponding peccary sequence commences with AG at the beginning of the alignment suggesting that this is the splice acceptor site at the end of a putative peccary cofilin intron. Thus it is assumed that the preceding peccary sequence is intronic (Fig 6.7).

The sequence of clone S5 is a SINE sequence having both forward and reverse complementary sequences in GenBank (See Appendix 6.3). Until now only three Collared peccary SINE sequences were available (Furutani *et al*., 1998; Sulandari *et al*., 1997). The comparison of these published peccary SINE sequences with S5 sequence is presented in Figure 6.8. The comparison among these peccary SINE sequences shows that they are not highly conserved, indicating that there are different types of SINEs in peccaries. Since PRE-1 is the most typical of porcine SINEs, Fig 6.9 shows the alignment of porcine PRE-1 with the S5 peccary SINE sequence. The sequence homology of pig and peccary SINEs is presented in Table 6.4. Interestingly all three peccary SINE sequences are more similar to the porcine PRE-1 SINE than they are to each other. The S5 peccary SINE sequence can be classified as a peccary PRE-1, and PRE-1 sequences in collared peccaries are more divergent than between peccaries and the pig sequence.

```
The best scores are:                                      initn init1 opt z-sc E(1017344)
    PIGCOFIL Pig cofilin mRNA, complete cds   (1390) 1116 1116 1147 1078.7 2.6e-53
    HSU21909 Human cofilin mRNA, partial cds  ( 960) 1080 1080 1101 1037.4 7.6e-51
    AW239319 xb38h05.y1 NCI_CGAP_Lu31 Homo s  ( 521) 1080 1080 1101 1040.7 9.1e-51
    HSNMCFL1 H.sapiens mRNA for non-muscle t  (1059) 1071 1071 1092 1028.3 2.2e-50
    AW249211 2821075.5prime NIH_MGC_7 Homo s  ( 643) 1071 1071 1092 1031.1 2.5e-50
    AW250264 2821137.5prime NIH_MGC_7 Homo s  ( 511) 1071 1071 1092 1032.3 2.7e-50
    AW247672 2820161.5prime NIH_MGC_7 Homo s  ( 504) 1071 1071 1092 1032.4 2.7e-50
    HUMCOF Human cofilin mRNA                 ( 501) 1071 1071 1092 1032.4 2.7e-50
    AW248338 2820462.5prime NIH_MGC_7 Homo s  ( 333) 1071 1071 1092 1034.7 3.1e-50
    AW322927 uo54b11.y1 NCI_CGAP_Lu29 Mus mu  ( 618) 1035 1035 1047 988.9 5.9e-48
    MUSCOF Mus musculus mRNA for cofilin, co  (1134) 1017 1017 1029 968.6 4.3e-47
    AW323540 uo73e01.y1 NCI_CGAP_Mam1 Mus mu  ( 623) 1017 1017 1029 971.9 5.2e-47
    AL022775 s7c19b51 Beddington mouse disse  ( 551) 1017 1017 1029 972.5 5.4e-47
    AW260598 um84a04.y1 Sugano mouse liver m  ( 542) 1017 1017 1029 972.6 5.4e-47
    RNCOFIL R.norvegicus mRNA for cofilin     (1039)  999  999 1020 960.6 1.3e-46


PIGCOFIL Pig cofilin mRNA, complete cds                 (1390 nt)
 initn: 1116 init1: 1116 opt: 1147 Z-score: 1078.7 expect() 2.6e-53
 92.857% identity in 266 nt overlap

         160       170       180       190       200       210
Clone S10  CCTTATAAGGTGCGTGTGGCTCGCCCTCTAGGCCTCCGGTGTGGCTGTCTCCGATGGGGT
                                          :::::::::::::::::::::: :: :::::
PIGCOF     TTGCTATCTCCTTTTCGCTTCCGGAAATATGGCCTCCGGTGTGGCTGTCTCTGACGGGGT
             340       350       360       370       380       390

         220       230       240       250       260       270
Clone S10  CATCAAAGTGTTCAATGACATGAAGGTGCGTAAGTCTTCTACACCGGAGGAGGTGAAGAA
           :::::::::::::::::::::::::::::::::: :: ::::::::::::::::::::
PIGCOF     CATCAAAGTGTTCAATGACATGAAGGTGCGTAAGTCCTCGACACCGGAGGAGGTGAAGAA
             400       410       420       430       440       450

         280       290       300       310       320       330
Clone S10  GCGCAAGAAGGCAGTGCTCTTCTGTCTGAGCGAGGACAAGAAGAACATCATCCTGGAGGA
           :::::::::::: :::::::::: ::::::::::::::::::::::::::::::::::
PIGCOF     GCGCAAGAAGGCGGTGCTCTTCTGCCTGAGCGAGGACAAGAAGAACATCATCCTGGAGGA
             460       470       480       490       500       510

         340       350       360       370       380       390
Clone S10  GGGCAAGGAGATCCTCGTGGGTGATGTGGGCCAGACTGTGGATGACCCCTACGCCACCTT
           ::::::::::::::::::::::: :::::::::::::: :: :::::::: ::::::::
PIGCOF     GGGCAAGGAGATCCTCGTGGGTGACGTGGGCCAGACTGTAGACGACCCCTATGCCACCTT
             520       530       540       550       560       570

         400       410       420       430       440       450
Clone S10  TGTCAAGATGCTGCCAGACAAGGACTGCCGCTACGGCCACTACGTCGACGACCTCCT
           :::::::::::::::: ::::::::::::::::: : :: ::: : :: : :::: :
PIGCOF     TGTCAAGATGCTGCCGGACAAGGACTGCCGCTATGCCCTCTATGACG-CCACCTACGAGA
             580       590       600       610       620        630
```

Fig 6.7 The GenBank search results and the best sequence alignment of clone S10 show that it contains peccary cofilin gene sequences. The start codon (shaded (ATG)) for the pig cofilin cDNA sequence shows that the preceding sequence is the 5' UTR. An apparent splice acceptor site at the end of a *cofilin* intron in the peccary sequence is indicated by a box (AG)

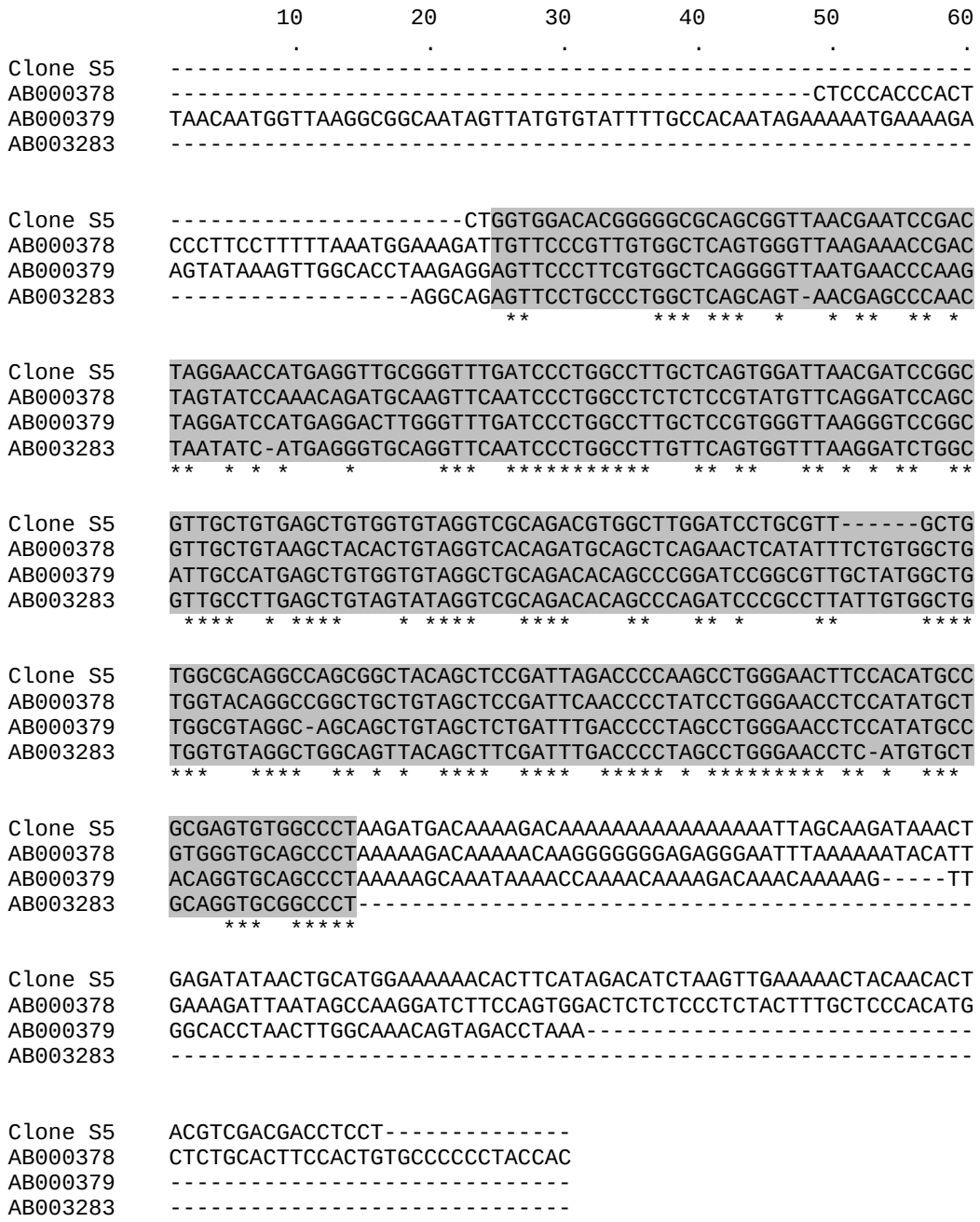


Fig 6.8 Alignment of three SINE sequences derived from Collared peccary. Published peccary SINE sequences (DDBJ accession numbers AB000378 and AB000379, Sulandari *et al.*, 1997; GenBank accession number AB003283, Furutani *et al.*, 1998) are aligned with a SINE sequence from clone S5 in this experiment. * indicates identical nucleotide sequences in the four SINEs. The shaded regions correspond to PRE-1 from pigs as defined by Singer *et al.* (1987).

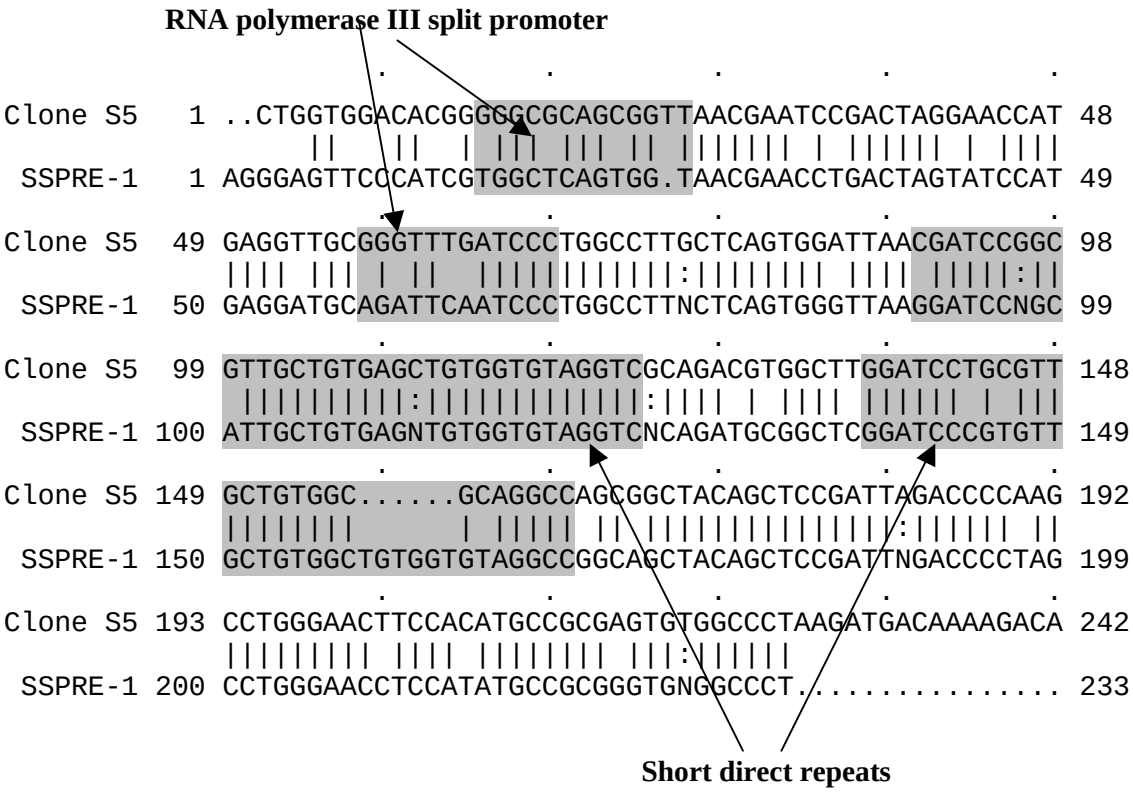


Fig 6.9 Comparison of novel peccary SINE sequence from Clone S5 with porcine PRE-1 sequence (GenBank accession number Y00104) showing the structural features based on Singer *et al*. (1987) (indicated by shading (…)).

Table 6.4 PRE-1 sequence homology of pig and peccary.

	Clone S5	[1]SSPRE1	[2]Peccary SINE1	[3]Peccary SINE2
[1]SSPRE1	85.2 %	-		
[2]Peccary SINE1	70.1 %	81.5 %	-	
[3]Peccary SINE2	77.2 %	84.1 %	70.4 %	-
[4]Peccary SINE3	76.0 %	83.5 %	71.9 %	79.1 %

[1]GenBank accession number Y00104.

[2,3]DDBJ accession numbers AB000378 and AB000379.

[4]GenBank accession number AB003283 and located in the *trappin* gene.

6.5. Discussion

Degenerate primers based on conserved motifs derived from Murine leukemia viruses (Tristem *et al*., 1996b) have been used to amplify fragments from peccary genomic DNA. Two clones show close homology, albeit with a large deletion, to mouse and pig retroviral sequences. These are the first retroviral sequences identified in peccary. They clearly contradict the observations of Benveniste and Todaro (1975) based on Southern hybridisation that pigs only, and not peccaries, contain C-type retroviruses which they claimed were derived from an invasion of the pig genome by murine retroviruses after the split of the Suidae and Tayassuidae lineages. This chapter clearly shows that endogenous retroviral sequences are present in the peccary genome and that these sequences are very similar to mouse retroviral sequences.

Pigs and peccaries are inarguably more closely related to each other than to mouse (Douzery and Catzeflis, 1995). However the FASTA alignment of the peccary endogenous retroviral sequences generated in this study suggested they are more similar to mouse retrovirus than pig retrovirus. To test this contradictory result more carefully, pig (GenBank accession number AF038600) and mouse (GenBank accession number MULV13893) endogenous retroviral sequences were aligned with peccary sequences, after removing the region corresponding to sequences deleted in peccary. This alignment showed 62%, 65% and 65% sequence identity between pig and mouse, pig and peccary, and mouse and peccary, respectively. Thus the pig and peccary endogenous retroviral sequences have identical levels of sequence identity as mouse and peccary, but the pig and mouse sequences have slightly less identity. Thus the FASTA results are misleading since they are based only on the alignments either side of the deletion in the peccary. This misleading impression is reinforced by the fact that there are so many mouse retroviral sequences in GenBank that a large number of mouse “hits” are observed before the first pig “hit”.

Only two clones (S1, S2) of six containing the smaller (putative 434 bp) inserts (S1, S2, S5, S9, S10) were verified to contain target peccary endogenous retroviral sequences. This raises the possibility that if additional clones containing the larger insert (840 bp PCR products) were analysed, a complete, undeleted retroviral sequence may have been found. However, due to time constraints, only one larger product clone was analysed in this study.

By using internal primers which span the 534 bp deletion point, PCR amplification has confirmed that this deletion is common to all available Collared peccary samples from Colombia. Also it is clear that the endogenous retroviruses in which the deletion occurred has at least some sequence differences from full-length retroviruses in the peccary, otherwise the Tristem (1996b) degenerate primers would have amplified full-length product. The 534 bp deletion in the *pol* gene

coding sequences is incompatible with this deleted endogenous retrovirus completing the retroviral life cycle without assistance from complete competent retrovirus. The sequence comparison of two retroviral clones shows that these two clones are not identical, implying that these two retroviral sequences were amplified from different copies in the peccary genome. Therefore, it can be assumed that the deletion occurred before this retrovirus was inserted into multiple sites in the peccary host genome. The amplification, albeit more weakly, of a similar sized PCR product from a pig control DNA sample using these same internal primers, indicates that a similar deleted retrovirus also is present in pigs. The relationship between the pig and peccary deleted viruses is unclear. If they are both nonfunctional and were incorporated before the separation of the Suidae and Tayassuidae lineage, then mutational decay might be expected to have caused substantial divergence, so that amplification of pig product with peccary-specific primers would be unlikely. Sequence analysis of the pig PCR product will give some clue to the origin of the deleted viruses found in both pigs and peccaries.

Four non-target sequences were amplified from peccary with the degenerate retroviral primers. Clone S10 contains part of the peccary cofilin gene sequence. Cofilin is a wide-distributed, intracellular, actin binding protein which is involved in the translocation of actin-cofilin complex from cytoplasm to nucleus. Two cofilin isoforms have been identified in human, which are called non-muscle type cofilin (*CFL1*) and muscle-type cofilin (*CFL2*) (Gillett *et al*., 1996). The cofilin sequence in this study could contribute to understanding the evolution of cofilin in pigs and their relatives. The microsatellite sequence in clone S6 can be used as a marker for population, parentage and mapping studies in the peccary. It may also be useful in the pig. Góngora *et al*. (unpublished data) found that 87% of eighteen porcine microsatellite primers amplified microsatellite products in Collared peccaries, suggesting the potential for reciprocal cross species use of peccary microsatellite primers in pigs.

The novel peccary SINE sequence found in this study could be a useful indicator of evolutionary history (Sulandari *et al*., 1997; Yasue and Wada, 1996). PRE-1 sequences are present in the genomes of Collared peccary (*Tayassu tajacu*) at almost the same frequency as in pig, at least as determined by dot blot hybridisation (Yasue and Wada, 1996). Sulandari *et al*. (1997) successfully amplified peccary SINE sequences using individual locus-specific porcine PRE-1 primers and used their sequences to calculate the divergence time for pigs and peccaries. In pig, PRE-1 loci have been tested as candidates for SSCP markers in genetic linkage analysis (Harumi *et al*., 1995). Also, length polymorphisms of the PRE-1 poly (A) tail have been evaluated as genetic markers (Ellegren, 1993) because of their abundance in the porcine genome.

Therefore this newly discovered peccary SINE sequence could also be used as a marker and for population studies.

Although the only peccary endogenous retroviruses analysed in this study have a large deletion, it is highly probable that complete functional retroviruses are present in the peccary genome, as they have been found in all other species extensively investigated. The degenerate primers of Tristem *et al*. (1996b) failed to amplify full-length retroviral sequences in peccaries although they have worked well in other mammalian species such as dog, red fox, American mink, Eurasian badger, grey seal, cow, sheep, horse, rabbit, free-tailed bat and pig. An alternative approach, other than the use of these degenerate primers, will be required to extract the full-length peccary endogenous retroviral sequences for the reverse transcriptase gene, in order to further advance the comparative evolutionary studies of peccary and porcine endogenous viruses. Do peccaries have a similar number of endogenous retroviruses as the pig? What proportion are functional? Are different classes of endogenous retroviruses recognisable by envelope sequences and analogous to PERV-A, -B and -C of pigs, found in peccaries? Such questions remain for future studies.

Chapter Seven

Final Discussion and Conclusion

7.1. Development of pig - human comparative map

Comparative mapping enables recognition of the conservation of genomic segments across species and will be a valuable tool for data-mining of the genomic and QTL mapping efforts in many species. Some time ago, Zoo-FISH established the broad chromosomal relationships between human and pig chromosomes (Fröncke *et al.*, 1996; Goureau *et al.*, 1996; Johansson *et al.*, 1995; Rettenberger *et al*., 1995b). As a supplement to this, an international effort has been made to map Comparative Anchor Tagged Sequences (CATS) markers in the pig. As part of this collaboration, consensus primers for 53 coding sequences (Type I markers), which map to human chromosome 9, 10, 20, and 22 were allotted to our Sydney group and used to generate PCR products from pig genomic DNA. Six of these CATS loci were previously mapped by Zhang (1997). In this thesis, 47 CATS primers comprising new primers for chromosomes 20 and 22 or primers for products not mapped by Zhang (1997) have been used to amplify porcine PCR products (Chapter 3). Sequencing of the porcine PCR products confirmed that 43% (23/53) of the products are from expected porcine homologues, including Zhang's (1997) sequencing results for unmapped products. Subsequently, 6 newly identified loci were physically assigned to pig chromosomes with a French somatic cell hybrid panel. The porcine locations of *OXT* and *ADRA1A* show inconsistency with the known syntenic relationships between pig and human chromosomes. In the former case, this reflects imperfect knowledge of pig/human synteny for pig chromosomes, but in the latter case, the result reflects the complications inherent in mapping members of a multi-gene family.

The major identified problem of the CATS primers is that they are designed from consensus nucleotide sequences including rodent species. However, rodents have been used as the background genome for the somatic cell hybrid panel. Therefore, rodent PCR products are frequently coamplified and often could not be distinguished from the porcine products. In such cases, it is impossible to map the gene.

Despite the problems, the comparative mapping results have provided a useful, albeit limited, contribution for exchanging information between species.

7.2. Characterising and mapping PERVs

PERVs have potential infectious risk in relation to xenotransplantation. As the inbred Westran line of pigs may be used as a xenotransplantation donor in future, it is very important that any PERVs in it be characterised and mapped (Chapter 4 and Chapter 5).

7.2.1. Characterising PERVs

The conserved region of the PERV-A and PERV-B *env* gene was used to make PCR primers to amplify all possible variants of PERVs. 410 PERV clones, amplified by *Taq* and *Pfu* polymerases, were initially analysed by restriction enzyme digestion pattern to provide a basis for the selection of clones for further analysis. 31 PERV clones were fully sequenced by single-pass sequencing. The sequence results show that 55% of the clones (17/31) have premature stop codons within the protein coding sequences, suggesting that they could not make infectious virus particles.

Five clone sequences were classified as recombinants between PERV-A and PERV-B. The implication of these recombinant clones is that defective PERVs could regain infectivity through recombination and potentially more dangerously PERVs could even recombine with HERVs. However, it has still not been unequivocally established that these recombinant clones are not artefacts of PCR. A very important job for the near future will be to attempt to amplify recombinant product directly from Westran genomic template using primers specific for these recombinant viruses.

7.2.2. Physically mapping of PERVs

FISH has established that there are at least 13 PERV-A and 9 PERV-B sites in Westran pigs. The comparison of these PERV locations between Westran and a Large White pig (Rogel-Gaillard *et al*., 1999) indicates that the PERV-A locations are quite different in the two breeds, but about half of the PERV-B locations are common. These mapping results corroborate the expectation that different breeds will have different PERV locations but may share some common PERV sites.

A method is evaluated in this thesis for cloning PERV insertion junctions using a strategy employing PERV and SINE primers. So far, only a single junction has been mapped. Somatic cell hybrid data show that this particular PERV junction fragment maps to a known PERV-B site on pig chromosome 17 (Chapter 5). The result of this method indicates that PERV inserts in Westran pigs can be mapped precisely, especially if the technique is applied using a radiation hybrid panel. However, it still remains to correlate each of the PERV sequences generated in

Chapter 4 with the chromosomal locations described in Chapter 5. This is an important task for the future.

7.3. Endogenous retroviruses in Collared peccaries

Peccaries are classified in a different mammalian family to pigs, although the Tayassuidae are the closest relatives of the Suidae. Benveniste and Todaro (1975) claimed that peccaries do not contain endogenous retroviruses and produced an elaborate scenario for the recent invasion of the pig genome by C-type retroviruses from rodents, after the divergence of Tayassuidae and Suidae. A study (Chapter 6) was carried out to determine the existence and nature of endogenous retroviral sequences in the Collared peccary (*Tayassu tajacu*) using the degenerate retroviral primers designed by Tristem *et al*. (1996b). The sequence of putative retroviral clones showed that some matched well with known retroviral sequences. Ironically, given the claims of Benveniste and Todaro (1975), the peccary retroviral sequence matched mouse sequence just as well as that of pig. In any case, the results established for the first time that peccaries have endogenous retroviruses in their genome. Strangely the peccary retroviral sequences have a 534 bp deletion, also confirmed by new set of internal primers. Pigs also appear to have the same deleted sequences in some of their endogenous viruses. The comparison of deleted retroviral sequences between pigs and peccaries will give some clues to the origin and evolution of the deletion. It is virtually certain that peccaries also have full-length endogenous retroviruses in their genome as in other species, but this remains to be established in future studies.

7.4. Future research

At this stage, there is no correlation of the PERV sequences and functional characterisation with the mapping results in Westran pigs. A priority for the future is to assign specific PERV sequences to their physical map locations. Ultimately it would be very useful to have a set of genotyping primers capable of recognising presence or absence of all known retroviral inserts in the pig.

Recombinant PERVs are a particular concern in relation to xenotransplantation. For this reason, it is vitally important to determine whether the recombinant PERV sequences are genuine or artefacts of PCR. Specific primers will be designed for use on Westran genomic DNA to test whether they are genuine recombinants or not.

Preliminary results indicate that Westran pigs are positive for PERV-C endogenous retroviruses. Further characterisation of the PERV-C insert(s) in Westran pigs is also an important task for the future work.

Finally, only endogenous retroviruses with 534 bp deleted from the protease gene have so far been amplified in Collared peccaries. It is very unlikely that there are no full-length retroviruses in peccaries. It will be highly desirable in future to extract full-length peccary endogenous retroviral sequences in order to better investigate the relationship between and evolution of peccary and porcine endogenous retroviruses.

Bibliography

Akiyoshi, D. E., Denaro, M., Zhu, H., Greenstein, J. L., Banerjee, P. and Fishman, J. A., 1998, Identification of a full-length cDNA for an endogenous retrovirus of miniature swine, *Journal of Virology* **72**(5)**:**4503-4507.

Aleyasin, A. and Barendse, W., 1999, Comparative mapping of genes from human chromosome 12 by genetic linkage mapping in cattle, *Journal of Heredity* **90**(5)**:**537-542.

Andersson, L., Archibald, A., Ashburner, M., Audun, S., Barendse, W., Bitgood, J., Bottema, C., Broad, T., Brown, S., Burt, D., Charlier, C., Copeland, N., Davis, S., Davisson, M., Edwards, J., Eggen, A., Elgar, G., Eppig, J. T., Franklin, I., Grewe, P., Gill, T., Graves, J. A. M., Hawken, R., Hetzel, J., Hillyard, A., Jacob, H., Jaswinska, L., Jenkins, N., Kunz, H., Levan, G., Lie, O., Lyons, L., Maccarone, P., Mellersh, C., Montgomery, G., Moore, S., Moran, C., Morizot, D., Neff, M., Nicholas, F. W., O'Brien, S. J., Parsons, Y., Peters, J., Postlethwait, J., Raymond, M., Rothschild, M., Schook, L., Sugimoto, Y., Szpirer, C., Tate, M., Taylor, J., Vandeberg, J., Wakefield, M., Wienberg, J. and Womack, J., 1996, Comparative genome organization of vertebrates, *Mammalian genome* **7**:717-734.

Andersson, L., Edfors-Lilja, I., Ellegren, H., Johansson, M. and Marklund, L., 1994a, Mapping trait loci by crossbreeding genetically divergent populations of domestic animals, *Animal Biotechnology* **5**(2)**:**225-231.

Andersson, L., Haley, C. S., Ellegren, H., Knott, S. A., Johansson, M. andersson, K. andersson-Eklund, L., Edfors-Lilja, I., Fredholm, M. and Hansson, I., 1994b, Genetic mapping of quantitative trait loci for growth and fatness in pigs, *Science* **263**(5154)**:**1771-4.

Archibald, A. L., 1994, From mapping to manipulating the vertebrate genome, *Animal Biotechnology* **5**(2)**:**233-242.

Archibald, A. L., Couperwhite, S., Haley, C. S., Beattie, C. W. and Alexander, L. J., 1994, RFLP and linkage analysis of the porcine casein loci--CASAS1, CASAS2, CASB and CASK, *Animal Genetics* **25**(5)**:**349-51.

Archibald, A. L. and Haley, C. S., 1998, Genetic linkage maps, in: *The genetics of the pig* (M. F. Rothschild and A. Ruvinsky, eds.), CAB International, New York, pp. 265-294.

Archibald, A. L., Haley, C. S., Brown, J. F., Couperwhite, S., McQueen, H. A., Nicholson, D., Coppieters, W., Van de Weghe, A., Stratil, A., Winterø, A. K., Fredholm, M., Larsen, N. J., Nielsen, V. H., Milan, D., Woloszyn, N., Robic, A., Dalens, M., Riquet, J., Gellin, J., Caritez, J.-C., Burgaud, G., Ollivier, L., Bidanel, J.-P., Vaiman, M., Renard, C., Geldermann, H., Davoli, R., Ruyter, D., Verstege, E. J. M., Groenen, M. A. M., Davies, W., Høyheim, B., Keiserud, A. andersson, L., Ellegren, H., Johansson, M., Marklund, L., Miller, J. R. anderson Dear, D. V., Signer, E., Jeffreys, A. J., Moran, C., Le Tissier, P., Muladno, Rothschild, M. F., Tuggle, C. K., Vaske, D., Helm, J., Liu, H.-C., Rahman, A., Yu, T.-P., Larson, R. G. and Schmitz, C.

B., 1995, The PiGMaP consortium linkage map of the pig (Sus scrofa), *Mammalian Genome* **6:**157-175.

Armstrong, J. A., Porterfield, J. S. and deMadrid, A. T., 1971, C-type virus particles in pig kidney cell lines, *Journal of General Virology* **10:**195-198.

Bailey, L. L., Nehlsen-Cannarella, S. L., Concepcion, W. and Jolley, W. B., 1985, Baboon-to-human cardiac xenotransplantation in a neonate, *JAMA* **254**(23)**:**3321-9.

Bebenek, K., Abbotts, J., Roberts, J. D., Wilson, S. H. and Kunkel, T. A., 1989, Specificity and machanism of error-prone replication by Human Immunodeficiency Virus-1 Reverse Transcriptase, *The Journal of Biological Chemistry* **264**(28)**:**16948-16956.

Beier, D. R., 1993, Single-strand conformation polymorphism (SSCP) analysis as a tool for genetic mapping, *Mammalian Genome* **4**(11)**:**627-31.

Benirschke, K. and Kumamoto, A. T., 1989, Further studies on the chromosomes of three species of peccary, in: *Advances in Neotropical Mammalogy* (K. H. Redford and J. F. Esenberg, eds.), Sandhill Crane Press, Gainesville, FL, pp. 309-316.

Benirschke, K., Kumamoto, A. T. and Meritt, D. A., 1985, Chromosomes of the Chacoan peccary, *Catagonus wagneri* (Rusconi), *The Journal of Heridity* **76:**95-98.

Benveniste, R. E. and Todaro, G. J., 1975, Evolution of type C viral genes: Preservation of ancestral murine type C viral sequences in pig cellular DNA, *Proceedings of the National Academy of Sciences of the United States of America* **72**(10)**:**4090-4094.

Bhatti, F. N., Zaidi, A., Schmoeckel, M., Cozzi, E., Chavez, G., Wallwork, J., White, D. J. and Friend, P. J., 1998, Survival of life-supporting HDAF transgenic kidneys in primates is enhanced by splenectomy, *Transplantation Proceedings* **30**(5)**:**2467.

Bird, A. P., 1995, Gene number, noise reduction and biological complexity, *Trends in Genetics* **11**(3)**:**94-100.

Birnboim, H. C. and Doly, J., 1979, A rapid alkaline extraction procedure for screening recombinant plasmid DNA, *Nucleic Acids Research* **7:**1513.

Blair, D. G., Mason, W. S., Hunter, E. and Vogt, P. K., 1976, Temperature-sensitive mutants of avian sarcoma viruses: genetic recombination between multiple or coordinate mutants and avian leukosis viruses, *Virology* **75**(1)**:**48-59.

Boeke, J. D. and Stoye, J. P., 1997, Retrotransposons, endogenous retroviruses and the evolution of retroelements, in: *Retroviruses* (J. M. Coffin, S. H. Hughes and H. E. Varmus, eds.), Cold Sprong Harbor Laboratory Press, New York, pp. 343-435.

Bosch, A., Nunes, V., Patterson, D. and Estivill, X., 1993, Isolation and characterization of 14 CA-repeat microsatellites from human chromosome 21, *Genomics* **18**(1)**:**151-5.

Bösch, S., Arnauld, C., Cariolet, R. and Jestin, A., 1999, Characterisation of endogenous retroviral sequences in specific-pathogen-free swine, in: *The 5th Congress of the International Xenotransplantation Association*, Nagoya, Japan, pp. 117, Abstract number 1156.

Bostock, D. E. and Owen, L. N., 1973, Porcine and ovine lymphosarcoma: a review, *Journal of the National Cancer Institute* **50**(4)**:**933-9.

Botstein, D., White, R. L., Skolnick, M. and Davis, R. W., 1980, Construction of a genetic linkage map in man using restriction fragment length polymorphisms, *American Journal of Human Genetics* **32**(3)**:**314-31.

Bradbury, J., 2000, Cloned pigs bring xenotransplatation nearer, *The Lancet* **355:**991.

Breese, S. S., 1970, Virus-like particles occurring in culture of stable pig kidney cell line, *Archiv für die Gesamte Virusforschung* **30:**401-404.

Broad, T. E., Lambeth, M., Burkin, D. J., Jones, C., Pearce, P. D., Maher, D. W. and Ansari, H. A., 1996, Physical mapping confirms that sheep chromosome 10 has extensive conserved synteny with cattle chromosome 12 and human chromosome 13, *Animal Genetics* **27**(4)**:**249-53.

Brookes, A. J., 1999, The essence of SNPs, *Gene* **234**(2)**:**177-86.

Bruch, J., Rettenberger, G., Leeb, T., Meier-Ewert, S., Klett, C., Brenig, B. and Hameister, H., 1996, Mapping of type I loci from human chromosome 7 reveals segments of conserved synteny on pig chromosomes 3, 9 and 18, *Cytogenetics & Cell Genetics* **73**(3)**:**164-7.

Bruford, M. W. and Wayne, R. K., 1993, Microsatellites and their application to population genetic studies, *Current Opinion in Genetics & Development* **3**(6)**:**939-43.

Byrne, G. W., McCurry, K. R., Martin, M. J., McClellan, S. M., Platt, J. L. and Logan, J. S., 1997, Transgenic pigs expressing human CD59 and decay-accelerating factor produce an intrinsic barrier to complement-mediated damage, *Transplantation* **63**(1)**:**149-55.

Caetano, A. R., Pomp, D., Murray, J. D. and Bowling, A. T., 1999, Comparative mapping of 18 equine type I genes assigned by somatic cell hybrid analysis, *Mammalian Genome* **10**(3)**:**271-6.

Cantor, C. R. and Smith, C. L., 1999, genomics: The science and technology behind the human genome project, John Wiley and Sons, Inc.

Chari, R. S., Collins, B. H., Magee, J. C., DiMaio, J. M., Kirk, A. D., Harland, R. C., McCann, R. L., Platt, J. L. and Meyers, W. C., 1994, Brief report: treatment of hepatic failure with ex vivo pig-liver perfusion followed by liver transplantation, *New England Journal of Medicine* **331**(4)**:**234-7.

Chee, M., Yang, R., Hubbell, E., Berno, A., Huang, X. C., Stern, D., Winkler, J., Lockhart, D. J., Morris, M. S. and Fodor, S. P., 1996, Accessing genetic information with high-density DNA arrays, *Science* **274**(5287)**:**610-4.

Chen, C. G., Salvaris, E. J., Romanella, M., Aminian, A., Katerelos, M., Fisicaro, N., d'Apice, A. J. and Pearse, M. J., 1998, Transgenic expression of human alpha1,2-fucosyltransferase (H-transferase) prolongs mouse heart survival in an ex vivo model of xenograft rejection, *Transplantation* **65**(6)**:**832-7.

Chen, Z. Q., Lautenberger, J. A., Lyons, L. A., McKenzie, L. and O'Brien, S. J., 1999, A human genome map of comparative anchor tagged sequences, *Journal of Heredity* **90**(4)**:**477-84.

Chevalet, C., Gouzy, J. and SanChristobal-Gaudy, M., 1997, Regional assignment of genetic markers using a somatic cell hybrid panel: a WWW interactive program available for the pig genome, *Computer Appincations in the Biosciences* **13:**69-73.

Chowdhary, B. P., Frönicke, L., Gustavsson, I. and Scherthan, H., 1996, Comparative analysis of the cattle and human genomes: detection of ZOO-FISH and gene mapping-based chromosomal homologies, *Mammalian Genome* **7**(4)**:**297-302.

Clavel, F., Hoggan, M. D., Willey, R. L., Strebel, K., Martin, M. A. and Repaske, R., 1989, Genetic recombination of human immunodeficiency virus, *Journal of Virology* **63**(3)**:**1455-9.

Coffin, J. M., 1992, Structure and classification of retroviruses, in: *The retroviridae* (J. A. levy, ed.), Plenum press, New York and London, pp. 19-49.

Cole, S. E., Wiltshire, T. and Reeves, R. H., 1998, Physical mapping of the evolutionary boundary between human chromosomes 21 and 22 on mouse chromosome 10, *Genomics* **50**(1)**:**109-11.

Collins, F. S., 1995, Positional cloning moves from perditional to traditional [published erratum appears in Nat Genet 1995 Sep;11(1):104], *Nature Genetics* **9**(4)**:**347-50.

Cooper, D. N., Smith, B. A., Cooke, H. J., Niemann, S. and Schmidtke, J., 1985, An estimate of unique DNA sequence heterozygosity in the human genome, *Human Genetics* **69**(3)**:**201-5.

Cooper, H. M., 1954, Kangaroo Island's wild pigs, *The South Australian Naturalist* **28**(5)**:**57-61.

Cozzi, E. and White, D. J., 1995, The generation of transgenic pigs as potential organ donors for humans, *Nature Medicine* **1**(9)**:**964-6.

Crawford, A. M., Dodds, K. G. and McEwan, J. C., 2000, DNA markers, genetic maps and the identification of QTL: general principles, in: *Breeding for disease resistance in farm animals* (R. F. E. Axfrod, S. C. Bishop, F. W. Nicholas and J. B. Owen, eds.), CABI Publishing, New York, pp. 3-26.

Dabkowski, P. L., Vaughan, H. A., McKenzie, I. F. C. and Sandrin, M. S., 1994, Isolation of a cDNA clone encoding the pig alpha1,3 Galactosyltransferase, *Transplantation Proceedings* **26**(3)**:**1335.

Daggett, C. W., Yeatman, M., Lodge, A. J., Chen, E. P., Van Trigt, P., Byrne, G. W., Logan, J. S., Lawson, J. H., Platt, J. L. and Davis, R. D., 1997, Swine lungs expressing human complement-regulatory proteins are protected against acute pulmonary dysfunction in a human plasma perfusion model, *Journal of Thoracic & Cardiovascular Surgery* **113**(2)**:**390-8.

Deacon, T., Schumacher, J., Dinsmore, J., Thomas, C., Palmer, P., Kott, S., Edge, A., Penney, D., Kassissieh, S., Dempsey, P. and Isacson, O., 1997, Histological evidence of fetal pig neural cell survival after transplantation into a patient with Parkinson's disease, *Nature Medicine* **3**(3)**:**350-3.

DeBry, R. W. and Seldin, M. F., 1996, Human/mouse homology relationships, *Genomics* **33:**337-351.

DeMartini, J. C., de la Concha-Bermejillo, A., Carlson, J. O. and Bowen, R. A., 2000, Diseases caused by Maedi-visna and other ovine lentiviruses, in: *Breeding for disease registance in farm animals* (R. F. E. Axford, S. C. Bishop, F. W. Nicholas and J. B. Owen, eds.), CABI Publishing, New York, pp. 301-324.

Deng, Y.-M., Lee, J.-H., Moran, C., Tuch, B. and Rawlinson, W. D., 2000, Mapping dispersed repetitive loci using semi-specific PCR cloning and somatic cell hybrid mapping, *submitted* .

Dodgson, J. B., Cheng, H. H. and Okimoto, R., 1997, DNA marker technology: a revolution in animal genetics, *Poultry Science* **76**(8)**:**1108-14.

Don, R. H., Cox, P. T., Wainwright, B. J., Baker, K. and Mattick, J. S., 1991, 'Touchdown' PCR to circumvent spurious priming during gene amplification, *Nucleic Acids Research* **19**(14)**:**4008.

Douzery, E. and Catzeflis, F. M., 1995, Molecular evolution of the mitochondrial 12S rRNA in Ungulata (mammalia), *Journal of Molecular Evolution* **41**(5)**:**622-36.

Drake, J. W., 1969, Comparative rates of spontaneous mutation, *Nature* **221**(5186)**:**1132.

Ducrocq, S., 1994, An Eocene peccary from Thailand and the biogeographical origins of the artiodactyl family Tayassuidae, *Palaeontology* **37**(Part 4)**:**765-779.

Dunham, I., Shimizu, N., Roe, B. A., Chissoe, S., Hunt, A. R., Collins, J. E., Bruskiewich, R., Beare, D. M., Clamp, M., Smink, L. J., Ainscough, R., Almeida, J. P., Babbage, A., Bagguley, C., Bailey, J., Barlow, K., Bates, K. N., Beasley, O., Bird, C. P., Blakey, S., Bridgeman, A. M., Buck, D., Burgess, J., Burrill, W. D. and O'Brien, K. P., 1999, The DNA sequence of human chromosome 22 [published erratum appears in Nature 2000 Apr 20;404(6780):904], *Nature* **402**(6761)**:**489-95.

Echard, G., 1990, The gene map of the pig (Sus Scrofa domestica L.), in: *Genetic maps - locus maps of complex traits* (S. J. O'Brien, ed.), Cold Spring Harbour Laboratory Press, pp. 4.110.

Echard, G., Milan, D., Yerle, M., Lahbib-Mansais, Y. and Gellin, J., 1992, The gene map of the pig (Sus scrofa domestica L.): a review, *Cytogenetics & Cell Genetics* **61**(2)**:**146-51.

Elder, J. H., Gautsch, J. W., Jensen, F. C., Lerner, R. A., Hartley, J. W. and Rowe, W. P., 1977, Biochemical evidence that MCF murine leukemia viruses are envelope (*env*) gene recombinants, *Proceedings of the National Academy of Sciences of the United States of America* **74**(10)**:**4676-80.

Ellegren, H., 1993, Variable SINE 3' poly (A) sequences:an abundant class of genetic markers in the pig genome, *Mammalian Genome* **4:**429-434.

Ellegren, H., Chowdhary, B. P., Johansson, M., Marklund, L., Fredholm, M., Gustavsson, I. and Andersson, L., 1994, A primary linkage map of the porcine genome reveals a low rate of genetic recombination, *Genetics* **137**(4)**:**1089-100.

Ellegren, H., Fredholm, M., Edfors-Lilja, I., Winterø, A. K. and Andersson, L., 1993, Conserved syntany between pig chromosome 8 and human chromisome 4 but rearranged and distorted linkage maps, *Genomics* **17:**599-603.

Ellermann, V. and Bang, O., 1908, Experimentelle Leukämie bei Hühnern, *Zentralblatt Fur Bakteriologie Parasitenkunde Infektionskrankheiten Und Hyg* **46:**595-609.

Eppig, J. T., 1996, Comparative maps:adding pieces to the mammalian jigsaw puzzle, *Current Opinion in Genetics and Development* **6:**723-730.

Eppig, J. T. and Nadeau, J. H., 1995, Comparative maps: the mammalian jigsaw puzzle, *Current Opinion in Genetics and Development* **5:**709-716.

Ewens, W. J., Griffiths, R. C., Ethier, S. N., Wilcox, S. A. and Graves, J. A., 1992, Statistical analysis of in situ hybridization data: derivation and use of the Z_{max} test, *Genomics* **12**(4)**:**675-82.

Faller, D. V. and Hopkins, N., 1978, T1 oligonucleotides that segregate with tropism and with properties of gp70 in recombinants between N- and B-tropic murine leukemia viruses, *Journal of Virology* **26**(1)**:**153-8.

Fodor, W. L., Williams, B. L., Matis, L. A., Madri, J. A., Rollins, S. A., Knight, J. W., Velander, W. and Squinto, S. P., 1994, Expression of a functional human complement inhibitor in a transgenic pig as a model for the prevention of xenogeneic hyperacute organ rejection, *Proceedings of the National Academy of Sciences of the United States of America* **91**(23)**:**11153-7.

Frankel, W. N., Stoye, J. P., Taylor, B. A. and Coffin, J. M., 1990, A linkage map of endogenous murine leukemia proviruses [published erratum appears in Genetics 1990 Jun;125(2):455], *Genetics* **124**(2)**:**221-36.

Frönicke, L., Chowdhary, B. P., Scherthan, H. and Gustavsson, I., 1996, A comparative map of the porcine and human genomes demonstrates ZOO-FISH and gene mapping-based chromosomal homologies, *Mammalian Genome* **7:**285-290.

Fujii, J., Otsu, K., Zorzato, F., de Leon, S., Khanna, V. K., Weiler, J. E., O'Brien, P. J. and MacLennan, D. H., 1991, Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia, *Science* **253**(5018)**:**448-51.

Furutani, Y., Kato, A., Yasue, H., Alexander, L. J., Beattie, C. W. and Hirose, S., 1998, Evolution of the trappin multigene family in the Suidae, *Journal of Biochemistry* **124**(3)**:**491-502.

Gao, Q. and Womack, J. E., 1997, Comparative mapping of anchor loci from HSA19 to cattle chromosomes 7 and 18, *Journal of Heredity* **88**(6)**:**524-7.

Gibbs, M. J., Armstrong, J. S. and Gibbs, A. J., 2000, Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences, *Bioinformatics* in press.

Gillett, G. T., Fox, M. F., Rowe, P. S., Casimir, C. M. and Povey, S., 1996, Mapping of human non-muscle type cofilin (*CFL1*) to chromosome 11q13 and muscle-type cofilin (*CFL2*) to chromosome 14, *Annals of Human Genetics* **60:**201-11.

Gillois, M., 1991, Gene mapping today: applications to farm animals, *Genetics, Selection, Evolution* **1**(suppl1)**:**19s-48s.

Gojobori, T. and Yokoyama, S., 1985, Rates of evolution of the retroviral oncogene of Moloney murine sarcoma virus and of its cellular homologues, *Proceedings of the National Academy of Sciences of the United States of America* **82**(12)**:**4198-201.

Gonda, M. A., 1992, The lentiviruses of cattle, in: *The retroviridae* (J. A. levy, ed.), Plenum press, New York and London, pp. 83-109.

Góngora, J., Bernal, J., Chen, Y., Moran, C. and Nicholas, F., Cytogenetic and molecular genetic studies of Colombian Collared peccaries, *Unpublished* .

Goureau, A., Yerle, M., Schmitz, A., Riquet, J., Milan, D., Pinton, P., Frelat, G. and Gellin, J., 1996, Human and porcine correspondence of chromosome segments using bidirectional chromosome painting, *Genomics* **36:**252-262.

Groth, C. G., Korsgren, O., Tibell, A., Tollemar, J., Moller, E., Bolinder, J., Ostman, J., Reinholt, F. P., Hellerstrom, C. and Andersson, A., 1994, Transplantation of porcine fetal pancreas to diabetic patients, *Lancet* **344**(8934)**:**1402-4.

Gustavsson, I., 1988, Standard karyotype of the domestic pig, *Heriditas* **109:**151-157.

Gyapay, G., Schmitt, K., Fizames, C., Jones, H., Vega-Czarny, N., Spillett, D., Muselet, D., Prud'Homme, J. F., Dib, C., Auffray, C., Morissette, J., Weissenbach, J. and Goodfellow, P. N., 1996, A radiation hybrid map of the human genome, *Human Molecular Genetics* **5**(3)**:**339-46.

Haldane, J. B. S., 1927, The comparative genetics of colour in rodents and carnivora, *Biological Reviews and Biological Proceedings of the Cambridge Philosophical Society* **2**(3)**:**199-212.

Haley, C. S., Archibald, A. andersson, L., Bosma, A. A., Davies, W., Fredholm, M., Geldermann, H., Groenen, M., Gustavsson, I., Ollivier, L., Tucker, E. M. and Van de Weghe, A., 1990, The pig gene mapping project - PiGMaP, in: *Proceeding of the 4th World Congress on Genetics Applied to Livestock Production*, Edinburgh, pp. 67-70.

Hallouin, F., Goupille, C., Bureau, V., Meflah, K. and Le Pendu, J., 1999, Increased tumorigenicity of rat colon carcinoma cells after alpha1,2-fucosyltransferase FTA anti-sense cDNA transfection, *International Journal of Cancer* **80**(4)**:**606-11.

Hamann, J., Vogel, B., van Schijndel, G. M. and van Lier, R. A., 1996, The seven-span transmembrane receptor CD97 has a cellular ligand (CD55, DAF), *Journal of Experimental Medicine* **184**(3)**:**1185-9.

Hameister, H., Klett, C., Bruch, J., Dixkens, C., Vogel, W. and Christensen, K., 1997, Zoo-FISH analysis: The American mink (Mustela vison) closely resembles the cat karyotype, *Chromosome Research* **5:**5-11.

Hardy, W. D., 1992, Feline oncoretroviruses, in: *The retroviridae* (J. A. levy, ed.), Plenum press, New York and London, pp. 109-180.

Harumi, T., Kimura, M. and Yasue, H., 1995, Survey on swine SINEs (PRE-1) as candidates for SSCP markers in genetic linkage analysis, *Animal Genetics* **26:**403-406.

Hawthorne, W. J., Burgess, J. S., Chen, Y., Walters, S., Patel, T., Clarke, J., Weston, L., O'Connell, P. J., Moran, C., Chapman, J. R. and Allen, R. D. M., 1999, A pig colony suitable for xenotransplantation, in: *7th World Congress of the International Pancreas and Islet Transplantation Association*, Sydney, Australia, pp. 143.

Hattori, M., Fujiyama, A., Taylor, T. D., Watanabe, H., Yada, T., Park, H. S., Toyoda, A., Ishii, K., Totoki, Y., Choi, D. K., Soeda, E., Ohki, M., Takagi, T., Sakaki, Y., Taudien, S., Blechschmidt, K., Polley, A., Menzel, U., Delabar, J., Kumpf, K., Lehmann, R., Patterson, D., Reichwald, K., Rump, A., Schillhabel, M. and Schudy, A., 2000, The DNA sequence of human chromosome 21. The chromosome 21 mapping and sequencing consortium, *Nature* **405**(6784)**:**311-9.

Hayes, H., 1995, Chromosome painting with human ahromosome specific DNA libraries reveals the extent and the distribution of conserved segments in the bovine chromosomes, *Cytogenetics and Cell Genetics* **71:**168-174.

Hecht, S. J., Stedman, K. E., Carlson, J. O. and DeMartini, J. C., 1996, Distribution of endogenous type B and type D sheep retrovirus sequences in ungulates and other mammals, *Proceedings of the National Academy of Sciences of the United States of America* **93**(8)**:**3297-302.

Hendey, Q. B., 1976, Fossil Peccary from the Pliocene of South Africa, *Science* **192:**787-789.

Heneine, W., Tibell, A., Switzer, W. M., Sandstrom, P., Rosales, G. V., Mathews, A., Korsgren, O., Chapman, L. E., Folks, T. M. and Groth, C. G., 1998, No evidence of infection with porcine endogenous retrovirus in recipients of porcine islet-cell xenografts [published erratum appears in Lancet 1998 Oct 31;352(9138):1478], *Lancet* **352**(9129)**:**695-9.

Herniou, E., Martin, J., Miller, K., Cook, J., Wilkinson, M. and Tristem, M., 1998, Retroviral diversity and distribution in vertebrates, *Journal of Virology* **72**(7)**:**5955-5966.

Herring, C., Whittam, A., Cunningham, D., Fernandez-Suarez, X., Gu, M.-L., Long, Z. and Langford, G., 1999, Analysis of porcine endogenous retrovirus variants in hDAF transgenic pigs, in: *The 5th Congress of the International Xenotransplantation Association*, Nagoya, Japan, pp. 116, Abstract number 1153.

Hoebee, B., Stoppelaar, J. M., Suijkerbuijk, R. F. and S. Monard, S., 1994, Isolation of rat chromosome-specific paint probes by bivariate flow sorting followed by degenerate oligonucleotide primed-PCR, *Cytogenetics and Cell Genetics* **66:**277-282.

Hourcade, D., Holers, V. M. and Atkinson, J. P., 1989, The regulators of complement activation (RCA) gene cluster, *Advances in Immunology* **45:**381-416.

Hu, W. S. and Temin, H. M., 1990, Genetic consequences of packaging two RNA genomes in one retroviral particle: pseudodiploidy and high rate of genetic recombination, *Proceedings of the National Academy of Sciences of the United States of America* **87**(4)**:**1556-60.

Hu, Z., Rohrer, G. A., Murtaugh, M. P., Stone, R. T. and Beattie, C. W., 1997, Mapping genes to swine X chromosome provides reference loci for comparative mapping, *Mammalian Genome* **8**(8)**:**608-10.

Hufty, M. P., Sedgwick, C. J. and Benirschke, K., 1973, The karyotypes of the White-lipped and Collared peccaries, aspects of their chromosomal evolution, *Genen Phaenen* **16**(3)**:**81-86.

Hutchison, C. A., Hardies, S. C., Loeb, D. D., Shehee, W. R. and Edgell, M. H., 1989, LINEs and related retroposons: long interspersed repeated sequences in the eukaryotic genome, in: *Mobile DNA* (D. E. Berg and M. M. Howe, eds.), Americal Society for Microbiology, Washington DC, pp. 593-617.

Iles, D. E., Segers, B., Heytens, L., Sengers, R. C. and Wieringa, B., 1992, High-resolution physical mapping of four microsatellite repeat markers near the RYR1 locus on chromosome 19q13.1 and apparent exclusion of the MHS locus from this region in two malignant hyperthermia susceptible families, *Genomics* **14**(3)**:**749-54.

Jeffreys, A. J., Wilson, V. and Thein, S. L., 1985, Hypervariable 'minisatellite' regions in human DNA, *Nature* **314**(6006)**:**67-73.

Jiang, Z., Priat, C., Galibert, F., 1998, Traced othologous amplified sequence tags (TOASTs) and mammalian comparative maps, *Mammalian Genome* **9:**577-587.

Johansson, M., Ellegren, H. and Andersson, L., 1995, Comparative mapping reveals extensive linkage conservation-but with gene order rearrangements-between the pig and the human genomes, *Genomics* **25:**682-690.

Juneja, R. K. and Vögeli, P., 1998, Biochemical genetics, in: *The genetics of the pig* (M. F. Rothschild and A. Ruvinsky, eds.), CAB International, New York, pp. 105-134.

Kawai, S. and Hanafusa, H., 1972, Genetic recombination with avian tumor virus, *Virology* **49**(1)**:**37-44.

Kettmann, R., Burny, A., Callebaut, I., Droogmans, L., Mammerickx, M., Willems, L. and Portetelle, D., 1992, Bovine leukemia virus, in: *The retroviridae* (J. A. levy, ed.), Plenum press, New York and London, pp. 39-81.

Kim, K.-I., Lee, J.-H., Li, K., Zhang, Y.-P., Lee, S.-S., Gongora, J. and Moran, C., Phylogenetic relationships of Asian and European pig breeds determined by mitochondrial DNA D-loop sequence polymorphism, *submitted to Animal Genetics*.

Kozak, C. A. and Ruscetti, S., 1992, Retroviruses in rodents, in: *The retroviridae* (J. A. levy, ed.), Plenum press, New York and London, pp. 405-481.

Kwok, C., Korn, R. M., Davis, M. E., Burt, D. W., Critcher, R., McCarthy, L., Paw, B. H., Zon, L. I., Goodfellow, P. N. and Schmitt, K., 1998, Characterization of whole genome radiation hybrid mapping resources for non-mammalian vertebrates, *Nucleic Acids Research* **26**(15)**:**3562-6.

Lahbib-Mansais, Y., Dalias, G., Milan, D., Yerle, M., Robic, A., Gyapay, G. and Gellin, J., 1999, A successful strategy for comparative mapping with human ESTs: 65 new regional assignments in the pig, *Mammalian Genome* **10**(2)**:**145-53.

Lahbib-Mansais, Y., Yerle, M., Pinton, P. and Gellin, J., 1996, Chromosomal localization of homeobox genes and associated markers on porcine chromosomes 3, 5, 12, 15, 16 and 18: comparative mapping study with human and mouse, *Mammalian Genome* **7**(3)**:**174-9.

Langford, G. A., Yannoutsos, N., Cozzi, E., Lancaster, R., Elsome, K., Chen, P., Richards, A. and White, D. J. G., 1994, Production of Pigs Transgenic For Human Decay Accelerating Factor, *Transplantation Proceedings* **26**(3)**:**1400-1401.

Lanneluc, I., Mulsant, P., Saidi-Mehtar, N. and Elsen, J. M., 1996, Synteny conservation between parts of human chromosome 4q and bovine and ovine chromosomes 6, *Cytogenetics & Cell Genetics* **72**(2-3)**:**212-4.

Lee, J. H., Zhang, W., Chen, Y., Lyons, L. A., Robic, A. and Moran, C., 1998, Developing a comparative porcine map relative to human chromosomes 9, 10, 20 and 22, *Animal Genetics* **29**(Suppl. 1)**:**39.

Leis, J., Baltimore, D., Bishop, J. M., Coffin, J., Fleissner, E., Goff, S. P., Oroszlan, S., Robinson, H., Skalka, A. M. and Temin, H. M., 1988, Standardized and simplified nomenclature for proteins common to all retroviruses, *Journal of Virology* **62**(5)**:**1808-9.

Lemieux, N., Dutrillaux, B. and Viegas, P. E., 1992, A simple method for simultaneous R or G-banding and fluorescence *in situ* hybridization of small single-copy genes, *Cytogenetics & Cell Genetics* **59**(4)**:**311-312.

Le Tissier, P., Stoye, J. P., Takeuchi, Y., Patience, C. and Weiss, R. A., 1997, Two sets of human-tropic pig retrovirus, *Nature* **389**(6652)**:**681-2.

Lewin, B., 1994, *Genes V*, Oxford University Press, Oxford.

Lieber, M. M., Sherr, C. J., Benveniste, R. E. and Todaro, G. J., 1975, Biological and immunological properties of porcine type C viruses, *Virology* **66:**616-619.

Loftus, S. K., Shiang, R., Warrington, J. A., Bengtsson, U., McPherson, J. D. and Wasmuth, J. J., 1994, Genes encoding adrenergic receptors are not clustered on the long arm of human chromosome 5, *Cytogenetics & Cell Genetics* **67**(2)**:**69-74.

Lomasney, J. W., Cotecchia, S., Lorenz, W., Leung, W. Y., Schwinn, D. A., Yang-Feng, T. L., Brownstein, M., Lefkowitz, R. J. and Caron, M. G., 1991, Molecular cloning and expression of the cDNA for the alpha 1A-adrenergic receptor. The gene for which is located on human chromosome 5, *Journal of Biological Chemistry* **266**(10)**:**6365-9.

Lyons, L. A., Kehler, J. S. and O'Brien, S. J., 1999, Development of comparative anchor tagged sequences (CATS) for canine genome mapping, *Journal of Heredity* **90**(1)**:**15-26.

Lyons, L. A., Laughlin, T. F., Copeland, N. G., Jenkins, N. A., Womack, J. E. and O'Brien, S. J., 1997, Comparative anchor tagged sequences (CATS) for integrative mapping of mammalian genomes, *Nature Genetics* **15:**47-56.

Mang, R., Goudsmit, J. and van der Kuyl, A. C., 1999, Novel endogenous type C retrovirus in baboons: complete sequence, providing evidence for baboon endogenous virus gag-pol ancestry, *Journal of Virology* **73**(8)**:**7021-6.

Marklund, L., Johansson Moller, M., Hoyheim, B., Davies, W., Fredholm, M., Juneja, R. K., Mariani, P., Coppieters, W., Ellegren, H. and Andersson, L., 1996, A comprehensive linkage map of the pig based on a wild pig-Large White intercross, *Animal Genetics* **27**(4)**:**255-69.

Martin, S. L., 1991, LINEs, *Current Opinion in Genetics & Development* **1**(4)**:**505-8.

Martin, J., Herniou, E., Cook, J., Waugh O'Neill, R. and Tristem, M., 1997, Human endogenous retrovirus type I-related viruses have an apparently widespread distribution within vertebrates, *Journal of Virology* **71**(1)**:**437-43.

Martin, U., Kiessig, V., Blusch, J. H., Haverich, A., von der Helm, K., Herden, T. and Steinhoff, G., 1998a, Expression of pig endogenous retrovirus by primary porcine endothelial cells and infection of human cells, *Lancet* **352**(9129)**:**692-4.

Martin, U., Steinhoff, G., Kiessig, V., Chikobava, M., Anssar, M., Morschheuser, T., Lapin, B. and Haverich, A., 1998b, Porcine endogenous retrovirus (PERV) was not transmitted from transplanted porcine endothelial cells to baboons in vivo, *Transplant International* **11**(4)**:**247-51.

McCarthy, L. C., 1996, Whole genome radiation hybrid mapping, *Trends in Genetics* **12**(12)**:**491-3.

McCurry, K. R., Kooyman, D. L., Diamond, L. E., Byrne, G. W., Martin, M. J., Logan, J. S. and Platt, J. L., 1995, Human complement regulatory proteins in transgenic animals regulate complement activation in xenoperfused organs, *Transplantation Proceedings* **27**(1)**:**317-8.

McGuire, T. C., O'Rourke, K. I., Knowles, D. P. and Cheevers, W. P., 1990, Caprine arthritis encephalitis lentivirus transmission and disease, *Current Topics in Microbiology & Immunology* **160:**61-75.

McIntosh, G. M. and Pointon, A., 1981, The Kangaroo Island strain of pig in biomedical research, *Australian Veterinary Journal* **57:**182-185.

Mead, R., Curnow, R. N. and Hasted, A. M., 1993, *Statistical Methods in Agriculture and Experimental Biology*, Chapman and Hall, London.

Melton, L., 1999, Inroads made into transplantation problems, *The Lancet* **354:**1272.

Michaels, M. G. and Simmons, R. L., 1994, Xenotransplant-associated zoonoses. Strategies for prevention, *Transplantation* **57**(1)**:**1-7.

Miller, J. R., 1994, Use of porcine interspersed repeat sequences in PCR-mediated genotyping, *Mammalian Genome* **5**(10)**:**629-32.

Miyamoto, M. M., Kraus, F., Laipis, P. J., Tanhauser, S. M. and Webb, S. D., 1993, Mitochondrial DNA phylogenies within Artiodactyla, in: *Mammal Phylogeny* (F. S. Szalay, M. J. Novacek and M. C. McKenna, eds.), Springer-Verlag, New York, pp. 268-281.

Mollnes, T. E. and Fiane, A. E., 1999, Xenotransplantation: how to overcome the complement obstacle?, *Molecular Immunology* **36**(4-5)**:**269-76.

Moran, C., 1998, Molecular genetics, in: *The genetics of the pig* (M. F. Rothschild and A. Ruvinsky, eds.), CAB International, New York, pp. 135-161.

Murphy, W. J., Menotti-Raymond, M., Lyons, L. A., Thompson, M. A. and O'Brien, S. J., 1999, Development of a feline whole genome radiation hybrid panel and comparative mapping of human chromosome 12 and 22 loci, *Genomics* **57**(1)**:**1-8.

Nadeau, J. H. and Sankoff, D., 1997, Landmarks in the Rosetta Stone of mammalian comparative maps, *Nature Genetics* **15:**6-7.

Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M. and Kumlin, E., 1987, Variable number of tandem repeat (VNTR) markers for human gene mapping, *Science* **235**(4796)**:**1616-22.

Needleman, S. B. and Wunsch, C. D., 1970, A general method applicable to the search for similarities in the amino acid sequence of two proteins, *Journal of Molecular Biology* **48**(3)**:**443-53.

Nicholas, F. W. and Harper, P., 1996, Inherited disorders: the comparative picture, *Australian Veterinay Journal* **73**(2)**:**64-66.

Nikaido, M., Rooney, A. P. and Okada, N., 1999, Phylogenetic relationships among cetartiodactyls based on insertions of short and long interspersed elements: hippopotamuses are the closest extant relatives of whales, *Proceedings of the National Academy of Sciences of the United States of America* **96**(18)**:**10261-6.

O'Brien, S. J., Cevario, S. J., Martenson, J. S., Thompson, M. A., Nash, W. G., Chang, E., Graves, J. A., Spencer, J. A., Cho, K. W., Tsujimoto, H. and Lyons, L. A., 1997, Comparative gene mapping in the domestic cat (*Felis catus*), *Journal of Heredity* **88**(5)**:**408-14.

O'Brien, S. J., Seuánez, H. N. and Womack, J. E., 1988, Mammalian genome organization: an evolutionary view, *Annual Review of Genetics.* **22:**323-351.

O'Brien, S. J., Womack, J. E., Lyons, L. A., Moore, K. J., Jekins, N. A. and Copeland, N. G., 1993, Anchored reference loci for comparative genome mapping in mammals, *Nature Genetics* **3:**103-112.

Ogawa, H., Kobayashi, T., Yokoyama, I., Nagasaka, T., Namii, Y., Hayashi, S., Kadomatsu, K., Muramatsu, T. and Takagi, H., 1999, Suppression of porcine xenoantigen expression by dorminant-negative effect of alpha-1,3-Galactosyltransferase (alpha-1,3-GT) splicing variants, *Transplantation Proceedings* **32:**58.

Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K. and Sekiya, T., 1989, Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms, *Proceedings of the National Academy of Sciences of the United States of America* **86**(8)**:**2766-70.

Paradis, K., Langford, G., Long, Z., Heneine, W., Sandstrom, P., Switzer, W. M., Chapman, L. E., Lockey, C., Onions, D. and Otto, E., 1999, Search for cross-species transmission of porcine endogenous retrovirus in patients treated with living pig tissue, *Science* **285**(5431)**:**1236-41.

Pathak, V. K., Hu, W.-S. and Temmin, H. M., 1990, Hypermutation and other variations in retroviruses, in: *Somatic hypermutation in V-regions* (E. J. Steele, ed.), CRC Press, pp. 149-157.

Pathak, V. K. and Temin, H. M., 1990a, Broad spectrum of in vivo forward mutations, hypermutations and mutational hotspots in a retroviral shuttle vector after a single replication cycle: deletions and deletions with insertions, *Proceedings of the National Academy of Sciences of the United States of America* **87**(16)**:**6024-8.

Pathak, V. K. and Temin, H. M., 1990b, Broad spectrum of in vivo forward mutations, hypermutations and mutational hotspots in a retroviral shuttle vector after a single replication cycle: substitutions, frameshifts and hypermutations, *Proceedings of the National Academy of Sciences of the United States of America* **87**(16)**:**6019-23.

Patience, C., Patton, G. S., Takeuchi, Y., Weiss, R. A., McClure, M. O., Rydberg, L. and Breimer, M. E., 1998, No evidence of pig DNA or retroviral infection in patients with short-term extracorporeal connection to pig kidneys, *Lancet* **352**(9129)**:**699-701.

Patience, C., Takeuchi, Y. and Weiss, R. A., 1997a, Infection of human cells by an endogenous retrovirus of pigs, *Nature Medicine* **3**(3)**:**282-6.

Patience, C., Wilkinson, D. A. and Weiss, R. A., 1997b, Our retroviral heritage, *Trends in Genetics* **13**(3)**:**116-20.

Payen, E., Saidi-Mehtar, N., Pailhoux, E. and Cotinot, C., 1995, Sheep gene mapping: assignment of ALDOB, CYP19, WT and SOX2 by somatic cell hybrid analysis, *Animal Genetics* **26**(5)**:**331-3.

Payne, L. N., 1992, Biology of avian retroviruses, in: *The retroviridae* (J. A. levy, ed.), Plenum press, New York and London, pp. 299-404.

Pearson, W. R. and Lipman, D. J., 1988, Improved Tools for Biological Sequence Analysis, *Proceedings of the National Academy of Sciences of the United States of America* **85:**2444-2448.

Pelzer, K. D. and Sprecher, D. J., 1993, Controlling BLV infection on dairy operations, *Veterinary Medicine* **88:**275-281.

Petropoulos, C., 1997, Appendix 2, Retroviral taxonomy, protein structure, sequences and genetic maps, in: *Retroviruses* (J. M. Coffin, S. H. Hughes and H. E. Varmus, eds.), Cold Sprong Harbor Laboratory Press, New York, pp. 757-805.

Pitkin, Z. and Mullon, C., 1999, Evidence of absence of porcine endogenous retrovirus (PERV) infection in patients treated with a bioartificial liver support system, *Artificial Organs* **23**(9)**:**829-33.

Platt, J. L. and Bach, F. H., 1991, The barrier to xenotransplantation, *Transplantation* **52**(6)**:**937-47.

Platt, J. L. and Lin, S. S., 1998, The future promises of xenotransplantation, in: *Annals of the New York Academy of Sciences* (J. Fishman, D. Sachs and R. Shaikh, eds.), The New York Academy of Sciences, New York, pp. 5-18.

Platt, J. L., Vercellotti, G. M., Lindman, B. J., Oegema, T. R., Jr., Bach, F. H. and Dalmasso, A. P., 1990, Release of heparan sulfate from endothelial cells. Implications for pathogenesis of hyperacute rejection, *Journal of Experimental Medicine* **171**(4)**:**1363-8.

Postlethwait, J. H., Yan, Y.-L., Gates, M. A., Horne, S., Amores, A., Brownlie, A., Donovan, A., Egan, E. S., Force, A., Gong, Z., Goutel, C., Fritz, A., Kelsh, R., Knapik, E., Liao, E., Paw, B., Ransom, D., Singer, A., Thomson, M., Abduljabbar, T. S., Yelick, P., Beier, D., Joly, J.-S., Larhammar, D., Rosa, F., Westerfield, M., Zon, L. I., Johnson, S. L. and Talbot, W. S., 1998, Vetebrate genome evolution and the zebrafish gene map, *Nature Genetics* **18:**345-349.

Preston, B. D., Poiesz, B. J. and Loeb, L. A., 1988, Fidelity of HIV-1 reverse transcriptase, *Science* **242**(4882)**:**1168-71.

Rabbitts, P., Impey, H., Heppell-Parton, A., Langford, C., Tease, C., Lowe, N., Bailey, D., Ferguson-Smith, M. and Carter, N., 1995, Chromosome specific paints from a high flow karyotype of the mouse, *Nature Genetics* **9:**369-375.

Randi, E., Lucchini, V. and Diong, C. H., 1996, Evolutionary genetics of the Suiformes as reconstructed using mtDNA sequencing, *Journal of Mammalian Evolution* **3**(2)**:**163-194.

Raudsepp, T., Fronicke, L., Scherthan, H., Gustavsson, I. and Chowdhary, B. P., 1996, Zoo-FISH delineates conserved chromosomal segments in horse and man, *Chromosome Research* **4:**218-225.

Rettenberger, G., Bruch, J., Beattie, C. W., Moran, C., Fries, R. and Hameister, H., 1995a, Chromosomal assignment of seventeen porcine microsatellites and genes by use of a somatic cell hybrid mapping panel, *Animal Genetics* **26**(4)**:**269-73.

Rettenberger, G., Bruch, J., Fries, R., Archibald, A. L. and Hameister, H., 1996, Assignment of 19 porcine type I loci by somatic cell hybrid analysis detects new regions of conserved synteny between human and pig, *Mammalian Genome* **7**(4)**:**275-9.

Rettenberger, G., Fries, R., Engel, W., Scheit, K. H., Dolf, G. and Hameister, H., 1994, Establishment of a partially informative porcine somatic cell hybrid panel and assignment of the loci for

transition protein 2 (TNP2) and protamine 1 (PRM1) to chromosome 3 and polyubiquitin (UBC) to chromosome 14, *Genomics* **21**(3)**:**558-66.

Rettenberger, G., Klett, C., Zechner, U., Bruch, J., Just, W., Vogel, W. and Hameister, H., 1995a, Zoo-FISH analysis: cat and human karyotypes closely resemble the putative ancestral mammalian karyotype, *Chromosome Research* **3:**479-486.

Rettenberger, G., Klett, C., Zechner, U., Kunz, J., Vogel, W. and Hameister, H., 1995b, Visualization of the conservation of synteny between humans and pig by heterologous chromosomal painting, *Genomics* **26:**372-378.

Robic, A., Riquet, J., Yerle, M., Milan, D., Lahbib-Mansais, Y., Dubut-Fontana, C. and Gellin, J., 1996, Porcine linkage and cytogenetic maps integrated by regional mapping of 100 microsatellites on somatic cell hybrid panel, *Mammalian Genome* **7**(6)**:**438-45.

Rogel-Gaillard, C., Bourgeaux, N., Billault, A., Vaiman, M. and Chardon, P., 1999, Construction of a swine BAC library: application to the characterization and mapping of porcine type C endoviral elements, *Cytogenetics & Cell Genetics* **85**(3-4)**:**205-11.

Rohrer, G. A., Alexander, L. J., Hu, Z., Smith, T. P., Keele, J. W. and Beattie, C. W., 1996, A comprehensive map of the porcine genome, *Genome Research* **6**(5)**:**371-91.

Rohrer, G. A., Alexander, L. J., Keele, J. W., Smith, T. P. and Beattie, C. W., 1994, A microsatellite linkage map of the porcine genome, *Genetics* **136**(1)**:**231-45.

Rønne, M., Stefanova, V., Di Berardino, D. and Strandby Poulsen, B., 1987, The R-banded karyotype of the domestic pig (Sus scrofa domestica L.), *Hereditas* **106**(2)**:**219-31.

Rosenberg, N. and Jolicoeur, P., 1997, Retroviral pathogenesis, in: *Retroviruses* (J. M. Coffin, S. H. Hughes and H. E. Varmus, eds.), Cold Sprong Harbor Laboratory Press, New York, pp. 475-585.

Rous, P., 1911, A sarcoma of the fowl transmissible by an agent separable from the tumor cells, *The Journal of Experimental Medicine* **13:**397-411.

Rous, P. and Murphy, J. B., 1913, Variations in a chicken sarcoma caused by a filterable agent, *The Journal of Experimental Medicine* **17:**219-231.

Ruddle, F. H., 1981, A new era in mammalian gene mapping: somatic cell genetics and recombinant DNA methodologies, *Nature* **294**(5837)**:**115-20.

Ruvinsky, A. and Rothschild, M. F., 1998, Systematics and evolution of the Pig, in: *The Genetics of the Pigs* (A. Ruvinsky and M. F. Rothschild, eds.), CAB International, Oxon and New York, pp. 1-16.

Sambrook, J., Fritsch, E. F. and Maniatis, T., 1989, *Molecular Cloning: a laboratory manual*, Cold Spring Harbour Laboratory Press.

Sanger, F., Nicklen, S. and Coulson, A. R., 1977, DNA sequencing with chain-terminating inhibitors, *Proceedings of the National Academy of Sciences of the United States of America* **74:**5463-5467.

Scherthan, H., Cremer, T., Arnason, U., Weier, H., Lima-de-Faria, A. and Frönicke, L., 1994, Comaprative chromosome painting disclosures homologous egments in distantly related mammals, *Nature Genetics* **6:**342-347.

Schlapfer, J., Gallagher, D. S., Jr., Burzlaff, J. D., Womack, J. E., Stelly, D. M., Taylor, J. F. and Davis, S. K., 1998, Comparative mapping of bovine chromosome 13 by fluorescence in situ hybridization, *Animal Genetics* **29**(4)**:**265-72.

Schmitz, A., Chardon, P., Gainche, I., Chaput, B., Guilly, M. N., Frelat, G. and Vaiman, M., 1992, Pig standard bivariate flow karyotype and peak assignment for chromosomes X, Y, 3 and 7, *Genomics* **14:**357-362.

Schmitz, A., Oustry, A., Chaput, B., Yerle, M., Milan, D., Frelat, G. and Cribui, E. P., 1995, The bovine bivariate flow karyotype and peak identification by chromosome painting with PCR-generated probes, *Mammalian Genome* **6:**415-420.

Schmoeckel, M., Bhatti, F. N., Zaidi, A., Cozzi, E., Waterworth, P. D., Tolan, M. J., Pino-Chavez, G., Goddard, M., Warner, R. G., Langford, G. A., Dunning, J. J., Wallwork, J. and White, D. J., 1998, Orthotopic heart transplantation in a transgenic pig-to-primate model [published erratum appears in Transplantation 1998 Oct 15;66(7):943], *Transplantation* **65**(12)**:**1570-7.

Schumacher, J. M., Ellias, S. A., Palmer, E. P., Kott, H. S., Dinsmore, J., Dempsey, P. K., Fischman, A. J., Thomas, C., Feldman, R. G., Kassissieh, S., Raineri, R., Manhart, C., Penney, D., Fink, J. S. and Isacson, O., 2000, Transplantation of embryonic porcine mesencephalic in patients with PD, *Neurology* **54**(5)**:**1042-1050.

Sharma, A., Okabe, J., Birch, P., McClellan, S. B., Martin, M. J., Platt, J. L. and Logan, J. S., 1996, Reduction in the level of Gal(alpha1,3)Gal in transgenic mice and pigs by the expression of an alpha(1,2)fucosyltransferase, *Proceedings of the National Academy of Sciences of the United States of America* **93**(14)**:**7190-5.

Singer, D. H., Parent, L. J. and Ehrlich, R., 1987, Identification and DNA sequence of an interspersed repetitive DNA element in the genome of the miniature swine, *Nucleic Acids Research* **15**(6)**:**2780.

Singer, M. F., 1982, SINEs and LINEs: highly repeated short and long interspersed sequences in mammalian genomes, *Cell* **28**(3)**:**433-4.

Smith, E. J., Lyons, L. A., Cheng, H. H. and Suchyta, S. P., 1997, Comparative mapping of the chicken genome using the East Lansing reference population, *Poultry Science* **76:**743-747.

Solinas-Toldo, S., Lengauer, C. and Fries, R., 1995, Comparative genome map of human and cattle, *Genomics* **27:**489-496.

Soller, M., 1994, Marker assisted selection - an review, *Animal Biotechnology* **5**(2)**:**193-207.

Spinardi, L., Mazars, R. and Theillet, C., 1991, Protocols for an improved detection of point mutations by SSCP, *Nucleic Acids Research* **19**(14)**:**4009.

Stephenson, J. R. anderson, G. R., Tronick, S. R. and Aaronson, S. A., 1974, Evidence for genetic recombination between endogenous and exogenous mouse RNA type C viruses, *Cell* **2**(2)**:**87-94.

Stewart, E. A., McKusick, K. B., Aggarwal, A., Bajorek, E., Brady, S., Chu, A., Fang, N., Hadley, D., Harris, M., Hussain, S., Lee, R., Maratukulam, A., O'Connor, K., Perkins, S., Piercy, M., Qin, F., Reif, T., Sanders, C., She, X., Sun, W. L., Tabar, P., Voyticky, S., Cowles, S., Fan, J. B. and Cox, D. R., 1997, An STS-based radiation hybrid map of the human genome, *Genome Research* **7**(5)**:**422-33.

Stoye, J. P. and Coffin, J. M., 1995, The dangers of xenotransplantation, *Nature Medicine* **1**(11)**:**1100.

Stoye, J. P., Le Tissier, P., Takeuchi, Y., Patience, C. and Weiss, R. A., 1998, Endogenous retroviruses: a potential problem for xenotransplantation?, in: *Annals of the New York Academy of Sciences* (J. Fishman, D. Sachs and R. Shaikh, eds.), The New York Academy of Sciences, New York, pp. 67-74.

Stuhlmann, H. and Berg, P., 1992, Homologous recombination of copackaged retrovirus RNAs during reverse transcription, *Journal of Virology* **66**(4)**:**2378-88.

Sulandari, S., Muladno, Harumi, T., Yanai, S., Wada, Y. and Yasue, H., 1997, Localization of swine PRE-1 homologues in 13 loci of Phacochoerus aethiopicus and Tayassu tajacu genomes and their sequence divergence, *Animal Genetics* **28:**210-215.

Sun, H. S., Yerle, M., Pinton, P., Chardon, P., Rogel-Gaillard, C., Lyons, L. A., Laughlin, T. F. and Tuggle, C. K., 1999, Physical assignments of human chromosome 13 genes on pig chromosome 11 demonstrate extensive synteny and gene order conservation between pig and human, *Animal Genetics* **30**(4)**:**304-308.

Suzuka, I., Sekiguchi, K. and Kodama, M., 1985, Some characteristics of a porcine retrovirus from a cell line derived from swine malignant lymphomas, *FEBS Letters* **183**(1)**:**124-8.

Suzuka, I., Shimizu, N., Sekiguchi, K., Hoshino, H., Kodama, M. and Shimotohno, K., 1986, Molecular cloning of unintegrated closed circular DNA of porcine retrovirus, *FEBS Letters* **198**(2)**:**339-43.

Swindle, M. M., 1998, Defining appropriate health status and magement program for specific pathogen free (SPF) swine for xenotransplantation, in: *Annals of the New York Academy of Sciences* (J. Fishman, D. Sachs and R. Shaikh, eds.), The New York Academy of Sciences, New York, pp. 111-120.

Switzer, W. M., Shanmugam, V., Chapman, L. and Heneine, W., 1999, Polymerase chain reaction assays for the diagnosis of infection with the porcine endogenous retrovirus and the detection of pig cells in human and nonhuman recipients of pig xenografts, *Transplantation* **68**(2)**:**183-8.

Takeuchi, Y., Patience, C., Magre, S., Weiss, R. A., Banerjee, P. T., Le Tissier, P. and Stoye, J. P., 1998, Host range and interference studies of three classes of pig endogenous retrovirus, *Journal of Virology* **72**(12)**:**9986-91.

Teich, N., 1982, Taxonomy of retroviruses, in: *RNA tumor viruses* (R. Weiss, N. Teich, H. Varmus and J. Coffin, eds.), Cold Sprong Harbor laboratory, pp. 25-207.

Temin, H., Hannink, M., Hu, W.-S. and Pathak , V. K., 1991, Retrovirus Variation and Regulation of *c-rel*, in: *Viruses That Affect the Immune System* (H. Fan, I. Chen, N. Rosenberg and W. Sugden, eds.), American Society for Microbiology, pp. 135-142.

Temin, H. M., 1991, Sex and recombination in retroviruses, *Trends in Genetics* **7**(3)**:**71-4.

Tereba, A., 1983, Asymmetric chromosomal distribution of endogenous retrovirus loci in chickens and mice, *Current Topics in Microbiology & Immunology* **107:**29-50.

Theimer, T. C. and Keim, P., 1998, Phylogenetic relationships of the peccaries based on mitochondrial cytochrome b DNA sequences, *Journal of Mammalogy* **79**(2)**:**566-572.

Thomas, J. W., Lee-Lin, S. Q. and Green, E. D., 1999, Human-mouse comparative mapping of the genomic region containing CDK6: localization of an evolutionary breakpoint, *Mammalian Genome* **10**(7)**:**764-767.

Thomsen, P. D. and Miller, J. R., 1996, Pig genome analysis: differential distribution of SINE and LINE sequences is less pronounced than in the human and mouse genomes, *Mammalian Genome* **7**(1)**:**42-6.

Todaro, G. J., Benveniste, R. E., Lieber, M. M. and Sherr, C. J., 1974, Characterization of a type C virus released from the porcine cell line PK (15), *Virology* **58:**65-74.

Todd, N. B., 1985, Significance of a diploid number of 20 in the peccary Catagonus wagneri, *Journal of Heredity* **76**(4)**:**310.

Trask, B. J., 1991a, Fluorescence in situ hybridization: applications in cytogenetics and gene mapping, *Trends in Genetics* **7**(5)**:**149-54.

Trask, B. J., 1991b, Gene mapping by in situ hybridization, *Current Opinion in Genetics & Development* **1**(1)**:**82-7.

Tristem, M., 1996, Amplification of divergent retroelements by PCR, *Biotechniques* **20**(4)**:**608-12.

Tristem, M., Herniou, E., Summers, K. and Cook, J., 1996a, Three retroviral sequences in amphibians are distinct from those in mammals and birds, *Journal of Virology* **70**(7)**:**4864-4870.

Tristem, M., Kabat, P., Lieberman, L., Linde, S., Karpas, A. and Hill, F., 1996b, Characterization of a novel murine leukemia virus-related subgroup within mammals, *Journal of Virology* **70**(11)**:**8241-6.

United Kingdom Transplant Services Association, 1998, *Transplant Statistics*.

van der Kuyl, A. C., Dekker, J. T. and Goudsmit, J., 1999, Discovery of a new endogenous type C retrovirus (FcEV) in cats: evidence for RD-114 being an FcEV(Gag-Pol)/baboon endogenous virus BaEV(Env) recombinant, *Journal of Virology* **73**(10)**:**7994-8002.

van der Kuyl, A. C., Mang, R., Dekker, J. T. and Goudsmit, J., 1997, Complete nucleotide sequence of simian endogenous type D retrovirus with intact genome organization: evidence for ancestry to simian retrovirus and baboon endogenous virus, *Journal of Virology* **71**(5)**:**3666-76.

van der Laan, L. J. W., Lockey, C., Griffeth, B. C., Frasier, F. S., Wilson, C. A., Onions, D. E., Hering, B. J., Long, Z., Otto, E., Torbett, B. E. and Salomon, D. R., 2000, Infection by porcine endogenous retrovirus after islet xenotransplantation in SCID mice, *Nature* **407:**501-504.

Vanhove, B., Renard, J.-P. and Soulillou, J.-P., 1998, Genetic engineering in the pig - Gene knockout and alternative technoloques, in: *Annals of the New York Academy of Sciences* (J. Fishman, D. Sachs and R. Shaikh, eds.), The New York Academy of Sciences, New York, pp. 28-36.

Van Poucke, M., Tornsten, A., Mattheeuws, M., Van Zeveren, A., Peelman, L. J. and Chowdhary, B. P., 1999, Comparative mapping between human chromosome 3 and porcine chromosome 13, *Cytogenetics and Cell Genetics* **85**(3-4)**:**279-284.

Venta, P. J., Brouillette, J. A., Yuzbasiyan-Gurkan, V., Brewer, G. J., 1996, Gene-specific universal mammalian sequence-tagged sites: Application to the canine genome, *Biochemical Genetics* **34**(7/8)**:**321-341.

Vogt, P. K., 1971, Genetically stable reassortment of markers during mixed infection with avian tumor viruses, *Virology* **46**(3)**:**947-52.

Vogt, P. K., 1997, Historical introduction to the general properties of retroviruses, in: *Retroviruses* (J. M. Coffin, S. H. Hughes and H. E. Varmus, eds.), Cold Sprong Harbor Laboratory Press, New York, pp. 1-25.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J. and Kuiper, M., 1995, AFLP: a new technique for DNA fingerprinting, *Nucleic Acids Research* **23**(21)**:**4407-14.

Wakefield, M. J. and Graves, J. A. M., 1996, Comparative maps of vertebrates, *Mammalian Genome* **7:**715-716.

Watson, J. D., 1992, *Recombinant DNA*, Scientific American Books, New York.

Weiss, R. A., Mason, W. S. and Vogt, P. K., 1973, Genetic recombinants and heterozygotes derived from endogenous and exogenous avian RNA tumor viruses, *Virology* **52**(2)**:**535-52.

Wieslander, L., 1979, A simple method to recover intact high molecular weight RNA and DNA after electrophoretic separation in low gelling temperature agarose gels, *Analytical Biochemistry* **98:**305-309.

Weiss, R. A., 1998, Transgenic pigs and virus adaptation, *Nature* **391**(6665)**:**327-8.

White, S. A. and Nicholson, M. L., 1999, Xenotransplantation, *British Journal of Surgery* **86**(12)**:**1499-514.

Williams, J. G., Kubelik, A. R., Livak, K. J., Rafalski, J. A. and Tingey, S. V., 1990, DNA polymorphisms amplified by arbitrary primers are useful as genetic markers, *Nucleic Acids Research* **18**(22)**:**6531-5.

Williamson, P., Holt, S., Townsend, S. and Boyd, Y., 1995, A somatic cell hybrid panel for mouse gene mapping characterized by PCR and FISH, *Mammalian Genome* **6**(6)**:**429-32.

Wilson, C. A., Wong, S., Muller, J., Davidson, C. E., Rose, T. M. and Burd, P., 1998, Type C retrovirus released from porcine primary peripheral blood mononuclear cells infects human cells, *Journal of Virology* **72**(4)**:**3082-7.

Wilson, C. A., Wong, S., VanBrocklin, M. and Federspiel, M. J., 2000, Extended analysis of the *in vitro* tropism of porcine endogenous retrovirus, *Journal of Virology* **74**(1)**:**49-56.

Womack, J. E., Johnson, J. S., Owens, E. K., Rexroad, C. E., 3rd, Schlapfer, J. and Yang, Y. P., 1997, A whole-genome radiation hybrid panel for bovine gene mapping, *Mammalian Genome* **8**(11)**:**854-6.

Womack, J. E. and Kata, S. R., 1995, Bovine genome mapping: evolutionary inference and the power of comparative genomics, *Current Opinion in Genetics and Development* **5:**725-733.

Wong, P. K. and McCarter, J. A., 1973, Genetic studies of temperature-sensitive mutants of Moloney-murine leukemia virus, *Virology* **53**(2)**:**319-26.

Wyke, J. A. and Beamand, J. A., 1979, Genetic recombination in Rous sarcoma virus: the genesis of recombinants and lack of evidence for linkage between pol, env and src genes in three factor crosses, *Journal of General Virology* **43**(2)**:**349-64.

Wyke, J. A., Bell, J. G. and Beamand, J. A., 1975, Genetic recombination among temperature-sensitive mutants of Rous sarcoma virus, *Cold Spring Harbor Symposia on Quantitative Biology* **39:**897-905.

Yang, F., Muller, S., Just, R., Ferguson-Smith, M. A. and Weinberg, J., 1997, Comparative chromosome painting in mammals: human and the Indian muntjak (*Muntiacus muntjak vaginalis*), *Genomics* **39:**396-401.

Yang, Y. P., Rexroad, C. E., 3rd, Schlapfer, J. and Womack, J. E., 1998, An integrated radiation hybrid map of bovine chromosome 19 and ordered comparative mapping with human chromosome 17, *Genomics* **48**(1)**:**93-9.

Yang-Feng, T. L., Xue, F. Y., Zhong, W. W., Cotecchia, S., Frielle, T., Caron, M. G., Lefkowitz, R. J. and Francke, U., 1990, Chromosomal organization of adrenergic receptor genes, *Proceedings of the National Academy of Sciences of the United States of America* **87**(4)**:**1516-20.

Yasue, H. and Wada, Y., 1996, A swine SINE (PRE-1 sequence) distribution in swine-related animal species and its phylogenetic analysis in swine genome, *Animal Genetics* **27:**95-98.

Yerle, M., Echard, G., Robic, A., Mairal, A., Dubut-Fontana, C., Riquet, J., Pinton, P., Milan, D., Lahbib-Mansais, Y. and Gellin, J., 1996, A somatic cell hybrid panel for pig regional gene mapping characterized by molecular cytogenetics, *Cytogenetics & Cell Genetics* **73**(3)**:**194-202.

Yerle, M., Lahbib-Mansais, Y., Pinton, P., Robic, A., Goureau, A., Milan, D. and Gellin, J., 1997, The cytogenetic map of the domestic pig (*Sus scrofa domestica*), *Mammalian Genome* **8:**592-607.

Yerle, M., Pinton, P., Robic, A., Alfonso, A., Palvadeau, Y., Delcros, C., Hawken, R., Alexander, L., Beattie, C., Schook, L., Milan, D. and Gellin, J., 1998, Construction of a whole-genome radiation hybrid panel for high-resolution gene mapping in pigs, *Cytogenetics & Cell Genetics* **82**(3-4)**:**182-8.

Zaidi, A., Schmoeckel, M., Bhatti, F., Waterworth, P., Tolan, M., Cozzi, E., Chavez, G., Langford, G., Thiru, S., Wallwork, J., White, D. and Friend, P., 1998, Life-supporting pig-to-primate renal xenotransplantation using genetically modified donors, *Transplantation* **65**(12)**:**1584-90.

Zhang, W., 1997, Pig gene mapping and genetic markers, *Ph.D. Thesis*, The University of Sydney.

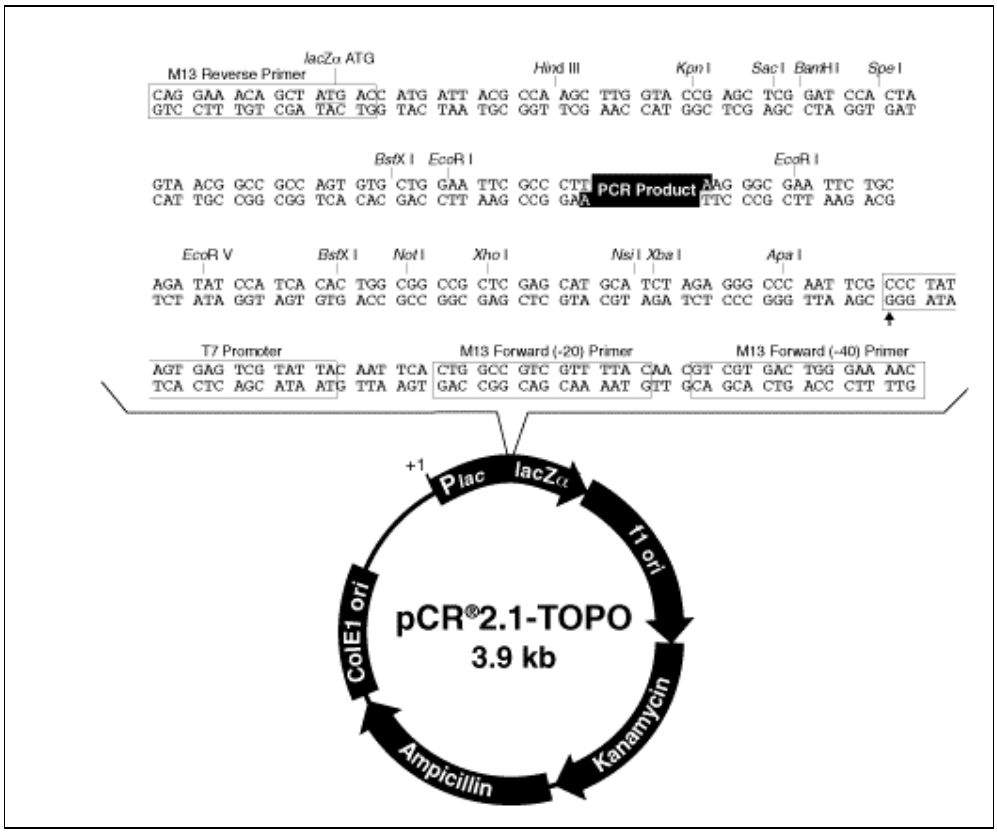
Zhang, W., Chen, Y. and Moran, C., 1996, Use of comparative anchor-tagged sequence (CATS) markers in pig gene mapping, *Animal Genetics* **27**(Suppl. 2)**:**84.

Zijlstra, C., Bosma, A. A., de Haan, N. A. and Mellink, C., 1996, Construction of a cytogenetically characterized porcine somatic cell hybrid panel and its use as a mapping tool, *Mammalian Genome* **7**(4)**:**280-4.

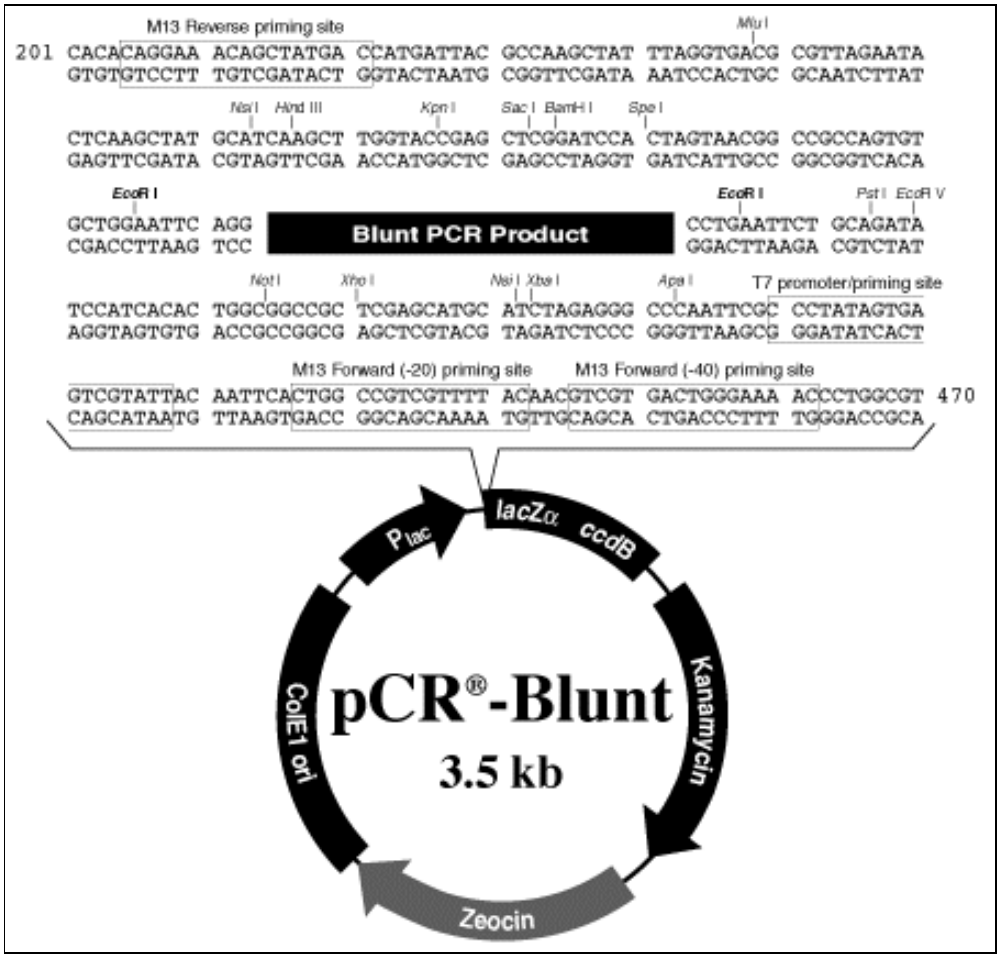
Zylstra, P., Rothenfluh, H. S., Weiller, G. F., Blanden, R. V. and Steele, E. J., 1998, PCR amplification of murine immunoglobulin germline V genes: strategies for minimization of recombination artefacts, *Immunology & Cell Biology* **76**(5)**:**395-405.

Appendices

Appendix 2.1 A physical map of the plasmid vector pCR®2.1-TOPO and the sequence surrounding the TOPO™ cloning site (source: Invitrogen, TOPO™ TA Cloning Instruction Manual, version E, pp 6).



Appendix 2.2 A physical map of the plasmid vector, pCR®-Blunt and the sequence surrounding the cloning site (source: Invitrogen, Zero Blunt™ PCR Cloning Kit Instruction Manual, version D, pp 4).



Appendix 2.3 Conversion of centrifugation force (*g*) to centrifugation speed (rpm).

(1) Beckman J2-21 M/E centrifuge

4,000 rpm = 1,935 *g* 5,000 rpm = 3,024 *g*

10,000 rpm = 12,096 *g* 15,000 rpm = 27,216 *g*

(2)Beckman microfuge E™ centrifuge

12,500 rpm = 14,000 *g*

Appendix 3.1 CATS sequences and verification of the PCR products for *ADRA1A* (a), *ADRA2A* (b), *ARSA* (c), *BMI1* (d), *CD40* (e), *GNAS1* (f), *GNAZ* (g), *IGL@* (h), *OXT* (i), *TOP1* (j) using the GenBank database.

(a) *ADRA1A*

```
The best scores are:                                      initn init1 opt z-sc E(283307)
CFGPCR5 Canis familiaris RDC5 mRNA for G p  (1695)  865  865  887 833.0 7.6e-40
HUMADRENA Homo sapiens (clone pmt2-huma1b)  (2669)  820  820  851 796.9 4.8e-38
HSU03865 Human adrenergic alpha-1b recepto  (1738)  820  820  851 799.1 5.6e-38
HUMA1AR Human alpha-1B-adrenergic receptor  (1560)  820  820  851 799.6 5.8e-38
HUMA1ADAR Human DNA for alphalA/D adrenerg  (2077)  720  720  831 779.4 5.9e-37
HUMA1AADR Human alpha-A1-adrenergic recept  (2002)  720  720  831 779.5 5.9e-37
HSU03864 Human adrenergic alpha-1a recepto  (1860)  720  720  831 779.9 6.1e-37
HUMA1DA Human alpha-1a/d adrenergic recept  (1831)  720  720  831 780.0 6.1e-37

HUMA1ADAR Human DNA for alphalA/D adrenergic receptor,  (2077 nt)
 initn: 720 init1: 720 opt: 831 Z-score: 779.4 expect() 5.9e-37
  75.4% identity in 276 nt overlap
                                                 10        20        30
/home/                                    CCCTTCTCNGCCNCCCTGGAGGTCCTGGGC
                                          :::::::: ::: :: ::::::: ::::::
HUMA1A GCCGACCTGCTGCTGAGCGCCACCGTACTGCCCTTCTCGGCCACCATGGAGGTTCTGGGC
          430       440       450       460       470       480

               40        50        60        70        80        90
/home/ TNCTGGGTNNTCGGCCGAGTCTTCTGTGACATCTGGGCCGCCGTGGACTTCCTGTGCTGN
       : :::::   : ::::: : ::::::: ::: : ::::::::::::::: : ::::::::
HUMA1A TTCTGGGCCTTTGGCCGCGCCTTCTGCGACGTATGGGCCGCCGTGGACGTGCTGTGCTGC
          490       500       510       520       530       540

              100       110       120       130       140       150
/home/ ACNGCCTCCATCCTCANCCTCTGCACCATCTCCNTCGACCGNTACATGGGTGTGCGCCAC
       :: :::::::::::: ::::::::::::::: : ::::: ::: :::: :::::::::
HUMA1A ACGGCCTCCATCCTCAGCCTCTGCACCATCTCCGTGGACCGGTACGTGGGCGTGCGCCAC
          550       560       570       580       590       600

              160       170       180       190       200       210
/home/ TCTCTCCAATACCCCNCCCTNGTCACCAAAAGGAAGGCNGCCTCNNTCCTCCCCNTNNTC
       :: ::: : :::::  :: :  : ::: :  : ::::: ::: :  ::::  :: :  ::
HUMA1A TCACTCAAGTACCCAGCCATCATGACCGAGCGCAAGGCGGCCGCCATCCTGGCCCTGCTC
          610       620       630       640       650       660

              220       230       240       250       260       270
/home/ TGGGTCTTNTCCACNGTCNTCTCCATGGGNCCTCTTCTTGGGTGGAAANGANCCNGCACC
       :::::: :  ::   ::  : ::: : :: :: :: :: :: ::: :: :: :: :  ::
HUMA1A TGGGTCGTAGCCCTGGTGGTGTCCGTAGGGCCCCTGCTGGGCTGG-AAGGAGCCCGTGCC
          670       680       690       700        710       720


/home/ CCCTGA
       ::::::
HUMA1A CCCTGACGAGCGCTTCTGCGGTATCACCGAGGAGGCGGGCTACGCTGTCTTCTCCTCCGT
           730       740       750       760       770       780
```

(b) *ADRA2A*

```
The best scores are:                                      initn init1 opt z-sc E(330345)
PIGA2AR Porcine alpha2A-adrenergic recep  (1728)  538  538  542 589.3 2.7e-26
HUMADRA2R Human alpha 2 adrenergic recep  (3604)  497  497  501 540.5 6.8e-24
CPU25722 Cavia porcellus alpha-2A adreno  (2291)  477  477  483 523.3 9.7e-23
HUMADRA Human platelet alpha-2-adrenergi  (1521)  470  470  474 515.7 3.9e-22
MUSALP2ADB Mouse alpha-2 adrenergic rece  (1454)  460  460  465 506.1 1.4e-21

PIGA2AR Porcine alpha2A-adrenergic receptor (PORA2AR) g (1728 nt)
 initn: 538 init1: 538 opt: 542 Z-score: 589.3 expect() 2.7e-26
 97.391% identity in 115 nt overlap

                                          10        20        30
/home/                                TCCAGGCCATAGAGTACAACCTGAAGCGCAC
                                      ::::::::::::::::::::::::::::::
PIGA2A CATCAGCTTGGATCGTTACTGGTCCATCACCCAGGCCATAGAGTACAACCTGAAGCGCAC
        510       520       530       540       550       560

              40        50        60        70        80        90
/home/ GCCNNGCCGCATCAAAGCAATCATCGTCACCGTGTGGGTCATCTCGGCCGTCATCTCCTT
       :::  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
PIGA2A GCCACGCCGCATCAAAGCAATCATCGTCACCGTGTGGGTCATCTCGGCCGTCATCTCCTT
        570       580       590       600       610       620

             100       110
/home/ CCCGCCGCTCATCTCCATCGGAGAAA
       :::::::::::::::::: :::::
PIGA2A CCCGCCGCTCATCTCCATC-GAGAAGAAGGCAGGCGGCGGTGGCCAGCAGCCGGCCGAAC
        630       640        650       660       670       680
```

(c) *ARSA*

```
The best scores are:                                      initn init1 opt z-sc
E(292158)
HUB384D8 Chromosome 22q13 BAC Clone CIT987 (139887)187  142  232 220.8 1.1e-07
MMDNAASFA M.musculus gene for arylsulfatas (4342)  116  116  232 238.9 3.3e-07
HSARYLA Human DNA for arylsulphatase A (EC (3637)  187  142  232 239.9 3.5e-07
MMRNAASFA M.musculus mRNA for arylsulfatas (2919)  116  116  232 241.0 3.8e-07
HSARYA Homo sapiens arylsulphatase A mRNA, (2022)  187  142  232 242.9 4.3e-07
HSARSA H.sapiens ARSA gene exon 2          ( 229)  187  142  173 191.8 0.0027

MMDNAASFA M.musculus gene for arylsulfatase A           (4342 nt)
 initn: 116 init1: 116 opt: 232 Z-score: 238.9 expect() 3.3e-07
  82.0% identity in 89 nt overlap

                                            10         20
/usr57                                AGTGG--ACTTGGGGTGGGG-CTGAGGGGG
                                      :::::   ::::: :::::: : :::::::
MMDNAA CTCGAGGCTACCTTACAGGGATGGCTGGCAAGTGGCATCTTGGAGTGGGGCCAGAGGGGG
     1120      1130      1140      1150      1160      1170

         30          40        50        60         70         80
/usr57 -CTTTCTGCCCC--CACCANGGCTTCCATCGATTCCTGGGCAT-CCATACT-CCATGACA
        ::: :::::::  :: :: :::::::: :::::::::::::: ::::: : :::::::
MMDNAA CCTTCCTGCCCCCGCATCAGGGCTTCCACCGATTCCTGGGCATCCCATATTCCCATGACC
     1180      1190      1200      1210      1220      1230
```

(d) *BMI1*

```
The best scores are:                                      initn init1 opt z-sc E(297024)
HUMBMI1X Human prot-oncogene (BMI-1) mRN (3203)   458  353  403 382.9 4.5e-15
S62198S1 flvi-2/bmi-1 [human, thymus, mR ( 867)   383  353  403 389.8 7e-15
MUSBMI1A Mouse B cell-specific Mo-MLV in (3600)   352  272  340 321.6 1.1e-11
MUSBMI1 Mouse zinc finger protein (bmi-1 (2959)   352  272  340 322.6 1.1e-11
MUSBMI1B Mouse zinc finger protein (bmi- (2791)   352  272  340 322.9 1.1e-11
XLU39959 Xenopus laevis proto-oncogene x (1777)   256  211  267 254.9 1.1e-07


HUMBMI1X Human prot-oncogene (BMI-1) mRNA, complete cds (3203 nt)
 initn: 458 init1: 353 opt: 403 Z-score: 382.9 expect() 4.5e-15
 92.632% identity in 95 nt overlap

                                    10        20         30
/home/                       CGTGNCGTNGTGAAGATAGANGAGA-ATTGCAGATGAAGA
                                      :::::::::: ::::  ::::::::::::
HUMBMI TCCTTCTGCTGATGCTGCCAATGGCTCTAATGAAGATAGAGGAGAGGTTGCAGATGAAGA
             790       800       810       820       830       840

      40        50        60        70        80        90
/home/ TAAGAGAATTATAACTGATGACGAGATAATAAGTTTATCCATTGAGTTCTTTGACCAGAA
       ::::::::::::::::::: :::::::::: :::::::::: ::::::::::::::
HUMBMI TAAGAGAATTATAACTGATGATGAGATAATAAGCTTATCCATTGAATTCTTTGACCAGAA
             850       860       870       880       890       900

     100       110       120       130       140       150
/home/ CAGGTAAATTCTCTAGGAAATGTATTNTATGCTAATATGTTTAGTAGGTACATTTTCCCC
       ::: :
HUMBMI CAGATTGGATCGGAAAGTAAACAAAGACAAAGAGAAATCTAAGGAGGAGGTGAATGATAA
             910       920       930       940       950       960
```

(e) *CD40*

```
The best scores are:                                      initn init1 opt z-sc E(281226)
HUMCD40L1 Human CD40 ligand (CD40L) gene,   (2395)  157  157  192 226.1 3.4e-06
N60998 TgESTzy23f11.r1 TgRH Tachyzoite cDN  ( 455)   68   68  114 140.1  1.1
N82885 TgESTzy35g02.r1 TgRH Tachyzoite cDN  ( 438)   68   68  114 140.3  1.1
N82175 TgESTzy44d06.r1 TgRH Tachyzoite cDN  ( 434)   68   68  114 140.4  1.1
W35539 TgESTzy90c10.r1 TgRH Tachyzoite cDN  ( 374)   68   68  114 141.2  1.2

HUMCD40L1 Human CD40 ligand (CD40L) gene, 5' flanking r (2395 nt)
 initn: 157 init1: 157 opt: 192 Z-score: 226.1 expect() 3.4e-06
  69.1% identity in 81 nt overlap

                                       10        20        30
/usr57                          TTTCTNCCNTTANNCAGTTTNCCTTCACAAAAAATTG
                                      : ::   : :: :  ::::: : ::   ::
HUMCD4 TTTCTTATCACCCAGATGATTGGGTCAGCACTTTTTGCTGTGTATCTTCATAGAAGGCTG
     2030      2040      2050      2060      2070      2080

        40        50        60        70        80        90
/usr57 GACAACGTAAGAAGAACCGTGGGCCTTTTGGAACTCAATTCCGGGTCCTTATCCTTACTG
       ::::: :::::: :::::    ::::::   :::: ::::  :::::::::
HUMCD4 GACAAGGTAAGATGAACCACAAGCCTTTATTAACTAAATTTGGGGTCCTTACTAATTCAT
     2090      2100      2110      2120      2130      2140
```

(f) *GNAS1*

```
The best scores are:                                      initn init1 opt z-sc E(467923)
HUMGNAS6 Human guanine nucleotide-binding  (2346)  521  330  633 676.0 5.4e-31
BTGSAR Bovine mRNA for adenylate cyclase-s (1599)  432  358  361 384.6 1.4e-14
BOVGTPA Bovine GTP-binding stimulatory pro (1493)  432  358  361 384.9 1.4e-14
SSGTPRNA S.scrofa mRNA for GTP-binding pro (1770)  462  358  359 381.9 1.7e-14
CFGTPBPAM C.familiaris mRNA for stimulator (1927)  444  340  341 362.0 2e-13

HUMGNAS6 Human guanine nucleotide-binding protein alpha (2346 nt)
 initn: 521 init1: 330 opt: 633 Z-score: 676.0 expect() 5.4e-31
  73.3% identity in 270 nt overlap

                                                 10         20
 267 a                                  TGCAGCCG--TCTNAA-CTCTTCANGAGCA
                                        :::::  :  ::: :: :::::::: :::::
HUMGNA TCATCCGGGAGGACAACCAGACCAACCGCCTGCAGGAGGCTCTGAACCTCTTCAAGAGCA
      810       820       830       840       850       860

        30        40        50        60        70        80
 267 a TCTGGAACAACAGGTTTGTGCGGTGACCCCCCGGCGTCCTCCCCACGGAGCTCGAGGCCC
       ::::::::::::::::::::  :::::: :::  :  ::: : :   :  :  :::::::
HUMGNA TCTGGAACAACAGGTTTGTGGAGTGACCGCCC-ACCCCCTGCGC-TTGCCCAGGAGGCCC
      870       880       890       900         910         920

        90       100       110       120       130       140
 267 a TGGATCTGCATGGTGCAGGGGAAGGAGTCCTGTNCAGGCACCCCANCCCCCT-GCANANA
       ::: ::::::  ::  :  :  : ::  :: :: :: :::  :::  ::::: ::  : :
HUMGNA TGG-TCTGCACTGTTTATAGAGAAGAACCCCGTGCAAGCATTCCAGACCCCTGGCCGAAA
         930       940       950       960       970       980

        150       160       170       180       190       200
 267 a GGGCGCCCCCCTTCCCGGGCGCTGANCCCGCTCTCCTCCGTGTTAAATGGCTGCGCACCA
       : ::::  :   ::::  ::  : :  : ::   ::: : :: :: ::::::::::::::
HUMGNA GCGCGCTTC---TCCCAAGCATTCACACGGCCTCCCTTCTTG-TAGATGGCTGCGCACCA
         990          1000      1010      1020       1030      1040

        210       220       230       240       250       260
 267 a TCTCTGTGATTCTGTTCCTCAACAAGCAANATCTGCTGGCTGANAAAGTCCTTGCTGGAA
       :::::::::: :::::::::::::::::: ::::::: ::::: :::::::::::::: :
HUMGNA TCTCTGTGATCCTGTTCCTCAACAAGCAAGATCTGCTCGCTGAGAAAGTCCTTGCTGGGA
            1050      1060      1070      1080      1090      1100
```

(g) *GNAZ*

```
The best scores are:                                      initn init1 opt z-sc E(483628)
HSAC000029 865e9 in GNAZ region of chrom22 (142887)432  432  496 578.8 2.3e-27
HUMGXA Human Gx-alpha gene                 (3957)  432  432  496 599.6 5.8e-27
HUMGNAZ Human transducin alpha-subunit (GN (2679)  432  432  496 601.8 6.4e-27
RATGXA Rat guanine nucleotide-binding regu (1529)  387  387  442 538.4 3.9e-23
CHKGPASA Gallus gallus Gi2 protein alpha-s (1400)  360  315  386 469.8 2.8e-19

HSAC000029 865e9 in GNAZ region of chrom22 (NO ACCESSIO (142887 nt)
 initn: 432 init1: 432 opt: 496 Z-score: 578.8 expect() 2.3e-27
  89.1% identity in 128 nt overlap

                                              10          20
 127 a                                 CGGGC-TCNTGGAGA--AAATTCACCTTCA
                                       ::::: :  ::::::  :: ::::::::::
HSAC00 AGGACATCCTGCGCTCCCGGGACATGACCACGGGCATTGTGGAGAACAAGTTCACCTTCA
   75840     75850     75860     75870     75880     75890

        30        40        50        60        70        80
 127 a AGGGAGCTCACCTTCAANATGGTGGACGTGGGCGGGCAGANGTCANAACGCAAAAAGTGG
       : ::::::::::::::: :::::::::::::: ::::::: :::: : ::::::::::::
HSAC00 A-GGAGCTCACCTTCAAGATGGTGGACGTGGGGGGGCAGAGGTCAGAGCGCAAAAAGTGG
   75900      75910     75920     75930     75940     75950

        90       100       110       120
 127 a ATCCNCTGCTTCGAGGGCGTNACAGCCATCATCTTCTGGT
       :::: ::::::::::::::: ::::::::::::::::
HSAC00 ATCCACTGCTTCGAGGGCGTCACAGCCATCATCTTCTGTGTGGAGCTCAGCGGCTACGAC
    75960     75970     75980     75990     76000     76010
```

(h) *IGL@*

```
The best scores are:                                      initn init1 opt z-sc
E(483402)
PIGIGLVJC Sus domesticus Ig rearranged lam  ( 681)  580  580  580 659.8 1.5e-29
MVIGLAM M.vison immunoglobulin lambda ligh  ( 805)  463  463  463 526.0 3.6e-22
S66247 Ig  lambda =immunoglobulin lambda c  ( 805)  463  463  463 526.0 3.6e-22
MVNAIL M.vison mRNA for non-active immunog  ( 606)  463  463  463 527.5 4e-22
MVIIGLVJCA Mink immunoglobulin lambda chai  ( 868)  454  454  454 515.4 1.3e-21

PIGIGLVJC Sus domesticus Ig rearranged lambda chain mRN (681 nt)
 initn: 580 init1: 580 opt: 580 Z-score: 659.8 expect() 1.5e-29
  92.4% identity in 131 nt overlap

                                            10        20        30
 134 a                                CACCCTGGTGTGTCTAATAAGTGACTTCTA
                                      ::::::::::::::::::::::::::::::
PIGIGL CTCCTCTGAGGAGCTCGGCACCAACAAGGCCACCCTGGTGTGTCTAATAAGTGACTTCTA
      270       280       290       300       310       320

               40        50        60        70        80        90
 134 a CCCGGGCGCCGTNACCGTGACCTGGAAAGCAGGCGGTACCACCGTCACCCATGGCGTNGA
       :::::::::::: :: :::::::::: :::::::: :::::::::::::: ::::: ::
PIGIGL CCCGGGCGCCGTGACGGTGACCTGGAAGGCAGGCGGCACCACCGTCACCCAGGGCGTGGA
      330       340       350       360       370       380

              100       110       120       130
 134 a AACCACCAANCCCTCGAAACAGANCAACAACAAGTNCGCGGGAA
        :::::::: :::::::::::: :::::::::: :::::
PIGIGL GACCACCAAGCCCTCGAAACAGAGCAACAACAAGTACGCGGCCAGCAGCTACCTGGCCCT
      390       400       410       420       430       440
```

(i) *OXT*

```
The best scores are:                                      initn init1 opt z-sc E(483060)
HUMOTNPI Human prepro-oxytocin-neurophysin (1338)  196  196  292 310.6 2.1e-10
OAON Sheep gene for oxytocin and neurophys (1632)  229  229  258 274.2 1.9e-08
BTHOR01 Bovine prepro-oxytocin-neurophysin (1167)  229  229  258 275.7 2.2e-08
RNVAOXY Rat genes for vasopressin, oxytoci (16257)  59   59  168 169.9 0.0012
MUSOXYNEUI Mouse oxytocin-neurophysin I ge (2003)   97   97  168 179.2 0.003

                                    10        20        30
 103 a                        GGTCCTTTNTGCCCCTGCGGNGGAAAGAANGGCCGTGCT
                                      ::::::   :: :: ::: : ::::: ::
HUMOTN GCGCTGACCTCCGCCTGCTACATCCAGAACTGCCCCCTGGGAGGCAAG-AGGGCCGCGCC
     460       470       480       490       500        510

      40        50        60        70        80        90
 103 a GGACCTCGACGTGCGCNAGGTGAGCGCCCANCCCTCGTCCCGCGGCGATCNGGGCTGGGG
       ::::::::::::::: :::::::  :::: :::: :::::::::: :: :::  :::
HUMOTN GGACCTCGACGTGCGCAAGGTGAGTCCCCAGCCCTGGTCCCGCGGCGCTCCGGGGAGGGA
      520       530       540       550       560       570

     100
 103 a GGAT
       ::
HUMOTN GGGACCCGCAGCCACAGGGGCGCGCCCCGCTCCGGCCTCGCCTGAGAACTCCAGGAGCTG
      590       600       610       620       630
```

(j) *TOP1*

```
The best scores are:                                      initn init1 opt z-sc
E(467806)
HUMTOP18 Homo sapiens type I DNA topoisome  ( 627)  208  110  520 573.4 1.1e-24
MMTOPIA M.musculus TOP gene for topoisomer  (2220)  135   83  318 342.8 2.1e-12
HUMTOPPG1 Human topoisomerase I pseudogene  (2416)   54   54  300 322.4 2.6e-11
I09478 Sequence 1 from Patent WO 8909222    (3645)   87   54  295 314.8 4.6e-11
HUMTOPI Human topoisomerase I mRNA, comple  (3645)   87   54  295 314.8 4.6e-11

HUMTOP18 Homo sapiens type I DNA topoisomerase gene, ex (627 nt)
 initn: 208 init1: 110 opt: 520 Z-score: 573.4 expect() 1.1e-24
  64.8% identity in 335 nt overlap

                                     10        20        30
 342 a                          AGCCTTTNAATATGTGCCTAGTGTGGTGTGAAGATGGG
                                      : ::::: ::: :: :  ::::::  : ::
HUMTOP CCAAGGTCATGAAGGATGCAAAGACGAAGAAGTATGTACCTGGTAT--TGTGAAAGTTGG
      200       210       220       230       240         250

       40        50        60        70        80              90
 342 a GGTTGGTTGAGGGAGAAAAATGGGCACAATATCAGCCAGAGGCCCC------AGATTTTC
       :: ::::     ::::::: :: :  ::  ::: : ::: ::::::       : :::::
HUMTOP GGCTGGT----AGAGAAAAGTGTG--CAGCATCTGTCAG-GGCCCCTGGGCTGGCTTTTC
        260           270         280        290       300

            100       110       120       130       140       150
 342 a TAAGGTTTCTGGGTGTTGTCTTTTAGAAATCTCTGTACCTGAGGCTTCTCTGCTTTTATT
        : :::::::: :   :::::::: ::::::::: :: ::  ::::: : :    : :
HUMTOP GATGGTTTCTGAGAAATGTCTTTTGGAAATCTCTATA-CTAGGGCTTTTATTGACTCAAA
     310       320       330       340         350       360

            160       170       180       190       200       210
 342 a GCCCCTAACTGGCAGCATGGGGTTACAGTGTTCCTNTCCAGAANCAAGTTCTTGATAACT
       :   :    :::  : :  : ::     :::      ::::   :  ::::::::  :::
HUMTOP GTGGCAGGATGG--GTACAGTGTGCTCTTGTCTAGAGCCAG-GCCTGGTTCTTGAGGACT
      370       380         390       400         410       420

            220       230       240       250       260       270
 342 a TCCTTTCTTCCCAGGGTGGTANANTCAAANAAAAAGGCCGTGCANANACTGGAGGANCA-
       : :: :::  : ::::: ::: : ::::: :: ::::: :: :: : ::::::::: ::
HUMTOP TGCTATCT--CTAGGGTAGTAGAGTCAAAGAAGAAGGCTGTTCAGAGACTGGAGGAACAG
         430         440       450       460       470       480

              280       290       300       310       320       330
 342 a CTGATGAA-CTCGAATTTCAGGCCCCNNACCGANANGAAAAATAACCNATTTGCTTTGGA
       ::::::: :: ::: :::: ::: :  ::::: : :::::  :::  :::  : : :::
HUMTOP TTGATGAAGCTGGAAGTTCAAGCCACAGACCGAGAGGAAAATAAACAGATTGCCCTGGGA
          490       500       510       520       530       540

              340
 342 a ACCCCACTNTAN
       ::: :
HUMTOP ACCTCCAAACTCAATTATCTGGACCCTAGGATCACAGTGGCTTGGTAAGGTTAGGCCTCT
          550       560       570       580       590       600
```

Appendix 3.2 *OXT* genotypes in PiGMaP reference families and their pedigree structure.

Genotype of reference family Britain-1

G-227LW	A
G-1104M	AB
G-497M	AB
G-521LW	A
P-9606 (227X1104)	A
P-9818 (497X521)	AB
P-9810 (497X521)	B
(9606X9818)	
O-6111	A
O-6112	AB
O-6113	A
O-6114	A
O-6115	A
O-6118	B
O-6119	A
O-6120	AB
O-6122	B
(9606X9810)	
O-5204	AB
O-5205	AB
O-5206	AB
O-5207	AB
O-5208	A
O-5209	A
O-5210	B
O-5211	B
O-5213	A
O-5214	A
O-5215	B

Appendix 3.2 continued

Genotype of reference family Britain-2

G-153M	A
G-833LW	A
G-956LW	B
G-433M	A

P-9591 (153X833)	A
P-9360 (956X433)	A
P-9365 (956X433)	B

(9591X9360)

O-4564	A
O-4565	?
O-4566	?
O-4568	?
O-4569	A
O-4570	?
O-4571	A
O-4572	?
O-4573	?
O-4574	A

(9591X9365)

O-5296	A
O-5297	B
O-5300	A
O-5303	B
O-5304	A
O-5306	?
O-5307	B
O-5308	?
O-5310	A
O-5311	A

Appendix 3.2 continued

Genotype of reference family German

G-181WB	?
G-128PT	AB
G-115PT	AB
G-113PT	B

P-281 (181X113)	?
P-203 (181X128)	A
P-232 (181X115)	B
P-233 (181X115)	A
P-282 (181X128)	B
P-202 (181X113)	B

(203X281)		(232X281)	
O-20301	B	O-23201	B
O-20302	B	O-23202	B
O-20303	B	O-23203	B
O-20304	AB	O-23204	B
O-20305	B	O-23205	B
O-20306	AB	O-23206	B
O-20307	B	O-23207	B
O-20316	AB	O-23208	B
O-20318	AB	O-23209	B
O-20319	B	O-23210	B
O-20320	B	O-23211	B
O-20322	B	O-23212	B
O-20323	AB		
O-20324	AB		

(233X281)		(282X202)	
O-23301	B	O-20201	B
O-23302	B	O-20202	B
O-23306	B	O-20204	B
O-23307	B	O-20205	B
O-23308	B	O-20206	B
O-23310	B		
O-23311	AB		
O-23315	AB		
O-23316	AB		

Appendix 3.2 continued

Genotype of reference family France

G-30607LW	A
G-20755M	A
G-30608LW	A
G-20690M	A
G-30848LW	A
G-20738M	A
P-10012(607x755)	A
P-10010(608x690)	A
P-10011(848x738)	A
P-10002(608x690)	A
(012x010)	
O-6748	N.T.
O-6749	N.T.
O-6750	N.T.
O-6751	N.T.
O-6752	N.T.
O-6753	N.T.
O-6754	N.T.
O-6755	N.T.
O-6756	N.T.
O-6757	N.T.
O-6758	N.T.
O-6759	N.T.
O-6760	N.T.
O-6761	N.T.
(002x011)	
O-6732	N.T.
O-6733	N.T.
O-6734	N.T.
O-6735	N.T.
O-6739	N.T.
O-6740	N.T.
O-6741	N.T.

Note: N.T. – Not Tested

Appendix 3.2 continued

Genotype of reference family Netherlands

G-12M	A
G-6818LW	?
G-17M	A
G-Z95V4LW	B
P-6664(12x6818)	A
P-ZSK43(17xZ95V4)	A
(6664xZSK43)	N.T.
O-7725	N.T.
O-7726	N.T.
O-7727	N.T.
O-7728	N.T.
O-7729	N.T.
O-7730	N.T.
O-7733	N.T.
O-7735	N.T.
O-7737	N.T.
O-7740	N.T.

Note: N.T. – Not Tested

Appendix 3.2 continued

Genotype of reference family Sweden

G-2WB	B
G-8LW	A
G-1WB	A
G-5LW	A
P-79(2x8)	A
P-51(1x5)	A
(79x51)	
O-199	N.T.
O-200	N.T.
O-201	N.T.
O-202	N.T.
O-203	N.T.
O-204	N.T.
O-205	N.T.
O-206	N.T.
O-207	N.T.
O-208	N.T.
O-368	N.T.
O-369	N.T.
O-371	N.T.
O-372	N.T.
O-373	N.T.
O-374	N.T.
O-375	N.T.
O-376	N.T.
O-377	N.T.
O-378	N.T.

Note: N.T. – Not Tested

Appendix 3.4 Interpreting PCR data of *ADRA1A* (a), *ADRA2A* (b), *ARSA* (c), *GNAS1* (d), *OXT* (e), *TOP1* (f) in somatic cell hybrid panel

(a) *ADRA1A*

```
 ***************   MARKER ADRA1A                ***************
 Supposed discordant rates :    False +: 0.10,  False -: 0.10
 clone  :   0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2
            1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7
 profile :  - + - + - + + - + + + + - - - + - - + - - - - - - - -
 Reliability of the following results:
 Error risk lower than 0.1%            Maximal Correlation = 0.92
 Chromosome probabilities
 1   0.20E-05   2  0.14E-08   3  0.75E-03   4  0.45E-09    5  0.10E-06
 6   0.28E-07   7  0.42E-08   8  0.11E-06   9  0.11E-06   10  0.26E-08
11   0.15E-10  12  0.54E-15  13  0.11E-05  14  0.32E-09   15  0.10E-05
16   0.10E+01  17  0.14E-08  18  0.18E-09   X  0.11E-06
Chrom 16 : P =1.00,  Region : Proba, Correl,  R+M+ / R-M+ / R+M- / R-M-
C16A    q11-q13              0.0006  0.6781      7      3      1     16
C16B    q14                  0.4730  0.9250     10      0      1     16
C16C    q21                  0.0526  0.8575     10      0      2     15
C16D    q22-q23              0.4730  0.9250     10      0      1     16
```

(b) *ADRA2A*

```
 ***************   MARKER ADRA2A                ***************
Supposed discordant rates :    False +: 0.10,  False -: 0.10
clone  :   0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2
           1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7
profile :  - - - - - - - - - - - - - - - - - - - + - - - - - + -

*** CAUTION : Low frequency of positive results  ( 0.07 )  ***

Reliability of the following results:

*** ERROR RISK ABOUT  5%              Maximal Correlation = 0.80

     RERUN YOUR PCR

Chromosome probabilities

 1  0.92E-06   2  0.66E-03    3  0.15E-09    4  0.79E-05    5  0.24E-13

 6  0.52E+00   7  0.27E-08    8  0.89E-06    9  0.46E-11   10  0.28E-10

11  0.16E-12  12  0.65E-04   13  0.14E-03   14  0.48E+00   15  0.17E-05

16  0.11E-08  17  0.64E-03   18  0.86E-04    X  0.25E-14

Chrom 6: P =0.52,  Region:  Proba,  Correl,  R+M+ / R-M+ / R+M- / R-M-

C06E1   1/2 q21             0.0465   0.6782     2      0      2     23

C06E2   1/2 q21             0.4181   0.8000     2      0      1     24

C06F    q22-q23             0.0465   0.4600     1      1      1     24

Chrom 14: P =0.48,  Region:  Proba,  Correl,  R+M+ / R-M+ / R+M- / R-M-

C14C    -                    0.0052   0.5933    2      0      3     22

C14E    -                    0.0052   0.5933    2      0      3     22

C14F    -                    0.0465   0.6782    2      0      2     23

C14G    -                    0.4181   0.8000    2      0      1     24
```

(c) *ARSA*

```
 ***************   MARKER ARSA                    ***************

 Supposed discordant rates :    False +: 0.10,  False -: 0.10

 clone  :   0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2
            1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7

 profile :  - + + + - + + - + + + - + + - - + + + - - + - - - - +

 Reliability of the following results:

 Error risk lower than 0.1%            Maximal Correlation = 0.93

 Chromosome probabilities

 1  0.65E-09   2  0.98E-13    3  0.36E-09    4  0.52E-11    5  0.10E+01

 6  0.73E-10   7  0.55E-13    8  0.80E-11    9  0.24E-07   10  0.39E-10

11  0.44E-12  12  0.61E-16   13  0.33E-10   14  0.33E-09   15  0.45E-12

16  0.23E-05  17  0.40E-11   18  0.40E-11    X  0.25E-07

Chrom 5: P =1.00,  Region :  Proba,  Correl,  R+M+ / R-M+ / R+M- / R-M-

C05A    p11-p15             0.9997   0.9282    14      1      0     12

C05B    q11                 0.0002   0.6236    13      2      3      9

C05C    q12-(1/2 q21)       0.0000   0.4807    11      4      3      9

C05D    (1/2 q21)-q24       0.0002   0.6298    12      3      2     10

C05E    q25                 0.0000   0.5635    11      4      2     10
```

(d) *GNSA1*

```
 ***************   MARKER GNAS1                 ***************
 Supposed discordant rates :    False +: 0.10,  False -: 0.10
 clone  :   0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2
            1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7
 profile :  - - - - - - - - - - - - - - - - - - - - + + - - - - -
 *** CAUTION : Low frequency of positive results  ( 0.07 )  ***
 Reliability of the following results:
 Error risk lower than 0.5%            Maximal Correlation = 1.00
 Chromosome probabilities
 1  0.21E-06   2  0.12E-01    3  0.35E-10    4  0.23E-07    5  0.46E-12
 6  0.18E-03   7  0.50E-07    8  0.28E-08    9  0.35E-11   10  0.31E-11
11  0.38E-13  12  0.15E-04   13  0.27E-02   14  0.14E-02   15  0.39E-06
16  0.25E-09  17  0.98E+00   18  0.20E-04    X  0.73E-17
Chrom 17: P =0.98,  Region:  Proba,  Correl,  R+M+ / R-M+ / R+M- / R-M-
C17A    q11-q14             0.0108   0.6782     2     0      2     23
C17B    1/2 q21             0.0972   0.8000     2     0      1     24
C17C    (1/2 q21)-q23       0.8752   1.0000     2     0      0     25
```

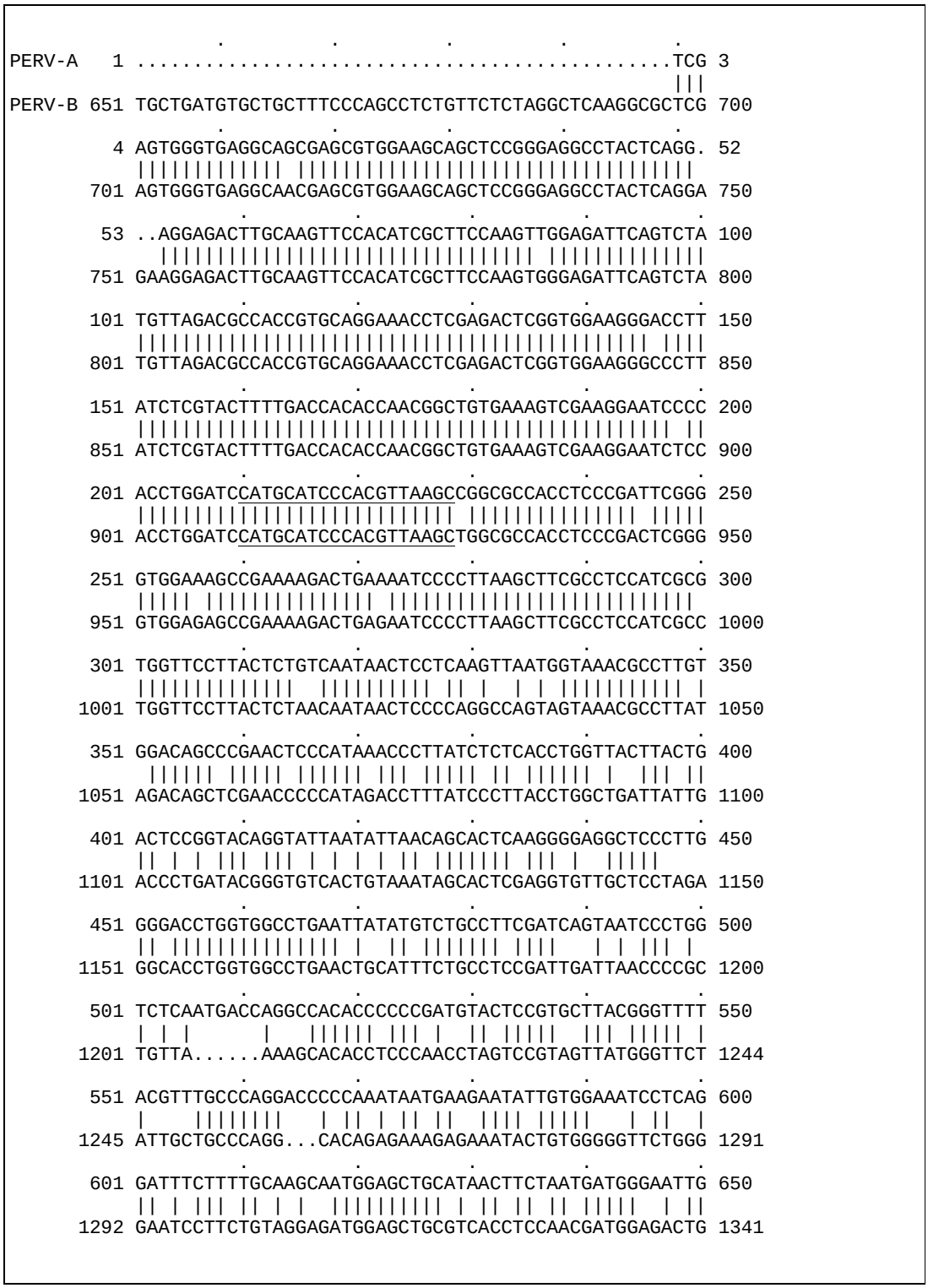
(e) *OXT*

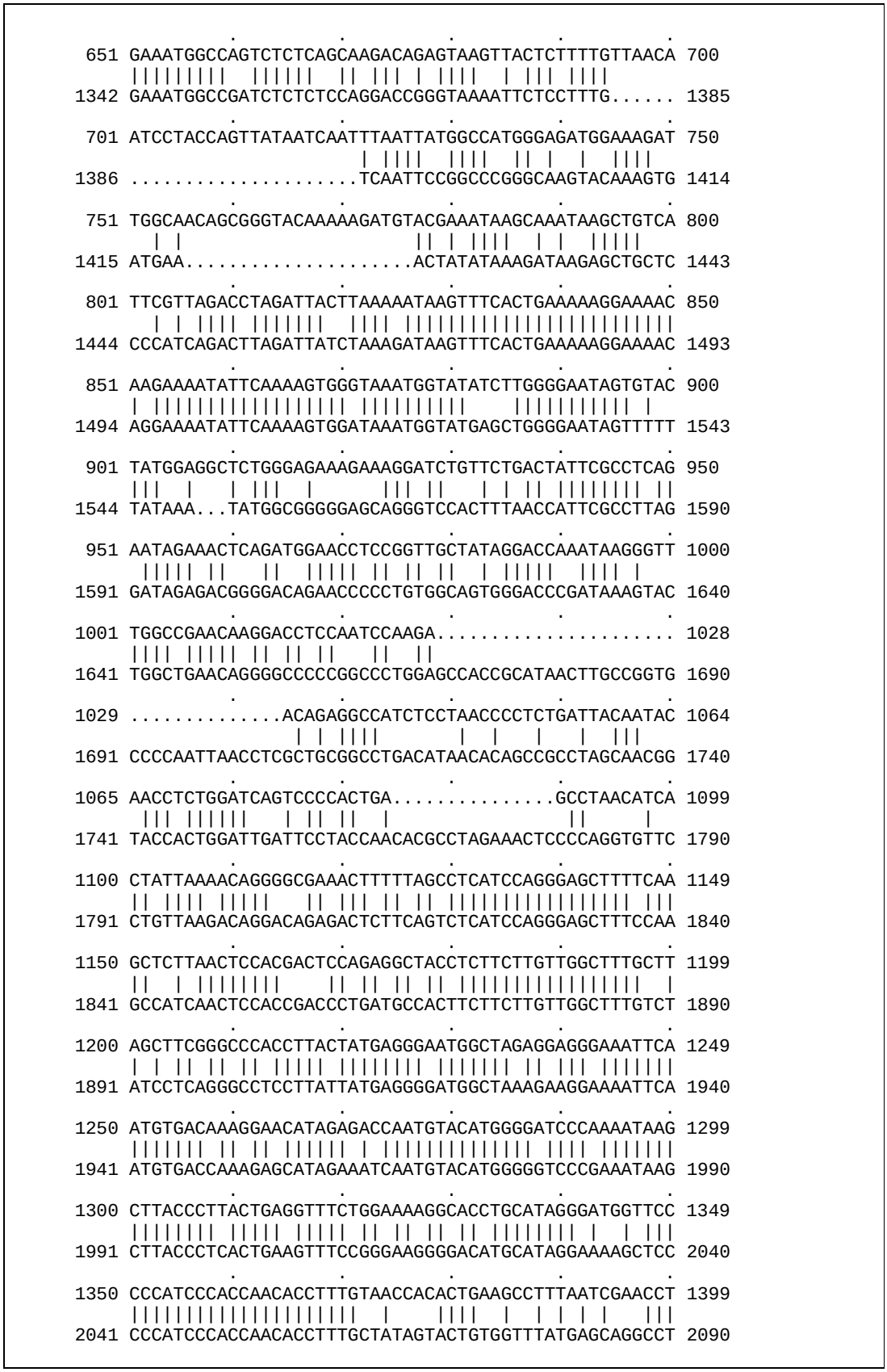
```
 ***************   MARKER OXT                    ***************
Supposed discordant rates :    False +: 0.10,  False -: 0.10
clone  :   0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2
           1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7
profile :  - - - - - - - - - - - - - - - - - - - - - - + - - - + -

*** CAUTION : Low frequency of positive results  ( 0.07 )  ***

Reliability of the following results:

*** ERROR RISK ABOUT  5%              Maximal Correlation = 0.80

     RERUN YOUR PCR

Chromosome probabilities

 1  0.15E-05   2  0.11E-02    3  0.25E-09    4  0.16E-06    5  0.33E-11

 6  0.10E+00   7  0.44E-08    8  0.16E-05    9  0.25E-10   10  0.58E-12

11  0.22E-10  12  0.11E-03   13  0.19E-01   14  0.79E+00   15  0.28E-05

16  0.18E-08  17  0.87E-01   18  0.14E-03    X  0.42E-14

Chrom 6: P =0.10,  Region :  Proba,  Correl,  R+M+ / R-M+ / R+M- / R-M-

C06E2   1/2 q21              0.0086  0.3500     1      1      2     23

C06F    q22-q23              0.0770  0.4600     1      1      1     24

C06G1   q24-(1/2 q31)        0.0086  0.3500     1      1      2     23

C06G2   1/2 q31              0.0086  0.3500     1      1      2     23

Chrom 14: P =0.79,  Region:  Proba,  Correl,  R+M+ / R-M+ / R+M- / R-M-

C14C    -                   0.0086   0.5933     2      0      3     22

C14D1   -                   0.0001   0.4781     2      0      5     20

C14D2   -                   0.0010   0.5292     2      0      4     21

C14E    -                   0.0086   0.5933     2      0      3     22

C14F    -                   0.0770   0.6782     2      0      2     23

C14G    -                   0.6933   0.8000     2      0      1     24
```

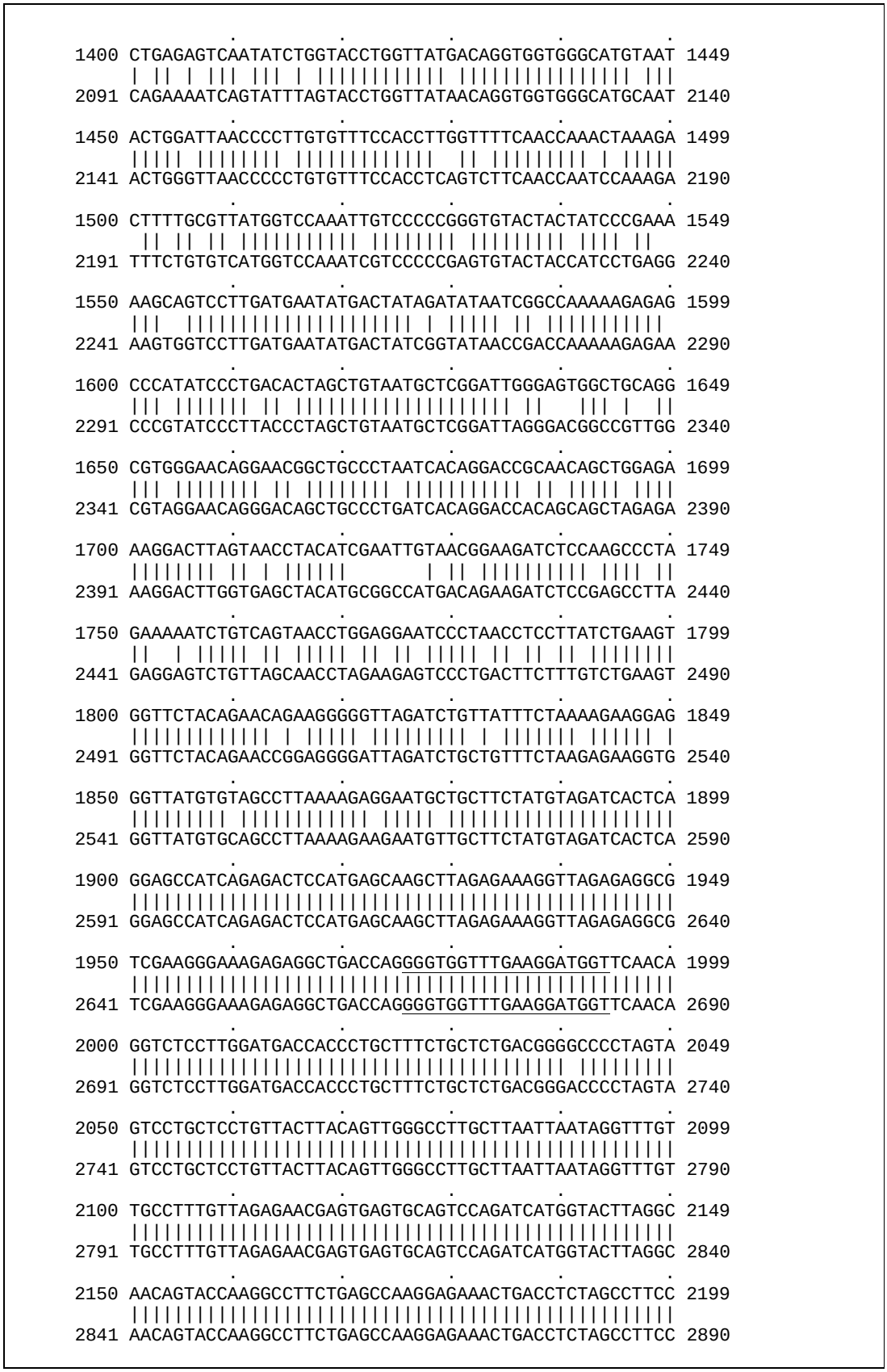
(f) *TOP1*

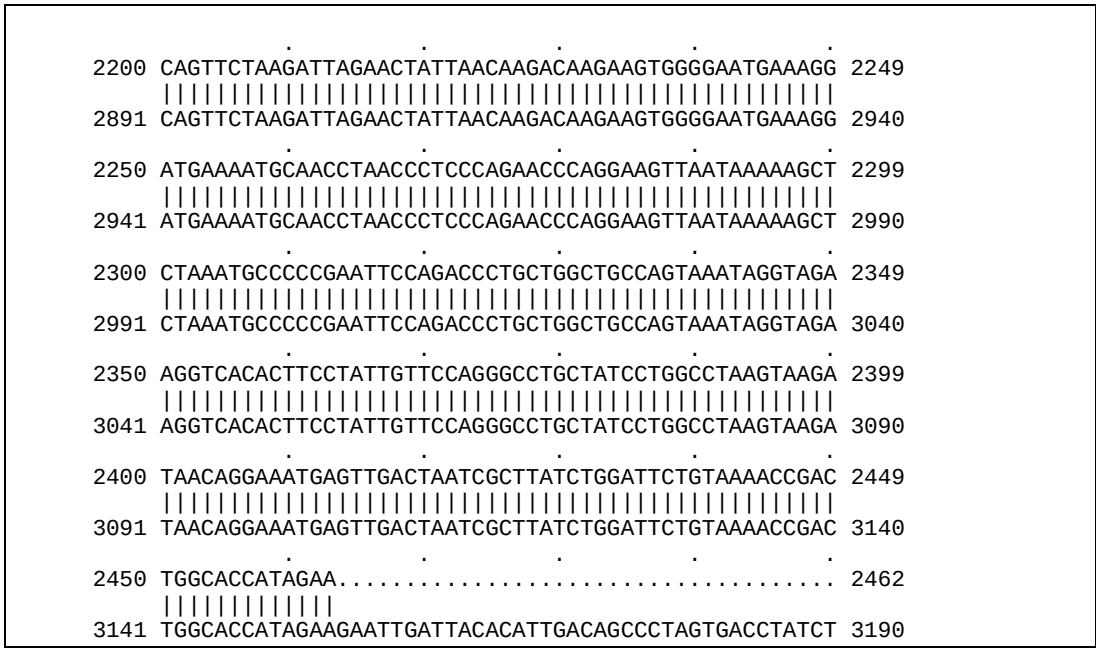
```
 ***************   MARKER TOP1                    ***************

 Supposed discordant rates :    False +: 0.10,  False -: 0.10

 clone  :   0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2
            1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7

 profile :  - - - - - - - - - - - - - - - - - - - - + + - - - - -

 *** CAUTION : Low frequency of positive results  ( 0.07 )  ***

 Reliability of the following results:

 Error risk lower than 0.5%            Maximal Correlation = 1.00

 Chromosome probabilities

 1  0.21E-06   2  0.12E-01    3  0.35E-10    4  0.23E-07    5  0.46E-12

 6  0.18E-03   7  0.50E-07    8  0.28E-08    9  0.35E-11   10  0.31E-11

11  0.38E-13  12  0.15E-04   13  0.27E-02   14  0.14E-02   15  0.39E-06

16  0.25E-09  17  0.98E+00   18  0.20E-04    X  0.73E-17

Chrom 17: P =0.98,  Region:  Proba,  Correl,  R+M+ / R-M+ / R+M- / R-M-

C17A    q11-q14             0.0108   0.6782     2      0      2     23

C17B    1/2 q21             0.0972   0.8000     2      0      1     24

C17C    (1/2 q21)-q23       0.8752   1.0000     2      0      0     25
```

Appendix 4.1. Alignment of published PERV-A and PERV-B *env* sequences (EMBL accession numbers Y12238 and Y12239) showing highly conserved regions are located at each end, flanking a variable region. Primer sequences were presented with underline (<u>underline</u>).






1400 CTGAGAGTCAATATCTGGTACCTGGTTATGACAGGTGGTGGGCATGTAAT 1449
2091 CAGAAAATCAGTATTTAGTACCTGGTTATAACAGGTGGTGGGCATGCAAT 2140
1450 ACTGGATTAACCCCTTGTGTTTCCACCTTGGTTTTCAACCAAACTAAAGA 1499
2141 ACTGGGTTAACCCCCTGTGTTTCCACCTCAGTCTTCAACCAATCCAAAGA 2190
1500 CTTTTGCGTTATGGTCCAAATTGTCCCCCGGGTGTACTACTATCCCGAAA 1549
2191 TTTCTGTGTCATGGTCCAAATCGTCCCCCGAGTGTACTACCATCCTGAGG 2240
1550 AAGCAGTCCTTGATGAATATGACTATAGATATAATCGGCCAAAAAGAGAG 1599
2241 AAGTGGTCCTTGATGAATATGACTATCGGTATAACCGACCAAAAAGAGAA 2290
1600 CCCATATCCCTGACACTAGCTGTAATGCTCGGATTGGGAGTGGCTGCAGG 1649
2291 CCCGTATCCCTTACCCTAGCTGTAATGCTCGGATTAGGGACGGCCGTTGG 2340
1650 CGTGGGAACAGGAACGGCTGCCCTAATCACAGGACCGCAACAGCTGGAGA 1699
2341 CGTAGGAACAGGGACAGCTGCCCTGATCACAGGACCACAGCAGCTAGAGA 2390
1700 AAGGACTTAGTAACCTACATCGAATTGTAACGGAAGATCTCCAAGCCCTA 1749
2391 AAGGACTTGGTGAGCTACATGCGGCCATGACAGAAGATCTCCGAGCCTTA 2440
1750 GAAAAATCTGTCAGTAACCTGGAGGAATCCCTAACCTCCTTATCTGAAGT 1799
2441 GAGGAGTCTGTTAGCAACCTAGAAGAGTCCCTGACTTCTTTGTCTGAAGT 2490
1800 GGTTCTACAGAACAGAAGGGGGTTAGATCTGTTATTTCTAAAAGAAGGAG 1849
2491 GGTTCTACAGAACCGGAGGGGATTAGATCTGCTGTTTCTAAGAGAAGGTG 2540
1850 GGTTATGTGTAGCCTTAAAAGAGGAATGCTGCTTCTATGTAGATCACTCA 1899
2541 GGTTATGTGCAGCCTTAAAAGAAGAATGTTGCTTCTATGTAGATCACTCA 2590
1900 GGAGCCATCAGAGACTCCATGAGCAAGCTTAGAGAAAGGTTAGAGAGGCG 1949
2591 GGAGCCATCAGAGACTCCATGAGCAAGCTTAGAGAAAGGTTAGAGAGGCG 2640
1950 TCGAAGGGAAAGAGAGGCTGACCAGGGGTGGTTTGAAGGATGGTTCAACA 1999
2641 TCGAAGGGAAAGAGAGGCTGACCAGGGGTGGTTTGAAGGATGGTTCAACA 2690
2000 GGTCTCCTTGGATGACCACCCTGCTTTCTGCTCTGACGGGGCCCCTAGTA 2049
2691 GGTCTCCTTGGATGACCACCCTGCTTTCTGCTCTGACGGGACCCCTAGTA 2740
2050 GTCCTGCTCCTGTTACTTACAGTTGGGCCTTGCTTAATTAATAGGTTTGT 2099
2741 GTCCTGCTCCTGTTACTTACAGTTGGGCCTTGCTTAATTAATAGGTTTGT 2790
2100 TGCCTTTGTTAGAGAACGAGTGAGTGCAGTCCAGATCATGGTACTTAGGC 2149
2791 TGCCTTTGTTAGAGAACGAGTGAGTGCAGTCCAGATCATGGTACTTAGGC 2840
2150 AACAGTACCAAGGCCTTCTGAGCCAAGGAGAAACTGACCTCTAGCCTTCC 2199
2841 AACAGTACCAAGGCCTTCTGAGCCAAGGAGAAACTGACCTCTAGCCTTCC 2890


2200 CAGTTCTAAGATTAGAACTATTAACAAGACAAGAAGTGGGGAATGAAAGG 2249
2891 CAGTTCTAAGATTAGAACTATTAACAAGACAAGAAGTGGGGAATGAAAGG 2940
2250 ATGAAAATGCAACCTAACCCTCCCAGAACCCAGGAAGTTAATAAAAAGCT 2299
2941 ATGAAAATGCAACCTAACCCTCCCAGAACCCAGGAAGTTAATAAAAAGCT 2990
2300 CTAAATGCCCCCGAATTCCAGACCCTGCTGGCTGCCAGTAAATAGGTAGA 2349
2991 CTAAATGCCCCCGAATTCCAGACCCTGCTGGCTGCCAGTAAATAGGTAGA 3040
2350 AGGTCACACTTCCTATTGTTCCAGGGCCTGCTATCCTGGCCTAAGTAAGA 2399
3041 AGGTCACACTTCCTATTGTTCCAGGGCCTGCTATCCTGGCCTAAGTAAGA 3090
2400 TAACAGGAAATGAGTTGACTAATCGCTTATCTGGATTCTGTAAAACCGAC 2449
3091 TAACAGGAAATGAGTTGACTAATCGCTTATCTGGATTCTGTAAAACCGAC 3140
2450 TGGCACCATAGAA..... 2462
3141 TGGCACCATAGAAGAATTGATTACACATTGACAGCCCTAGTGACCTATCT 3190

Appendix 4.2.

(a) Restriction enzyme digestion patterns for PERV *env* gene clones amplified by *Taq* polymerase.

Clone No.	KpnI type	MboI Type	Combined Types	Clone No.	KpnI type	MboI Type	Combined Types
1	A	K	AK	66	B	L	BL
3	B	L	BL	68	B	L	BL
4	B	L	BL	72	B	L	BL
6	A	M	AM	73	A	P	AP
8	A	N	AN	74	B	L	BL
9	B	L	BL	76	B	L	BL
10	A	M	AM	77	B	L	BL
11	A	M	AM	78	B	L	BL
12	B	L	BL	79	B	L	BL
14	B	L	BL	80	B	L	BL
16	A	M	AM	81	B	L	BL
17	A	O	AO	82	A	N	AN
20	B	L	BL	83	B	L	BL
21	C	L	CL	84	B	L	BL
22	A	M	AM	86	B	L	BL
24	B	N	BN	87	B	L	BL
29	D	L	DL	89	B	L	BL
30	B	L	BL	90	B	L	BL
34	B	L	BL	92	B	L	BL
35	A	M	AM	93	B	L	BL
36	B	L	BL	95	A	P	AP
37	B	L	BL	96	B	L	BL
38	A	M	AM	97	A	N	AN
39	B	L	BL	98	A	N	AN
40	A	M	AM	99	B	L	BL
41	B	L	BL	100	B	L	BL
45	B	L	BL				
48	B	L	BL				
49	A	M	AM				
50	A	M	AM				
53	B	L	BL				
54	A	P	AP				
55	B	L	BL				
56	A	P	AP				
60	B	L	BL				
61	B	L	BL				
62	B	L	BL				
65	B	L	BL				

(b) Restriction enzyme digestion patterns for PERV *env* gene clones amplified by *Pfu* proof reading polymerase.

Clone No.	KpnI type	MboI Type	Combined Types	Clone No.	KpnI type	MboI Type	Combined Types
1	B	L	BL	43	B	L	BL
2	B	L	BL	44	B	L	BL
3	A	P	AP	45	A	P	AP
4	B	L	BL	46	B	L	BL
5	A	L	AP	47	B	L	BL
6	A	M	AM	48	B	L	BL
7	B	L	BL	49	B	L	BL
8	B	L	BL	50	B	L	BL
10	B	L	BL	51	B	L	BL
11	A	N	AN	52	B	L	BL
12	B	L	BL	53	B	L	BL
13	B	L	BL	54	B	L	BL
14	B	L	BL	55	B	L	BL
15	B	L	BL	56	A	Q	AQ
16	A	P	AP	57	A	M	AM
17	B	L	BL	58	A	P	AP
18	B	L	BL	60	A	M	AM
20	B	L	BL	61	B	L	BL
21	B	L	BL	62	A	R	AR
22	B	L	BL	63	A	M	AM
23	B	L	BL	64	B	L	BL
24	B	L	BL	65	B	L	BL
25	A	N	AN	66	A	M	AM
26	B	L	BL	67	A	M	AM
27	B	L	BL	68	B	L	BL
28	B	L	BL	69	B	L	BL
29	B	L	BL	71	A	N	AN
30	B	L	BL	72	A	N	AN
31	B	L	BL	73	A	P	AP
32	B	L	BL	74	B	L	BL
33	B	L	BL	75	A	P	AP
34	B	L	BL	76	B	L	BL
35	B	L	BL	77	B	L	BL
36	B	L	BL	78	B	L	BL
37	B	L	BL	79	B	L	BL
38	B	L	BL	80	B	L	BL
39	B	L	BL	81	A	P	AP
40	B	L	BL	82	B	L	BL
42	B	L	BL	83	B	L	BL

Clone No.	KpnI type	MboI Type	Combined Types	Clone No.	KpnI type	MboI Type	Combined Types
84	B	L	BL	128	B	L	BL
85	B	L	BL	129	B	L	BL
86	B	L	BL	130	B	L	BL
87	B	L	BL	131	A	M	AM
88	A	N	AN	132	B	L	BL
89	A	P	AP	133	B	L	BL
90	B	L	BL	134	A	P	AP
91	B	L	BL	135	B	L	BL
92	B	L	BL	136	B	L	BL
93	A	P	AP	137	B	L	BL
94	B	L	BL	138	B	L	BL
95	B	L	BL	139	B	L	BL
97	B	L	BL	140	A	P	AP
98	A	P	AP	141	B	L	BL
99	B	L	BL	142	B	L	BL
100	B	L	BL	143	B	L	BL
101	B	L	BL	144	A	P	AP
102	B	L	BL	146	B	L	BL
103	A	P	AP	147	B	L	BL
104	B	L	BL	148	B	L	BL
105	B	L	BL	149	B	L	BL
106	B	L	BL	150	B	L	BL
107	A	M	AM	151	B	L	BL
109	B	L	BL	152	B	L	BL
110	B	L	BL	153	B	L	BL
111	B	L	BL	154	A	M	AM
112	A	S	AS	155	B	L	BL
115	A	T	AT	156	B	L	BL
116	A	M	AM	157	A	P	AP
117	B	L	BL	158	B	L	BL
118	B	L	BL	159	B	L	BL
119	A	P	AP	160	B	L	BL
120	A	P	AP	161	B	L	BL
121	A	P	AP	162	B	L	BL
122	B	L	BL	163	B	L	BL
123	B	L	BL	164	A	M	AM
124	B	L	BL	165	B	L	BL
125	A	P	AP	166	B	L	BL
127	B	L	BL	167	B	L	BL

Clone No.	KpnI type	MboI Type	Combined Types	Clone No.	KpnI type	MboI Type	Combined Types
168	B	L	BL	213	A	P	AP
169	B	L	BL	214	A	P	AP
170	A	P	AP	215	A	M	AM
171	B	L	BL	216	A	N	AN
172	A	P	AP	217	B	L	BL
174	B	L	BL	218	B	L	BL
175	B	L	BL	219	B	L	BL
176	B	L	BL	220	A	N	AN
177	A	N	AN	221	B	L	BL
178	B	L	BL	222	A	P	AP
179	A	P	AP	224	A	N	AN
181	A	P	AP	225	B	L	BL
182	B	L	BL	226	A	N	AN
183	A	P	AP	227	B	L	BL
184	B	L	BL	228	B	L	BL
186	B	L	BL	229	B	L	BL
187	B	L	BL	230	A	P	AP
189	B	L	BL	231	A	P	AP
190	B	L	BL	232	A	U	AU
192	B	L	BL	233	B	L	BL
193	A	Q	AQ	234	B	L	BL
194	B	L	BL	235	B	L	BL
195	B	L	BL	236	A	N	AN
196	B	L	BL	237	B	L	BL
197	A	Q	AQ	238	A	N	AN
198	B	L	BL	239	B	L	BL
199	B	L	BL	240	A	P	AP
200	B	L	BL	241	B	L	BL
201	B	L	BL	242	A	Q	AQ
202	B	L	BL	243	B	L	BL
203	B	L	BL	244	B	L	BL
204	B	L	BL	246	A	P	AP
205	A	P	AP	247	B	L	BL
206	B	L	BL	248	B	L	BL
207	A	P	AP	249	B	L	BL
208	B	L	BL	250	A	N	AN
209	B	L	BL	251	A	V	AV
211	B	L	BL	252	B	L	BL
212	B	L	BL	254	A	L	AP

Clone No.	KpnI type	MboI Type	Combined Types	Clone No.	KpnI type	MboI Type	Combined Types
255	B	L	BL	304	B	L	BL
257	B	L	BL	305	B	L	BL
260	A	X	AX	306	B	L	BL
261	B	L	BL	307	A	P	AP
263	B	L	BL	308	B	L	BL
264	B	L	BL	311	A	P	AP
265	B	L	BL	312	A	P	AP
266	B	L	BL	314	A	M	AM
267	B	L	BL	315	B	L	BL
268	B	L	BL	316	B	L	BL
270	A	N	AN	317	B	L	BL
271	A	Q	AQ	318	B	L	BL
273	A	N	AN	319	A	N	AN
274	B	L	BL	320	B	L	BL
275	B	L	BL	321	A	P	AP
276	A	P	AP	322	B	L	BL
277	B	L	BL	323	B	L	BL
279	B	L	BL	324	B	L	BL
281	B	L	BL	325	B	L	BL
282	B	L	BL	326	A	P	AP
284	B	L	BL	327	A	S	AS
285	B	L	BL	328	B	L	BL
286	A	P	AP	329	B	L	BL
287	B	L	BL	331	B	L	BL
288	A	P	AP	332	B	L	BL
289	B	L	BL	333	B	L	BL
290	B	L	BL	334	B	L	BL
291	B	L	BL	336	B	L	BL
293	B	L	BL	337	B	L	BL
294	B	L	BL	338	A	P	AP
295	A	W	AW	339	B	L	BL
296	B	L	BL	340	B	L	BL
297	B	L	BL	341	B	L	BL
298	B	L	BL	342	B	L	BL
299	B	L	BL	343	B	L	BL
300	B	L	BL	344	B	L	BL
301	B	L	BL	345	D	U	DU
302	B	L	BL	347	B	L	BL
303	A	P	AP	348	A	N	AN

Clone No.	KpnI type	MboI Type	Combined Types
349	B	L	BL
350	B	L	BL
351	A	V	AV
352	B	L	BL
354	B	L	BL
355	B	L	BL
357	B	L	BL
358	A	M	AM
359	A	M	AM
360	B	L	BL
362	A	P	AP
363	A	S	AS
364	B	L	BL
365	B	L	BL
366	B	L	BL
367	A	P	AP
368	B	L	BL
369	B	L	BL
370	A	R	AR
371	B	L	BL
372	A	S	AS
373	B	L	BL
374	A	P	AP
375	B	L	BL
376	B	L	BL
377	B	L	BL
379	B	L	BL
380	B	L	BL
381	B	L	BL
382	B	L	BL
383	B	L	BL
384	B	L	BL
385	B	L	BL
386	B	L	BL

Appendix 4.3.

(a) Alignment of nucleotide sequences of 17 PERV-A clones with PERV-A published sequences. The start codon is indicated with box. Sequences undersigned with + are the primer used for PCR and sequences undersigned with * are consensus sequences.

```
                    10        20        30        40        50        60
                     .         .         .         .         .         .
PERV-A          CATGCATCCCACGTTAAGCCGGCGCCACCTCCCGATTCGGGGTGGAAAGCCGAAAAGACT
Pfu-11          CATGCATCCCACGTTAAGCCGGTGCCACCTCCCGATTCGGGGTGGAGAGCTGAAAAGACT
Pfu-112         CATGCATCCCACGTTAAGCCGGTGCCACCTCCCGATTCGGGGTGGAGAGCTGAAAAGACT
Pfu-115         CATGCATCCCACGTTAAGCTGGCGCCACCTCCCGATTCGGGGTGGAGAGCCGAAAAGACT
Pfu-251         CATGCATCCCACGTTAAGCCGGTGCCACCTCCCGATTCGGGGTGGAGAGCTGAAAAGACT
Pfu-295         CATGCATCCCACGTTAAGCCGGTGCCACCTCCCGATTCGGGGTGGAGAGCTGAAAAGACT
Pfu-3           CATGCATCCCACGTTAAGCCGGTGCCACCTCCCGATTCGGGGTGGAGAGCTGAAAAGACT
Pfu-56          CATGCATCCCACGTTAAGCCGGTGCCACCTCCCGATTCGGGGTGGAGAGCCGAAAAAACT
Pfu-6           CATGCATCCCACGTTAAGCCGGCGCCACCTCCCGATTCAGGGTGGAGAGCTGAAAAGACT
Pfu-62          CATGCATCCCACGTTAAGCCGGCGCCACCTCCCGATTCGGGGTGGAGAGCCGAAAAAACT
Taq-10          CATGCATCCCACGTTAAGCCGGCGCCACCTCCCGATTCAGGGTGGAGAGCTGAAAAGACT
Taq-11          CATGCATCCCACGTTAAGCCGGCGCCACCTCCCGATTCAGGGTGGAGAGCTGAAAAGACT
Taq-17          CATGCATCCCACGTTAAGCCGGCGCCACCTCCCGATTCGGGGTGGAGAGCCGAAAAGACT
Taq-54          CATGCATCCCACGTTAAGCCGGCGCCACCTCCCGATTCGGGGTGGAGAGCCGAAAAAACT
Taq-56          CATGCATCCCACGTTAAGCCGGCGCCACCTCCCGATTCGGGGTGGAGAGCCGAAAAAACT
Taq-6           CATGCATCCCACGTTAAGCCGGCGCCACCTCCCGATTCAGGGTGGAGAGCTGAAAAGACT
Taq-8           CATGCATCCCACGTTAAGCCGGTGCCACCTCCCGATTCGGGGTGGAGAGCTGAAAAGACT
Taq-82          CATGCATCCCACGTTAAGCCGGTGCCACCTCCCGATTCGGGGTGGAGAGCTGAAAAGACT
                ++++++++++++++++++ ** *************** ******* *** ***** ***
                                   Codon start
                    70        80        90       100       110       120
                     .         .         .         .         .         .
PERV-A          GAAAATCCCCTTAAGCTTCGCCTCCATCGCGTGGTTCCTTACTCTGTCAATAACTCCTCA
Pfu-11          GAAAATCCCCTTAAGCTTCGCCTCCATCGCTTGGTTCCTTACTCTGTCAATAACTCCTCA
Pfu-112         GAAAATCCCCTTAAGCTTCGCCTCCATCGCTTGGTTCCTTACTCTGTCAATAACTCCTCA
Pfu-115         GAAAATCCCCTTAAGCTTCGCCTCCATCGCTTGGTTCCTTACTCTGTCAATAACTCCTCA
Pfu-251         GAAAATCCCCTTAAGCTTCGCCTCCATCGCTTGGTTCCTTACTCTGTCAATAACTCCTCA
Pfu-295         GAAAATCCCCTTAAGCTTCGCCTCCATCGCTTGGTTCCTTACTCTGTCAATAACTCCTCA
Pfu-3           GAAAATCCCCTTAAGCTTCGCCTCCATCGCTTGGTTCCTTACTCTGTCAATAACTCCTCA
Pfu-56          GAAAATCCCCTTAAGCTTCGCCTCCATCGCTTGGTTCCTTACTCTGTCAATAACTCCTCA
Pfu-6           GAAAATCCCCTTAAGCTTCGCCTCCATCGCTTGGTTCCTTACTCTGTCAATAACTCCTCA
Pfu-62          GAAAATCCCCTTAAGCTTCGCCTCCATCGCTTGGTTCCTTACTCTGTCAATAACTCCTCA
Taq-10          GAAAATCCCCTTAAGCTTCGCCTCCATCGCTTGGTTCCTTACTCTGTCAATAACTCCTCA
Taq-11          GAAAATCCCCTTAAGCTTCGCCTCCATCGCTTGGTTCCTTACTCTGTCAATAACTCCTCA
Taq-17          GAAAATCCCCTTAAGCTTCGCCTCCATCGCTTGGTTCCTTACTCTGTCAATAACTCCTCA
Taq-54          GAAAATCCCCTTAAGCTTCGCCTCCATCGCTTGGTTCCTTACTCTGTCAATAACTCCTCA
Taq-56          GAAAATCCCCTTAAGCTTCGCCTCCATCGCTTGGTTCCTTACTCTGTCAATAACTCCTCA
Taq-6           GAAAATCCCCTTAAGCTACGCCTCCATCGCTTGGTTCCTTACTCTGTCAATAACTCCTCA
Taq-8           GAAAATCCCCTTAAGCTTCGCCTCCATCGCTTGGTTCCTTACTCTGTCAATAACTCCTCA
Taq-82          GAAAATCCCCTTAAGCTTCGCCTCCGTCGCTTGGTTCCTTACTCTGTCAATAACTCCTCA
                ***************** ******* **** ****************************

                   130       140       150       160       170       180
                     .         .         .         .         .         .
PERV-A          AGTTAATGGTAAACGCCTTGTGGACAGCCCGAACTCCCATAAACCCTTATCTCTCACCTG
Pfu-11          AGTTAATGGTAAACGCCTTGTGGACAGCCCGAACTCCCATAAACCCTTATCTCTCACCTG
Pfu-112         AGTTAATGGTAAACGCCTTGTGGACAGCCCGAACTCCCATAAACCCTTATCTCTCACCTG
Pfu-115         AGTTAATGGTAAACGCCTTGCGGACAGCCCGAACTTCCATAAACCCTTATCTCTCACCTG
Pfu-251         AGTTAATGGTAAACGCCTTGTGGACAGCCCGAACTCCCATAAACCCTTATCTCTCACCTG
Pfu-295         AGTTAATGGTAAACGCCTTGTGGACAGCCCGAACTCCCATAAACCCTTATCTCTCACCTG
Pfu-3           AGTTAATGGTAAACGCCTTGCGGACAGCCCGAACTTCCATAAACCCTTATCTCTCACCTG
Pfu-56          AGTTAATGGTAAACGCCTTGCGGACAGCCCGAACTTCCATAAACCCTTATCTCTCACCTG
Pfu-6           AGTTAATGGTAAACGCCTTGTGGACAGCCCGAACTCCCATAAACCCTTATCTCTCACCTG
```

```
Pfu-62          AGTTAATGGTAAACGCCTTGCGGACAGCCCGAACTTCCATAAACCCTTATCTCTCACCTG
Taq-10          AGTTAATGGTAAACGCCTTGTGGACAGCCCGAACTCCCATAAACCCTTATCTCTCACCTG
Taq-11          AGTTAATGGTAAACGCCTTGTGGACAGCCCGAACTCCCATAAACCCTTATCTCTCACCTG
Taq-17          AGTTAATGGTAAACGCCTTGCGGACAGCCCGAACTTCCATAAACCCTTATCTCTCACCTG
Taq-54          AGTTAATGGTAAACGCCTTGCGGACAGCCCGAACTTCCATAAACCCTTATCTCTCACCTG
Taq-56          AGTTAATGGTAAACGCCTTGCGGACAGCCCGAACTTCCATAAACCCTTATCTCTCACCTG
Taq-6           AGTTAATGGTAAACGCCTTGTGGACAGCCCGAACTCCCATAAACCCTTATCTCTCACCTG
Taq-8           AGTTAATGGTAAACGCCTTGTGGACAGCCCGAACTCCCATAAACCCTTATCTCTCACCTG
Taq-82          AGTTAATGGTAAACGCCTTGTGGACAGCCCGAACTCCCATAAACCCTTATCTCTCACCTG
                ******************* ************** ************************

                       190       200       210       220       230       240
                         .         .         .         .         .         .
PERV-A          GTTACTTACTGACTCCGGTACAGGTATTAATATTAACAGCACTCAAGGGGAGGCTCCCTT
Pfu-11          GTTACTTACTGACTCCAGTACAGGTATTACTATTAACAGCACTCAAGGGGAGGCTCCCTT
Pfu-112         GTTACTTACTGACTCCAGTACAGGTATTACTATTAACAGCACTCAAGGGGAGGCTCCCTT
Pfu-115         GTTACTTACTGACTCCAGTACAGGTATTACTATTAACAGCACTCAAGGGGAGGCTCCCTT
Pfu-251         GTTACTTACTGACTCCAGTACAGGTATTACTATTAACAGCACTCAAGGGGAGGCTCCCTT
Pfu-295         GTTACTTACTGACTCCAGTACAGGTATTACTATTAACAGCACTCAAGGGGAGGCTCCCTT
Pfu-3           GTTACTTACTGACTCCAGTACAGGTATTACTATTAACAGCACTCAAGGGGAGGCTCCCTT
Pfu-56          GTTACTTACTGACTCCAGTACAGGTATTACTATTAACAGCACTCAAGGGGAGGCTCCCTT
Pfu-6           GTTACTTACTGACTCCAGTACAGGTATTACTATTAACAGCACTCAAGGGGAGGCTCCCTT
Pfu-62          GTTACTTACTGACTCCAGTACAGGTATTACTATTAACAGCACTCAAGGGGAGGCTCCCTT
Taq-10          GTTACTTACTGACTCCAGTACAGGTATTACTATTATCAGCACTCAAGGGGAGGCTCCCTT
Taq-11          GTTACTTACTGACTCCAGTACAGGTATTACTATTAACAGCACTCAAGGGGAGGCTCCCTT
Taq-17          GTTACTTACTGACTCCAGTACAGGTATTACTATTAACAGCACTCAAGGGGAGGCTCCCTT
Taq-54          GTTACTTACTGACTCCAGTACAGGTATTACTATTAACAGCACTCAAGGGGAGGCTCCCTT
Taq-56          GTTACTTACTGACTCCAGTACAGGTATTACTATTAACAGCACTCAAGGGGAGGCTCCCTT
Taq-6           GTTACTTACTGACTCCAGTACAGGTATTACTATTAACAGCACTCAAGGGGAGGCTCCCTT
Taq-8           GTTACTTACTGACTCCAGTACAGGTATTACTATTAACAGCACTCAAGGGGAGGCTCCCTT
Taq-82          GTTACTTACTGACTCCAGTACAGGTATTACTATTAACAGCACTCAAGGGGAGGCTCCCTT
                *************** ************ ***** ************************

                       250       260       270       280       290       300
                         .         .         .         .         .         .
PERV-A          GGGGACCTGGTGGCCTGAATTATATGTCTGCCTTCGATCAGTAATCCCTGGTCTCAATGA
Pfu-11          AGGGACCTGGTGGCCTGAGTTATATGTCTGCCTTCGATTGGTAATCCCTGGTCTCAACGA
Pfu-112         AGGGACCTGGTGGCCTGAGTTATATGTCTGCCTTCGATTGGTAATCCCTGGTCTCAACGA
Pfu-115         AGGGACCTGGTGGCCTGAGTTATATGTCTGCCTTCGATCGGTAATCCCTGGTCTCAACGA
Pfu-251         AGGGACCTGGTGGCCTGAGTTATATGTCTGCCTTCGATTGGTAATCCCTGGTCTCAACGA
Pfu-295         AGGGACCTGGTGGCCTGAGTTATATGTCTGCCTTCGATTGGTAATCCCTGGTCTCAACGA
Pfu-3           AGGGACCTGGTGGCCTGAGTTATATGTCTGCCTTCGATCGGTAATCCCTGGTCTCAACGA
Pfu-56          AGGGACCTGGTGGCCTGAGTTATATGTCTGCCTTCGATCGGTAATCCCTGGTCTCAACGA
Pfu-6           AGGGACCTGGTGGCCTGAGTTATATGTCTGCCTTCGATCGGTAATCCCTGGTCTCAACAA
Pfu-62          AGGGACCTGGTGGCCTGAGTTATATGTCTGCCTTCGATCGGTAATCCCTGGTCTCAACAA
Taq-10          AGGGACCTGGTGGCCTGAGTTATATGTCTGCCTTCGATCGGTAATCCCTGGTCTCAACAA
Taq-11          AGGGACCTGGTGGCCTGAGTTATATGTCTGCCTTCGATCGGTAATCCCTGGTCTCAACAA
Taq-17          AGGGACCTGGTGGCCTGAGTTATATGTCTGCCTTCGATCGGTAATCCCTGGTCTCAACGA
Taq-54          AGGGACCTGGTGGCCTGAGTTATATGTCTGCCTTCGATCGGTGATCCCTGGTCTCAACGA
Taq-56          AGGGACCTGGTGGCCTGAGTTATATGTCTGCCTTCGATCGGTAATCCCTGGTCTCAACGA
Taq-6           AGGGACCTGGTGGCCTGAGTTATATGTCTGCCTTCGATCGGTAATCCCTGGTCTCAACAA
Taq-8           AGGGACCTGGTGGCCTGAGTTATATGTCTGCCTTCGATTGGTAATCCCTGGTCTCAACGA
Taq-82          AGGGACCTGGTGGCCTGAGTTATATGTCTGCCTTCGATTGGTAATCCCTGGTCTCAACGA
                 ***************** ******************  ** **************  *

                       310       320       330       340       350       360
                         .         .         .         .         .         .
PERV-A          CCAGGCCACACCCCCCGATGTACTCCGTGCTTACGGGTTTTACGTTTGCCCAGGACCCCC
Pfu-11          CCAGGCCACACCCCCTGATGTACTCTGTGCTTACAGGTTTAATGTTTGCCCAGGACCCCC
Pfu-112         CCAGGCCACACCCCCTGATGTACTCTGTGCTTACAGGTTTAATGTTTGCCCAGGACCCCC
Pfu-115         CCAGGCCACACCCCCTGATGTACTCTGTGCTTACAGGTTTAATGTTTGCCCAGGACCCCC
Pfu-251         CCAGGCCACACCCCCTGATGTACTCTGTGCTTACAGGTTTAATGTTTGCCCAGGACCCCC
Pfu-295         CCAGGCCACACCCCCTGATGTACTCTGTGCTTACAGGTTTAATGTTTGCCCAGGACCCCC
Pfu-3           CCAGGCCACACCCCCTGATGTACTCCGTGCTTACGGGTTTAATGTTTGCCCAGGACCCCC
Pfu-56          CCAGGCCACACCCCCTGATGTACTCCGTGCTTACGGGTTTAATGTTTGCCCAGGACCCCC
Pfu-6           CCAGGCCACACCCCCCGATGTACTCCGTGCTTACAGGTTTAATGTTTGCCCAGGACCCCC
```

```
Pfu-62          CCAGGCCACACCCCCTGATGTACTCCGTGCTTACAGGTTTAATGTTTGCCCAGGACCCCC
Taq-10          CCAGGCCACACCCCCCGATGTACTCCGTGCTTACAGGTTTAATGTTTGCCCAGGACCCCC
Taq-11          CCAGGCCACACCCCCCGATGTACTCCGTGCTTACAGGTTTAATGTTTGCCCAGGACCCCC
Taq-17          CCAGGCCACACCCCCTGATGTACTCTGTGCTTACAGGTTTAATGTTTGCCCAGGACCCCC
Taq-54          CCAGGCCACACCCCCTGATGTACTCCGTGCTTACGGGTTTAATGTTTGCCCAGGACCCCC
Taq-56          CCAGGCCACACCCCCTGATGTACTCCGTGCTTACGGGTTTAATGTTTGCCCAGGACCCCC
Taq-6           CCAGGCCACACCCCCCGATGTACTCCGTGCTTACAGGTTTAATGTTTGCCCAGGACCCCC
Taq-8           CCAGGCCACACCCCCTGATGTACTCTGTGCTTACAGGTTTAATGTTTGCCCAGGACCCCC
Taq-82          CCAGGCCACACCCCCTGATGTACTCTGTGCTTACAGGTTTAATGTTTGCCCAGGACCCCC
                ********** ********* ******** ***** * *****************

                       370       380       390       400       410       420
                         .         .         .         .         .         .
PERV-A          AAATAATGAAGAATATTGTGGAAATCCTCAGGATTTCTTTTGCAAGCAATGGAGCTGCAT
Pfu-11          AAATAATGAAGAATATTGTGGAAATCCTCAGGATTTCTTTTGCAAGCAAGGGAGCTGCAT
Pfu-112         AAATAATGAAGAATATTGTGGAAATCCTCAGGATTTCTTTTGCAAGCAAGGGAGCTGCAT
Pfu-115         AAATAATGAAGAATATTGTGGAAATCCTCAGGATTTCTTTTGCAAGCAAGGGAGCTGCGT
Pfu-251         AAATAATGAAGAATATTGTGGAAATCCTCAGGATTTCTTTTGCAAGCAAGGGAGCTGCAT
Pfu-295         AAATAATGAAGAATATTGTGGAAATCCTCAGGATTTCTTTTGCAAGCAAGGGAGCTGCAT
Pfu-3           AAATAATGAAGAATATTGTGGAAATCCTCAGGATTTCTTTTGCAAGCAAGGGAGCTGCGT
Pfu-56          AAATAATGAAGAATATTGTGGAAATCCTCAGGATTTCTTTTGCAAGCAAGGGAGCTGCGT
Pfu-6           AAATAATGAAGAATATTGTGGAAATCCTCAGGATTTCTTTTGCAAGCAAGGGAGCTGCGT
Pfu-62          AAATAATGAAGAATATTGTGGAAATCCTCAGGATTTCTTTTGCAAGCAAGGGAGCTGCGT
Taq-10          AAATAATGAAGAATATTGTGGAAATCCTCAGGATTTCTTTTGCAAGCAAGGGAGCTGCGT
Taq-11          AAATAATGAAGAATATTGTGGAAATCCTCAGGATTTCTTTTGCAAGCAAGGGAGCTGCGT
Taq-17          AAATAATGAAGAATATTGTGGAAATCCTCAGGATTTCTTTTGCAAGCAAGGGAGCTGCGT
Taq-54          AAATAATGGAGAATATTGTGGAAATCCTCAGGATTTCTTTTGCAAGCAAGGGAGCTGCGT
Taq-56          AAATAATGAAGAATATTGTGGAAATCCTCAGGATTTCTTTTGCAAGCAAGGGAGCTGCGT
Taq-6           AAATAATGAAGAATATTGTGGAAATCCTCAGGATTTCTTTTGCAAGCAAGGGAGCTGCGT
Taq-8           AAATAATGAAGAATATTGTGGAAATCCTCAGGATTTCTTTTGCAAGCAAGGGAGCTGCAT
Taq-82          AAATAATGAAGAATATTGTGGAAATCCTCAGGATTTCTTTTGCAAGCAAGGGAGCTGCAT
                ******** ****************************************** ******** *

                       430       440       450       460       470       480
                         .         .         .         .         .         .
PERV-A          AACTTCTAATGATGGGAATTGGAAATGGCCAGTCTCTCAGCAAGACAGAGTAAGTTACTC
Pfu-11          AACTTCTAATGATGGGAATTGGAAATGGCCAATCTCTCAGCAAGACAGAGTAAGTTACTC
Pfu-112         AACTTCTAATGATGGGAATTGGAAATGGCCAGTCTCTCAGCAAGACAGAGTAAGTTACTC
Pfu-115         AACTTCTAATGATGGGAATTGGAAATGGCCAGTCTCTCAGCAAGACAGAGTAAGTTACTC
Pfu-251         AACTTCTAATGATGGGAATTGGAAATGGCCAATCTCTCAGCAAGACAGAGTAAGTTACTC
Pfu-295         AACTTCTAATGATGGGAATTGGAAATGGCCAATCTCTCAGCAAGACAGAGTAAGTTACTC
Pfu-3           AACTTCTAATGATGGGAATTGGAAATGGCCAGTCTCTCAGCAAGACAGAGTAAGTTACTC
Pfu-56          AACTTCTAATGATGGGAATTGGAAATGGCCAGTCTCTCAGCAAGACAGAGTAAGTTACTC
Pfu-6           AACTTCTAATGATGGGAATTGGAAATGGCCAATCTCTCAGCAAGACAGAGTAAGTTACTC
Pfu-62          AACTTCTAATGATGGGAATTGGAAATGGCCAATCTCTCAGCAAGACAGAGTAAGTTACTC
Taq-10          AACTTCTAATGATGGGAATTGGAAATGGCCAATCTCTCAGCAAGACAGAGTAAGTTACTC
Taq-11          AACTTCTAATGATGGGAATTGGAAATGGCCAATCTCTCAGCAAGACAGAGTAAGTTACTC
Taq-17          AACTTCTAATGATGGGAATTGGAAATGGCCAGTCTCTCAGCAAGACAGAGTAAGTTACTC
Taq-54          AACTTCTAATGATGGGAATTGGAAATGGCCAGTCTCTCAGCAAGACAGAGTAAGTTACTC
Taq-56          AACTTCTAATGATGGGAATTGGAAATGGCCAGTCTCTCAGCAAGACAGAGTAAGTTACTC
Taq-6           AACTTCTAATGATGGGAATTGGAAATGGCCAATCTCTCAGCAAGACAGAGTAAGTTACTC
Taq-8           AACTTCTAATGATGGGAATTGGAAATGGCCAATCTCTCAGCAAGACAGAGTAAGTTACTC
Taq-82          AACTTCCAATGATGGGAATTGGAAATGGCCAATCTCTCAGCAAGACAGAGTAAGTTACTC
                ****** ********************** ****************************

                       490       500       510       520       530       540
                         .         .         .         .         .         .
PERV-A          TTTTGTTAACAATCCTACCAGTTATAATCAATTTAATTATGGCCATGGGAGATGGAAAGA
Pfu-11          TTTTGTTAACAATCCTACCAGTTATAATCAATTTAATTATGGCCATGGGAGATGGAAAAA
Pfu-112         TTTTGTTAACAGTCCTACCAGTTATAATCAATTTAATTATGGCCATGGGAGATGGAAAGA
Pfu-115         TTTTGTTAACAGTCCTACCAGTTATAATCAATTTAATTATGGCCATGGGAGATGGAAAGA
Pfu-251         TTTTGTTAACAATCCTACCAGTTATAATCAATTTAATTATGGCCATGGGAGATGGAAAAA
Pfu-295         TTTTGTTAACAATCCTACCAGTTATAATCAATTTAATTATGGCCATGGGAGATGGAAAGA
Pfu-3           TTTTGTTAACAGTCCTACCAGTTATAATCAATTTAATTATGGCCATGGGAGATGGAAAGA
Pfu-56          TTTTGTTAACAGTCCTACCAGTTATAATCAATTTAATTATGGCCATGGGAGATGGAAAGA
Pfu-6           TTTTGTTAACAATCCTACCAGTTATAATCAATTTAATTATGGCCATGGGAGATGGAAAGA
```

```
Pfu-62          TTTTGTTAACAATCCTACCAGTTATAATCAATTTAATTATGGCCATGGGAGATGGAAAGA
Taq-10          TTTTGTTAACAATCCTACCAGTTATAATCAATTTAATTATGGCCATGGGAGATGGAAAGA
Taq-11          TTTTGTTAACAATCCTACCAGTTACAATCAATTTAATTATGGCCATGGGAGATGGAAAGA
Taq-17          TTTTGTTAACAGTCCTACCAGTTATAATCAATTTAATTATGGCCATGGGAGATGGAAAGA
Taq-54          TTTTGTTAACAGTCCTACCAGTTATAATCAATTTAATTATGGCCATGGGAGATGGAAAGA
Taq-56          TTTTGTTAACAGTCCTACCAGTTATAATCAATTTAATTATGGCCATGGGAGATGGAAAGA
Taq-6           TTTTGTTAACAATCCTACCAGTTATAATCAATTTAATTATGGCCATGGGAGATGGAAAGA
Taq-8           TTTTGTTAACAATCCTACCAGTTATAATCAATTTAATTATGGCCATGGGAGATGGAAAAA
Taq-82          TTTTGTTAACAATCCTACCAGTTATAATCAATTTAATTATGGCCATGGGAGATGGAAAAA
                *********** ************ ********************************** *

                       550       560       570       580       590       600
                         .         .         .         .         .         .
PERV-A          TTGGCAACAGCGGGTACAAAAAGATGTACGAAATAAGCAAATAAGCTGTCATTCGTTAGA
Pfu-11          TTGGCAACAGCAGGTACAAAAAGATGTACGAAATAAGCAAATAAGCTGTAATTTGTTAGA
Pfu-112         TTGGCAACAGCGGGTACAAAAAGATGTACGAAATAAGCAAATAAGCTGTAATTCGTTAGA
Pfu-115         TTGGCAACAGCGGGTACAAAAAGATGTACGAAATAAGCAAATAAGCTGTAATTCGTTAGA
Pfu-251         TTGGCAACAGCAGGTACAAAAAGATGTACGAAATAAGCAAATAAGCTGTAATTTGTTAGA
Pfu-295         TTGGCAACAGCGGGTACAAAAATATGTACGAAATAAGCAACGAAGCTGTAATTCGTTAGA
Pfu-3           TTGGCAACAGCGGGTACAAAAAGATGTACGAAATAAGCAAATAAGCTGTAATTCGTTAGA
Pfu-56          TTGGCAACAGCGGGTACAAAAAGATGTACGATATAAGCAAATAAGCTGTAATTCGTTAGA
Pfu-6           TTGGCAACAGCGGGTACAAAAATATGTACGAAATAAGCAACGAAGCTGTAATTCGTTAGA
Pfu-62          TTGGCAACAGCGGGTACAAAAATATGTACGAAATAAGCAACGAAGCTGTAATTCGTTAGA
Taq-10          TTGGCAACAGCGGGTACAAAAATATGTACGAAATAAGCAACGAAGCTGTAATTCGTTAGA
Taq-11          TTGGCAACAGCGGGTACAAAAATATGTACGAAATAAGCAACGAAGCTGTAATTCGTTAGA
Taq-17          TTGGCAACAGCGGGTACAAAAAGATGTACGAAATAAGCAAATAAGCTGTAATTCGTTAGA
Taq-54          TTGGCAACAGCGGGTACAAAAAGATGTACGAAATAAGCAAATAAGCTGTAATTCGTTAGA
Taq-56          TTGGCAACAGCGGGTACAAAAAGATGTACGAAATAAGCAAATAAGCTGTAATTCGTTAGA
Taq-6           TTGGCAACAGCGGGTACAAAAATATGTACGAAATAAGCAACGAAGCTGTAACTCGTTAGA
Taq-8           TTGGCAACAGCAGGTACAAAAAGATGTACGAAATAAGCAAATAAGCTGTAATTTGTTAGA
Taq-82          TTGGCAACAGCAGGTACAAAAAGATGTACGAAATAAGCAAATAAGCTGTAATTTGTTAGA
                *********** ********** ******** ********  ******* * * ******

                       610       620       630       640       650       660
                         .         .         .         .         .         .
PERV-A          CCTAGATTACTTAAAAATAAGTTTCACTGAAAAAGGAAAACAAGAAAATATTCAAAAGTG
Pfu-11          CCTAGATTACTTAAAAATAAGTTTCACTGAAAAAGGAAAACAAGAAAATATTCAAAAGTG
Pfu-112         TCTAGATTACTTAAAAATAAGTTTTACTGAAAAAGGAAAACAAGAAAATATTCAAAAGTG
Pfu-115         TCTAGATTACTTAAAAATAAGTTTTACTGAAAAAGGAAAACAAGAAAATATTCAAAAGTG
Pfu-251         CCTAGATTACTTAAAAATAAGTTTCACTGAAAAAGGAAAACAAGAAAATATTCAAAAGTG
Pfu-295         CCTAGATTACTTAAAAATAAGTTTCACTGAAAAAGGAAAACAAGAAAATATTCAAAAGTG
Pfu-3           TCTAGATTACTTAAAAATAAGTTTTACTGAAAAAGGAAAACAAGAAAATATTCAAAAGTG
Pfu-56          TCTAGATTACTTAAAAATAAGTTTTACTGAAAAAGGAAAACAAGAAAATATTCAAAAGTG
Pfu-6           CCTAGATTACTTAAAAATAAGTTTCACTGAAAAAGGAAAACAAGAAAATATTCAAAAGTG
Pfu-62          CCTAGATTACTTAAAAATAAGTTTCACTGAAAAAGGAAAACAAGAAAATATTCAAAAGTG
Taq-10          CCTAGATTACTTAAAAATAAGTTTCACTGAAAAAGGAAAACAAGAAAATATTCAAAAGTG
Taq-11          CCTAGATTACTTAAAAATAAGTTTCACTGAAAAAGGAAAACAAGAAAATATTCAAAAGTG
Taq-17          TCTAGATTACTTAAAAATAAGTTTTACTGAAAAAGGAAAACAAGAAAATATTCAAAAGTG
Taq-54          TCTAGATTACTTAAAAATAAGTTTTACTGAAAAAGGAAAACAAGAAAATATTCAAAAGTG
Taq-56          TCTAGATTACTTAAAAATAAGTTTTACTGAAAAAGGAAAACAAGAAAATATTCAAAAGTG
Taq-6           CCTAGATTACTTAAAAATAAGTTTCACTGAAAAAGGAAAACAAGAAAATATTCAAAAGTG
Taq-8           CCTAGATTACTTAAAAATAAGTTTCACTGAAAAAGGAAAACAAGAAAATATTCAAAAGTG
Taq-82          CCTAGATTACTTAAAAATAAGTTTCACTGAAAAAGGAAAACAAGAAAATATTCAAAAGTG
                 ********************** ***********************************

                       670       680       690       700       710       720
                         .         .         .         .         .         .
PERV-A          GGTAAATGGTATATCTTGGGGAATAGTGTACTATGGAGGCTCTGGGAGAAAGAAAGGATC
Pfu-11          GGTAAATGGTATGTCTTGGGGAATAATGTACTATGGAGGCTCTGGGAGAAGGAAAGGATT
Pfu-112         GGTAAATGGTATGTCTTGGGGAATAATGTACTATGGAGGCACTGGGAGAAGGAAAGGATC
Pfu-115         GGTAAATGGTATGTCTTGGGGAATAATGTACTATGGAGGCTCTGGGAGAAGGAAAGGATT
Pfu-251         GGTAAATGGTATGTCTTGGGGAATAATGTACTATGGAGGCTCTGGGAGAAGGAAAGGATT
Pfu-295         GGTAAATGGTATGTCTTGGGGAATAATGTACTATGGAGGTTCTGGGAGGAGGAAAGGATC
Pfu-3           GGTAAATGGTATGTCTTGGGGAATAATGTACTATGGAGGCACTGGGAGAAGGAAAGGATC
Pfu-56          GGTAAATGGTATGTCTTGGGGAATAATGTACTATGGAGGCACTGGGAGAAGGAAAGGATC
Pfu-6           GGTAAATGGTATGTCTTGGGGAATAATGTACTATGGAGGTTCTGGGAGGAGGAAAGGATC
```

```
Pfu-62          GGTAAATGGTATGTCTTGGGGAATAATGTACTATGGAGGCACTGGGAGAAGGAAAGGATC
Taq-10          GGTAAATGGTATGTCTTGGGGAATAATGTACTATGGAGGTTCTGGGAGGAGGAAAGGATC
Taq-11          GGTAAATGGTATGTCTCGGGGAATAATGTACTATGGAGGTTCTGGGAGGAGGAAAGGATC
Taq-17          GGTAAATGGTATGTCTTGGGGAATAATGTACTATGGAGGCACTGGGAGAAGGAAAGGATC
Taq-54          GGTAAATGGTATGTCTTGGGGAATAATGTACTATGGAGGCACTGGGAGAAGGAAAGGATC
Taq-56          GGTAAATGGTATGTCTTGGGGAATAATGCACTATGGAGGCACTGGGAGAAGGAAAGGATC
Taq-6           GGTAAATGGTATGTCTTGGGGAATAATGTACTATGGAGGTTCTGGGAGGAGGAAAGGATC
Taq-8           GGTAAATGGTATGTCTTGGGGAATAATGTACTATGGAGGCTCTGGGAGAAGGAAAGGATT
Taq-82          GGTAAATGGTATGTCTTGGGGAATAATGTACTATGGAGGCTCTGGGAGAAGGAAAGGATT
                ************ *** ******** ** **********  ******* * ********

                       730       740       750       760       770       780
                         .         .         .         .         .         .
PERV-A          TGTTCTGACTATTCGCCTCAGAATAGAAACTCAGATGGAACCTCCGGTTGCTATAGGACC
Pfu-11          TGTTCTGACTATTCGCCTCAGAATAGAAACTCAGATGGAACCTCCGGTTGCTATAGGACC
Pfu-112         TGTTCTGACTATTCGCCTCAGAATAGAAACTCAGATGGAACCTCCGGTTGCTATAGGACC
Pfu-115         TGTTCTGACTATTCGCCTCAGAATAGAAACTCAGATGGAACCTCCGGTTGCTATAGGACC
Pfu-251         TGTTCTGACTATTCGCCTCAGAATAGAAACTCAGATGGAACCTCCGGTTGCTATAGGACC
Pfu-295         TGTTCTGACTATTCGCCTCAGAATAGAAACTCAGATGGAACCTCCGGTTGCTATAGGACC
Pfu-3           TGTTCTGACTATTCGCCTCAGAATAGAAACTCAGATGGAACCTCCGGTTGCTATAGGACC
Pfu-56          TGTTCTGACTATTCGCCTCAGAATAGAAACTCAGATGGAACCTCCGGTTGCTATAGGACC
Pfu-6           TGTTCTGACTATTCGCCTCAGAATAGAAACTCAGATGGAACCTCCGGTTGCTATAGGACC
Pfu-62          TGTTCTGACTATTCGCCTCAGAATAGAAACTCAGATGGAACCTCCGGTTGCTATAGGACC
Taq-10          TGTTCTGACTATTCGCCTCAGAATAGAAACTCAGATGGAACCTCCGGTTGCTATAGGACC
Taq-11          TGTTCTGACTATTCGCCTCAGAATAGAAACTCAGATGGAACCTCCGGTTGCTATAGGACC
Taq-17          TGTTCTGACTATTCGCCTCAGAATAGAAACTCAGATGGAACCTCCGGTTGCTATAGGACC
Taq-54          TGTTCTGACTATTCGCCTCAGAATAGAAACTCAGATGGAACCTCCGGTTGCTATAGGACC
Taq-56          TGTTCTGACTATTCGCCTCAGAATAGAAACTCAGATGGAACCTCCGGTTGCTATAGGACC
Taq-6           TGTTCTGACTATTCGCCTCAGAATAGAAACTCAGATGGAACCTCCGGTTGCTATAGGACC
Taq-8           TGTTCTGACTATTCGCCTCAGAATAGAAACTCAGATGGAACCTCCGGTTGCTATAGGACC
Taq-82          TGTTCTGACTATTCGCCTCAGAATAGAAACTCAGATGGAACCTCCGGTTGCTATAGGACC
                ************************************************************

                       790       800       810       820       830       840
                         .         .         .         .         .         .
PERV-A          AAATAAGGGTTTGGCCGAACAAGGACCTCCAATCCAAGAACAGAGGCCATCTCCTAACCC
Pfu-11          AAATAAGGGTTTGGCCGAACAAGGACCTCTAATCCAAGAACAGAGGCCATCTCCTAACCC
Pfu-112         AAATAAGGGTTTGGCTGAACAAGGACCTCCAATCCAAGAACAGAGGCCATCTCCTAACCC
Pfu-115         AAATAAGGGTTTGGCCGAACAAGGACCTCCAATCCAAGAACAGAGGCCATCTCCTAACCC
Pfu-251         AAATAAGGGTTTGGCCGAACAAGGACCTCTAATCCAAGAACAGAGGCCATCTCCTAACCC
Pfu-295         AAATAAGGGTTTGGCTGAACAAGGACCTCCAATCCAAGAACAGAGGCCATCTCCTAACCC
Pfu-3           AAATAAGGGTTTGGCTGAACAAGGACCTCCAATCCAAGAACAGAGGCCATCTCCTAACCC
Pfu-56          AAATAAGGGTTTGGCTGAACAAGGACCTCCAATCCAAGAACAGAGGCCATCTCCTAACCC
Pfu-6           AAATAAGGGTTTGGCCGAACAAGGACCTCCAATCCAAGAACAGAGGCCATCTCCTAACCC
Pfu-62          AAATAAGGGTTTGGCTGAACAAGGACCTCCAATCCAAGAACAGAGGCCATCTCCTAACCC
Taq-10          AAATAAGGGTTTGGCCGAACAAGGACCTCCAATC-AAGAACAGAGGCCATCTCCTAACCC
Taq-11          AAATAAGGGTTTGGCCGAACAAGGACCTCCAATCCAAGAACAGAGGCCATCTCCTAACCC
Taq-17          AAATAAGGGTTTGGCTGAACAAGGACCTCCAATCCAAGAACAGAGGCCATCTCCTAACCC
Taq-54          AAATAAGGGTTTGGCTGAACAAGGACCTCCAATCCAAGAACAGAGGCCATCTCCTAACCC
Taq-56          AAATAAGGGTTTGGCTGAACAAGGACCTCCAATCCAAGAACAGAGGCCATCTCCTAACCC
Taq-6           AAATAAGGGTTTGGCCGAACAAGGACCTCCAATCCAAGAACAGAGGCCATCTCCTAACCC
Taq-8           AAATAAGGGTTTGGCCGAACAAGGACCTCTAATCCAAGAACAGAGGCCATCTCCTAACCC
Taq-82          AAATAAGGGTTTGGCCGAACAAGGACCTCTAATCCAAGAACAGAGGCCATCTCCTAACCC
                *************** ************* **** ************************

                       850       860       870       880       890       900
                         .         .         .         .         .         .
PERV-A          CTCTGATTACAATACAACCTCTGGATCAGTCCCCACTGAGCCTAACATCACTATTAAAAC
Pfu-11          CTCTGATTACAATACAACCTCTGGATCAGTCCCCACTGAGCCTAACATCACTATTAAAAC
Pfu-112         CTCTGTTTACAATACAACCTCTGGATTAGTCCCCCCTGAGCCTAACTTCACTATTAAAAC
Pfu-115         CTCTGATTACAATACAACCTCTGGATCAGTCCCCACTGAGCCTAACATCACTATTAAAAC
Pfu-251         CTCTGTTTACAATACAACCTCTGGATTAGTCCCCCCTGAGCCTAACTTCACTATTAAAAC
Pfu-295         CTCTGTTTACAATACAACCTCTGGATTAGTCCCCCCTGAGCCTAACTTCACTATTAAAAC
Pfu-3           CTCTGTTTACAATACAACCTCTGGATTAGTCCCCCCTGAGCCTAACTTCACTATTAAAAC
Pfu-56          CTCTGATTACAATACAACCTCTGGATCAGTCCCCACTGAGCCTAACATCACTATTAAAAC
Pfu-6           CTCTGATTACAATACAACCTCTGGATCAGTCCCCACTGAGCCTAACATCACTATTAAAAC
```

```
Pfu-62          CTCTGTTTACAATACAACCTCTGGATTAGTCCCCCCTGAGCCTAACTTCACTATTAAAAC
Taq-10          CTCTGATTACAATACAACCTCTGGATCAGTCCCCACTGAGCCTAACATCACTATTAAAAC
Taq-11          CTCTGATTACAATACAACCTCTGGATCAGTCCCCACTGAGCCTAACATCACTATTAAAAC
Taq-17          CTCTGTTTACAATACAACCTCTGGATTAGTCCCCCCTGAGCCTAACTTCACTATTAAAAC
Taq-54          CTCTGTTTACAATACAACCTCTGGATTAGTCCCCCCTGAGCCTAACTTCACTATTAAAAC
Taq-56          CTCTGTTTACAATACAACCTCTGGATTAGTCCCCCCTGAGCCTAACTTCACTATTAAAAC
Taq-6           CTCTGATTACAATACAACCTCTGGATCAGTCCCCACTGAGCCTAACATCACTATTAAAAC
Taq-8           CTCTGATTACAATACAACCTCTGGATCAGTCCCCACTGAGCCTAACATCACTATTAAAAC
Taq-82          CTCTGATTACAATACAACCTCTGGATCAGTCCCCACTGAGCCTAACATCACTATTAAAAC
                ***** ******************** ******* *********** *************

                       910       920       930       940       950       960
                         .         .         .         .         .         .
PERV-A          AGGGGCGAAACTTTTTAGCCTCATCCAGGGAGCTTTTCAAGCTCTTAACTCCACGACTCC
Pfu-11          AGTGGTGAAACTTTTTAGCCTCATCCAGGGAGATTTTCAGGCTCTTAACTCCACGACTCC
Pfu-112         AGGGGCGAAACTTTTTAGCCTCATCCAGGGAGCTTTTCAAGCTCTTAACTCCACGACTCC
Pfu-115         AGCGGCAAAACTTTTTAGCCTCATCCAGGGAGATTTTCAGGCTCTTAACTCCACGACTCC
Pfu-251         AGGGGCGAAACTTTTTAGCCTCATCCAGGGAGCTTTTCAAGCTCTTAACTCCACGACTCC
Pfu-295         AGGGGCGAAACTTTTTAGCCTCATCCAGGGAGCTTTTCAAGCTCTTAACTCCACGACTCC
Pfu-3           AGGGGCGAAACTTTTTAGCCTCATCCAGGGAGCTTTTCAAGCTCTTAACTCCACGACTCC
Pfu-56          AGTGGTGAAACTTTTTAGCCTCATCCAGGGAGATTTTCAGGCTCTTAACTCCACGACTCC
Pfu-6           AGCGGCAAAACTTTTTAGCCTCATCCAGGGAGATTTTCAGGCTCTTAACTCCACGACTCC
Pfu-62          AGGGGCGAAACTTTTTAGCCTCATCCAGGGAGCTTTTCAAGCTCTTAACTCCACGACTCC
Taq-10          AGCGGCAAAACTTTTTAGCCTCATCCAGGGAGATTTTCAGGCTCTTAACTCCACGACTCC
Taq-11          AGCGGCAAAACTTTTTAGCCTCATCCAGGGAGATTTTCAGGCTCTTAACTCCACGACTCC
Taq-17          AGGGGCGAAACTTTTTAGCCTCATCCAGGGAGCTTTTCAAGCTCTTAACTCCACGACTCC
Taq-54          AGGGGCGAAACTTTTTAGCCTCATCCAGGGAGCTTTCCAAGCTCTTAACTCCACGACTCC
Taq-56          AGGGGCGAAACTTTTTAGCCTCATCCAGGGAGCTCTTCAAGCTCTTAACTCCACGACTCC
Taq-6           AGCGGCAAAACTTTTTAGCCTCATCCAGGGAGGTTTTCAGGCTCTTAACTCCACGACTCC
Taq-8           AGTGGTGAAACTTTTTAGCCTCATCCAGGGAGATTTTCAGGCTCTTAACTCCACGACTCC
Taq-82          AGTGGTGAAACTTTTTAGCCTCATCCAGGGAGATTTTCAGGCTCTTAACTCCACGACTCC
                ** **  ************************* * * ** ********************

                       970       980       990      1000      1010      1020
                         .         .         .         .         .         .
PERV-A          AGAGGCTACCTCTTCTTGTTGGCTTTGCTTAGCTTCGGGCCCACCTTACTATGAGGGAAT
Pfu-11          AGAGGCTACCTCTTCTTGTTGGCTTTGCTTAGCTTCGGGCCCACCTTACTATGAGGGAAT
Pfu-112         AGAGGCTACCTCTTCTTGTTGGCTTTGCTTAGCTTTGGGCCCACCTTACTATGAGGGAAT
Pfu-115         AGAGGCTACCTCTTCTTGTTGGCTTTGCTTAGCTTCGGGCCCACCTTACTATGAGGGAAT
Pfu-251         AGAGGCTACCTCTTCTTGTTGGCTTTGCTTAGCTTTGGGCCCACCTTACTATGAGGGAAT
Pfu-295         AGAGGCTACCTCTTCTTGTTGGCTTTGCTTAGCTTTGGGCCCACCTTACTATGAGGGAAT
Pfu-3           AGAGGCTACCTCTTCTTGTTGGCTTTGCTTAGCTTTGGGCCCACCTTACTATGAGGGAAT
Pfu-56          AGAGGCTACCTCTTCTTGTTGGCTTTGCTTAGCTTCGGGCCCACCTTACTATGAGGGAAT
Pfu-6           AGAGGCTACCTCTTCTTGTTGGCTTTGCTTAGCTTCGGGCCCACCTTACTATGAGGGAAT
Pfu-62          AGAGGCTACCTCTTCTTGTTGGCTTTGCTTAGCTTTGGGCCCACCTTACTATGAGGGAAT
Taq-10          AGAGGCTACCTCTTCTTGTTGGCTTTGCTTAGCTTTGGGCCCACCTTACTATGAGGGAGT
Taq-11          AGAGGCTACCTCTTCTTGTTGGCTTTGCTTAGCTTCGGGCCCACCTTACTATGAGGGAAT
Taq-17          AGAGGCTACCTCTTCTTGTTGGCTTTGCTTAGCTTTGAGCCCACCTTACTATGAGGGAAT
Taq-54          AGAGGCTACCTCTTCTTGTTGGCTTTGCTTAGCTTTGGGCCCACCTTACTATGAGGGAAT
Taq-56          AGAGGCTACCTCTTCTTGTTGGCTTTGCTTAGCTTTGGGCCCACCTTACTATGAGGGAAT
Taq-6           AGAGGCTACCTCTTCTTGTTGGCTTTGCTTAGCTTCGGGCCCACCTTACTATGAGGGAAT
Taq-8           AGAGGCTACCTCTTCTTGTTGGCTTTGCTTAGCTTCGGGCCCACCTTACTATGAGGGAAT
Taq-82          AGAGGCTACCTCTTCTTGTTGGCTTTGCTTAGCTTCGGGCCCACCTTACTATGAGGGAAT
                ********************************** * ******************** *

                      1030      1040      1050      1060      1070      1080
                         .         .         .         .         .         .
PERV-A          GGCTAGAGGAGGGAAATTCAATGTGACAAAGGAACATAGAGACCAATGTACATGGGGATC
Pfu-11          GGCTAGAGGAGGGAAATTCAATGTGACAAAGGAACATAGAGACCAATGTACATGGGGATC
Pfu-112         GGCTAGAGGAGGGAAATTCAATGTGACAAAGGAACATAGAGACCAATGTACATGGGGATC
Pfu-115         GGCTAGAGGAGGGAAATTCAATGTGACAAAGGAACATAGAGACCAATGTACATGGGGATC
Pfu-251         GGCTAGAGGAGGGAAATTCAATGTGACAAAGGAACATAGAGACCAATGTACATGGGGATC
Pfu-295         GGCTAGAGGAGGGAAATTCAATGTGACAAAGGAACATAGAGACCAATGTACATGGGGATC
Pfu-3           GGCTAGAGGAGGGAAATTCAATGTGACAAAGGAACATAGAGACCAATGTACATGGGGATC
Pfu-56          GGCTAGAGGAGGGAAATTCAATGTGACAAAGGAACATAGAGACCAATGTACATGGGGATC
Pfu-6           GGCTAGAGGAGGGAAATTCAATGTGACAAAGGAACATAGAGACCAATGTACATGGGGATC
```

```
Pfu-62          GGCTAGAGGAGGGAAATTCAATGTGACAAAGGAACATAGAGACCAATGTACATGGGGATC
Taq-10          GGCTAGAGGAGGGAAATTCAATGTGACAAAGGAACATAGAGACCAATGTACATGGGGATC
Taq-11          GGCTAGAGGAGGGAAATTCAATGTGACAAAGGAACATAGAGACCAATGTACATGGGGATC
Taq-17          GGCTAGAGGAGGGAAATTCAATGTGACAAAGGAACATAGAGACCAATGTACATGGGGATC
Taq-54          GGCTAGAGGAGGGAAATTCAATGTGACAAAGGAACATAGAGACCAATGTACATGGGGATC
Taq-56          GGCTAGAGGAGGGAAATTCAATGTGACAAAGGAACATAGAGACCAATGTACATGGGGATC
Taq-6           GGCTAGAGGAGGGAAATTCAATGTGACAAAGGAACATAGAGACCAATGTACATGGGGATC
Taq-8           GGCTAGAGGAGGGAAATTCAATGTGACAAAGGAACATAGAGACCAATGTACATGGGGATC
Taq-82          GGCTAGAGGAGGGAAATTCAATGTGACAAAGGAACATAGAGACCAATGTACATGGGGATC
                ************************************************************

                      1090      1100      1110      1120      1130      1140
                         .         .         .         .         .         .
PERV-A          CCAAAATAAGCTTACCCTTACTGAGGTTTCTGGAAAAGGCACCTGCATAGGGATGGTTCC
Pfu-11          CCAAAATAAGCTTACCCTTACTGAGGTTTCTGGAAAAGGCACCTGCATAGGGAGGGTTCC
Pfu-112         CCAAAATAAGCTTACCCTTACTGAGGTTTCTGGAAAAGGCACCTGCATAGGGA-GGTTCC
Pfu-115         CCAAAATAAGCTTACCCTTACTGAGGTTTCTGGAAAAGGCACCTGCATAGGGAAGGTTCC
Pfu-251         CCAAAATAAGCTTACCCTTACTGAGGTTTCTGGAAAAGGCACCTGCATAGGGAGGGTTCC
Pfu-295         CCAAAATAAGCTTACCCTTACTGAGGTTTCTGGAAAAGGCACCTGCATAGGGA-GGTTCC
Pfu-3           CCAAAATAAGCTTACCCTTACTGAGGTTTCTGGAAAAGGCACCTGCATAGGGAAGGTTCC
Pfu-56          CCAAAATAAGCTTACACTTACTGAGGTTTCTGGAAAAGGCACCTGCATAGGGA-GGTTCC
Pfu-6           CCAAAATAAGCTTACCCTTACTGAGGTTTCTGGAAAAGGCACCTGCATAGGGAGGGTTCC
Pfu-62          CCAAAATAAGCTTACCCTTACTGAGGTTTCTGGAAAAGGCACCTGCATAGGGAGGGTTCC
Taq-10          CCAAAATAAGCTTACCCTTACTGAGGTTTCTGGAAAAGGCACCTGCATAGGGA-GGTTCC
Taq-11          CCAAAATAAGCTTACCCTTACTGAGGTTTCTGGAAAAGGCACCTGCATAGGGA-GGTTCC
Taq-17          CCAAAATAAGCTTACCCTTACTGAGGTTTCTGGAAAAGGCACCTGCATAGGGAGGGTTCC
Taq-54          CCAAAATAAGCTTACCCTTACTGAGGTTTCTGGAAAAGGCACCTGCATAGGGA-GGTTCC
Taq-56          CCAAAATAAGCTTACTCTTACTGAGGTTTCTGGAAAAGGCACCTGCATAGGGAGGGTTCC
Taq-6           CCAAAATAAGCTTACCCTTACTGAGGTTTCTGGAAAAGGCACCTGCATAGGGAGGGTTCC
Taq-8           CCAAAATAAGCTTACACTTACTGAGGTTTCTGGAAAAGGCACCTGCATAGGGA-GGTTCC
Taq-82          CCAAAATAAGCTTACACTTACTGAGGTTTCTGGAAAAGGCACCTGCATAGGGA-GGTTCC
                *************** ************************************ * ****

                      1150      1160      1170      1180      1190      1200
                         .         .         .         .         .         .
PERV-A          CCCATCCCACCAACACCTTTGTAACCACACTGAAGCCTTTAATCGAACCTCTGAGAGTCA
Pfu-11          CCCATCCCACCAACACCTTTGTAACCACACTGAAGCCTTTAATCGAACCTCTGAGAGTCA
Pfu-112         CCCATCCCACCAACACCTTTGTAACCACACTGAAGCCTTTAATCGAACCTCTGAGAGTCA
Pfu-115         CCCATCCCACCAACACCTTTGTAACCACACTGAAGCCTTTAATCGAACCTCTGAGAGTCA
Pfu-251         CCCATCCCACCAACACCTTTGTAACCACACTGAAGCCTTTAATCGAACCTCTGAGAGTCA
Pfu-295         CCCATCCCACCAACACCTTTGTAACCACACTGAAGCCTTTAATCGAACCTCTGAGAGTCA
Pfu-3           CCCATCCCACCAACACCTTTGTAACCACACTGAAGCCTTTAATCGAACCTCTGAGAGTCA
Pfu-56          CCCATCCCACCAACACCTTTGTAACCACACTGAAGCCTTTAATTGAACCTCTGAGAGTCA
Pfu-6           CCCATCCCACCAACACCTTTGTAACCACACTGAAGCCTTTAATCGAACCTCTGAGAGTCA
Pfu-62          CCCATCCCACCAACACCTTTGTAACCACACTGAAGCCTTTAATCGAACCTCTGAGAGTCA
Taq-10          CCCATCCCACCAACACCTTTGTAACCACACTGAAGCCTTTAATCGAACCTCTGAGAGTCA
Taq-11          CCCATCCCACCAACACCTTTGTAACCACACTGAAGCCTTTAATCGAACCTCTGAGAGTCA
Taq-17          CCCATCCCACCAACACCTTTGTAACCACACTGAAGCCTTTAATTGAACCTCTGAGAGTCA
Taq-54          CCCATCCCACCAACACCTTTGTAACCACACTGAAGCCTTTAATCGAACCTCTGAGAGTCA
Taq-56          CCCATCCCACCAACACCTTTGTAACCACACTGAAGCCTTTAATCGAACCTCTGAGAGTCA
Taq-6           CCCATCCCACCAACACCTTTGTAACCACACTGAAGCCTTTAATCGAACCTCTGAGAGTCA
Taq-8           CCCATCCCACCAACACCTTTGTAACCACACTGAAGCCTTTAATTGAACCTCTGAGAGTCA
Taq-82          CCCATCCCACCAACACCTTTGTAACCACACTGAAGCCTTTAATTGAACCTCTGAGAGTCA
                ****************************************** ****************

                      1210      1220      1230      1240      1250      1260
                         .         .         .         .         .         .
PERV-A          ATATCTGGTACCTGGTTATGACAGGTGGTGGGCATGTAATACTGGATTAACCCCTTGTGT
Pfu-11          GTACCTGGTACCTGGTTATGACAGGTGGTGGGCATGTAATACTGGATTAACCCCTTGTGT
Pfu-112         GTACCTGGTACCTGGTTATGACAGGTGGTGGGCATGTAATACTGGATTAACCCCTTGTGT
Pfu-115         GTACCTGGTACCTGGTTATGACAGGTGGTGGGCATGTAATACTGGATTAACCCCTTGTGT
Pfu-251         GTACCTGGTACCTGGTTATGACAGGTGGTGGGCATGTAATACTGGATTAACCCCTTGTGT
Pfu-295         GTACCTGGTACCTGGTTATGACAGGTGGTGGGCATGTAATACTGGATTAACCCCTTGTGT
Pfu-3           GTACCTGGTACCTGGTTATGACAGGTGGTGGGCATGTAATACTGGATTAACCCCTTGTGT
Pfu-56          GTATCTGGTACCTGGTTATGACAGGTGGTGGGCATGTAATACTGGATTAACCCCTTGTGT
Pfu-6           GTATCTGGTACCTGGTTATGACAGGTGGTGGGCATGTAATACTGGATTAACCCCTTGTGT
```

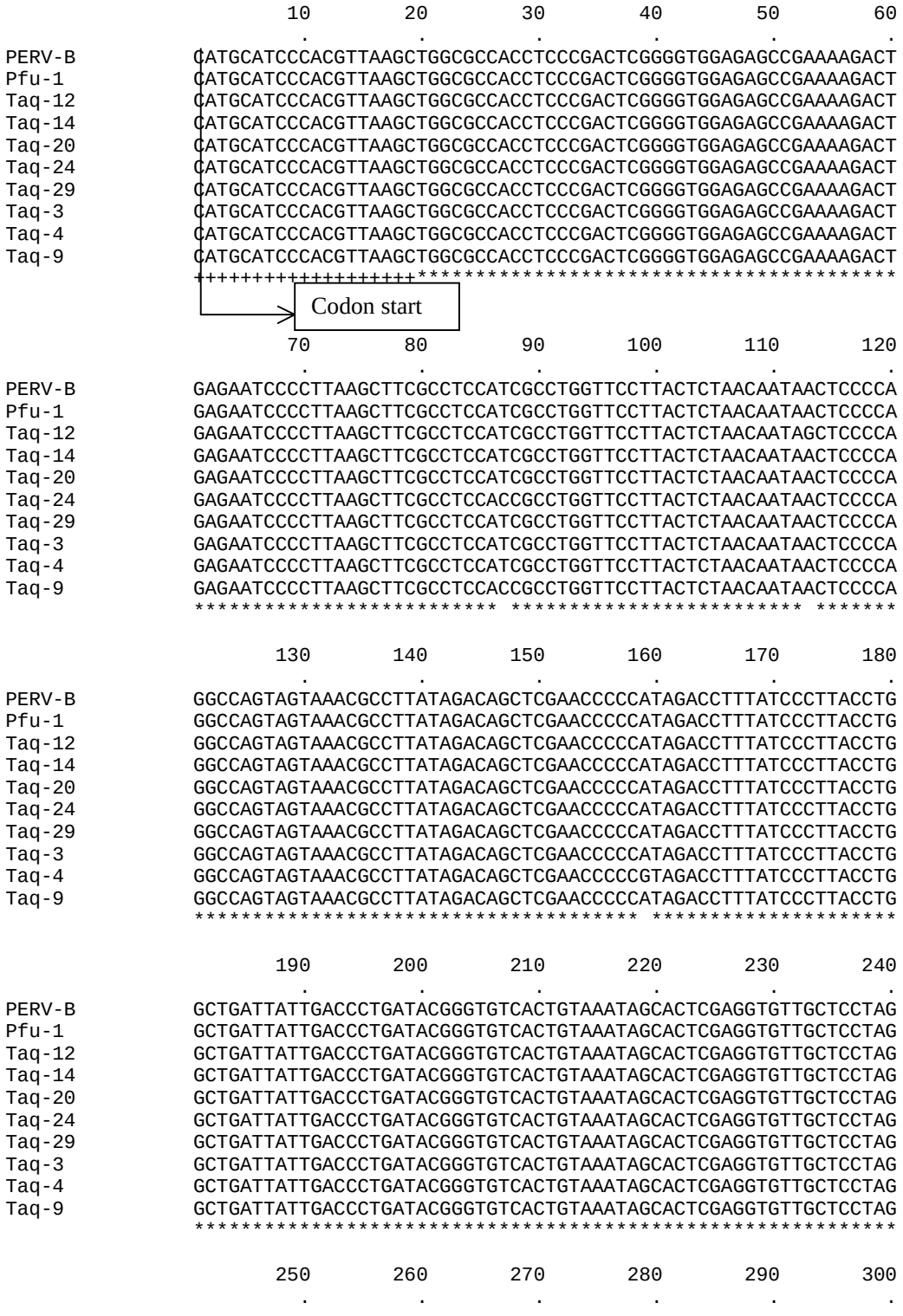
```
Pfu-62          GTACCTGGTACCTGGTTATGACAGGTGGTGGGCATGTAATACTGGATTAACCCCTTGTGT
Taq-10          GTACCTGGTACCTGGTTATGACAGGTGGTGGGCATGTAATACTGGATTAACCCCTTGTGT
Taq-11          GTATCTGGTACCTGGTTATGACAGGTGGTGGGCATGTAATACTGGATTA-CCCCTTGTGT
Taq-17          GTATCTGGTACCTGGTTATGACAGGTGGTGGGCATGTAATACTGGATCAACCCCTTGTGT
Taq-54          GTACCTGGTACCTGGTTATGACAGGTGGTGGGCATGTAATACTGGATTAACCCCTTGTGT
Taq-56          GTACCTGGTACCTGGTTATGACAGGTGGTGGGCATGTAATACTGGATTAACCCCTTGTGT
Taq-6           GTATCTGGTACCTGGTTATGACAGGTGGTGGGCATGTAATACTGGATTA-CCCCTTGTGT
Taq-8           GTATCTGGTACCTGGTTATGACAGGTGGTGGGCATGTAATACTGGATTAACCCCTTGTGT
Taq-82          GTATCTGGTACCTGGTTATGACAGGTGGTGGGCATGTAATACTGGATTAACCCCTTGTGT
                 ** ****************************************** * **********

                      1270      1280      1290      1300      1310      1320
                         .         .         .         .         .         .
PERV-A          TTCCACCTTGGTTTTCAACCAAACTAAAGACTTTTGCGTTATGGTCCAAATTGTCCCCCG
Pfu-11          TTCCACCCTGGTTTTCAACCAAACTAAAGACTTTTGTGTTATGGTCCAAATTGTCCCCCG
Pfu-112         TTCCACCCTGGTTTTCAACCAAACTAAAGACTTTTGTGTTATGGTCCAAATTGTCCCCCG
Pfu-115         TTCCACCCTGGTTTTCAACCAAACTAAAGACTTTTGTGTTATGGTCCAAATTGTCCCCCG
Pfu-251         TTCCACCCTGGTTTTCAACCAAACTAAAGACTTTTGTGTTATGGTCCAAATTGTCCCCCG
Pfu-295         TTCCACCCTGGTTTTCAACCAAACTAAAGACTTTTGTGTTATGGTCCAAATTGTCCCCCG
Pfu-3           TTCCACCCCTGGTTTCAACCAAACTAAAGACTTTTGTGTTATGGTCCAAATTGTCCCCCG
Pfu-56          TTCCACCCTGGTTTTCAACCAAACTAAAGACTTTTGCTTTATGGTCCAAATTGTCCCCCG
Pfu-6           TTCCACCCTGGTTTTCAACCAAACTAAAGACTTTTGTGTTATGGTCCAAATTGTCCCCCG
Pfu-62          TTCCACCCTGGTTTTCAACCAAACTAAAGACTTTTGTGTTATGGTCCAAATTGTCCCCCG
Taq-10          TTCCACCCTGGTTTTCAACCAAACTAAAGACTTTTGTGTTATGGTCCAAATTGTCCCCCG
Taq-11          TTCCACCCTGGTTTTCAACCAAACTAAAGACTTTTGCGTTATGGTCCAAATTGTCCCCCG
Taq-17          TTCCACCCTGGTTTTCAACCAAACTAAAGACTTTTGTGTTACGGTCCAAATTGTCCCCCG
Taq-54          TTCCACCCTGGTTTTCAACCAGACTAAAGACTTTTGTGTTATGGTCCAAATTGTCCCCCG
Taq-56          TTCCACCCTGGTTTTCAACCAAACTAAAGACTTTTGTGTTATGGTCCAAATTGTCCCCCG
Taq-6           TTCCACCCTGGTTTTCAACCAAACTAAAGACTTTTGCGTTATGGTCCAAATTGTCCCCCG
Taq-8           TTCCACCCTGGTTTTCAACCAAACTAAAGACTTTTGCTTTATGGTCCAAATTGTCCCCCG
Taq-82          TTCCACCCTGGTTTTCAACCAAACTAAAGACTTTTGCTTTATGGTCCAAATTGTCCCCCG
                *******   * ********* **************  *** ******************

                      1330      1340      1350      1360      1370      1380
                         .         .         .         .         .         .
PERV-A          GGTGTACTACTATCCCGAAAAAGCAGTCCTTGATGAATATGACTATAGATATAATCGGCC
Pfu-11          GGTGTACTACTATCCCGAAAAAGCAGTCCTTGATGAATATGACTATAGATATAATCGGCC
Pfu-112         GGTGTACTACTATCCCGAAAAAGCAGTCCTTGATGAATATGACTATAGATATAATCGGCC
Pfu-115         GGTGTACTACTATCCCGAAAAAGCAGTCCTTGATGAATATGACTATAGATATAATCGGCC
Pfu-251         GGTGTACTACTATCCCGAAAAAGCAGTCCTTGATGAATATGACTATAGATATAATCGGCC
Pfu-295         GGTGTACTACTATCCCGAAAAAGCAGTCCTTGATGAATATGACTATAGATATAATCGGCC
Pfu-3           GGTGTACTACTATCCCGAAAAAGCAGTCCTTGATGAATATGACTATAGATATAATCGGCC
Pfu-56          GGTGTACTACTATCCCAAAAAAGCGGTCCTTGATGAATATGACTATAGATATAATCGGCC
Pfu-6           GGTGTACTACTATCCCGAAAAAGCGGTCCTTGATGAATATGACTATAGATATAATCGGCC
Pfu-62          GGTGTACTACTATCCCGAAAAAGCAGTCCTTGATGAATATGACTATAGATATAATCGGCC
Taq-10          GGTGTACTACTATCCCGAAAAAGCAGTCCTTGATGAATATGACTATAGATATAATCGGCC
Taq-11          GGTGTACTACTATCCCAAAAAAGCGGTCCTTGATGAATATGGCTATAGATATAATCGGCC
Taq-17          GGTGTACTACTATCCCAAAAAAGCGGTCCTTGATGAATATGACTATAGATATAATCGGCC
Taq-54          GGTGTACTACTATCCCGAAAAAGCAGTCCTTGATGAATATGACTATAGATATAATCGGCC
Taq-56          GGTGTACTACTATCCCGAAAAAGCAGTCCTTGATGAATATGACTATAGATATAATCGGCC
Taq-6           GGTGTACTACTATCCCAAAAAAGCGGTCCTTGATGAATATGGCTGTAGATATAATCGGCC
Taq-8           GGTGTACTACTATCCCAAAAAAGCGGTCCTTGATGAATATGACTATAGATATAATCGGCC
Taq-82          GGTGTACTACTATCCCAAAAAAGCGGTCCTTGATGAATATGACTATAGATATAATCGGCC
                **************** ******* **************** ** ***************

                      1390      1400      1410      1420      1430      1440
                         .         .         .         .         .         .
PERV-A          AAAAAGAGAGCCCATATCCCTGACACTAGCTGTAATGCTCGGATTGGGAGTGGCTGCAGG
Pfu-11          AAAAAGAGAACCCATATCCCTGACACTAGCTGTAATGCTCGGATTGGGAGTGGCTGCAGG
Pfu-112         AAAAAGAGAACCCATATCCCTGACACTAGCTGTAATGCTCGGATTGGGAGTGGCTGCAGG
Pfu-115         AAAAAGAGAACCCATATCCCTGACACTAGCTGTAATGCTCGGATTGGGAGTGGCTGCAGG
Pfu-251         AAAAAGAGAACCCATATCCCTGACACTAGCTGTAATGCTCGGATTGGGAGTGGCTGCAGG
Pfu-295         AAAAAGAGAACCCATATCCCTGACACTAGCTGTAATGCTCGGATTGGGAGTGGCTGCAGG
Pfu-3           AAAAAGAGAACCCATATCCCTGACACTAGCTGTAATGCTCGGATTGGGAGTGGCTGCAGG
Pfu-56          AAAAAGAGAACCCATATCCCTGACACTAGCTGTAATGCTCGGATTGGGAGTGGCTGCAGG
Pfu-6           AAAAAGAGAACCCATATCCCTGACACTAGCTGTAATGCTCGGATTGGGAGTGGCTGCAGG
```

```
Pfu-62          AAAAAGAGAACCCATATCCCTGACACTAGCTGTAATGCTCGGATTGGGAGTGGCTGCAGG
Taq-10          AAAAAGAGAACCCATATCCCTGACACTAGCTGTAATGCTCGGATTGGGAGAGGCTGCAGG
Taq-11          AAAAAGAGAACCCATATCCCTGACACTAGCTGTAATGCTCGGATTGGGAGTGGCTGCAGG
Taq-17          AAAAAGAGAACCCATATCCCTGACACTAGCTGTAATGCTCGGATTGGGAGTGGCTGCAGG
Taq-54          AAAAAGAGAACCCATATCCCTGACACTAGCTGTAATGCTCGGATTGGGAGTGGCTGCAGG
Taq-56          AAAAAGAGAACCCATATCCCTGACACTAGCTGTAATGCTCGGATTGGGAGTGGCTGCAGG
Taq-6           AAAAAGAGAACCCATATCCCTGACACTAGCTGTAATGCTCGGATTGGGAGTGGCTGCAGG
Taq-8           AAAAAGAGAACCCATATCCCTGACACTAGCTGTAATGCTCGGATTGGGAGTGGCTGCAGG
Taq-82          AAAAAGAGAACCCATATCCCTGACACTAGCTGTAATGCTCGGATTGGGAGTGGCTGCAGG
                ********* ****************************************** *********

                      1450      1460      1470      1480      1490      1500
                         .         .         .         .         .         .
PERV-A          CGTGGGAACAGGAACGGCTGCCCTAATCACAGGACCGCAACAGCTGGAGAAAGGACTTAG
Pfu-11          CGTGGAAACAGGAACGGCTGCCCTAATCACAGGACCACAACAGCTGGAGAAAGGACTTAG
Pfu-112         CGTGGAAACAGGAACGGCTGCCCTAATCACAGGACCACAACAGCTGGAGAAAGGACTTAG
Pfu-115         CGTGGGAACAGGAATGGCTGCCCTAATCACAGGACCACAACAGCTGGAGAAAGGACTTAG
Pfu-251         CGTGGAAACAGGAACGGCTGCCCTAATCACAGGACCACAACAGCTGGAGAAAGGACTTAG
Pfu-295         CGTGGGAACAGGAACGGCTGCCCTAATCACAGGACCACAACAGCTGGAGAAAGGACTTAG
Pfu-3           CGTGGAAACAGGAACGGCTGCCCTAATCACAGGACCACAACAGCTGGAGAAAGGACTTAG
Pfu-56          CATGGGAACAGGAACGGCTGCCCTAATCACAGGACCACAACAGCTGGAGAAAGGACTTAG
Pfu-6           CGTGGGAACAGGAACGGCTGCCCTAATCACAGGACCACAACAGCTGGAGAAAGGACTTAG
Pfu-62          CGTGGAAACAGGAACAGCTGCCCTAATCACAGGACCACAACAGCTGGAGAAAGGACTTAG
Taq-10          CGTGGAAACAGGAACGGCTGCCCTAATCACAGGACCACAACAGCTGGAGAAAGGACTTAG
Taq-11          CATGGGAACAGGAACGGCTGCCCTAATCACAGGACCACAACAGCTGGAGAAAGGACTTAG
Taq-17          CATGGGAACAGGAACGGCTGCCCTAATCACAGGACCACAACAGCTGGAGAAAGGACTTAG
Taq-54          CGTGGAAACAGGAACGGCTGCCCTAATCACAGGACCACAACAGCTGGAGAAAGGACTTAG
Taq-56          CGTGGAAACAGGAACGGCTGCCCTAATCACAGGACCACAACAGCTGGAGAAAGGACTTAG
Taq-6           CATGGGAACAGGAACGGCTGCCCTAATCACAGGACCACAACAGCTGGAGAAAGGACTTAG
Taq-8           CATGGGAACAGGAACGGCTGCCCTAATCACAGGACCACAACAGCTGGAGAAAGGACTTAG
Taq-82          CATGGGAACAGGAACGGCTGCCCTAATCACAGGACCACAACAGCTGGAGAAAGGACTTAG
                * *** ********  ******************** ************************

                      1510      1520      1530      1540      1550      1560
                         .         .         .         .         .         .
PERV-A          TAACCTACATCGAATTGTAACGGAAGATCTCCAAGCCCTAGAAAAATCTGTCAGTAACCT
Pfu-11          TGACCTACATCGAATTGTAACGGAAGATCTCCAAGCCCTAGAAAAATCTGTCAGTAACCT
Pfu-112         TGACCTACATCGAATTGTAACGGAAGATCTCCAAGCCCTAGAAAAATCTGTCAGTAACCT
Pfu-115         TGACCTACATCGAATTGTAATGGAAAATCTCTAAGCCCTAGAAAAATCTGTCTGTAACCA
Pfu-251         TGACCTACATCGAATTGTAACGGAAGATCTCCAAGCCCTAGAAAAATCTGTCAGTAACCT
Pfu-295         TGACCTACATCGAATTGTAACGGAAGATCTCCAAGCCCTAGAAAAATCTGTCAGTAACCT
Pfu-3           TGACCTACATCGAATTGTAACGGAAGATCTCCAAGCCCTAGAAAAATCTGTCAGTAACCT
Pfu-56          TGACCTACATCGAATTGTAATGGAAGATCTCCAAGCCCTAGAAAAATCTGTCAGTAACCT
Pfu-6           TGACCTACATCGAATTGTAACGGAAGATCTCCAAGCCCTAGAAAAATCTGTCTGTAACCA
Pfu-62          TGACCTACATCGAATTGTAACGGAAGATCTCCAAGCCCTAGAAAAATCTGTCAGTAACCT
Taq-10          TGACCTACATCGAATTGTAACGGAAGATCTCCAAGCCCTAGAAAAATCTGTCAGTAACCT
Taq-11          TGACCTACATCGAATTGTAATGGAAGATCTCCAAGCCCTAGAAAAATCTGTCAGTAACCT
Taq-17          TGACCTACATCGAATTGTAATGGAAGATCTCCAAGCCCTAGAAAAATCTGTCAGTAACCT
Taq-54          TGACCTACATCGAATTGTAACGGAAGATCTCCAAGCCCTAGAAAAATCTGTCAGTAACCT
Taq-56          TGACCTACATCGAATTGTAACGGAAGATCTCCAAGCCCTAGAAAAGTCTGTCAGTAACCT
Taq-6           TGACCTACATCGAATTGTAATGGAAGATCTCCAAGCCCTAGAAAAATCTGTCAGTAACCT
Taq-8           TGACCTACATCGAATTGTAATGGAAGATCTCCAAGCCCTAGAAAAATCTGTCAGTAACCT
Taq-82          TGACCTACATCGAATTGTAATGGAAGATCTCCAAGCCCTAGAAAAATCTGTCAGTAACCT
                * ****************** **** ***** ************* ****** ******

                      1570      1580      1590      1600      1610      1620
                         .         .         .         .         .         .
PERV-A          GGAGGAATCCCTAACCTCCTTATCTGAAGTGGTTCTACAGAACAGAAGGGGGTTAGATCT
Pfu-11          AGAGGAATCCCTAACCTCCTTATCTGAAGTGGTTCTACAGAACAGAAGAGGGTTAGATCT
Pfu-112         AGAGGAATCCCTAACCTCCTTATCTGAAGTGGTTCTACAGAACAGAAGAGGGTTAGATCT
Pfu-115         AGAGGAATCCCTAACCTCCTTATCTGAAGTGGTTCTACAGAACAGAAGGGGGTTAGATCT
Pfu-251         AGAGGAATCCCTAACCTCCTTATCTGAAGTGGTTCTACAGAACAGAAGAGGGTTAGATCT
Pfu-295         AGAGGAATCCCTAACCTCCTTATCTGAAGTGGTTCTACAGAACAGAAGAGGGTTAGATCT
Pfu-3           AGAGGAATCCCTAACCTCCTTATCTGAAGTGGTTCTACAGAACAGAAGAGGGTTAGATCT
Pfu-56          AGAGGAATCCCTAACCTCCTTATCTGAAGTGGTTCTACAGAACAGAAGGGGGTTAGATCT
Pfu-6           AGAGGAATCCCTAACCTCCTTATCTGAAGTGGTTCTACAGAACAGAAGGGGGTTAGATCT
```

```
Pfu-62          AGAGGAATCCCTAACCTCCTTATCTGAAGTGGTTCTACAGAACAGAAGAGGGTTAGATCT
Taq-10          AGAGGAATCCCTAACCTCCTTATCTGAAGTGGTTCTACAGAACAGAAGAGGGTTAGATCT
Taq-11          AGAGGAATCCCTAACCTCCTTATCTGAAGTGGTTCTACAGAACAGAAGGGGGTTAGATCT
Taq-17          AGGGGAATCCCTAACCTCCTTATCTGAAGTGGTTCTACAGAACAGAAGGGGGTTAGATCT
Taq-54          AAAGGAATCCCTAACCTCCTTATCTGAAGTGGTTCTACAGAACAGAAGAGGGTTAGATCT
Taq-56          AGAGGAATCCCTAACCTCCTTATCTGAAGTGGTTCTACAGAACAGAAGAGGGTTAGATCT
Taq-6           AGAGGAATCCCTAACCTCCTTATCTGAAGTGGTTCTACAGAACAGAAGGGGGTTAGATCT
Taq-8           AGAGGAATCCCTAACCTCCTTATCTGAAGTGGTTCTACAGAACAGAAGGGGGTTAGATCT
Taq-82          AGAGGAATCCCTAACCTCCTTATCTGAAGTGGTTCTACAGAACAGAAGGGGGTTAGATCT
                   ******************************************** ***********

                       1630      1640      1650      1660      1670      1680
                          .         .         .         .         .         .
PERV-A          GTTATTTCTAAAAGAAGGAGGGTTATGTGTAGCCTTAAAAGAGGAATGCTGCTTCTATGT
Pfu-11          GTTATTTCTAAAAGAAGGTGGGTTATGTGTAGCCTTAAAAGAAGAATGTTGCTTCTATGT
Pfu-112         GTTATTTCTAAAAGAAGGTGGGTTATGTGTAGCCTTAAAAGAAGAATGTTGCTTCTATGT
Pfu-115         GTTATTTCTAAAAGAAGGTGGGTTATGTGTAGCCTTAAAAGAAGAATGTTGCTTCTATGT
Pfu-251         GTTATTTCTAAAAGAAGGTGGGTTATGTGTAGCCTTAAAAGAAGAATGTTGCTTCTATGT
Pfu-295         GTTATTTCTAAAAGAAGGTGGGTTATGTGTAGCCTTAAAAGAAGAATGTTGCTTCTATGT
Pfu-3           GTTATTTCTAAAAGAAGGTGGGTTATGTGTAGCCTTAAAAGAAGAATGTTGCTTCTATGT
Pfu-56          GTTATTTCTAAAAGAAGGTGGGTTATGTGTAGCCTTAAAAGAAGAATGTTGCTTCTATGT
Pfu-6           GTTATTTCTAAAAGAAGGTGGGTTATGTGTAGCCTTAAAAGAAGAATGTTGCTTCTATGT
Pfu-62          GTTATTTCTAAAAGAAGGTGGGTTATGTGTAGCCTTAAAAGAAGAATGTTGCTTCTATGT
Taq-10          GTTATTTCTAAAAGAAGGTGGGTTATGTGTAGCCTTAAAAGAAGAATGTTGCTTCTATGT
Taq-11          GTTATTTCTAAAAGAAGGTGGGTTATGTGTAGCCTTAAAAGAAGAATGTTGCTTCTATGT
Taq-17          GTTATTTCTAAAAGAAGGTGGGTTATGTGTAGCCTTAAAAGAAGAATGTTGCTTCTATGT
Taq-54          GTTATTTCTAAAAGAAGGTGGGTTATGTGTAGCCTTAAAAGAAGAATGTTGCTTCTATGT
Taq-56          GTTATTTCTAAAAGAAGGTGGGTTATGTGTAGCCTTAAAAGAAGAATGTTGCTTCTATGT
Taq-6           GTTATTTCTAAAAGAAGGTGGGTTATGTGTAGCCTTAAAAGAAGAATGTTGCTTCTATGT
Taq-8           GTTATTTCTAAAAGAAGGTGGGTTATGTGTAGCCTTAAAAGAAGAATGTTGCTTCTATGT
Taq-82          GTTATTTCTAAAAGAAGGTGGGTTATGTGTAGCCTTAAAAGAAGAATGTTGCTTCTATGT
                ****************** ********************** ***** ***********

                       1690      1700      1710      1720      1730      1740
                          .         .         .         .         .         .
PERV-A          AGATCACTCAGGAGCCATCAGAGACTCCATGAGCAAGCTTAGAGAAAGGTTAGAGAGGCG
Pfu-11          AGATCACTCAGGAGCCATCAGAGACTCCATGAGCAAGCTTAGAGAAAGGTTAGAGAGGCG
Pfu-112         AGATCACTCAGGAGCCATCAGAGACTCCATGAGCAAGCTTAGAGAAAGGTTAGAGAGGCG
Pfu-115         GGATCACTCAGGAGCCATCAGAGACTCCATGAGCAAGCTTAGAGAAAGGTTAGAGAGGCG
Pfu-251         AGATCACTCAGGAGCCATCAGAGACTCCATGAGCAAGCTTAGAGAAAGGTTAGAGAGGCA
Pfu-295         AGATCACTCAGGAGCCATCAGAGACTCCATGAGCAAGCTTAGAGAAAGGTTAGAGAGGCA
Pfu-3           AGATCACTCAGGAGCCATCAGAGACTCCATGAGCAAGCTTAGAGAAAGGTTAGAGAGGCG
Pfu-56          GGATCACTCAGGAGCCATCAGAGACTCCATGAGCAAGCTTAGAGAAAGGTTAGAGAGGCA
Pfu-6           GGATCACTCAGGAGCCATCAGAGACTCCATGAGCAAGCTTAGAGAAAGGTTAGAGAGGCG
Pfu-62          AGATCACTCAGGAGCCATCAGAGACTCCATGAGCAAGCTTAGAGAAAGGTTAGAGAGGCG
Taq-10          AGATCACTCAGGAGCCATCAGAGACTCCATGAGCAAGCTTAGAGAAAGGTTAGAGAGGCG
Taq-11          GGATCACTCAGGAGCCATCAGAGACTCCATGAGCAAGCTTAGAGAAAGGTTAGAGAGGCA
Taq-17          GGATCACTCAGGAGCCATCAGAGACTCCATGAGCAAGCTTAGAGAAAGGTTAGAGAGGCA
Taq-54          AGATCACTCAGGAGCCATCAGAGACTCCATGAGCAAGCTTAGAGAAAGGTTAGAGAGGCG
Taq-56          AGATCACTCAGGAGCCAACAGAGACTCCATGAGCAAGCTTAGAGAAAGGTTAGAGAGGCG
Taq-6           GGATCACTCAGGAGCCATCAGAGACTCCATGAGCAAGCTTAGAGAAAGGTTAGAGAGGCA
Taq-8           GGATCACTCAGGAGCCATCAGAGACTCCATGAGCAAGCTTAGAGAAAGGTTAGAGAGGCG
Taq-82          GGATCACTCAGGAGCCATCGGAGACTCCATGAGCAAGCTTAGAGAAAGGTTAGAGAGGCA
                 **************** * ************************************

                       1750      1760      1770      1780
                          .         .         .         .
PERV-A          TCGAAGGGAAAGAGAGGCTGACCAGGGGTGGTTTGAAGGATGGT
Pfu-11          TCGAAGGGAAAGAGAGGCTGACCAGGGGTGGTTTGAAGGATGGT
Pfu-112         TCGAAGGGAAAGAGAGGCTGACCAGGGGTGGTTTGAAGGATGGT
Pfu-115         TCGAAGGGAAAGAGAGGCTGACCAGGGGTGGTTTGAAGGATGGT
Pfu-251         TCGAAGGGAAAGAGAGGCTGACCAGGGGTGGTTTGAAGGATGGT
Pfu-295         TCGAAGGGAAAGAGAGGCTGACCAGGGGTGGTTTGAAGGATGGT
Pfu-3           TCGAAGGGAAAGAGAGGCTGACCAGGGGTGGTTTGAAGGATGGT
Pfu-56          TCAAAGGGAAAGAGAGGCTGACCAGGGGTGGTTTGAAGGATGGT
Pfu-6           TCGAAGGGAAAGAGAGGCTGACCAGGGGTGGTTTGAAGGATGGT
```

```
Pfu-62          TCGAAGGGAAAGAGAGGCTGACCAGGGGTGGTTTGAAGGATGGT
Taq-10          TCGAAGGGAAAGAGAGGCTGACCAGGGGTGGTTTGAAGGATGGT
Taq-11          TCAAAGGGAAAGAGAGGCTGACCAGGGGTGGTTTGAAGGATGGT
Taq-17          TCAAAGGGAAAGAGAGGCTGACCAGGGGTGGTTTGAAGGATGGT
Taq-54          TCGAAGGGAAAGAGAGGCTGACCAGGGGTGGTTTGAAGGATGGT
Taq-56          TCGAAGGGAAAGAGAGGCTGACCAGGGGTGGTTTGAAGGATGGT
Taq-6           TCAAAGGGAAAGAGAGGCTGACCAGGGGTGGTTTGAAGGATGGT
Taq-8           TCGAAGGGAAAGAGAGGCTGACCAGGGGTGGTTTGAAGGATGGT
Taq-82          TCAAAGGGAAAGAGAGGCTGACCAGGGGTGGTTTGAAGGATGGT
                ** **********************++++++++++++++++++
```

(b) Alignment of nucleotide sequences of 9 PERV-B clones with PERV-B published sequences. The start codon is indicated with box. Sequences undersigned with + are the primer used for PCR and sequences undersigned with * are consensus sequences.



```
PERV-B          AGGCACCTGGTGGCCTGAACTGCATTTCTGCCTCCGATTGATTAACCCCGCTGTTAAAAG
Pfu-1           AGGCACCTGGTGGCCTGAACTGCATTTCTGCCTCCGATTGATTAACCCCGCTGTTAAAAG
Taq-12          AGGCACCTGGTGGCCTGAACTGCATTTCTGCCTCCGATTGATTAACCCCGCTGTTAAAAG
Taq-14          AGGCACCTGGTGGCCTGAACTGCATTTCTGCCTCCGATTGATTAACCCCGCTGTTAAAAG
Taq-20          AGGCACCTGGTGGCCTGAACTGCATTTCTGCCTCCGATTGATTAACCCCGCTGTTAAAAG
Taq-24          AGGCACCTGGTGGCCTGAACTGCATTTCTGCCTCCGATTGATTAACCCCGCTGTTAAAAG
Taq-29          AGGCACCTGGTGGCCTGAACTGCATTTCTGCCTCCGATTGATTAACCCCGCTGTTAAAAG
Taq-3           AGGCACCTGGTGGCCTGAGCTGCATTTCTGCCTCCGATTGATTAACCCCGCTGTTAAAAG
Taq-4           AGGCACCTGGTGGCCTGAACTGCATTTCTGCCTCCGATTGATTAACCCCGCTGTTAAAAG
Taq-9           AGGCACCTGGTGGCCTGAACCGCATTTCTGCCTCCGATTGATTAACCCCGCTGTTAAAAG
                ***************** * ************************************

                       310       320       330       340       350       360
                         .         .         .         .         .         .
PERV-B          CACACCTCCCAACCTAGTCCGTAGTTATGGGTTCTATTGCTGCCCAGGCACAGAGAAAGA
Pfu-1           CACACCTCCCAACCTAGTCCGTAGTTATGGGTTCTATTGCTGCCCAGGCACAGAGAAAGA
Taq-12          CACACCTCCCAACCTAGTCCGTAGTTATGGGTTCTATTGCTGCCCAGGCACAGAGAAAGA
Taq-14          CACACCTCCCAACCTAGTCCGTAGTTATGGGTTCTATTGCTGCCCAGGCACAGAGAAAGA
Taq-20          CACACCTCCCAACCTAGTCCGTAGTTATGGGTTCTATTGCTGCCCAGGCACAGAGAAAGA
Taq-24          CACACCTCCCAACCTAGTCCGTAGTTATGGGTTCTATTGCTGCCCAGGCACAGAGAAAGA
Taq-29          CACACCTCCCAACCTAGTCTGTAGTTATGGGTTCTATTGCTGCCCAGGCACAGAGAAAGA
Taq-3           CACACCTCCCAACCTAGTCCGTAGTTATGGGTTCTATTGCTGCCCAGGCACAGAGAAAGA
Taq-4           CACACCTCCCAACCTAGTCCGTAGTTATGGGTTCTATTGCTGCCCAGGCACAGAGAAAGA
Taq-9           CACACCTCCCAACCTAGTCCGTAGTTATGGGTTCTATTGCTGCCCAGGCACAGAGAAAGA
                ****************** ************************************

                       370       380       390       400       410       420
                         .         .         .         .         .         .
PERV-B          GAAATACTGTGGGGGTTCTGGGGAATCCTTCTGTAGGAGATGGAGCTGCGTCACCTCCAA
Pfu-1           GAAATACTGTGGGGGTTCTGGGGAATCCTTCTGTAGGAGATGGAGCTGCGTCACCTCCAA
Taq-12          GAAATACTGTGGGGGTTCTGGGGAATCCTTCTGTAGGAGATGGAGCTGCGTCACCTCCAA
Taq-14          GAAATACTGTGGGGGTTCTGGGGAATCCTTCTGTAGGAGATGGAGCTGCGTCACCTCCAA
Taq-20          GAAATACTGTGGGGGTTCTGGGGAATCCTTCTGTAGGAGATGGAGCTGCGTCACCTCCAA
Taq-24          GAAATACTGTGGGGGTTCTGGGGAATCCTTCTGTAGGAGATGGAGCTGCGTCACCTCCAA
Taq-29          GAAATACTGTGGGGGTTCTGGGGAATCCTTCTGTAGGAGATGGAGCTGCGTCACCTCCAA
Taq-3           GAAATACTGTGGGGGTTCTGGGGAATCCTTCTGTAGGAGATGGAGCTGCGTCACCTCCAA
Taq-4           GAAATACTGTGGGGGTTCTGGGGAATCCTTCTGTAGGAGATGGAGCTGCGTCACCTCCAA
Taq-9           GAAATACTGTGGGGGTTCTGGGGAATCCTTCTGTAGGAGATGGAGCTGCGTCACCTCCAA
                ************************************************************

                       430       440       450       460       470       480
                         .         .         .         .         .         .
PERV-B          CGATGGAGACTGGAAATGGCCGATCTCTCTCCAGGACCGGGTAAAATTCTCCTTTGTCAA
Pfu-1           CGATGGAGACTGGAAATGGCCGATCTCTCTCCAGGACCGGGTAAAATTCTCCTTTGTCAA
Taq-12          CGATGGAGACTGGAAATGGCCGATCTCTCTCCAGGACCGGGTAAAATTCTCCTTTGTCAA
Taq-14          CGATGGAGACTGGAAATGGCCGATCTCTCTCCAGGACCGGGTAAAATTCTCCTTTGTCAA
Taq-20          CGATGGAGACTGGAAATGGCCGATCTCTCTCCAGGACCGGGTAAAATTCTCCTTTGTCAA
Taq-24          CGATGGAGACTGGAAATGGCCGATCTCTCTCCAGGACCGGGTAAAATTCTCCTTTGTCAA
Taq-29          CGATGGAGACTGGAAATGGCCGATCTCTCTCCAGGACCGGGTAAAATTCTCCTTTGTCAA
Taq-3           CGATGGAGACTGGAAATGGCCGATCTCTCTCCAGGACCGGGTAAAATTCTCCTTTGTCAA
Taq-4           CGATGGAGACTGGAAATGGCCGATCTCTCTCCAGGACCGGGTAAAATTCTCCTTTGTCAA
Taq-9           CGATGGAGACTGGAAATGGCCGATCTCTCTCCAGGACCGGGTAAAATTCTCCTTTGTCAA
                ************************************************************

                       490       500       510       520       530       540
                         .         .         .         .         .         .
PERV-B          TTCCGGCCCGGGCAAGTACAAAGTGATGAAACTATATAAAGATAAGAGCTGCTCCCCATC
Pfu-1           TTCCGGCCCGGGCAAGTACAAAGTGATGAAACTATATAAAGATAAGAGCTGCTCCCCATC
Taq-12          TTCCGGCCCGGGCAAGTACAAAGTGATGAAACTATATAAAGATAAGAGCTGCTCCCCATC
Taq-14          TTCCGGCCCGGGCAAGTACAAAGTGATGAAACTATATAAAGATAAGAGCTGCTCCCCATC
Taq-20          TTCCGGCCCGGGCAAGTACAAAGTGATGAAACTATATAAAGATAAGAGCTGCTCCCCATC
Taq-24          TTCCGGCCCGGGCAAGTACAAAGTGATGAAACTATATAAAGATAGGAGCTGCTCCCCATC
Taq-29          TTCCGGCCCGGGCAAGTACAAAGTGATGAAACTATATAAAGATAAGAGCTGCTCCCCATC
Taq-3           TTCCGGCCCGGGCAAGTACAAAGTGATGAAACTATATAAAGATAAGAGCTGCTCCCCATC
Taq-4           TTCCGGCCCGGGCAAGTACAAAGTGATGAAACTATATAAAGATAAGAGCTGCTCCCCATC
Taq-9           TTCCGGCCCGGGCAAGTACAAAGTGATGAAACTATATAAAGATAAGAGCTGCTCCCCATC
```

```
                ****************************************** ***************

                       550       560       570       580       590       600
                         .         .         .         .         .         .
PERV-B          AGACTTAGATTATCTAAAGATAAGTTTCACTGAAAAAGGAAAACAGGAAAATATTCAAAA
Pfu-1           AGACTTAGATTATCTAAAGATAAGTTTCACTGAAAAAGGAAAACAGGAAAATATTCAAAA
Taq-12          AGACTTAGATTATCTAAAGATAAGTTTCACTGAAAAAGGAAAACAGGAAAATATTCAAAA
Taq-14          AGACTTAGATTATCTAAAGATAAGTTTCACTGAAAAAGGAAAACAGGAAAATATTCAAAA
Taq-20          AGACTTAGATTATCTAAAGATAAGTTTCACTGAAAAAGGAAAACAGGAAAATATTCAAAA
Taq-24          AGACTTAGATTATCTAAAGATAAGTTTCACTGAAAAAGGAAAACAGGAAAATATTCAAAA
Taq-29          AGACTTAGATTATCTAAAGATAAGTTTCACTGAAAAAGGAAAACAGGAAAATATTCAAAA
Taq-3           AGACTTAGATTATCTAAAGATAAGTTTCACTGAAAAAGGAAAACAGGAAATTATTCAAAA
Taq-4           AGACTTAGATTATCTAAAGATAAGTTTCACTGAAAAAGGAAAACAGGAAAATATTCAAAA
Taq-9           AGACTTAGATTATCTAAAGATAAGTTTCACTGAAAAAGGAAAACAGGAAAATATTCAAAA
                ************************************************ *********

                       610       620       630       640       650       660
                         .         .         .         .         .         .
PERV-B          GTGGATAAATGGTATGAGCTGGGGAATAGTTTTTTATAAATATGGCGGGGGAGCAGGGTC
Pfu-1           GTGGATAAATGGTATGAGCTGGGGAATAGTTTTTTATAAATATGGCGGGGGAGCAGGGTC
Taq-12          GTGGATAAATGGTATGAGCTGGGGAATAGTTTTTTATAAATATGGCGGGGGAGCAGGGTC
Taq-14          GTGGATAAATGGTATGAGCTGGGGAATAGTTTTTTATAAATATGGCGGGGGAGCAGGGTC
Taq-20          GTGGATAAATGGTATGAGCTGGGGAATAGTTTTTTATAAATATGGCGGGGGGGCAGGGTC
Taq-24          GTGGATAAATGGTATGAGCTGGGGAATAGTTTTTTATAAATATGGCGGGGGAGCAGGGTC
Taq-29          GTGGATAAATGGCATGAGCTGGGGAATAGTTTTTTATAAATATGGCGGGGGAGCAGGGTC
Taq-3           GTGGATAAATGGTATGAGCTGGGGAATAGTTTTTTATAAATATGGCGGGGGAGCAGGGTC
Taq-4           GTGGATAAATGGTATGAGCTGGGGAATAGTTTTTTATAAATATGGCGGGGGAGCAGGGTC
Taq-9           GTGGATAAATGGTATGAGCTGGGGAATAGTTTTTTATAAATATGGCGGGGGAGCAGGGTC
                ************ ************************************ ********

                       670       680       690       700       710       720
                         .         .         .         .         .         .
PERV-B          CACTTTAACCATTCGCCTTAGGATAGAGACGGGGACAGAACCCCCTGTGGCAGTGGGACC
Pfu-1           CACTTTAACCATTCGCCTTAGGATAGAGACGGGGACAGAACCCCCTGTGGCAGTGGGACC
Taq-12          CACTTTAACCATTCGCCTTAGGATAGAGACGGGGACAGAACCCCCTGTGGCAGTGGGACC
Taq-14          CACTTTAACCATTCGCCTTAGGATAGAGACGGGGACAGTACCCCCTGTGGCAGTGGGACC
Taq-20          CACTTTAACCATTCGCCTTAGGATAGAGACGGGGACAGAACCCCCTGTGGCAGTGGGACC
Taq-24          CACTTTAACCATTCGCCTTAGGATAGAGACGGGGACAGAACCCCCTGTGGCAGTGGGACC
Taq-29          CACTTTAACCATTCGCCTTAGGATAGAGACGGGGACAGAACCCCCTGTGGCAGTGGGACC
Taq-3           CACTTTAACCATTCGCCTTAGGATAGAGACGGGGACGGAACCCCCTGTGGCAGTGGGACC
Taq-4           CACTTTAACCATTCGCCTTAGGATAGAGACGGGGACAGAACCCCCTGTGGCAGTGGGACC
Taq-9           CACTTTAACCATTCGCCTTAGGATAGAGACGGGGACAGAACCCCCTGTGGCAGTGGGACC
                ********************************** * *********************

                       730       740       750       760       770       780
                         .         .         .         .         .         .
PERV-B          CGATAAAGTACTGGCTGAACAGGGGCCCCCGGCCCTGGAGCCACCGCATAACTTGCCGGT
Pfu-1           CGATAAAGTACTGGCTGAACAGGGGCCCCCGGCCTGGGAGCCACCGCATAACTTGCCGGT
Taq-12          CGATTAAGTACTGGCTGAACAGGGGCCCCCGGCCCTGGAGCCACCGCATAACTTGCCGGT
Taq-14          CGATAAAGTACTGGCTGAACAGGGGCCCCCGGCCCTGGAGCCACCGCATAACTTGCCGGT
Taq-20          CGATAAAGTACTGGCTGAACAGGGGCCCCCGGCCCTGGAGCCACCGCATAACTTGCCGGT
Taq-24          CGATAAAGTACTGGCTGAACAGGGGCCTCCGGCCCTGGAGCCACCGCATAACTTGCCGGT
Taq-29          CGATAAAGTACTGGCTGAACGGGGGCCCCCGGCCCTGGAGCCACCGCATAACTTGCCGGT
Taq-3           CGATAAAGTACTGGCTGAACAGGGGCCCCCGGCCCTGGAGCCACCGCATAACTTGCCGGT
Taq-4           CGATAAAGTACTGGCTGAACAGGGGCCTCCGGCCCTGGAGCCACCGCATAACTTGCCGGT
Taq-9           CGATAAAGTACTGGCTGAACAGGGGCCTCCGGCCCTGGAGCCACCGCATAACTTGCCGGT
                **** *************** ****** ******  ************************

                       790       800       810       820       830       840
                         .         .         .         .         .         .
PERV-B          GCCCCAATTAACCTCGCTGCGGCCTGACATAACACAGCCGCCTAGCAACGGTACCACTGG
Pfu-1           GCCCCAATTAACCTCGCTGCGGCCTGACATAACACAGCCGCCTAGCAACGGTACCACTGG
Taq-12          GCCCCAATTAACCTCGCTGCGGCCTGACATAACACAGCCGCCTAGCAACGGTACCACTGG
Taq-14          GCCCCAATTAACCTCGCTGCGGCCTGACATAACACAGCCGCCTAGCAACGGTACCACTGG
Taq-20          GCCCCAATTAACCTCGCTGCGGCCTGACATAACACAGCCGCCTAGCAACGGTACCACTGG
Taq-24          GCCCCAATTAACCTCGCTGCGGCCTGACATAACACAGCCGCCTAGCAACGGTACCACTGG
```

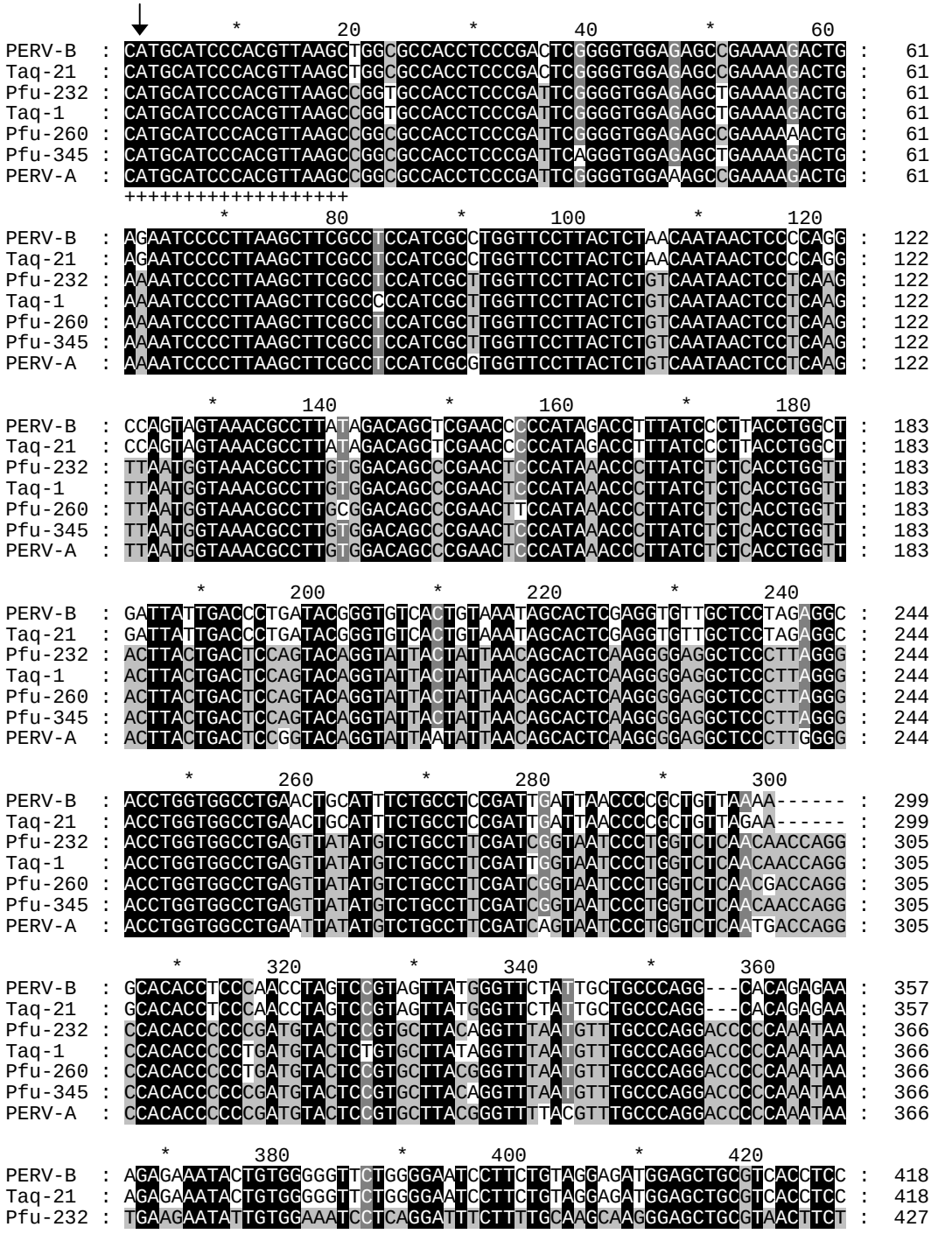
```
Taq-29          GCCCCAATTAACCTCGCTGCGGCCTGACATAACACAGCCGCCTAGCAACGGAACCACTGG
Taq-3           GCCCCAATTAACCTCGCTGCGGCCTGACATAACACAGCCACCTAGCAACGGTACCACTGG
Taq-4           GCCCCAATTAACCTCGCTGCGGCCTGACATAACACAGCCGCCTAGCAACGGTACCACTGG
Taq-9           GCCCCAATTAACCTCGCTGCGGCCTGACATAACACAGCCACCTAGCAACGGTACCACTGG
                ************************************ ********** ********

                       850       860       870       880       890       900
                         .         .         .         .         .         .
PERV-B          ATTGATTCCTACCAACACGCCTAGAAACTCCCCAGGTGTTCCTGTTAAGACAGGACAGAG
Pfu-1           ATTGATTCCTACCAACACGCCTAGAAACTCCCCAGGTGTTCCTGTTAAGACAGGACAGAG
Taq-12          ATTGATTCCTACCAACACGCCTAGAAACTCCCCAGGTGTTCCTGTTAAGACAGGACAGAG
Taq-14          ATTGATTCCTACCAACACGCCTAGAAACTCCCCAGGTGTTCCTGTTAAGACAGGACAGAG
Taq-20          ATTGATTCCTACCAACACGCCTAGAAACTCCCCAGGTGTTCCTGTTAAGACAGGACAGAG
Taq-24          ATTGATTCCTACCAACACGCCTAGAAACTCCCCAGGTGTTCCTGTTAAGACAGGACAGAG
Taq-29          ATTGATTCCTACCAACACGCCTAGAAACTCCCCAGGTGTTCCTGTTAAGACAGGACAGAG
Taq-3           ATTGATTCCTACCAACACGCCTAGAAACTCCCCAGGTGTTCCTGTTAAGACAGGACAGAG
Taq-4           ATTGATTCCTACCAACACGCCTAGAAACTCCCCAGGTGTTCCTGTTAAGACAGGACAGAG
Taq-9           ATTGATTCCTACCAACACGCCTAGAAACTCCCCAGGTGTTCCTGTTAAGACAGGACAGAG
                ************************************************************

                       910       920       930       940       950       960
                         .         .         .         .         .         .
PERV-B          ACTCTTCAGTCTCATCCAGGGAGCTTTCCAAGCCATCAACTCCACCGACCCTGATGCCAC
Pfu-1           ACTCTTCAGTCTCATCCAGGGAGCTTTCCAAGCCATCAACTCCACCGACCCTGATGCCAC
Taq-12          ACTCTTCAGTCTCATCCAGGGAGCTTTCCGAGCCATCAACTCCACCGACCCTGATGCCAC
Taq-14          ACTCTTCAGTCTCATCCAGGGAGCTTTCCAAGCCATCAACTCCACCGACCCTGATGCCAC
Taq-20          ACTCTTCAGTCTCATCCAGGGAGCTTTCCAAGCCATCAACTCCACCAACCCTGATGCCAC
Taq-24          ACTCTTCAGTCTCATCCAGGGAGCTTTCCAAGCCATCAACTCCACCGACCCTGATGCCAC
Taq-29          ACTCTTCAGTCTCATCCAGGGAGCTTTCCAAGCCATCAACTCCACCGACCCTGATGCCAC
Taq-3           ACCCTTCAGTCTCATCCAGGGAGCTTTCCAAGCCATCAACTCCACCGACCCCGATGCCAC
Taq-4           ACTCTTCAGTCTCATCCAGGGAGCTTTCCAAGCCATCAACTCCACCGACCCTGATGCCAC
Taq-9           ACCCTTCAGTCTCATCCAGGGAGCTTTCCAAGCCATCAACTCCACCGACCCCGATGCCAC
                ** ************************* *************** **** ********

                       970       980       990      1000      1010      1020
                         .         .         .         .         .         .
PERV-B          TTCTTCTTGTTGGCTTTGTCTATCCTCAGGGCCTCCTTATTATGAGGGGATGGCTAAAGA
Pfu-1           TTCTTCTTGTTGGCTTTGTCTATCCTCAGGGCCTCCTTATTATGAGGGGATGGCTAAAGA
Taq-12          TTCTTCTTGTTGGCTTTGTCTATCCTCAGGGCCTCCTTATTATGAGGGGATGGCTAAAGA
Taq-14          TTCTTCTTGTTGGCTTTGTCTATCCTCAGGGCCTCCTTATTATGAGGGGATGGCTAAAGA
Taq-20          TTCTTCTTGTTGGCTTTGTCTATCCTCAGGGCCTCCTTATTATGAGGGGATGGCTAAAGA
Taq-24          TTCTTCTTGTTGGCTTTGTCTATCCTCAGGGCCTCCTTATTATGAGGGGATGGCTAAAGA
Taq-29          TTCTTCTTGTTGGCTTTGTCTATCCTCAGGGCCTCCTTATTATGAGGGGATGGCTAAAGA
Taq-3           TTCTTCTTGTTGGCTTTGTCTATCCTCAGGGCCTCCTTATTATGAGGGGATGGCTAAAAA
Taq-4           TTCTTCTTGTTGGCTTTGTCTATCCTCAGGGCCTCCTTATTATGAGGGGATGGCTAAAGA
Taq-9           TTCTTCTTGTTGGCTTTGTCTATCCTCAGGGCCTCCTTATTATGAGGGGATGGCTAAAAA
                ************************************************************ *

                      1030      1040      1050      1060      1070      1080
                         .         .         .         .         .         .
PERV-B          AGGAAAATTCAATGTGACCAAAGAGCATAGAAATCAATGTACATGGGGGTCCCGAAATAA
Pfu-1           AGGAAAATTCAATGTGACCAAAGAGCATAGAAATCAATGTACATGGGGGTCCCGAAATAA
Taq-12          AGGAAAACTCAATGTGACCAAAGAGCATAGAAATCAATGTACATGGGGGTCCCGAAATAA
Taq-14          AGGAAAATTCAATGTGACCAAAGAGCATAGAAATCAATGTACATGGGGGTCCCGAAATAA
Taq-20          AGGAAAATTCAATGTGACCAAAGAGCATAGAAATCAATGTACATGGGGGTCCCGAAATAA
Taq-24          AGGAAAATTCAATGTGACCAAAGAGCATAGAAATCAATGTACATGGGGGTCCCGAAATAA
Taq-29          AGGAAAATTCAATGTGACCAA-GAGCATAGAAATCAATGTACATGGGGGTCCCGAAATAA
Taq-3           AGGAAAATTCAATGTGACCAAAGAGCATAGAAATCAATGTACATGGGGGTCCCGAAATAA
Taq-4           AGGAAAATTCAATGTGACCAAAGAGCATAGAAATCAATGTACATGGGGGTCCCGAAATAA
Taq-9           AGGAAAATTCAATGTGACCAAAGAGCATAGAAATCAATGTACATGGGGGTCCCGAAATAA
                ******* ************* ************************************

                      1090      1100      1110      1120      1130      1140
                         .         .         .         .         .         .
PERV-B          GCTTACCCTCACTGAAGTTTCCGGGAAGGGGACATGCATAGGAAAAGCTCCCCCATCCCA
Pfu-1           GCTTACCCTCACTGAAGTTTCCGGGAAGGGGACATGCATAGGAAAAGCTCCCCCATCCCA
```

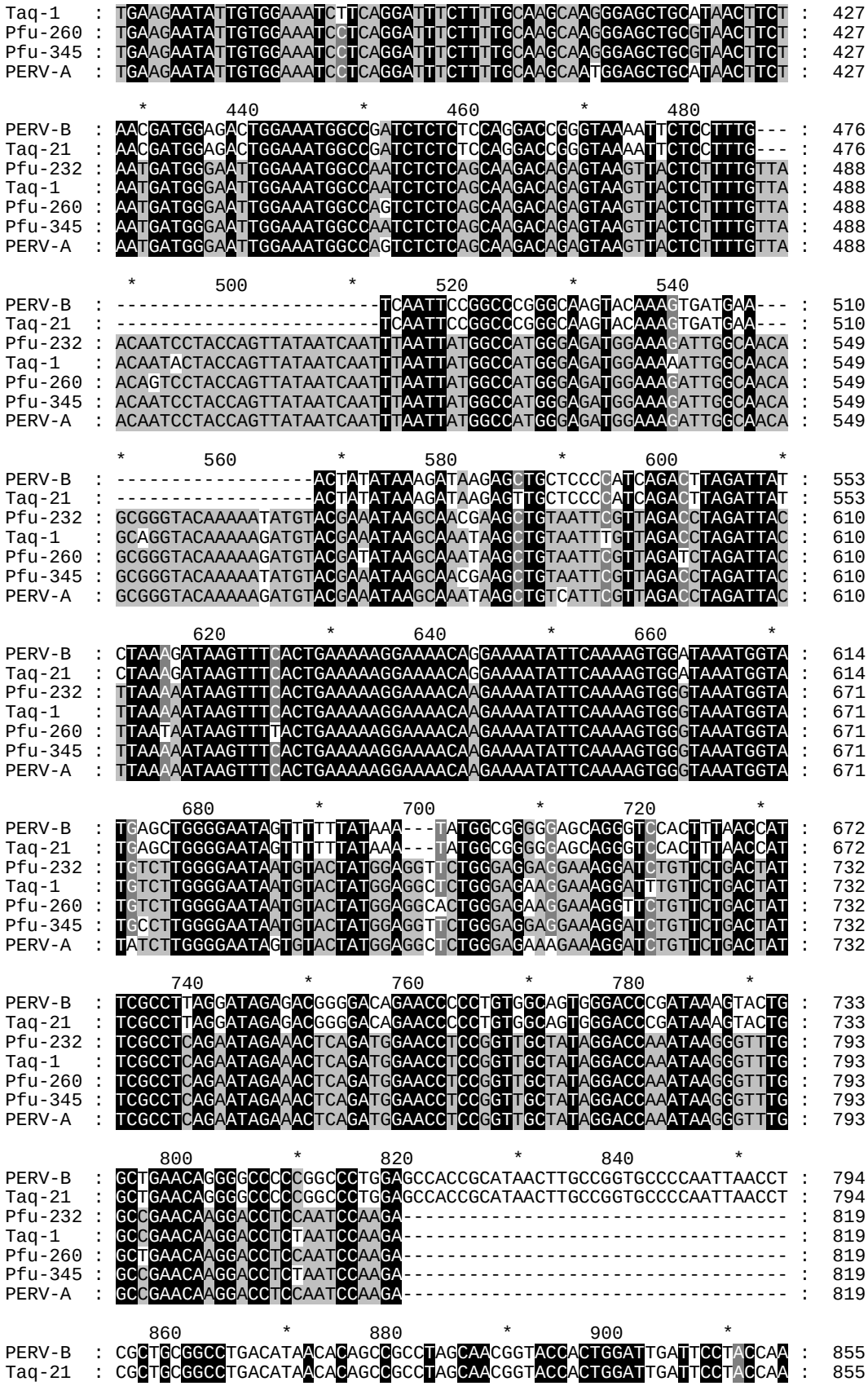
```
Taq-12          GCTTACCCTCACTGAAGTTTCCGGGAAGGGGACATGCATAGGAAAAGCTCCCCCATCCCA
Taq-14          GCTTACCCTCACTGAAGTTTCCGGGAAGGGGACATGCATAGGAAAAGCTCCCCCATCCCA
Taq-20          GCTTACCCTCACTGAAGTTTCCGGGAAGGGGACATGCATAGGAAAAGCTCCCCCATCCCA
Taq-24          GCTTACCCTCACTGAAGTTTCCGGGAAGGGGACATGCATAGGAAAAGCTCCCCCATCCCA
Taq-29          GCTTACCCTCACTGAAGTTTCCGGGAAGGGGACATGCATAGGAAAAGCTCCCCCATCCCA
Taq-3           GCTTACCCTCACTGAAGTTTCCGGGAAGGGGACATGCATAGGAAAAGCTCCCCCATCCCA
Taq-4           GCTTACCCTCACTGAAGTTTCCGGGAAGGGGACATGCATAGGAAAAGCTCCCCCATCCCA
Taq-9           GCTTACCCTCACTGAAGTTTCCGGGAAGGGGACATGCATAGGAAAAGCTCCCCCATCCCA
                ************************************************************

                      1150      1160      1170      1180      1190      1200
                         .         .         .         .         .         .
PERV-B          CCAACACCTTTGCTATAGTACTGTGGTTTATGAGCAGGCCTCAGAAAATCAGTATTTAGT
Pfu-1           CCAACACCTTTGCTATAGTACTGTGGTTTATGAGCAGGCCTCAGAAAATCAGTATTTAGT
Taq-12          CCAACACCTTTGCTATAGTACTGTGGTTTATGAGCAGGCCTCAGAAAATCAGTATTTAGT
Taq-14          CCAACACCTTTGCTATAGTACTGTGGTTTATGAGCAGGCCTCAGAAAATCAGTATTTAGT
Taq-20          CCAACACCTTTGCTATAGTACTGTGGTTTATGAGCAGGCCTCAGAAAATCAGTATTTAGT
Taq-24          CCAACACCTTTGCTATAGTACTGTGGTTTATGAGCAGGCCTCAGAAAATCAGTATTTAGT
Taq-29          CCAACACCTTTGCTATAGTACTGTGGTTTATGAGCAGGCCTCAGAAAATCAGTATTTAGT
Taq-3           CCAACACCTTTGCTATAGTACTGTGGTTTATGAGCAGGCCTCAGAAAATCAGTATTTAGT
Taq-4           CCAACACCTTTGCTATAGTACTGTGGTTTATGAGCAGGCCTCAGAAAATCAGTATTTAGT
Taq-9           CCAACACCTTTGCTATAGTACTGTGGTTTATGAGCAGGCCTCAGAAAATCAGTATTTAGT
                ************************************************************

                      1210      1220      1230      1240      1250      1260
                         .         .         .         .         .         .
PERV-B          ACCTGGTTATAACAGGTGGTGGGCATGCAATACTGGGTTAACCCCCTGTGTTTCCACCTC
Pfu-1           ACCTGGTTATAACAGGTGGTGGGCATGCAATACTGGGTTAACCCCCTGTGTTTCCACCTC
Taq-12          ACCTGGTTATAACAGGTGGTGGGCATGCAATACTGGGTTAACCCCCTGTGTTTCCACCTC
Taq-14          ACCTGGTTATAACAGGTGGTGGGCATGCAATACTGGGTTAACCCCCTGTGTTTCCACCTC
Taq-20          ACCTGGTTATAACAGGTGGTGGGCATGCAATACTGGGTTAACCCCCTGTGTTTCCACCTC
Taq-24          ACCTGGTTATAACAGGTGGTGGGCATGCAATACTGGGTTAACCCCCTGTGTTTCCACCTC
Taq-29          ACCTGGTTATAACAGGTGGTGGGCATGCAATACTGGGTTAACCCCCTGTGTTTCCACCTC
Taq-3           ACCTGGTTATAACAGGTGGTGGGCATGCAATACTGGGTTAACCCCCTGTGTTTCCACCTC
Taq-4           ACCTGGTTATAACAGGTGGTGGGCATGCAATACTGGGTTAACCCCCTGTGTTTCCACCTC
Taq-9           ACCTGGTTATAACAGGTGGTGGGCATGCAATACTGGGTTAACCCCCTGTGTTTCCACCTC
                ************************************************************

                      1270      1280      1290      1300      1310      1320
                         .         .         .         .         .         .
PERV-B          AGTCTTCAACCAATCCAAAGATTTCTGTGTCATGGTCCAAATCGTCCCCCGAGTGTACTA
Pfu-1           AGTCTTCAACCAATCCAAAGATTTCTGTGTCATGGTCCAAATCGTCCCCCGAGTGTACTA
Taq-12          AGTCTTCAACCAATCCAAAGATTTCTGTGTCATGGTCCAAATCGTCCCCCGAGTGTACTA
Taq-14          AGTCTTCAACCAATCCAAAGATTTCTGTGTCATGGTCCAAATCGTCCCCCGAGTGTACTA
Taq-20          AGTCTTCAACCAATCCAAAGATTTCTGTGTCATGGTCCAAATCGTCCCCCGAGTGTACTA
Taq-24          AGTCTTCAACCGATCCAAAGATTTCTGTGTCATGGTCCAAATCGTCCCCCGAGTGTACTA
Taq-29          AGTCTTCAACCAATCCAAAGATTTCTGTGTCATGGTCCAGCTCGTCCCCCGGGTATACTA
Taq-3           AGTCTTCAACCAATCCAAAGATTTCTGTGTCATGGTCCAAATCGTCCCCCGAGTGTACTA
Taq-4           AGTCTTCAACCAATCCAAAGATTTCTGTGTCATGGTCCAAATCGTCCCCCGAGTGTACTA
Taq-9           AGTCTTCAACCAATCCAAAGATTTCTGTGTCATGGTCCAAATCGTCCCCCGAGTGTACTA
                *********** ************************** ********** ** *****

                      1330      1340      1350      1360      1370      1380
                         .         .         .         .         .         .
PERV-B          CCATCCTGAGGAAGTGGTCCTTGATGAATATGACTATCGGTATAACCGACCAAAAAGAGA
Pfu-1           CCACCCTGAGGAAGTGGTCCTTGATGAATATGACTATCGGTATAACCGACCAAAAAGAGA
Taq-12          CCACCCTGAGGAAGTGGTCCTTGATGAATATGACTATCGGTATAACCGACCAAAAAGAGA
Taq-14          CCATCCTGAGGAAGTGGTCCTTGATGAATATGACTATCGGTATAACCGACCAAAAAGAGA
Taq-20          CCATCCTGAGGAAGTGGTCCTTGATGAATATGACTATCGGTATAACCGACCAAAAAGAGA
Taq-24          CCATCCTGAGGAAGTGGTCCTTGATGAATATGACTATCGGTATAACCGACCAAAAAGAGA
Taq-29          CCATCCTGAGGAAGTGGTCCTTGATGAATATGACTATCGGTATAACCGACCAAAAAGAGA
Taq-3           CCATCCCGAGGAAGTGGTCCTTGATGAATATGACTATCGGTATAACCGACCGAAAAGAGA
Taq-4           CCATCCTGAGGAAGTGGTCCTTGATGAATATGACTATCGGTATAACCGACCAAAAAGAGA
Taq-9           CCATCCCGAGGAAGTGGTCCTTGATGAATATGACTATCGGTATAACCGACCGAAAAGAGA
                *** ** ****************************************** ********
```

```
                1390      1400      1410      1420      1430      1440
                   .         .         .         .         .         .
PERV-B          ACCCGTATCCCTTACCCTAGCTGTAATGCTCGGATTAGGGACGGCCGTTGGCGTAGGAAC
Pfu-1           ACCCGTATCCCTTACCCTAGCTGTAATGCTCGGATTAGGGACGGCCGTTGGCGTAGGAAC
Taq-12          ACCCGTATCCCTTACCCTAGCTGTAATGCTCGGATTAGGGACGGCCGTTGGCGTAGGAAC
Taq-14          ACCCGTATCCCTTACCCTAGCTGTAATGCTCGGATTAGGGACGGCCGTTGGCGTAGGAAC
Taq-20          ACCCGTATCCCTTACCCTAGCTGTAATGCTCGGATTAGGGACGGCCGTTGGCGTAGGAAC
Taq-24          ACCCGTATCCCTTACCCTAGCTGTGATGCTCGGATTAGGGACGGCCGTTGGCGTAGGAAC
Taq-29          ACCCGTATCCCTTACCCTAGCTGTAATGCTCGGATTAGGGACGGCCGTTGGCGTAGGAAC
Taq-3           ACCCGTATCCCTTACCCTAGCTGTAATGCTCGGATTAGGGACGGCCGTTGGTGTAGGAAC
Taq-4           ACCCGTATCCCTTACCCTAGCTGTAATGCTCGGATTAGGGACGGCCGTTGGCGTAGGAAC
Taq-9           ACCCGTATCCCTTACCCTAGCTGTAATGCTCGGATTAGGGACGGCCGTTGGTGTAGGAAC
                ********************** ************************* ********

                1450      1460      1470      1480      1490      1500
                   .         .         .         .         .         .
PERV-B          AGGGACAGCTGCCCTGATCACAGGACCACAGCAGCTAGAGAAAGGACTTGGTGAGCTACA
Pfu-1           AGGGACAGCTGCCCTGATCACAGGACCACAGCAGCTAGAGAAAGGACTTGGTGAGCTACA
Taq-12          AGGGACAGCTGCCCTGATCACAGGACCACAGCAGCTAGAGAAAGGACTTGGTGAGCTACA
Taq-14          AGGGACAGCTGCCCTGATCACAGGACCACAGCAGCTAGAGAAAGGACTTGGTGAGCTACA
Taq-20          AGGGACAGCTGCCCTGATCACAGGACCACAGCAGCTAGAGAAAGGACTTGGTGAGCTACA
Taq-24          AGGGACAACTGCCCTGATCACAGGACCACAGCAGCTAGAGAAAGGACTTGGTGAGCTACA
Taq-29          AGGGACAGCTGCCCTGATCACAGGACCACAGCAGCTAGAGAAAGGACTTGGTGAGCTACA
Taq-3           AGGGACAGCTGCCCTGATCACAGGACCACAGCAGCTAGAGAAAGGACTTGGTGAGCTACA
Taq-4           AGGGACAGCTGCCCTGATCACAGGACCACAGCAGCTAGAGAAAGGACTTGGTGAGCTACA
Taq-9           AGGGACAGCTGCCCTGATCACAGGACCACAGCAGCTAGAGAAAGGACTTGGTGAGCTACA
                ******* **************************************************

                1510      1520      1530      1540      1550      1560
                   .         .         .         .         .         .
PERV-B          TGCGGCCATGACAGAAGATCTCCGAGCCTTAGAGGAGTCTGTTAGCAACCTAGAAGAGTC
Pfu-1           TGCGGCCATGACAGAAGATCTCCGAGCCTTAGAGGAGTCTGTTAGCAACCTAGAAGAGTC
Taq-12          TGCGGCCATGACAGAAGATCTCCGAGCCTTAGAGGAGTCTGTTAGCAACCTAGAAGAGTC
Taq-14          TGCGGCCATGACAGAAGATCTCCGAGCCTTAGAGGAGTCTGTTAGCAACCTAGAAGAGTC
Taq-20          TGCGGCCATGACAGAAGATCTCCGAGCCTTAGAGGAGTCTGTTAGCAACCTAGAAGAGTC
Taq-24          TGCGGCCATGACAGAAGATCTCCGAGCCTTAGAGGAGTCTGTTAGCAACCTAGAAGAGTC
Taq-29          TGCGGCCATGACAGAAGATCTCCGAGCCTTAGAGGAGTCTGTTAGCAACCTAGAAGAGTC
Taq-3           TGCGGCCATGACAGAAGATCTCCGAGCCTTAGAGGAATCTGTTAGCAACCTAGAAGAGTC
Taq-4           TGCGGCCATGACAGAAGATCTCCGAGCCTTAGAGGAGTCTGTTAGCAACCTAGAAGAGTC
Taq-9           TGCGGCCATGACAGAAGATCTCCGAGCCTTAGAGGAGTCTGTTAGCAACCTAGAAGAGTC
                ********************************** ***********************

                1570      1580      1590      1600      1610      1620
                   .         .         .         .         .         .
PERV-B          CCTGACTTCTTTGTCTGAAGTGGTTCTACAGAACCGGAGGGGATTAGATCTGCTGTTTCT
Pfu-1           CCTGACTTCTTTGTCTGAAGTGGTTCTACAGAACCGGAGGGGATTAGATCTGCTGTTTCT
Taq-12          CCTGACTTCTTTGTCTGAAGTGGTTCTACAGAACCGGAGGGGATTAGATCTGCTGTTTCT
Taq-14          CCTGACTTCTTTGTCTGAAGTGGTTCTACAGAACCGGAGGGGATTAGATCTGCTGTTTCT
Taq-20          CCTGACTTCTTTGTCTGAAGTGGTTCTACAGAACCGGAGGGGATTAGATCTGCTGTTTCT
Taq-24          CCTGACTTCTTTGTCTGAAGTGGTTCTACAGAACCGGAGGGGATTAGATCTGCTGTTTCT
Taq-29          CCTGACTTCTTTGTCTGAAGTGGTTCTACAGAACCGGAGGGGATTAGATCTGCTGTTTCT
Taq-3           CCTGACTTCTTTGTCTGAAGTGGTTCTACAGAAGTGGAGGGGATTAGATCTGCTGTTTCT
Taq-4           CCTGACTTCTTTGTCTGAAGTGGTTCTACAGAACCGGAGGGGATTAGATCTGCTGTTTCT
Taq-9           CCTGACTTCTTTGTCTGAAGTGGTTCTACAGAACCGGAGGGGATTAGATCTGCTGTTTCT
                ********************************  ************************

                1630      1640      1650      1660      1670      1680
                   .         .         .         .         .         .
PERV-B          AAGAGAAGGTGGGTTATGTGCAGCCTTAAAAGAAGAATGTTGCTTCTATGTAGATCACTC
Pfu-1           AAGAGAAGGTGGGTTATGTGCAGCCTTAAAAGAAGAATGTTGCTTCTATGTAGATCACTC
Taq-12          AAGAGAAGGTGGGTTATGTGCAGCCTTAAAAGAAGAATGTTGCTTCTATGTAGATCACTC
Taq-14          AAGAGAAGGTGGGTTATGTGCAGCCTTAAAAGAAGAATGTTGCTTCTATGTAGATCACTC
Taq-20          AAGAGAAGGTGGGTTATGTGCAGCCTTAAAAGAAGAATGTTGCTTCTACGTAGATCACTC
Taq-24          AAGAGAAGGTGGGTTATGTGCAGCCTTAAAAGAAGAATGTTGCTTCTATGTAGATCACTC
Taq-29          AAGAGAAGGTGGGTTATGTGCAGCCTTAAAAGAAGAATGTTGCTTCTACGTAGATCACTC
Taq-3           AAGAGAAGGTGGGTTATGTGCAGCCTTAAAAGAAGAATGTTGCTTCTATGTAGATCACTC
```

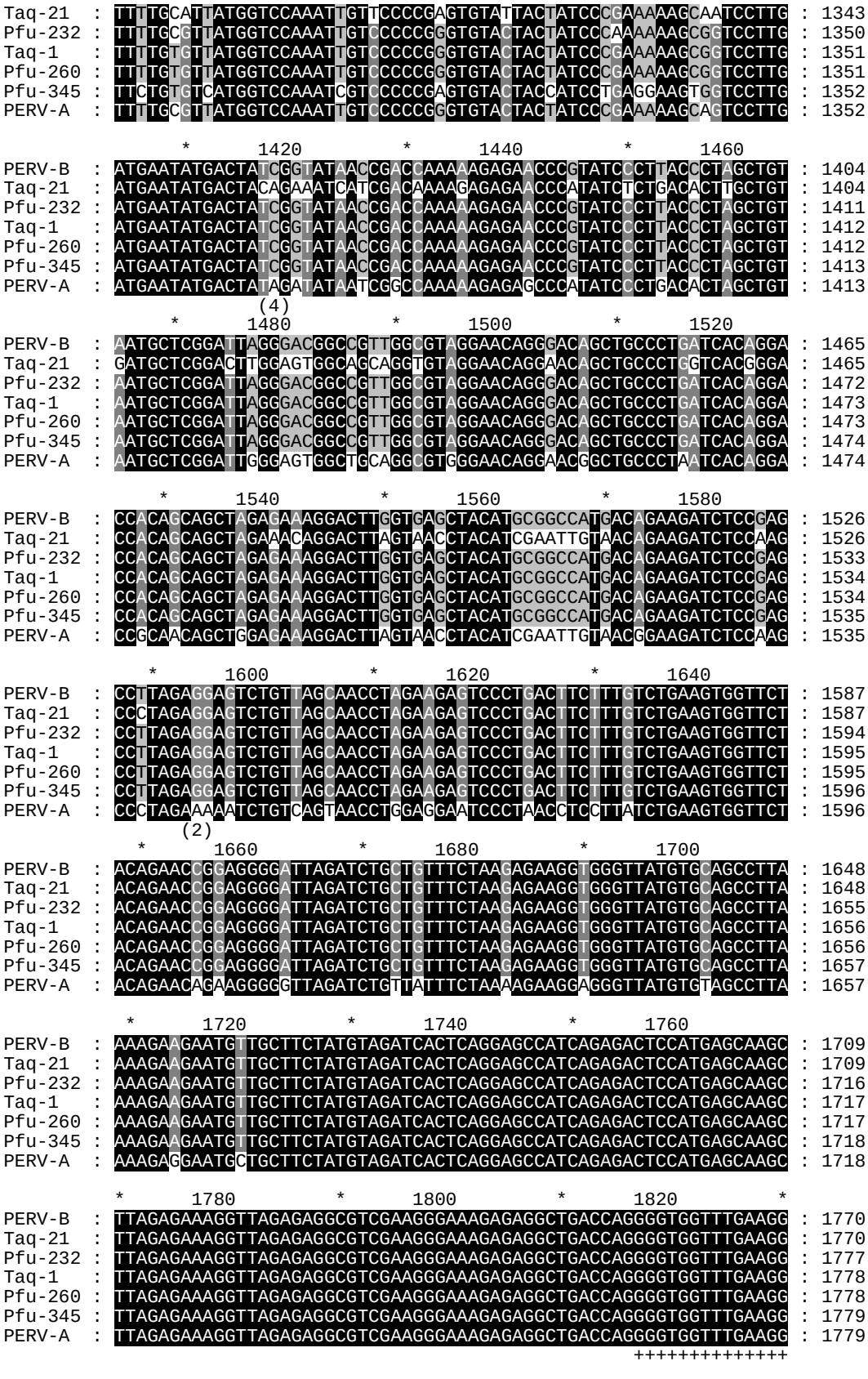
```
Taq-4           AAGAGAAGGTGGGTTATGTGCAGCCTTAAAAGAAGAATGTTGCTTCTATGTAGATCACTC
Taq-9           AAGAGAAGGTGGGTTATGTGCAGCCTTAAAAGAAGAATGTTGCTTCTATGTAGATCACTC
                ********************************************** ***********

                      1690      1700      1710      1720      1730      1740
                         .         .         .         .         .         .
PERV-B          AGGAGCCATCAGAGACTCCATGAGCAAGCTTAGAGAAAGGTTAGAGAGGCGTCGAAGGGA
Pfu-1           AGGAGCCATCAGAGACTCCATGAGCAAGCTTAGAGAAAGGTTAGAGAGGCGTCGAAGGGA
Taq-12          AGGAGCCATCAGAGACTCCATGAGCAAGCTTAGAGAAAGGTTAGAGAGGCGTCGAAGGGA
Taq-14          AGGAGCCATCAGAGACTCCATGAGCAAGCTTAGAGAAAGGTTAGAGAGGCGTCGAAGGGA
Taq-20          AGGAGCCATCAGAGACTCCATGAGCAAGCTTAGAGAAAGGTTAGAGAGGCGTCGAAGGGA
Taq-24          AGGAGCCATCAGAGACTCCATGAGCAAGCTTAGAGAAAGGTCAGAGAGGCGTCGAAGGGA
Taq-29          AGGAGCCATCAGAGACTCCATGAGCAAGCTTAGAGAAAGGTTAGAGAGGCGTCGAAGGGA
Taq-3           AGGAGCCATCAGAGACTCCATGAGCAAGCTTAGAGAAAGGTTAGAGAGGCGTTGAAGGGA
Taq-4           AGGAGCCATCAGAGACTCCATGAGCAAGCTTAGAGAAAGGTTAGAGAGGCGTCGAAGGGA
Taq-9           AGGAGCCATCAGAGACTCCATGAGCAAGCTTAGAGAAAGGTTAGAGAGGCGTCGAAGGGA
                ************************************* ********** *******

                      1750      1760      1770
                         .         .         .
PERV-B          AAGAGAGGCTGACCAGGGGTGGTTTGAAGGATGGT
Pfu-1           AAGAGAGGCTGACCAGGGGTGGTTTGAAGGATGGT
Taq-12          AAGAGAGGCTGACCAGGGGTGGTTTGAAGGATGGT
Taq-14          AAGAGAGGCTGACCAGGGGTGGTTTGAAGGATGGT
Taq-20          AAGAGAGGCTGACCAGGGGTGGTTTGAAGGATGGT
Taq-24          AAGAGAGGCTGACCAGGGGTGGTTTGAAGGATGGT
Taq-29          AAGAGAGGCTGACCAGGGGTGGTTTGAAGGATGGT
Taq-3           AAGAGAGGCTGAGCAGGGGTGGTTTGAAGGATGGT
Taq-4           AAGAGAGGCTGACCAGGGGTGGTTTGAAGGATGGT
Taq-9           AAGAGAGGCTGACCAGGGGTGGTTTGAAGGATGGT
                ************ ***++++++++++++++++++
```

(c) Alignment of nucleotide sequences of 5 recombinant clones with PERV-A and PERV-B published sequences. The start codon is indicated with an arrow. Sequences undersigned with + are the primer used for PCR and sequences. Sequences undersigned with (1), (2), (3), (4) are the breaking point between PERV-A and PERV-B. (1): *Taq*-21 PERV-B sequences change to PERV-A, (2): *Taq*-21 PERV-A sequences change to PERV-B, (3): *Pfu*-345 PERV-A sequences change to PERV-B, (4):*Taq*-1, Pfu-232, Pfu-260 PERV-A sequences change to PERV-B.




Taq-1 : TGAAGAATATTGTGGAAATCTTCAGGATTTCTTTTGCAAGCAAGGGAGCTGCATAACTTCT : 427
Pfu-260 : TGAAGAATATTGTGGAAATCCTCAGGATTTCTTTTGCAAGCAAGGGAGCTGCGTAACTTCT : 427
Pfu-345 : TGAAGAATATTGTGGAAATCCTCAGGATTTCTTTTGCAAGCAAGGGAGCTGCGTAACTTCT : 427
PERV-A : TGAAGAATATTGTGGAAATCCTCAGGATTTCTTTTGCAAGCAATGGAGCTGCATAACTTCT : 427
440 460 480
PERV-B : AACGATGGAGACTGGAAATGGCCGATCTCTCTCCAGGACCGGGTAAAATTCTCCTTTG--- : 476
Taq-21 : AACGATGGAGACTGGAAATGGCCGATCTCTCTCCAGGACCGGGTAAAATTCTCCTTTG--- : 476
Pfu-232 : AATGATGGGAATTGGAAATGGCCAATCTCTCAGCAAGACAGAGTAAGTTACTCTTTTGTTA : 488
Taq-1 : AATGATGGGAATTGGAAATGGCCAATCTCTCAGCAAGACAGAGTAAGTTACTCTTTTGTTA : 488
Pfu-260 : AATGATGGGAATTGGAAATGGCCAGTCTCTCAGCAAGACAGAGTAAGTTACTCTTTTGTTA : 488
Pfu-345 : AATGATGGGAATTGGAAATGGCCAATCTCTCAGCAAGACAGAGTAAGTTACTCTTTTGTTA : 488
PERV-A : AATGATGGGAATTGGAAATGGCCAGTCTCTCAGCAAGACAGAGTAAGTTACTCTTTTGTTA : 488
500 520 540
PERV-B : ----------TCAATTCCGGCCCGGGCAAGTACAAAGTGATGAA--- : 510
Taq-21 : ----------TCAATTCCGGCCCGGGCAAGTACAAAGTGATGAA--- : 510
Pfu-232 : ACAATCCTACCAGTTATAATCAATTTAATTATGGCCATGGGAGATGGAAAGATTGGCAACA : 549
Taq-1 : ACAATACTACCAGTTATAATCAATTTAATTATGGCCATGGGAGATGGAAAAATTGGCAACA : 549
Pfu-260 : ACAGTCCTACCAGTTATAATCAATTTAATTATGGCCATGGGAGATGGAAAGATTGGCAACA : 549
Pfu-345 : ACAATCCTACCAGTTATAATCAATTTAATTATGGCCATGGGAGATGGAAAGATTGGCAACA : 549
PERV-A : ACAATCCTACCAGTTATAATCAATTTAATTATGGCCATGGGAGATGGAAAGATTGGCAACA : 549
560 580 600
PERV-B : ----------ACTATATAAAGATAAGAGCTGCTCCCCATCAGACTTAGATTAT : 553
Taq-21 : ----------ACTATATAAAGATAAGAGTTGCTCCCCATCAGACTTAGATTAT : 553
Pfu-232 : GCGGGTACAAAAATATGTACGAAATAAGCAACGAAGCTGTAATTCGTTAGACCTAGATTAC : 610
Taq-1 : GCAGGTACAAAAAGATGTACGAAATAAGCAAATAAGCTGTAATTTGTTAGACCTAGATTAC : 610
Pfu-260 : GCGGGTACAAAAAGATGTACGATATAAGCAAATAAGCTGTAATTCGTTAGATCTAGATTAC : 610
Pfu-345 : GCGGGTACAAAAATATGTACGAAATAAGCAACGAAGCTGTAATTCGTTAGACCTAGATTAC : 610
PERV-A : GCGGGTACAAAAAGATGTACGAAATAAGCAAATAAGCTGTCATTCGTTAGACCTAGATTAC : 610
620 640 660
PERV-B : CTAAAGATAAGTTTCACTGAAAAAGGAAAACAGGAAAATATTCAAAAGTGGATAAATGGTA : 614
Taq-21 : CTAAAGATAAGTTTCACTGAAAAAGGAAAACAGGAAAATATTCAAAAGTGGATAAATGGTA : 614
Pfu-232 : TTAAAAATAAGTTTCACTGAAAAAGGAAAACAAGAAAATATTCAAAAGTGGGTAAATGGTA : 671
Taq-1 : TTAAAAATAAGTTTCACTGAAAAAGGAAAACAAGAAAATATTCAAAAGTGGGTAAATGGTA : 671
Pfu-260 : TTAATAATAAGTTTTACTGAAAAAGGAAAACAAGAAAATATTCAAAAGTGGGTAAATGGTA : 671
Pfu-345 : TTAAAAATAAGTTTCACTGAAAAAGGAAAACAAGAAAATATTCAAAAGTGGGTAAATGGTA : 671
PERV-A : TTAAAAATAAGTTTCACTGAAAAAGGAAAACAAGAAAATATTCAAAAGTGGGTAAATGGTA : 671
680 700 720
PERV-B : TGAGCTGGGGAATAGTTTTTTATAAA---TATGGCGGGGGAGCAGGGTCCACTTTAACCAT : 672
Taq-21 : TGAGCTGGGGAATAGTTTTTTATAAA---TATGGCGGGGGAGCAGGGTCCACTTTAACCAT : 672
Pfu-232 : TGTCTTGGGGAATAATGTACTATGGAGGTTCTGGGAGGAGGAAAGGATCTGTTCTGACTAT : 732
Taq-1 : TGTCTTGGGGAATAATGTACTATGGAGGCTCTGGGAGAAGGAAAGGATTTGTTCTGACTAT : 732
Pfu-260 : TGTCTTGGGGAATAATGTACTATGGAGGCACTGGGAGAAGGAAAGGTTCTGTTCTGACTAT : 732
Pfu-345 : TGCCTTGGGGAATAATGTACTATGGAGGTTCTGGGAGGAGGAAAGGATCTGTTCTGACTAT : 732
PERV-A : TATCTTGGGGAATAGTGTACTATGGAGGCTCTGGGAGAAAGAAAGGATCTGTTCTGACTAT : 732
740 760 780
PERV-B : TCGCCTTAGGATAGAGACGGGGACAGAACCCCCTGTGGCAGTGGGACCCGATAAAGTACTG : 733
Taq-21 : TCGCCTTAGGATAGAGACGGGGACAGAACCCCCTGTGGCAGTGGGACCCGATAAAGTACTG : 733
Pfu-232 : TCGCCTCAGAATAGAAACTCAGATGGAACCTCCGGTTGCTATAGGACCAAATAAGGGTTTG : 793
Taq-1 : TCGCCTCAGAATAGAAACTCAGATGGAACCTCCGGTTGCTATAGGACCAAATAAGGGTTTG : 793
Pfu-260 : TCGCCTCAGAATAGAAACTCAGATGGAACCTCCGGTTGCTATAGGACCAAATAAGGGTTTG : 793
Pfu-345 : TCGCCTCAGAATAGAAACTCAGATGGAACCTCCGGTTGCTATAGGACCAAATAAGGGTTTG : 793
PERV-A : TCGCCTCAGAATAGAAACTCAGATGGAACCTCCGGTTGCTATAGGACCAAATAAGGGTTTG : 793
800 820 840
PERV-B : GCTGAACAGGGGCCCCCGGCCCTGGAGCCACCGCATAACTTGCCGGTGCCCCAATTAACCT : 794
Taq-21 : GCTGAACAGGGGCCCCCGGCCCTGGAGCCACCGCATAACTTGCCGGTGCCCCAATTAACCT : 794
Pfu-232 : GCCGAACAAGGACCTCCAATCCAAGA---------- : 819
Taq-1 : GCCGAACAAGGACCTCTAATCCAAGA---------- : 819
Pfu-260 : GCTGAACAAGGACCTCCAATCCAAGA---------- : 819
Pfu-345 : GCCGAACAAGGACCTCTAATCCAAGA---------- : 819
PERV-A : GCCGAACAAGGACCTCCAATCCAAGA---------- : 819
860 880 900
PERV-B : CGCTGCGGCCTGACATAACACAGCCGCCTAGCAACGGTACCACTGGATTGATTCCTACCAA : 855
Taq-21 : CGCTGCGGCCTGACATAACACAGCCGCCTAGCAACGGTACCACTGGATTGATTCCTACCAA : 855

```
Pfu-232 : -ACAGAGGCCATCTCCTAACCCCTCTGATTACAATACACCTCTGGATCAGTCCCCACTGA :  879
Taq-1   : -ACAGAGGCCATCTCCTAACCCCTCTGATTACAATACACCTCTGGATCAGTCCCCACTGA :  879
Pfu-260 : -ACAGAGGCCATCTCCTAACCCCTCTGTTTACAATACACCTCTGGATTAGTCCCCCCTGA :  879
Pfu-345 : -ACAGAGGCCATCTCCTAACCCCTCTGATTACAATACACCTCTGGATCAGTCCCCACTGA :  879
PERV-A  : -ACAGAGGCCATCTCCTAACCCCTCTGATTACAATACACCTCTGGATCAGTCCCCACTGA :  879

            920         *       940         *       960         *
PERV-B  : CACGCCTAGAAACTCCCCAGGTGTTCCTGTTAAGACAGGACAGAGACTCTTCAGTCTCATC :  916
Taq-21  : CACGCCTAGAAACTCCCCAGGTGTTCCTGTTAAGACAGGACAGAGACTCTTCAGTCTCATC :  916
Pfu-232 : ---------------GCCTAACATCACTATTAAAACAGCGGCAAAACTTTTTAGCCTCATC :  925
Taq-1   : ---------------GCCTAACATCACTATTAAAACAGTGGTGAAAACTTTTTAGCCTCATC :  925
Pfu-260 : ---------------GCCTAACTTCACTATTAAAACAGGGGCGAAAACTTTTTAGCCTCATC :  925
Pfu-345 : ---------------GCCTAACATCACTATTAAAACAGTGGTGAAACTTTTTAGCCTCATC :  925
PERV-A  : ---------------GCCTAACATCACTATTAAAACAGGGGCGAAACTTTTTAGCCTCATC :  925

             980         *      1000         *      1020         *      1
PERV-B  : CAGGGAGCTTTCCAAGCCATCAACTCCACCGACCCTGATGCCACTTCTTCTTGTTGGCTTT :  977
Taq-21  : CAGGGAGCTTTCCAAGCCATCAACTCCACCGACCCTGATGCCACTTCTTCTTGTTGGCTTT :  977
Pfu-232 : CAGGGAGATTTTCAGGCTCTTAACTCCACGACTCCAGAGGCTACCTCTTCTTGTTGGCTTT :  986
Taq-1   : CAGGGAGATTTTCAGGCTCTTAACTCCACGACTCCAGAGGCTACCTCTTCTTGTTGGCTTT :  986
Pfu-260 : CAGGGAGCTTTTCAAGCTCTTAACTCCACGACTCCAGAGGCTACCTCTTCTTGTTGGCTTT :  986
Pfu-345 : CAGGGAGATTTTCAGGCTCTTAACTCCACGACTCCAGAGGCTACCTCTTCTTGTTGGCTTT :  986
PERV-A  : CAGGGAGCTTTTCAAGCTCTTAACTCCACGACTCCAGAGGCTACCTCTTCTTGTTGGCTTT :  986

            040         *      1060         *      1080         *      11
PERV-B  : GTCTATCCTCAGGGCCTCCTTATTATGAGGGGATGGCTAAAGAAGGAAAATTCAATGTGAC : 1038
Taq-21  : GTCTATCCTCAGGGCCTCCTTATTATGAGGGGATGGCTAAAGAAGGAAAATTCAATGTGAC : 1038
Pfu-232 : GCTTAGCTTCGGGCCCACCTTACTATGAGGGAATGGCTAGAGGAGGGAAATTCAATGTGAC : 1047
Taq-1   : GCTTAGCTTCGGGCCCACCTTACTATGAGGGAATGGCTAGAGGAGGGAAATTCAATGTGAC : 1047
Pfu-260 : GCTTAGCTTTGAGCCCACCTTACTATGAGGGAATGGCTAGAGGAGGGAAATTCAATGTGAC : 1047
Pfu-345 : GCTTAGCTTCGGGCCCACCTTACTATGAGGGAATGGCTAGAGGAGGGAAATTCAATGTGAC : 1047
PERV-A  : GCTTAGCTTCGGGCCCACCTTACTATGAGGGAATGGCTAGAGGAGGGAAATTCAATGTGAC : 1047

            00          *      1120          *      1140          *      116
PERV-B  : CAAAGAGCATAGAAATCAATGTACATGGGGGTCCCGAAATAAGCTTACCCTCACTGAAGTT : 1099
Taq-21  : CAAAGAGCATAGAAATCAATGTACATGGGGGTCCCGAAATAAGCTTACCCTTACTGAGGTT : 1099
Pfu-232 : AAAGGAACATAGAGACCAATGTACATGGGGATCCCAAAATAAGCTTACCCTTACTGAGGTT : 1108
Taq-1   : AAAGGAACATAGAGACCAATGTACATGGGGATCCCAAAATAAGCTTACACTTACTGAGGTT : 1108
Pfu-260 : AAAGGAACATAGAGACCAATGTACATGGGGATCCCAAAATAAGCTTACCCTTACTGAGGTT : 1108
Pfu-345 : AAAGGAACATAGAGACCAATGTACATGGGGATCCCAAAATAAGCTTACACTTACTGAGGTT : 1108
PERV-A  : AAAGGAACATAGAGACCAATGTACATGGGGATCCCAAAATAAGCTTACCCTTACTGAGGTT : 1108
                                                              (1)
            0           *      1180          *      1200          *      1220
PERV-B  : TCCGGGAAGGGGACATGCATAGGAAAAAGCTCCCCCATCCCACCAACACCTTTGCTATAGTA : 1160
Taq-21  : TCTGGAAAAGGCACCTGCATAGGAAAAGGTTCCCCCATCCCACCAACACCTTTGTAACCACA : 1160
Pfu-232 : TCTGGAAAAGGCACCTGCATAGG-GAGGTTCCCCCATCCCACCAACACCTTTGTAACCACA : 1168
Taq-1   : TCTGGAAAAGGCACCTGCATAGG-GAGGTTCCCCCATCCCACCAACACCTTTGTAACCACA : 1168
Pfu-260 : TCTGGAAAAGGCACCTGCATAGG-GAGGTTCCCCCATCCCACCAACACCTTTGTAACCACA : 1168
Pfu-345 : TCCGGGAAGGGGACATGCATAGGAAAAGCTCCCCCATCCCACCAACACCTTTGCTATAGTA : 1169
PERV-A  : TCTGGAAAAGGCACCTGCATAGGGATGGTTCCCCCATCCCACCAACACCTTTGTAACCACA : 1169
             (3)
                        *      1240          *      1260          *      1280
PERV-B  : CTGTGGTTTATGAGCAGGCCTCAGAAAATCAGTATTTAGTACCTGGTTATAACAGGTGGTG : 1221
Taq-21  : CTGAAGCCTTTAATCAAACCTCTGAGAGTCAATATCTGGTACCTGGTTATGACAGGTGGTG : 1221
Pfu-232 : CTGAAGCCTTTAATCGAACCTCTGAGAGTCAGTATCTGGTACCTGGTTATGACAGGTGGTG : 1229
Taq-1   : CTGAAGCCTTTAATTGAACCTCTGAGAGTCAGTATCTGGTACCTGGTTATGACAGGTGGTG : 1229
Pfu-260 : CTGAAGCCTTTAATCGAACCTCTGAGAGTCAGTACCTGGTACCTGGTTATGACAGGTGGTG : 1229
Pfu-345 : CTGTGGTTTATGAGCAGGCCTCAGAAAATCAGTATTTAGTACCTGGTTATAACAGGTGGTG : 1230
PERV-A  : CTGAAGCCTTTAATCGAACCTCTGAGAGTCAATATCTGGTACCTGGTTATGACAGGTGGTG : 1230

                       *      1300          *      1320          *      1340
PERV-B  : GGCATGCAATACTGGGTTAACCCCCTGTGTTTCCACCTCAGTCTTCAACCAATCCAAAGAT : 1282
Taq-21  : GGCATGTAATACTGGATTAACCCCTTGTGTTTCCACCTTGGTTTTTAACCAAACTAAAGAT : 1282
Pfu-232 : GGCATGTAATACTGGATT-ACCCCTTGTGTTTCCACCCTGGTTTTCAACCAAACTAAAGAC : 1289
Taq-1   : GGCATGTAATACTGGATTAACCCCTTGTGTTTCTACCCTGGTTTTCAACCAAACTAAAGAC : 1290
Pfu-260 : GGCATGTAATACTGGATTAACCCCTTGTGTTTCCACCCTGGTTTTCAACCAAACTAAAGAC : 1290
Pfu-345 : GGCATGCAATACTGGGTTAACCCCCTGTGTTTCCACCTCAGTCTTCAACCAATCCAAAGAT : 1291
PERV-A  : GGCATGTAATACTGGATTAACCCCTTGTGTTTCCACCTTGGTTTTCAACCAAACTAAAGAC : 1291

                       *      1360          *      1380          *      1400
PERV-B  : TTCTGTGTCATGGTCCAAATCGTCCCCCGAGTGTACTACCATCCTGAGGAAGTGGTCCTTG : 1343
```



```
PERV-B  : ATGGT : 1775
Taq-21  : ATGGT : 1775
Pfu-232 : ATGGT : 1782
Taq-1   : ATGGT : 1783
Pfu-260 : ATGGT : 1783
Pfu-345 : ATGGT : 1784
PERV-A  : ATGGT : 1784
          +++++
```

Appendix 4.4. Alignment of deduced amino acid sequences 31 PERV clones with PERV-A, PERV-B and PERV-C published sequences. * indicates a stop codon.

```
PileUp of: @/home/usas00/jhlee/.WAG/pileup-22212.22230

 Symbol comparison table: GenRunData:pileuppep.cmp  CompCheck: 1254

                 GapWeight: 3.000
            GapLengthWeight: 0.100

 pileup-allpep.msf  MSF: 631  Type: P  March  6, 19100 00:42  Check: 1256 ..

 Name: PERVB           Len:   631  Check: 6763  Weight:  1.00
 Name: Taq-20          Len:   631  Check: 7273  Weight:  1.00
 Name: Pfu-1           Len:   631  Check: 7247  Weight:  1.00
 Name: Taq-4           Len:   631  Check: 7293  Weight:  1.00
 Name: Taq-14          Len:   631  Check: 7188  Weight:  1.00
 Name: Taq-12          Len:   631  Check: 5090  Weight:  1.00
 Name: Taq-24          Len:   631  Check: 8497  Weight:  1.00
 Name: Taq-9           Len:   631  Check: 7454  Weight:  1.00
 Name: Taq-3           Len:   631  Check: 5563  Weight:  1.00
 Name: Taq-21          Len:   631  Check: 6323  Weight:  1.00
 Name: Pfu-11          Len:   631  Check: 2202  Weight:  1.00
 Name: Pfu-251         Len:   631  Check:  375  Weight:  1.00
 Name: Pfu-62          Len:   631  Check: 1903  Weight:  1.00
 Name: Taq-56          Len:   631  Check: 1360  Weight:  1.00
 Name: Pfu-3           Len:   631  Check:   77  Weight:  1.00
 Name: Taq-17          Len:   631  Check: 9939  Weight:  1.00
 Name: Pfu-115         Len:   631  Check: 9273  Weight:  1.00
 Name: Pfu-6           Len:   631  Check: 3502  Weight:  1.00
 Name: PERVA           Len:   631  Check: 2822  Weight:  1.00
 Name: Pfu-345         Len:   631  Check: 4965  Weight:  1.00
 Name: PERVC           Len:   631  Check: 5439  Weight:  1.00
 Name: Pfu-112         Len:   631  Check: 5363  Weight:  1.00
 Name: Pfu-295         Len:   631  Check: 4600  Weight:  1.00
 Name: Taq-54          Len:   631  Check: 4895  Weight:  1.00
 Name: Taq-8           Len:   631  Check: 5672  Weight:  1.00
 Name: Taq-82          Len:   631  Check: 5747  Weight:  1.00
 Name: Pfu-56          Len:   631  Check: 5800  Weight:  1.00
 Name: Pfu-260         Len:   631  Check: 1845  Weight:  1.00
 Name: Taq-1           Len:   631  Check: 2904  Weight:  1.00
 Name: Taq-6           Len:   631  Check: 2090  Weight:  1.00
 Name: Pfu-232         Len:   631  Check: 4863  Weight:  1.00
 Name: Taq-11          Len:   631  Check: 1885  Weight:  1.00
 Name: Taq-10          Len:   631  Check: 6453  Weight:  1.00
 Name: Taq-29          Len:   631  Check: 8591  Weight:  1.00

//

        1                                                  50
  PERVB  MHPTLSWRHL PTRGGEPKRL RIPLSFASIA WFLTLTITPQ ASSKRLIDSS
 Taq-20  MHPTLSWRHL PTRGGEPKRL RIPLSFASIA WFLTLTITPQ ASSKRLIDSS
  Pfu-1  MHPTLSWRHL PTRGGEPKRL RIPLSFASIA WFLTLTITPQ ASSKRLIDSS
  Taq-4  MHPTLSWRHL PTRGGEPKRL RIPLSFASIA WFLTLTITPQ ASSKRLIDSS
 Taq-14  MHPTLSWRHL PTRGGEPKRL RIPLSFASIA WFLTLTITPQ ASSKRLIDSS
 Taq-12  MHPTLSWRHL PTRGGEPKRL RIPLSFASIA WFLTLTIAPQ ASSKRLIDSS
 Taq-24  MHPTLSWRHL PTRGGEPKRL RIPLSFASTA WFLTLTITPQ ASSKRLIDSS
  Taq-9  MHPTLSWRHL PTRGGEPKRL RIPLSFASTA WFLTLTITPQ ASSKRLIDSS
  Taq-3  MHPTLSWRHL PTRGGEPKRL RIPLSFASIA WFLTLTITPQ ASSKRLIDSS
 Taq-21  MHPTLSWRHL PTRGGEPKRL RIPLSFASIA WFLTLTITPQ ASSKRLIDSS
 Pfu-11  MHPTLSRCHL PIRGGELKRL KIPLSFASIA WFLTLSITPQ VNGKRLVDSP
Pfu-251  MHPTLSRCHL PIRGGELKRL KIPLSFASIA WFLTLSITPQ VNGKRLVDSP
 Pfu-62  MHPTLSRRHL PIRGGEPKKL KIPLSFASIA WFLTLSITPQ VNGKRLADSP
 Taq-56  MHPTLSRRHL PIRGGEPKKL KIPLSFASIA WFLTLSITPQ VNGKRLADSP
  Pfu-3  MHPTLSRCHL PIRGGELKRL KIPLSFASIA WFLTLSITPQ VNGKRLADSP
```

```
 Taq-17  MHPTLSRRHL PIRGGEPKRL KIPLSFASIA WFLTLSITPQ VNGKRLADSP
Pfu-115  MHPTLSWRHL PIRGGEPKRL KIPLSFASIA WFLTLSITPQ VNGKRLADSP
  Pfu-6  MHPTLSRRHL PIQGGELKRL KIPLSFASIA WFLTLSITPQ VNGKRLVDSP
  PERVA  MHPTLSRRHL PIRGGKPKRL KIPLSFASIA WFLTLSITPQ VNGKRLVDSP
Pfu-345  MHPTLSRRHL PIQGGELKRL KIPLSFASIA WFLTLSITPQ VNGKRLVDSP
  PERVC  MHPTLNRRHL PIRGGKPKRL KIPLSFASIA WFLTLSITSQ TNGMRIGDSL
Pfu-112  MHPTLSRCHL PIRGGELKRL KIPLSFASIA WFLTLSITPQ VNGKRLVDSP
Pfu-295  MHPTLSRCHL PIRGGELKRL KIPLSFASIA WFLTLSITPQ VNGKRLVDSP
 Taq-54  MHPTLSRRHL PIRGGEPKKL KIPLSFASIA WFLTLSITPQ VNGKRLADSP
  Taq-8  MHPTLSRCHL PIRGGELKRL KIPLSFASIA WFLTLSITPQ VNGKRLVDSP
 Taq-82  MHPTLSRCHL PIRGGELKRL KIPLSFASVA WFLTLSITPQ VNGKRLVDSP
 Pfu-56  MHPTLSRCHL PIRGGEPKKL KIPLSFASIA WFLTLSITPQ VNGKRLADSP
Pfu-260  MHPTLSRRHL PIRGGEPKKL KIPLSFASIA WFLTLSITPQ VNGKRLADSP
  Taq-1  MHPTLSRCHL PIRGGELKRL KIPLSFAPIA WFLTLSITPQ VNGKRLVDSP
  Taq-6  MHPTLSRRHL PIQGGELKRL KIPLSYASIA WFLTLSITPQ VNGKRLVDSP
Pfu-232  MHPTLSRCHL PIRGGELKRL KIPLSFASIA WFLTLSITPQ VNGKRLVDSP
 Taq-11  MHPTLSRRHL PIQGGELKRL KIPLSFASIA WFLTLSITPQ VNGKRLVDSP
 Taq-10  MHPTLSRRHL PIQGGELKRL KIPLSFASIA WFLTLSITPQ VNGKRLVDSP
 Taq-29  MHPTLSWRHL PTRGGEPKRL RIPLSFASIA WFLTLTITPQ ASSKRLIDSS

         51                                                  100
  PERVB  NPHRPLSLTW LIIDPDTGVT VNSTRGVAPR GTWWPELHFC LRLINPAV..
 Taq-20  NPHRPLSLTW LIIDPDTGVT VNSTRGVAPR GTWWPELHFC LRLINPAV..
  Pfu-1  NPHRPLSLTW LIIDPDTGVT VNSTRGVAPR GTWWPELHFC LRLINPAV..
  Taq-4  NPRRPLSLTW LIIDPDTGVT VNSTRGVAPR GTWWPELHFC LRLINPAV..
 Taq-14  NPHRPLSLTW LIIDPDTGVT VNSTRGVAPR GTWWPELHFC LRLINPAV..
 Taq-12  NPHRPLSLTW LIIDPDTGVT VNSTRGVAPR GTWWPELHFC LRLINPAV..
 Taq-24  NPHRPLSLTW LIIDPDTGVT VNSTRGVAPR GTWWPELHFC LRLINPAV..
  Taq-9  NPHRPLSLTW LIIDPDTGVT VNSTRGVAPR GTWWPEPHFC LRLINPAV..
  Taq-3  NPHRPLSLTW LIIDPDTGVT VNSTRGVAPR GTWWPELHFC LRLINPAV..
 Taq-21  NPHRPLSLTW LIIDPDTGVT VNSTRGVAPR GTWWPELHFC LRLINPAV..
 Pfu-11  NSHKPLSLTW LLTDSSTGIT INSTQGEAPL GTWWPELYVC LRLVIPGLND
Pfu-251  NSHKPLSLTW LLTDSSTGIT INSTQGEAPL GTWWPELYVC LRLVIPGLND
 Pfu-62  NFHKPLSLTW LLTDSSTGIT INSTQGEAPL GTWWPELYVC LRSVIPGLNN
 Taq-56  NFHKPLSLTW LLTDSSTGIT INSTQGEAPL GTWWPELYVC LRSVIPGLND
  Pfu-3  NFHKPLSLTW LLTDSSTGIT INSTQGEAPL GTWWPELYVC LRSVIPGLND
 Taq-17  NFHKPLSLTW LLTDSSTGIT INSTQGEAPL GTWWPELYVC LRSVIPGLND
Pfu-115  NFHKPLSLTW LLTDSSTGIT INSTQGEAPL GTWWPELYVC LRSVIPGLND
  Pfu-6  NSHKPLSLTW LLTDSSTGIT INSTQGEAPL GTWWPELYVC LRSVIPGLNN
  PERVA  NSHKPLSLTW LLTDSGTGIN INSTQGEAPL GTWWPELYVC LRSVIPGLND
Pfu-345  NSHKPLSLTW LLTDSSTGIT INSTQGEAPL GTWWPELYVC LRSVIPGLNN
  PERVC  NSHKPLSLTW LITDSGTGIN INNTQGEAPL GTWWPDLYVC LRSVIPSL..
Pfu-112  NSHKPLSLTW LLTDSSTGIT INSTQGEAPL GTWWPELYVC LRLVIPGLND
Pfu-295  NSHKPLSLTW LLTDSSTGIT INSTQGEAPL GTWWPELYVC LRLVIPGLND
 Taq-54  NFHKPLSLTW LLTDSSTGIT INSTQGEAPL GTWWPELYVC LRSVIPGLND
  Taq-8  NSHKPLSLTW LLTDSSTGIT INSTQGEAPL GTWWPELYVC LRLVIPGLND
 Taq-82  NSHKPLSLTW LLTDSSTGIT INSTQGEAPL GTWWPELYVC LRLVIPGLND
 Pfu-56  NFHKPLSLTW LLTDSSTGIT INSTQGEAPL GTWWPELYVC LRSVIPGLND
Pfu-260  NFHKPLSLTW LLTDSSTGIT INSTQGEAPL GTWWPELYVC LRSVIPGLND
  Taq-1  NSHKPLSLTW LLTDSSTGIT INSTQGEAPL GTWWPELYVC LRLVIPGLNN
  Taq-6  NSHKPLSLTW LLTDSSTGIT INSTQGEAPL GTWWPELYVC LRSVIPGLNN
Pfu-232  NSHKPLSLTW LLTDSSTGIT INSTQGEAPL GTWWPELYVC LRSVIPGLNN
 Taq-11  NSHKPLSLTW LLTDSSTGIT INSTQGEAPL GTWWPELYVC LRSVIPGLNN
 Taq-10  NSHKPLSLTW LLTDSSTGIT IISTQGEAPL GTWWPELYVC LRSVIPGLNN
 Taq-29  NPHRPLSLTW LIIDPDTGVT VNSTRGVAPR GTWWPELHFC LRLINPAV..

         101                                                 150
  PERVB  KSTPPNLVRS YGFYCCPG.T EKEKYCGGSG ESFCRRWSCV TSNDGDWKWP
 Taq-20  KSTPPNLVRS YGFYCCPG.T EKEKYCGGSG ESFCRRWSCV TSNDGDWKWP
  Pfu-1  KSTPPNLVRS YGFYCCPG.T EKEKYCGGSG ESFCRRWSCV TSNDGDWKWP
  Taq-4  KSTPPNLVRS YGFYCCPG.T EKEKYCGGSG ESFCRRWSCV TSNDGDWKWP
 Taq-14  KSTPPNLVRS YGFYCCPG.T EKEKYCGGSG ESFCRRWSCV TSNDGDWKWP
 Taq-12  KSTPPNLVRS YGFYCCPG.T EKEKYCGGSG ESFCRRWSCV TSNDGDWKWP
 Taq-24  KSTPPNLVRS YGFYCCPG.T EKEKYCGGSG ESFCRRWSCV TSNDGDWKWP
  Taq-9  KSTPPNLVRS YGFYCCPG.T EKEKYCGGSG ESFCRRWSCV TSNDGDWKWP
  Taq-3  KSTPPNLVRS YGFYCCPG.T EKEKYCGGSG ESFCRRWSCV TSNDGDWKWP
```

```
 Taq-21  RSTPPNLVRS YGFYCCPG.T EKEKYCGGSG ESFCRRWSCV TSNDGDWKWP
 Pfu-11  QATPPDVLCA YRFNVCPGPP NNEEYCGNPQ DFFCKQGSCI TSNDGNWKWP
Pfu-251  QATPPDVLCA YRFNVCPGPP NNEEYCGNPQ DFFCKQGSCI TSNDGNWKWP
 Pfu-62  QATPPDVLRA YRFNVCPGPP NNEEYCGNPQ DFFCKQGSCV TSNDGNWKWP
 Taq-56  QATPPDVLRA YGFNVCPGPP NNEEYCGNPQ DFFCKQGSCV TSNDGNWKWP
  Pfu-3  QATPPDVLRA YGFNVCPGPP NNEEYCGNPQ DFFCKQGSCV TSNDGNWKWP
 Taq-17  QATPPDVLCA YRFNVCPGPP NNEEYCGNPQ DFFCKQGSCV TSNDGNWKWP
Pfu-115  QATPPDVLCA YRFNVCPGPP NNEEYCGNPQ DFFCKQGSCV TSNDGNWKWP
  Pfu-6  QATPPDVLRA YRFNVCPGPP NNEEYCGNPQ DFFCKQGSCV TSNDGNWKWP
  PERVA  QATPPDVLRA YGFYVCPGPP NNEEYCGNPQ DFFCKQWSCI TSNDGNWKWP
Pfu-345  QATPPDVLRA YRFNVCPGPP NNEEYCGNPQ DFFCKQGSCV TSNDGNWKWP
  PERVC  .TSPPDILHA HGFYVCPGPP NNGKHCGNPR DFFCKQWNCV TSNDGYWKWP
Pfu-112  QATPPDVLCA YRFNVCPGPP NNEEYCGNPQ DFFCKQGSCI TSNDGNWKWP
Pfu-295  QATPPDVLCA YRFNVCPGPP NNEEYCGNPQ DFFCKQGSCI TSNDGNWKWP
 Taq-54  QATPPDVLRA YGFNVCPGPP NNGEYCGNPQ DFFCKQGSCV TSNDGNWKWP
  Taq-8  QATPPDVLCA YRFNVCPGPP NNEEYCGNPQ DFFCKQGSCI TSNDGNWKWP
 Taq-82  QATPPDVLCA YRFNVCPGPP NNEEYCGNPQ DFFCKQGSCI TSNDGNWKWP
 Pfu-56  QATPPDVLRA YGFNVCPGPP NNEEYCGNPQ DFFCKQGSCV TSNDGNWKWP
Pfu-260  QATPPDVLRA YGFNVCPGPP NNEEYCGNPQ DFFCKQGSCV TSNDGNWKWP
  Taq-1  QATPPDVLCA YRFNVCPGPP NNEEYCGNLQ DFFCKQGSCI TSNDGNWKWP
  Taq-6  QATPPDVLRA YRFNVCPGPP NNEEYCGNPQ DFFCKQGSCV TSNDGNWKWP
Pfu-232  QATPPDVLRA YRFNVCPGPP NNEEYCGNPQ DFFCKQGSCV TSNDGNWKWP
 Taq-11  QATPPDVLRA YRFNVCPGPP NNEEYCGNPQ DFFCKQGSCV TSNDGNWKWP
 Taq-10  QATPPDVLRA YRFNVCPGPP NNEEYCGNPQ DFFCKQGSCV TSNDGNWKWP
 Taq-29  KSTPPNLVCS YGFYCCPG.T EKEKYCGGSG ESFCRRWSCV TSNDGDWKWP

         151                                                   200
  PERVB  ISLQDRVKFS FVNS...... ...GPGKYK. ....VMKLYK DK..SCSPSD
 Taq-20  ISLQDRVKFS FVNS...... ...GPGKYK. ....VMKLYK DK..SCSPSD
  Pfu-1  ISLQDRVKFS FVNS...... ...GPGKYK. ....VMKLYK DK..SCSPSD
  Taq-4  ISLQDRVKFS FVNS...... ...GPGKYK. ....VMKLYK DK..SCSPSD
 Taq-14  ISLQDRVKFS FVNS...... ...GPGKYK. ....VMKLYK DK..SCSPSD
 Taq-12  ISLQDRVKFS FVNS...... ...GPGKYK. ....VMKLYK DK..SCSPSD
 Taq-24  ISLQDRVKFS FVNS...... ...GPGKYK. ....VMKLYK DR..SCSPSD
  Taq-9  ISLQDRVKFS FVNS...... ...GPGKYK. ....VMKLYK DK..SCSPSD
  Taq-3  ISLQDRVKFS FVNS...... ...GPGKYK. ....VMKLYK DK..SCSPSD
 Taq-21  ISLQDRVKFS FVNS...... ...GPGKYK. ....VMKLYK DK..SCSPSD
 Pfu-11  ISQQDRVSYS FVNNPTSYNQ FNYGHGRWKN WQQQVQKDVR NKQISCNLLD
Pfu-251  ISQQDRVSYS FVNNPTSYNQ FNYGHGRWKN WQQQVQKDVR NKQISCNLLD
 Pfu-62  ISQQDRVSYS FVNNPTSYNQ FNYGHGRWKD WQQRVQKYVR NKQRSCNSLD
 Taq-56  VSQQDRVSYS FVNSPTSYNQ FNYGHGRWKD WQQRVQKDVR NKQISCNSLD
  Pfu-3  VSQQDRVSYS FVNSPTSYNQ FNYGHGRWKD WQQRVQKDVR NKQISCNSLD
 Taq-17  VSQQDRVSYS FVNSPTSYNQ FNYGHGRWKD WQQRVQKDVR NKQISCNSLD
Pfu-115  VSQQDRVSYS FVNSPTSYNQ FNYGHGRWKD WQQRVQKDVR NKQISCNSLD
  Pfu-6  ISQQDRVSYS FVNNPTSYNQ FNYGHGRWKD WQQRVQKYVR NKQRSCNSLD
  PERVA  VSQQDRVSYS FVNNPTSYNQ FNYGHGRWKD WQQRVQKDVR NKQISCHSLD
Pfu-345  ISQQDRVSYS FVNNPTSYNQ FNYGHGRWKD WQQRVQKYVR NKQRSCNSLD
  PERVC  TSQQDRVSFS YVNTYTSSGQ FNY....... .....LTWIR TGSPKCSPSD
Pfu-112  VSQQDRVSYS FVNSPTSYNQ FNYGHGRWKD WQQRVQKDVR NKQISCNSLD
Pfu-295  ISQQDRVSYS FVNNPTSYNQ FNYGHGRWKD WQQRVQKYVR NKQRSCNSLD
 Taq-54  VSQQDRVSYS FVNSPTSYNQ FNYGHGRWKD WQQRVQKDVR NKQISCNSLD
  Taq-8  ISQQDRVSYS FVNNPTSYNQ FNYGHGRWKN WQQQVQKDVR NKQISCNLLD
 Taq-82  ISQQDRVSYS FVNNPTSYNQ FNYGHGRWKN WQQQVQKDVR NKQISCNLLD
 Pfu-56  VSQQDRVSYS FVNSPTSYNQ FNYGHGRWKD WQQRVQKDVR YKQISCNSLD
Pfu-260  VSQQDRVSYS FVNSPTSYNQ FNYGHGRWKD WQQRVQKDVR YKQISCNSLD
  Taq-1  ISQQDRVSYS FVNNTTSYNQ FNYGHGRWKN WQQQVQKDVR NKQISCNLLD
  Taq-6  ISQQDRVSYS FVNNPTSYNQ FNYGHGRWKD WQQRVQKYVR NKQRSCNSLD
Pfu-232  ISQQDRVSYS FVNNPTSYNQ FNYGHGRWKD WQQRVQKYVR NKQRSCNSLD
 Taq-11  ISQQDRVSYS FVNNPTSYNQ FNYGHGRWKD WQQRVQKYVR NKQRSCNSLD
 Taq-10  ISQQDRVSYS FVNNPTSYNQ FNYGHGRWKD WQQRVQKYVR NKQRSCNSLD
 Taq-29  ISLQDRVKFS FVNS...... ...GPGKYK. ....VMKLYK DK..SCSPSD

         201                                                   250
  PERVB  LDYLKISFTE KGKQENIQKW INGMSWGIVF YKYGG.GAGS TLTIRLRIET
 Taq-20  LDYLKISFTE KGKQENIQKW INGMSWGIVF YKYGG.GAGS TLTIRLRIET
  Pfu-1  LDYLKISFTE KGKQENIQKW INGMSWGIVF YKYGG.GAGS TLTIRLRIET
```

```
  Taq-4  LDYLKISFTE KGKQENIQKW INGMSWGIVF YKYGG.GAGS TLTIRLRIET
 Taq-14  LDYLKISFTE KGKQENIQKW INGMSWGIVF YKYGG.GAGS TLTIRLRIET
 Taq-12  LDYLKISFTE KGKQENIQKW INGMSWGIVF YKYGG.GAGS TLTIRLRIET
 Taq-24  LDYLKISFTE KGKQENIQKW INGMSWGIVF YKYGG.GAGS TLTIRLRIET
  Taq-9  LDYLKISFTE KGKQENIQKW INGMSWGIVF YKYGG.GAGS TLTIRLRIET
  Taq-3  LDYLKISFTE KGKQEIIQKW INGMSWGIVF YKYGG.GAGS TLTIRLRIET
 Taq-21  LDYLKISFTE KGKQENIQKW INGMSWGIVF YKYGG.GAGS TLTIRLRIET
 Pfu-11  LDYLKISFTE KGKQENIQKW VNGMSWGIMY YGGSGRRKGF VLTIRLRIET
Pfu-251  LDYLKISFTE KGKQENIQKW VNGMSWGIMY YGGSGRRKGF VLTIRLRIET
 Pfu-62  LDYLKISFTE KGKQENIQKW VNGMSWGIMY YGGTGRRKGS VLTIRLRIET
 Taq-56  LDYLKISFTE KGKQENIQKW VNGMSWGIMH YGGTGRRKGS VLTIRLRIET
  Pfu-3  LDYLKISFTE KGKQENIQKW VNGMSWGIMY YGGTGRRKGS VLTIRLRIET
 Taq-17  LDYLKISFTE KGKQENIQKW VNGMSWGIMY YGGTGRRKGS VLTIRLRIET
Pfu-115  LDYLKISFTE KGKQENIQKW VNGMSWGIMY YGGSGRRKGF VLTIRLRIET
  Pfu-6  LDYLKISFTE KGKQENIQKW VNGMSWGIMY YGGSGRRKGS VLTIRLRIET
  PERVA  LDYLKISFTE KGKQENIQKW VNGISWGIVY YGGSGRKKGS VLTIRLRIET
Pfu-345  LDYLKISFTE KGKQENIQKW VNGMPWGIMY YGGSGRRKGS VLTIRLRIET
  PERVC  LDYLKISFTE KGKQENILKW VNGMSWGMVY YGGSGKQPGS ILTIRLKIN.
Pfu-112  LDYLKISFTE KGKQENIQKW VNGMSWGIMY YGGTGRRKGS VLTIRLRIET
Pfu-295  LDYLKISFTE KGKQENIQKW VNGMSWGIMY YGGSGRRKGS VLTIRLRIET
 Taq-54  LDYLKISFTE KGKQENIQKW VNGMSWGIMY YGGTGRRKGS VLTIRLRIET
  Taq-8  LDYLKISFTE KGKQENIQKW VNGMSWGIMY YGGSGRRKGF VLTIRLRIET
 Taq-82  LDYLKISFTE KGKQENIQKW VNGMSWGIMY YGGSGRRKGF VLTIRLRIET
 Pfu-56  LDYLKISFTE KGKQENIQKW VNGMSWGIMY YGGTGRRKGS VLTIRLRIET
Pfu-260  LDYLIISFTE KGKQENIQKW VNGMSWGIMY YGGTGRRKGS VLTIRLRIET
  Taq-1  LDYLKISFTE KGKQENIQKW VNGMSWGIMY YGGSGRRKGF VLTIRLRIET
  Taq-6  LDYLKISFTE KGKQENIQKW VNGMSWGIMY YGGSGRRKGS VLTIRLRIET
Pfu-232  LDYLKISFTE KGKQENIQKW VNGMSWGIMY YGGSGRRKGS VLTIRLRIET
 Taq-11  LDYLKISFTE KGKQENIQKW VNGMSRGIMY YGGSGRRKGS VLTIRLRIET
 Taq-10  LDYLKISFTE KGKQENIQKW VNGMSWGIMY YGGSGRRKGS VLTIRLRIET
 Taq-29  LDYLKISFTE KGKQENIQKW INGMSWGIVF YKYGG.GAGS TLTIRLRIET

         251                                                 300
  PERVB  GTEPPVAVGP DKVLAEQGPP ALEPPHNLPV PQLTSLRPDI TQPPSNGTTG
 Taq-20  GTEPPVAVGP DKVLAEQGPP ALEPPHNLPV PQLTSLRPDI TQPPSNGTTG
  Pfu-1  GTEPPVAVGP DKVLAEQGPP AWEPPHNLPV PQLTSLRPDI TQPPSNGTTG
  Taq-4  GTEPPVAVGP DKVLAEQGPP ALEPPHNLPV PQLTSLRPDI TQPPSNGTTG
 Taq-14  GTVPPVAVGP DKVLAEQGPP ALEPPHNLPV PQLTSLRPDI TQPPSNGTTG
 Taq-12  GTEPPVAVGP D*VLAEQGPP ALEPPHNLPV PQLTSLRPDI TQPPSNGTTG
 Taq-24  GTEPPVAVGP DKVLAEQGPP ALEPPHNLPV PQLTSLRPDI TQPPSNGTTG
  Taq-9  GTEPPVAVGP DKVLAEQGPP ALEPPHNLPV PQLTSLRPDI TQPPSNGTTG
  Taq-3  GTEPPVAVGP DKVLAEQGPP ALEPPHNLPV PQLTSLRPDI TQPPSNGTTG
 Taq-21  GTEPPVAVGP DKVLAEQGPP ALEPPHNLPV PQLTSLRPDI TQPPSNGTTG
 Pfu-11  QMEPPVAIGP NKGLAEQGPL IQEQ..... ......RPSP NPSDYNTTSG
Pfu-251  QMEPPVAIGP NKGLAEQGPL IQEQ..... ......RPSP NPSVYNTTSG
 Pfu-62  QMEPPVAIGP NKGLAEQGPP IQEQ..... ......RPSP NPSVYNTTSG
 Taq-56  QMEPPVAIGP NKGLAEQGPP IQEQ..... ......RPSP NPSVYNTTSG
  Pfu-3  QMEPPVAIGP NKGLAEQGPP IQEQ..... ......RPSP NPSVYNTTSG
 Taq-17  QMEPPVAIGP NKGLAEQGPP IQEQ..... ......RPSP NPSVYNTTSG
Pfu-115  QMEPPVAIGP NKGLAEQGPP IQEQ..... ......RPSP NPSDYNTTSG
  Pfu-6  QMEPPVAIGP NKGLAEQGPP IQEQ..... ......RPSP NPSDYNTTSG
  PERVA  QMEPPVAIGP NKGLAEQGPP IQEQ..... ......RPSP NPSDYNTTSG
Pfu-345  QMEPPVAIGP NKGLAEQGPL IQEQ..... ......RPSP NPSDYNTTSG
  PERVC  QLEPPMAIGP NTVLTGQRPP TQGP..... ......GPSS .....NITSG
Pfu-112  QMEPPVAIGP NKGLAEQGPP IQEQRP.... ....SPNPSV YNTTS....G
Pfu-295  QMEPPVAIGP NKGLAEQGPP IQEQRP.... ....SPNPSV YNTTS....G
 Taq-54  QMEPPVAIGP NKGLAEQGPP IQEQRP.... ....SPNPSV YNTTS....G
  Taq-8  QMEPPVAIGP NKGLAEQGPL IQEQRP.... ....SPNPSD YNTTS....G
 Taq-82  QMEPPVAIGP NKGLAEQGPL IQEQRP.... ....SPNPSD YNTTS....G
 Pfu-56  QMEPPVAIGP NKGLAEQGPP IQEQRP.... ....SPNPSD YNTTS....G
Pfu-260  QMEPPVAIGP NKGLAEQGPP IQEQRP.... ....SPNPSV YNTTS....G
  Taq-1  QMEPPVAIGP NKGLAEQGPL IQEQRP.... ....SPNPSD YNTTS....G
  Taq-6  QMEPPVAIGP NKGLAEQGPP IQEQRP.... ....SPNPSD YNTTS....G
Pfu-232  QMEPPVAIGP NKGLAEQGPP IQEQRP.... ....SPNPSD YNTTS....G
 Taq-11  QMEPPVAIGP NKGLAEQGPP IQEQRP.... ....SPNPSD YNTTS....G
 Taq-10  QMEPPVAIGP NKGLAEQGPP IKNRGHLL.. TPLITIQPLD QSPLS...LT
```

```
 Taq-29  GTEPPVAVGP DKVLAERGPP ALEPPHNLPV PQLTSLRPDI TQPPSNGTTG

         301                                                 350
  PERVB  LIPTNTPRNS PGVPVKTGQR LFSLIQGAFQ AINSTDPDAT SSCWLCLSSG
 Taq-20  LIPTNTPRNS PGVPVKTGQR LFSLIQGAFQ AINSTNPDAT SSCWLCLSSG
  Pfu-1  LIPTNTPRNS PGVPVKTGQR LFSLIQGAFQ AINSTDPDAT SSCWLCLSSG
  Taq-4  LIPTNTPRNS PGVPVKTGQR LFSLIQGAFQ AINSTDPDAT SSCWLCLSSG
 Taq-14  LIPTNTPRNS PGVPVKTGQR LFSLIQGAFQ AINSTDPDAT SSCWLCLSSG
 Taq-12  LIPTNTPRNS PGVPVKTGQR LFSLIQGAFR AINSTDPDAT SSCWLCLSSG
 Taq-24  LIPTNTPRNS PGVPVKTGQR LFSLIQGAFQ AINSTDPDAT SSCWLCLSSG
  Taq-9  LIPTNTPRNS PGVPVKTGQR PFSLIQGAFQ AINSTDPDAT SSCWLCLSSG
  Taq-3  LIPTNTPRNS PGVPVKTGQR PFSLIQGAFQ AINSTDPDAT SSCWLCLSSG
 Taq-21  LIPTNTPRNS PGVPVKTGQR LFSLIQGAFQ AINSTDPDAT SSCWLCLSSG
 Pfu-11  SVPTE..... PNITIKTVVK LFSLIQGDFQ ALNSTTPEAT SSCWLCLASG
Pfu-251  LVPPE..... PNFTIKTGAK LFSLIQGAFQ ALNSTTPEAT SSCWLCLALG
 Pfu-62  LVPPE..... PNFTIKTGAK LFSLIQGAFQ ALNSTTPEAT SSCWLCLALG
 Taq-56  LVPPE..... PNFTIKTGAK LFSLIQGALQ ALNSTTPEAT SSCWLCLALG
  Pfu-3  LVPPE..... PNFTIKTGAK LFSLIQGAFQ ALNSTTPEAT SSCWLCLALG
 Taq-17  LVPPE..... PNFTIKTGAK LFSLIQGAFQ ALNSTTPEAT SSCWLCLALS
Pfu-115  SVPTE..... PNITIKTAAK LFSLIQGDFQ ALNSTTPEAT SSCWLCLASG
  Pfu-6  SVPTE..... PNITIKTAAK LFSLIQGDFQ ALNSTTPEAT SSCWLCLASG
  PERVA  SVPTE..... PNITIKTGAK LFSLIQGAFQ ALNSTTPEAT SSCWLCLASG
Pfu-345  SVPTE..... PNITIKTVVK LFSLIQGDFQ ALNSTTPEAT SSCWLCLASG
  PERVC  SDPTE..... SSSTTKMGAK LFSLIQGAFQ ALNSTTPEAT SSCWLCLASG
Pfu-112  LVPPE..... PNFTIKTGAK LFSLIQGAFQ ALNSTTPEAT SSCWLCLALG
Pfu-295  LVPPE..... PNFTIKTGAK LFSLIQGAFQ ALNSTTPEAT SSCWLCLALG
 Taq-54  LVPPE..... PNFTIKTGAK LFSLIQGAFQ ALNSTTPEAT SSCWLCLALG
  Taq-8  SVPTE..... PNITIKTVVK LFSLIQGDFQ ALNSTTPEAT SSCWLCLASG
 Taq-82  SVPTE..... PNITIKTVVK LFSLIQGDFQ ALNSTTPEAT SSCWLCLASG
 Pfu-56  SVPTE..... PNITIKTVVK LFSLIQGDFQ ALNSTTPEAT SSCWLCLASG
Pfu-260  LVPPE..... PNFTIKTGAK LFSLIQGAFQ ALNSTTPEAT SSCWLCLALS
  Taq-1  SVPTE..... PNITIKTVVK LFSLIQGDFQ ALNSTTPEAT SSCWLCLASG
  Taq-6  SVPTE..... PNITIKTAAK LFSLIQGGFQ ALNSTTPEAT SSCWLCLASG
Pfu-232  SVPTE..... PNITIKTAAK LFSLIQGDFQ ALNSTTPEAT SSCWLCLASG
 Taq-11  SVPTE..... PNITIKTAAK LFSLIQGDFQ ALNSTTPEAT SSCWLCLASG
 Taq-10  SLLKQ..... RQNFLASSRE IFRLLTPRLQ RLPLLVGFA* LWAHLTM...
 Taq-29  LIPTNTPRNS PGVPVKTGQR LFSLIQGAFQ AINSTDPDAT SSCWLCLSSG

         351                                                 400
  PERVB  PPYYEGMAKE GKFNVTKEHR NQCTWGSRNK LTLTEVSGKG TCIGKAPPSH
 Taq-20  PPYYEGMAKE GKFNVTKEHR NQCTWGSRNK LTLTEVSGKG TCIGKAPPSH
  Pfu-1  PPYYEGMAKE GKFNVTKEHR NQCTWGSRNK LTLTEVSGKG TCIGKAPPSH
  Taq-4  PPYYEGMAKE GKFNVTKEHR NQCTWGSRNK LTLTEVSGKG TCIGKAPPSH
 Taq-14  PPYYEGMAKE GKFNVTKEHR NQCTWGSRNK LTLTEVSGKG TCIGKAPPSH
 Taq-12  PPYYEGMAKE GKLNVTKEHR NQCTWGSRNK LTLTEVSGKG TCIGKAPPSH
 Taq-24  PPYYEGMAKE GKFNVTKEHR NQCTWGSRNK LTLTEVSGKG TCIGKAPPSH
  Taq-9  PPYYEGMAKK GKFNVTKEHR NQCTWGSRNK LTLTEVSGKG TCIGKAPPSH
  Taq-3  PPYYEGMAKK GKFNVTKEHR NQCTWGSRNK LTLTEVSGKG TCIGKAPPSH
 Taq-21  PPYYEGMAKE GKFNVTKEHR NQCTWGSRNK LTLTEVSGKG TCIGKVPPSH
 Pfu-11  PPYYEGMARG GKFNVTKEHR DQCTWGSQNK LTLTEVSGKG TCIGRVPPSH
Pfu-251  PPYYEGMARG GKFNVTKEHR DQCTWGSQNK LTLTEVSGKG TCIGRVPPSH
 Pfu-62  PPYYEGMARG GKFNVTKEHR DQCTWGSQNK LTLTEVSGKG TCIGRVPPSH
 Taq-56  PPYYEGMARG GKFNVTKEHR DQCTWGSQNK LTLTEVSGKG TCIGRVPPSH
  Pfu-3  PPYYEGMARG GKFNVTKEHR DQCTWGSQNK LTLTEVSGKG TCIGKVPPSH
 Taq-17  PPYYEGMARG GKFNVTKEHR DQCTWGSQNK LTLTEVSGKG TCIGRVPPSH
Pfu-115  PPYYEGMARG GKFNVTKEHR DQCTWGSQNK LTLTEVSGKG TCIGKVPPSH
  Pfu-6  PPYYEGMARG GKFNVTKEHR DQCTWGSQNK LTLTEVSGKG TCIGRVPPSH
  PERVA  PPYYEGMARG GKFNVTKEHR DQCTWGSQNK LTLTEVSGKG TCIGMVPPSH
Pfu-345  PPYYEGMARG GKFNVTKEHR DQCTWGSQNK LTLTEVSGKG TCIGKAPPSH
  PERVC  PPYYEGMARR GKFNVTKEHR DQCTWGSQNK LTLTEVSGKG TCIGKVPPSH
Pfu-112  PPYYEGMARG GKFNVTKEHR DQCTWGSQNK LTLTEVSGKG TCIGRFPHPT
Pfu-295  PPYYEGMARG GKFNVTKEHR DQCTWGSQNK LTLTEVSGKG TCIGRFPHPT
 Taq-54  PPYYEGMARG GKFNVTKEHR DQCTWGSQNK LTLTEVSGKG TCIGRFPHPT
  Taq-8  PPYYEGMARG GKFNVTKEHR DQCTWGSQNK LTLTEVSGKG TCIGRFPHPT
 Taq-82  PPYYEGMARG GKFNVTKEHR DQCTWGSQNK LTLTEVSGKG TCIGRFPHPT
 Pfu-56  PPYYEGMARG GKFNVTKEHR DQCTWGSQNK LTLTEVSGKG TCIGRFPHPT
```

```
Pfu-260  PPYYEGMARG GKFNVTKEHR DQCTWGSQNK LTLTEVSGKG TCIGRFPHPT
  Taq-1  PPYYEGMARG GKFNVTKEHR DQCTWGSQNK LTLTEVSGKG TCIGRFPHPT
  Taq-6  PPYYEGMARG GKFNVTKEHR DQCTWGSQNK LTLTEVSGKG TCIGRVPPSH
Pfu-232  PPYYEGMARG GKFNVTKEHR DQCTWGSQNK LTLTEVSGKG TCIGRFPHPT
 Taq-11  PPYYEGMARG GKFNVTKEHR DQCTWGSQNK LTLTEVSGKG TCIGRFPHPT
 Taq-10  REWLEEGNSM *QRNIETNVH G....DPKIS LPLLRFLEKA PA*GGSPIPP
 Taq-29  PPYYEGMAKE GKFNVTKSIE INVHGGPEIS LPSLKFPGRG HA*EKLPHPT

         401                                                 450
  PERVB  QHLCYSTVVY E......QAS ENQYLVPGYN RWWACNTGLT PCVSTSVFNQ
 Taq-20  QHLCYSTVVY E......QAS ENQYLVPGYN RWWACNTGLT PCVSTSVFNQ
  Pfu-1  QHLCYSTVVY E......QAS ENQYLVPGYN RWWACNTGLT PCVSTSVFNQ
  Taq-4  QHLCYSTVVY E......QAS ENQYLVPGYN RWWACNTGLT PCVSTSVFNQ
 Taq-14  QHLCYSTVVY E......QAS ENQYLVPGYN RWWACNTGLT PCVSTSVFNQ
 Taq-12  QHLCYSTVVY E......QAS ENQYLVPGYN RWWACNTGLT PCVSTSVFNQ
 Taq-24  QHLCYSTVVY E......QAS ENQYLVPGYN RWWACNTGLT PCVSTSVFNR
  Taq-9  QHLCYSTVVY E......QAS ENQYLVPGYN RWWACNTGLT PCVSTSVFNQ
  Taq-3  QHLCYSTVVY E......QAS ENQYLVPGYN RWWACNTGLT PCVSTSVFNQ
 Taq-21  QHLCNHTEAF N......QTS ESQYLVPGYD RWWACNTGLT PCVSTLVFNQ
 Pfu-11  QHLCNHTEAF N......RTS ESQYLVPGYD RWWACNTGLT PCVSTLVFNQ
Pfu-251  QHLCNHTEAF N......RTS ESQYLVPGYD RWWACNTGLT PCVSTLVFNQ
 Pfu-62  QHLCNHTEAF N......RTS ESQYLVPGYD RWWACNTGLT PCVSTLVFNQ
 Taq-56  QHLCNHTEAF N......RTS ESQYLVPGYD RWWACNTGLT PCVSTLVFNQ
  Pfu-3  QHLCNHTEAF N......RTS ESQYLVPGYD RWWACNTGLT PCVSTPGFNQ
 Taq-17  QHLCNHTEAF N......*TS ESQYLVPGYD RWWACNTGST PCVSTLVFNQ
Pfu-115  QHLCNHTEAF N......RTS ESQYLVPGYD RWWACNTGLT PCVSTLVFNQ
  Pfu-6  QHLCNHTEAF N......RTS ESQYLVPGYD RWWACNTGLT PCVSTLVFNQ
  PERVA  QHLCNHTEAF N......RTS ESQYLVPGYD RWWACNTGLT PCVSTLVFNQ
Pfu-345  QHLCYSTVVY E......QAS ENQYLVPGYN RWWACNTGLT PCVSTSVFNQ
  PERVC  QHLCNHTEAF N......QTS ESQYLVPGYD RWWACNTGLT PCVSTLVFNQ
Pfu-112  NTF...VTTL KPLIEPLRVS TWYLVMTGGG H...VILD*P LVFPPWFSTK
Pfu-295  NTF...VTTL KPLIEPLRVS TWYLVMTGGG H...VILD*P LVFPPWFSTK
 Taq-54  NTF...VTTL KPLIEPLRVS TWYLVMTGGG H...VILD*P LVFPPWFSTR
  Taq-8  NTF...VTTL KPLIEPLRVS IWYLVMTGGG H...VILD*P LVFPPWFSTK
 Taq-82  NTF...VTTL KPLIEPLRVS IWYLVMTGGG H...VILD*P LVFPPWFSTK
 Pfu-56  NTF...VTTL KPLIEPLRVS IWYLVMTGGG H...VILD*P LVFPPWFSTK
Pfu-260  NTF...VTTL KPLIEPLRVS TWYLVMTGGG H...VILD*P LVFPPWFSTK
  Taq-1  NTF...VTTL KPLIEPLRVS IWYLVMTGGG H...VILD*P LVFLPWFSTK
  Taq-6  QHL...CNHT EAFN...RTS ESQYLVPGYD RWWACNTGLP LVFPPWFSTK
Pfu-232  NTF...VTTL KPLIEPLRVS IWYLVMTGGG H...VILDYP LCFHPGFQPN
 Taq-11  NTF...VTTL KPLIEPLRVS IWYLVMTGGG H...VILDYP LCFHPGFQPN
 Taq-10  TPL*PH*SL* SNL*ESVPGT WL*QVV...G M...*YWINP LCFHPGFQPN
 Taq-29  NTF...AIVL WFMSRPQKIS I*YLVITGGG H...AILG*P PVFPPQSSTN

         451                                                 500
  PERVB  SKDFCVMVQI VPRVYYHPEE VVLDEYDYRY NRPKREPVSL TLAVMLGLGT
 Taq-20  SKDFCVMVQI VPRVYYHPEE VVLDEYDYRY NRPKREPVSL TLAVMLGLGT
  Pfu-1  SKDFCVMVQI VPRVYYHPEE VVLDEYDYRY NRPKREPVSL TLAVMLGLGT
  Taq-4  SKDFCVMVQI VPRVYYHPEE VVLDEYDYRY NRPKREPVSL TLAVMLGLGT
 Taq-14  SKDFCVMVQI VPRVYYHPEE VVLDEYDYRY NRPKREPVSL TLAVMLGLGT
 Taq-12  SKDFCVMVQI VPRVYYHPEE VVLDEYDYRY NRPKREPVSL TLAVMLGLGT
 Taq-24  SKDFCVMVQI VPRVYYHPEE VVLDEYDYRY NRPKREPVSL TLAVMLGLGT
  Taq-9  SKDFCVMVQI VPRVYYHPEE VVLDEYDYRY NRPKREPVSL TLAVMLGLGT
  Taq-3  SKDFCVMVQI VPRVYYHPEE VVLDEYDYRY NRPKREPVSL TLAVMLGLGT
 Taq-21  TKDFCIMVQI VPRVYYYPEK AILDEYDYRN HRQKREPISL TLAVMLGLGV
 Pfu-11  TKDFCVMVQI VPRVYYYPEK AVLDEYDYRY NRPKREPISL TLAVMLGLGV
Pfu-251  TKDFCVMVQI VPRVYYYPEK AVLDEYDYRY NRPKREPISL TLAVMLGLGV
 Pfu-62  TKDFCVMVQI VPRVYYYPEK AVLDEYDYRY NRPKREPISL TLAVMLGLGV
 Taq-56  TKDFCVMVQI VPRVYYYPEK AVLDEYDYRY NRPKREPISL TLAVMLGLGV
  Pfu-3  TKDFCVMVQI VPRVYYYPEK AVLDEYDYRY NRPKREPISL TLAVMLGLGV
 Taq-17  TKDFCVTVQI VPRVYYYPKK AVLDEYDYRY NRPKREPISL TLAVMLGLGV
Pfu-115  TKDFCVMVQI VPRVYYYPEK AVLDEYDYRY NRPKREPISL TLAVMLGLGV
  Pfu-6  TKDFCVMVQI VPRVYYYPEK AVLDEYDYRY NRPKREPISL TLAVMLGLGV
  PERVA  TKDFCVMVQI VPRVYYYPEK AVLDEYDYRY NRPKREPISL TLAVMLGLGV
Pfu-345  SKDFCVMVQI VPRVYYHPEE VVLDEYDYRY NRPKREPVSL TLAVMLGLGT
  PERVC  TKDFCIMVQI VPRVYYYPEK AILDEYDYRN HRQKREPISL TLAVMLGLGV
```

```
Pfu-112  LKTFVLWSKL SPGCTTIPKK QSLMNMTIDI IGQKENPYP* H.........
Pfu-295  LKTFVLWSKL SPGCTTIPKK QSLMNMTIDI IGQKENPYP* H.........
 Taq-54  LKTFVLWSKL SPGCTTIPKK QSLMNMTIDI IGQKENPYP* H.........
  Taq-8  LKTFALWSKL SPGCTTIPKK RSLMNMTIDI IGQKENPYP* H.........
 Taq-82  LKTFALWSKL SPGCTTIPKK RSLMNMTIDI IGQKENPYP* H.........
 Pfu-56  LKTFALWSKL SPGCTTIPKK RSLMNMTIDI IGQKENPYP* H.........
Pfu-260  LKTFVLWSKL SPGCTTIPKK RSLMNMTIGI TDQKENPYPL P.........
  Taq-1  LKTFVLWSKL SPGCTTIPKK RSLMNMTIGI TDQKENPYPL P.........
  Taq-6  LKTFALWSKL SPGCTTIPKK RSLMNMAVDI IGQKENPYP* H.........
Pfu-232  *R.LLRYGPN CPPGVLLSQK SGP**I*LSV *PTKKRT.RI P.........
 Taq-11  *R.LLRYGPN CPPGVLLSQK SGP**IWL*I *SAKKRT.HI P.........
 Taq-10  *R.LLCYGPN CPPGVLLSRK SSP**I*L*I *SAKKRT.HI P.........
 Taq-29  PKISVSWSSS SPGYTTILRK WSLMNMTIGI TDQKENPYPL P.........

         501                                                 550
  PERVB  AVGVGTGTAA LITGPQQLEK GLGELHAAMT EDLRALEESV SNLEESLTSL
 Taq-20  AVGVGTGTAA LITGPQQLEK GLGELHAAMT EDLRALEESV SNLEESLTSL
  Pfu-1  AVGVGTGTAA LITGPQQLEK GLGELHAAMT EDLRALEESV SNLEESLTSL
  Taq-4  AVGVGTGTAA LITGPQQLEK GLGELHAAMT EDLRALEESV SNLEESLTSL
 Taq-14  AVGVGTGTAA LITGPQQLEK GLGELHAAMT EDLRALEESV SNLEESLTSL
 Taq-12  AVGVGTGTAA LITGPQQLEK GLGELHAAMT EDLRALEESV SNLEESLTSL
 Taq-24  AVGVGTGTTA LITGPQQLEK GLGELHAAMT EDLRALEESV SNLEESLTSL
  Taq-9  AVGVGTGTAA LITGPQQLEK GLGELHAAMT EDLRALEESV SNLEESLTSL
  Taq-3  AVGVGTGTAA LITGPQQLEK GLGELHAAMT EDLRALEESV SNLEESLTSL
 Taq-21  AAGVGTGTAA LVTGPQQLET GLSNLHRIVT EDLQALEESV SNLEESLTSL
 Pfu-11  AAGVETGTAA LITGPQQLEK GLSDLHRIVT EDLQALEKSV SNLEESLTSL
Pfu-251  AAGVETGTAA LITGPQQLEK GLSDLHRIVT EDLQALEKSV SNLEESLTSL
 Pfu-62  AAGVETGTAA LITGPQQLEK GLSDLHRIVT EDLQALEKSV SNLEESLTSL
 Taq-56  AAGVETGTAA LITGPQQLEK GLSDLHRIVT EDLQALEKSV SNLEESLTSL
  Pfu-3  AAGVETGTAA LITGPQQLEK GLSDLHRIVT EDLQALEKSV SNLEESLTSL
 Taq-17  AAGMGTGTAA LITGPQQLEK GLSDLHRIVM EDLQALEKSV SNLGESLTSL
Pfu-115  AAGVGTGMAA LITGPQQLEK GLSDLHRIVM ENL*ALEKSV CNQEESLTSL
  Pfu-6  AAGVGTGTAA LITGPQQLEK GLSDLHRIVT EDLQALEKSV CNQEESLTSL
  PERVA  AAGVGTGTAA LITGPQQLEK GLSNLHRIVT EDLQALEKSV SNLEESLTSL
Pfu-345  AVGVGTGTAA LITGPQQLEK GLGELHAAMT EDLRALEESV SNLEESLTSL
  PERVC  AAGVGTGTAA LVTGPQQLET GLSNLHRIVT EDLQALEKSV SNLEESLTSL
Pfu-112  ..*L*CSDWE WLQAWKQERL P*SQDHNSWR KDLVTYIEL* RKISKP*KNL
Pfu-295  ..*L*CSDWE WLQAWEQERL P*SQDHNSWR KDLVTYIEL* RKISKP*KNL
 Taq-54  ..*L*CSDWE WLQAWKQERL P*SQDHNSWR KDLVTYIEL* RKISKP*KNL
  Taq-8  ..*L*CSDWE WLQAWEQERL P*SQDHNSWR KDLVTYIEL* WKISKP*KNL
 Taq-82  ..*L*CSDWE WLQAWEQERL P*SQDHNSWR KDLVTYIEL* WKISKP*KNL
 Pfu-56  ..*L*CSDWE WLQAWEQERL P*SQDHNSWR KDLVTYIEL* WKISKP*KNL
Pfu-260  ..*L*CSD*G RPLA*EQGQL P*SQDHSS*R KDLVSYMRP* QKISEP*RSL
  Taq-1  ..*L*CSD*G RPLA*EQGQL P*SQDHSS*R KDLVSYMRP* QKISEP*RSL
  Taq-6  ..*L*CSDWE WLQAWEQERL P*SQDHNSWR KDLVTYIEL* WKISKP*KNL
Pfu-232  ..YPSCNARI RDGRWRRNRD SCPDHRTTAA RERTW*ATCG HDRRSPSLRG
 Taq-11  ..DTSCNARI GSGCRHGNRN GCPNHRTTTA GERT**PTSN CNGRSPSPRK
 Taq-10  ..DTSCNARI GRGCRRGNRN GCPNHRTTTA GERT**PTSN CNGRSPSPRK
 Taq-29  ..*L*CSD*G RPLA*EQGQL P*SQDHSS*R KDLVSYMRP* QKISEP*RSL

         551                                                 600
  PERVB  SEVVLQNRRG LDLLFLREGG LCAALKEECC FYVDHSGAIR DSMSKLRERL
 Taq-20  SEVVLQNRRG LDLLFLREGG LCAALKEECC FYVDHSGAIR DSMSKLRERL
  Pfu-1  SEVVLQNRRG LDLLFLREGG LCAALKEECC FYVDHSGAIR DSMSKLRERL
  Taq-4  SEVVLQNRRG LDLLFLREGG LCAALKEECC FYVDHSGAIR DSMSKLRERL
 Taq-14  SEVVLQNRRG LDLLFLREGG LCAALKEECC FYVDHSGAIR DSMSKLRERL
 Taq-12  SEVVLQNRRG LDLLFLREGG LCAALKEECC FYVDHSGAIR DSMSKLRERL
 Taq-24  SEVVLQNRRG LDLLFLREGG LCAALKEECC FYVDHSGAIR DSMSKLRERS
  Taq-9  SEVVLQNRRG LDLLFLREGG LCAALKEECC FYVDHSGAIR DSMSKLRERL
  Taq-3  SEVVLQKWRG LDLLFLREGG LCAALKEECC FYVDHSGAIR DSMSKLRERL
 Taq-21  SEVVLQNRRG LDLLFLREGG LCAALKEECC FYVDHSGAIR DSMSKLRERL
 Pfu-11  SEVVLQNRRG LDLLFLKEGG LCVALKEECC FYVDHSGAIR DSMSKLRERL
Pfu-251  SEVVLQNRRG LDLLFLKEGG LCVALKEECC FYVDHSGAIR DSMSKLRERL
 Pfu-62  SEVVLQNRRG LDLLFLKEGG LCVALKEECC FYVDHSGAIR DSMSKLRERL
 Taq-56  SEVVLQNRRG LDLLFLKEGG LCVALKEECC FYVDHSGANR DSMSKLRERL
  Pfu-3  SEVVLQNRRG LDLLFLKEGG LCVALKEECC FYVDHSGAIR DSMSKLRERL
```

```
 Taq-17  SEVVLQNRRG LDLLFLKEGG LCVALKEECC FYVDHSGAIR DSMSKLRERL
Pfu-115  SEVVLQNRRG LDLLFLKEGG LCVALKEECC FYVDHSGAIR DSMSKLRERL
  Pfu-6  SEVVLQNRRG LDLLFLKEGG LCVALKEECC FYVDHSGAIR DSMSKLRERL
  PERVA  SEVVLQNRRG LDLLFLKEGG LCVALKEECC FYVDHSGAIR DSMSKLRERL
Pfu-345  SEVVLQNRRG LDLLFLREGG LCAALKEECC FYVDHSGAIR DSMSKLRERL
  PERVC  SEVVLQNRRG LDLLFLKEGG LCVALKEECC FYVDHSGAIR DSMNKLRERL
Pfu-112  SVT*RNP*PP YLKWFYRTEE G*ICYF*KKV GYV*P*KKNV ASM*ITQEPS
Pfu-295  SVT*RNP*PP YLKWFYRTEE G*ICYF*KKV GYV*P*KKNV ASM*ITQEPS
 Taq-54  SVT*RNP*PP YLKWFYRTEE G*ICYF*KKV GYV*P*KKNV ASM*ITQEPS
  Taq-8  SVT*RNP*PP YLKWFYRTEG G*ICYF*KKV GYV*P*KKNV ASMWITQEPS
 Taq-82  SVT*RNP*PP YLKWFYRTEG G*ICYF*KKV GYV*P*KKNV ASMWITQEPS
 Pfu-56  SVT*RNP*PP YLKWFYRTEG G*ICYF*KKV GYV*P*KKNV ASMWITQEPS
Pfu-260  LAT*KSP*LL CLKWFYRTGG D*ICCF*EKV GYVQP*KKNV ASM*ITQEPS
  Taq-1  LAT*KSP*LL CLKWFYRTGG D*ICCF*EKV GYVQP*KKNV ASM*ITQEPS
  Taq-6  SVT*RNP*PP YLKWFYRTEG G*ICYF*KKV GYV*P*KKNV ASMWITQEPS
Pfu-232  VC*QPRRVPD FFV*SGSTEP EGIRSAVSKR RWVMCSLKRR MLLLCRSLRS
 Taq-11  ICQ*PRGIPN LLI*SGSTEQ KGVRSVISKR RWVMCSLKRR MLLLCGSLRS
 Taq-10  ICQ*PRGIPN LLI*SGSTEQ KRVRSVISKR RWVMCSLKRR MLLLCRSLRS
 Taq-29  LAT*KSP*LL CLKWFYRTGG D*ICCF*EKV GYVQP*KKNV AST*ITQEPS

         601                             631
  PERVB  ERRRREREAD QGWFEGW... .......... .
 Taq-20  ERRRREREAD QGWFEGW... .......... .
  Pfu-1  ERRRREREAD QGWFEGW... .......... .
  Taq-4  ERRRREREAD QGWFEGW... .......... .
 Taq-14  ERRRREREAD QGWFEGW... .......... .
 Taq-12  ERRRREREAD QGWFEGW... .......... .
 Taq-24  ERRRREREAD QGWFEGW... .......... .
  Taq-9  ERRRREREAD QGWFEGW... .......... .
  Taq-3  ERR*REREAE QGWFEGW... .......... .
 Taq-21  ERRRREREAD QGWFEGW... .......... .
 Pfu-11  ERRRREREAD QGWFEGW... .......... .
Pfu-251  ERHRREREAD QGWFEGW... .......... .
 Pfu-62  ERRRREREAD QGWFEGW... .......... .
 Taq-56  ERRRREREAD QGWFEGW... .......... .
  Pfu-3  ERRRREREAD QGWFEGW... .......... .
 Taq-17  ERHQREREAD QGWFEGW... .......... .
Pfu-115  ERRRREREAD QGWFEGW... .......... .
  Pfu-6  ERRRREREAD QGWFEGW... .......... .
  PERVA  ERRRREREAD QGWFEGW... .......... .
Pfu-345  ERRRREREAD QGWFEGW... .......... .
  PERVC  EKRRREKETT QGWFEGW... .......... .
Pfu-112  ETP*ASLEKG *RGVEGKERL TRGGLKDG.. .
Pfu-295  ETP*ASLEKG *RGIEGKERL TRGGLKDG.. .
 Taq-54  ETP*ASLEKG *RGVEGKERL TRGGLKDG.. .
  Taq-8  ETP*ASLEKG *RGVEGKERL TRGGLKDG.. .
 Taq-82  ETP*ASLEKG *RGIKGKERL TRGGLKDG.. .
 Pfu-56  ETP*ASLEKG *RGIKGKERL TRGGLKDG.. .
Pfu-260  ETP*ASLEKG *RGVEGKERL TRGGLKDG.. .
  Taq-1  ETP*ASLEKG *RGVEGKERL TRGGLKDG.. .
  Taq-6  ETP*ASLEKG *RGIKGKERL TRGGLKDG.. .
Pfu-232  HQRLH..EQA *RKVREASKG KRG*PGVV*R M
 Taq-11  HQRLH..EQA *RKVREASKG KRG*PGVV*R M
 Taq-10  HQRLH..EQA *RKVREASKG KRG*PGVV*R M
 Taq-29  ETP*ASLEKG *RGVEGKERL TRGGLKDG.. .
```

Appendix 4.5. Chi-square test for the number of clones in each class.

Chi-Square Test

Expected counts are printed below observed counts

	Taq	Pfu	Total
A	19	95	114
	17.80	96.20	
B	43	248	291
	45.42	245.58	
Recom	2	3	5
	0.78	4.22	
Total	64	346	410

ChiSq = 0.082 + 0.015 + 0.129 + 0.024 + 1.905 + 0.352 = **2.508**
df = 2
* WARNING * 1 cells with expected counts less than 1.0
* Chisquare approximation probably invalid
2 cells with expected counts less than 5.0

1.1.1 Cumulative Distribution Function

Chisquare with 2 d.f.

x	P(X <= x)
2.5080	**0.7146**

Appendix 4.6. Chi-square homogeneity analysis for the cycle sequence comparison error rate based on 8 duplicate sequences from 4 clones. The non-significant result justifies pooling of the data to estimate sequence comparison error rate (28/7,020).

```
Expected counts are printed below observed counts

            C1       C2    Total
    1        5      895      900
          3.59   896.41

    2        7      893      900
          3.59   896.41

    3        4      896      900
          3.59   896.41

    4        3      647      650
          2.59   647.41

    5        3      997     1000
          3.99   996.01

    6        4      896      900
          3.59   896.41

    7        1      899      900
          3.59   896.41

    8        1      869      870
          3.47   866.53

Total       28     6992     7020

ChiSq =  0.554 +  0.002 +
         3.240 +  0.013 +
         0.047 +  0.000 +
         0.064 +  0.000 +
         0.245 +  0.001 +
         0.047 +  0.000 +
         1.868 +  0.007 +
         1.758 +  0.007 = 7.855
df = 7, p = 0.347
8 cells with expected counts less than 5.0
```

Appendix 5.1. Hybridisation data for animals 115, 167, and the pooled results across two animals for the PERV-A and PERV-B probes at each chromosomal location.

PERV-A, animal 115

location	grains	location	grains	location	grains
1p2.2	1	7p1.3	1	13q3.6	1
1p1.1	4	7p1.3	15	13q4.1	8
1q1.2	16	7p1.2	1	13q4.1	46
1q1.4	1	7p1.2	3	13q4.3	6
1q1.5	1	7p1.1	1	13q4.7	1
1q1.7	4	7q1.1	3	13q4.8	1
1q1.8	11	7q1.4	3	14q1.2	1
1q2.1	1	7q1.5	9	14q1.3	1
1q2.3	1	7p2.6	1	14q1.5	1
1q2.6	1	8p2.3	1	15q1.4	1
2p1.4	10	8p2.2	1	15q2.1	1
2q1.2	1	8p1.1	1	15q2.5	2
2q2.1	2	8q1.2	1	15q2.6	1
2q2.5	1	8q2.3	1	16q1.3	1
3p1.5	2	8q2.5	1	16q2.1	6
3p1.4	11	9p2.2	1	17q1.2	4
3p1.1	1	9p1.1	1	17q1.4	11
3p2.6	2	9p2.3	1	17q2.1	5
4q1.5	1	9q2.4	15	17q2.1	2
5p1.4	4	9q2.6	7	17q2.3	2
5p1.3	2	10p1.5	1	18q1.3	1
5p1.2	5	10p1.4	2	18q2.2	1
5p1.1	8	10p1.1	1	Xp2.4	1
5q1.1	7	10q1.3	1	Xp2.2	2
5q1.2	5	10q1.6	1	Xp2.1	6
5q2.1	22	11p1.4	1	Xp2.1	2
5q2.2	4	11p1.3	3	Xq1.1	1
5q2.4	5	11p1.3	1	Yp1.3	1
6p1.5	1	11q1.1	1	Yp1.1	8
6p1.5	8	12p1.4	9	Yq	9
6p1.4	5	12p1.3	18	Yq	14
6p1.3	7	12p1.2	12	Yq	5
6p1.2	4	12q1.1	9		
6q2.1	1	12q1.2	7		
6q2.7	1	12q1.3	6		
6q3.4	8	12q1.4	4		
6q3.5	10	13q3.4	1		

PERV-A, animal 167

location	grains	location	grains	location	grains	location	grains
1p2.2	1	5q2.4	2	9q1.4	4	15q1.3	1
1p1.3	4	6p1.5	1	9q2.2	4	15q2.2	1
1p1.2	2	6p1.5	5	9q2.4	36	15q2.3	3
1q1.1	12	6p1.4	15	9q2.5	15	15q2.4	2
1q1.2	16	6p1.3	5	10p1.5	1	15q2.5	3
1q1.4	5	6q2.2	1	10p1.4	2	16q1.3	1
1q1.6	2	6q2.4	2	10p1.1	1	16q2.1	1
1q1.7	4	6q3.3	1	10q1.3	1	16q2.1	6
1q1.8	7	6q3.4	4	10q1.6	1	17q1.2	4
1q2.7	1	6q3.5	19	11p1.4	1	17q1.4	30
1q2.8	1	7p1.3	16	11p1.3	4	17q2.1	4
2p1.4	13	7p1.3	6	11q1.1	2	17q2.1	3
2q1.3	2	7p1.2	2	11q1.7	1	18q1.1	2
2q1.4	6	7p1.1	1	12p1.5	1	18q1.3	1
2q2.1	3	7q1.4	8	12p1.4	16	Xp2.4	1
2q2.6	2	7q1.5	27	12p1.3	32	Xp2.3	2
3p1.5	7	7q2.3	1	12p1.2	17	Xp2.2	4
3p1.4	13	7q2.5	1	12q1.1	13	Xp2.1	7
3p1.3	1	7p2.6	3	12q1.2	9	Xq1.2	2
3p1.1	2	8p2.2	3	12q1.3	2	Xq1.3	4
3q1.1	1	8p2.1	4	12q1.4	2	Xq2.1	1
3q2.2	1	8p2.1	2	12q1.5	2	Yp1.3	6
4p1.4	4	8p1.2	1	13q1.3	1	Yp1.1	22
4p1.2	1	8q1.1	1	13q1.4	1	Yq	14
4q1.3	1	8q1.2	1	13q3.1	1	Yq	1
4q1.4	1	8q2.1	3	13q3.2	1	Yq	1
4q2.3	1	8q2.2	4	13q4.1	4		
5p1.4	8	8q2.5	4	13q4.1	46		
5p1.3	13	8q2.6	1	13q4.3	8		
5p1.2	5	9p2.3	2	13q4.5	2		
5p1.1	17	9p2.1	3	14q1.1	2		
5q1.1	13	9p1.3	2	14q1.3	7		
5q1.2	19	9p1.2	1	14q1.5	4		
5q2.1	24	9p1.1	2	14q2.2	1		
5q2.2	3	9q1.2	2	14q2.5	1		

PERV-A, pooled results across two animals

location	grains	location	grains	location	grains	location	grains
1p2.2	2	5q2.4	7	9p1.2	2	14q1.5	5
1p1.3	4	6p1.5	2	9p1.1	2	14q2.2	1
1p1.2	2	6p1.5	13	9q1.2	2	14q2.5	1
1p1.1	16	6p1.4	20	9q1.4	4	15q1.3	1
1q1.2	32	6p1.3	12	9q2.3	5	15q1.4	1
1q1.4	6	6p1.2	4	9q2.4	51	15q2.1	1
1q1.5	3	6q2.1	1	9q2.5	22	15q2.2	1
1q1.7	8	6q2.2	1	10p1.5	2	15q2.4	3
1q1.8	18	6q2.4	2	10p1.4	4	15q2.5	4
1q2.1	1	6q2.7	1	10p1.1	2	15q2.6	4
1q2.3	1	6q3.3	1	10q1.3	2	16q1.3	1
1q2.6	1	6q3.4	12	10q1.6	2	16q1.3	1
1q2.7	1	6q3.5	29	11p1.4	1	16q2.1	1
1q2.8	1	7p1.3	1	11p1.3	4	16q2.1	12
2p1.4	23	7p1.3	31	11p1.3	5	17q1.2	8
2q1.2	1	7p1.2	7	11q1.1	3	17q1.4	41
2q1.3	2	7p1.1	5	11q1.7	1	17q2.1	9
2q2.1	8	7q1.1	2	12p1.5	1	17q2.1	5
2q2.1	3	7q1.3	3	12p1.4	25	17q2.3	2
2q2.5	3	7q1.4	11	12p1.3	50	18q1.1	2
3p1.5	9	7q1.5	36	12p1.2	29	18q1.3	2
3p1.4	24	7q2.3	1	12q1.1	22	18q2.2	1
3p1.3	1	7q2.5	1	12q1.2	16	Xp2.4	1
3p1.1	3	7q2.6	4	12q1.3	8	Xp2.4	3
3q1.1	1	8p2.3	1	12q1.4	6	Xp2.2	6
3q2.2	1	8p2.2	4	12q1.5	2	Xp2.1	13
3p2.6	2	8p2.1	4	13q1.3	1	Xp2.1	2
4p1.4	4	8p2.1	3	13q1.4	1	Xq1.1	1
4p1.2	1	8p1.2	1	13q3.1	1	Xq1.2	2
4q1.4	1	8q1.1	1	13q3.2	1	Xq1.3	4
4q1.5	2	8q1.2	2	13q3.4	1	Xq2.1	1
4q2.3	1	8q2.1	3	13q3.6	1	Yp1.3	7
5p1.4	12	8q2.2	4	13q4.1	12	Yp1.1	30
5p1.3	15	8q2.3	1	13q4.1	92	Yq	23
5p1.2	10	8q2.5	5	13q4.3	14	Yq	15
5p1.1	25	8q2.6	1	13q4.5	2	Yq	6
5q1.1	20	9p2.3	2	13q4.7	1		
5q1.2	24	9p2.2	1	13q4.8	1		
5q2.1	46	9p2.1	3	14q1.2	3		
5q2.2	7	9p1.3	2	14q1.3	8		

PERV-B, animal 115

location	grains	location	grains	location	grains
1p2.2	9	7q1.3	3	14q1.4	6
1p2.1	4	7q1.5	1	14q2.8	1
1p1.1	1	7q1.5	1	14q2.9	1
1q1.2	1	7q2.5	1	15q1.4	1
1q1.3	1	7q2.6	1	15q1.5	1
1q1.7	1	8p2.2	1	16q2.1	4
1q1.8	1	8p2.1	1	16q2.1	21
2p1.4	3	8p1.2	1	16q2.2	5
2q1.1	2	8q2.1	2	17q1.2	3
2q1.4	1	9p2.1	2	17q1.4	2
2q2.5	1	9q2.1	1	17q2.1	13
2q2.9	1	9q2.2	3	17q2.1	18
3p1.5	1	9q2.4	19	17q2.2	6
3p1.4	1	9q2.5	4	18q1.2	1
3p1.3	1	10p1.3	2	18q2.1	1
3p1.1	2	10p1.2	1	Xp2.3	1
3q1.1	2	10q1.3	1	Xp2.2	3
3q1.4	1	10q1.4	1	Yp1.1	1
3q2.5	1	10q1.6	1	Yq	2
3q2.6	1	11p1.3	3		
4q2.1	1	11p1.2	9		
5p1.3	1	11p1.1	2		
5p1.2	1	11q1.1	1		
5p1.1	1	11q1.3	3		
5q1.1	2	11q1.6	1		
5q1.2	7	12p1.4	2		
5q2.1	9	12p1.3	2		
6p1.5	1	12p1.2	8		
6p1.4	1	12q1.1	38		
6p1.3	1	12q1.2	11		
6q2.1	1	12q1.4	1		
6q3.1	1	12q1.5	1		
6q3.3	1	13q3.1	1		
6q3.4	1	13q4.1	1		
7p1.3	3	13q4.3	2		
7p1.3	5	13q4.8	2		
7p1.2	7	14q1.2	1		
7p1.1	3	14q1.3	32		

PERV-B, animal 167

location	grains	location	grains	location	grains	location	grains
1p2.3	4	6p1.5	1	10q1.5	3	18q1.3	1
1p2.2	22	6p1.5	2	11p1.5	1	18q2.1	1
1p2.1	4	6p1.4	3	11p1.3	1	18q2.3	1
1p1.4	2	6p1.3	1	11p1.2	17	Xp2.1	4
1p1.1	1	6p1.2	4	11p1.1	11	Xp2.3	10
1q1.6	2	6q2.8	1	11q1.3	1	Xq1.1	1
1q2.1	1	6q3.2	2	11q1.5	1	Yp1.1	2
1q2.3	1	6q3.3	1	12p1.4	1	Yq	5
2p1.6	1	6q3.4	3	12p1.3	5	Yq	2
2p1.5	2	7p1.3	1	12p1.2	12		
2p1.4	2	7p1.3	4	12q1.1	50		
2q1.1	1	7p1.2	4	12q1.3	12		
2q1.2	1	7p1.1	7	12q1.4	5		
2q1.3	1	7q1.1	8	12q1.5	1		
2q2.1	2	7q1.3	2	13q2.1	1		
2q2.2	1	7q1.5	5	13q2.2	1		
2q2.6	1	7q2.1	1	13q3.4	1		
3p1.6	4	7q2.4	1	13q4.1	1		
3p1.4	2	7q2.5	1	13q4.1	15		
3p1.3	2	8p2.3	1	13q4.3	3		
3p1.1	3	8p2.2	9	14q1.2	1		
3q1.1	1	8p2.1	1	14q1.3	33		
3q1.3	1	8p1.2	1	14q1.5	8		
3q1.4	1	8p1.1	1	14q2.6	1		
3q2.5	2	8q2.1	1	14q2.9	1		
4p1.4	2	8q2.1	4	15q1.2	1		
4p1.3	1	8q2.2	1	15q1.5	1		
4p1.2	1	8q2.3	2	15q2.3	1		
4p1.1	1	8q2.5	1	15q2.5	4		
4q2.1	1	8q2.7	1	15q2.6	1		
4q2.5	1	9p1.3	1	16q1.4	3		
5p1.4	1	9q1.4	2	16q2.1	33		
5p1.1	2	9q2.1	1	16q2.2	6		
5q1.1	6	9q2.3	7	17q1.2	6		
5q1.2	16	9q2.4	31	17q1.3	10		
5q2.1	7	9q2.5	12	17q2.1	15		
5q2.2	1	10p1.4	1	17q2.1	18		
5q2.3	3	10p1.3	1	17q2.2	2		
5q2.4	4	10p1.2	1	18q1.2	1		

PERV-B, pooled results across two animals

location	grains	location	grains	location	grains	location	grains
1p2.3	4	5p1.2	1	8q2.3	2	13q4.3	3
1p2.2	31	5p1.1	3	8q2.5	1	13q4.8	2
1p2.1	8	5q1.1	8	8q2.7	1	14q1.2	2
1p1.4	2	5q1.2	23	9p2.1	2	14q1.3	65
1p1.1	2	5q2.1	16	9p1.3	1	14q1.5	14
1q1.2	1	5q2.2	1	9q1.4	2	14q2.6	1
1q1.3	1	5q2.3	3	9q2.1	2	14q2.8	1
1q1.7	3	5q2.4	4	9q2.3	10	14q2.9	2
1q1.8	1	6p1.5	1	9q2.4	50	15q1.2	1
1q2.1	1	6p1.5	3	9q2.5	16	15q1.5	1
1q2.7	1	6p1.4	4	10p1.4	1	15q1.5	2
2p1.6	1	6p1.3	2	10p1.3	3	15q2.3	1
2p1.4	5	6p1.2	4	10p1.2	2	15q2.5	4
2p1.3	2	6q2.3	1	10q1.3	1	15q2.6	1
2q1.1	3	6q2.8	1	10q1.4	1	16q2.1	7
2q1.2	1	6q3.1	1	10q1.5	3	16q2.1	54
2q1.3	2	6q3.2	2	10q1.6	1	16q2.2	11
2q2.1	2	6q3.3	2	11p1.5	1	17q1.2	9
2q2.2	1	6q3.4	4	11p1.3	4	17q1.4	12
2q2.5	2	7p1.3	1	11p1.2	26	17q2.1	28
2q2.9	1	7p1.3	7	11p1.1	13	17q2.2	36
3p1.5	5	7p1.2	9	11q1.1	1	17q2.3	8
3p1.4	3	7p1.2	14	11q1.3	4	18q1.2	2
3p1.3	3	7p1.1	11	11q1.5	1	18q2.1	1
3p1.1	5	7q1.3	5	11q1.6	1	18q2.1	2
3q1.1	3	7q1.5	1	12p1.4	3	18q2.4	1
3q1.3	1	7q1.5	6	12p1.3	7	Xp2.1	4
3q1.4	2	7q2.1	1	12p1.2	20	xp1.3	11
3q2.5	3	7q2.4	1	12q1.1	88	Xp1.1	4
3q2.6	1	7q2.5	2	12q1.2	23	Yp1.1	3
4p1.4	2	7q2.6	1	12q1.3	5	Yp1.1	7
4p1.3	1	8p2.3	1	12q1.4	2	Yq	2
4p1.2	1	8p2.2	10	12q1.5	1		
4p1.1	1	8p2.1	2	13q2.1	1		
4q1.5	1	8p1.2	2	13q2.2	1		
4q2.1	1	8q1.1	1	13q3.1	1		
4q2.5	1	8q2.1	1	13q3.4	1		
5p1.4	1	8q2.1	6	13q4.1	2		
5p1.3	1	8q2.2	1	13q4.1	17		

Appendix 5.2. Interpreting PCR data of a PERV junction fragment in somatic cell hybrid panel.


```
 ***************   MARKER PERV-Junction         ***************
 Supposed discordant rates :    False +: 0.10,  False -: 0.10
 clone  :   0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2
            1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7
 profile :  - - - - - - - - - - - - - - - - - - - - + + - - - - -

 *** CAUTION : Low frequency of positive results  ( 0.07 )  ***

 Reliability of the following results:

 Error risk lower than 0.5%             Maximal Correlation = 1.00

 Chromosome probabilities

1  0.21E-06   2  0.12E-01   3  0.35E-10   4  0.23E-07     5  0.46E-12
6  0.18E-03   7  0.50E-07   8  0.28E-08   9  0.35E-11    10  0.31E-11
11 0.38E-13  12  0.15E-04  13  0.27E-02  14  0.14E-02    15  0.39E-06
16  0.25E-09  17  0.98E+00  18  0.20E-04   X  0.73E-17

Chrom 17 : P =0.98, Region: Proba, Correl,  R+M+ / R-M+ / R+M- / R-M-
C17A  q11-q14               0.0108  0.6782    2     0      2     23
C17B  1/2 q21               0.0972  0.8000    2     0      1     24
C17C  (1/2 q21)-q23         0.8752  1.0000    2     0      0     25
```

Appendix 6.1. (a) The sequence from Collared peccary 840 bp PCR products (clone L1), (b) GenBank search results and the best sequence alignment of this sequence show that this sequence is not a retroviral sequence.

(a)

```
CTGTTGGATACGGGGGCGCACGTTATCTCTTTGAATTAAGAGTTTTCATCTTTTCCCAAACTCTCTTTTATTGCAGAGTCTTTCTCTTA
TCCCATGGGCAAAATATGGTGGTGAGCACACTGATTCCCAGCTGAAGCTTCTGCTCAGTGTGCAGGGGCTGCCCAGCTTGCTGGCTGAT
CTGGCCAACGGACATCCAATCTTAGAAGAAAGGGGCCATCATGGATTGATTAGCGATGGCTGATGGGGGTGGCGGGCTGGGGCTAGAAG
TGAAAGGTGCTGAGCTGCCCAATCCTGATCTGGAGTCATGAGAGTGTCTTGAGCTGTAGGTTTGGTTGTATACATGAAATCACACTGAA
GTTAGGACTGCCTGTCTGTGTCTAACACTGCTGCCTCCTCCTCTTGCTCCTCCCCTCCCCTTCATCCTCTTGCAGCATAAAATAGGATG
CCCTGGCTTGAAGGAGGAGCTCAGAGCTCTGACTGTGGTCCACCACCCTGGGTGCTCCTCTTACCACTAGATGCCTTGCATTCCCAGTT
CTGAAGAGCCTGGCATCTAAAGTAAGTCTGCTTGCAGCTTCACTGATGAGTCCAGACTTCAAGCTAAATGCATCAAATGCCAAAATTAT
TGACTTGTCTAGCCTGATGCCAGGCCTTCAGTGAAAATTTCTGTTCTAATGTGTTTTTCTCAATTCATACTTACTATTTCAGTTTAATT
AAGTTTTTTGAGGACAAAACAGGAAGGGAAAGCCAAAGTAAAAGATGAGATCTCTGCTTTGAACCTGAGACGGGAAGCCTTTTTCCTT
CCTAAGAGGGAGTCATAAACCCTACGTCGACGACCTCCT
```

(b)

```
The best scores are:                                      initn init1 opt z-sc E(1074264)
    AP000330 Homo sapiens genomic DNA, chrom  (171706)  189  106  185 141.7 0.0033
    AP000171 Homo sapiens genomic DNA, chrom  (100000)  189  106  185 144.9 0.0037
    AP000056 Homo sapiens genomic DNA, chrom  (100000)  189  106  185 144.9 0.0037
    AP000124 Homo sapiens genomic DNA of 21q  (100000)  189  106  185 144.9 0.0037
    HS501N12 Homo sapiens DNA sequence from   (170952)  189  100  183 139.9 0.0041
    RNLALB01 Rat gene for alpha-lactalbumin   (3829)   82   82  172 151.8 0.04
    HS424L16 Human DNA sequence from clone 4  (107484)  228  116  164 124.7 0.046
    AC003042 Homo sapiens chromosome 17, clo  (102818)  267  144  158 119.3 0.096
    AW121956 UI-M-BH2.3-aoi-h-05-0-UI.s1 NIH  ( 421)   77   77  167 160.1 0.13
    MUSPTPT9 Mouse mRNA for protein tyrosine  (6734)  134   63  161 138.2 0.13

AP000330 Homo sapiens genomic DNA, chromosome 21q22.1,  (171706 nt)
 initn: 189 init1: 106 opt: 185 Z-score: 141.7 expect() 0.0033
 62.431% identity in 181 nt overlap

     250       260       270       280       290          300
    L1 GCGGGCTGGGGCTAGAAGTGAAAGGTGCTGAGCTGCCCAATCC---TGATCTGGAGTCAT
                                     ::::::: :: ::     :::::: ::::
AP0003 TTTCTGTCTTTCCTTTCTTCCCAGAAAAGCAGCTGCCTAAACCACAGCATCTGGTGTCA-
        101860    101870    101880    101890    101900    101910

        310        320       330        340          350       360
    L1 GAGAGTGTCT-TGAGCTGTAGGTTTGG-TTGTATACATGAAATC---ACACTGAAGTTAG
        : : : :::     :: : : :::::     :: : ::   ::   :::  : ::: ::
AP0003 -ACATTCTCTGCCCTCTTTGGCTTTGGCCAAGATGCCTGTGGTCCCAACATGGCAGTCAG
          101920    101930    101940    101950    101960    101970

              370       380       390       400       410       420
    L1 GAC-TGCCTGTCTGTGTCTAACACTGCTGCCTCCTCCTCTTGCTCCTCCCCTCCCCTTCA
        :: :: :::::::::: : :: :: :: : :::::::: : :::::  : :   :::::
AP0003 CACATGGCTGTCTGTGTTTTACTCTTCTCCTTCCTCCTCCTCCTCCT-TCTTTTTCTTCA
          101980    101990    102000    102010     102020    102030

                  430          440       450       460       470
    L1 TCCT----CTTGCAG---CATAAAATAGGATGCCCTGGCTTGAAGGAGGAGCTCAGAGCT
       :: :     : ::::   :: : ::: ::::
AP0003 TCTTTTGGATCGCAGGTACACAGAATTGGATTAATCATAAACTTTGGGTCATGCCAAACT
           102040    102050    102060    102070    102080    102090

           480       490       500       510       520       530
    L1 CTGACTGTGGTCCACCACCCTGGGTGCTCCTCTTACCACTAGATGCCTTGCATTCCCAGT

AP0003 CTCACCATGGCCCACTTCTATTTAGTTAATTATTTATAATAATGATTAAAATAATTAGTT
           102100    102110    102120    102130    102140    102150
```

Appendix 6.2. (a) The GenBank search results for clone S6 and its best alignment. The (AG)n microsatellite sequence is indicated by shading (…).

(b) The GenBank search results of clone S9 and its best alignment.

(a)

```
The best scores are:                                      initn init1 opt z-sc E(1011497)
    AC018910 Homo sapiens clone RP11-17G13,   (179941)  316  136  425 423.0 6.7e-19
    AC007590 Homo sapiens BAC 161A6, complet  (100364)  157  157  401 400.4 2.2e-17
    AP000070 Homo sapiens genomic DNA, chrom  (100000)  157  157  401 400.4 2.2e-17
    AP000070 Homo sapiens genomic DNA, chrom  (100000)  157  157  401 400.4 2.2e-17
    AC005099 Homo sapiens BAC clone RG351J01  (131611)  172   79  382 378.2 2.9e-16
    AC005099 Homo sapiens BAC clone CTA-351J  (131611)  172   79  382 378.2 2.9e-16
    U73479 Homo sapiens cosmid clone U138C3   (36029)  227  124  368 370.7 2.8e-15
    HS269M15 Human DNA sequence from clone 2  (177562)  169  169  361 353.7 4.9e-15
    HS269M15 Human DNA sequence from clone R  (177562)  169  169  361 353.7 4.9e-15
    AC007878 Homo sapiens clone NH0236P02, c  (185134)  244  104  360 352.3 5.6e-15
    AC018552 Homo sapiens chromosome 16 clon  (182714)  146  146  343 334.0 6e-14
    HS390O13 Human DNA sequence from clone 3  (144676)  378  145  340 332.1 9.6e-14
    AC007056 Homo sapiens clone from human c  (182481)  159  159  335 325.3 1.8e-13
    HS438L4 Human DNA sequence from clone 43  (87100)  157   96  336 330.8 1.9e-13
    AC003691 Human Chromosome 11q23 PAC clon  (65020)  183  151  327 322.7 7.1e-13
    AC004100 Homo sapiens chromosome 17, clo  (95604)  263  107  324 317.2 9.9e-13
    AC006328 Homo sapiens clone NH0102O05, c  (177769)  178  133  321 310.3 1.3e-12
    AC003668 Homo sapiens Xp22 Bin 95 PAC 23  (146058)  202  115  320 310.4 1.6e-12
    HS34B21 Human DNA sequence from clone 34  (154394)  269  110  317 306.8 2.3e-12
    AC016883 Homo sapiens chromosome 8 clone  (180479)  149  149  316 304.8 2.6e-12


AC018910 Homo sapiens clone RP11-17G13, WORKING DRAFT S (179941 nt)
 initn: 316 init1: 136 opt: 425 Z-score: 423.0 expect() 6.7e-19
 69.626% identity in 214 nt overlap

                                      10        20        30
Clone S6                         AGGAGGTCGTCAACGTAGCGTATCTCACAACATGGCAGC
                                  ::: : :   : : :::::: :::::::::
AC0189   GGCCTCCTCTGATGCCTTGCCCCATGGATCTCACCATGAGGCAGCTCACAGCATGGCAGC
        62790     62800     62810     62820     62830     62840

        40        50        60        70        80        90
Clone S6 TTGCCTCGCCAGAGAGAGAGACAGAGAGAGAGAGAGACAGAGAGCTAGCAAGATGGAAGC
         : :: ::  ::::: ::    : :: :  ::  ::: :::: ::: :::: :  ::::
AC0189   TGGCTTCATCAGAGTAAGCAGTAAAGTGTCAGTAAGAGAGAGTGCTGGCAA-ACAGAAGT
        62850     62860     62870     62880     62890      62900

       100       110         120           130       140       150
Clone S6 CAGTCTTTTGTAACCTAA--TCAGAAGTGACA----ACCACTTTGGCTGTGCTCTCTTTC
         :::::::::::::::::  ::::::::::::    : :::::: :::    ::: :::
AC0189   CAGTCTTTTGTAACCTAATCTCAGAAGTGACATCCCATCACTTTTGCTTCATTCTGTTTA
         62910     62920     62930     62940     62950     62960

             160       170        180       190       200       210
Clone S6 TTAGAATCGAGTCACCAGGTCCAG-CTACACTCAAGACGACCTGGGAATCACACAGGGCC
         ::: :: : :::::: :::::::: : ::::: :::  :    ::  :: ::::: :
AC0189   TTACAAGCAAGTCACTAGGTCCAGCCCACACTAAAGGGG---CGGTGATTACACAAGAAT
         62970     62980     62990     63000         63010     63020

              220       230       240       250       260       270
Clone S6 GTGACCCCCGGGAGTCACAGCGGCCACGCTCCCTCATGAATCCATCCCACAGTTAAGCTC
         ::::
AC0189   GTGATGTAGAGATACCGGAGGCAGGAATCATTGGAGACTATATCAGAAGGCTACCTATCC
            63030     63040     63050     63060     63070     63080
```

(b)

```
The best scores are:                                      initn init1 opt z-sc E(1017047)
    AC018964 Homo sapiens chromosome 2 clone  (183943)  472  254  352 316.7 5.5e-13
    AC008588 Homo sapiens chromosome 5 clone  (191447)  229  136  176 142.6 0.0026
    BGU09214 Borrelia garinii p93 gene, comp  (2082)  163   83  179 170.8 0.0064
    BBP97 B.burgdorferi gene for P97-protein  (2082)  163   83  179 170.8 0.0064
    BBK48P93 B.burgdorferi p93 gene (partial  (2081)  163   83  179 170.8 0.0064
    BBTNP100 B.burgdorferi (TN) gene for p10  (2019)  163   83  179 171.0 0.0065
    AC006288 Homo sapiens chromosome 9, clon  (220218)  172   96  166 132.0 0.0089
    AL136129 Homo sapiens chromosome 6 clone  (183474)  190   90  162 129.1 0.015
    AL137018 Homo sapiens chromosome 9 clone  (199521)   82   82  161 127.6 0.017
    AC003084 Human BAC clone RG084D04 from 7  (166219)  201   94  161 128.6 0.018
    AC021519 Homo sapiens chromosome 11 clon  (65193)  113   82  163 135.8 0.018
    AC021877 Homo sapiens clone RP11-17L8, *  (185967)  164   99  158 125.0 0.026
    AC003992 Human BAC clone RG250D13 from 7  (176109)  122   80  154 121.4 0.043
    DMC66A1 Drosophila melanogaster cosmid 6  (34127)   81   81  157 133.5 0.047

AC018964 Homo sapiens chromosome 2 clone RP11-483P24 ma (183943 nt)
 initn: 472 init1: 254 opt: 352 Z-score: 316.7 expect() 5.5e-13
 72.258% identity in 155 nt overlap

                230       240       250       260       270       280
Clone S9 TCACACTGTATGCATAGAGTGAAAAGAAAAGTAGTCTGCTGTTATGGGAAGACCAAGAAA
                                        ::::::: :: :: :::::::::::: :::
AC0189   AAAGAAAATCCCACGAGCTGTATGCTCAGTGTAGTCTACTTTTGTGGGAAGACCAACAAA
           33620     33630     33640     33650     33660     33670

                290       300       310       320       330       340
Clone S9 TCATCCTCATCAGGGTCTCAGGAAAAGCTATAATGCCACCAAAGGAGCAGCACAAAGCCA
         :: : :: :::  :::: :: : :::: :::: :::::::: :  :::::: : ::: : :
AC0189   TCCTTCTGATCCTGGTCCCATGCAAAGGTATATTGCCACCTAGAGAGCAGGA-AAATCTA
           33680     33690     33700     33710     33720      33730

                350       360       370        380        390
Clone S9 TCTGTTCCAGAGCAGGGAGCTTTCACTCAGCATATTCT-CTGATATCAA-CACTACGTCG
         ::::  :: :::   : :::::  :: ::: ::  : : :::::: ::: :: :::  :
AC0189   TCTGCCCCTGAGAGAGAAGCTTGGACCCAGAATGCTGTACTGATACCAACCAATACTGCT
            33740     33750     33760     33770     33780     33790

       400
Clone S9 ACGACCTCCT
         : :::
AC0189   ATGACTGAGGGAAGGATAGAAAATGTTCTCTCTTCTGAGACCCTCCACACACGCAATAAA
            33800     33810     33820     33830     33840     33850
```

Appendix 6.3. Comparison of the clone S5 sequences with nucleotide sequences in the GenBank database. The inquiry sequences are matched with different kinds of porcine sequences both forward sequences (a) and reverse complementary sequences (b) suggesting that the clone S5 sequences are Peccary SINE sequences.

(a)

The best scores are:		initn	init1	opt	z-sc	E(1016413)
SSC251914 Sus scrofa MHC class I SLA gen	(158063)	623	623	902	740.0	1.7e-36
SSC251829 Sus scrofa MHC class I SLA gen	(152211)	849	849	878	719.9	2.3e-35
SSRYRA S.scrofa gene for skeletal muscle	(29699)	815	623	882	730.8	2.9e-35
AF170527 Sus scrofa glycoprotein IIIa (G	(4511)	609	609	864	724.3	4.4e-34
SSU28757 Sus scrofa lysozyme gene, compl	(12975)	864	588	855	711.9	7.5e-34
SSCRCRYR1 S.scrofa gene for skeletal mus	(28080)	770	506	847	701.6	1.3e-33
SSEPSILG S.scrofa epsilon-globin gene	(2397)	789	507	855	719.6	1.5e-33
SSTMP835 S.scrofa tmp83.5 gene	(21737)	823	565	839	696.0	3.4e-33
SSC131112 Sus scrofa MHC class I SLA gen	(154867)	712	712	829	678.6	4.5e-33
SSC131112 Sus scrofa MHC class I SLA gen	(154867)	712	712	829	678.6	4.5e-33
SSRYRCRC1 S.scrofa gene for skeletal mus	(14910)	809	569	836	695.2	5.5e-33
AF036005 Sus scrofa interleukin-2 recept	(8480)	821	517	833	695.3	9.6e-33
I64617 Sequence 32 from patent US 566556	(3588)	863	623	836	701.8	9.9e-33
I87899 Sequence 32 from patent US 571681	(3588)	863	623	836	701.8	9.9e-33
I21911 Sequence 32 from patent US 552548	(3588)	863	623	836	701.8	9.9e-33
I05233 Sequence 15 from Patent EP 022249	(3588)	863	623	836	701.8	9.9e-33
SSINHBAR Porcine mRNA for inhibin beta (	(3573)	863	623	836	701.8	9.9e-33
I01838 Sequence 5 from Patent US 4798885	(1629)	863	623	836	705.4	1.4e-32
SSCFRIXWB S.scrofa coagulation factor IX	(498)	805	555	840	714.2	1.5e-32
SSCFRIXCM S.scrofa coagulation factor IX	(498)	796	546	831	706.6	3.8e-32
AF009673 Sus scrofa Na/nucleoside cotran	(3048)	596	508	823	691.6	4.3e-32
AF009673 Sus scrofa Na/nucleoside cotran	(3048)	596	508	823	691.6	4.3e-32
SSTNFAB Porcine TNF-alpha and TNF-beta g	(10240)	753	536	818	681.8	4.5e-32
PIGPRE1EC Pig EcoRI/ClaI fragment of gen	(732)	689	583	820	695.6	1.1e-31
SSU23954 Sus scrofa skeletal muscle calp	(3772)	602	518	808	677.9	2e-31
SSLILH S.scrofa gene for leydig Insulin-	(2706)	701	535	806	677.8	2.9e-31
PIGFSHBS Sus scrofa gene for follicle st	(10172)	595	595	792	659.9	7.5e-31
SSLPLRNA S.scrofa LPL mRNA for lipoprote	(2963)	507	472	792	665.6	1.2e-30
SSAPOAIG S.scrofa apoAI gene	(3641)	731	544	771	646.9	1.1e-29
SSC9912 Sus scrofa plp gene	(20957)	504	461	761	630.5	1.6e-29
SSPRE S.scrofa DNA for SINE sequence SSP	(233)	529	529	770	658.7	3.8e-29
SSIFNG S.scrofa DNA for interferon-gamma	(5568)	514	274	751	628.2	8.1e-29
SSDCAPN1 Sus scrofa domestica calpain la	(502)	696	575	760	646.8	8.2e-29
SSPRE1 Sus scrofa repetitive dna sequenc	(233)	478	478	736	630.1	1.5e-27
SSPPK S.scrofa ppk98 gene	(19298)	684	407	718	594.7	1.7e-27
PIGMHDR6 Swine MHC class I PD6-glycoprot	(3797)	574	435	717	601.3	3.7e-27
SSR236939 Sus scrofa mRNA for hypothetic	(2496)	581	336	718	604.0	4e-27
SSCAPUM2 Porcine sarcoplasmic/endoplasmi	(7510)	551	310	712	593.9	4.8e-27
AB005545 Sus scrofa mRNA for C-reactive	(1762)	490	442	713	601.4	7.8e-27
AF035405 Sus scrofa Na+/H+ exchanger (NH	(1618)	542	271	711	600.1	1e-26
SSU70881 Sus scrofa alpha(1,2)fucosyltra	(2922)	418	371	706	593.2	1.4e-26
SSU66254 Sus scrofa leptin (ob) gene, co	(5920)	438	318	698	583.2	2.4e-26
AB001831 Porcine clone P430 DNA, PRE-1 r	(507)	675	338	704	599.6	3.5e-26
PIGP450SCC Sus scrofa P-450 cholesterol	(2376)	585	462	698	587.4	3.5e-26
AB001829 Porcine clone P200 DNA, PRE-1 r	(402)	595	409	702	598.9	4.7e-26
SSR236935 Sus scrofa mRNA for hypothetic	(2290)	601	458	681	573.3	2.2e-25
PIGAPAI Pig apolipoprotein A-1 gene, com	(3333)	506	392	671	563.1	5.6e-25
SSC6076 Sus scrofa HSL gene, exons 6 to	(4911)	469	327	663	554.6	1.1e-24
SSU27706 Sus scrofa P-glycoprotein class	(986)	630	340	635	538.4	4.5e-23
AB001830 Porcine clone P252 DNA, PRE-1 r	(499)	459	425	634	540.7	6.7e-23

(b)

```
The best scores are:                                      initn init1 opt z-sc E(1014421)
    SSRYRA S.scrofa gene for skeletal muscle  (29699)  614  614  937 890.2 3.9e-44
    SSC131112 Sus scrofa MHC class I SLA gen  (154867)  629  629  921 865.4 1.8e-43
    SSC131112 Sus scrofa MHC class I SLA gen  (154867)  629  629  921 865.4 1.8e-43
    SSC251914 Sus scrofa MHC class I SLA gen  (158063)  611  611  918 862.3 2.6e-43
    SSC251829 Sus scrofa MHC class I SLA gen  (152211)  620  620  879 824.7 3.4e-41
    SSC237667 Sus scrofa SPP1 gene, exons 6   (1295)  835  602  883 855.3 7.8e-41
    SSC237667 Sus scrofa SPP1 gene, exons 6   (1295)  835  602  883 855.3 7.8e-41
    PIGB2AR Sus scrofa beta-2-adrenergic rec  (5288)  739  360  850 815.3 3.2e-39
    AF036005 Sus scrofa interleukin-2 recept  (8480)  796  602  833 796.2 2.3e-38
    SSCRCRYR1 S.scrofa gene for skeletal mus  (28080)  807  535  830 786.5 2.4e-38
    SSAJ5412 Sus scrofa DNA for endogenous r  ( 574)  825  553  831 809.3 6.4e-38
    SSTMP835 S.scrofa tmp83.5 gene            (21737)  553  553  817 775.3 1.3e-37
    PIGPRE1H Pig HincII fragment of genomic   (2161)  772  531  817 788.3 2.5e-37
    PIGCYP1 Sus scrofa 17-alpha-hydroxylase   (1759)  682  413  811 783.6 5.6e-37
    SSRYRCRC1 S.scrofa gene for skeletal mus  (14910)  740  575  795 756.1 2.3e-36
    AB018743 Sus scrofa mRNA for 25-hydroxyv  (2316)  818  593  778 750.0 3.2e-35
    SSY16039 Sus scrofa A-FABP gene for fatt  (8144)  667  492  775 740.0 3.2e-35
    SSY16039 Sus Scrofa A-FABP gene           (8144)  667  492  775 740.0 3.2e-35
    PIGUTBIND Sus scrofa uteroferrin gene, 5  (1958)  565  535  773 746.1 6.2e-35
    SSBAT1G S.scrofa BAT1 gene                (10674)  443  443  734 698.7 5e-33
    AB003363 Sus scrofa S100C gene, complete  (7650)  571  465  694 661.7 8e-31
    SSU28757 Sus scrofa lysozyme gene, compl  (12975)  553  463  688 652.9 1.5e-30
    SSAJ3752 Sus scrofa V-ATPase gene, exon   (2091)  391  391  683 658.3 4.5e-30
    SSC9912 Sus scrofa plp gene               (20957)  491  430  676 638.6 5.7e-30
    PIGAPOLIP Sus scrofa apolipoprotein B ge  (10686)  462  275  669 635.5 1.6e-29
    SSMSAT27 Sus scrofa microsatellite DNA i  ( 499)  604  464  669 652.7 3.9e-29
    PIGAPOB02 Pig apolipoprotein B gene (Lpb  (7117)  453  266  660 629.1 5.6e-29
    PIGAPOLPB7 Sus scrofa apolipoprotein B (  (7117)  453  266  660 629.1 5.6e-29
    SSSP835 S.scrofa sp83.5 mRNA              (2374)  594  508  657 632.3 1.1e-28
    AW315575 13018 MARC 2PIG Sus scrofa cDNA  ( 407)  437  437  658 643.2 1.6e-28
    SSFABP2 S.scrofa H-FABP protein, exons 2  (3344)  479  278  651 624.6 2.1e-28
    AW312312 4010 MARC 1PIG Sus scrofa cDNA   ( 341)  434  434  653 639.3 3.2e-28
    SSU00793 Sus scrofa POU-domain protein (  (2695)  584  584  643 618.0 6.2e-28
    SSGPIE8 S.scrofa gpi gene for glucosepho  ( 790)  493  398  642 623.9 9.8e-28
    SSJ001202 Sus scrofa mRNA for apoptosis-  (2600)  451  451  639 614.3 1e-27
    AW314207 10134 MARC 2PIG Sus scrofa cDNA  ( 348)  426  386  628 614.9 7.1e-27
    SSR236928 Sus scrofa mRNA for hypothetic  (1452)  614  324  621 600.1 1.1e-26
    AW307624 1706 MARC 1PIG Sus scrofa cDNA   ( 412)  482  313  620 606.2 1.8e-26
    AB003281 Phacochoerus aethiopicus gene f  (1128)  445  275  603 584.0 1.1e-25
    AB003283 Pecari tajacu gene for trappin,  (1169)  412  357  602 582.8 1.3e-25
    SSR236936 Sus scrofa mRNA for hypothetic  (1576)  451  356  596 575.3 2.5e-25
    AF021874 Sus scrofa carboxypeptidase E g  (1034)  497  305  585 567.0 1.1e-24
    AW308493 3630 MARC 1PIG Sus scrofa cDNA   ( 261)  662  466  581 570.8 2.7e-24
    AB003282 Phacochoerus aethiopicus gene f  (1325)  398  288  576 556.9 3.2e-24
    SSR236939 Sus scrofa mRNA for hypothetic  (2496)  536  328  573 550.4 3.9e-24
    SSU96150 Sus scrofa tear lipocalin/von E  (4631)  465  353  571 545.0 4.2e-24
    SSU14331 Sus scrofa myogenin gene, compl  (6511)  457  284  563 535.3 1e-23
    PIGWAPC Pig DNA for elafin family member  (3670)  382  294  557 532.7 2.5e-23
    AB003285 Sus scrofa gene for elafin homo  (1054)  393  232  558 540.7 3.2e-23
    AW307727 1643 MARC 1PIG Sus scrofa cDNA   ( 339)  502  408  559 548.0 3.9e-23
```