Karyotype Evolution:

Evolutionary Chromosome Rearrangements in

Material Homologous to Human Chromosome

22q Studied in the Dog and Gibbon

by

Cordelia F. Langford

Thesis submitted for the degree of

Doctor of Philosophy

Cordelia Langford

This thesis is dedicated to my husband, Gavin,
to my parents, Edward and Patricia,
and to my sister, Deborah

Cordelia Langford

Abstract

Following divergence from the common ancestor, mammalian karyotype evolution has been an ongoing process, by which chromosomes have been structurally and numerically reorganised by various intra- and inter-chromosomal rearrangements. Each mammalian species has a unique karyotype, but despite millions of years of divergent evolution, a balance has occurred between karyotype diversity and conservation. As a result, each mammalian species has a unique arrangement of homologous chromosome blocks known as evolutionarily conserved chromosome segments (ECCSs) within their karyotype.

The study of karyotype evolution requires the definition of ECCSs and the junctions between them, which can lead to an understanding of the underlying mechanisms, which drive evolutionary rearrangements. It is important for our understanding of phylogeny as well as normal and abnormal chromosome structure and sequence organisation in the mammals.

There are several different possible approaches to generating comparative maps of ECCSs, such as comparative sequence analysis. But, in the absence of genome sequence, alternative approaches are required, such as comparative chromosome painting (zoo-FISH), high-resolution cross-species FISH, and cloning and sequencing.

The aim of the work towards this thesis was to study evolutionary chromosome rearrangements involving material homologous to human chromosome 22 in the dog and gibbon. The dog is distantly related to the human and, due to its complex karyotype, had not previously been included in evolutionary studies. For the purposes of studying the dog, it was necessary to produce a standard karyotype, which was a useful outcome for the canine genetics research community. The gibbon is the primate most closely related to the human, which has material homologous to human chromosome 22 in two ECCSs within its karyotype.

Cordelia Langford

The two species were studied by reciprocal heterologous chromosome painting and cross-species FISH. The gibbon rearrangement was ultimately studied at the sequence level for a detailed analysis of the sequences surrounding the rearrangement breakpoint junctions involving material homologous to human chromosome 22q.

Cordelia Langford

Acknowledgements

Cordelia Langford

Table of Contents   Page

Chapter One: Introduction

Chapter Two: Materials and Methods

Cordelia Langford

Chapter Three: Reciprocal Zoo-FISH Analysis

Chapter Four: FISH analysis of the Gibbon and Dog

Chapter Five: Cloning the homologous block rearrangement breakpoints

Cordelia Langford

Chapter Six: Sequence analysis of the evolutionary rearrangement points

Chapter Seven: Discussion

Cordelia Langford

Glossary of Abbreviations

1^0	primary
2^0	secondary
BAC	bacterial artificial chromosome
BLAST	Basic Local Alignment Search Tool
BrdU	5-bromodeoxyuridine
BSA	bovine serum albumen
CCD	charge-coupled device
CFA	*Canis familiaris*
CIAP	calf intestinal alkaline phosphatase
CISS	chromosomal *in situ* suppression (hybridisation)
C_0t	concentration $\times$ time (DNA reassociation kinetics)
cv	coefficient of variation
DAPI	4', 6-diamidino-2-phenylindole
dATP	2' deoxyadenosine 5'-triphosphate
dCTP	2' deoxycytidine 5'-triphosphate
dGTP	2' deoxyguanosine 5'-triphosphate
DMEM	Dulbecco's Minimal Essential Medium
DMSO	dimethyl sulphoxide
dNTP	any deoxynucleoside triphosphate, or a mixture of all four
dTTP	2' deoxythymidine 5'-triphosphate
dUTP	2' deoxyuridine 5'-triphosphate
DNA	deoxyribonucleic acid
DOP-PCR	degenerate oligonucleotide-primed polymerase chain reaction
DTT	dithiothreitol
EBV	Epstein-Barr virus
ECACC	European Collection of Animal Cell Cultures
EDTA	ethylenediamine tetraacetic acid (disodium salt)
EGTA	ethylene glycol-*bis*(β-aminoethyl ether)-N,N,N',N'-tetraacetic acid

FBS	foetal bovine serum
FISH	fluorescence *in situ* hybridisation
FITC	fluorescein isothiocyanate
HSA	*Homo sapiens*
HSRE	high salt restriction endonuclease buffer
HSY	*Hylobates syndactylus*
kb	kilobase (DNA)
LINE	long interspersed nuclear element
LSRE	low salt restriction endonuclease buffer
Mb	megabase (DNA)
NTA	nitrilotriacetic acid
OMIM	Online Mendelian Inheritance in Man
o/n	overnight
PAC	P1-derived artificial chromosome
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PHA	phytohaemagglutinin
PMSF	phenylmethylsulfonyl fluoride
rpm	revolutions per minute
RPMI	Roswell Park Memorial Institute (tissue culture medium)
SAM	*S*-adenosyl methionine
SDS	sodium dodecyl sulphate
SINE	short interspersed nuclear element
SSC	standard saline citrate
STS	sequence tagged site
TAPS	N-tris(hydroxymethyl)methyl-3-amino-propanesulphonic acid
Tris	tris(hydroxymethyl)aminomethane
U	units
UV	ultraviolet
v/v	volume/volume

Cordelia Langford

w/v   weight/volume

Cordelia Langford

Publications

Parts of the work presented in this thesis have appeared previously in the following publications:

Langford, C. F., Fischer, P. E., Binns, M.M., *et al.* (1996) Chromosome-specific paints from a high-resolution flow karyotype of the dog. *Chromosome Res* 4(2):115-23

Breen, M., Bullerdiek, J., Langford, C. F. (1999a) The DAPI banded karyotype of the domestic dog *Canis familiaris* generated using chromosome-specific paint probes, *Chromosome Res* 7:401-406

Breen, M., Thomas, R., Binns, M. M., Carter, N. P., Langford, C. F. (1999b) Reciprocal chromosome painting reveals detailed regions of conserved synteny between the karyotypes of the domestic dog (Canis familiaris) and human. *Genomics*. 61(2):145-55

Langford, C.F. and Breen, M**.** (2003). Chromosome Structure and Function - Comparative Cytogenetics. In: Nature Encyclopedia of the Human Genome. Macmillan Publishers Ltd. Nature Publishing Group, Hampshire, U.K. www.ehgonline.net

CHAPTER 1

Introduction

1.1 Evolution and speciation

1.1.1 The Class Mammalia

1.2 Mammalian Genomes

1.2.1 Chromosome Structure

1.2.2 Sequence Architecture

1.2.3 The Karyotype

1.3 Karytoype Evolution

1.3.1 Chromosome Rearrangements

1.3.2 Phenotypic Effects of Germline Chromosome Rearrangements

1.4 Methods of Studying Karyotype Evolution

1.4.1 Comparative Banding

1.4.2 Comparative Genome Mapping

1.5 Approaches for Constructing Comparative Maps

1.5.1 Genetic linkage analysis

1.5.2 Somatic cell hybrid (SCH) analysis

1.5.3 Radiation Hybrid (RH) analysis

1.5.4 In situ hybridisation analysis

    1.5.4.1 Comparative FISH mapping

    1.5.4.2 Comparative Chromosome Painting and Associated Techniques

1.6 Zoo-FISH

1.6.1 Limitations of zoo-FISH Using DNA from Chromosome-Specific Plasmid Libraries

1.6.2 Zoo-FISH Using DOP-PCR Generated Chromosome-Specific Paints

1.7 Patterns of Comparative Karyotype Organisation

1.8 Defining ECCS Boundaries

1.9 Aims of this thesis

1.1 Evolution and speciation

The world contains a rich diversity of species adapted to their environment and sharing genetic and phenotypic characteristics. In most cases the members of each species are reproductively isolated from the members of other species. It has become widely accepted that the characters of organisms are variable and that diversity and adaptability develop progressively with time by a dynamic process termed *evolution*. Darwin initiated the view that evolution is driven by natural selection (Darwin, 1859), and the evolution of a new species results from the proliferation of hereditary mutants, leading to changes in allele frequencies and chromosome combinations in populations over time. The accumulation of genetic and phenotypic differences in sexually reproducing populations results in reproductive isolation and, consequently, speciation. New species, thus, possess inherited variants of genes not found in their ancestors.

1.1.1 The Class Mammalia

Mammals are homoeothermic vertebrates with hair or fur, and the females secrete milk for the nourishment of their young. Mammals diverged from a branch of reptiles (the synapsids) during the Jurassic period approximately 200 million years ago. It is believed that the abrupt extinction of the dinosaurs during the Cretaceous period facilitated the rapid adaptive radiation of the mammals (Novacek, 1992). Fossil records suggest that tens of thousands of mammalian species have emerged, diverged and disappeared in this time interval, and it is difficult to determine accurately the precise sequence of their divergence. There are more than 5,000 extant mammalian genera, distributed in 425 families and 46 orders within the three major infraclasses: the Protheria (egg-laying monotremes (platypus and echidna)), Metatheria (the marsupials) and the Eutheria (placental mammals). The Eutheria and Metatheria diverged from a rat-sized insectivorous common ancestor about 130 million years ago, whereas the Protheria diverged about 180 million years ago. A summary of mammalian phylogeny is presented in figure 1.1.

A Summary of Mammalian Phylogeny
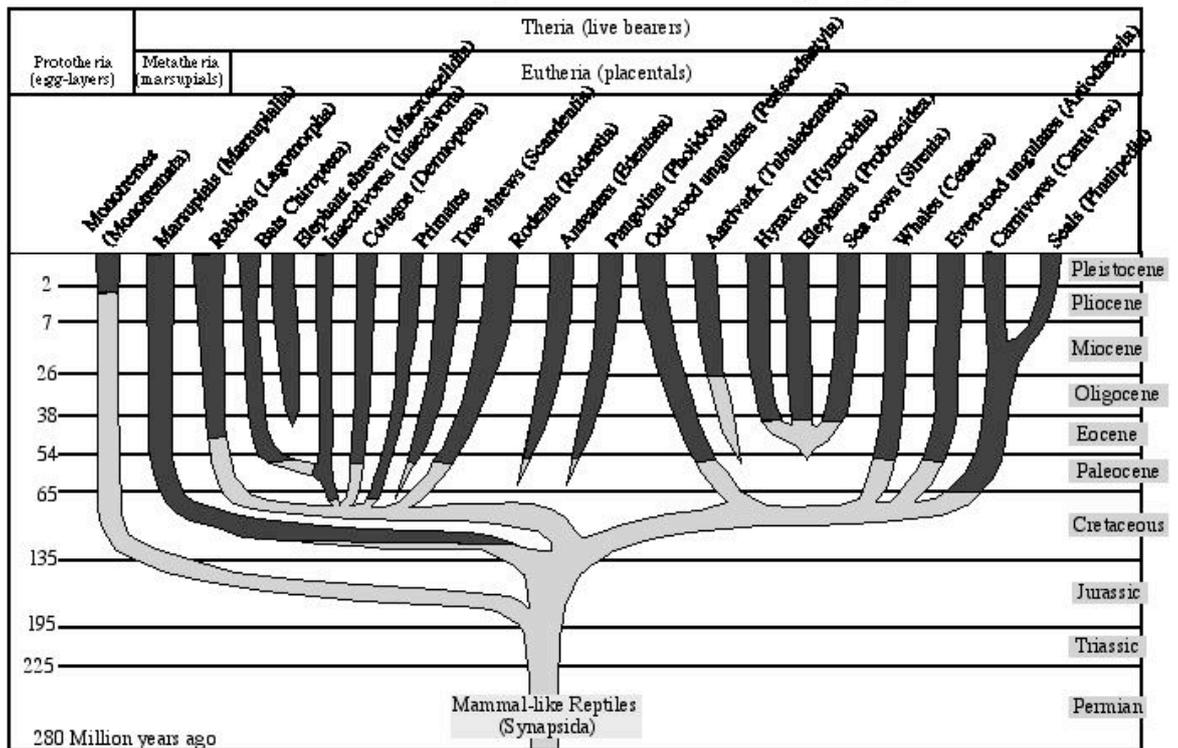
Figure 1.1 The divergent relationship between the Protheria, the Metatheria and the

Eutheria is shown along the horizontal axis in the context of geological era and timescale

(on the vertical). Reproduced from

http://www.qmw.ac.uk/~ugbt991/mammals/week6slides/sld002.htm

1.2 Mammalian Genomes

Despite millions of years of divergent evolution, mammalian genomes appear to be highly

conserved across the extant genera, which have been studied. The physical size of the

haploid genome is approximately 3,000 million base pairs (megabase pairs, Mb), and the

number of coding genes has been estimated to be in the region of 30,000 (IHGSC, 2001).

The mammalian genome is divided up and organised into chromosomes, and there are

differences between species in the number of chromosomes they posses.

1.2.1 Chromosome Structure

In diploid organisms (such as mammals) there are two copies of each chromosome type, one inherited maternally and the other inherited paternally (except for the sex chromosomes in males, where a Y chromosome is inherited from the father and an X from the mother). A typical human cell contains 46 chromosomes, 22 pairs of autosomes (non-sex chromosomes) and two sex chromosomes (Franke, 1981). Each chromosome is a single DNA molecule packaged in a protein scaffold and contains a centromere (to attach the DNA to the mitotic spindle during cell division), replication origins and a telomere located at each end of the linear molecule. Stretches of double-helical DNA wrap around associated histone proteins to form regularly repeating nucleosome "beads-on-a-string" units of chromatin (illustrated in figure 1.2). Chromatin fibres (11 nm in diameter) are packed and coiled together into a fibre 30 nm in diameter. The 30-nm fibres are also elaborately folded and organised by other non-histone proteins into a series of looped domains. Each loop contains 20,000-100,000 nucleotide pairs of double-stranded DNA extending up to approximately 300 nm in diameter. During cell division, the chromatin further condenses into microscopically distinct chromosomes.

short region of
DNA double helix   2 nm

"beads-on-a-string"
form of chromatin   11 nm

30-nm chromatin
fiber of packed
nucleosomes   30 nm

section of
chromosome in an
extended form   300 nm

condensed section
of chromosome   700 nm

                                     centromere
entire
mitotic   1400 nm
chromosome
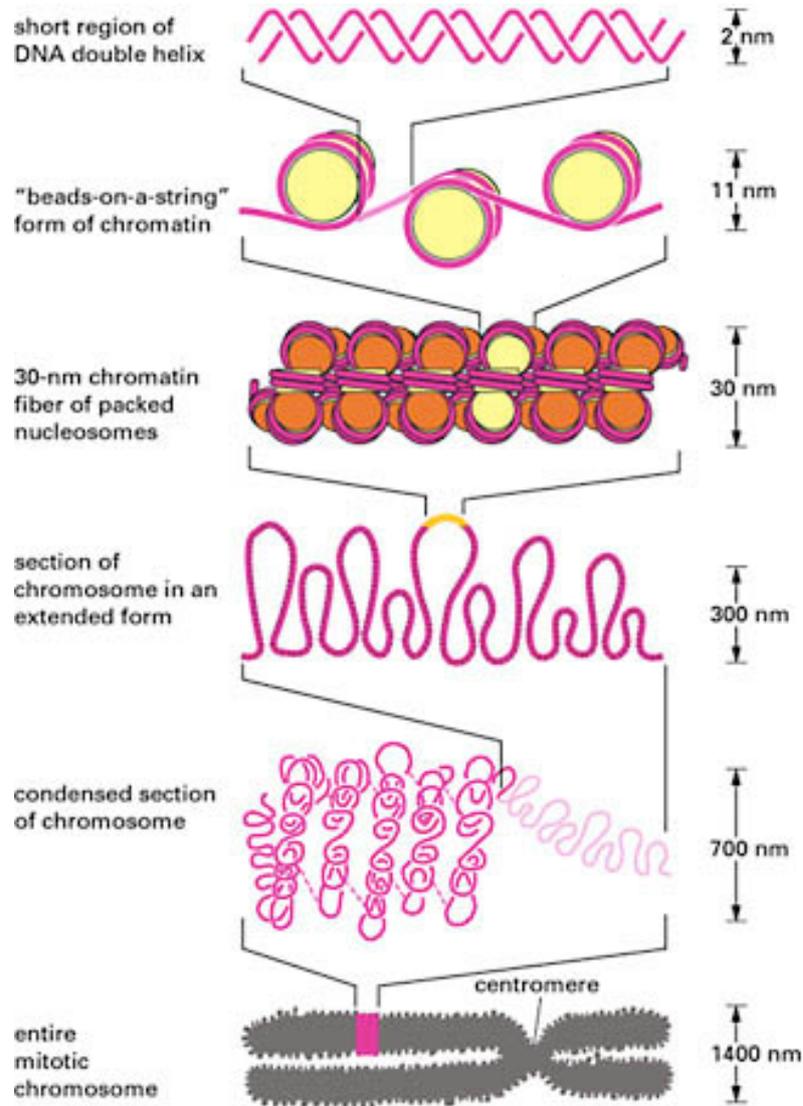
Figure 1.2 Schematic illustrating some of the orders of chromatin packing thought to give rise to the highly condensed mitotic chromosome. Reproduced from Alberts, Bray, Johnson, Lewis, Raff, Roberts and Walter, 1998 © Garland Publishing http://www.garlandscience.com/ECB/about.html

After duplication, each chromosome consists of two sister chromatids and the looped domains of each chromatid are further coiled and supercoiled into condensed sections approximately 700 nm in diameter. Although the lengths of chromosomes can vary, an entire mammalian metaphase chromosome (consisting of two sister chromatids joined at the centromere) is approximately 1.5 μm wide and up to 10 μm long.

During mitosis two daughter cells are produced from a single parent cell, each with a

diploid set of chromosomes. During the production of germ cells, single parent cells undergo meiotic division, which produces four haploid daughter cells. The processes of cell division result in the sister chromatids of each chromosome moving apart to opposite spindle poles to become daughter chromosomes. The movements depend on the attachment of spindle microtubules to the centromere. Metaphase chromosomes can be visualised microscopically and the chromosomes are distinguished and classified by their size and by the position of the centromere (Figure 1.3). Thus metacentric chromosomes have two distinct chromosome arms with a centromere midway between the ends. Submetacentric chromosomes have the centromere somewhat closer to one end. Acrocentric chromosomes have either a single arm or have the centromere positioned very close to one end. The short and long arms are referred to as the p arm and the q arm, respectively (Franke, 1981)

Figure 1.3 The ordered G-banded chromosomes of a male human cell.

In order to replicate, a DNA molecule requires a specific nucleotide sequence to act as a DNA replication origin recognised by DNA polymerase (Abdurashidova, *et al.* 2003). The replication origins, which consist of core consensus sequences several nucleotides in

length, are spaced at intervals of several thousand nucleotide pairs. The ends of chromosomes have simple repeating sequences, telomeres, that provide long-term stability (Pathak, *et al.* 2002). Without telomeres, each replication cycle of the chromosome would cause the DNA strand to become shorter. However, to prevent this, telomere sequences are extended periodically by an enzyme called telomerase. Such additions compensate for the loss of a few nucleotides of telomeric DNA in each replication cycle and help to ensure that chromosome ends do not gradually erode on replication.

1.2.2 Sequence Architecture

In the human, coding sequences comprise approximately 2% of the genome, whereas repeat sequences account for at least 50% (IHGSC, 2001). Repeat sequences also account for between 35% and 55% of other mammalian genomes. The repeats provide a palaeontological record and their inheritance patterns hold clues about evolutionary events and forces. It is possible to study groups of repeats and to follow their fates in different regions of the genome and in different species. Some repeats in different parts of the genome have recombined and fostered genome rearrangements in germlines, thus reshaping the genome and creating new genes. Although most is known about repeat elements in the human, a certain amount of information has also been generated about repeats in other mammals (for example, Demattei, *et al.* 2000). Generally, repetitive sequences can be divided into five classes:

  A. Transposon-derived interspersed repeats;

  B. Inactive partially retroposed copies of cellular genes (including protein-coding genes and small structural RNAs) usually referred to as processed pseudogenes;

  C. Simple sequence repeats, consisting of direct repetitions of relatively short k-mers such as $(A)_n$, $(CA)_n$ or $(CCG)_n$;

  D. Segmental duplications, (Low-copy repeats - LCRs) consisting of blocks of around 10-300 kb that have been copied from one region of the genome into another region;

  E. Blocks of tandemly repeated sequences (with a variation in the repeat unit up to

several thousand bases) such as those located at centromeres, telomeres, the short arms of acrocentric chromosomes and ribosomal gene clusters.

A. <u>Transposon-derived interspersed repeats</u>

Transposons are segments of DNA that can move around to different positions in the genome of a single cell. In the process of moving, they may cause mutations in several ways:

1. If a transposon inserts itself into a functional gene, it will probably destroy or alter the gene's activity.

2. Faulty repair at the gap left at the old site (by a transposon) can lead to mutation there.

3. The presence of a string of identical repeated sequences presents a problem for precise pairing during meiosis. This can lead to unequal crossing over and cause duplications and deletions.

Most of the repetitive human sequence is derived from transposable elements, and in fact 45% of the genome sequence has been identified as such (IHGSC, 2001). In mammals there are four main types of transposable element, which can be divided into two classes: DNA transposons (one type, consisting only of DNA that moves directly from place to place) and retrotransposons (three types, which first transcribe the DNA into RNA and then use reverse transcriptase to make a DNA copy of the RNA to insert in a new location (Prak and Kazazian, 2000).

A.1 DNA transposons

DNA transposons move by excision from the original location and integration into a new location in the genome without an RNA intermediate. This process requires a transposase enzyme that is encoded by some transposons. The main characteristics of DNA transposons are the Terminal Inverted Repeats (TIRs) at both ends, which are identical sequences 10-500 bp long reading in opposite directions. The transposase recognises and binds specifically to the TIRs or a sequence of DNA that makes up the

target site. Some transposases require a specific sequence as their target site whereas others can insert the transposon anywhere in the genome. Thus, the transposase catalyses the excision and subsequent splicing of the transposable element. The DNA at the target site is cut in such a manner that over-hanging "sticky ends" are produced. After the transposon is ligated to the host DNA, the gaps (caused by the single-strand overhangs) are repaired resulting in identical short direct repeats (target site duplications) at each end of the integrated transposon. These target site duplications (illustrated in figure 1.4) are evident as repeats flanking the element (Smit and Riggs, 1996).
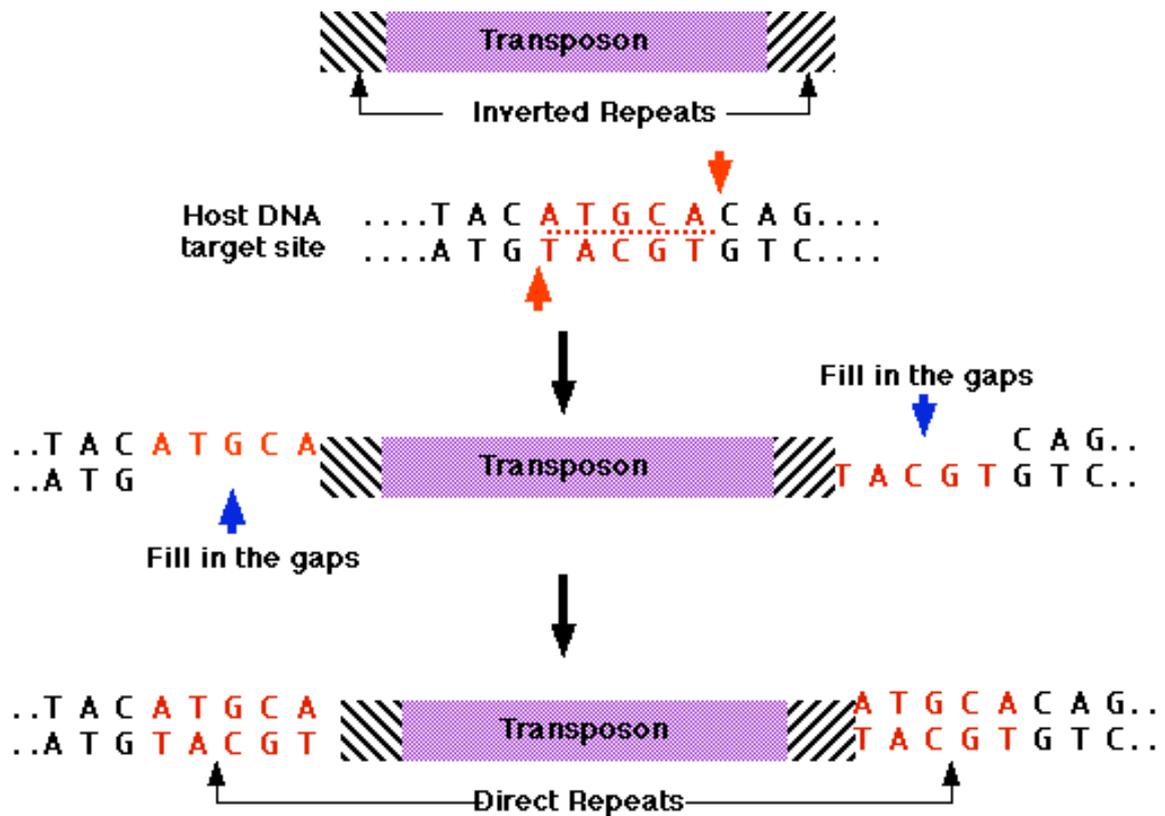
Figure 1.4 Illustration of the mechanism by which a transposon integrates into its target site. Reproduced from http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/T/.

A.2-4 Retrotransposons

Whereas transposons move by excision from the original location and ligation into the new location, retrotransposons move by the ligation of a copy of the original element. In contrast to the transposons, the duplication and transposition of retrotransposons occurs through an RNA intermediate. The original retrotransposon is maintained *in situ*, where it is transcribed. The RNA copy is then transcribed back into DNA using a reverse

transcriptase and this is integrated into a new genomic location. Many retrotransposons have long terminal repeats (LTRs) at their ends that may contain over 1000 base pairs each. Like DNA transposons, retrotransposons also generate short target-site duplications at their new insertion sites. The three types of retrotransposons are described below.

Long Interspersed Nuclear Elements (LINEs) are the most ancient repeats identified in eukaryotic genomes and the human genome contains over 500,000. LINEs are long DNA sequences that represent messenger RNAs originally transcribed by RNA polymerase II. Some LINEs encode a functional reverse transcriptase and/or endonuclease, which enable them to mobilise not only themselves, but also other retrotransposons (LINEs, Alu sequences and other SINEs, see below). Because of the mode of transposition, the number of LINEs can increase in the genome.

LINEs can be divided into three distantly related families, namely LINE1, LINE2 and LINE3. Of these only LINE1 is active in human and other mammals (IHGSC, 2001). A full length (6 kb) LINE1 element consists of a 5' untranslated region (5' UTR) that harbours an RNA polymerase II promoter and two open reading frames (ORF1 and ORF2) followed by a 3' UTR and a PolyA tail. ORF1 encodes an endonuclease, whereas ORF2 encodes a reverse transcriptase. Once a LINE1 element has been translated, the LINE RNA assembles with its own encoded proteins and moves back to the nucleus. The endonuclease makes a single-stranded DNA nick at the site of integration and the reverse transcriptase uses the nicked DNA to prime reverse transcription from the 3' end of the LINE RNA. The enzyme frequently fails to reach the 5' end, resulting in many truncated, non-functional insertions (IHGSC, 2001). In fact, the average size of a LINE-derived repeat is 900 bp. The LINE retrotransposon machinery is believed to be responsible for most reverse transcription in mammalian genomes, including the retrotransposition of the non-autonomous SINEs and the creation of processed pseudogenes (see description below).

Short Interspersed Nuclear Elements (SINEs) are short DNA sequences that range in size between 100-400 bp and represent reverse-transcribed RNA molecules originally transcribed by RNA polymerase III; that is, molecules of tRNA and 5S rRNA. SINEs do not encode any proteins and are characterised by an internal RNA polymerase III promoter that ensures transcriptional activity in new copies (Smit, 1996). These non-autonomous transposons are thought to use the LINE machinery for transposition. In most cases, the promoter regions of SINEs are derived from tRNA sequences. But the one exception is a single family of SINEs derived from the Signal Recognition Particle (SRP) component 7SL, which also happens to include the only active SINE in the human genome: the *Alu* element.

SINEs can be divided into three distinct families in the human genome: the aforementioned active Alu family and the inactive MIR and Ther/MIR3 families. MIRs (mammalian-wide interspersed repeats) are approximately 260 bp long, tRNA-derived interspersed repeats. MIRs are thought to be the most ancient mammalian SINE family and are believed to have spread through the genome prior to the Cretaceous radiation of mammals (Jurka *et al*., 1995).

The most abundant SINEs are those belonging to the Alu family, which is primate-specific but has counterparts in the genomes of several other mammals. Alus are named after the *Alu*I restriction site they carry and there are over one million copies in the human genome (Mighell, *et al.* 1997). A typical human Alu element, which consists of a 300 bp head-to-tail dimer, which appear to be reverse transcripts of 7S RNA, part of the Signal Recognition Particle (SRP). The left monomer has significant similarity with a RNA Pol III promoter; an A-rich linker connects the right and left monomers (Rogozin *et al*., 2000).

Based on the presence of diagnostic nucleotide substitutions, Alus are divided into three branches, which are further classified into sub-branches reflecting the age of individual elements from the oldest (J), to intermediate (S), to the youngest (Y) (Mighell, *et al.* 1997). The AluJ repeats are divided into the Jo and Jb sub-branches and it is estimated

that they evolved in the mammalian genome 50 to 80 million years ago. The AluS repeats are divided into the Sq, Sp, Sx, Sc, Sg and Sg1 sub-branches. It is estimated that they evolved 35 million years ago (Jurka and Milosavljevic, 1991, Mighell, *et al.* 1997). The AluY repeats (Y, Ya5, Ya8 and Yb8) probably date back 20 million years (Mighell *et al.*, 1997).

LINE elements have been proposed to be the main generators of Alu expansion (Smit, 1999). LINEs are thought to mobilise Alus because of the similarity of their target site duplications and the similarity of their insertion sites (the DNA nick for Alu insertions is probably made by LINE1 endonuclease). The "piggyback" parasitism of LINEs by SINEs remains difficult to reconcile with the observation that LINEs seem to insert preferentially into AT rich regions, whereas SINEs such as Alus accumulate in GC regions. One theory suggests that Alu elements integrate either randomly or preferentially in AT-rich regions but those that are actively transcribed under conditions of stress (and likely to reside in GC rich regions of the genome) are more likely to become fixed in the population. This explanation predicts that Alu RNA may have some advantageous function (Smit, 1999, Prak and Kazazian, 2000).

SINEs and LINEs have been found to be the cause of the mutations responsible for some cases of human genetic disease, including Haemophilia A (Factor VIII gene) and Haemophilia B (Factor IX gene), X-linked severe combined immunodeficiency (SCID, gene for part of the IL-2 receptor), predisposition to colon polyps and cancer (APC gene) and Duchenne muscular dystrophy (dystrophin gene).

Long Terminal Repeat (LTR) retrotransposons contain genes, which encode a protease, reverse transcriptase, RNAse H and integrase. They are flanked on both ends by LTRs with promoter activity. The transcript is reverse transcribed in a cytoplasmic virus-like particle, primed by a tRNA. The vertebrate-specific endogenous retroviruses (ERVs) appear to be the only LTR retrotransposons with activity in the mammalian genome. Most of the remnants of LTR retrotransposons consist only of an isolated LTR – the internal

sequence having been lost by homologous recombination between the flanking LTRs (IHGSC, 2001).

B. <u>Processed pseudogenes</u>

Pseudogenes have close sequence similarity to one or more paralogous genes but are non-functional due to the failure of either transcription or translation (Mighell *et al*., 2000). Pseudogenes arise either by retrotransposition or duplication of genomic DNA. Pseudogenes that arise by retrotransposition are called processed pseudogenes and their main characteristics include a lack of introns and 5' promoter sequences (Maestre *et al*., 1995).

C. <u>Simple sequence repeats</u>

Simple sequence repeats (SSRs) are near-perfect tandem repeats of a particular k-mer. SSRs with a short repeat unit (n = 1-13 bp) are called microsatellites, whereas those with longer repeat units (n = 14-500 bp) are called minisatellites. SSRs comprise about 3% of the human genome (IHGSC, 2001) and are thought to arise by slippage of DNA polymerase during replication.

D. <u>Segmental duplications (LCRs)</u>

Low-copy repeats (LCRs) or paralogous segmental duplications are unlike highly repetitive sequences. They are region-specific blocks of DNA ranging from 10 kb to 1.5 Mb in size with 95-97% sequence similarity. It is believed that they have arisen within the past 35-50 Myr and might have played an important role in human and great ape genome evolution by mediating chromosome rearrangements and creating novel fusion genes (Eichler, 2001, Samonte and Eichler, 2002, Inoue *et al*., 2001, Stankiewicz *et al*., 2001). Interchromosomal duplications involve blocks of sequence duplicated among non-homologous chromosomes, particularly near the centromeric and telomeric regions of human chromosomes (IHGSC, 2001). Intrachromosomal duplications involve blocks of sequence duplicated within a particular chromosome or chromosome arm.

E. <u>Blocks of localised tandem repeats</u>

Whereas the previously described repeats are generally distributed throughout the genome, certain tandem repeats have specific locations. For example, one type (α-satellites), of the Satellite repeats first observed by Sueoka (1961), are primarily found in the centromeric regions of chromosomes. The term satellite DNA was coined because the physical structure of repetitive DNA generates a buoyancy different to that of standard DNA (visualised as satellite bands after density-gradient centrifugation of genomic DNA). The amount of satellite DNA in mammalian genomes can vary widely between species. In humans less than 5% of the genome is made up of satellite DNA while in cattle up to 25% is satellite DNA and in some mammals a single type of satellite DNA sequence may occupy a whole chromosome arm. Satellite DNAs seem to have undergone comparatively rapid evolution such that there can be marked differences in the satellite DNA sequences of two closely related species (Alexandrov, *et al.* 2001).

Telomeres have unique structures that include another distinct class of short nucleotide sequences present as tandemly repeated units. Although the sequences are variable between species, the basic repeat unit in all species studied to date has the pattern 5'-$T_{1-4}A_{0-1}G_{1-8}$-3'. For example, the repeat unit in mammals is TTAGGG, which is repeated several thousand times. The number of copies of the basic repeat unit in telomeres varies between species, between chromosomes within a species, or on different homologues of the same chromosome and even on the same chromosome at different stages of the life cycle (Pathak *et al.,* 2002)

1.2.3 The Karyotype

The ordered chromosome complement of an organism is referred to as its karyotype. Chromosomes are orientated in karyotypes so that the shorter arm (p arm) is towards the top and the longer arm (q arm) is towards the bottom. Stains such as Giemsa generate specific differential patterns of dark and light bands along a chromosome's length allowing visualisation of the linear differentiation of each chromosome in a karyotype.

Giemsa (G) and reverse (R) banding are two of the most frequently used cytogenetic techniques for staining metaphase chromosomes (Craig and Bickmore, 1993). The banding patterns reflect the underlying DNA sequence organisation and condensation, and have been correlated with variations in gene density, time of replication and density of repeat sequences. For example, Giemsa-induced dark chromosome bands represent A-T rich and gene poor regions of DNA, whereas G-light bands represent G-C rich and gene rich regions of DNA (summarised in Table 1.1).

Table 1.1 The properties of Giemsa (G) and Reverse (R) bands (adapted from Gardiner, 1995)

G-bands	R-bands
Dark-staining Giemsa bands	Light-staining Giemsa bands
AT rich	GC rich
Replicate late	Replicate early
Early condensation	Late condensation
DNase insensitive	DNase sensitive
SINE/Alu poor, LINE rich	SINE/Alu rich, LINE poor
Gene poor	Gene rich

Up to 850 different G-bands can be visualised in the human karyotype. Consequently, bands can be diagnostic for each chromosome and are consistent within each typical individual of a species (see figure 1.3). The standard karyotype is often also represented by a stylised ideogram (Franke, 1994).

1.3 Karytoype Evolution

Each mammalian species studied has a unique karyotype and it has been speculated that karyotype evolution has had a role to play in the process of speciation. Mammalian karyotype evolution is an ongoing process following divergence from the common ancestral karyotype (Benton, M. J. 1990). During this time, chromosomes have been structurally and numerically reorganised by chromosome rearrangements. Despite the similarities in genome size and gene content, the diploid chromosome number in extant mammals ranges from 6 in the female Indian muntjac deer (*Muntiacus muntjak vaginalis*) to 134 in the black rhinoceros (*Diceros bicornis*) (Marshall Graves, 1998).

The number of chromosomes in karyotypes can vary enormously not just between but also within mammalian families, indicating that there is no trend of increasing or decreasing chromosome numbers during evolution. For example, although the female Indian muntjac deer has 6 chromosomes in a diploid cell, the Chinese muntjac deer (*Muntiacus muntjak reevesi*) has 46 chromosomes (Yang, *et al., 1997*). Also, within the family *Carnivora* the cat (*Felis cattus*) has 19 pairs of chromosomes whereas the dog (*Canis familiaris*) has 39 pairs in a diploid cell (Langford, *et al.,* 1996).

Mammalian karyotype evolution has proceeded to different degrees in the different groups since they diverged from the common ancestor. Thus, karyotype evolution has been rapid with extensive chromosomal rearrangements in lesser apes, rodents and equids (Ryder, *et al.,* 1978, Qumsiyeh, 1994, Andersson, *et al.,* 1996), but has been quite conservative in bovids and cetaceans (Buckland and Evans, 1978, Arnason, 1977, Gallagher and Womack, 1992, Gallagher, *et al.,* 1994). A balance has occurred between karyotype diversity and conservation between mammals. There has been ample opportunity for chromosomal rearrangements to occur during the evolution of mammalian species, but there has evidently been strong selection against total genome scrambling. As a result of karyotype evolution, each mammalian species has a unique arrangement of homologous chromosome segments known as evolutionarily conserved chromosome

segments (ECCS) (Langford and Breen, 2003).

1.3.1 Chromosome Rearrangements

Various intra- and inter-chromosomal rearrangement types (explained below and illustrated in figure 1.5) have occurred during mammalian karyotype evolution such as:

1   Intra-chromosomal inversions

2   Non-homologous inter-chromosomal translocations

3   Centromere-centromere or telomere-telomere fusions

<u>Inversions</u>

Inversions involve the detachment of a chromosome segment, its rotation through 180 degrees and its subsequent reattachment. As a result the order of the genes in that segment are reversed with respect to the rest of the chromosome. Intra-chromosomal pericentric (including the centromere) or paracentric (not including the centromere) inversions of chromosome blocks do not affect the overall size of the chromosome but they do affect the arrangement of segments within it and may well change the relative lengths of the two arms. For example, if an acrocentric chromosome acquires a pericentric inversion, it can be transformed into a metacentric chromosome, whereas if an acrocentric or metacentric chromosome acquires a paracentric inversion, the morphology of the chromosome will not be changed. Such reorganisations may increase or decrease the number of evolutionarily conserved chromosome segments in a karyotype as well as change their arrangement.

There is evidence that inversions are produced through the activity of transposable elements (Tuddenham, *et al.,* 1994). Segmental duplications occurring as a result of the insertion of transposable elements could sponsor chromosomal inversions by the process of recombination.

Types of mutation
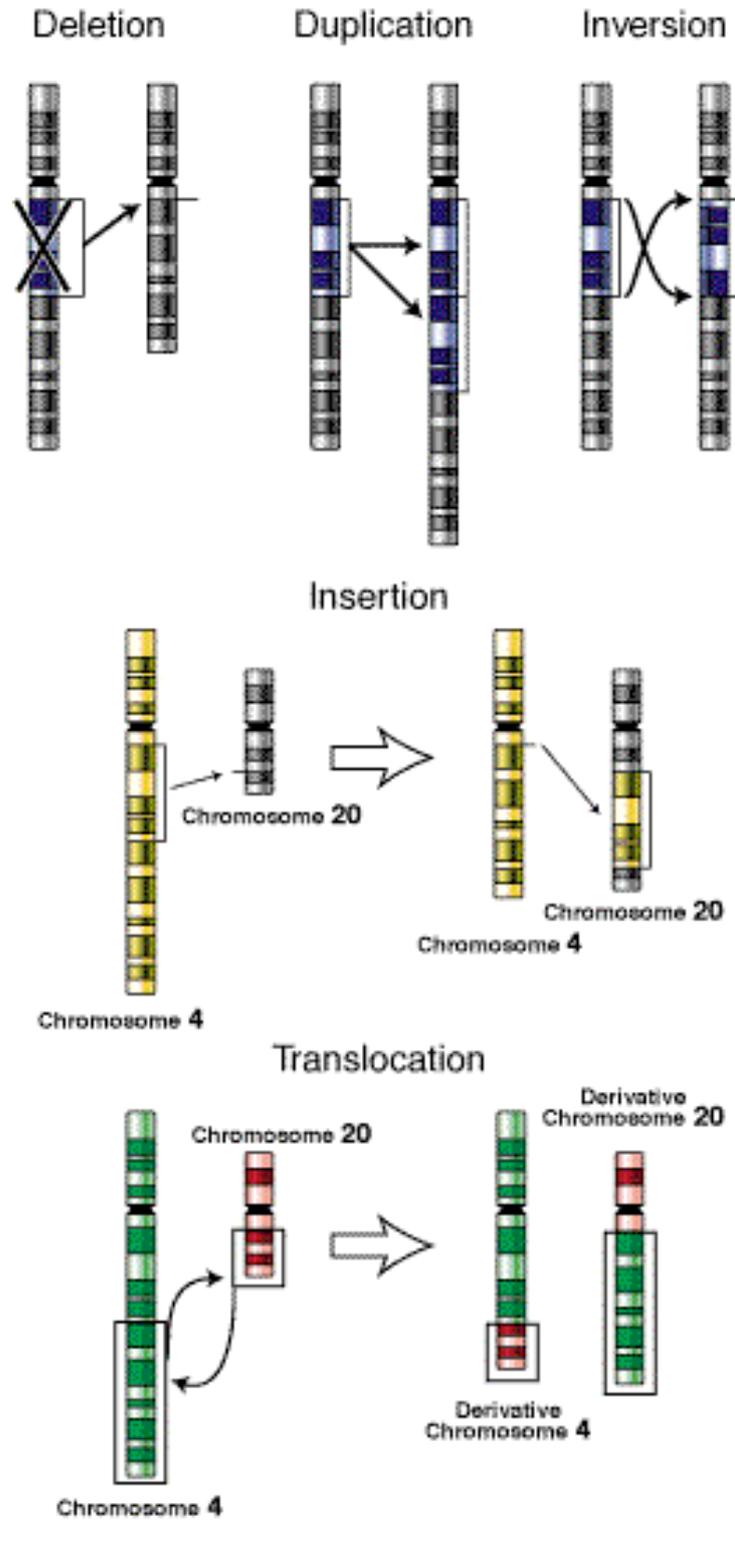
Figure 1.5 Schemtaic illustration of chromosome rearrangements and mutations

Translocations

Translocations involve the detachment of a segment from one chromosome and its

attachment to a different (non-homologous) chromosome. The significance of this is that genes from one chromosome are transferred to another chromosome and their linkage relationships are altered. When pieces of two non-homologous chromosomes are interchanged without any net loss of genetic material, the event is referred to as a reciprocal translocation. Segmental duplications caused by the activity of transposable elements may cause translocations by recombination. During meiosis, heterozygous translocated chromosomes could be expected to pair with their non-translocated homologues in a cross-like pattern. The two translocated chromosomes face each other opposite the centre of the cross, and the two non-translocated chromosomes do likewise. To maximise pairing, the translocated and non-translocated chromosomes alternate with each other, forming the arms of the cross. This configuration is diagnostic of a translocation heterozygote. Cells in which the translocated chromosomes are homozygous do not form crosses. Instead, each of the translocated chromosomes pairs smoothly with its structurally identical partner.

<u>Fusions</u>

Non-homologous chromosomes can fuse at their centromeres, creating structures called Robertsonian translocation chromosomes. For example, if two acrocentric chromosomes fuse, they will produce a metacentric chromosome; the tiny short arms of the participating chromosomes are lost in this process. Such chromosome fusions have apparently occurred quite often in the course of karyotype evolution (Ward, *et al.,* 1987). For example, G-banding studies suggest that each of the large chromosomes of the Indian muntjac deer evolved by the fusion of numerous small ancestral acrocentric chromosomes. Even though it is a common form of chromosome rearrangement in mammals, changes in chromosomal number, caused by fusions, significantly reduce the fertility of hybrid intermediates. An analysis of published data on 1170 mammalian karyotypes provided strong evidence that karyotype evolution is driven by the non-random segregation of chromosomes during female meiosis (Pardo-Manuel de Villena and Sapienza, 2001). Heterozygous carriers of Robertsonian translocations possess different numbers of centromeres on paired homologous chromosomes. The authors

proposed that, whenever this occurs, asymmetry in female meiosis and polarity of the meiotic spindle dictate that the chromosome with the greater number of centromeres will attach preferentially to the pole that is most efficient at capturing centromeres. This mechanism could explain how chromosomal variants become fixed in populations and how non-random segregation could affect karyotype evolution across a broad phylogenetic range.

Chromosomes can also fuse end-to-end (a telomere-telomere fusion) to form a structure with two centromeres. If one of these is subsequently inactivated, the chromosome fusion will be stable. Such a fusion evidently occurred in the evolution of our own species. Human chromosome 2 (_Homo sapiens_ (HSA) 2), which is metacentric, has arms that correspond to two different acrocentric chromosomes in the genomes of the great apes (chimpanzee, gorilla and orangutan). Detailed comparative cytological banding analysis indicated that the telomeres of the short arms of these two ancestral chromosomes (corresponding to chimpanzee chromosomes 12 and 13) apparently fused to create HSA2 (Yunis and Prakash, 1982).

1.3.2 Phenotypic Effects of Germline Chromosome Rearrangements

Homozygous segmental deletions that remove several genes are usually lethal because at least some of the missing genes are likely to be essential for life. Duplications, in contrast, may be viable in the homozygous condition, provided they are not too large. In the heterozygous condition, deletions and duplications could affect the phenotype by altering the dosage of groups of genes. Usually, the larger the chromosome segment involved, the greater the phenotypic effect. In fact, aneuploidy for very large chromosome segments typically is lethal. However, sometimes small heterozygous deletions or duplications can have a lethal effect, indicating that the aneuploid region contains at least one gene with a strict requirement for proper dosage. For example the loss of one copy of some developmental genes can cause severe problems because of haploinsufficiency, where a single copy of a gene cannot produce enough protein.

Inversions and translocations may also affect the phenotype. Sometimes the rearrangement breakpoints disrupt genes, rendering them mutant. The mutant phenotype appears if the rearrangements then become homozygous. It is also possible to get the mutant phenotype where the translocation is heterozygous, for example where parts of two separate genes fused to create a gene whose product is damaging and/or inappropriately expressed. In other cases, the breakpoints are not themselves disruptive, but the genes near them are put into a different chromosome environment, where they may not function normally. Such a gene is influenced by chromosome position effect. If an euchromatic gene is juxtaposed near heterochromatin, the heterochromatin could exert a repressing effect on the gene function.

1.4 Methods of Studying Karyotype Evolution

Evidence that chromosomal segments could be conserved during evolution was obtained early in the history of mammalian genetic studies. Thus, in 1927, Haldane observed that phenotypically similar traits (albinism and pink eyes) were linked together in more than one species (Haldane, 1927). Haldane recognised that, if these phenotypes in different species resulted from mutations in homologous genes, linkage between albino and pink-eyed genes may represent a chromosomal segment conserved since the divergence of lineages leading to the species.

The study of karyotype evolution requires the definition of ECCSs by comparing the karyotypes of each species being analysed.

1.4.1 Comparative Banding

Before the 1970s, most comparative karyotype studies were carried out by the painstaking analysis of banded metaphase chromosomes from each species. Almost identical cytogenetic banding patterns of the X chromosome among many mammals demonstrated that some long-term evolutionary conservation of chromosome structure had occurred (Ohno, *et al*., 1964). More recent banding studies of mammalian autosomes

illustrated ECCSs between species belonging to even distantly related groups, such as rodents and humans, (Sawyer and Hozier, 1986).

The broadest karyotype evolution study to date based on cytogenetic banding alone was carried out by Dutrilleaux on the primates from lemur to man (Dutrillaux, 1979). He was able (sometimes speculatively) to find great ape, old world and new world monkey, and lemur chromosome homologues for each human chromosome by matching up the bands with each primate species studied.

1.4.2 Comparative Genome Mapping

Since the chromosome banding studies of the 1970s, other methods have been developed to compare genomes for the identification of ECCS and to study karyotype evolution. Comparative genomic mapping studies can involve physical and genetic techniques for the molecular comparison of landmarks to map ECCS between mammalian genomes, but comparisons between the genomes of different species can only be carried out if each of them already has a "map" of comparable parameters. A physical map consists of an ordered set of clones or markers located on the genome. A genetic map defines the order and genetic separation of polymorphic landmarks (markers) by virtue of their linkage to other markers, defined indirectly through the tendency of markers to segregate together during meiosis.

Because their homology can be detected over considerable evolutionary distances, genes are reliable as anchor loci for following chromosome segments during evolution. Mapping the Haemophilia A and B genes on the X chromosome in humans and dogs provided the first comparative mapping information for loci on chromosome X (Hutt, *et al.,* 1948). However, it was only when accurate chromosome numbers became known for different species that organised comparative mapping was carried out, and in 1993, O'Brien and co-workers proposed a list of 321 evenly spaced gene loci from man and mouse, which would be suitable for comparative gene mapping in mammals and other vertebrates (O'Brien, *et al*., 1993).

Comparative mapping data are defined as either conserved syntenies or conserved linkages. Two genes are syntenic if they occur on the same chromosome of a species. Conserved synteny refers to two or more orthologous genes that are syntenic in two or more species regardless of gene order on each chromosome. Conserved linkage refers to conservation of both synteny and gene order of homologous genes between species. Large stretches of conserved synteny have been inferred by comparisons of gene maps of various mammals including human, mouse, pig and sheep. Many conserved linkages have also been found and have been used to estimate rates of chromosome rearrangement during mammalian evolution. For example, by using the average length of all conserved linkages, it was estimated that approximately 144 chromosome rearrangements (in the form of inversions or translocations) had occurred since the divergence of the lineages leading to humans and mice (Waterston, *et al.,* 2002).

In order to distinguish specific genes as the main landmarks of a comparative map (distinct from other sets of markers), the term "Type I" markers was introduced (O'Brien, *et al*., 1993). Due to their polymorphic nature, Type II markers, such as microsatellites, minisatellites, SINEs, and LINEs, were initially considered unsuitable for cross species genome comparisons. However, more recently, Type II markers have been used for comparative mapping between closely related species, for example, within the order Artiodactyla (Prakash, *et al.,* 1996).

Sequence Tagged Sites (STS) provide another set of comparable markers (approximately 25-400 bp long) used to map ECCS across genomes. When these markers originate from coding sequences, they are referred to as Expressed Sequence Tags (ESTs). STSs and ESTs can be assayed and mapped by filter hybridisation, by *in situ* hybridisation, or by using the polymerase chain reaction (PCR). Comparative anchored tagged sequences (CATs (Lyons, *et al*., 1997)) and traced orthologous amplified sequence tags (TOASTs (Jiang, *et al.,* 1998)) represent PCR primer based comparative markers, which have been assayed across species to generate information

about the correspondence between genomes.

1.5 Approaches for Constructing Comparative Maps

Several mapping approaches have contributed towards comparative genome analysis. Some techniques indicate the relative order of genes, and others assign genes to chromosomes or specific regions of chromosomes. The following five sections provide an overview of comparative mapping techniques.

1.5.1 Genetic linkage analysis

The relative order of gene loci within a genome can be represented in a linkage map. Distances between loci do not correspond to physical distances but to the frequency of recombination between the pair or set of loci investigated. The closer the loci are to each other, the greater their chances of co-segregating during meiosis. Linked loci can be assigned to a specific chromosome or 'linkage group' if one or more are physically mapped to a chromosome.

1.5.2 Somatic cell hybrid (SCH) analysis

Loci residing on the same chromosome are syntenic and a synteny map represents a list of loci, which reside on the same chromosome in a particular species. Synteny maps are built through the use of somatic cell hybrid panels constructed by fusing cell lines from two species, one of which (the donor) is the species to be mapped (Gross and Harris, 1975). During the process of the hybrid stabilising under the culture conditions, some of the donor chromosomes will be lost. Analysis of pairs of genes in a panel of SCH lines reveals concordance or discordance of their retention in the SCH, thus indicating synteny or asynteny, respectively.

The main technique now for carrying out SCH panel analysis is by PCR assays with species-specific primers. Several SCH panels are available for human and all the main livestock species and the physical assignment of genes, ESTs, microsatellites and STSs

has been rapidly progressed using the PCR approach.

Although SCH analysis shows synteny relationships between loci, it does not generate information about genetic distances. However, like linkage maps, synteny maps can play a significant role in carrying out comparisons between the genomes of different species.

1.5.3 Radiation Hybrid (RH) analysis

Radiation Hybrid mapping is a technique similar in principle to SCH mapping. However, prior to the fusion of two cell lines, the genome of the species being interrogated is exposed to high doses of X-ray irradiation, which causes chromosomal fragmentation (Thomas, *et al.,* 2001). The RH panels are analysed by PCR with species-specific primers.

As well as generating information about synteny between loci, RH mapping can also indicate the physical distance between them. The farther apart two markers are on a chromosome the greater are the chances that they will be separated onto different fragments by X-ray treatment and vice versa. RH mapping has proved to be a powerful tool for high-resolution mapping in human and mouse (Deloukas, *et al.,* 1997), farm animals such as pigs (Yerle, *et al.,* 1998) and the dog (Spriggs, *et al.,* 2003, Thomas, *et al.,* 2001). Parallel RH mapping studies (e.g. between human chromosome 17 and bovine chromosome 19) have been conducted to generate comparative mapping information (Yang, *et al.,* 1998).

1.5.4 Comparative Sequence analysis

Comparison of orthologous genes in human and mouse and their function has shown that sequence similarity across much of the coding regions of genes and some of the regulatory elements that control them has been maintained since their divergence from a common ancestor. For example, regions of conservation have been identified upstream of the SCL gene in human, mouse and chicken, and have been shown to be associated with active regulatory regions (Gottgens, *et al*., 2001). Comparative mapping and

sequencing could aid the identification of conserved genomic regions between other genera and human, which are likely to correspond to exonic or regulatory sequences. The argument for the applicability of such analyses is that functionally important sequences have been conserved at the sequence level, whereas other regions will differ as a result of accumulated mutations since their divergence. As significant amounts of the mouse genome have now been sequenced, the opportunity to use the mouse sequence as an analytical tool to study the human genome has become increasingly utilised.

1.5.5 *In situ* hybridisation analysis

Specific DNA sequences can be localised to cytogenetically prepared metaphase chromosomes by *in situ* hybridisation (ISH). In this technique, a mixture of the chromosomal DNA and the probe are denatured and then re-annealed to allow the probe to hybridise to complementary sequences in the chromosomes. After hybridisation, unbound probe is washed away and the site of hybridisation is detected and analysed microscopically. Single nucleotides can be modified and incorporated into the probe enzymatically. After hybridisation, the modified nucleotides in the probe are detected immunologically or histochemically by procedures taking less than a day to complete. The detection of non-isotopic *in situ* hybridisation probe hybridisation is direct or relies on affinity reagents, such as avidin or antibodies against the probe hapten conjugated to fluorochromes (fluorescence *in situ* hybridisation (FISH)). Currently, the most widely used non-isotopic *in situ* hybridisation systems involve nucleotides conjugated to biotin, digoxigenin or a fluorochrome (Langer-Safer, *et al*., 1982).

FISH experiments are analysed using a fluorescence microscope. In order to locate precisely the position of the hybridisation signals, the metaphase chromosomes are usually counter-stained after hybridisation with a fluorescent DNA dye such as propidium iodide (PI) or 4', 6-diamidino-2-phenylindole (DAPI). The metaphase chromosome-banding pattern generated by DAPI is analogous to G-banding. The counter-stains are not just chosen for the banding patterns they generate, but also for the wavelength of their fluorescence, which must not interfere with the specific probe signals.

FISH probes can be generated from complex sources, such as bacterial clones (Ambros *et al.,* 1986, Landegent *et al*., 1985). However, these clones inevitably contain repetitive sequences, which give rise to low overall non-specific signals on the metaphase chromosomes. Such non-specific fluorescence can potentially obscure the specific FISH signal. To overcome the problem, Landegent *et al*., (1987), developed a competitive hybridisation strategy of including unlabelled total human DNA or $C_0t=1$ DNA (containing the most abundant repetitive fraction of the genome) in the hybridisation mixture with a labelled cosmid probe. The probe mixture and the metaphase chromosomes are denatured together. Theoretically, during hybridisation, the unlabelled competitor DNA will bind to repetitive sequences in both the probe and the target chromosomes more rapidly than the repetitive elements in the probe bind to the target. Therefore, the chromosomal hybridisation of the repeat sequences present in the probe is substantially reduced and the signal from the specific probe is clear.

FISH probe signal intensification can be achieved using fluorescein isothiocyanate (FITC) conjugates in multiple amplification layers for the detection of biotinylated probes (Langer-Safer *et al*., 1982, Pinkel *et al*., 1986). The use of digital imaging systems also greatly enhances the power of FISH-mapping (Viegas-Péquignot *et al.,* 1989, Lichter *et al*., 1990, Albertson *et al*., 1991). Digital images can be taken with a fluorescence microscope equipped with a thermo-electronically cooled charge-coupled device (CCD) camera controlled by a computer. Grey scale source images are captured separately with filter sets for each fluorochrome used (including the counter-stain). Source images are saved as grey scale data files using the image capture software. The images from one metaphase can be merged and each fluorescence signal displayed in a different computer-generated pseudo-colour (Lichter, *et al.*, 1991).

1.5.4.1 <u>Comparative FISH mapping</u>

The feasibility of rapidly producing high-resolution maps of human chromosomes by FISH was reported by Lichter *et al* (1990), when they mapped 50 cosmids to human

chromosome 11 using digital imaging microscopy (Lichter, *et al.,* 1990). It was later theorised that mammalian chromosome homology maps could be refined by detailed cross-species FISH using, for example, human large-insert clones as probes on animal chromosomes (Haaf and Bray-Ward, 1996). Sub-regional clones are available for each human chromosome band. There are several hundred non-chimaeric yeast artificial chromosome (YAC) clones from the Centre d'Etude du Polymorphisme Humain (CEPH) and several thousand BAC and PAC clones from the Human Genome Mapping Project available with sequence tagged site (STS) markers, which have been FISH-mapped to human metaphase chromosomes (Haaf and Bray-Ward 1996, IHGSC, 2001).

1.5.4.2 <u>Comparative Chromosome Painting</u>

The FISH mapping of individual genes for comparative purposes is time consuming and gives only patchy information on chromosome homology between species. However, this problem can be overcome if chromosome paints are used for FISH. Chromosome paints are complex mixtures of probes, which can be synthesised from whole or parts of flow-sorted or micro-dissected chromosomes (see section on flow sorting and micro-dissection below). Chromosome paints can be used for FISH to highlight whole chromosomes or sub-regions of chromosomes (Carter, 1994) As illustrated in figure 1.6, when a whole chromosome paint (WCP) is denatured and applied to denatured metaphase spreads from the *same* species, the two copies of that chromosome type in each metaphase spread hybridise with the paint probe. On fluorescence microscopy, the regions hybridised to the paint appear as brightly coloured chromosomes in the metaphase spread.

When a WCP is hybridised to the metaphase chromosomes of a *different* mammalian species, blocks of ECCSs on various chromosomes are highlighted (see figure 1.6). Thus, comparative chromosome painting (also called heterologous chromosome painting or zoo-FISH (Scherthan *et al*. 1994), has revolutionised the field of comparative karyotype analysis because it permits the direct visualisation of regions of chromosomal homology to a resolution of 5 to 7 Mb (half a cytogenetic band) between even distantly related

mammalian species (Scherthan *et al*. 1994, Wienberg and Stanyon 1995, Andersson *et al.,* 1996, O'Brien *et al*. 1997, Wienberg *et al*. 1997, Chowdhary 1998). Furthermore, reciprocal zoo-FISH studies provide confirmation of chromosome homologies in two independent experiments as well as additional information about sub-regional homology between two species (Müller *et al.* 1997), (see figure 1.6).

Figure 1.6 (next page) illustrates forward and reciprocal chromosome painting schematically. In a standard forward painting experiment, a whole-chromosome paint from one species (species A) highlights homologous segments in the chromosomes of another species (species B). But the sub-regional origin of each homologous segment is unknown. In a reciprocal painting experiment, whole-chromosome paints from species B are hybridised back onto the metaphase chromosomes of species A.

FORWARD
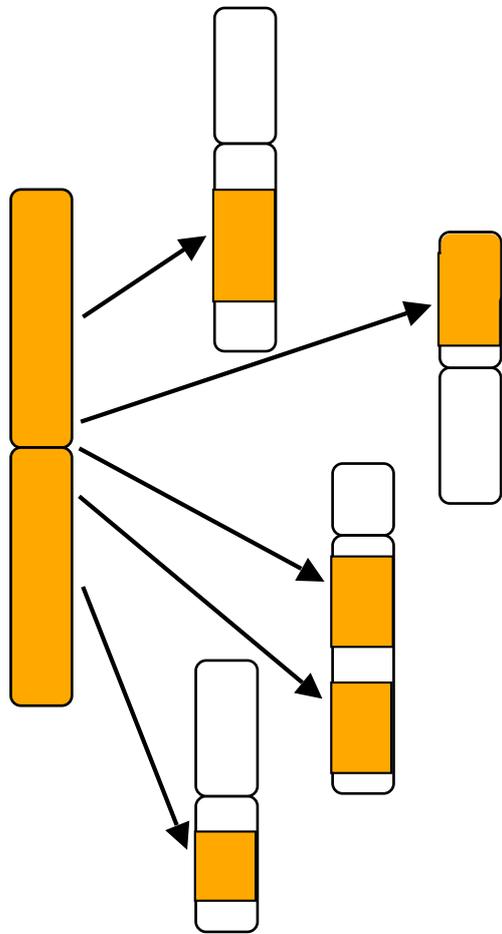
REVERSE
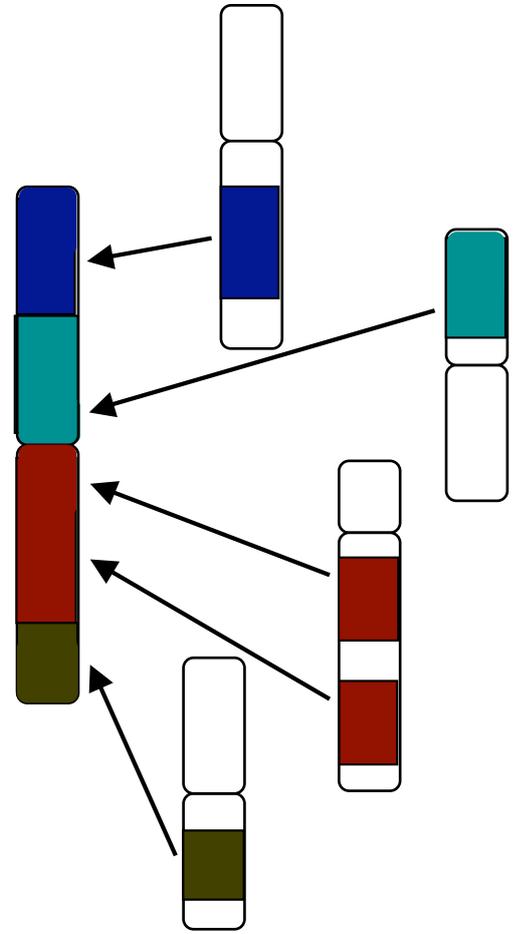
Species A

Species B

Species A

Species B

1.5.4.2.1 <u>Chromosome flow sorting</u>

This technique can produce highly pure samples of individual chromosomes. Chromosomes, which have been stained with two fluorescent dyes (Hoechst 33258 and Chromomycin A3), are forced to flow in sheath fluid one-by-one through the focus of two lasers. The lasers excite the fluorescent dyes and the emitted light signals from each chromosome are presented as co-ordinates on a bivariate plot (flow karyotype) of Hoechst 33258 versus Chromomycin A3. These two dyes bind to DNA differentially: Hoechst 33258 binds preferentially to AT-rich regions and Chromomycin A3 to GC-rich regions. Therefore, the chromosomes can be resolved on the flow karyotype based on their DNA content (size) and base pair ratios (van den Engh *et al.,* 1985). Any discrete chromosome peak on the flow karyotype can be selected using the cytometer workstation software and sorted to a high degree of purity (>95%) (Ross and Langford, 1997). The sorting process uses electrostatic deflection to direct charged droplets of the sheath fluid containing the chromosome of choice into a collection tube. Since droplets can be charged either positively or negatively (and hence deflected to one side or the other), it is possible to sort two chromosome types simultaneously into separate collection tubes.

The lay out of a typical commercially available dual-laser flow cytometer is shown in Figure 1.7

Figure 1.7 Lay out of a typical dual-laser flow cytometer. (Only one laser beam is illustrated.) The laser beam is shown focused onto the stream of cells or chromosomes. Both forward angle scattered light and emitted fluorescence can be detected. The fluorescence events are converted into electronic signals and processed before being displayed by the sorter workstation software.
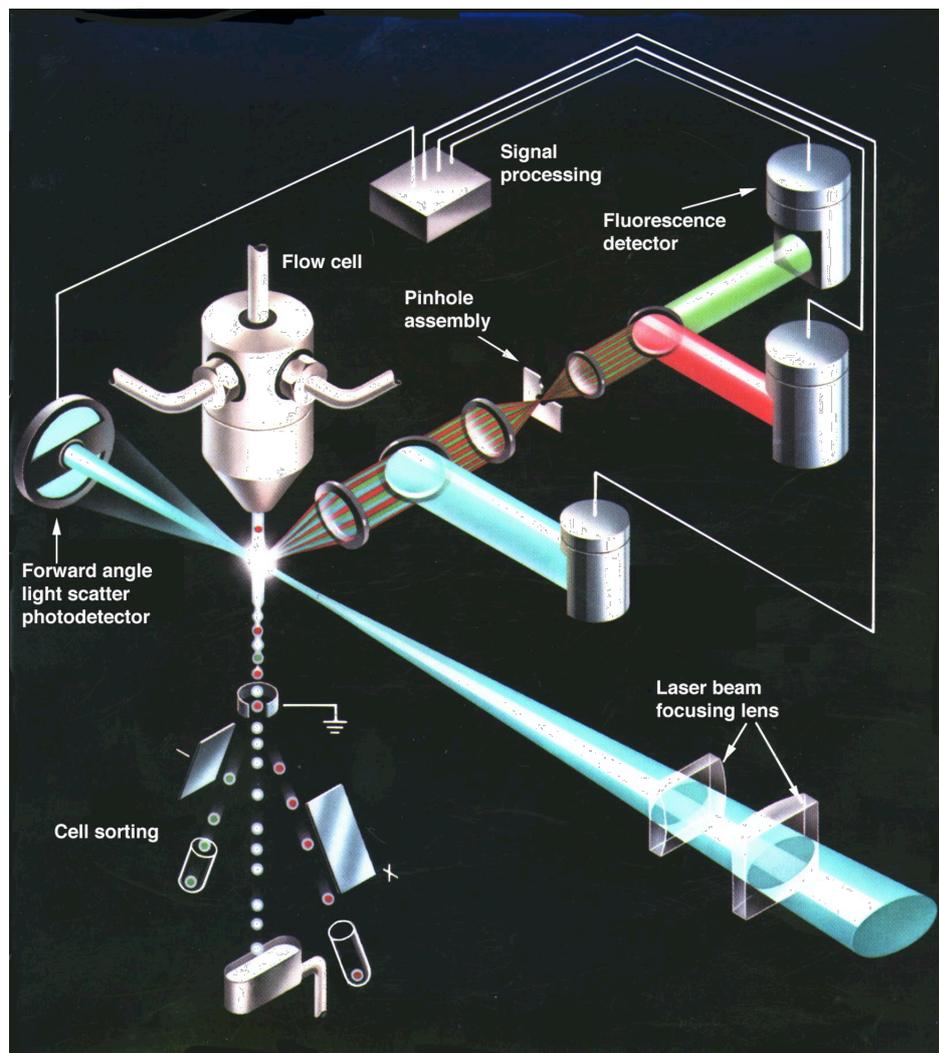
Human chromosomes lend themselves well to flow-cytometric analysis and sorting because of their large range of sizes and base pair compositions. All but chromosomes 9-12 of man can be resolved on the bivariate flow karyotype (Figure 1.8).

Figure 1.8 Human bivariate flow karyotype. Chromomycin A3 and Hoechst 33258 fluorescence intensities are plotted in arbitrary units. Each cluster of points corresponds to one chromosome type, with the exception of chromosomes 9-12, which appear as a single cluster.
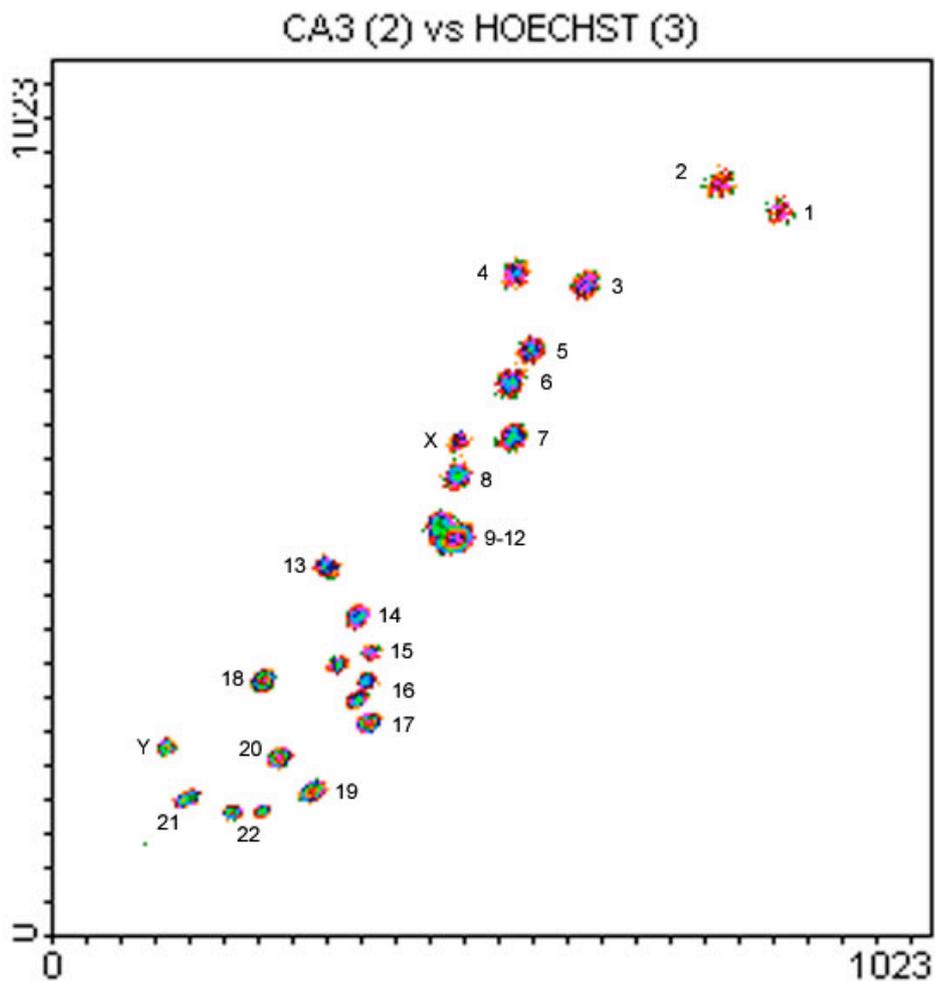
1.5.4.2.2 Chromosome microdissection

An alternative to flow sorting for generating chromosome specific probes is microdissection of cytogenetically prepared metaphase chromosomes. A glass needle attached to a micromanipulator is used to dissect a whole chromosome, a chromosome arm or regions of arms ranging from 5-10 Mb in size. Several dissected chromosome fragments are transferred to a collection tube, where the material undergoes PCR amplification (Cannizzaro, 1996).

1.5.4.2.3 Chromosome Paint Generation

Once isolated, DNA from each chromosome type can be either directly amplified using partially degenerate primers (e.g. degenerate oligonucleotide primed PCR (DOP-PCR; (Telenius *et al*. 1992a; Telenius *et al*. 1992b; Carter 1994), or used for library construction (Collins *et al* 1991). In both cases, whole chromosome-specific DNA is available as a complex probe for FISH. DOP-PCR employs partially degenerate oligonucleotides for the general, species-independent amplification of target DNA. The degeneracy, coupled with a PCR protocol utilising a low annealing temperature for the first few cycles, ensures priming from multiple (e.g. approximately 10^6 in human) dispersed sites within a given genome. The DOP-PCR method of probe generation is not reliant on cloning and produces highly representative chromosome paints, which improves the potential accuracy of interpreting Zoo-FISH results.

1.6 Zoo-FISH studies in the mammals

The first cross-species chromosome painting studies were reported among the genomes of evolutionarily closely related hominids (Wienberg *et al*. 1990). Jauch and co-workers then described the hybridisation of human chromosome-specific paints onto the metaphase spreads of the great apes (chimpanzee, gorilla and orangutan) and some of the lesser apes (gibbons) (Jauch *et al*. 1992). Wienberg and colleagues extended the study to compare the human genome organisation with that of the relatively primitive New

World monkey *Macaca fuscata* (Wienberg *et al*. 1992). The high degree of sequence homology among primate genomes facilitated the identification of homologies between their chromosomes by chromosome painting (Wienberg *et al*. 1994; Koehler *et al*. 1995a,b; Consigliere *et al*. 1996; Wienberg and Stanyon 1997). These studies were carried out using biotinylated DNA isolated from chromosome-specific plasmid libraries from the Lawrence Livermore collection (Collins *et al*. 1991) or PCR-generated linker-adapter library DNA probes (Vooijs *et al*. 1993). The researchers deduced that, as predicted by G-banding studies, there was a considerable level of conserved chromosomal synteny between the karyotypes of the great apes and man and less synteny between the karyotypes of lesser apes and man.

It was reported that by changing the methodology of hybridisation to reduce stringency and increase hybridisation time, it was possible to extend comparative chromosome painting studies of human to more distantly related mammals such as the whale (Scherthan *et al*. 1994). Subsequently, Raudsepp and co-workers published the first comparative genome map by zoo-FISH between the human and the horse (Raudsepp *et al*. 1996).

1.6.1 Limitations of zoo-FISH Using DNA from Chromosome-Specific Plasmid Libraries
The early zoo-FISH studies provided valuable new information regarding comparative genome organisation between human and other mammals. However, it became evident that the representation of each of the Lawrence Livermore chromosome-specific libraries was inconsistent. It was observed that paint probes representing some human chromosomes generated only weak hybridisation signals and that certain chromosome regions in others were under-represented by the libraries. Weak or absent hybridisation signals potentially could lead to the misinterpretation of zoo-FISH results.

The limitations of the libraries were most probably caused by contamination of human with hamster chromosomes during flow sorting and/or deletions of the human chromosome hybrid cell lines. This, coupled with the extra potential problem of biases

introduced during library amplification, means that each library may under-represent certain chromosome sequences or blocks of sequences.

1.6.2 Zoo-FISH Using DOP-PCR Generated Chromosome-Specific Paints

The majority of problems in chromosome probe representation were alleviated when researchers conducting zoo-FISH studies began to utilise chromosome-specific paint probes generated from degenerate oligonucleotide-primed PCR (DOP-PCR) amplified flow-sorted chromosomes. Only a few hundred chromosomes were required as template for DOP-PCR amplification. It is undoubtedly much easier to maintain a high degree of purity during the few minutes required to sort a few hundred chromosomes for DOP-PCR compared to the weeks required to isolate sufficient chromosome material for the Lawrence Livermore libraries.

A considerable number of zoo-FISH studies have been carried out (Ferguson-Smith *et al.*, 1998). They span (at least) five mammalian orders (Primates, Artiodactyla, Carnivora, Perissodactyla and Cetacea), and involve the hybridisation of (usually) human chromosome specific paints onto metaphase preparations of at least twenty-four species. A summary of the results of many of those studies is presented in the pull-out poster (figure 1.9), which was published in the 15 October 1999 issue of *Science* and is reproduced with kind permission from Jennifer Marshall Graves.

Figure 1.9 (next page) Comparative Genomics and Mammalian Radiations, published in the 15 October 1999 issue of Science.

The number of homologous autosomal segments in primates detected by the 22 human autosomal chromosome specific paints ranges from 23 in the chimpanzee, orangutan and the macaque (Jauch *et al.* 1992) to 63 in the concolor gibbon (Jauch *et al.* 1992, Koehler *et al.* 1995b). At the time of this study, the number of human homologous autosomal segments detected in non-primates ranges from 30 in the dolphin (Bielec *et al*. 1998) and harbour seal (Rettenberger *et al*. 1995b; Frönicke *et al*. 1997) to 49 in cattle (Hayes 1995). This information is summarised in Table 1.2, (see over page).

Table 1.2 (next page) The number of homologous autosomal segments detected by the 22 human autosomal chromosome specific paints in twenty-four mammals, from five mammalian orders (Primates, Artiodactyla, Carnivora, Perissodactyla and Cetacea).

Mammal	Number of autosomal homologous segments
Chimpanzee *Pan troglodytes*[1]	23
Gorilla *Gorilla gorilla*[1]	25
Orangutan *Pongo pygmaeus*[1]	23
White handed Gibbon *Hylobates lar*[1]	51
Concolor Gibbon *Hylobates concolor*[1, 2]	63
Siamang Gibbon *Hylobates syndactylus*[3]	59
Capuchin *Cebus capuchinus*[4]	33
Marmoset *Callithrix jacchus*[5]	30
Macaque *Macaca fuscata*[6]	23
Black-handed spider monkey *Ateles geoffroy*[7]	48
Silvered leaf monkey *Presbytis cristata*[8]	30
Red howler monkey *Alouatta seniculus arctoidea*[9]	42
Red howler monkey *Aluoatta seniculus sara*[9]	41
Lemur *Eulemur fulvus mayottensis*[10]	38
Cat *Felis catus*[11, 12]	31
American mink *Mustela vison*[13]	32
Harbour seal *Phoca vitulina*[14]	30
Cattle *Bos taurus*[15, 16, 17]	49
Sheep *Ovis aries*[18]	47
Pig *Sus scrofa*[19, 20, 21, 22]	46
Horse *Equus caballus*[23, 24, 25]	42
Indian muntjac *Muntiacus muntjak vaginalis*[26,27,28,29]	47
Common shrew *Sorex araneus*[30]	32
Dolphin *Tursiops truncatus*[31]	30

[1]Jauch *et al*. 1992, [2]Koehler *et al*. 1995b, [3]Koehler *et al*. 1995a, [4]Richard *et al*. 1996, [5]Sherlock *et al*. 1996, [6]Wienberg *et al*. 1992, [7]Morescalchi *et al*. 1997, [8]Bigoni *et al*. 1997, [9]Consigliere *et al*. 1996, [10]Muller *et al*. 1997, [11]Rettenberger *et al*. 1995b, [12]Wienberg *et al*. 1997, [13]Hameister *et al*. 1997, [14]Frönicke *et al*. 1997, [15]Hayes *et al*. 1995, [16]Solinas-Toldo *et al*. 1995, [17]Chowdhary *et al*. 1996, [18]Iannuzzi *et al*. 1999, [19]Rettenberger *et al*. 1995a, [20]Frönicke *et al*. 1996, [21]Goureau *et al*. 1996, [22]Milan *et al*. 1996, [23]Raudsepp *et al*. 1996, 1997, [24]Rettenberger *et al*. 1996, [25]Lear and Bailey 1997, [26]Scherthan *et al*. 1994, 1995, [27]Frönicke and Scherthan 1997, [28]Wienberg and Stanyon 1997, [29]Yang *et al*. 1997, [30]Dickens *et al*. 1998, [31]Bielec *et al*. 1998

1.7 Patterns of Comparative Karyotype Organisation

As more zoo-FISH studies have been carried out, patterns of comparative karyotype organisation have emerged. Conservation of whole chromosome synteny and conservation of ancestral neighbouring segment combinations have been observed (Chowdhary *et al*. 1998). The former involves chromosome types that tend to be conserved as a single chromosome or a single ECCS in most of the species studied. Chromosomes corresponding to human chromosomes 13, 17, 20 and X demonstrate conservation of whole chromosome synteny. In nearly all the species studied to date by zoo-FISH, these chromosomes are either represented as a single chromosome or as a whole chromosome arm. The only possible exception has been found in the Indian muntjac (2n = 6/7), where the region corresponding to HSA20 is disrupted by a small segment homologous to HSA10 (Yang, *et al.,* 1997).

Of all mammalian chromosomes, the X stands out as the most conserved between mammals. The majority of the genes on the human X that have been mapped in other mammalian species are also on the X. There are, however, several genes on human X that are on autosomes in the marsupial (Marshall Graves 1998). The exceptional conservation of chromosome X was recognised in the 1960s by Ohno and was proposed to be the result of selection against disruption of the chromosome-wide X inactivation system (Ohno 1964).

Regions corresponding to (parts of) human chromosomes 3 and 21, 14 and 15, 12 and 22, and 16 and 19, tend to be neighbouring in the genomes of most of the species studied. This tendency indicates that these combinations probably represent ancestral chromosome arrangements (Chowdhary, *et al.,* 1998). The ancestral combinations were probably disrupted during the relatively recent chromosome fission events during the evolution of the primate karyotype. An alternative explanation may be that these combinations arose by the convergent (or *de novo*) fusion of independent ancestral genomic fragments during evolution. However, this seems highly unlikely considering that

the neighbouring segments have been consistently observed in numerous divergent species.

1.8 Defining ECCS Boundaries

High-resolution cross-species FISH using sub-regional probes can be used to define the boundaries of ECCSs on a finer scale than that provided by chromosome paints. Clones that span ECCSs contain sequences that define evolutionary rearrangement points. Fine mapping of these regions may provide clues to understanding the DNA sequence and the rearrangement processes that have contributed to ancestral genome evolution. Having access to genome sequences for many different mammals will allow many such rearrangement points to be studied, but until that time targeted analyses will have value.

1.9 Aims of this thesis

The aim of this work was to carry out a study of evolutionary chromosome rearrangements involving material homologous to human chromosome 22 in two mammals: the domestic dog and the Siamang gibbon, with a view to understanding the underlying mechanisms by which they occurred. The work follows a targeted approach including reciprocal chromosome painting (chapter 3), high-resolution cross-species FISH (chapter 4), and the construction, characterisation and screening of a gibbon genomic cosmid library (chapter 5). The most detailed possible analysis of one evolutionary rearrangement event involving HSA22 material was carried out at the sequence level, where the sequences of two gibbon cosmids spanning HSA22 syntenic block junctions were analysed (chapter 6). The reasons for choosing human chromosome 22, the dog and the gibbon for analysis are described below.

Human chromosome 22

Human chromosome 22 is the second smallest of the human autosomes, being 48 to 54

megabase pairs in size (Mayall *et al.* 1984), and comprising some 1.6-1.8 % of the genomic DNA. It was also the first human chromosome for which the complete reference sequence was determined (Dunham, *et al.* 1999). Chromosome 22 is a recently formed chromosome that is only found in higher primates. Numerous comparative banding and painting studies have revealed that, apart from in the mouse, material homologous to HSA22 is found in only two or three separate blocks within 1, 2 or 3 different chromosome types in lemurs and all other mammalian karyotypes studied (summarised in figures 1.10 and 1.11). In contrast, blocks of HSA22 homologies are found at 21 different sites within the murine genome on eight different chromosome types. The most parsimonious interpretation of this evidence is that the state of HSA22-homologous material within the ancestral mammalian karyotype is in two blocks, which have undergone a fusion event during the evolution of the primates. In fact it has been suggested that HSA22 was formed from a single reciprocal translocation event involving two ancestral chromsomes (Haig 1999).

As well as being involved in relatively simple rearrangements during mammalian karyotype evolution, and having been fully sequenced, the human chromosome 22 material was a suitable candidate for analysis because of the other considerable resources available for molecular analysis including contiguous yeast (YAC) and bacterial (BAC, PAC, cosmid, fosmid) clones spanning almost the entire chromosome.

Figure 1.10 and 1.11 (next pages) Schematic summary of zoo-FISH studies indicating regions of human chromosome 22 homology in the chromsomes of mammals and primates (modified from Glas, *etal.,* 1998). The mammalian branching order is based on a molecular phylogenetic analysis reported in Novacek, 1992, and the primate branches are based on Dutrillaux, 1979.
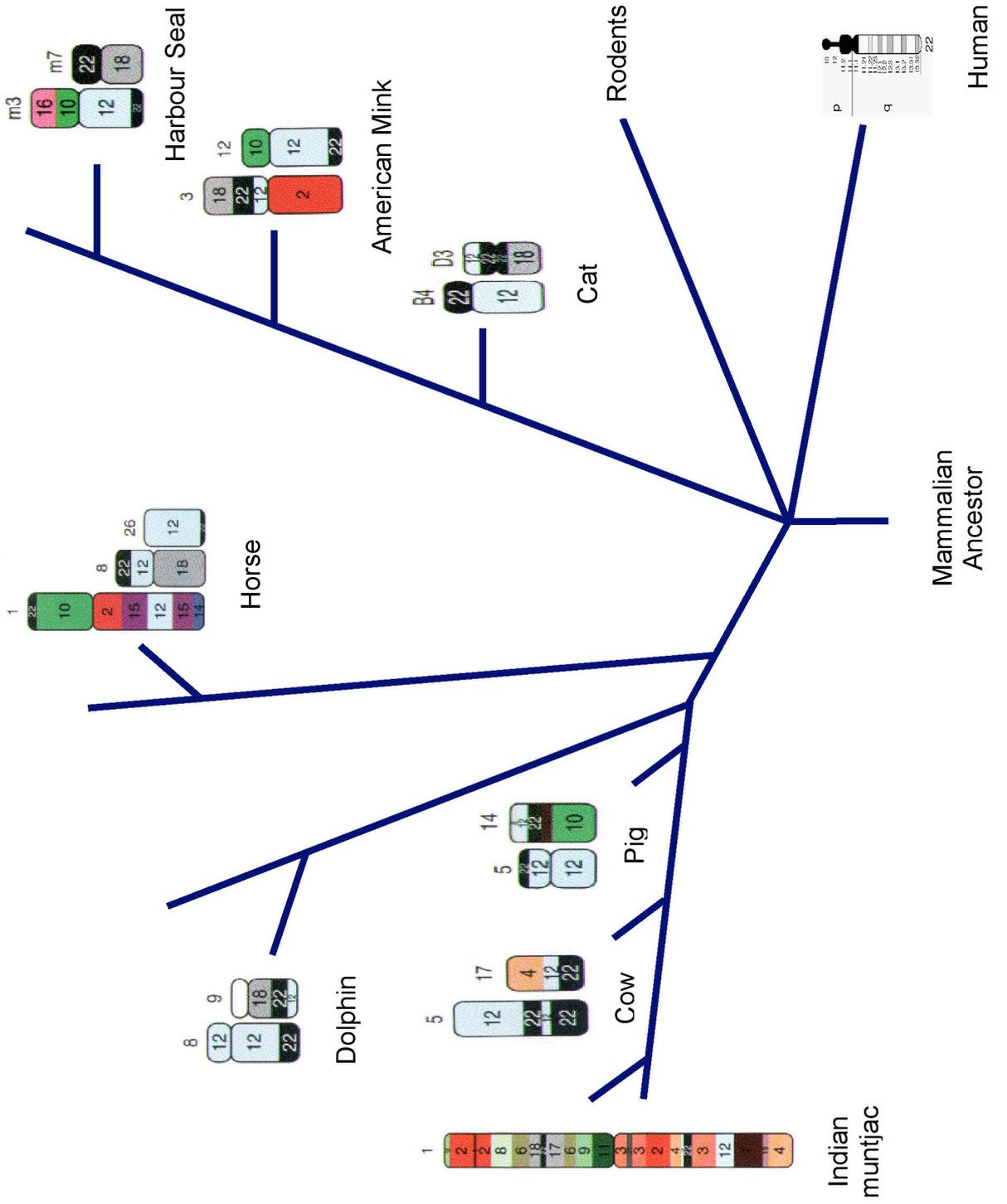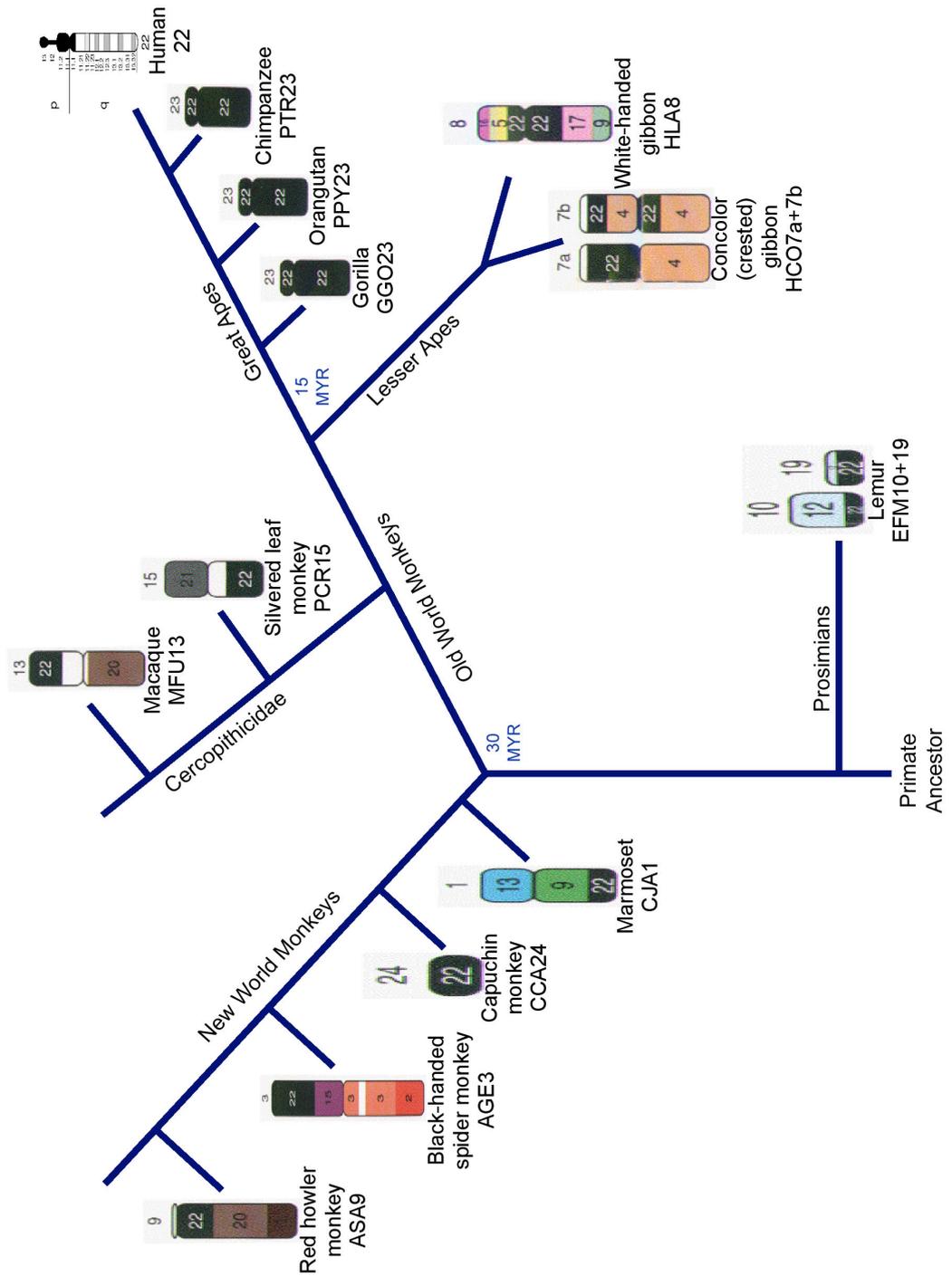
In planning the experiments, of the two mammals selected for karyotype analysis, one was from a family distantly related to humans (i.e. carnivora) and one from a closely related primate (i.e. lesser ape). The distantly related mammal chosen for study was the

carnivorous domestic dog. The closely related primate chosen was a lesser ape, the Siamang gibbon. The reasons for choosing those mammals are described below.

The Domestic Dog

The dog and human diverged from a common ancestor approximately 70 million years ago (Novacek, 1992). The domestic dog is used as an animal model for many human diseases, and several genetic disorders in dogs have been shown to be models of human inherited diseases, including X-linked severe combined immunodeficiency (SCID) (Henthorn *et al.* 1994), Duchenne muscular dystrophy (Schatzberg *et al.* 1999) and narcolepsy (Kadotani *et al.* 1998). The dog has 78 chromosomes: 76 acrocentric autosomes and two sex chromosomes (Selden *et al.* 1975). The large submetacentric X and the minute metacentric Y are the longest and shortest of the chromosome complement, respectively. The largest autosome is almost equal in length to the X chromosome, with the remaining autosomes diminishing gradually in size.

At the time of the research for this thesis, the dog was the only mammal among the common domestic and laboratory animals for which there was no standard karyotype. Attempts to establish an accepted karyotype had been frustrated by the similarity in size and banding morphology of several of the smaller chromosomes. In 1995, the Committee for the Standardisation of the Canine Karyotype agreed upon the order and banding pattern of the first 21 chromosomes, plus X and Y (Switonski *et al.,* (1996). It was generally accepted that the unequivocal cytogenetic identification of the remaining 17 undesignated autosomes would be dependent on chromosome painting or the mapping of specific probes to each. Because only limited cytogenetic studies had previously been carried out on the dog, it was an appropriate candidate for karyotype analysis by chromosome painting.

The Siamang Gibbon

There is a close analogy of chromosome G-banding between most of the great apes and man, and at least 70% of bands are common to Simians and the Prosimian lemurs.

Studies on banded primate karyotypes have gone some way to reveal the sequence of chromosomal rearrangements, which have occurred during their evolution and have allowed the proposal of a precise geneaology of many primates (Dutrillaux, 1979). However, chromosomal conservation in primates has some striking exceptions. The gibbons, for example, exhibit extensive chromosome rearrangements away from the great ape ancestral karyotype, despite a relatively recent divergence of only 18 to 25 million years ago. Almost none of the *Hylobates syndactylus* (Siamang) gibbon chromosomes can be identified, by banding, as being homologous to the human chromosome complement (Van Tuinen and Ledbetter 1983, Koehler *et al.* 1995, O'Brien *et al.* 1998).

The Siamang gibbon (Figure 1.12) is a primate closely related to the great apes and has had some previous cytogenetic study by chromosome painting (Koehler *et al.* 1995). It was chosen for study because previous chromosome painting studies indicated that it is the closest primate relation to the human with material homologous to human chromosome 22 distributed into two discrete ECCS, which are on different arms of gibbon chromosome 18.

The studies carried out for this thesis are described in the following pages.

Figure 1.12 Hylobates syndactylus the Siamang or Great Gibbon. Photographed by S. Hoffman, reproduced from Animal Diversity Web, University of Michigan, http://animaldiversity.ummz.umich.edu.

CHAPTER 2

Materials and Methods

<u>Materials</u>

2.1 Composition of solutions

2.2 Reagents and stains

2.3 Media

2.4 Cells and cell lines

2.5 Bacterial clones

<u>Methods</u>

2.6 Tissue culture

2.6.1 Lymphoblastoid cell culture

2.6.2 Fibroblast cell culture

2.6.3 Cell cryopreservation

2.7 Flow karyotype analysis and chromosome sorting

2.7.1 Lymphoblastoid cell chromosome isolation and staining for flow cytometric analysis

2.7.2 Generation of a bivariate flow karyotype and chromosome sorting

2.8 Generation of chromosome-specific paints by DOP-PCR

2.11.2 Preparation of Lawrist16 vector arms

2.11.3 Ligation and packaging of partially -digested DNA

2.11.4 DH5αMCR E. coli plating cell preparation

2.11.5 Assessing library titres

2.11.6 Making gibbon cosmid library filters

2.11.7 Filter screening with gibbon STSs

2.11.8 Isolation of unknown gibbon sequences by Vectorette PCR

<u>Materials</u>

2.1 Composition of solutions

<u>Denaturation solution</u>

- 0.5 M NaOH

- 1.5 M NaCl

<u>dNTP mix for 1^0 DOP-PCR:</u>

- 2.5 mM each dNTP

<u>dNTP mix for 2^0 DOP-PCR:</u>

- 2.5 mM each dATP, dCTP, dGTP

- 1.25 mM dTTP

<u>dNTP mix for nick translation:</u>

- 0.5 mM each dATP, dCTP, dGTP

<u>Flow sorter sheath fluid:</u>

- 100 mM NaCl

- 10 mM Tris-base

- 1 mM Na_2EDTA

- 0.5 mM sodium azide

<u>HSRE:</u>

- 500 mM Tris-HCl pH 7.5

- 100 mM $MgCl_2$

- 1.5 M NaCl

Hybridisation buffer (FISH):

- 50% (v/v) deionised formamide

- 10% (w/v) dextran sulphate

- 2× SSC

Hybridisation buffer (library filter screening):

- 0.2% (w/v) ficoll

- 0.2% BSA

- 0.2% polyvinylpyrollidone

- 6× SSC

- 50 mM Tris-base

- 10% (w/v) dextran sulphate

- 1% (w/v) N-lauroyl sarcosine

Filtered through one layer of Whatman No. 4 filter paper

Hypotonic swelling solution:

- 75 mM KCl

- 0.2 mM spermine

- 0.5 mM spermidine

Lambda diluent:

- 10 mM Tris-HCl pH 7.5

- 10 mM $MgSO_4$

10× ligase buffer:

- 400 mM Tris-HCl pH 7.6

- 100 mM $MgCl_2$

- 10 mM DTT

6× loading buffer:

- 0.25% (w/v) bromophenol blue

- 0.25% (w/v) xylene cyanol

- 30% (v/v) glycerol

Made up in $T_{0.1}E$

LSRE:

- 50 mM Tris-HCl pH 7.5

- 10 mM $MgCl_2$

Lysis buffer:

- 50 mM glucose

- 10 mM Na_2EDTA

- 25 mM Tris-HCl pH 8.0

1× NDS:

- 500 mM Na_2EDTA

- 10 mM Tris-base pH 9.5

- 1% (w/v) N-Lauroylsarcosine

10× NEB buffer (modified):

- 66 mM Tris-HCl pH 8.8

- 16.6 mM $(NH_4)_2SO_4$

- 67 mM $MgCl_2$

Made up in $T_{0.1}E$

Neutralisation solution

- 0.5 M Tris-HCl pH7.4

- 1.5 M NaCl

10× nick translation buffer:

- 0.5 M Tris-HCl pH 7.5

- 0.1 M $MgSO_4$

- 1 mM DTT

- 500 µg/ml BSA

PBS:

- 2.6 mM KH_2PO_4

- 26 mM Na_2HPO_4

- 145 mM NaCl pH 7.2

- 2.7 mM KCl

Polyamine buffer:

- 80 mM KCl

- 20 mM NaCl

- 2 mM Na_2EDTA

- 0.5 mM EGTA

- 15 mM Tris

- 0.2 mM spermine

- 0.5 mM spermidine

- 3 mM DTT

- 0.25% (v/v) Triton X-100

Adjusted to pH 7.2 with HCl

Restriction enzyme diluent:

- 20 mM Tris-HCl pH 7.5

- 100 mM KCl

- 50 µg/ml gelatine

- 0.1% (v/v) 2-mercaptoethanol

- 50% (v/v) glycerol

5× SM:

- 500 mM NaCl

- 50 mM $MgSO_4$

- 250 mM Tris-HCl pH 7.5

- 0.05% (w/v) gelatine

- 50% (v/v) glycerol

20× SSC:

- 3 M NaCl

- 0.3 M tri-sodium citrate pH 7.0

50× TAE:

- 2 M Tris-acetate

- 0.05 M Na_2EDTA

10× TAK buffer:

- 300 mM Tris-HCl pH 7.9

- 600 mM potassium acetate

- 90 mM magnesium acetate

- 3 mg/ml BSA

- 5 mM DTT

- 800 µM SAM

Make up the solution omitting the DTT and SAM and store frozen in 1 ml aliquots.

Add DTT and SAM just before use.

10× TAPS2 buffer:

- 250 mM TAPS pH 9.3

- 166 mM $(NH_4)_2SO_4$

- 25 mM $MgCl_2$

- 0.165% (w/v) BSA

- 0.1 mM 2-mercaptoethanol

10× TBE:

- 0.89 M Tris-borate

- 2 mM Na_2EDTA pH 8.3

TE:

- 10 mM Tris-HCl pH 8.0

- 1 mM Na_2EDTA

$T_{0.1}E$:

- 10 mM Tris-HCl pH 8.0

- 0.1 mM Na_2EDTA

4× TNFM

- 4× SSC

- 0.05% (v/v) Tween 20

- 5% (w/v) non-fat milk powder

  Filtered through several layers of Whatman No.4 filter paper

2.2 Reagents and stains

Note: All stains were made up in sterile distilled water, unless stated otherwise.

Restriction enzymes:

- *Mbo*I 5 U/µl (NEB)

- Dam methylase

- Calf Intestinal Alkaline Phosphatase

- *Sca*I 10 U/µl (NEB)

- *BamH*I

Haptens for 2^0 DOP-PCR:

- Biotin-16-dUTP (Boehringer)

- Digoxygenin-11-dUTP (Boehringer)

- FluoroRed-dUTP (Amersham)

- FluoroGreen-dUTP (Amersham)

- FluorX-dCTP (Biological Detection Systems)

$T_{0.1}E$:

- Cy3-dUTP (Amersham)

- Oregon Green-dUTP (Molecular Probes Europe)

All at 1 nmol/µl

Chromomycin A3:

- 2 mg/ml in ethanol

DAPI:

- 80 ng/ml in 2× SSC

Hoechst 33258:

- 1 mg/ml

Turck's stain:

- 1% (v/v) acetic acid

- 0.1 mg/ml gentian violet

2.3 Media

LB broth:

- 1% (w/v) Bacto tryptone

- 0.5% (w/v) Bacto yeast extract

- 1% (w/v) NaCl

2× TY broth:

- 1.6% (w/v) Bacto tryptone

- 1% (w/v) Bacto yeast extract

- 0.5% (w/v) NaCl

2.4 Cells and cell lines

Note: The lymphoblastoid cell lines used had previously been immortalised by EBV transformation. The transformation process induces the EBV to randomly integrate into the host genome.

An anonymous normal human male lymphoblastoid cell line ("HRC575") was obtained from ECACC, Wiltshire, U.K.

A female Siamang gibbon, (*Hylobates syndactylus* , "HSY") lymphoblastoid cell line was kindly donated by Dr. Johannes Wienberg, Cambridge University Department of Pathology, Cambridge, U.K.

Canine blood samples for chromosome isolation and metaphase spread preparations were kindly donated by Dr. Matthew Breen, Animal Health Trust, Newmarket, U.K.

2.5 Bacterial clones

All BAC, PAC, cosmid and fosmid clones used for the high resolution FISH analysis of Siamang and dog metaphase chromosomes (were identified from the human chromosome 22 data base ("22ace", available online at http://webace.sanger.ac.uk/cgi-bin/webace?db=acedb22&frame=&.cgifields=db&Browse+DB=Browse+DB) and the Molecular Cytogenetics data base ("FISHLog") at The Sanger Centre. DNA from each clone was either isolated and subsequently labelled for FISH by nick translation by myself, or was obtained from frozen stocks previously isolated and labelled by members of the Molecular Cytogenetics team. Clones identified from FISHLog had been previously FISH-mapped to specific regions of human chromosome 22.

Methods

2.6 Tissue culture

Note: All cell types were cultured under 5% CO_2 at 37˚C.

2.6.1 Lymphoblastoid cell culture

Lymphoblastoid cells were cultured in suspension in 50 ml of RPMI-1640 (Sigma) supplemented with 16% (v/v) FBS (GibcoBRL), 100 U/ml Penicillin (Sigma), 100 µg/ml Streptomycin (Sigma) and 2 mM L-glutamine (Sigma) in 75 cm^2 tissue culture flasks. Once confluent, sub-culturing was carried out as follows:

1. Add 25 ml of fresh media to each flask and gently break up any cell clumps.

2. Distribute the media between three new 75 cm^2 flasks and add a further 25 ml of fresh media to each flask, effecting a 1/3 dilution of the confluent starting culture.

2.6.2 Fibroblast cell culture

Fibroblast cells were cultured in 50 ml of DMEM (Sigma) supplemented with 8% (v/v) FBS, 100 U/ml Penicillin, 100 µg/ml Streptomycin and 2 mM L-glutamine in 162 cm^2 tissue culture flasks. Once confluent, the adherent fibroblasts were sub-cultured as follows:

1. Discard the tissue culture medium and rinse the cell layer with PBS.

2. Discard the PBS and incubate the cell layer with 2.5 ml of Trypsin/EDTA solution (Sigma) for 2 to 5 minutes at room temperature. Tap the flasks to facilitate cell release.

3. Resuspend the cells in 150 ml of fresh media and distribute between three new 162 cm^2 flasks, effecting a 1/3 dilution of the confluent starting culture.

2.6.3 Cell cryopreservation

For frozen storage, cells were pelleted at 200 × *g*, and resuspended at approximately 1x10^7 cells/ml in 10% (v/v) DMSO in FBS. Polypropylene tubes containing the cell suspension were cooled overnight to −70˚C, then transferred to the gas phase of a liquid nitrogen vessel (approximately −180˚C) for permanent storage. To recover from liquid nitrogen storage, cells were thawed rapidly at 37˚C then washed and resuspended in 10 ml of fresh media.

2.7 Flow karyotype analysis and chromosome sorting

2.7.1 Lymphoblastoid cell chromosome isolation and staining for flow cytometric analysis

Chromosomes were isolated from lymphoblastoid cells and stained for flow cytometric analysis by the following method:

1. To a 40% confluent lymphoblastoid culture add Colcemid Karyomax (GibcoBRL) to 0.1 µg/ml. Gently mix the culture and incubate for 6 hours at 37˚C.

2. Transfer the culture to a 50 ml centrifuge tube and pellet the cells at 200 × *g* for 10 minutes. Discard the supernatant and invert the tube on a tissue to drain.

3. Gently resuspend the cell pellet in 10 ml of hypotonic swelling solution and incubate at room temperature for 15 minutes.

4. Assess the proportion of cells arrested in metaphase by staining 10 µl of the cell suspension with 10 µl of Turck's stain and viewing in a haemocytometer with a phase-contrast microscope[a].

5. Transfer the swollen cell suspension to a 25 ml centrifuge tube and pellet the cells at 400 × *g* for 8 minutes. Discard the supernatant and drain the tubes as before.

6. Resuspend the cell pellet in 1.5 ml of ice-cold polyamine buffer and incubate on ice for 10 minutes.

7. Release the chromosomes into the solution by vortexing for 15 seconds at a speed which causes the suspension to swirl up around the wall of the tube.

8. Remove 10 µl of the suspension onto a microscope slide and stain it with 1 µl of 1 mg/ml propidium iodide. Check that the chromosomes are free in solution using a fluorescence microscope.

9. Centrifuge the chromosome suspension at $100 \times g$ for 1 minute to pellet any chromosome clumps and cellular debris.

10. Transfer 1.4 ml of the supernatant to a tube suitable for use on the flow sorter. Add 70 µl of Chromomycin A3 stain and mix immediately.

11. Add 3.5 µl of 1 M $MgSO_4$ and 7 µl of Hoechst 33258 stain to the chromosome suspension. Mix well and incubate the preparation for at least 1 hour on ice.

12. Add 175 µl of 0.1 M sodium citrate and 175 µl of 0.25 M sodium sulphite solution 15 minutes prior to flow analysis and sorting. Mix the solution well and incubate on ice.

[a]Metaphase chromosomes, stained purple, are visible filling the cell.

2.7.2 Generation of a bivariate flow karyotype and chromosome sorting

Bivariate flow karyotypes were generated and chromosomes sorted by the following method:

1. Set up a modified commercial flow sorter (Coulter ELITE-ESP, Coulter Electronics, see Figure), equipped with two argon-ion lasers, for chromosome analysis and sorting following the manufacturer's instructions.

2. Operate the lasers at a power of 300 mW and align them on the sample stream such that the chromosomes pass in the sheath fluid first through the UV (351-364 nm) beam (exciting the Hoechst 33258 stain), then through the 457.9 nm beam (exciting the Chromomycin A3 stain).

3. Use fluorescent microspheres (Coulter Electronics) to align the optics in preparation for chromosome analysis, such that the cv of signals at the Hoechst and Chromomycin detectors is 1.6 or less.

4. Adjust the sample pressure such that the stained chromosomes flow in single file through the focus point of each laser at a rate of approximately 1,000 per second.

5. Set up the flow sorter software to display the fluorescent signals emitted from each chromosome event as a co-ordinate on a bivariate plot of Hoechst 33258 versus Chromomycin A3 fluorescence. The chromosome types can then be resolved on the plot by their size and AT to GC base pair ratios.

6. Use the software to define the chromosome type to be sorted and switch on the high-voltage deflection plates.

7. Utilising the electrostatic deflection of charged droplets containing the chromosome type of interest, sort 500 to 1000 copies directly into a 0.5 ml PCR tube containing 33 µl of sterile distilled water.

8. With the lid secure flick the bottom of the tube to suspend any chromosomes which may have adhered to the side of the tube. Spin the tube briefly in a microfuge to bring down the solution.

9. Store the flow-sorted chromosomes on ice if they are to be used for DOP-PCR the same day. Alternatively, they will remain stable for over 3 years stored at –20˚C.

2.8 Generation of chromosome-specific paints by DOP-PCR

Chromosome-specific paints were generated from flow-sorted chromosomes by the following methods:

2.8.1 Primary DOP-PCR amplification of flow-sorted chromosomes

Flow-sorted chromosomal DNA was randomly amplified by primary DOP-PCR by the following method:

1. For a 50 µl reaction in a 0.5 ml PCR tube, mix the following:

- X µl sterile distilled water (to a final volume of 50 µl)

- 500 to 1000 flow-sorted chromosomes

- 5.0 µl 10× TAPS2 buffer

- 4.0 µl 1^0 DOP-PCR dNTP mix

- 5.0 µl 20 µM DOP-PCR primer (see Appendix)

2. Mix the tube contents and spin briefly in a microfuge to bring down the solution.

3. Place the tube on a PCR block and incubate at 94˚C for 8 minutes.

4. Add 0.5 µl (2.5 U) of *Taq* polymerase and continue the program with nine cycles of:

- 94˚C for 1 minute

- 30˚C for 1.5 minutes

- Transition at 0.23˚C per second to 72˚C

- 72˚C for 3 minutes

followed by 30 cycles of:

- 94˚C for 1 minute

- 62˚C for 1 minute

- 72˚C for 1 minute

followed by:

- 72˚C for 7 minutes

5. Mix 5 µl of the PCR products with 1 µl of 6× loading buffer and analyse by electrophoresis on a 1% w/v agarose gel.

2.8.2 Secondary DOP-PCR amplification

The products of a primary DOP-PCR amplification were further amplified and haptenised by the following method:

1. For a 50µl reaction in a 0.5 ml PCR tube, mix the following:

    - X µl sterile distilled water (to a final volume of 50 µl)

    - 5.0 µl 10× TAPS2 buffer

    - 3.5 µl 2^0 DOP-PCR dNTP mix

    - 5.0 µl 2^0 DOP-PCR hapten (see reagents and stains, this chapter)

    - 5.0 µl 20 µM DOP-PCR primer (see Appendix)

    - 2 µl (approximately 150 ng) 1^0 DOP-PCR product

2. Mix the tube contents and spin briefly in a microfuge to bring down the solution.

3. Place the tube on a PCR block and incubate at 94˚C for 8 minutes.

4. Add 0.5 µl (2.5 U) of *Taq* polymerase and continue the program with 25 cycles of:

    - 94˚C for 1 minute

    - 62˚C for 1 minute

    - 72˚C for 1 minute

    followed by:

    - 72˚C for 7 minutes

5. Mix 3 µl of the PCR products with 1 µl of 6× loading buffer and analyse by electrophoresis on a 1% agarose gel.

2.9 DNA Preparation

2.9.1 Preparation of high molecular weight gibbon genomic DNA from lymphoblastoid cells

High molecular weight gibbon genomic DNA was prepared by the following method:

Safety Note: Carry out all manipulations involving phenol, chloroform and diethyl ether in a fume cabinet.

1. Culture a flask of gibbon lymphoblastoid cells until almost confluent, so that there are approximately 1×10^6 cells/ml in 50 ml of medium[a].

2. Transfer the culture to a polypropylene tube and pellet the cells by centrifugation at $400 \times g$ for 10 minutes, discard the supernatant and drain the tube briefly by inverting it on a tissue.

3. Wash the cell pellet in 20 ml of sterile PBS and fully resuspend the cells to 1×10^7 cells/ml in $1 \times$ NDS + 100 µg/ml Proteinase K (GibcoBRL). Incubate the mixture o/n in a 50°C water bath.

4. Add 4 volumes of 0.2 µm filtered $T_{0.1}E$. Invert the tube gently to mix and allow the solution to equilibrate to room temperature.

5. Add an equal volume of phenol (equilibrated with Tris) and mix gently, to form an emulsion, by slowly inverting the tube end-over-end for 10 minutes.

6. Separate the aqueous phase from the organic phase by centrifugation at $6000 \times g$ for 10 minutes. Collect the aqueous phase into a fresh tube.

7. Carry out two further phenol extractions, until the phase interface is clear of white protein debris.

8. Add an equal volume of 1:1 phenol:chloroform and mix gently for 10 minutes. After centrifugation at $3000 \times g$ for 15 minutes, collect the aqueous phase into a fresh tube.

9. Add an equal volume of chloroform and mix gently for 10 minutes. After centrifugation at $3000 \times g$ for 15 minutes, collect the aqueous phase into a fresh tube.

10. Add an equal volume of diethyl ether and mix gently for 10 minutes. After centrifugation at $1000 \times g$ for 10 minutes, discard the ether.

11. To precipitate the DNA, add 0.2 volumes of 10 mM ammonium acetate and mix gently. Add 2 volumes of 100% ethanol and invert the tube slowly end-over-end for 2-3 minutes.

12. Hook and swirl the precipitated high molecular weight DNA around a sterile inoculating loop and wash it by swirling in 40 ml of 70% ethanol for 2-3 minutes.

13. Transfer the DNA to a fresh 1.5 ml microfuge tube and air-dry for 1 hour at room temperature.

14. Dissolve the DNA in 500 µl TE by standing the tube o/n at 4°C[b].

[a]For high molecular weight DNA extractions, it is possible to yield approximately 4 pg of DNA per cell.

[b]A high molecular weight genomic DNA solution should be stored at 4°C.

2.9.2 Preparation of DNA from BAC, PAC, fosmid and cosmid clones

DNA from BAC, PAC, fosmid and cosmid clones was prepared by the following alkaline lysis "mini-prep" method:

Safety Note: Carry out all manipulations involving phenol in a fume cabinet.

1. Inoculate 10 ml of 2× TY broth in a 25 ml centrifuge tube with 1 µl of the appropriate clone glycerol stock, or a single colony. For cosmids and PACs include 30 µg/ml of kanamycin in the broth. For BACs in the vector pBACc3.6 and fosmids include 25 µg/ml of chloramphenicol, and for BACs in the vector pBeloBAC11 or pBAC108L include 12.5 µg/ml of chloramphenicol.

2. Allow the cultures to grow by incubating o/n at 37°C with shaking at 210 rpm.

3. Make at least one back-up glycerol stock for each culture by mixing 200 µl of sterile 100% glycerol with 800 µl of culture[a].

4. Pellet the overnight culture by centrifuging at 3400 × *g* for 10 minutes at room temperature, discard the supernatant and drain the tube briefly by inverting on a tissue.

5. Gently resuspend the cell pellet in 200 µl of lysis buffer, transfer the suspension to a 1.5 ml microfuge tube and incubate at room temperature for 10 minutes.

6. Add 400 µl of fresh 0.2 M NaOH/1% (w/v) SDS (for 1 ml, mix 50 µl of 4 M NaOH with 100 µl of 10% (w/v) SDS and 850 µl of sterile distilled water). Invert the tube several times to mix and incubate on ice for 5 minutes.

7. Add 300 µl of 3 M sodium acetate (pH 5.2), mix the solution and incubate on ice for 10-30 minutes until the host cell debris forms a precipitate.

8. Pellet the debris in a microfuge at 12,000 × *g* for 5 minutes. Transfer the supernatant to a fresh tube. Microfuge the supernatant for a further 5 minutes, and transfer the clear supernatant again to a fresh tube.

9. Add 600 µl of isopropanol (propan-2-ol) and either place at -70˚C for 10 minutes, or -20˚C for between 30 minutes and o/n.

10. Pellet the precipitate in a microfuge at 12,000 × *g* for 5 minutes, discard the supernatant and briefly drain the tube on a tissue.

11. On ice, resuspend the pellet in 200 µl of 0.3 M sodium acetate (pH 7.0).

12. Add an equal volume of 1:1 phenol:chloroform mixture and vortex for 30 seconds to form an emulsion. Separate the aqueous from the organic phases by microfuging at 12,000 × *g* for 3 minutes and transfer 150 µl of the aqueous phase to a fresh tube.

13. Back-extract the phenol:chloroform by adding 50 µl of 0.3 M sodium acetate (pH 7.0), vortexing to form an emulsion and microfuging at 12,000 × *g* for 2 minutes. Transfer 50 µl of the aqueous phase to the other 150 µl.

14. Add an equal volume of isopropanol invert the tube to mix and place at -70°C for 10 minutes.

15. After pelleting the DNA in a microfuge at 12,000 × g for 10 minutes, discard the supernatant and wash the DNA pellet with 500 µl of ice-cold 70% ethanol.

16. Immediately microfuge at 12,000 × g for 5 minutes. Discard the supernatant and air-dry the pellet for 25 minutes in a 37°C oven.

17. On ice, resuspend the DNA in 50 µl of $T_{0.1}E$ with 1 µl of RNAseA (stock 10 mg/ml). Incubate the preparation at 55°C for 15 minutes then store permanently at –20°C.

[a]Store glycerol stocks permanently at -70°C.

2.10 Procedures for fluorescence *in situ* hybridisation (FISH)

2.10.1 Fixed metaphase preparations from lymphoblastoid cell lines

Fixed metaphase preparations were prepared from lymphoblastoid cell lines by the following method:

1. To a 50% confluent culture, add BrdU to 10 µg/ml, mix well and incubate the cells for a total of 3 hours at 37°C.

2. 1.25 hours prior to harvest add ethidium bromide to 10 µg/ml and mix well.

3. 45 minutes prior to harvest add Colcemid Karyomax (GibcoBRL) to 0.05 µg/ml and mix well.

4. At the time of harvest, transfer the contents of the flask to a 25 ml centrifuge tube and pellet the cells by centrifugation at 200 × g for 10 minutes.

5. Discard the supernatant and resuspend the cells in 10 ml of 75 mM KCl pre-warmed to 37°C. Transfer the cell suspension to a 15 ml polystyrene tube and incubate at 37°C for 15-20 minutes.

6. Add 1 ml of fresh 3:1 methanol:acetic acid fix, invert the tube to mix then immediately pellet the cells by centrifugation at 400 × g for 10 minutes.

7. After carefully discarding the supernatant, resuspend the cell pellet in 2-3 ml of 3:1 fix. Centrifuge at 400 × g for 5 minutes and discard the supernatant.

8. Repeat the fixation and centrifugation steps a further 3 times. After the last centrifugation step, remove as much of the supernatant as possible.

9. Resuspend the fixed cell pellet in a small volume (1 to 2 ml) of 3:1 fix so that the solution remains cloudy.

10. Assess the fixed metaphase preparation by dropping a small aliquot onto a microscope slide from a pipette tip. Metaphase spreads can be detected under phase contrast using a light microscope.

For long-term storage, resuspend the fixed cells in approximately 15 ml of 3:1 fix. Seal the lid of the tube with parafilm and store the preparation at -20˚C.

2.10.2 Fixed metaphase preparations from fibroblast cell lines

Fixed metaphase preparations were prepared from fibroblast cell lines by the following method:

1. To a 50-60% confluent culture, add BrdU to 10 μg/ml, mix well and incubate the cells for a total of 4 hours at 37˚C.

2. 2.5 hours prior to harvest add Colcemid Karyomax to 0.05 μg/ml and mix well.

3. At the time of harvest, collect the culture media and rinse the adherent cell surface with 10 ml of sterile PBS. Collect the PBS and pool it with the media.

4. Incubate the adherent cells with Trypsin/EDTA solution (Sigma) and monitor cell release from the flask surface by viewing under a light microscope until 50% of the adherent cells have released.

5. Resuspend the released cells in the media and PBS collected previously, and transfer the suspension to a 50 ml polypropylene tube.

6. Pellet the cells by centrifugation at 200 × g for 10 minutes.

7. Continue the preparation from step 5 of the method for fixing lymphoblastoid cells *(2.10.1)* above.

2.10.3 Metaphase spread slide preparation

Fixed metaphase spreads were prepared on glass microscope slides by the following procedure:

1. Bring the tube containing the fixed metaphase suspension to room temperature.

2. After discarding approximately 12 ml of fix, resuspend the cell suspension in the 3 ml of fix remaining by flicking the tube.

3. Drop a single drop of the fixed preparation from a glass pipette onto a clean glass microscope slide, immediately followed with 1 drop of fresh 3:1 fix.

4. Allow the slide to air dry and examine it for metaphase spreads under phase contrast using a light microscope.

5. Mark the area of spread cells on the slide with a diamond pen.

6. Incubate the slides in 3:1 fix at room temperature for 30-60 minutes.

7. After air-drying, dehydrate the slides through a fresh 70%, 70%, 90%, 90% and 100% ethanol series for 1 minute each.

8. After air-drying, incubate the slides in acetone at room temperature for 10 minutes. Air-dry the slides.

9. If the slides are to be used the same day, incubate them at 65°C in an oven for at least 1 hour.

10. If the slides are to be used the next day, incubate them at 42°C o/n.

11. For long-term storage seal the slides in a box at room temperature.

2.10.4 Probe labelling by nick translation

Single-copy BAC, PAC, cosmid and fosmid DNA probes for use in FISH were labelled by nick translation by the following method:

1. For a 25 µl reaction to label 1 µg of DNA, add the following to a 1.5 ml microfuge tube on ice:

    - X µl sterile distilled water (to a final volume of 25 µl)

    - 2.5 µl 10× nick translation buffer

    - 1.9 µl nick translation dNTP mix

    - 0.7 µl Biotin-16-dUTP (or other hapten, see reagents and stains, 2.2, above)

    - 0.5 µl DNase I (1 µg/ml working solution)

    - 0.5 µl (5 U) DNA polymerase I

    - Y µl DNA (1µg)

2. Mix the tube contents and spin briefly in a microfuge to bring down the solution.

3. Incubate at 14˚C for 40 minutes to 1 hour.

4. To stop the reaction, add 2.5 µl of 0.5 M EDTA, mix well and transfer the tube to ice.

5. Mix 3 µl of the sample with 1 µl of 6× loading buffer and analyse by electrophoresis on a 1% agarose gel.

6. Meanwhile, precipitate the cut and labelled DNA by the addition of 2.5 µl of 3 M sodium acetate (pH 7.0) followed by 1 ml of 100% ice-cold ethanol. Mix the solution well and incubate at -70˚C for 30 minutes or -20˚C o/n.

7. Pellet the DNA by spinning in a microfuge at 12,000 × *g* for 10 minutes. Wash the pellet with 1 ml of 70% ice-cold ethanol and microfuge at 12,000 × *g* for 10 minutes.

8. Discard the supernatant and air-dry the pellet by incubating at 37°C for 25 minutes with the tube lid off.

9. Resuspend the DNA on ice in 10 µl of $T_{0.1}E$ so that the concentration of DNA will be approximately 100 ng/µl.

2.10.5 Hybridisation of human and gibbon single-copy probes and chromosome-specific paints onto human, pygmy chimpanzee, gibbon and lemur metaphases

Human and gibbon DNA probes and chromosome-specific paints were hybridised onto human, pygmy chimpanzee, gibbon and lemur metaphases by the following method:

Safety Note: Carry out all manipulations involving formamide in a fume cabinet.

1. Add the following to a 1.5 ml microfuge tube on ice:

    - 0.5 µl labelled DNA (30-50 ng)

    - 1 µl human C_0t-1 DNA (1 µg)

    - 14 µl hybridisation buffer

2. Mix the tube contents, and spin briefly in a microfuge to bring down the solution.

3. Denature the probe mix by incubating at 65°C for 10 minutes.

4. Pre-anneal the probe by incubating at 37°C for 20 minutes to 1 hour. Meanwhile, denature the metaphase-spread slides in 70% formamide at 65°C for 2 minutes.

5. Quench the denatured slides in 70% ice-cold ethanol for 2 minutes then dehydrate through a 70%, 70%, 90%, 90% and 100% ethanol series, allowing 1 minute in each.

6. Air-dry the slides

7. Pipette the probe mix onto the slide and cover it with a cover slip. Seal the edges of the cover slip with rubber cement.

8. Incubate the slides o/n at 37°C.

2.10.6 Hybridisation of human single-copy probes and chromosome-specific paints onto canine metaphases

Human DNA probes and chromosome-specific paints were hybridised onto canine metaphases by the following method:

1. Add the following to a 1.5 ml microfuge tube on ice:

    - 1.5 µl labelled DNA probe (90-150 ng)

    - 1.0 µl sonicated dog genomic DNA (10 µg)

    - 14 µl hybridisation buffer

2. Follow the protocol above *(2.10.5)* from steps 2. to 7.

3. Incubate the slides over two nights at 37˚C.

2.10.7 Hybridisation of canine chromosome-specific paints onto human metaphases

Canine chromosome-specific paints were hybridised onto human metaphases by the following method:

1. Add the following to a 1.5 ml microfuge tube on ice:

    - 1.5 µl labelled DNA probe (90-150 ng)

    - 1.0 µl human C_0t-1 DNA (1 µg)

    - 0.9 µl sonicated salmon sperm DNA (9 µg)

2. Ethanol-precipitate the tube contents and resuspend the DNA pellet in 14 µl of hybridisation buffer.

3. Follow protocol *(2.10.5)* from steps 2. to 7.

4. Incubate the slides over two nights at 37˚C.

2.10.8 Direct detection for fluorescently-labelled probes

After hybridisation, fluorescently-labelled probes were detected by the following method:

1. Pre-warm 3 coplin jars of 2× SSC and 2 coplin jars of 50% formamide/1× SSC to 42˚C. Remove the dried rubber cement sealing the cover slips and rinse the cover slips off in the first jar of 2× SSC.

2. Stringency wash the slides by incubating for 2× 5 minutes in 50% formamide/2× SSC at 42˚C.

3. Wash the slides by incubating for 2× 5 minutes in 2× SSC at 42˚C.

4. Rinse the slides in 4× SSC, 0.05% (v/v) Tween 20 for 5 minutes at room temperature.

5. Incubate the slides for 3 minutes in 50 ml DAPI stain in the dark.

6. Briefly rinse the slides in 2× SSC and quickly dehydrate through a fresh series of 70%, 70%, 90%, 90% and 100% ethanol. Air-dry the slides in the dark.

7. Apply 13 µl of citifluor AF1 antifade slide mount to each slide and overlay with a 22 x 32 mm cover slip.

8. Seal the cover slips with nail varnish and store in the dark.

2.10.9 Three-layer immunochemical detection for biotinylated probes

After hybridisation, biotinylated probes were detected imunochemically by the following three-layer method:

1. Pre-warm 3 coplin jars of 2× SSC and 2 coplin jars of 50% formamide/1× SSC to 42˚C. Remove the dried rubber cement sealing the cover slips and rinse the cover slips off in the first jar of 2× SSC.

2. Stringency wash the slides by incubating for 2× 5 minutes in 50% formamide/1× SSC at 42˚C.

3. Wash the slides by incubating for 2× 5 minutes in 2× SSC at 42˚C.

4. Incubate the slides in 4× TNFM at 37˚C for 10 minutes.

5. Meanwhile, dilute the detection reagents as follows (make 100 μl per slide plus at least 50 μl excess for each layer) in 1.5 ml microfuge tubes:

   - Avidin-FITC DCS or Avidin-Cy3 diluted 1/500 in 4× TNFM

   - Biotinylated anti-avidin diluted 1/250 in 4× TNFM

6. Incubate the detection solutions for 10 minutes at room temperature in the dark.

7. Pellet any debris by microfuging at 12,000 g for 10 minutes and use the supernatants.

8. Incubate the slides with 100 μl diluted Avidin-FITC or Avidin-Cy3 under a parafilm cover slip at 37˚C in a dark humidified box for 20 to 60 minutes.

9. Wash the slides for 3× 5 minutes in 4× TNFM at 42˚C.

10. Incubate the slides with 100 μl of diluted biotinylated anti-avidin for 20 to 60 minutes in a dark humidified box at room temperature.

11. Wash the slides for 3× 5 minutes in 4× TNFM at 42˚C.

12. Incubate the slides with 100 μl of diluted Avidin-FITC or Avidin-Cy3 for 20 to 60 minutes in a dark humidified box at room temperature.

13. Wash the slides for 2× 5 minutes in 4× TNFM at 42˚C.

14. Wash the slides for 2× 5 minutes in 4× SSC, 0.05% (v/v) Tween 20 at room temperature.

15. Follow method *(2.10.8)* from step 5.

2.11 Procedures for the construction of a genomic cosmid library

2.11.1 Partial restriction enzyme digestion and phosphatasing of high molecular weight genomic DNA

High molecular weight gibbon genomic DNA was partially digested with *Mbo*I and phosphatased with CIAP in preparation for cloning, by the following method:

Note: Care was taken to minimise shearing of the high molecular weight gibbon DNA by carrying out all manipulations slowly and with cut-off pipette tips.

1. Using a cut-off pipette tip transfer 30 μg of high molecular weight gibbon DNA (dissolved in $T_{0.1}E$) to a 1.5 ml microfuge tube. Add 0.2 μm filtered 10 mM Tris-HCl pH 8.0 to a total volume of 450 μl. Resuspend the DNA by stirring gently with the pipette tip.

2. Add 50 μl of 10× TAK buffer without SAM or DTT added and mix by stirring gently with the pipette tip.

3. Allow the DNA to fully equilibrate by incubating o/n at 4˚C.

4. Add 0.25 μl of 0.2 μm filtered 1 M DTT (to give a 0.5 mM final concentration) and mix by stirring gently with the pipette tip. Incubate for 30 to 40 minutes at 4˚C.

5. On ice, aliquot the equilibrated high molecular weight DNA solution into ten 1.5 ml microfuge tubes, so that tube 1 has 60 μl and tubes 2 to 10 have 30 μl (1.8 μg DNA) each.

6. Make a 1/50 dilution of *Mbo*I to 0.1 U/μl in 1× TAK buffer without SAM or DTT added.

7. Serially dilute the *Mbo*I in tubes 1 to 9 as follows:

    I. To tube 1, add 0.1 U (1 μl of 0.1 U/μl) of *Mbo*I and stir gently with the pipette tip twenty times.

II.   Using a cut-off tip transfer 30 μl of the solution from tube 1 to tube 2 and stir gently with the pipette tip twenty times.

III.   Repeat the procedure in step ii. up to tube 9, then discard 30 μl of solution from tube 9 so that the quantity of *Mbo*I per tube will be as follows:

- Tube 1   0.05 U

- Tube 2   0.025 U

- Tube 3   0.0125 U

- Tube 4   0.00625 U

- Tube 5   0.003125 U

- Tube 6   0.0015625 U

- Tube 7   0.00078125 U

- Tube 8   0.000390625 U

- Tube 9   0.0001953125 U

- Tube 10   no-enzyme control

8.   Incubate the tubes for 1 h at 37˚C in an oven.

9.   Inactivate the restriction enzyme by incubating for 15 minutes at 70˚C on a hot block. After cooling to room temperature transfer the tubes to ice for 2 to 3 h.

10. Dilute CIAP from 10 U/μl to 0.6 U/μl in 1× TAK buffer with no SAM or DTT added.

11. Add 0.6 U (1 μl) of CIAP to each reaction tube and incubate for 30 minutes at 37˚C in a water bath.

12. At room temperature, using a cut-off pipette tip, remove 10 μl from each digested and phosphorylated sample, mix with 2 μl of 6× loading buffer and analyse by electrophoresis on a 0.3% agarose/1× TAE gel. Include 100 ng of *Hind*III-digested

Bacteriophage λ DNA and 25 ng of undigested Bacteriophage λ DNA as DNA size markers[a]. Run the gel at 1.5 V/cm o/n at room temperature.

13. Meanwhile, to inactivate the CIAP, add 2.2 µl of 150 mM NTA to the remaining 20 µl of each sample and incubate for 20 minutes at 68˚C on a hot block.

14. Immerse the tubes in ice to cool rapidly and spin the tubes briefly in a microfuge to bring down the solution.

15. Add 1.2 µl of 5 M NaCl to each sample and gently mix well.

16. Add 60 µl of 100% ethanol to each sample and gently mix well.

17. Allow the DNA to precipitate by incubating the tubes o/n at -20˚C[b].

[a]Denature the size markers at 65˚C for 2 minutes before use.

[b]Digested DNA in 100% ethanol may be stored for several months at -20˚C.

2.11.2 Preparation and testing of Lawrist16 vector arms

Lawrist 16 vector arms were prepared and tested by the following methods:

Safety Note: Carry out all manipulations involving phenol, chloroform and ether in a fume cabinet.

Method A: Preparation of Lawrist 16 vector arms

1. In a 1.5 ml microfuge tube combine the following:

    - 10 µl 10× NEB ScaI buffer

    - X µl DNA (20 µg)

    - Y µl sterile distilled water (to a total volume of 100 µl)

    - 9 µl (90 U) ScaI

2. Mix the tube contents, and spin briefly in a microfuge to bring down the solution.

3. Incubate for 2 h at 37°C in an oven.

4. Remove a 1 µl aliquot and mix it with 2 µl of 6× loading buffer. Analyse the digestion products by electrophoresis on a 0.7% agarose/1× TBE gel. Include 200 ng of undigested Lawrist 16 as a control, and *Hind*III-digested Bacteriophage λ DNA as a size marker.

5. From the remainder of the *Sca*I digestion mix remove a 5 µl aliquot, (*Sca*I control) and store at –20°C for use in Method B (see below).

6. Meanwhile, to the remainder of the *Sca*I digestion add 1× LSRE buffer to a volume of 370 µl.

7. Add 35 µl (35 U) of CIAP, pipette up and down a few times to mix and incubate for 45 minutes at 37°C in a water bath.

8. To inactivate the CIAP add 45 µl of 150 mM NTA and incubate for 25 minutes at 68°C in a hot-block.

9. Extract the DNA with an equal volume of 1:1 phenol/chloroform, then with chloroform alone, each time re-extracting the organic phase with an equal volume of TE. Finally, extract the aqueous phase with ether until no whiteness appears at the phase interface.

10. Place the tube at 68°C for a few minutes with the cap open to evaporate the ether, then precipitate the DNA o/n at –20°C with 0.1 volume of 3 M sodium acetate pH 5.2 and 2 volumes of absolute ethanol.

11. Pellet the DNA in a microfuge at 12,000 *g* for 15 minutes at room temperature and discard the supernatant. Wash the DNA pellet with 1 ml of 70% ethanol, then microfuge for 2 minutes, discard the supernatant and air-dry the pellet.

12. Dissolve the DNA in 177.5 µl of TE. Remove a 5 µl aliquot (*Sca*1/CIAP control) and store at –20°C for use in Method B.

13. To the remaining *Sca*I- and CIAP-treated DNA, add 20 µl of 10× *Bam*HI buffer and
    7.5 µl (150 U) of *Bam*HI. Pipette up and down a few times to mix.

14. Incubate for 90 minutes at 37°C in a water bath, then cool the reaction to room
    temperature.

15. Remove a 2.5 µl aliquot and mix with 2 µl of 6× loading buffer and analyse the
    digestion products by electrophoresis on a 0.7% agarose/1× TBE gel. Include *Hind*III-
    digested Bacteriophage λ DNA as a size marker.

16. To the remainder of the *Sca*I/CIAP/*Bam*HI reaction mix, carry out organic extractions
    and precipitate the DNA as described in steps 9, 10 and 11.

17. Dissolve the vector arms in 50 µl of TE and remove a 5 µl aliquot (*Sca*I/CIAP/*Bam*HI
    control) for use in Method B. Store the remainder of the prepared arms at –70°C in 10
    µl aliquots (approximately 200 ng/µl) in 0.5 ml microfuge tubes.

Method B: Testing of Lawrist 16 vector arms

1. To the 5 µl sample from step A5 (*Sca*I control) add 45 µl of TE.

2. Carry out organic extractions and precipitate the DNA as described in steps A9, A10
   and A11.

3. Dissolve the *Sca*I control DNA in 5 µl of TE.

4. To this sample and those from steps A12 (*Sca*1/CIAP control) and A17
   (*Sca*I/CIAP/*Bam*HI control) add 11 µl of TE.

5. Add 2 µl of 10× Nizetic ligase buffer and 0.2 µl of 100 mM ATP to each sample. Dilute
   the stock T4 DNA ligase to 50 U/µl in 1× Nizetic ligase buffer.

6. Divide each sample from step 5 into two 9 µl aliquots.

7. To one add 1 µl (50 U) of T4 DNA ligase. To the other, add 1 µl of 1× Nizetic ligase buffer.

8. Incubate o/n at 14˚C.

9. Analyse the quality of the vector arms by electrophoresis on a 0.7% agarose/1× TAE gel. Include *Hind*III digested λ DNA as a size marker and undigested vector DNA as a gel control.

2.11.3 Ligation and packaging of partially -digested DNA

Partially digested gibbon genomic DNA (from Method *2.11.1)* was ligated to Lawrist 16 vector arms (as prepared in Method *2.11.2*) and packaged into infective λ Bacteriophage particles by the following methods:

Method A: Ligation of vector arms to partially digested DNA

1. Pellet the partially digested DNA (stored in 100% ethanol at –20˚C) from Method *2.11.1* in a microfuge at 12,000 *g* for 15 minutes at room temperature. Align the tube hinge in the microfuge rotor so that it is possible to predict where the DNA pellet will be.

2. Carefully remove the supernatant with a P200 pipette tip. The DNA pellet will be visible as "speckles" at the bottom and up the side of the tube.

3. Wash the DNA pellet with 1 ml of 70% ethanol and microfuge at 12,000 *g* for 7 minutes.

4. Carefully remove the supernatant with a P1000, then a P200 pipette tip. Allow the DNA pellet to air dry with the tube lid off for 10 minutes at room temperature.

5. Resuspend the DNA in 10 µl of TE by working beads of TE over the surface of the DNA with a pipette tip. Incubate the sample on ice for 4 to 5 h, working the TE over the DNA pellet surface every hour.

6. Set up Test and Control ligation reactions as follows:

    Test reaction:

    - X µl vector arms (80-1520 ng)

    - Y µl genomic DNA (50-800 ng)

    - Z µl TE (to a final volume of 16 µl)

    - 1.6 µl 10× ligase buffer

    - 1.4 µl T4 DNA ligase (400 U/µl)

    - 1.0 µl 6 mM ATP

    Control reaction:

    - X µl vector arms (400 ng)

    - Y µl TE (to a final volume of 8 µl)

    - 0.8 µl 10× ligase buffer

    - 0.7 µl T4 DNA ligase

    - 0.5 µl 6 mM ATP

7. Flick the tubes to mix the reactants and spin briefly in a microfuge to bring down the solution. Incubate o/n at 14°C on a PCR block.

8. Either store the ligation products at -70°C or (if packaging the same day) keep them on ice.

Method B: Packaging ligated DNA

1. Package ligated DNA from Method A for 2 h using the Stratagene Gigapack Gold II, IIXL or Gold III packaging extract systems, according to the manufacturer's instructions.

2. Stop the reaction by adding 500 µl of lambda diluent, followed by 132 µl of 5× SM. Mix the samples gently by inversion.

3. Assess the library titre of each packaging reaction (see *2.11.4* below).

4. Freeze small (30 μl) aliquots of the packaged bacteriophage on dry ice, and store them at –70˚C.

2.11.4 DH5αMCR E.coli plating cell preparation

DH5αMCR cells were prepared by the following method:

1. Allow an antibiotic-free LB agar plate to equilibrate to room temperature.

2. Collect a DH5αMCR glycerol stock from -70˚C onto dry ice.

3. In a safety cabinet, gently scratch the surface of the (still frozen) glycerol stock with the tip of a sterile cocktail stick and streak the cells across the plate surface.

4. Take a fresh cocktail stick and cross hatch across the first streak. Repeat this process with a fresh cocktail stick each time, cross-hatching across the previous streaks, until the bottom of the plate is reached.

5. Allow the cell suspension to dry on the surface of the plate for a few minutes with the lid off.

6. Replace the lid and incubate the plates inverted at 37˚C o/n.

7. In a safety cabinet, use a sterile loop to transfer a small single separate colony to 40 ml of LB media in a sterile 250 ml conical flask.

8. Incubate the 250 ml flask with the lid loosened, shaking, o/n at 37˚C. Include a centrifuge tube containing 40 ml of LB media alone to act as a contamination control.

9. Transfer the cell suspension to a 50 ml centrifuge tube. Pellet the cells by centrifugation at 4,500 *g* for 15 minutes at room temperature.

10. Discard the supernatant into disinfectant and gently resuspend the cell pellet in 20 ml of autoclaved 10 mM MgSO$_4$.

11. Store the cells at 4°C for up to four weeks.

2.11.5 Plating libraries on E. coli and assessing library titres

Library titres were assessed by the following method:

1. In a safety cabinet set up seven transduction reactions in 1.5 ml microfuge tubes as follows:

    I.   1 µl library packaged extract   + 99 µl lambda diluent

    II. 50 µl packaged self-ligation   + 50 µl lambda diluent

    III. 10 µl 5× SM   + 90 µl lambda diluent

    IV. 100 µl lambda diluent

    V.   100 µl DH5αMCR plating cells

    VI. 30 µl lambda diluent alone

    VII. 20 µl packaged extract alone

2. Add 100 µl of fresh (less than four weeks old) DH5αMCR plating cells to reactions i. to iv. and mix the tube contents gently. Incubate all the tubes at 37°C in a hot block for 20 minutes.

3. Dilute the infected cells in reaction tubes i. and ii. with 1 ml of LB media each and invert several times to mix.

4. Incubate the reactions for 45 minutes at 37°C in a hot block.

5. Pellet the cells by microfuging tubes i. and ii. for 2 minutes at 6000 rpm.

6. Remove the supernatant with a pipette tip leaving approximately 30 µl in the tube.

7. Gently resuspend each cell pellet in the residual media by pipetting up and down.

8. Using a sealed, bent glass pasteur pipette, spread each sample on a separate 8x20 cm plate containing 1% LB agar + 30 µg/ml kanamycin.

9. Allow 10 minutes for the plate surface to air-dry with the lid off.

10. Replace the lids and incubate the plates upside down for approximately 16 h, at 37°C.

11. Count the number of colonies on each plate. (The test plates iii. to vii. should have no colony growth.)

12. Assess the library titre by multiplying the number of colonies on plate i. by the total volume (µl) of the library (see *2.11.4*, Method B), to give the number of potential colonies of that library. Although some colonies from the packaging reaction of self-ligated vector arms are to be expected, the number of colonies should only be a fraction of that obtained by the genomic DNA packaging.

2.11.6 Plating libraries onto filters and their replication for screening by hybridisation

Method A:

Master filters, each supporting approximately 20,000 colonies from a library, were set up by the following method:

Note: Carry out all manipulations up to step 14 in a safety cabinet.

1. Estimate the volume of packaged library extract required to generate 20,000 colonies (see *2.11.5* above).

2. Calculate the number of transduction reactions required, allowing 10 µl of packaged extract per reaction (e.g. if 60 µl of packaged extract is required for 20,000 colonies, then 6 transduction reactions are required, each containing 10 µl of packaged extract and 90 µl of lambda diluent.)

3. For each transduction reaction aliquot 90 µl of lambda diluent into a 1.5 ml microfuge tube.

4. Add 10 µl of packaged library extract and rinse the pipette tip by pipetting up and down gently 3 or 4 times.

5. Add 100 µl of DH5αMCR cells and mix gently by pipetting up and down twenty times.

6. Set up control reactions containing lambda diluent alone, DH5αMCR cells alone and LB broth alone.

7. Incubate all the tubes at 37°C in a hot block for 20 minutes.

8. Add 1ml of LB broth to each tube and invert them several times to mix.

9. Incubate for 45 minutes in a hot block at 37°C.

10. Pellet the cells by microfuging for 2 minutes at 6000 r.p.m.

11. Discard the supernatant from all but one of the tubes where approximately 200 µl should be retained.

12. Pool and resuspend all the pellets in the 200 µl of supernatant.

13. Using a sealed, bent pasteur pipette, spread the cells (or control reaction) onto a Hybond N+ (Amersham) nylon transfer membrane (8x20 cm) supported on a 1% LB agar plate including 30 µg/ml of kanamycin.

14. Allow the membrane surface to air dry for 10 minutes.

15. Incubate the plates upside down for approximately 16 h, at 30°C and then for approximately 7 h at 37°C, until the colonies have grown to 0.1-0.5 mm in diameter.

16. Transfer the plates to 4°C to delay colony growth. Store the plates at 4°C until further use.

Method B:

Master filters were replicated for hybridisation by the following method:

Note: Carry out all filter manipulations in a safety cabinet.

1. Bring the plate and filter to be replicated to room temperature (the colonies are difficult to replicate if too cold).

2. Using clean forceps, carefully peel the filter off the plate and lay it colony-side up on a clean plate of glass.

3. Label the replica filter appropriately and carefully place it writing-side down on top of the master filter. Allow a slight over-hang of the replica filter over the master filter to ease their separation in step 6.

4. To assist in cell transfer, apply firm pressure to the filters (e.g. by placing a 2.5 kg metal thermal cycler insert on top) for 2 to 4 minutes.

5. Using a 24-guage syringe needle, punch holes at the corners of both filters in distinct patterns to allow alignment of the master and replica when picking colonies at a later date.

6. Transfer the replica filter to a fresh 1% LB agar plate including 30 µg/ml of kanamycin.

7. Incubate the replica plate upside down for 3 to 4 h at 37°C, until the colonies have grown to an appropriate size.

8. Transfer the master filter to a fresh 1% LB agar plate containing 30 µg/ml kanamycin and 25% (v/v) glycerol. Incubate the plate upside down for 2 h at 37°C.

9. Wrap the plate tightly in Parafilm, and store it upside down at -70°C. This plate will be used for picking positive colonies.

10. Process the replica filter as described in Method *2.11.7* before use for hybridisation screening.

2.11.7 Library filter processing

Library filters were processed by the following method:

1. Place two sheets of 3MM paper on two biohazard trays. Saturate one tray with 10% SDS and the second tray with denaturation solution.

2. Ensure that there are no air bubbles under the saturated paper, and drain away the excess solution.

3. Place the filters on the 3MM paper with 10% SDS, colony-side up for 4 minutes.

4. Transfer the filters onto the tray with denaturation solution for 10 minutes.

5. Air-dry the filters on a clean piece of 3MM for 10 minutes.

6. Neutralise the filters by washing in 1 l of neutralisation solution for 5 minutes on an orbital shaker.

7. Repeat step 6.

8. Wash the filters in 1 l of 50 mM Tris-HCl pH 7.4/ 0.15 M NaCl (a 1/10 dilution of neutralisation solution) for 5 minutes, shaking on an orbital shaker.

9. Rinse the filters in 1 l of 2× SSC/ 0.1% SDS for 5 minutes, shaking on an orbital shaker.

10. Rinse the filters in 1 l of 2× SSC for 5 minutes, shaking on an orbital shaker.

11. Rinse the filters twice in 1 l of 50 mM Tris-HCl pH 7.4 for 5 minutes each, shaking on an orbital shaker.

12. Air-dry the filters, colony-side up on clean 3MM paper.

13. Cross-link the DNA to the filters by exposing the filters colony-side down on Saran-wrap to UV light on a transilluminator for 2 minutes.

14. Store the filters at room temperature in a sealed plastic bag.

2.11.8 Filter Screening

Method A: PCR-labelling of gibbon STSs

1. Using the appropriate primer pairs, perform a PCR using gibbon genomic DNA as a template. Electrophorese the products on a 2.5% agarose gel and stain with ethidium bromide to visualise the bands.

2. Over a UV transilluminator wearing a face shield and gloves, excise each product and transfer to 100 µl of $T_{0.1}E$ in a 1.5 ml microfuge tube.

3. Store o/n at 4˚C.

4. Make a PCR master mixture of an appropriate volume by combining multiples of the following single reaction volumes:

    - 6.83 µl $T_{0.1}E$

    - 1.5 µl 10× Perkin Elmer PCR buffer (Perkin Elmer)

    - 0.12 µl (0.6 U) AmpliTaq polymerase

    - 0.3 µl mixture of 3 unlabelled dNTPs (5 mM each)

5. Aliquot 8.75 µl of this mixture into PCR tubes and add 5 µl of $T_{0.1}E$ from the agarose slice from step 2.

6. Add 0.75 µl of primer mixture (100 ng/µl each). Add one drop of mineral oil.

7. With the protection of appropriate safety shielding, pipette 0.5 µl of α-^{32}P dCTP (3,000 Ci/mmol) under the mineral oil and mix slightly with the pipette tip.

8. Place the tubes into a thermal cycler and perform PCR using the following cycling profile:

    - 94˚C for 5 minutes

    followed by 20 cycles of:

    - 93˚C for 30 seconds

- 55˚C for 30 seconds

- 72˚C for 30 seconds

followed by:

- 72˚C for 5 minutes

Method B:

1. Tightly roll each processed filter for analysis by hybridisation and insert them in pairs into 15 ml polypropylene tubes. Add 15 ml of prehybridisation buffer to the tubes, replace the lids and incubate for at least 1 h at 65˚C with shaking. (If a large number of filters are to be probed, these may be stacked colony-side up and probed in a small sandwich box with a tight lid.)

2. Prepare the following competition mixture by combining the following in a 1.5 ml microfuge tube:

- 125 µl 20× SSC

- 125 µl (1.25 mg) sonicated human placental DNA (stock 10 µg/µl)

- X µl (50 µg) gibbon genomic DNA

- Y µl $T_{0.1}E$ (to a final volume of 500 µl)

- 15 µl radiolabelled probe (from Method A above)

For multiple probes to be hybridised, pool together in the competition mix and modify the volume of $T_{0.1}E$.

3. Mix by inverting the tube and place in a boiling water bath for 5 minutes.

4. Snap chill on ice, then add to the filters in the prehybridisation solution.

5. Incubate o/n at 65˚C with shaking.

Method C: Post-hybridisation filter washes and autoradiography

1. Taking the appropriate precautions to prevent radioactive exposure, discard the hybridisation buffer and briefly rinse each filter in 2× SSC.

2. Wash the filters in 1 l of 2× SSC for 5 minutes at room temperature on an orbital shaker.

3. Wash the filters twice for 30 minutes in 1 l of 0.5× SSC/ 1% sarcosyl at 65°C on an orbital shaker.

4. Wash the filters twice for 5 minutes in 1 l of 0.2× SSC at room temperature on an orbital shaker.

5. After blotting off any excess liquid, lay the filters face-down on a clean sheet of Saran-wrap. Place a second sheet of Saran-wrap on top of the filters and fold over the edges to prevent the filters from over-drying.

6. Expose the filters to () film in an autoradiograph cassette with a () screen for at least 4 h at −70°C.

2.11.9 Colony PCR

Library colonies were analysed for the presence of STSs by colony PCR using the following method:

1. In a safety cabinet, touch a sterile toothpick onto a well-separated colony.

2. Suspend the cell sample by swizzling the toothpick 100 µl of $T_{0.1}E$ in a 96-well microtitre plate. Store the microtitre plate at 4°C until use.

3. Make a colony PCR master mixture of an appropriate volume by combining multiples of the following single reaction volumes:

   - 6.175 µl $T_{0.1}E$

   - 1.5 µl 10× NEB PCR buffer

   - 0.75µl dNTP

- 0.495 µl BSA (5 mg/ml stock)

- 0.21 µl 2-mercaptoethanol (1/20 stock)

- 0.12 µl (0.6 U) AmpliTaq Polymerase

4. Aliquot 9.25 µl per reaction in a microtitre plate and add 5 µl of $T_{0.1}E$ from each colony sample.

5. Add 0.75 µl of primer mixture (100 ng/µl) and seal the wells with a rubber plate-sealer.

6. Place the plate on a thermal cycler and perform PCR using the following cycling profile:

- 94˚C for 5 minutes

followed by 20 cycles of:

- 93˚C for 30 seconds

- 60˚C for 30 seconds

- 72˚C for 30 seconds

followed by:

- 72˚C for 5 minutes

7. Add 3µl of 6x loading buffer to each reaction well. Analyse 10 µl of the PCR products by electrophoresis on a 2.5% agarose gel.

2.12 Procedures for the construction of a vectorette library from a gibbon cosmid clone

2.12.1 Restriction enzyme digestion of a gibbon cosmid clone

In preparation for library construction, gibbon cosmid clones were restriction digested by the following method:

1. In a 1.5 ml microfuge tube, mix the following:

94

- X µl sterile distilled water (to a final volume of 18 µl)

- 2 µl 10× manufacturer's recommended restriction enzyme buffer

- 0.2 µl 100 mM spermidine

- Y µl (200 ng) DNA

2. Pipette the tube contents up and down to mix and briefly spin in a microfuge to bring the liquid down.

3. Add 1 µl of restriction enzyme and pipette up and down to rinse the tip.

4. Incubate for 1 h at 37˚C in an oven.

5. Add a further 1 µl of restriction enzyme, again pipetting up and down to rinse the tip.

6. Incubate for 2 h at 37˚C in an oven.

7. At room temperature, remove 10 µl of the digested sample and mix it with 3 µl of 6× loading buffer. Analyse the sample by electrophoresis on a 1% agarose gel. Include a control sample of 150 ng of uncut DNA on the gel.

8. To the remainder of the digested DNA, add 40 µl of sterile distilled water.

9. Add 5 µl of 3 M sodium acetate and mix well. Add 100 µl of ice-cold absolute ethanol and mix well. Incubate o/n at –20˚C to precipitate the DNA.

2.12.2 Ligation of digested cosmid DNA to vectorette bubbles

Digested gibbon cosmid DNA was ligated to annealed vectorette bubbles by the following method:

1. Pellet the DNA from *2.12.1* by microfuging at 12,000 *g* for 15 minutes at room temperature. Discard the supernatant and wash the pellet with 500 µl of ice-cold 70% ethanol.

2. Microfuge at 12,000 g for 10 minutes, discard the supernatant and air-dry the DNA by incubating at room temperature for 30 minutes with the tube lid open.

3. Resuspend the DNA in 100 µl of 1× vectorette ligation buffer on ice.

4. Add 10 µl of 1 pmol/µl annealed vectorette bubbles (appropriate for the restriction enzyme used) and pipette up and down to mix.

5. Add 1.1 µl of rATP, followed by 0.5 µl (2.5 U) of T4 DNA ligase (Boehringer Mannheim).

6. Pipette up and down gently to mix, then incubate for 1 h at 37˚C in a hot block.

7. Dilute the reaction by adding $T_{0.1}E$ to a final volume of 500 µl.

8. Store the vectorette library frozen at –20˚C.

2.12.3 Rescue of unknown gibbon sequences by vectorette PCR

Unknown gibbon sequences were rescued by vectorette PCR by the following method:

1. Make a vectorette PCR master mixture of an appropriate volume by combining multiples of the following single reaction volumes:

   - 9.425 µl $T_{0.1}E$

   - 1.5 µl 10× NEB PCR buffer

   - 0.75µl dNTP

   - 0.495 µl BSA (5 mg/ml stock)

   - 0.21 µl 2-mercaptoethanol (1/20 stock)

   - 0.12 µl (0.6 U) AmpliTaq Polymerase

2. Aliquot 13.25 µl per reaction in a microtitre plate and add 1 µl of vectorette library template.

3. Add 0.75 µl of each primer (100 ng/µl).

4. Place the plate on a thermal cycler and perform PCR using the following cycling profile:

- 94°C for 5 minutes

  followed by 35 cycles of:

- 94°C for 30 seconds

- 55°C/60°C/65°C for 30 seconds

- 72°C for 3 minutes

  followed by:

- 72°C for 5 minutes

5. Add 3 µl of 6× loading buffer to each microtitre well, and analyse 10 µl of the vectorette PCR products by electrophoresis on a 2.5% agarose gel.

6. With the agarose gel on a transilluminator, excise vectorette PCR product bands of appropriate size and specificity using sterile inoculating loops. Transfer the excised gel pieces to PCR tubes containing 50 µl of $T_{0.1}E$, and store at 4°C.

CHAPTER 3

Reciprocal Zoo-FISH analysis of DNA homologous to human chromosome 22 in the domestic dog and Siamang gibbon

3.1   Introduction

3.1.1 The Domestic Dog

3.1.2 The Siamang Gibbon

3.2   Establishing the Standard Karyotype of the Domestic Dog

3.2.1   Establishing the canine flow karyotype

3.2.2 Characterisation of canine chromosome specific paint probes

3.2.3 Standardisation of the canine DAPI-banded karyotype

3.3   Reciprocal chromosome painting analysis between human and the dog

3.3.1 Generation of human chromosome-specific paints

3.3.2   Hybridisation of human chromosome 22 specific paint onto dog metaphase chromosomes

3.3.3 Reciprocal hybridisation of dog chromosome paints onto human metaphase chromosomes

3.4   Reciprocal chromosome painting analysis between human and gibbon

3.4.1   Hybridisation of human chromosome 22 paint onto gibbon metaphase chromosomes

3.4.2   Generation of the Siamang flow karyotype and chromosome isolation

3.4.3 Identification of the Siamang chromosome 18 specific paint

3.4.4   Hybridisation of Siamang chromosome 18 paint onto human metaphase chromosomes

3.4.5   Hybridisation of human chromosome 22, 16, 5 and 2 paints onto Siamang metaphase chromosomes

3.5   Discussion

3.1 Introduction

The aim of the work reported in this chapter was to carry out reciprocal zoo-FISH analyses of the metaphase chromosomes of the dog and gibbon to identify regions corresponding to evolutionarily conserved chromosome segments (ECCS) in material homologous to human chromosome 22.

For the dog, the first step towards a reciprocal zoo-FISH study was the production of a bivariate flow karyotype, a panel of canine chromosome-specific paints and the establishment of a standard karyotype of the dog, which are described in this chapter. At the time of this work I was involved in a collaboration to standardise the dog karyotype, and the information generated by that collaboration is bound into this thesis (Breen, *et al.,* 1999).

For the gibbon study, it was necessary to generate metaphase chromosome spreads from cultured cells, a bivariate flow karyotype and a chromosome 18-specific paint from flow-sorted chromosomes, which are described here.

RESULTS

3.2 Establishing the Standard Karyotype of the Domestic Dog

3.2.1 Establishing the canine flow karyotype

Before staining for flow cytometric analysis, chromosomes may be isolated from a range of cell types, which must be healthy and growing optimally. A healthy, dividing population of cells is arrested in metaphase before chromosome isolation, by the addition of a spindle inhibitor (Colcemid). For the chromosome preparation, arrested cells are swollen in a hypotonic buffer, then the cell membrane is removed using a combination of

detergent and mechanical disruption to release the chromosomes into a stabilising buffer (Sillar and Young, 1981, Ross and Langford, 1997).

Peripheral blood lymphocytes from karyotypically normal Red Setter dogs were stimulated to divide in short-term cultures with phytohaemagglutinin and pokeweed mitogen. Previous protocols for the *in vitro* stimulation of isolated lymphocytes had often relied on the use of only one mitogen. The two mitogens used in this study had a synergistic effect on each other. After incubation with Colcemid the mitotic indexes of the cultures were high (up to 35%) which led to good chromosome preparations.

Chromosome suspensions were prepared and stained with Hoecsht 33258 and Chromomycin A3 and analysed on a dual-laser Elite ESP flow cytometer (Coulter Electronics). The chromosomes were resolved by DNA content and AT to GC base pair composition. Despite the complexity of the canine karyotype, the good quality chromosome preparations coupled with the high calibre instrumentation available resulted in the first-reported production of reproducible high-resolution bivariate flow karyoytpes of the dog from short-term cultures of peripheral blood lymphocytes (Langford, *et al* 1996, bound into this thesis).

The bivariate flow karyotype of a male red setter dog is shown in Figure 3.1. In the male dog, the 76 autosomes and two sex chromosomes were resolved into 32 peaks. These were labelled arbitrarily from A to Z, followed by aa to ff.

The number of homologues represented by a flow karyotype peak is indicated by the relative number of chromosome events in that peak. Using the sorter workstation software it is possible to display the number of events as a count frequency. The count frequency of a normal chromosome peak representing two homologues of the same size will be double that of the chromosome X and Y peaks in a male flow karyotype, which represent only one homologue each.

Figure 3.1 (next page) Bivariate flow karyotype of the male dog. The expanded panel shows the karyotype for the smaller chromosome peaks (Q to ff)
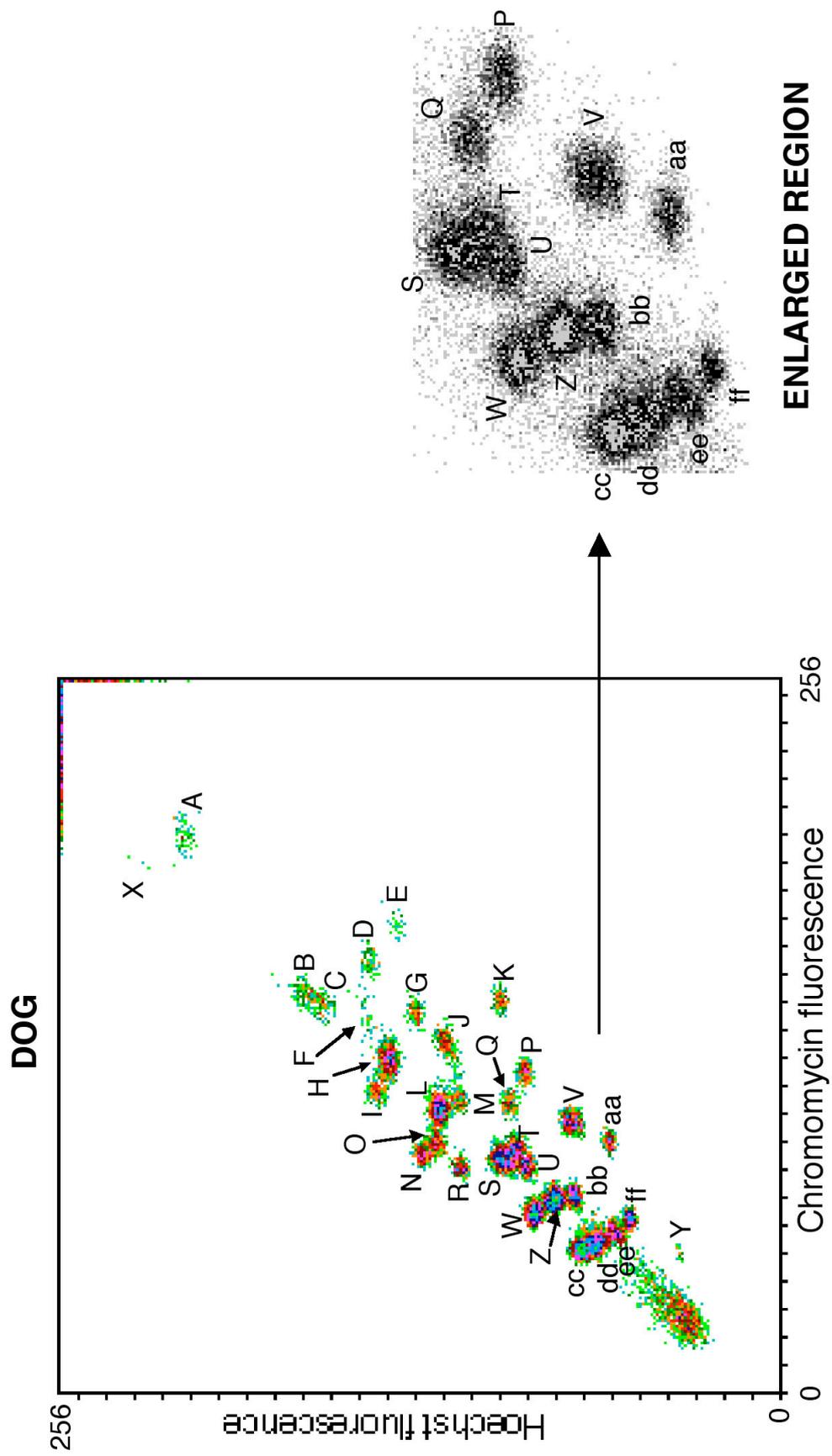
The sum of the count frequencies for all the peaks in the male flow karyotype was divided by the total number of chromosomes (78) to calculate the count frequency of one chromosome homologue. The number of homologues represented by each peak in the male flow karyotype was then estimated (to the nearest integer) from the relative count frequencies of each peak and are presented in Table 3.1 (see over).

Eight of the 32 peaks contained four chromosome homologues, which was not surprising considering the similarity in size of many of the smaller autosomes. The peaks representing chromosomes X and Y (peaks X and Y, respectively) were identified by their position in the flow karyotype (size) and their relative count frequencies, which were each equivalent to a single chromosome homologue.

In the male flow karyotype, all the chromosomes were accounted for in the peaks (39 pairs).

Table 3.1 The number of chromosome homologues represented by each peak in the male dog flow karyotype.

Peak	Number of homologues	Peak	Number of homologues
A	2	Q	2
B	2	R	2
C	2	S	4
D	2	T	2
E	2	U	2
F	2	V	4
G	2	W	4
H	4	X	1
I	2	Y	1
J	4	Z	4
K	2	aa	2
L	4	bb	2
M	2	cc	4
N	2	dd	2
O	2	ee	2
P	2	ff	2

In the female flow karyotype, an additional peak (labelled V') was observed close to peak V with a slightly reduced DNA content. (In the male flow karyotype peak V contained 4 chromosome homologues). The female peak V' contained a count frequency equivalent to a single chromosome homologue, and the concomitant count frequency for female peak V was equivalent to only three chromosome homologues. It is most likely that the

homologue in peak V' was a smaller size to the other peak V chromosomes due to a heterochromatic polymorphism (heteromorphism).

3.2.2 Characterisation of canine chromosome specific paint probes

Five hundred chromosomes were sorted from each resolved peak from the male and female flow karyotypes into sterile 0.5 ml PCR tubes for the generation of biotinylated chromosome specific paint probes by DOP-PCR.

Figure 3.2 (next page) Chromosome painting pattern for bivariated peaks A-T. The DAPI counterstain is pseudocoloured in red and the hybridisation signal is pseudocoloured in green.

Figure 3.3 (subsequent page) Chromosome painting pattern for bivariated peaks U-ff. The DAPI counterstain is pseudocoloured in red and the hybridisation signal is pseudocoloured in green.
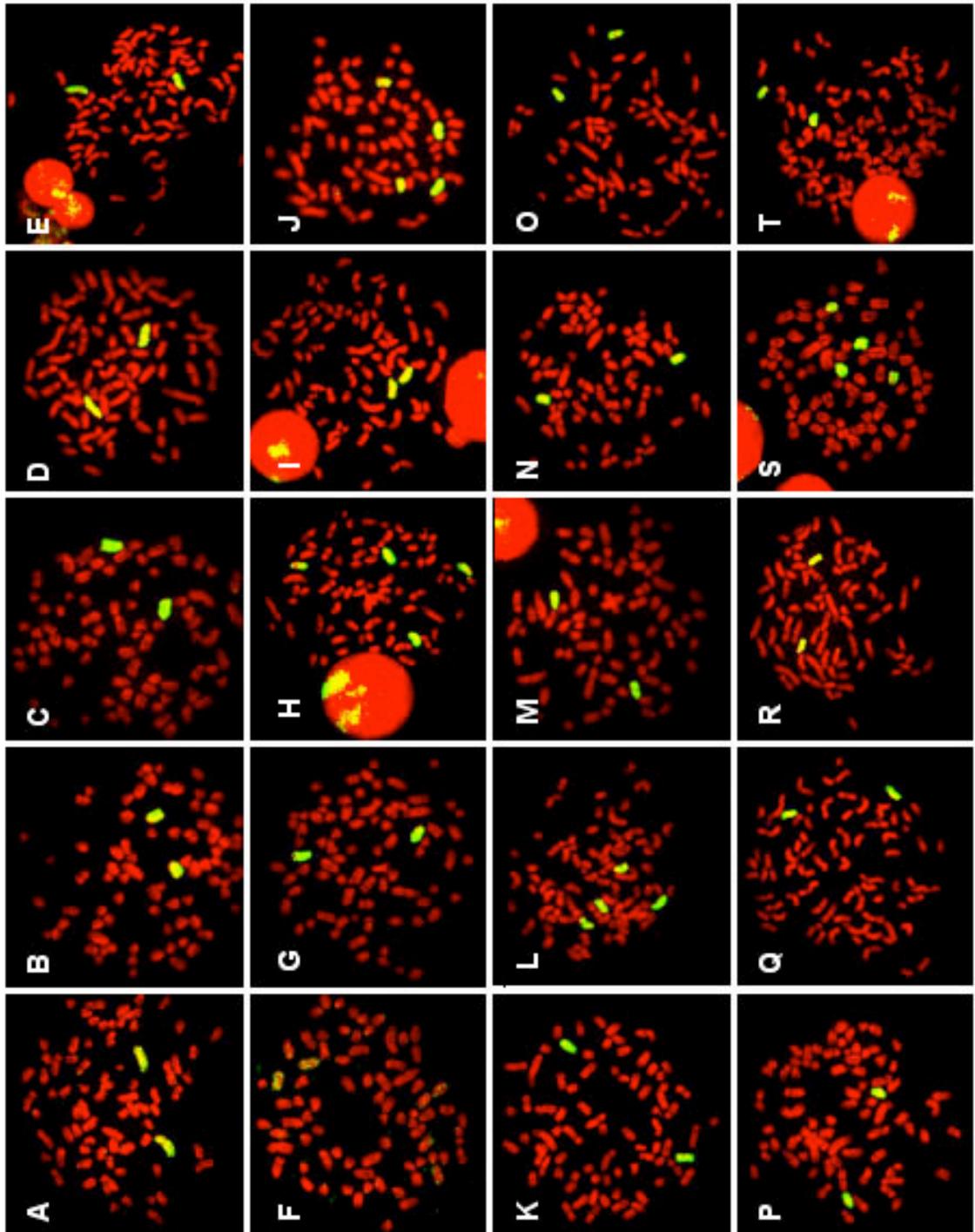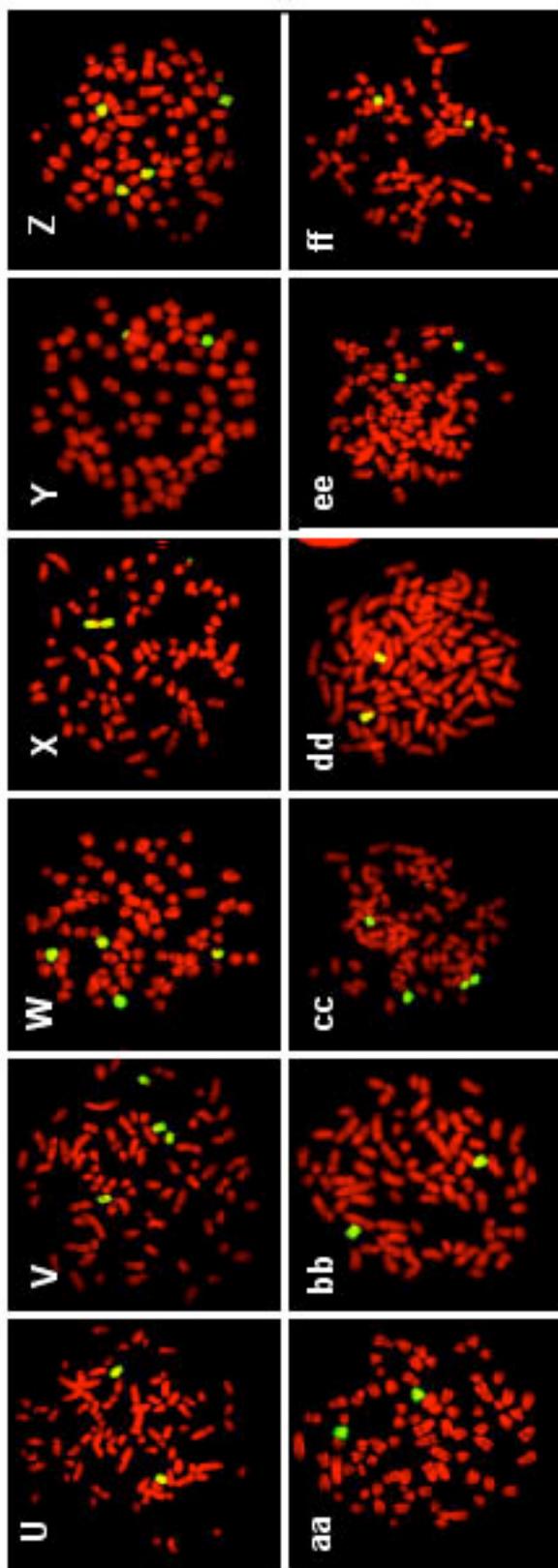
The canine chromosome specific paints were assessed by hybridisation to the metaphase spreads of a normal male red setter dog (Figures 3.2 and 3.3). Images from 20 metaphase spreads were acquired and processed using a FISH workstation comprising a fluorescence microscope (Axiophot, Carl Zeiss) equipped with an 83000 triple dichroic mirror block and separate excitation filter set (Chroma Technologies), a cooled CCD camera (KAF1400, Photometrics, Tucson, AZ, USA) driven by a Macintosh PowerMac 8100 computer and dedicated software (SmartCapture, Digital Scientific, Cambridge, UK).

The estimated number of homologues contained in each peak was confirmed in each case by chromosome painting. Twenty-two of the 32 paints (generated from the male flow karyotype) each hybridised to individual pairs of chromosomes. Eight of the paints each hybridised to two pairs of chromosomes and two of the paints represented one of the sex chromosomes each (Figure 3.2 and 3.3). The paint generated from the female peak V' hybridised to one homologous pair, which was also one of the pairs hybridised by the paint from peak V.

In most cases, hybridisation to the metaphase chromosomes was strong and specific although weak labelling just above background was found on six additional homologues for peak F and two additional homologues each for peaks J, O, R, T, U, cc and dd. This can be explained by the proximity of other flow karyotype peaks to F, J, O, R, T, U, cc and dd. It is quite possible that in this situation, co-sorting of a small proportion of chromosomes from the nearby peak(s) occurred, and were represented to a low degree in the paint.

The chromosome X-specific paint also hybridised to a region of Yp and, similarly, the Y-specific paint hybridised to a region of Xp. A similar pattern of hybridisation is seen with human X and Y paints on metaphase spreads. It is assumed that these Xp and Yp regions that show homology are the short arm pairing segments (pseudoautosomal regions), which pair and recombine during male meiosis.

The image of each metaphase spread was processed using a high-pass digital filter (SmartCapture, Digital Scientific, Cambridge, UK) to reveal enhanced DAPI bands (Figure 3.4).

Figure 3.4 (next page) Hybridisation image of the paint from flow karyotype peak K and the corresponding enhanced DAPI banded metaphase image are shown. Analysis of the enhanced DAPI band image, revealed that peak K contains dog chromosome 9.
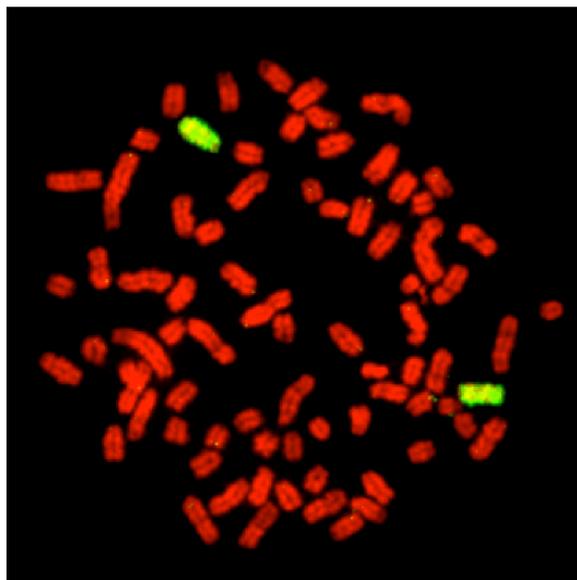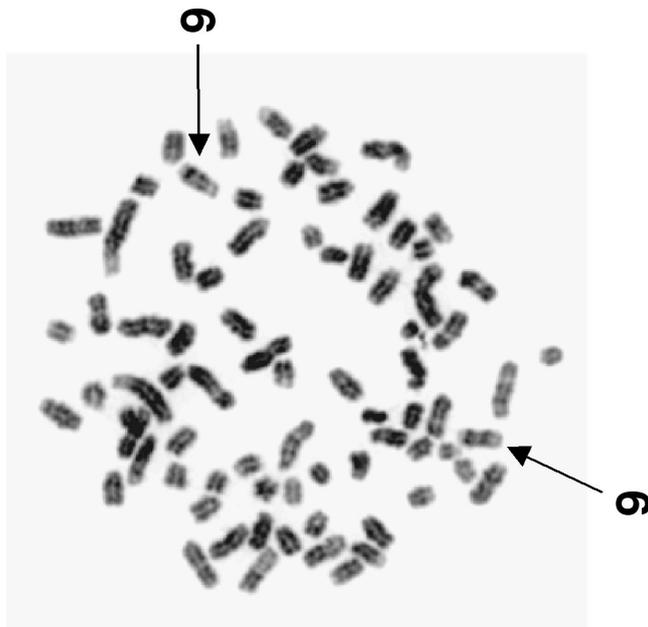
An initial study of our enhanced DAPI-banded images enabled the identification (P. Fischer (The Animal Health Trust, Newmarket, UK and N. Reiman (Heidelberg, Germany)) of canine metaphase chromosomes 1-21 plus X and Y and their corresponding flow karyotype peaks (Langford *et al*. 1996, bound into this thesis).

In a subsequent analysis the identification of the remaining 17 undesignated chromosomes and their corresponding flow karyotype peaks, was carried out by M. Breen (Newmarket, UK) (Breen *et al.* 1999a, bound into this thesis).

Although eight of the flow karyotype peaks each contained two similarly-sized pairs of homologous chromosomes (H = 8 + 11, J = 10 + 17, L = 13 + 15, S = 21 + 23, V = 24 + 28, W = 29 + 32, Z = 31 + 34, cc = 33 + 36), the enhanced DAPI banding patterns were sufficiently different to distinguish each homologous pair.

The assignment of each dog chromosome to its corresponding flow karyotype peak is presented in Table 3.2.

Table 3.3 (Next page) Chromosome assignment of 33 canine flow karyotype peaks

Peak	Number of homologues	Chromosome assignment	Peak	Number of homologues	Chromosome assignment
A	2	1	R	2	19
B	2	3	S	4	21,23
C	2	4	T	2	25
D	2	2	U	2	27
E	2	5	V	4 (3)	24, 28
F	2	7	V'	1	28
G	2	6	W	4	29, 32
H	4	8, 11	X	1	X
I	2	12	Y	1	Y
J	4	10, 17	Z	4	31, 34
K	2	9	aa	2	26
L	4	13, 15	bb	2	30
M	2	16	cc	4	33, 36
N	2	22	dd	2	37
O	2	14	ee	2	38
P	2	20	ff	2	35
Q	2	18			

3.2.3 Standardisation of the canine DAPI-banded karyotype

In a collaborative study with Matthew Breen (The Animal Health Trust, Newmarket, UK) the panel of 33 canine chromosome-specific paints were used to identify unequivocally each chromosome type in a normal canine metaphase spread. After hybridising the first paint to be tested and following image capture from 30 metaphases, the same metaphases were re-used for repeat hybridisations for all paints (following successive re-denaturation steps for 15 to 60 seconds at 65°C). Using this process all chromosome

pairs were conclusively identified in each metaphase spread. Accurate karyotypes were produced from 30 mid-metaphase spreads and used to derive a 460-band DAPI ideogram, which has been incorporated into the karyotyper software of Vysis Quips™ Image Analysis Software (Breen *et al.,* 1999a). We were able to produce the first complete DAPI-banded karyotype of the dog in which each chromosome was accurately placed, together with a 460-band ideogram. The data formed the basis for a proposed standard for the dog karyotype (Breen *et al.* 1999a).

3.3   Reciprocal chromosome painting analysis between human and the dog

3.3.1 Generation of human chromosome-specific paints

A chromosome suspension from a normal anonymous female lymphoblastoid cell line (HRC 160) was prepared and analysed as described in 2.*. In the resulting bivariate flow karyotype (Figure 3.5), all the human chromosomes except 9, 10, 11 and 12 were resolved into individual peaks. Human chromosomes 9, 10 11 and 12 share a similar size and base pair ratio and, thus, usually remain as one large peak in the centre of the flow karyotype. Three chromosome types (15, 16 and 22) were each represented by two small adjacent peaks, which each contained a count frequency equivalent to a single chromosome homologue. The separation of the individual homologue peaks were due to resolvable size differences caused by heterochromatic polymorphisms, as described earlier.

Five hundred copies of each resolved human chromosome type were isolated into sterile 0.5 ml PCR tubes for DOP-PCR. The chromosome paint probes were generated and labelled as described in section 2.8, with either biotin-16-dUTP (for use in reciprocal zoo-FISH experiments between human and dog) or FITC- or Cy3-dUTP (for use in the reciprocal chromosome painting experiments between human and gibbon, described later).

Figure 3.5 Bivariate flow karyotype of a normal male human
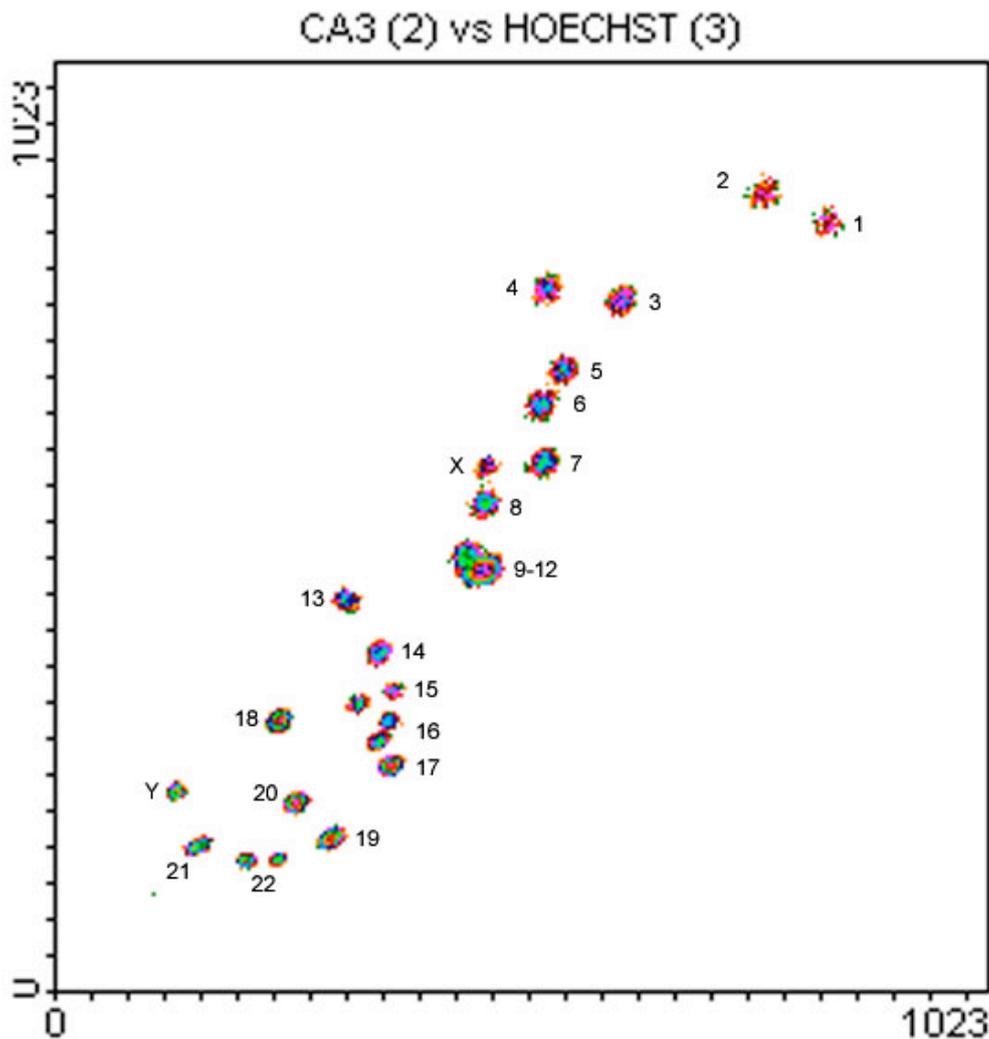
As the dog is distantly related to the human it was deemed likely that there would be considerable sequence divergence leading to weak hybridisation signals after zoo-FISH. The human chromosome paint probes that were to be hybridised onto canine metaphase spreads were, therefore, labelled with biotin-16-dUTP because the biotin-avidin detection system allowed for signal amplification after hybridisation.

Each human chromosome-specific paint was checked for specificity on normal human metaphase spreads by FISH. At least 10 metaphases were analysed for each test-paint. Each of the human chromosome-specific paints hybridised only to the two chromosome homologues of origin in each human metaphase spread studied.

3.3.2 *Hybridisation of human chromosome 22 specific paint onto dog metaphase chromosomes*

The biotinylated human chromosome-22 paint was hybridised onto normal dog metaphase spreads, and data from 20 metaphases were analysed. The human chromosome 22 paint hybridised onto four blocks per canine metaphase, representing segments in two different chromosome types (Figure 3.6, next page). Paint hybridisation was detected on both homologues of canine chromosome 10q21-23.1 and both homologues of canine 26q21-24, identified according to the nomenclature by Breen *et al*. (1999).

Figure 3.6 Hybridisation pattern of human chromosome 22 paint on dog metaphase spread. The paint is pseudocoloured green, the chromosomes blue. Canine chromosomes 10 and 26 are indicated.
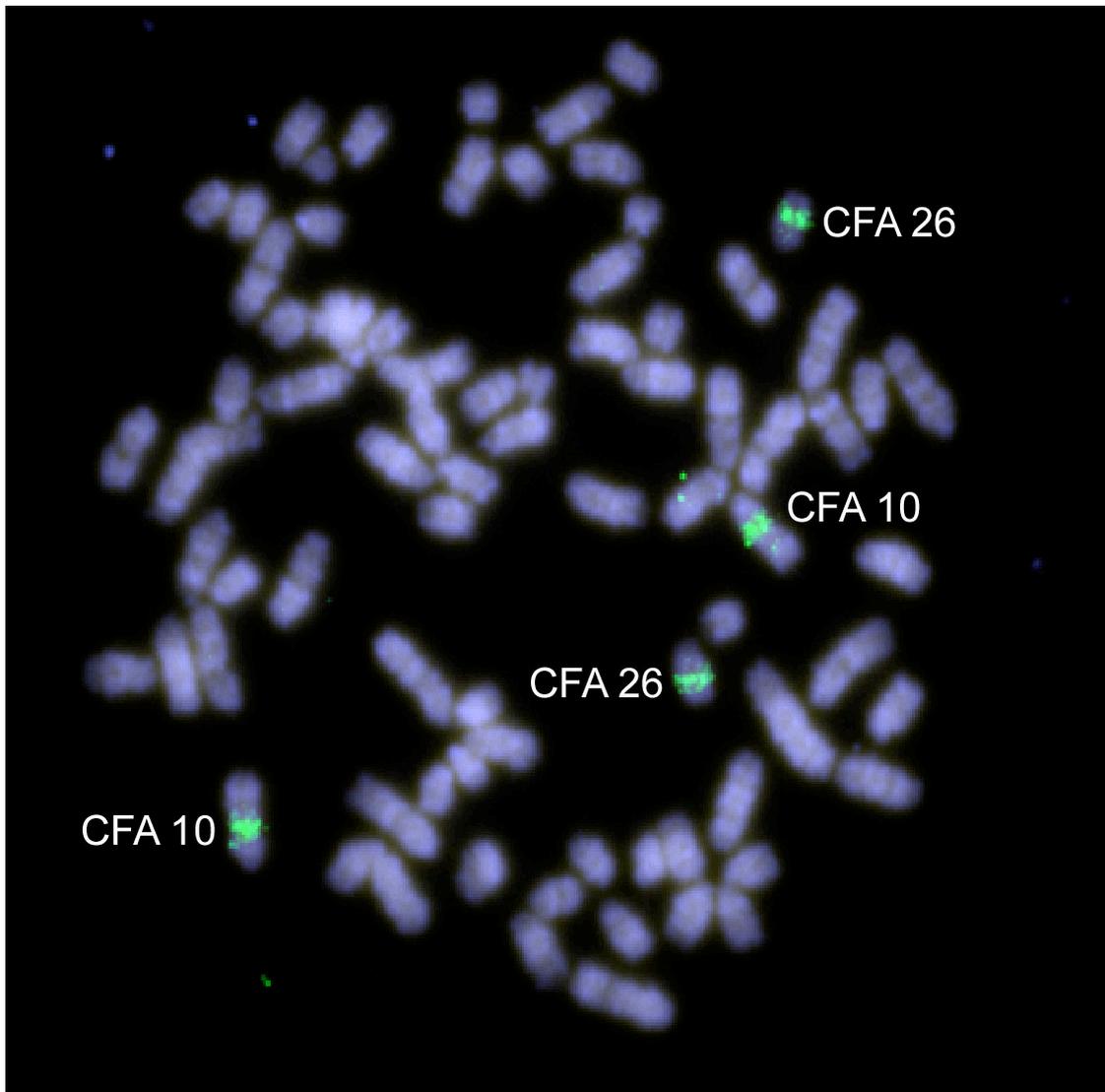
3.3.3 Reciprocal hybridisation of dog chromosome paints onto human metaphase chromosomes

In order to establish which part of each canine chromosome was syntenic with which part of human chromosome 22 a reciprocal painting study was carried out. Canine chromosome 10 and 26 paints were hybridised onto normal human metaphase spreads, and data from 20 metaphases were analysed for each paint probe.

The canine chromosome 10 paint hybridised onto 6 blocks per human metaphase, both homologues of 2pter-q21.1, 12p15-21.2 and 22q13.1-qter. The canine chromosome 26 paint hybridised to 4 blocks per human metaphase, both homologues of 12p24.1-pter and 22cen-q13.1 (Figure 3.7, and Breen *et al.* 1999).

Figure 3.7 Ideogram illustrating hybridisation patterns of canine chromosome 10 and 26 paints on human metaphase chromosomes. From Breen *et al.* 1999.
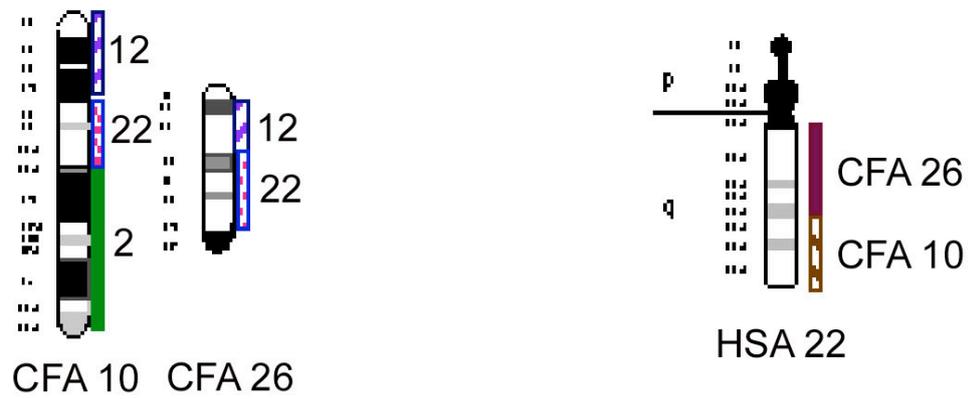
3.4   Reciprocal chromosome painting analysis between human and gibbon

3.4.1   Hybridisation of human chromosome 22 paint onto gibbon metaphase chromosomes

Various different protocols for producing good quality fixed metaphase spread preparations were investigated. These involved variations in the starting cultured cell concentration, duration of incubation with the spindle inhibitor Colcemid and duration in the hypotonic potassium chloride swelling solution. The most satisfactory fixed metaphase preparations were achieved following the protocol described in chapter 2.

Standard hybridisation and signal detection techniques carried out routinely in the FISH laboratory use biotinylated single-copy probes or chromosome paints. The biotinylated probes are hybridised and the signal is amplified and detected using the three-layer technique described in chapter 2. Although the probe is pre-annealed with $C_0t = 1$ DNA prior to hybridisation, the period of hybridisation will facilitate the non-specific binding of probe to repetitive elements in the metaphase chromosomes as well as to the surface of the microscope slide. Any binding will be amplified by a three-layer detection step. In certain systems it is possible to minimise the occurrence and subsequent amplification of background hybridisation by using probes or paints which have been directly labelled with a fluorescent molecule, rather than with biotin. The directly labelled probes do not usually need signal amplification as long as the binding is quite specific. As well as reducing the appearance of background non-specific hybridisation signals, the directly labelled probes also use fewer reagents per experiment compared to biotinylated probes that need a three-layer detection.

Because the Siamang gibbon is so closely related to the human, it was believed that a direct labelling of the paints would be sufficient to visualise the hybridisation signals, and would minimise the occurrence of background non-specific hybridisation signals.

Therefore, the gibbon metaphase spreads were analysed using a human chromosome 22-specific paint directly labelled with the fluorescent molecule FluoroGreen™-dUTP.

The FluoroGreen™-labelled human chromosome 22-specific paint was hybridised to gibbon metaphase chromosomes, and data from 20 metaphases were analysed. In each metaphase spread the human chromosome 22 paint hybridised onto four blocks, two on each homologue, one on each p arm and one on each q arm (Figure 3.8). The Siamang chromosome 18 was identified according to Van Tuinen and Ledbetter (1983).

Figure 3.8 (next page) Hybridisation signals obtained by hybridising human chromosome 22-specific paint to Siamang metaphase spreads. The metaphase chromosomes have been pseudo-coloured red and the paint signals pseudo-coloured green. The corresponding digital DAPI-banded metaphase image is in black and white.
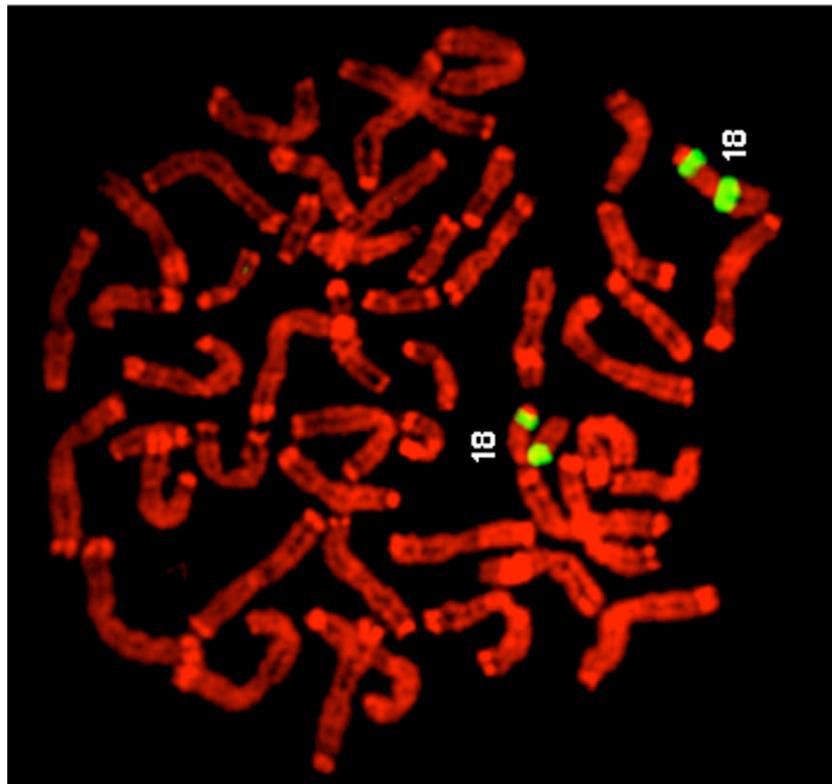
There were clearly regions on both arms of Siamang chromosome 18, which were not painted by the human chromosome 22 probe, implying that those regions were syntenic with one or more other human chromosomes. Previously published painting experiments had also indicated that those regions were not homologous to human chromosome 22, but shared homology with (unspecified) regions of human chromosomes 5, 2 and 16 (Koehler *et al.* 1995).

3.4.2 *Generation of the Siamang flow karyotype and chromosome isolation*

Confirmation of the human chromosome homology map of Siamang chromosome 18 by a reciprocal chromosome painting study required the generation of a Siamang chromosome 18-specific paint and its hybridisation to human metaphase chromosomes.

To establish a Siamang bivariate flow karyotype, a suspension of chromosomes was isolated, stained with Hoechst 33258 and Chromomycin A3 and analysed on a dual laser flow cytometer. In the resulting bivariate flow karyotype, twenty-one peaks representing the 25 homologous chromosome pairs were resolved (Figure 3.9). The peaks were labelled alphabetically from A to U. Sorting regions were set up and 500 to 1000 chromosomes from each resolvable peak were isolated directly into sterile 0.5 ml PCR tubes for the generation of chromosome-specific paints labelled with FluoroGreen™-dUTP by DOP-PCR.
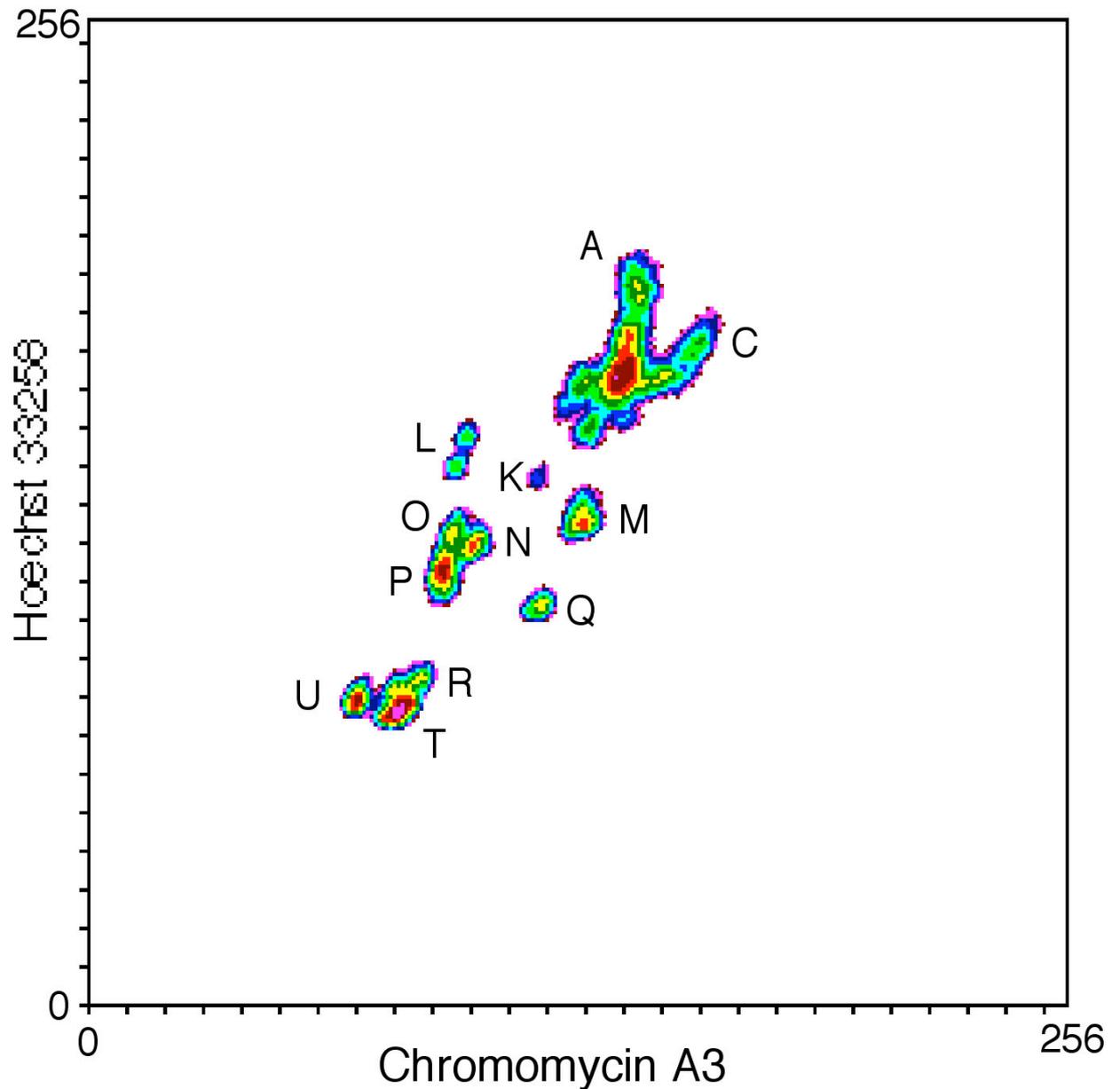
Figure 3.9 Bivariate flow karyotype of the gibbon.

3.4.3 Identification of the Siamang chromosome 18 specific paint

To identify which flow karyotype peak represented Siamang chromosome 18 each of the

21 paints was hybridised onto Siamang metaphase chromosomes. For each paint probe,

data from 20 metaphases were analysed. The paint generated from GC base pair rich

flow karyotype peak "Q" hybridised to both Siamang chromosome 18 homologues,

according to Van Tuinen and Ledbetter, (1983) indicating that flow karyotype peak Q contained Siamang chromosome 18 alone (Figure 3.10).

A narrow region at the telomere of the short arm of Siamang chromosome 18 was unlabelled by FISH with paint Q. Previous Q-banding studies, which highlighted heterochromatin in chromosomes, indicated that there is a block of heterochromatin at the telomeres of nearly every Siamang chromosome, including chromosome 18 (Koehler *et al.* 1995). It is likely that the heterochromatic sequences were either not amplified by DOP-PCR and were not represented by the chromosome 18 paint, or the repetitive heterochromatic sequences were present in the paint but were self-annealed by the human $C_0t=1$ DNA during the pre-annealing step and thus, did not hybridise to that region.

Figure 3.10 (next page) Hybridisation pattern of the paint from peak Q. In the left panel, the paint signals are pseudocoloured green and the chromosomes red Siamang chromosome 18 is marked on the DAPI image in the right panel.
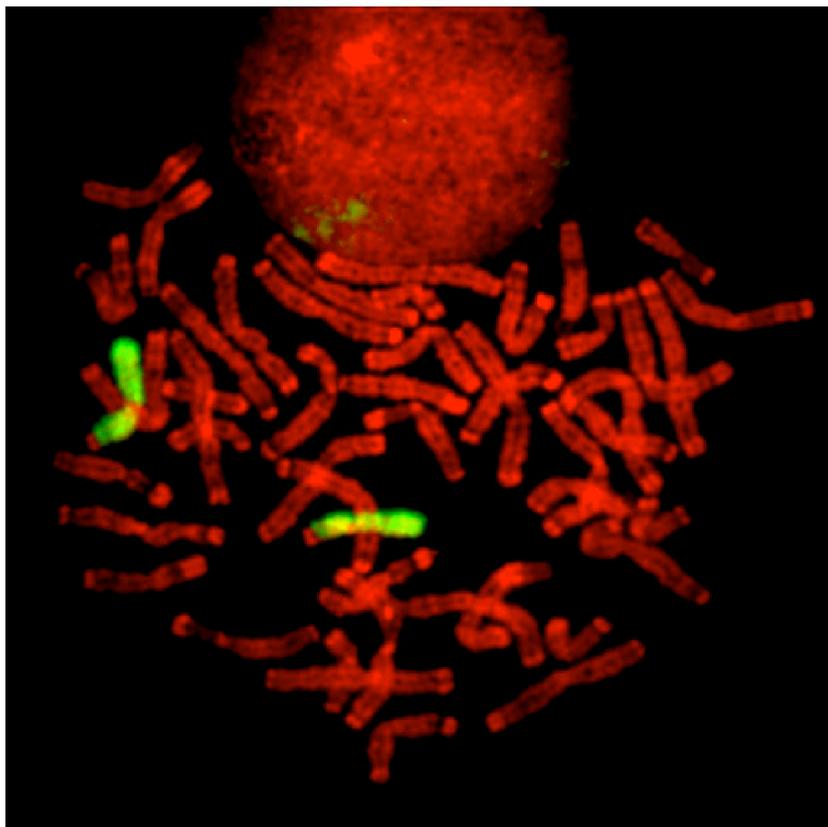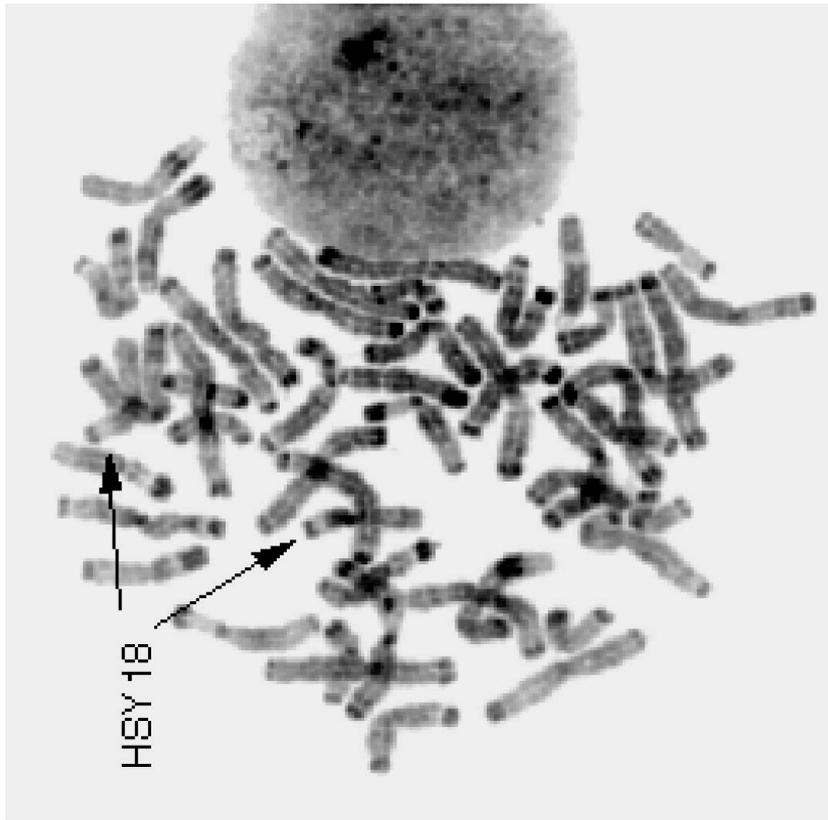
3.4.4   Hybridisation of Siamang chromosome 18 paint onto human metaphase chromosomes

The FluoroGreen™-labelled Siamang chromosome 18 paint was hybridised onto human metaphase spreads, and data from 20 metaphases were analysed. As well as human chromosome 22, regions of the other human chromosomes sharing homology with Siamang chromosome 18 were highlighted. The Siamang chromosome 18 paint hybridised to parts of four different human chromosome types: the whole of 22q, a narrow band in 16p12-13.2, a block in 5q11.2-13 close to the centromere and a terminal block in 2p22-pter (Figure 3.11).

3.4.5   Hybridisation of human chromosome 22, 16, 5 and 2 paints onto Siamang metaphase chromosomes

In order to establish the precise arrangement in Siamang chromosome 18 of the various human-homologous blocks, FluoroGreen™-labelled human chromosome 16-, 5- and 2-specific paints were hybridised individually to Siamang metaphase spreads. Data from 20 metaphases were analysed for each paint. The homologous regions corresponding to each of the human chromosomes 16, 5 and 2 were highlighted on other Siamang chromosomes in the metaphase spread, as well as on chromosome 18.

Figure 3.11 (next page) Hybridisation pattern of gibbon chromosome 18 paint on human metaphases. The paint signal is pseudocoloured green and the chromosomes red. The human chromosomes with hybridisation signal are indicated with arrows.
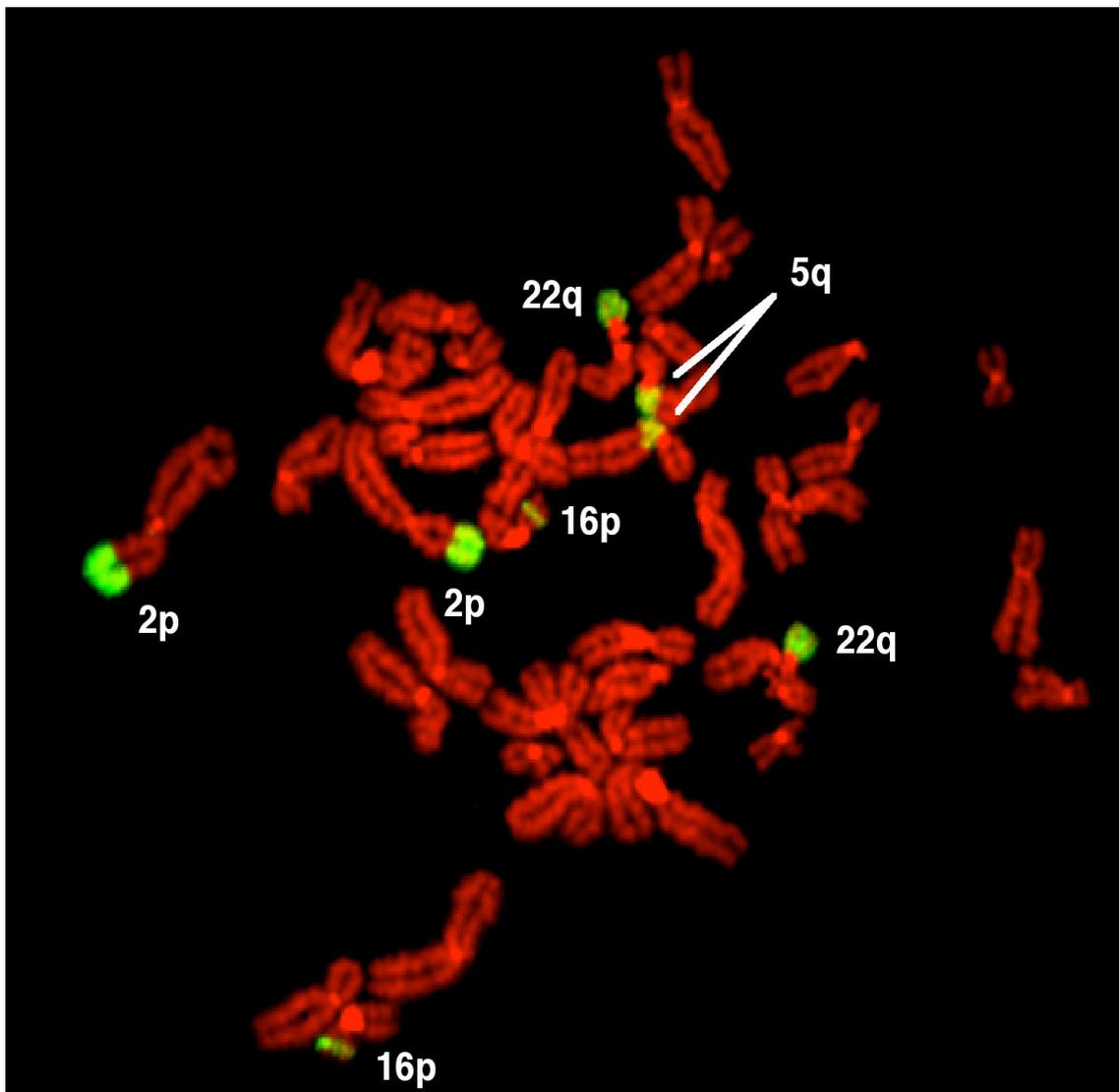
The human chromosome 16 paint hybridised to two blocks on Siamang 11 (p and q), one block on 14q and to two narrow blocks on Siamang 18, one in the p arm adjacent and centromeric to the human chromosome 22-homologous block, and one on the q arm, adjacent and telomeric to the human chromosome 22-homologous block.

The human chromosome 5 paint hybridised to one block in 7p, one block on 11q, one block on 16q and a single block on Siamang 18p, between the centromere and the human chromosome 16-homologous block.

The human chromosome 2 paint hybridised to one block on 8q, one block on 9p, two blocks on 14 (p and q), one block on 22q and a single block on Siamang 18q between the human chromosome 16-homologous block and the telomere.

A homology map was constructed indicating the regions of homology between Siamang chromosome 18 and human chromosomes 22, 16, 5 and 2. (Figure-3.12)

HSA22   HSA16   HSA5   HSA2
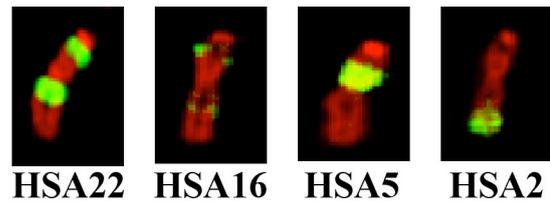
HSY18 IDEOGRAM INDICATING
HUMAN-HOMOLOGOUS BLOCKS

Figure 3.12 Above, hybridisation patterns of paints for human chromosomes 22, 16, 5 and 2, respcetively, on gibbon chromsome 18. The hybridisation pattern is pseudocoloured green and the chromosome red. Below, Ideogram of Siamang chromosome 18 and the corresponding mapped human homology.

3.5   Discussion

The aim of this work was to conduct a reciprocal zoo-FISH analysis of DNA homologous to human chromosome 22 in the domestic dog as well as a study of human homologies with Siamang chromosome 18. Work towards this aim has involved the production of dog and gibbon flow karyotypes and the generation of chromosome specific paints by DOP-PCR. Reciprocal zoo-FISH analyses have led to the construction of homology maps for human chromosome 22, dog chromosomes 10 and 26, and Siamang chromosome 18.

The Canine Study

Flow-karyotyping the dog presented particular challenges. Several pairs of canine chromosome types, which shared size and base pair similarities, did not resolve into individual flow karyotype peaks. But despite the complexity of the canine karyotype, good quality chromosome preparations resulted in the generation of reproducible high-resolution flow karyotypes of the male and female dog.

The ability to assess the number of chromosome events represented by each dog flow karyotype peak assisted in the identification of the peaks containing chromosomes X and Y in the male. It also allowed an explanation of the additional peak V' in the female flow karyotype, as it was most likely caused by a heterochromatic polymorphism (heteromorphism) in one chromosome 28 homologue (as determined by chromosome painting). In the male, flow karyotype peak V contains 4 homologues: 2 of chromosome 24 and 2 of chromosome 28. All four homologues are in the same peak because the size and AT to GC base pair ratios of the two chromosome types are similar. A heteromorphism of one chromosome homologue may affect its size, and, because heterochromatin is AT-rich, its AT to GC ratio. In the case of peak V', the heteromorphic chromosome 28 homologue is reduced in size enough to be resolved as an extra peak on the flow karyotype.

131

To identify the dog chromosome types in each peak, chromosomes were flow sorted, amplified and labelled by DOP-PCR and painted onto metaphase spreads of a normal male dog. Twenty-two of the peaks from the male flow karyotype each hybridised to single homologous pairs, and eight of the peaks hybridised to two pairs. Paints from the remaining two male flow karyotype peaks hybridised to only one homologue each in male metaphase spreads, thus corresponding to the sex chromosomes X and Y. All of the 38 pairs of autosomes and the two sex chromosomes of the dog were accounted for in the flow karyotype.

Prior to the work towards this thesis, there was no reliable method of identifying every chromosome in a canine metaphase spread. The complex nature of the dog karyoytpe, coupled with difficulties in stimulating lymphocytes and optically aligning a flow cytometer for chromosome analysis, had also previously frustrated attempts to flow sort dog chromosomes. Work carried out by VanDevanter *et al.* (1994), to generate a limited number of canine chromosome paints, avoided the need to obtain good resolution of flow karyotype peaks by sorting only single chromosomes. Ligation-adapter PCR was then used to make paint probes of limited representation from these single chromosomes.

The canine chromosome paints generated by DOP-PCR for this study were of good quality. The eight pairs of chromosome types, which were represented by paints from peaks H, J, L, S, V, W, Z and cc, had sufficiently different banding patterns to enable the distinction between each comprising pair.

Microscopy and flow karyotyping used alone could not easily distinguish each individual canine chromosome. However, using the two technologies together has resolved the problem and has enabled the unequivocal identification of each chromosome pair in a metaphase spread, and the proposal of a DAPI banded standard karyotype for the domestic dog (Breen *et al.*, 1999).

The characterisation of the canine chromosome paint probes, as well as the proposed DAPI-banded karyotype permitted the reciprocal zoo-FISH analysis between human and the dog. The study revealed detailed regions of conserved synteny between human chromosome 22 and canine chromosomes 10 and 26.

One of the most striking widely conserved ancestral neighbouring chromosome segment combinations across a number of mammalian orders is that of human chromosome 12 and chromosome 22 (Chowdhary, *et al.,* 1998). Regions homologous to human chromosomes 12q15-21.3 and 22q13.1-qter are associated on canine chromosome 10, and regions homologous to human 12q24.1-qter and 22q11.1-13.1 are associated on canine chromosome 26. This provides more evidence that these combinations probably represent ancestral chromosome arrangements.

Segments of human chromosomes 12 and 22 appear as neighbours on two chromosomes in cattle (BTA 5 and BTA 17), pig (SSU 5 and SSU 14), dolphin (TTR 8 and TTR 9), cat (FCA B4 and FCA D3), horse (ECA 8 and ECA 26), lemur (EMA 10 and EMA 19), American mink (MVI 3 and MVI 12) (summarised in Glas *et al.*, 1998), and giant panda (AME 12 and AME 15) (Nash *et al.*, 1998). It has also been observed on just one chromosome of the harbour seal (PVI m3) (Frönicke *et al.*, 1997).

Unidirectional chromosome painting involving the dog was first reported by Werner *et al.* (1997), using a human chromosome 17 paint probe to identify homologous segments on two dog chromosomes. The only previous report of reciprocal chromosome painting involving human and the dog was by Thomas *et al.* (1999), who used the paints generated by this study to identify conserved segments of synteny between human chromosomes 1p31.2-p32.3, 11q23-q25, 16q21-q24 and 17p12-p13 and dog chromosome 5.

A unidirectional approach to the work for this thesis, would only have yielded the information that human chromosome 22 has shared homology with dog chromosomes 10

and 26. It was only due to the reciprocal nature of the study, that it was possible to identify which segment of human chromosome 22 is homologous to which dog chromosome. For this thesis, a detailed reciprocal zoo-FISH analysis between the domestic dog and the human has generated metaphase chromosome homology maps for human chromosome 22 and canine chromosomes 10 and 26. (For full canine/human chromosome homology maps, see Breen *et al.*, 1999b). Canine chromosome 10 shares homology with human 22q13.1-qter and canine chromosome 26 shares homology with human 22cen-q13.1.

These results strongly suggest that a site of evolutionary rearrangement (fusion) is present in human chromosome 22q13.1. At the low level of resolution afforded by these and other zoo-FISH studies, it might be suggested that the arrangement of human chromosome 22 homologous material in the canine karyotype represented the ancestral state. During the karyotype evolution of the primates a rearrangement event probably lead to the fusion of the human chromosome 22 homologous material at a point now identified within 22q13.1.

The Siamang Study

In the second part of this chapter, Siamang metaphase chromosomes were analysed with a human chromosome 22 specific paint probe. In each spread analysed Siamang chromosome 18 was identified by its DAPI-banding pattern. Two distinct blocks of synteny were identified on Siamang chromosome 18 p and q, as described by Koehler *et al.* (1995).

For further analysis, a high-resolution flow karyotype of the gibbon was produced and a paint probe for Siamang chromosome 18 was generated. Reciprocal chromosome painting between the human and Siamang was used to generate a homology map for Siamang chromosome 18 in relation to human chromosomes 22, 16, 5 and 2.

A previous FISH study had been carried out which involved the unidirectional hybridisation of human paints onto Siamang metaphase chromosomes (Koehler *et al.* 1995). The homologies between Siamang 18 and human chromosomes 22, 16, 5 and 2 were reported in different orientations to the arrangement I found. However, the metaphase spreads for the two studies had been prepared from the same Siamang individual. After discussion, the authors commented that their hybridisations had been difficult to interpret due to the fact that they had been using paints generated from chromosome specific libraries. The authors agreed that their results, if repeated using DOP-PCR paints, could in fact have looked the same as mine (Johannes Weinberg, personal communication).

From the information in the homology map between Siamang chromosome 18 and human chromosomes, a proposed homology map of the chromosome ancestral to Siamang chromosome 18 could be suggested (Figure 3.13).
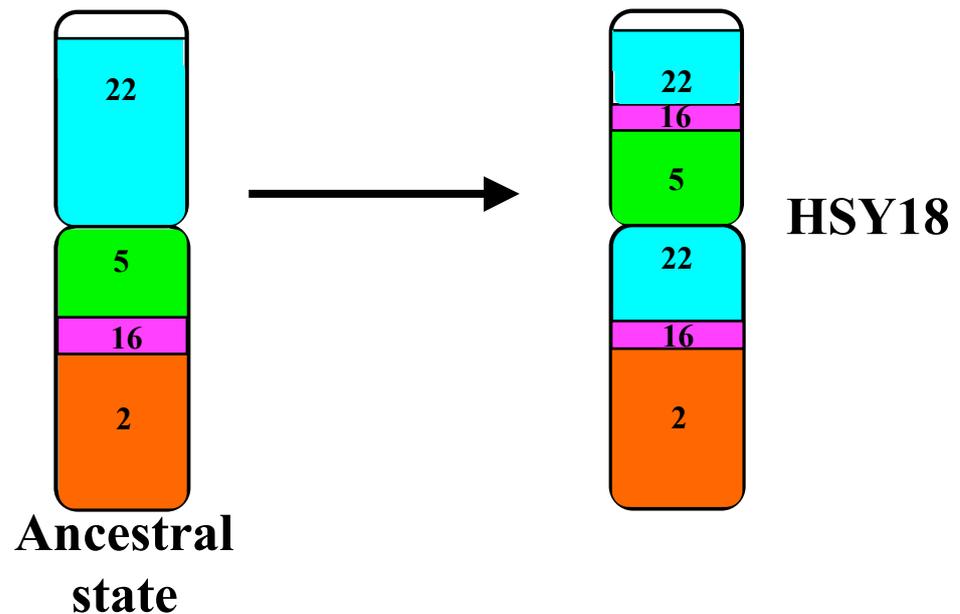
Figure 3.13 Homology map of the current Siamang chromosome 18 (right) and a proposed homology map of the ancestral chromosome (left). The human chromosome 22-homologous material is intact in one block on 18p, the human chromosome 5-homologous block is in the proximal q arm, the human chromosome 16-homologous material is intact in one block on 18q adjacent to the human chromosome 5-homologous block and the human chromosome 2-homologous block is on 18qter. Fission in the blocks homologous to human chromosomes 22 and 16 and a pericentric inversion in the ancestral chromosome could have lead to the current state of Siamang chromosome 18.

Summary

A detailed reciprocal zoo-FISH analysis between the human and the dog was used to identify a region corresponding to an evolutionary chromosome fusion point in human chromosome 22q13.1. At the low level of resolution afforded by these and other zoo-FISH studies, it might be suggested that the arrangement of human chromosome 22 homologous material in the canine karyotype represented the ancestral state. During the

karyotype evolution of the primates a rearrangement event probably lead to the fusion of the human chromosome 22 homologous material at a point now identified within 22q13.1.

The reciprocal zoo-FISH study between human and Siamang did not lead to the identification of a chromosome rearrangement site in human chromosome 22 because both 22q-homologous blocks reside in Siamang chromosome 18. It was possible to suggest that a pericentric inversion might have lead to the fission of human chromosome 22 homologous material in the ancestor to Siamang chromosome 18.

To further refine the analysis of the 22q rearrangement point, a higher resolution zoo-FISH study was carried out. This work is described in the following chapter.

CHAPTER 4

FISH analysis of the Siamang and Dog using human chromosome 22q bacterial clones

4.1 **Introduction**

4.1.1 Cross-species FISH

The first FISH analyses in apes were investigations of the fusion event that lead to the generation of human chromosome 2. Detailed comparative cytological banding analysis had indicated that human chromosome 2 arose through the fusion of two ancestral ape chromosomes. Human chromosome regions 2p and 2q are homologous to chromosomes 12 and 13 respectively in the chimpanzee (*Pan troglodytes*) and chromosomes 11 and 12 respectively in the gorilla (*Gorilla gorilla)* and orangutan (*Pongo pygmaeus*) (Yunis and Prakash, 1982). This was confirmed by zoo-FISH of a human chromosome 2 paint to primate metaphase chromosomes (Wienberg, *et al.,* (1990).

By carrying out FISH, using two human chromosome 2-specific cosmids containing the vertebrate telomeric repeat, the nature of the evolutionary rearrangement was confirmed as a telomere-telomere fusion by Ijdo *et al.* (1991). The two clones were FISH-mapped to human chromosome 2q12 as well as to the ends of other chromosomes. It was concluded that the human chromosome 2 locus cloned was the relic of the telomere-telomere fusion and marks the point at which two ancestral ape chromosomes fused to give rise to human chromosome 2 (Ijdo *et al.,* 1991).

Subsequent studies involved the use of a human chromosome 21 alpha-satellite plasmid clone for FISH on great ape and human metaphase chromosomes, (Baldini *et al.,* 1993). The plasmid containing the alphoid sequence hybridised to the centromeric region of all the human, chimpanzee, gorilla and orangutan chromosomes. The clone also identified sequences on human chromosome 2p21. As that region is not centromeric, it was suggested that those sequences might have been derived from an ancestral centromere, which was subsequently inactivated after the evolutionary rearrangement event in order to maintain chromosome stability.

To further extend the study, a panel of human chromosome 2 YACs was used for FISH to investigate the fragmented human chromosome 2 homologues in four chromosomes of the lesser ape *Hylobates lar (H.lar)* (Arnold *et al.*, 1996). Previous chromosome painting studies had revealed that human chromosome 2 is homologous to five distinct regions on *H. lar* chromosomes 1, 10, 12 and two parts of 16. But it had not been established which segment of human chromosome 2 was homologous to which *H. lar* chromosome. As well as hybridising *H. lar* chromosome specific paints back on to human chromosome 2, Arnold and colleagues also hybridised YACs specific for the major bands on human chromosome 2 to *H. lar* metaphase chromosomes. From its hybridisation pattern, a YAC was found that identified the rearrangement point between human chromosome 2-homologous material present on *H. lar* chromosomes 10 and 16. This study demonstrated that a combination of reciprocal heterologous chromosome painting and FISH of specific probes, such as YACs, could be used to identify homologies between closely related species, to construct detailed comparative chromosome maps rapidly and to identify evolutionary rearrangement points.

YACs have also been used to identify chromosome rearrangements within the genomes of higher primates. High-resolution G-banding analyses revealed the high degree of morphological conservation of great ape chromosomes (Nickerson and Nelson, 1998). The distribution of heterochromatin and the occurrence of pericentric inversions were the most notable differences. Pericentric inversions may have played an important role in the establishment of reproductive isolation and speciation of the hominoids as they diverged from a common ancestor (Nickerson and Nelson, 1998). Human YAC clones were used for FISH to identify pericentric inversions when comparing the human karyotype to that of the chimpanzee (Nickerson and Nelson, 1998). Five evolutionary pericentric inversion points were identified on the chimpanzee chromosomes homologous to human chromosomes 4, 9 and 12. The YACs spanning an inversion point showed hybridisation signals on both the p and q arms of the corresponding chimpanzee chromosome.

A FISH study of chromosome homologies between more distantly related mammalian species has also been reported (Cole *et al.*, 1998). Physical mapping using contiguous human YAC and PAC clones was carried out between the human and laboratory mouse, *Mus musculus*. A human YAC from chromosome 22 was identified which spans the evolutionary rearrangement point defining the boundary between material homologous to regions of human chromosome 21 and human chromosome 22 on mouse chromosome 10.

It was noted by the authors that detailed analysis of sequences across evolutionary rearrangement points on chromosomes would provide insight into the processes involved both in chromosome evolution and maintaining regions of conserved synteny.

4.1.2 Aim of this chapter

The aim of the work described in this chapter was to further refine the analysis of the regions corresponding to evolutionary chromosome rearrangement points in material homologous to human chromosome 22q on Siamang chromosome 18 and on dog chromosomes 10 and 26. In order to carry out the analysis, a high-resolution cross-species FISH study was to be carried out by hybridising human chromosome 22q bacterial clones onto Siamang and dog metaphases.

The high-resolution map of human chromosome 22 (Collins *et al.,* 1995) provided a framework for the sequencing effort (Dunham *et al.,* 1999). To identify genomic clones for sequencing, extensive clone maps of the chromosome were constructed using cosmids, fosmids, bacterial artificial chromosomes (BACs) and P1-derived artificial chromosomes (PACs). Clones were identified by screening BAC and PAC libraries using chromosome 22-specific STS markers, or by using cosmid and fosmid libraries derived from flow-sorted DNA from chromosome 22. Overlapping clone contigs were assembled by restriction fingerprinting and ordered relative to each other using the established framework map (Collins *et al.,* 1995).

The high quality of the map, which has been verified by sequencing (http://www.sanger.ac.uk/HGP/Chr22, or http://www.genome.ou.edu/Chr22.html), provided a powerful tool for selecting clones evenly distributed along chromosome 22q for the FISH studies described in this chapter.

RESULTS

4.2   FISH Analysis of Siamang Chromosomes with Human Bacterial Clones

4.2.1 Analysis of Siamang 18 with human 22q BACs

Thirteen BAC clones evenly distributed along human chromosome 22q (from 22q11.23 to 22q13.33) were selected from the Sanger Institute chromosome 22 mapping resource. This equates to an average spacing of approximately one BAC per 3.5 to 4 Mb. The DNA from each BAC clone was isolated by plasmid preparation and biotinylated by nick translation, then hybridised to human and Siamang metaphase chromosomes. Ten human and Siamang metaphase spreads were analysed for each BAC probe.

There is no standard banded karyotype for the Siamang, and high-resolution localisations were not possible. Therefore the localisation of the BAC clones to Siamang chromosome 18 were scored as "proximal" or "distal" within the human chromosome 22 homologous regions on HSY 18p or 18q. The BAC clones used, their human chromosome 22q localisations and their corresponding Siamang chromosome 18 arm localisations are tabulated in Table 4.1.

Table 4.1 Human chromosome 22q localisation (HSA22) and corresponding Siamang chromosome 18 arm localisation (HSY18) of thirteen BAC clones. The bK prefix on the clone names relates to the library from which the clones originated. NA = not assessed

BAC clone	HSA22 localisation	HSY18 localisation
bK65A6	22q11.23	Proximal 18q
bK125H2	22q12.1	Proximal 18q
bK282F2	22q12.3	Proximal 18q
bK415G2	22q12.3	Distal 18q
bK221H1	22q12.3	Distal 18q
bK212A2	22q12	Distal 18q
bK236H12	22q13.1	Proximal 18p
bK206C7	22q13.1	NA
bK229A8	22q13.1	NA
bK216E10	22q13.2	NA
bK989H11	22q13	Proximal 18p
bK1109B5	22q13.31-13.32	NA
bK799F10	22q13.33	Distal 18p

All of the BAC probes gave strong, clean and informative hybridisation signals on human chromosome 22q. Nine of the BAC probes (bK65A6, bK125H2, bK282F2, bK415G2, bK221H1, bK212A2, bK236H12, bK989H11 and bK799F10) gave strong, clean and informative hybridisation signals on Siamang chromosome 18. Four of the BAC probes (bK206C7, bK229A8, bK216E10 and bK1109B5) gave dissipated signals with high background so that it proved difficult to assess confidently signal localisation (NA in table).

Six of the BAC probes (bK65A6, bK125H2, bK282F2, bK415G2, bK221H1 and bK212A2) hybridised to the long arm of Siamang chromosome 18 and three of the BAC probes

(bK236H12, bK989H11 and bK799F10) hybridised to the short arm (see table 4.1). These two groups of clones are located proximally and distally to the rearrangement breakpoint, respectively, on human chromosome 22. Some of the results of the first round FISH analysis are summarised in Figure 4.1.

Figure 4.1 (next page) Summary of the BAC clone analysis of Siamang metaphase chromosomes. On the left, the location on human chromosome 22 of eight of the tested BAC clones is indicated by their position adjacent to the chromosome 22 ideogram. On the right, eight corresponding Siamang chromosome 18 images show the FISH signal location from each BAC clone. The images are ordered according to the location of the BAC signal. The BACs in dark blue are proximal and the BACs in light blue are distal to the breakpoint on chromosome 22.

13

12

p

11.2

11.1

11.1

11.21

11.22

11.23 — bK65A6

12.1 — bK125H2

q

12.2

12.3 — bK282F2
bK415G2

bK212A2

13.1 — bK236H12

13.2 — bK989H11

13.31

13.32

13.33

bK799F10

22
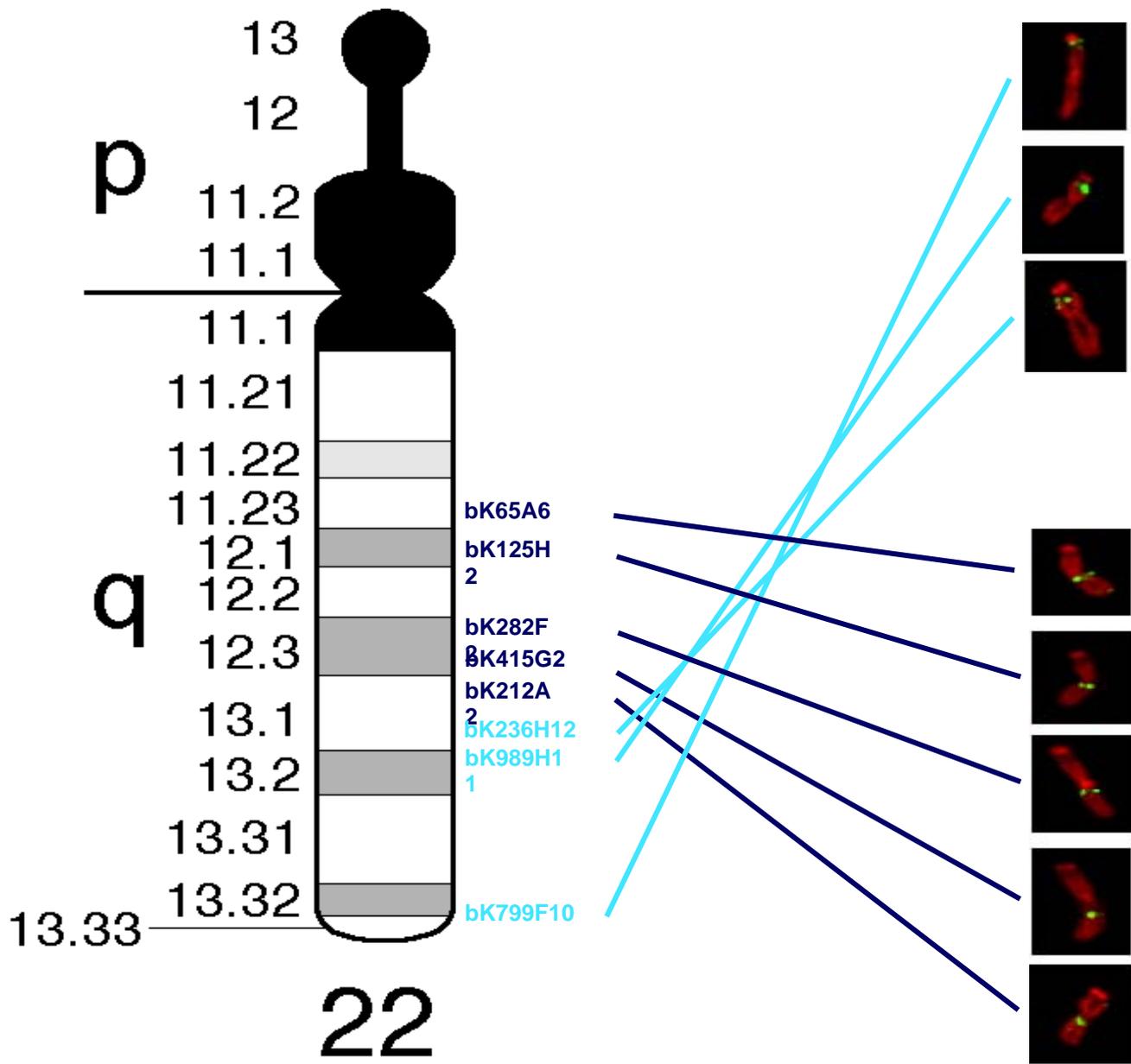
It was possible to discern that the BAC clones hybridised to a similar position with respect to centromere, telomere and each other within the regions on Siamang chromosome 18 as they did to human chromosome 22. For example, the clones bK415G2 and bK221H2 are located on human 22q distal to clones bK65A6 and bK125H2. On Siamang 18q, clones bK415G2 and bK221H2 are clearly distal to clones bK65A6 and bK125H2 within the human 22q homologous region. Similarly, the clone bK236H12 is located on human 22q proximal to clone bK799F10. On Siamang 18p, bK236H12 is proximal to bK799F10 within the human 22q homologous region.

The BAC probes bK212A2 and bK236H12 are located in 22q12 and 22q13.1, respectively. The hybridisation of BAC probe bK212A2 to Siamang 18q and bK236H12 to Siamang 18p indicated that the rearrangement breakpoint is located in the region (approximately 3-4 Mb) of 22q between those two clones. The BAC probes proximal to bK212A2 on human 22q all hybridised to Siamang 18q and the BAC probes distal to bK236H12 all hybridised to Siamang 18p.

4.2.2 Analysis of Siamang 18 with clones from human 22q12-13.1

In order to narrow down the position of the human 22q homologous block rearrangement point in Siamang chromosome 18, eleven new bacterial clones were selected for a second round of FISH screening. The one BAC, three PAC, four fosmid and three cosmid clones were selected because they are located on human 22q evenly distributed between the BAC clones bK212A2 and bK236H12. These clones sampled the entire interval and the distance between them was approximately 300 to 400 kb. The DNA from each clone was isolated, biotinylated by nick translation then hybridised to human and Siamang metaphase chromosomes as described earlier (4.2.1). Ten human and Siamang metaphase spreads were analysed for each probe. All of the probes gave informative hybridisation signals on human chromosome 22 and Siamang 18.

The probes used and the corresponding Siamang chromosome 18 localisation, are tabulated in Table 4.2.

Table 4.2 Human bacterial clone probes from 22q12-q13.1 and their corresponding Siamang chromosome 18 localisation. Clones are listed according to their order from centromere to telomere on chromosome 22.

Human 22q clone	Siamang 18 localisation
BAC bK212A2	18q
Fosmid fF24E5	18q
Fosmid fF126G10	18q
Cosmid cE132D12	18q
PAC dJ293L6	18q
Fosmid fF4G12	18q
Fosmid fF45C1	18q + 18p
Cosmid cE81G9	18p
Cosmid cE146D10	18p
PAC dJ151B14	18p

Six of the clones (bK212A2, fF24E5, fF126G10, cE132D12, dJ293L6 and fF4G12) hybridised entirely to Siamang 18q and three (cE81G9, cE146D10 and dJ151B14) hybridised entirely to Siamang 18p. Those clones lie proximal and distal to the human 22q homologous block rearrangement breakpoint, respectively. The fosmid clone fF45C1, which lies distal to fF4G12 and partially overlaps cE81G9 on human 22q13.1, hybridised to both the long arm and the short arm of Siamang 18 (see Figure 4.2).

*Figure 4.2 (*next page) Summary of the second round FISH analysis of Siamang metaphase chromosomes. At the bottom of the figure coloured bars indicate the location of fF4G12, fF45C1 and cE81G9 on human chromosome 22 and the overlap between fF45C1 and cE81G9. The dark blue and light blue bar colours indicate the regions proximal and distal to the breakpoint on chromosome 22, respectively. At the top of the figure four corresponding Siamang chromosome 18 images show the FISH signal location from each clone. The chromosome 18 images are ordered from left to right according to the location of the clone on human chromosome 22.

cE81G9 hybridises
to
HSY 18p

fF45C1 hybridises
to
HSY 18p and 18q

fF4G12 hybridises
to
HSY 18q

cE81G9

FF45C1

fF4G12

HSA22q13.1

CENTROMERE

TELOMERE
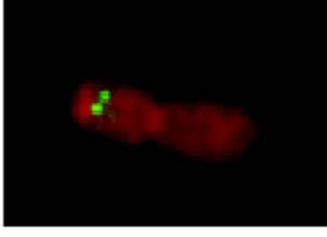
In order to produce hybridisation signals on both Siamang chromosome 18 arms, fF45C1 must contain sequences homologous to sections of both 18q and 18p. It seemed possible that fF45C1 contains the sequence in human 22q13.1 that spans the homologous block rearrangement point in Siamang chromosome 18. On further careful analysis of the Siamang gibbon metaphase spreads after hybridisation with fF45C1, it was apparent by eye that the FISH signal in 80% of metaphases analysed was weaker on 18p than on 18q. In the other cases, the signals appeared to be of equal intensities.

4.3   FISH Analysis of Dog Chromosomes with Human 22q13.1 Bacterial Clones

4.3.1   FISH analysis of Dog chromosomes with individual bacterial clones

In Chapter 3, the reciprocal zoo-FISH analysis between human chromosome 22 and dog chromosomes 10 and 26 indicated that the junction between material homologous to dog chromosomes 10 and 26 was located in 22q13.1. In the previous section of this chapter, human fosmid clone fF45C1 was identified as containing sequences spanning the evolutionary homologous block rearrangement point in Siamang chromosome 18. Fosmid fF45C1 was localised to human chromosome 22q13.1. Based on those results, clones from the same region of human 22q13.1 were selected for FISH analysis of dog metaphase chromosomes to investigate the possibility that the rearrangement in Siamang chromosome 18 had reverted the human chromosome 22-specific material back to it's ancestral state, represented in the dog.

Three human chromosome 22 clones (BAC bk256C5, fosmid fF45C1 and cosmid cE81G9) were selected for hybridisation individually to dog metaphases. The BAC clone bk256C5 contains sequences, which span those cloned in fF45C1 and cE81G9. The reason for selecting the larger-insert BAC clone was to provide a longer stretch of sequence for hybridisation to the dog chromosomes, which might prove to be more successful than hybridising the shorter fosmid and cosmid clones.

Up to 250 ng of biotinylated DNA from each clone were hybridised to dog metaphase chromosomes, but no clear FISH signals were visualised for any of the individual clones.

4.3.2 FISH analysis of Dog chromosomes with contiguous bacterial clones from human chromosome 22q13.1

As the hybridisation of individual human chromosome 22q13.1 clones to dog metaphases was unsuccessful, an alternative approach was attempted by hybridising several contiguous bacterial clones from the region simultaneously.

A mixture of DNA from three bacterial clones contiguous on human 22q13.1 and overlapping fF45C1 was used for FISH analysis of dog metaphase chromosomes. The three contiguous clones used were BAC bK833B7, PAC dJ394H8 and PAC dJ1170K4. The hybridisation of these three clones together provided the equivalent of a probe 450 kb in length. The sequence of clone fF45C1 is located near the centre of the three contiguous clones.

If the human 22q homologous block rearrangement breakpoint identified in fF45C1 is also present in the canine karyotype the mixed clones could generate hybridisation signals on dog chromosomes 10 and 26. Each signal would be generated by DNA at least 200 kb in length, which is of sufficient size to visualise by fluorescence microscopy after cross-species FISH (Matthew Breen, personal communication).

Biotinylated DNA from the three clones was mixed and tested by hybridisation to human and Siamang metaphase spreads. After detection, ten metaphase spreads were analysed each. A single bright hybridisation signal on HSA 22q13.1 and two bright hybridisation signals on Siamang 18p and 18q were revealed. The contiguous clone signals appeared in the same location on Siamang 18 as the two signals generated by fF45C1.

Biotinylated DNA from the contiguous clones was mixed and hybridised to canine metaphase spreads. After detection, one bright specific hybridisation signal was revealed, only on dog chromosome 26. (Figure 4.3)
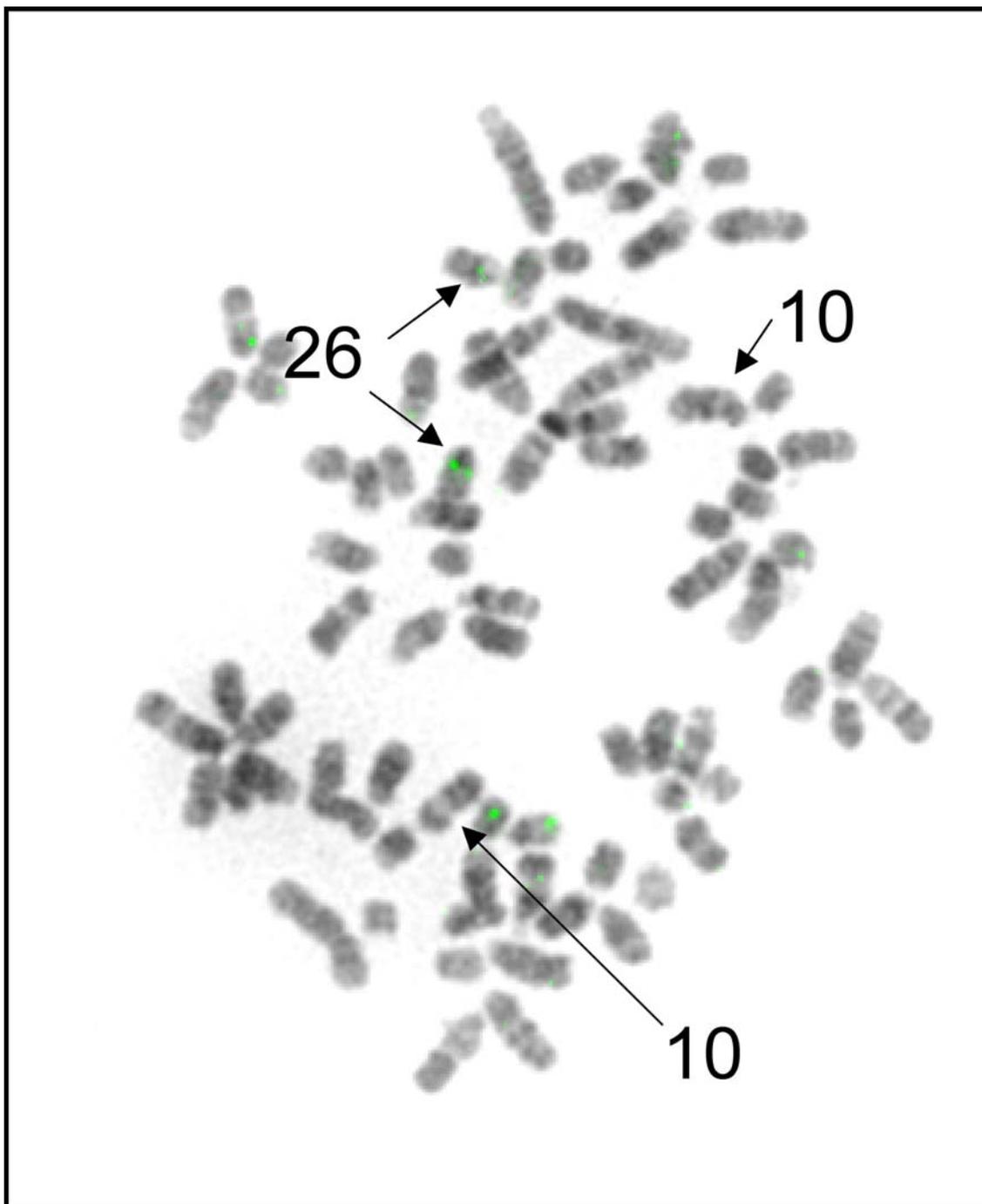
Figure 4.3 Hybridisation signal (pseudocoloured green) of three contiguous human

chromosome 22 BAC clones onto a canine metaphase spread. Canine chromosomes 10

and 26 are indicated (arrows).

Thirty metaphase spreads were analysed and although there were several non-specific signals on other dog chromosomes (see Figure 4.3), no specific signal was detected reproducibly on dog chromosome 10.

One possible explanation for this observation might be that the sequence on dog chromosome 10 has diverged sufficiently so that the probe did not hybridise well at the stringency used for the study. Alternatively, it could be explained if the rearrangement breakpoint identified in human chromosome 22q13.1 was not present in the dog karyotype.

4.3   Discussion

In Chapter 3, low-resolution cytogenetic analysis by reciprocal zoo-FISH indicated that the rearrangement points between human chromosome 22 and its two syntenic blocks in the dog lie within 22q13.1. In this chapter, a higher resolution FISH analysis has been carried out on Siamang chromosome 18 using bacterial clones from human chromosome 22q. This resulted in the identification of a fosmid clone from human 22q13.1, which hybridised to both Siamang chromosome 18 arms, which therefore is likely to contain sequences, which span the homologous block rearrangement point.

The evolutionary rearrangements that gave rise to the situation in the three species are most likely to have occurred as independent events. Nevertheless, the similar location of the ends of the gibbon and dog syntenic blocks on human chromosome 22q13.1 meant that a higher resolution analysis of the dog was needed to provide more information. In fact, the hybridisation of clones containing sequences spanning the rearrangement point in the gibbon identified only a single location in the dog genome.

The majority of previous cross-species FISH studies have relied on the use of YAC clones due to their availability and the size of their inserts. Distantly related species are likely to possess more genome sequence divergence than closely related species.

Therefore, the use of large-insert clones for cross-species FISH studies probably increases the chances of successful probe hybridisation and subsequent specific FISH signals. Although there are a considerable number of YACs available, which form part of the human chromosome 22 framework map (Dunham, *et al.,* 1999), BACs were selected to carry out the first cross-species FISH analyses of Siamang chromosomes. It was a well-established fact that at least 30% of YAC clones are chimaeric. The aim of the work for this chapter was to map the human 22q homologous block evolutionary rearrangement point in Siamang chromosome 18. One of the key indicators that a clone contains sequences, which span a breakpoint would be the occurrence of two FISH signals on Siamang 18. Without a guarantee that the YACs were not chimaeric, it was felt that their use could have reduced the reliability of the cross-species FISH results. The occurrence of chimaerism in BAC clones had not been reported.

The size of insert in a BAC clone can range from 150 to 500 kb. PAC inserts range from 30 to 300 kb, whilst cosmid and fosmid clone inserts range from 35 to 45 kb. The clones used for this study harbour inserts only a fraction of the size of the average CEPH megaYAC clone. However, as long as reliable signals were generated by FISH on to Siamang metaphase chromosomes, there was an advantage to using smaller insert clones. The fact that one clone was identified, which generated signals on both arms of Siamang 18, and, therefore contained sequences spanning the homologous block rearrangement point, meant that the location of breakpoint in human chromosome 22 was narrowed down to within a smaller interval than would have been possible with a larger-insert clone.

For the FISH studies, clones were selected which had been used successfully before as part of the mapping process of human 22. For the first round of screening, the thirteen BAC clones were selected on the basis of their mapped location on human 22q and the quality of the FISH signal. Although all the BACs gave good strong signals on human metaphases, after hybridisation to Siamang metaphase chromosomes, four of the BAC clones gave dissipated FISH signals with high background. Therefore, it was concluded

that the hybridisation problems for the four clones was probably due to sequence differences which might have arisen since the ancestors to the lesser apes and great apes diverged.

At the time of these studies, there was no standard karyotype for the Siamang gibbon. However, for the purposes of these studies, it was sufficient to map the BAC clones relative to the centromere and each other within the human 22q homologous blocks on Siamang 18p or 18q. All of the BAC clones appeared to have retained their orientation with respect to the centromere, telomere and each other in the 22q homologous regions on Siamang 18. These results are consistent with there not having been any intrachromosomal rearrangements within the two blocks after the gross rearrangement.

By mapping the ordered BAC clones to Siamang 18p or 18q, it was possible to narrow down the location of the homologous block rearrangement breakpoint to between two BACs (bK212A2 and bK236H12) located in 22q12 and 22q13.1, respectively. In order to narrow down further the position of the breakpoint, eleven new bacterial clones, distributed between bK212A2 and bK236H12, were selected for the second round of FISH screening. Seven of those clones were fosmids or cosmids with relatively small insert sizes of 35 to 45 kb each. There was concern that those clones would be too small to generate meaningful FISH signals. However, all of the clones in the second round of screening gave informative hybridisation signals.

The fosmid clone fF45C1 (mapped to human chromosome 22q13.1) generated signals on Siamang 18p and 18q. The signals on each arm were adjacent to the junction between the human 22q- and human 16p-homologous segments. Fosmid fF45C1 was thus identified as likely to contain sequences spanning the 22q-homologous block breakpoint. From the observations that the FISH signal in 80% of the Siamang metaphases was weaker on 18p than on 18q, it could be speculated that the sequence spanning the homologous block breakpoint was positioned towards the distal end of the fosmid.

Having identified a clone containing sequences spanning an evolutionary chromosome rearrangement point in the Siamang, the next stage of the work towards analysing the nature of the syntenic block boundaries was to identify and analyse the Siamang sequences, which span the homologous block junctions. This work is described in the next two chapters.

CHAPTER 5

Cloning the homologous block rearrangement breakpoints in Siamang chromosome 18

5.1 Introduction

In chapter 4, the human chromosome 22q clone fF45C1 was identified by FISH, which contains sequences homologous to those that spanned an evolutionary rearrangement breakpoint in Siamang chromosome 18. In order to analyse the underlying mechanism which caused the rearrangement, the homologous block junctions need to be analysed at the sequence level.

5.1.1. Strategies for Analysing Sequences at the Homologous Block Junctions

The aim of the work described in this chapter was to use two different approaches to map the homologous block junctions in Siamang chromosome 18 and isolate the gibbon sequences spanning the fusion points between chromosome 22-homologous and non-homologous regions.

The first approach was to use PCR to assay for Siamang STSs on chromosome 18, using primer pairs defining STSs at 1 kb intervals in fF45C1. Long-range PCR would be used to amplify overlapping "tile-paths" of sequences from the two gibbon chromosome 18 segments. This would be carried out by using alternate sense and anti-sense primers, so that the sense primer of one STS would be used for PCR with the anti-sense primer of an adjacent STS. The rationale for this was that if alternate primers primed successfully in the human then priming might also be successful between the same alternate primer pairs in the gibbon. Assuming that was the case it seemed reasonable to expect that a failure of priming in the gibbon might indicate a discontinuity in the homologous sequences and that the primers lie either side of the homologous block breakpoint. In that case the two homologous sequences of DNA will be too far apart for extension to take place.

A prerequisite for the PCR approach was the design and use of primers defining STSs from fosmid fF45C1.

The second approach was to use radiolabelled Siamang STSs to identify and isolate Siamang cosmid clones containing sequences spanning the homologous block junction points. That approach required the construction and screening of a Siamang genomic cosmid library from high molecular weight genomic DNA.

The cloning capacity of the cosmid system is only a fraction of that which can be achieved using the large-insert PAC and BAC systems. But due to the high efficiencies of ligation, packaging and transfection, a cosmid library is much easier to prepare than a PAC or BAC library, even from small quantities of starting genomic DNA. A cosmid cloning system was chosen to be the most appropriate for this study as the target regions were small and it was envisaged that a Siamang genomic library would have only limited use.

The cosmid cloning system exploits the ability of λ bacteriophage to introduce a DNA molecule into an *E. coli* host cell (Collins and Hohn, 1978). The cosmid vector is a plasmid incorporating a 12 bp λ *cos* signal sequence, which is required for packaging the DNA into the bacteriophage pre-head structure. Most of the phage genome is excluded from the cosmid because there is no need to produce infectious virions following transfection of the host cell. The viral particles containing linear recombinant DNA molecules are assembled *in vitro*. Under the correct conditions these particles "infect" the host and introduce their DNA contents. Within the cell, the linear DNA is circularised and maintained as an extra-chromosomal plasmid. The bacteriophage head exerts a size limitation on the system as it will not accommodate a DNA molecule larger than 52 kb or smaller than 38 kb. Therefore there is no requirement to size-select the genomic DNA for ligation.

The Lawrist 16 vector (Yokobata *et al.,* 1991) has dual *cos* sites (Bates and Swift, 1983). Thus the recombinant DNA molecule comprises a fragment of genomic DNA flanked by vector arms, each with its own *cos* site. Vector concatemer formation is minimal, and so the genomic DNA is dephosphorylated instead of the vector. As well as dual *cos* sites, Lawrist 16 also has a *Bam*HI site for cloning fragments obtained by partial *Mbo*I or *Sau*3AI digestion, an origin of replication that will function in the host cell and a Neo^R gene for kanamycin selection

in transfected *E. coli*. These features are accommodated in only 8 kb of DNA; thus the cloned genomic DNA fragment can be as large as 44 kb.

An *E. coli* host strain that is suitable for the propagation of cosmid libraries has disabled recombination and restriction systems and also can be infected at high efficiency by λ phage. An *E. coli* strain with the appropriate genotypic features, which has been widely used for cosmid library propagation is DH5αMCR.

For libraries with very limited usage, the primary transfectants can be plated directly onto membranes, grown overnight, then used for preparation of replica membranes. The master membrane is stored at -70°C in the presence of glycerol, and positive clones are isolated from this membrane following hybridisation to its replica.

RESULTS

5.2   PCR analysis of Siamang chromosome 18

Fosmid fF45C1 is 45.681 kb in length and has now been fully sequenced at the Sanger Institute. At the time of this study, the clone had been through the sequencing pipeline but, due to secondary structures causing problems with the process, it was sequenced in two separate pieces A and B (Stephen Dodsworth, personal communication). Strand A was 6.355 kb and strand B was 38.926 kb in length. Based on the size of the insert and the two sequences, the gap between the two strands was estimated to be approximately 400 bp in length (Stephen Dodsworth, personal communication). Fosmid fF45C1 contains both Alu and LINE repeats, as well as a Colony Stimulating Factor (CSF2RB) gene and CSF-like pseudogene.

5.2.1 STS design and verification

Forty STSs with an average size of 141 bp and spaced at intervals of approximately 1 kb were identified (Sarah Hunt, Human Genetics Informatics, Sanger Institute) from the known sequence of strand A (A1 to A6) and strand B (B1 to B38) of fF45C1 (the sequences of the primers, plus the size of each STS are listed in Appendix I). The intervals between B16 and B18, B23 and B25, B25 and B27, and B28 and B30 were approximately 2 kb. For PCR assays of the STSs, primer pairs ("sense" and "anti-sense") were designed (Sarah Hunt, Human Genetics Informatics, Sanger Institute). Due to the repetitive nature of some of the fF45C1 strand B sequence, STSs B17, B24, B26 and B29 could not be designed. However, primer pairs defining 40 STSs were considered sufficient to carry out the first round of PCR analysis of the regions of Siamang chromosome 18 homologous to fF45C1.

The STSs were tested for their specificity by PCR on human genomic DNA, human chromosome 22-hybrid DNA and hamster genomic DNA. The appropriate reaction conditions were established to ensure that each assay amplified the expected size product prior to their use in characterising Siamang chromosome 18.

All 40 primer pairs amplified the predicted sized product from human genomic and human chromosome 22-hybrid DNA.

5.2.2 STS-PCR mapping of Siamang DNA

The STSs were tested for their presence on Siamang genomic DNA by PCR. Thirty-six of the primer pairs (90%) generated a specific product. Nine of the 36 successful primers (A1, B5, B11, B19, B20, B33, B35, B36 and B37) generated a specific product, as well as at least one other product. Five of the primer pairs (A3, B3, B5, B16 and B19) generated a product, which was a different size to that generated from human DNA, suggesting that a degree of sequence divergence had taken place. The STS-PCR results are summarised in Figure 5.1. The annealing temperatures most suited to each primer pair are tabulated below (Table 5.1).

Table 5.1 (next page) Optimal annealing temperature for each primer pair for human chromosome 22 and gibbon genomic DNA as PCR template. The optimal annealing temperature was selected as that which generated a single product. In the cases where more than one product was generated at all three temperatures, the optimal annealing temperature was chosen that gave the most specific product.

O.A.T.   – optimal annealing temperature (˚C)

nsp   – no specific product

[a] More than one product generated from human genomic DNA

[b] More than one product generated from human chromosome 22 hybrid DNA

[c] More than one product generated from Siamang genomic DNA

[d] Product from gibbon DNA is a different size to that from human DNA

Primer Pair	O.A.T. Human (˚C)	O.A.T. Gibbon (˚C)	Primer Pair	O.A.T. Human (˚C)	O.A.T. Gibbon (˚C)
A1	65[a]	55[c]	B15	65	65
A2	60	65	B16	65[a]	65[d>]
A3	65	65[d>]	B18	65	nsp
A4	60	60	B19	65	65[c,d<]
A5	65	60	B20	65	60[c]
A6	65	65	B21	65	nsp
B1	65	65	B22	65[a,b]	nsp
B2	65	65	B23	65	nsp
B3	65	65[d<]	B25	65	65
B4	60	60	B27	60	60
B5	65	60[c,d<]	B28	60	60
B6	65	65	B30	60	60
B7	60	60	B31	60	60
B8	65	65	B32	60	60
B9	65	65	B33	60	60[c]
B10	60	60	B34	60	60
B11	65	65[c]	B35	60	60[c]
B12	60	60	B36	65[a,b]	65[c]
B13	65	65[c]	B37	65	65[c]
B14	65	65	B38	65	65

5.2.3 Long-range PCR assays of Siamang DNA

Based on the results from the STS analysis of Siamang DNA, alternate sense and anti-sense primers from fF45C1 STSs were used for longer-range PCR assays to amplify overlapping sequences from the two Siamang chromosome 18 segments. Individual oligonucleotides were mixed to collectively define overlapping sequences. For example, A1 sense was mixed with A2 anti-sense (A1 + A2), A3 sense with A4 anti-sense (A3 + A4), and so on.

For one reaction (A6 + B1) the expected PCR product size in human was approximately 1.4 kb. For four reactions (B16 + B18, B23 + B25, B25 + B27 and B28 + B30), the expected PCR product sizes were approximately 2 kb. For all other reactions, the expected PCR product sizes in human were approximately 1 kb. Long-range PCR assays were carried out on high molecular weight human and Siamang genomic DNA. The results of the long-range PCR assays are summarised in Figure 5.1 and presented as STS maps of human chromosome 22 clone fF45C1 and part of Siamang chromosome 18.

Figure 5.1 (next page) STS-PCR and long-range PCR map of fF45C1 and Siamang chromosome 18. The fF45C1 and Siamang chromosome 18 STSs are indicated by orange and purple boxes, respectively. The successful long-range PCR amplification between an alternate primer pair is indicated by a brown bar beneath two STSs for human DNA and by a green bar above two STSs for Siamang DNA. Stretches of gibbon sequence which failed to amplify are indicated by red arrows.
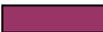
Twenty-nine of the 39 alternate primer pairs generated product of the expected size from human genomic DNA, and twenty-four from Siamang DNA. Two of the alternate primer pairs (A5 + A6 and A6 + B1) generated PCR product from Siamang DNA but not from human under long-range conditions. It might be that the sequence-feature in human that caused the original problem with sequencing fF45C1 also caused a problem in PCR but is not found in the gibbon sequence at that point. Alternatively, it could be that the PCR failures might be due to sequence divergence between the human and gibbon genomes.

The rationale for this long-range PCR analysis was that if an STS failed to amplify in the gibbon it might be implicated in the rearrangement point. But there were in fact five stretches of gibbon sequence, ranging in size from approximately 1 to 7 kb, which failed to amplify, listing below:

1. The expected size product from human DNA was generated by B1 + B2 primers, but no product was generated from gibbon DNA.
2. No product was generated from human or gibbon DNA by B4 + B5 or B5 + B6 primers.
3. No product was generated from human DNA by B15 + B16 or B16 + B18 primers. No product was generated from gibbon DNA by B15 + B16, B16 + B18 or B18 + B19 primers.
4. No product was generated from human DNA by B23 + B25 or B25 + B27. No product was generated from gibbon DNA from B20 + B21, B21 + B22, B22 + B23, B23 + B25 or B25 + B27.
5. No product was generated from human DNA by B28 + B30. No product was generated from gibbon DNA by B28 + B30, B30 + B31 or B31 + B32.

The individual primer pairs defining STSs B1, B2, B4, B5, B6, B15, B16, B19, B20, B25, B27, B28, B30, B31, and B32 generated the expected size product from Siamang DNA under normal PCR conditions.

Although the rationale was that a region, which fails to amplify in the gibbon, might be implicated in the rearrangement point, a region which fails in both human and gibbon could also be implicated. The only regions that can be excluded as the breakpoint are those which amplify the long product in the gibbon. Therefore, from the results of the long-range PCR analysis of Siamang DNA it was only possible to exclude 24 kb of fF45C1 sequence.

Because the long-range PCR approach did not provide a definitive answer in locating the position of the breakpoint in Siamang chromosome 18, an alternative approach was required, which was to generate a genomic gibbon library and screen for clones containing sequences spanning the fusion points.

5.3   Construction of Siamang genomic cosmid library

5.3.1 Partial digestion of high molecular weight Siamang DNA

High molecular weight Siamang gibbon genomic DNA was extracted from cultured lymphoblastoid cells, and the OD_{260}/OD_{280} value was 1.875 indicating that there was no significant protein contamination in the DNA.

To assess its suitability for a competitive restriction digestion (see below), the high molecular weight DNA was subjected to restriction endonuclease digestion with *Mbo*I, both before and after methylation treatment with *dam* Methylase. The enzymes were included in excess to ensure that the reaction went to completion and that no *Mbo*I restriction sites remained which had not been cut or methylated. After gel electrophoresis it was deduced that the Siamang DNA had behaved similarly to human DNA. It was clear that *Mbo*I cut as predicted and that *dam* Methylase methylated and, thus, protected the Siamang DNA from digestion (Figure 5.2).

Figure 5.2 (next page) Photograph of 0.7% agarose gel of electrophoresed restriction digestion products. M= marker consisting of *Hind*III digested λ DNA (marker band sizes are indicated). U/T= untreated starting DNA. A= DNA treated with *Mbo*I. B= untreated control DNA. C= *Mbo*I enzyme only (no DNA). D= DNA treated with *dam* Methylase and *Mbo*I.
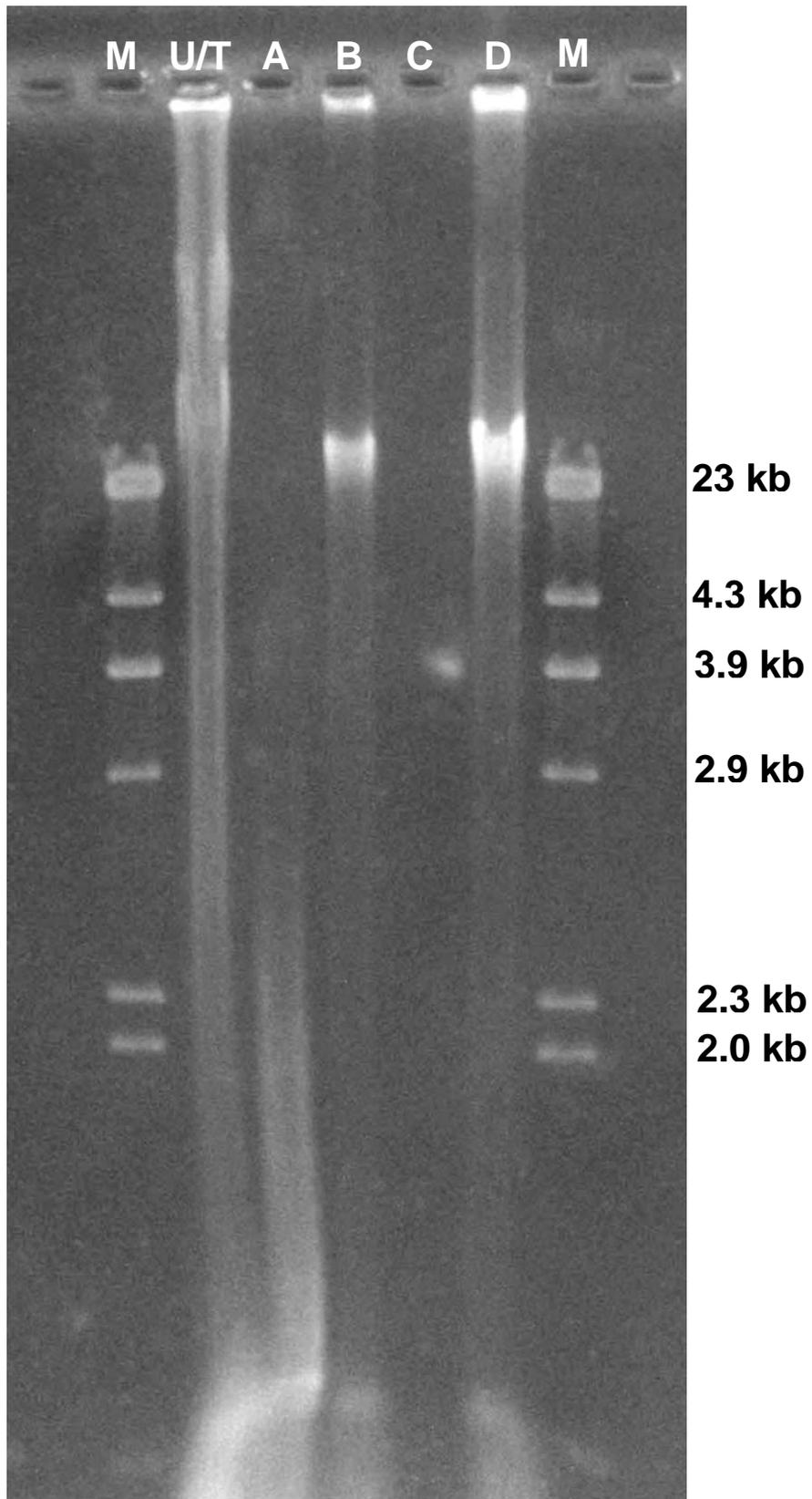
The DNA was cut by *Mbo*I to generate fragment sizes up to 4 kb. The DNA without enzyme resolved into a band, greater than 23 kb in size, as did the methylated DNA. Hoheisel *et al.* (1989) state that 6.25 units of *Mbo* I digests 1 µg of human genomic DNA to completion i.e. one cut in every 400 to 500 bp. Based on the current knowledge of human DNA, it was assumed that the Siamang DNA also has a *Mbo*I site approximately every 400 bases. In order to generate DNA fragments of sufficient size to clone in a cosmid library, DNA fragments of 40 to 50 kb were required. Therefore, it was necessary to digest the Siamang DNA at every 100th site to generate appropriately sized products.

5.3.1.1 Competitive digestion

The first approach used to digest the DNA was the competing *Mbo*I/*dam* methylase method described by Hoheisel *et al.* (1989). This was tried because competitive digestion was reported to be more controllable than methods relying on the use of limiting time or enzyme concentration. The technique was being used routinely in-house for the construction of flow-sorted chromosome-specific libraries (Ross and Langford, 1997).

A range of digestion and methylation conditions was tested, involving different unit ratios of *Mbo*I to *dam* Methylase (1:150, 1:75 and 1:300) and containing approximately 150 ng of DNA. The competing reactions of cleavage and methylation ran to completion. Using those conditions it was found that the Siamang genomic DNA cut too readily and fragment sizes were actually quite difficult to control. Even the products of the reaction containing the greatest ratio of *Mbo*I to *dam* Methylase (1:300) ranged in size from 2.3 to 23 kb, and were over-digested to consider for cloning.

The volume of *dam* Methylase in the 1:300 unit ratio reaction (9.4 µl) was approximately one fifth of the total reaction volume (50 µl). The combination of 0.25 U of *Mbo*I with 75 U of *dam* Methylase resulted in a 10.4% v/v final concentration of glycerol. In order for the restriction enzymes to work effectively it was important to include less than or equal to 10% v/v glycerol in the reaction mix. It would have been impractical to set up reactions using higher ratios of *Mbo*I to *dam* Methylase without further increasing the proportion of glycerol. Therefore, it was decided not

to attempt further competitive reactions at higher ratios. Instead, a limiting enzyme approach was used by titrating *Mbo*I. This is described in the following section.

5.3.1.2 <u>Limiting enzyme</u>

The second approach to digest the DNA was to use a dilution series of the *Mbo* I restriction enzyme alone for a fixed length of time (1 hr). Initially, nine partial digestion reactions were carried out involving a titration of *Mbo* I from 2 U down to 0.0078125 U, but in all of those reactions the DNA was over digested and ran ahead of the 23.1 kb size marker (Figure 5.3).

Figure 5.3 (next page) 0.3% agarose gel of electrophoresed restriction digestion products. M= marker consisting of *Hind*III digested λ DNA (marker band sizes are indicated). Samples as indicated. C= control lane, untreated DNA.
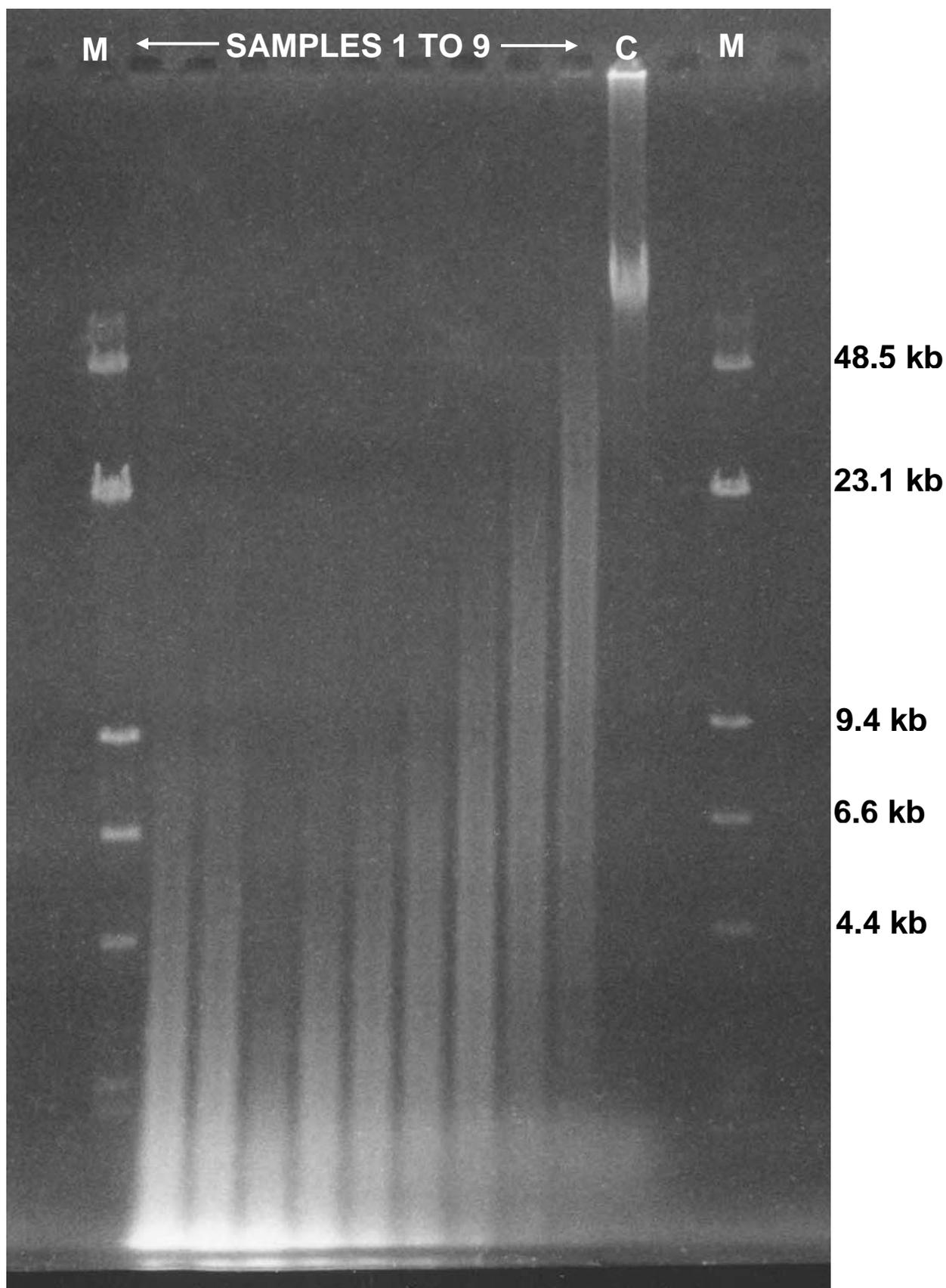
A further nine reactions were carried out titrating *Mbo*I from 0.05 U down to 0.0001953125 U. Reactions 5 and 6 (0.003125 U and 0.0015625 U of *Mbo* I, respectively) produced optimal sized Siamang DNA restriction fragments for use in library construction (Figure 5.4). The optimal restriction reactions showed clear digestion, as in lanes 5 and 6, but with no evidence of DNA running ahead of the 23.1 kb λ *Hin*d III fragment.

Figure 5.4 (next page) 0.3% agarose gel of electrophoresed restriction digestion products. M= marker consisting of Hind*III digested* λ *DNA (marker band sizes are indicated). Samples as indicated. C= control lane, untreated DNA.*
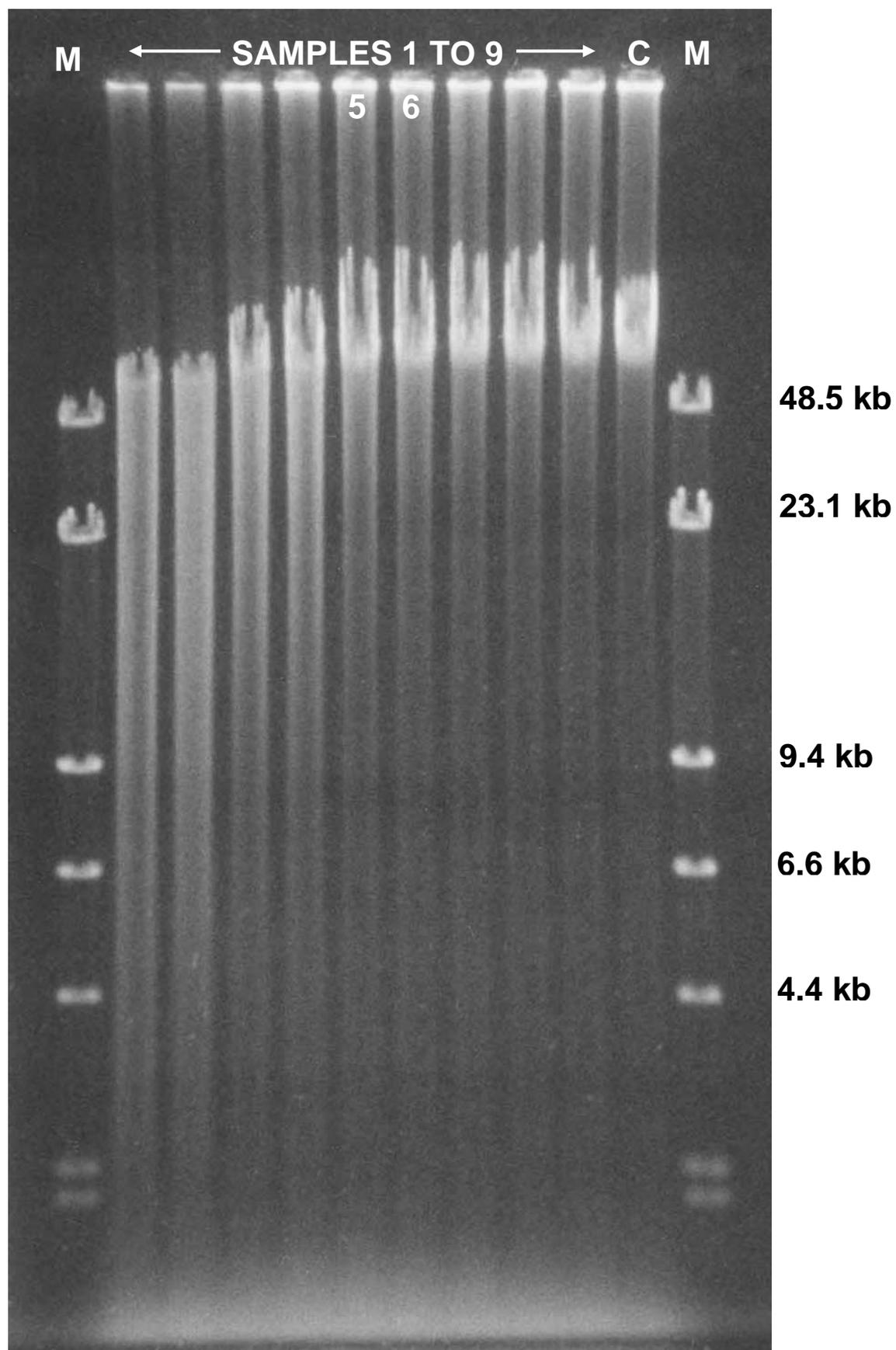
5.3.2 Ligation and packaging

Four reactions (ligation 1 to 4) were carried out to ligate the partially digested Siamang DNA to linearised Lawrist 16 vector arms. DNA from partial digestion reaction 6 (see *5.3.1.2*) was used for ligation 1 and 2, and DNA from digestion reaction 5 was used for ligation 3 and 4. In order to increase the chance of a successful out-come, each ligation reaction used a different nanogram ratio of vector arms to DNA. The ratios of arms to DNA for Ligation reactions 1 to 4 were 1:1, 8:1, 1.8:1 and 6:1, respectively.

Nine packaging reactions (HSY 1 to 9) of the ligated DNA were subsequently carried out to package the ligations into infective particles ready for plating on *E. coli*. 400 ng of DNA from ligation 1 was used for each packaging reaction HSY1 and HSY2. 63.3 ng of DNA from ligation 2 was used for HSY3, HSY4 and HSY5. 50 ng of DNA from ligation 3 was used for HSY6, HSY7 and HSY9 and 50 ng of DNA from ligation 4 was used for HSY8. In order to analyse the viability of commercially available packaging kits three different kits were used for the reactions. Gigapack Gold II was used for HSY1, HSY2, HSY3, HSY4, HSY5 and HSY6. Gigapack Gold IIXL was used for HSY7 and Gigapack Gold III was used for HSY8 and HSY9. (All three packaging kits were purchased from Stratagene).

5.3.3 Assessing the titre of the library

The best measure of the success of a library is determined by the genome coverage value. This requires the calculation of the total number of recombinant clones and the average clone insert size. The efficiency of cosmid library preparation, expressed as colony forming units (cfu) per µg of starting genomic DNA, is expected to lie between $5x10^5$ and $1x10^7$ for human DNA. The vector-arms self-ligation test (Section 2.11.3) gives a measure of the expected background of non-recombinants in the experiment. These non-recombinants can arise from vector concatemers caused by a failure to completely CIAP-modify the Lawrist 16 *Sca*I restriction sites.

Plating an aliquot of the packaged phage on *E. coli* gives the titre of the library and provides clones for an assessment of library quality (by FISH and restriction digestion). To test the titre of the nine packaging reactions, a sample of each was plated on *E. coli* and the results of the library titre assessments are summarised in Table 5.2.

Table 5.2 Assessment of the titres of each packaging reaction.

Packaging reaction	Reaction Volume (µl)	Ligated DNA (ng)	Colony Count	cfu[a] per µg DNA
HSY1	8	400	330	5.4×10^5
HSY2	8	400	260	4.3×10^5
HSY3	6.6	63.3	11	1.1×10^5
HSY4	6.6	63.3	9	9.4×10^4
HSY5	6.6	63.3	6	6.2×10^4
HSY6	8	50	67	8.8×10^5
HSY7	8	50	68	8.9×10^5
HSY8	4	50	32	4.2×10^5
HSY9	8	50	43	5.7×10^5

[a] cfu – colony forming units per µg of starting genomic DNA was calculated as follows: First, the total number of potential colonies for each packaging reaction was calculated by multiplying the number of colonies by 660 (the total available volume of each packaged extract). That value was then multiplied by 2.5 (for 400 ng ligated DNA), 15.798 (for 63.3 ng) or 20 (for 50 ng).

The background level of non-recombinants for ligation reactions 1 and 3 were defined by the number of colonies generated by the vector arms self-ligation reactions. The non-recombinants for ligations 2 and 4 were not assessed. The result obtained with the vector self-ligation should

indicate that less than 1% of clones are non-recombinant (Ross and Langford 1997). Less than 0.3% of clones from ligation reaction 1 and less than 0.1% of clones from ligation 3 were non-recombinant (data not shown).

The two packaging reactions HSY6 and HSY7 produced the highest titres (8.8×10^5 and 8.9×10^5 cfu's, respectively). They were both carried out in a reaction volume of 8 µl with 50 ng of ligated DNA (1.8:1 ratio of arms to DNA), using Gigapack Gold II (HSY6) and Gigapack Gold IIXL (HSY7) packaging kits. That result indicated that there was no advantage over the Gigapack Gold II in using the Gigapack Gold IIXL kit.

The next highest titre was produced by HSY9 (5.7×10^5 cfu's). That was also carried out in a reaction volume of 8 µl with 50 ng of DNA from the same ligation reaction (1.8:1 ratio) as for HSY6 and HSY7. The only difference, which may account for the lower titre, was the use of the Gigapack Gold III kit.

The packaging reactions HSY6 and HSY7 were approximately twice as efficient as reactions HSY1, HSY2 and HSY8 (5.4×10^5, 4.3×10^5 and 4.2×10^5 cfu's, respectively). The HSY1 and HSY2 reactions were also carried out in a volume of 8 µl, using the Gigapack Gold II kit, but a total of 400 ng of DNA ligated in a ratio of 1:1 was used, which may explain the lower titre.

The Gigapack Gold III kit was used for the HSY8 reaction and was carried out in a volume of only 4 µl, using 50 ng of DNA ligated in a ratio of 6:1.

Packaging reactions HSY3, HSY4 and HSY5 generated the lowest titres (1.1×10^5, 9.4×10^4 and 6.2×10^4 cfu's, respectively). Those reactions were carried out using the Gigapack Gold II kit, but in a reaction volume of 6.6 µl using 63.3 ng of DNA ligated in a ratio of 8:1.

It was concluded that the highest library titres were generated using 50 ng of DNA from a ligation reaction with a low ratio of vector arms to DNA (e.g. 1:1 or 1.8:1) and packaged in a volume of 8 µl using either Gigapack Gold II or IIXL kits.

5.3.4   *Assessing the integrity of Siamang genomic cosmid clones*

5.3.4.1 <u>EcoR I digestion to estimate insert size</u>

The average insert size was estimated for thirty clones picked from the test plating for packaging reaction HSY1. Colonies (HSY1A1 to HSY1C6) were picked into LB freezing broth and individually grown up overnight. DNA was prepared by a standard alkaline-lysis procedure then digested with *EcoR* I. Samples of the cultures were stored frozen at −70˚C.

The products of the clone restriction digestion were analysed by electrophoresis together with size standards (1kb ladder and *Hin*d III digestion of λ DNA), on a 0.7% agarose gel. The gel was stained with 0.5 µg/ml Ethidium Bromide and the bands visualised under U.V. light (Figure 5.5).

Figure 5.5 (next page) 0.7% agarose gel with *Eco*RI digestion products for gibbon genomic cosmid clones. M= size marker. Samples as indicated.

178

M   M ← Samples 1 to 15 → MM ← Samples 16 to 30 → MM
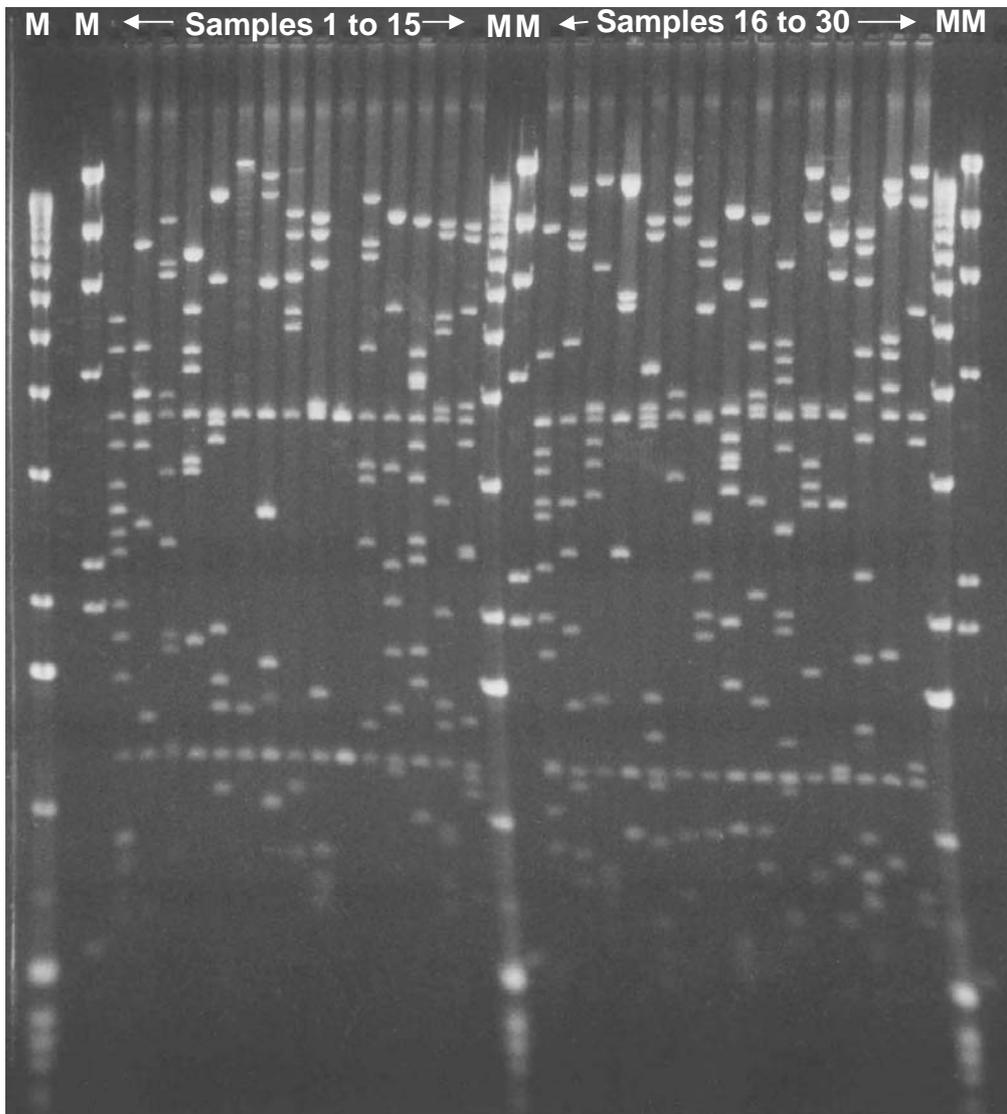
Due to the complexity of the gel it was difficult to manually estimate the total size of the inserts for each of the clones analysed. Further analysis of the clones was carried out by another *EcoR* I digestion followed by electrophoresis on a 0.7% agarose gel. Following staining with a 1:10,000 dilution of Vistra Green™ in TE, the gel was scanned automatically using a FluorImager scanner (Molecular Dynamics Inc). The scanner analysis software provided an estimate of the band sizes for each clone analysed (data not shown).

For Lawrist 16, insert sizes normally lie between 33 and 44 kb. One of the clones, (A10), had no insert and only vector bands were visible on the gel in the corresponding lane. The average estimated size of insert of the remaining 29 clones analysed was 37.1 kb, which is within the range expected.

Although the insert size figure of 37.1 kb was calculated from clones from a single library (HSY1), the same value was used to estimate the coverage of all the libraries, based on the assumption that HSY2 to HSY9 all contained similar sized inserts to HSY1. Coverage (C) was calculated from the equation:

$$C = [N \times S]/G \text{ where}$$

N =the total number of clones in the library

S = the average clone insert size, and

G =the size of the starting genome

The estimated genome coverage of each library is tabulated below (table 5.3).

Table 5.3 Estimated genome coverage of each library

Packaging reaction	Genome coverage
HSY1	2.7
HSY2	2.1
HSY3	0.09
HSY4	0.07
HSY5	0.05
HSY6	0.55
HSY7	0.56
HSY8	0.26
HSY9	0.35

Based on the assumption above, the total coverage of all the libraries together was calculated to be 6.73. However, as different genomic digests were used for the different libraries, the calculation was treated as an estimate. Library HSY7, which used the Gigapack Gold IIXL packaging system might be expected to have a higher average insert size, although this was not assessed.

The size of the starting genome determines the number of clones required for a given coverage, which in turn determines the probability of identifying a specific clone. The theoretical probabilities of finding at least one clone containing a given target sequence are 0.632, 0.993 and 0.999, respectively, for libraries with C values of 1, 5 and 7 (Mark Ross, personal communication). The overall Siamang genomic library coverage was 6.73, which provided a probability between 0.993 and 0.999 of identifying a clone from a specific point in the gibbon genome. It was concluded that

enough clones were present in the libraries to progress onto the screening part of the project (section 5.4).

5.3.4.2 <u>FISH analysis to test clone specificity</u>

FISH analysis of the cosmid DNA samples on metaphase chromosome spreads provides a measure of the frequency of chimaeric clones. DNA isolated from cosmid clones HSY1A1 to HSY1A10 was biotinylated by nick translation and hybridised to Siamang and human metaphase chromosomes. As would be expected, the probe for clone A10 gave no hybridisation signal on the Siamang chromosomes. One of the other clones hybridised to the end of every Siamang chromosome, but did not hybridise to any human chromosomes. It seems likely that it contains a heterochromatic/telomeric repeat, specific to the Siamang. The remaining eight clones hybridised with good, clear, single signals to both Siamang and human chromosomes.

5.4   Identification of Siamang cosmid clones spanning the homologous block junctions

5.4.1 Screening high-density filters

5.4.1.1 <u>Plating high-density filters</u>

The most suitable approach for screening the cosmid library was to hybridise radiolabelled gibbon STS probes to membrane filters carrying the DNA from lysed colonies. As a first attempt, it was decided to plate out and screen 2×10^5 colonies, which represented approximately 2.5 Siamang genome equivalents. The primary transfectants were randomly plated directly onto membranes on agar and grown overnight. The packaging reactions HSY1 and HSY2 were plated on *E. coli* to generate approximately 20,000 colonies on each of 10 master filters of 7.5 x 11 cm (SCHSY1 to SCHSY10). SCHSY1, 2, 3, 4 and 5 were generated from library HSY1, whereas SCHSY6, 7, 8, 9 and 10 were generated from library HSY2.

After the preparation of replica membranes, the master membrane was stored at -70°C in the presence of glycerol, so that clones could be isolated from this membrane following hybridisation to its replica. The replica filters, "SCHSYRep1" to "SCHSYRep10" were processed prior to screening.

5.4.1.2 High-density filter screening

The SCHSYRep filters were screened by hybridisation with pools of radiolabelled gibbon STSs. The PCR products were generated using the fF45C1 A4, B1, B10, B20 and B38 primer pairs on Siamang genomic DNA template. These primer pairs were chosen as they were distributed along the length of fF45C1 and spaced approximately 10 kb apart. The PCR products were excised from agarose gels then used as template in a second PCR containing a single radiolabelled dNTP (α-^{32}P dATP). The labelled products were pre-reassociated with gibbon genomic DNA, in order to compete out any repetitive sequences within the probes.

The probe pool was hybridised to the colony filters overnight at 65°C. Filters were then washed to remove any un-bound probe. After a 3.5-hour exposure, the autoradiograph showed a low background with numerous signals ranging in intensity. There were five very intense signals SCHSY4.1, SCHSY4.2, SCHSY5.1, SCHSY7.2 and SCHSY10.1. There were eighteen medium-intensity signals on SCHSY2, SCHSY4, SCHSY5, SCHSY6, SCHSY7, SCHSY8, SCHSY9 and SCHSY10. (see Figure 5.6 (left-hand panel) for an example filter). SCHSY1 and SCHSY3 had only numerous low-intensity signals.

Figure 5.6 (next page) From left to right, autoradiographs of high-density and low-density colony filters after probing with radiolabelled gibbon STS. In the left panel, strong signals are indicated with arrows. In the right panel, the signal from a colony, which was picked for subsequent analysis is indicated an arrow.
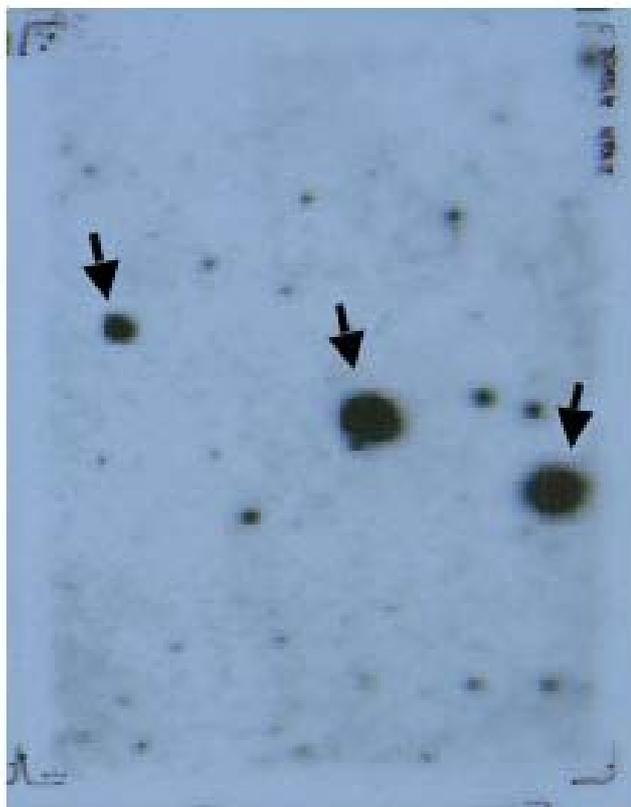
5.4.1.3 <u>Colony verification by PCR</u>

In an initial verification analysis, eleven "mixed colony" regions (SCHSY10.1 to SCHSY10.11), representing the full range of autoradiograph signal intensities, were picked from master filter SCHSY10, and streaked on LB agarose plates containing kanamycin. The entire areas demarcated by the autoradiograph signals (incorporating several colonies) were sampled from the master filter, to avoid the possibility of missing the positive colonies. After an overnight incubation, each mixed colony streak was sampled and analysed for the presence of gibbon STSs by colony PCR using fF45C1 A4, B1, B10, B20 and B38 primer pairs. Only SCHSY10.1 was positive containing A4 and B1 STSs. SCHSY10.1 was sampled from the region corresponding to the largest autoradiograph signal for filter SCHSYRep10.

Bearing the results of the initial PCR screen in mind, another twenty-one mixed colony regions were sampled and streaked from master filters SCHSY2, 4, 5, 6, 7, 8 and 9, representing the large and medium-sized autoradiograph signals. Each mixed colony streak was sampled and analysed by colony PCR as before.

After sampling for colony PCR, the mixed colony streaks were grown in liquid culture overnight prior to being frozen as "SCHSY mixed colony glycerols". SCHSY4.1, 4.3, 5.1 and 7.2 contained STSs A4 and B1. SCHSY4.2 contained STSs B10 and B20. All the other cultures tested were negative.

5.4.2 Screening low-density filters

5.4.2.1 <u>Estimating cell viability in SCHSY mixed colony glycerols</u>

From a $1/10^6$ dilution in LB broth containing kanamycin, the concentration of viable cells for each SCHSY4.1, 4.2, 4.3, 5.1, 7.2, 8.4 and 10.1 mixed colony glycerol was estimated. The mixed colony dilutions were stored at 4°C until used for plating out. The cell counts per ml for each mixed glycerol are tabulated below (Table 5.4).

Table 5.4 Cell counts per ml of each mixed colony glycerol*.*

SCHSY mixed colony glycerol	Cell count per ml	Vol. for 500 colonies[a] (µl)
4.1	9.9×10^8	505
4.2	9.1×10^8	550
4.3	9.6×10^8	521
5.1	7.2×10^8	694
7.2	7.5×10^8	667
8.4	9.0×10^8	555
10.1	5.2×10^8	961

[a] Volume of the mixed colony glycerol $1/10^6$ dilution (described above) containing 500 colonies

5.4.2.2 <u>Plating low-density colonies</u>

Approximately 500 colonies from each mixed colony $1/10^6$ dilution were spread on an aged LB agar plate containing kanamycin and incubated for 12 hours at 37°C. Colony material was transferred from each plate onto a hybridisation membrane, which was processed immediately to prevent drying out. The colony plates were incubated for a further 2 to 3 hours at 37°C to allow the colonies to re-grow before being stored at 4°C.

5.4.2.3 <u>Low-density filter screening</u>

The low-density filters were screened by hybridisation with radiolabelled Siamang STSs as described above. SCHSY4.1, 4.3, 5.1, 7.2, 8.4 and 10.1 were screened with B1, and SCHSY4.2 was screened with B10.

After a 4.5-hour exposure, the autoradiograph showed a low background. There were numerous signals of similar intensities from filters 4.1 and 7.2. There were four high-intensity signals from 4.2 and only one low-intensity signal from 10.1. There were no signals from filters 4.3, 5.1 and 8.4. (Figure 5.6, see above)

5.4.2.4 <u>Verifying colonies by PCR</u>

Single positive colonies (SCHSY4.11, 4.21, 7.21 and 10.11) were picked from the plates representing filters SCHSY4.1, 4.2, 7.2 and 10.1 and streaked to single colonies on fresh plates. One well-separated colony from each streak was picked and analysed for STS content by colony PCR using fF45C1 B1 and B10 primer pairs. SCHSY4.11, 7.21 and 10.11 contained the B1 STS but not the B10 STS. SCHSY4.21 contained the B10 STS, but not the B1 STS.

5.5   Construction of a cosmid map defining the homologous block rearrangement point

PCR analysis was carried out using fF45C1 primer pairs A2 to B38 to determine the STS marker content of each isolated cosmid and hence to identify overlaps between the clones. Clones SCHSY4.11, 7.21 and 10.11 contained STSs A2 to B1. Clone SCHSY4.21 contained STSs B2 to B20, inclusive, plus B25, B27 and B32. The four clones were organised into a map based on their regions of homology with human fF45C1 (summarised in Figure 5.7). From the results of the STS-PCR analysis of the four clones generated to this point it was possible to exclude the regions A2-B1 and B2-B32 from containing sequences spanning the breakpoint. The region B32-B38 had been previously excluded from containing breakpoint sequences based on the long-range PCR assays reported in section *5.2.3*.

Figure 5.7 (next page) STS map based on regions of homology of four gibbon cosmid clones with human chromosome 22 clone fF45C1. SCHSY4.11, 7.21 and 10.1 overlapped each other in the map as they all contained sequence homologous to the region in fF45C1 from STS A2 to B1. The region highlighted in red represents a section in fF45C1 not represented in the clones analysed.

Positive clones analysed for STS marker content and contig constructed

5.6   FISH analysis of breakpoint clones

Based on the results described above and the original FISH results of fF45C1, it was reasoned that DNA from the three clones SCHSY4.11, 7.21 and 10.11 would hybridise by FISH to Siamang 18p, and that DNA from SCHSY4.21 would hybridise to Siamang 18q.

DNA from the cosmids SCHSY4.11, 4.21, 7.21 and 10.11 was isolated and biotinylated by nick translation and hybridised to Siamang and human metaphase chromosomes in separate experiments. At least ten metaphase spreads were analysed for each hybridisation experiment. The DNA from SCHSY4.11, 7.21 and 10.11 hybridised to Siamang 18p. The DNA from SCHSY7.21 and 10.11 also hybridised to human 22q13.1. The DNA from SCHSY4.11 hybridised to human 22q13.1 and human 16p. These results are summarised in Table 5.5.

The DNA from SCHSY4.21 hybridised to Siamang 18q, human 22q13.1 and human 16p (see Table 5.5).

Table 5.5 Summary of FISH localisation of gibbon cosmids after hybridisation to gibbon and human metaphase chromosomes. A tick in a box indicates a positive hybridisation signal.

SCHSY cosmid	HSY 18p	HSY 18q	HSA 22q13.1	HSA 16p
4.11	3		3	3
7.21	3		3	
10.11	3		3	
4.21		3	3	3

These FISH results confirm that cosmid clones SCHSY4.11 and 4.21 contain gibbon sequences homologous to human 22q13.1 (4.21 proximal and 4.11 distal to the homologous block breakpoint) as well as containing sequences homologous to human 16p. The most likely explanation for these observations is that SCHSY4.11 contains the gibbon sequence, which spans the junction between the human chromosome 22- and 16-homologous blocks on Siamang 18p and SCHSY4.21 spans the junction on Siamang 18q.

The lack of signal on human chromosome 16p from clones SCHSY7.21 and 10.11 does not mean that they do not contain sequences that span the junction and don't contain sequences homologous to human 16p. They may well contain those sequences, but the stretch of DNA homologous to human chromosome 16p may be too short to generate a visible FISH signal.

Gibbon cosmid SCHSY4.11 contained STS markers A2 to B1 and it was FISH-mapped to HSY 18p, HSA 22q13.1 and HSA 16p. Cosmid SCHSY4.21 contained STS markers B2 to B20, B25, B27 and B32 and was FISH mapped to HSY 18q, HSA 22q13.1 and HSA 16p. In view of all the evidence, it was concluded that the evolutionary rearrangement breakpoint lies in material

homologous to human chromosome 22q13.1 within the 1 kb region defined by markers B1 and B2.

5.7 Discussion

This chapter describes the cloning and mapping of two homologous block evolutionary rearrangement junctions in Siamang chromosome 18.This involved the design of PCR primer pairs defining STSs in the human chromosome 22 fosmid fF45C1. Homologous STSs were mapped on Siamang chromosome 18 and five stretches of Siamang sequence failed to amplify by long-range PCR.

A Siamang genomic cosmid library of 6.73 genome equivalents was constructed from high molecular weight DNA. Part of the library was screened with radiolabelled Siamang STSs generated using fF45C1 PCR primers. The isolated clones were screened for STS marker content and organised into a map indicating homology with fF5C1. Two of the clones SCHSY4.11 and SCHSY4.21 were shown by FISH to contain sequences homologous to human 22q and human 16p. SCHSY4.11 was from Siamang chromosome 18p and 4.21 was from chromosome 18q. The evidence of the FISH results coupled with the STS content of each clone defined the human chromosome 22q breakpoint position to within 1 kb between STS markers B1 and B2.

Because the majority of the fosmid clone fF45C1 had been sequenced at the time of this work, STSs could be defined and primers designed to assay for them. In order to test the primers for their specificity, hamster genomic DNA was included as a control because the human chromosome 22-hybrid DNA was isolated from a human x hamster hybrid cell line. If the primers generated product from human chromosome 22-hybrid DNA, but not from hamster genomic DNA, it was assumed that the primers were amplifying human chromosome 22 specific sequences. The fact that primers for A1, B16, B22 and B36 generated more than one product from human

191

genomic DNA was possibly due to the repetitive nature of the sequence in that region of the clone.

When assayed with Siamang DNA, 90% of the fF45C1 primer pairs generated PCR products. Those results implied that the particular region of Siamang chromosome 18 being investigated shares considerable sequence homology with human fosmid fF45C1. Not only is the homology close enough to allow the hybridisation by FISH of fF45C1, but also to allow the amplification of Siamang STSs using fF45C1 PCR primers.

Although 10% of the primer pairs failed to generate specific product from the gibbon, long range PCR was carried out to amplify overlapping sequences from Siamang DNA. The assumption that priming would be successful in the gibbon except when the primers lay either side of the breakpoint was a big one, especially as some of the primers failed under standard PCR conditions. Nonetheless from the results of the long-range PCR it was possible to exclude 24 kb of fF45C1 sequence as the location of the breakpoint.

Because the long-range PCR approach did not provide a definitive answer in locating the position of the breakpoint and limited the exclusion to only 24 kb of sequence, it was necessary to apply an alternative approach to identify and analyse the breakpoint. Therefore, a genomic gibbon library was constructed and screened for clones containing sequences spanning the fusion points.

After analysing the high molecular weight Siamang genomic DNA using *Mbo* I and *dam* Methylase, it seemed to behave in a similar way to human DNA, and a competitive restriction digestion was set up to cut the DNA prior to cloning. The Siamang DNA cut much more readily than had been anticipated. It was only possible to generate the correctly sized fragments by titrating the *Mbo* I in a limiting enzyme restriction digestion. After screening the low-density library

filters by hybridisation the colonies selected from the regions generating large positive signals were verified by STS-PCR

There are two possible explanations for the results observed following STS-PCR analysis of the gibbon cosmids. One explanation could be that each of the four cosmids had their ends lying in the 1 kb region between B1 and B2, although this seems unlikely. The other explanation could be that the sequences spanning the rearrangement junctions lie in the 1 kb region between B1 and B2, which could be why none of the four clones contain both STSs B1 with B2.

From the STS maps, it could be seen that clones SCHSY4.11, 7.21 and 10.11 contained at least 5.4 kb of DNA homologous to human chromosome 22q13.1. Clone SCHSY4.21 contained at least 31 kb of DNA homologous to human chromosome 22q13.1 proximal to the breakpoint. The fact that, after FISH, clone SCHSY4.11 hybridised to Siamang 18p, human chromosome 22q13.1 and 16p confirmed that it contained sequences spanning the homologous block fusion point. Similarly, SCHSY4.21 hybridised to Siamang 18q, human chromosome 22q13.1 and 16p, confirming that it also contained sequences spanning the homologous block fusion point.

Having identified the two cosmid clones containing sequences spanning the rearrangement junctions in Siamang chromosome 18, it was necessary to study them at the sequence level in order to analyse the underlying structure and mechanism, which might have lead to the evolutionary chromosome rearrangement. The work towards that analysis is described in Chapter 6.

CHAPTER 6

Sequence Analysis of the Evolutionary Rearrangement Points in the Gibbon

6.1   Introduction

6.2   Isolating gibbon breakpoint fragments and sequence analysis

6.2.1 Restriction digestion of breakpoint clones and ligation to vectorette bubbles

6.2.2 Vectorette PCR

6.2.3 Sequence analysis of vectorette PCR products

6.3   Sequence analysis of SCHSY4.11 and SCHSY4.21 cosmids

6.3.1 Shotgun sequencing

6.3.2 Gibbonace

6.3.3 Gibbon primer design and verification

6.4   Discussion

6.1   Introduction

The work described so far in this thesis has involved the mapping of evolutionary rearrangement points in gibbon chromosome 18 at increasing levels of resolution. The boundaries of the human chromosome 22-homologous blocks were defined in gibbon chromosome 18 by cross-species chromosome painting to a resolution of approximately 7 Mb. The identification of a human chromosome 22 fosmid clone hybridising to both arms of gibbon chromosome 18 increased the resolution of the analysis of one of the rearrangement points to within 40 kb of sequence. The rearrangement point was further narrowed down to within 1 kb of sequence by STS PCR carried out on gibbon cosmid clones spanning the fusion points between human chromosome 22- and human chromosome 16-homologous blocks on gibbon 18 p and q.

In order to analyse the fusion regions at the highest resolution possible, it was necessary to generate sequence from the gibbon cosmid clones isolated in chapter 5. The initial strategy used was to amplify fragments containing gibbon sequences spanning the fusion points by vectorette PCR. These products would then be used as sequencing templates. This approach was taken because of the advantages of the vectorette system in terms of speed and resources: it was an economical and rapid method to confirm that the cosmids do contain sequences spanning the fusion points, perhaps before embarking on the more costly and time-consuming sequencing of two entire cosmids. Vectorette PCR was originally developed by Riley et al., (1990) for the rescue of the ends of YAC, and it is an efficient method for isolating unknown DNA adjacent to any known sequence of 20 bp or more in length. However, there are potential drawbacks to using this method as it generates short sequences and, thus, provides only a limited analysis of the region of interest. For that reason, the sequencing of the entire cosmids was planned as an alternative strategy.

RESULTS

6.2   Isolating gibbon fusion point fragments for sequence analysis

6.2.1 <u>Breakpoint fragment isolation by vectorette PCR</u>

According to STS mapping (Chapter 5), the sequence on human chromosome 22 corresponding to the gibbon ancestral chromosome 18 rearrangement breakpoint could be narrowed down to a 1 kb stretch of DNA between STSs B1 and B2. In order to provide additional information about the sequences spanning the fusion points on gibbon chromosome 18, fragments of SCHSY4.11 and SCHSY4.21 were isolated using a vectorette system modified to use human primers to carry out PCR on gibbon cosmids. For SCHSY4.11 the B1 sense primer was used with the vectorette primer 224. For SCHSY4.21 the B2 antisense was used with 224. A single product of approximately 400 bp was generated from SCHSY4.11 digested with *Rsa*I. For SCHSY4.21 single products were generated following digestion with *Pst*I*, Rsa*I and *Hinc*II of sizes 1.0 kb, 550 bp and 900 bp, respectively (figure 6.1). The identity of the four PCR products was confirmed by hemi-nested PCR, which also served to generate large amounts of DNA for sequencing. For the SCHSY4.11 *Rsa*I B1 + 224 product a nested primer "B1 nest" was designed (Sarah Hunt) from the human sequence distal to B1. For the three SCHSY4.21 B2 + 224 products, a nested primer "B2 nest" was designed proximal to B2. Primer 224 was used in combination with B1 nest or B2 nest for these secondary PCRs (see figure 6.1).

The vectorette PCR products from SCHSY4.11 *Rsa*I B1 + 224 (approximately 400 bp) and SCHSY4.21 *Pst*I B2 + 224 (approximately 1.2 kb) were electrophoresed and visualised on a preparative agarose gel. The appropriate bands were excised and the DNA extracted from the gel using the Geneclean ™ kit.

Figure 6.1 (next page) Agarose gel analysis of vectorette PCR products. M= 1 kb DNA size marker, sizes as indicated. Lanes as follows:

1. SCHSY4.11 *Rsa*I product from B1sense + 224

2. SCHSY4.11 *Rsa*I product from B1nest + 224

3. SCHSY4.11 *Rsa*I product from B1sense + B1antisense

4. SCHSY4.11 *Rsa*I product from B1nest + B1antisense

5. Gibbon genomic DNA control product from B1sense + B1antisense

6. Negative control

7. SCHSY4.21 *Pst*I product from B2antisense +224

8. SCHSY4.21 *Pst*I product from B2nest + 224

9. SCHSY4.21 *Pst*I product from B2antisense + B2sense

10. SCHSY4.21 *Pst*I product from B2nest + B2sense

11. SCHSY4.21 *Rsa*I product from B2antisense + 224

12. SCHSY4.21 *Rsa*I product from B2nest + 224

13. SCHSY4.21 *Rsa*I product from B2antisense + B2sense

14. SCHSY4.21 *Rsa*I product from B2nest + B2sense

15. SCHSY4.21 *Hinc*II product from B2antisense + 224

16. SCHSY4.21 *Hinc*II product from B2nest + 224

17. SCHSY4.21 *Hinc*II product from B2antisense + B2sense

18. SCHSY4.21 *Hinc*II product from B2nest + B2sense

6.2.2 <u>Sequence analysis of vectorette PCR fragments</u>

The vectorette-PCR fragments were sequenced by Elizabeth Huckle of the Sanger Institute Sequencing Development Team. Each product was sequenced from both directions using primers 224 and B1 nest for SCHSY4.11 or 224 and B2 nest for SCHSY4.21.

The SCHSY4.21 reaction primed by B2 nest yielded a stretch of 508 bases of sequence homologous to human chromosome 22 distal to the B2 STS. The SCHSY4.21 reaction primed by 224 yielded a stretch of 460 bp of non-human homology (closest to the priming site) followed by a 132 bp stretch of sequence homologous to human chromosome 22.

The SCHSY4.11 sequencing reaction primed by B1 nest yielded a stretch of 243 bp of sequence homologous to human chromosome 22 proximal to the B1 STS. The SCHSY4.11 reaction primed by 224 yielded a stretch of 111 bases of sequence homologous to human chromosome 22 proximal to the B1-primed region. Preceding this sequence, there were a further 15 bases of sequence which did not match any known human sequence by BLAST.

It was possible that the un-matched sequences from both vectorette products were actually homologous to unsequenced regions of human chromosome 16 (or another human chromosome), or that they represented chromosomal material present in the great ape ancestral genome, which was lost since divergence of the lesser apes occurred. However, at this stage the possibility could not be excluded that there are gibbon-specific sequences flanking chromosome 22 homologous material on both sides of the fusion points. The gibbon sequence may have been inserted at the rearrangement points at the time of or after the rearrangement event.

The last seven bases before the homology with human chromosome 22 is lost are common to SCHSY4.21 and SCHSY4.11. The region of commonality and loss of homology was in an Alu repeat.

From the sequence generated from the vectorette PCR products, it was difficult to speculate about the rearrangement mechanism. As the PCR products generated such short sequencing reads, limitations of the vectorette PCR approach have been demonstrated. For example, if the rearrangement were sponsored by complex sequence motifs farther than a few hundred bp away from the junction points, it would be impossible to explore the mechanism.

In order to describe the rearrangement points in detail (and generate further evidence for a possible mechanism) a full sequence analysis of clones SCHSY4.11 and SCHSY4.21 was carried out.

6.3   Sequence analysis of SCHSY4.11 and SCHSY4.21 cosmids

DNA from the two cosmids was prepared using a standard alkaline lysis procedure and was submitted to the Sanger Centre sequencing pipeline (overseen by Matt Jones, David Willey, and Kirsten McClay). Highly accurate, finished sequence was generated for each clone, and the sequences were submitted to the EMBL database. The sequence data were analysed for repeat elements (RepeatMasker), DNA homologies (BLASTN) and gene predictions (FgenesH) by Sarah Hunt (Sanger Institute Informatics Group). Along with other annotation information, the results of the analysis were stored in Gibbonace (created by Carol Scott and Sarah Hunt), which is an implementation of ACeDB.

The DNA insert of clone SCHSY4.11 is 40,516 bp long and the insert of clone SCHSY4.21 is 34,056 bp long. The sequences from both junction point cosmids support the lower-resolution analyses, described previously, which placed the HSA 16 and HSA 22 homologous blocks adjacent to each other on either side of the rearrangement junctions on HSY 18p and 18q. The gross features of each clone are summarised in Figure 6.2 and 6.3.

Figure 6.2 and 6.3 (next two pages) Screen capture from Gibbonace illustrating major sequence landmarks of clones SCHSY4.11 and 4.21.

6.2.2 <u>Sequence analysis of vectorette PCR fragments</u>

The vectorette-PCR fragments were sequenced by Elizabeth Huckle of the Sanger Institute Sequencing Development Team. Each product was sequenced from both directions using primers 224 and B1 nest for SCHSY4.11 or 224 and B2 nest for SCHSY4.21.

The SCHSY4.21 reaction primed by B2 nest yielded a stretch of 508 bases of sequence homologous to human chromosome 22 distal to the B2 STS. The SCHSY4.21 reaction primed by 224 yielded a stretch of 460 bp of non-human homology (closest to the priming site) followed by a 132 bp stretch of sequence homologous to human chromosome 22.

The SCHSY4.11 sequencing reaction primed by B1 nest yielded a stretch of 243 bp of sequence homologous to human chromosome 22 proximal to the B1 STS. The SCHSY4.11 reaction primed by 224 yielded a stretch of 111 bases of sequence homologous to human chromosome 22 proximal to the B1-primed region. Preceding this sequence, there were a further 15 bases of sequence which did not match any known human sequence by BLAST.

It was possible that the un-matched sequences from both vectorette products were actually homologous to unsequenced regions of human chromosome 16 (or another human chromosome), or that they represented chromosomal material present in the great ape ancestral genome, which was lost since divergence of the lesser apes occurred. However, at this stage the possibility could not be excluded that there are gibbon-specific sequences flanking chromosome 22 homologous material on both sides of the fusion points. The gibbon sequence may have been inserted at the rearrangement points at the time of or after the rearrangement event.

The last seven bases before the homology with human chromosome 22 is lost are common to SCHSY4.21 and SCHSY4.11. The region of commonality and loss of homology was in an Alu repeat.

From the sequence generated from the vectorette PCR products, it was difficult to speculate about the rearrangement mechanism. As the PCR products generated such short sequencing reads, limitations of the vectorette PCR approach have been demonstrated. For example, if the rearrangement were sponsored by complex sequence motifs farther than a few hundred bp away from the junction points, it would be impossible to explore the mechanism.
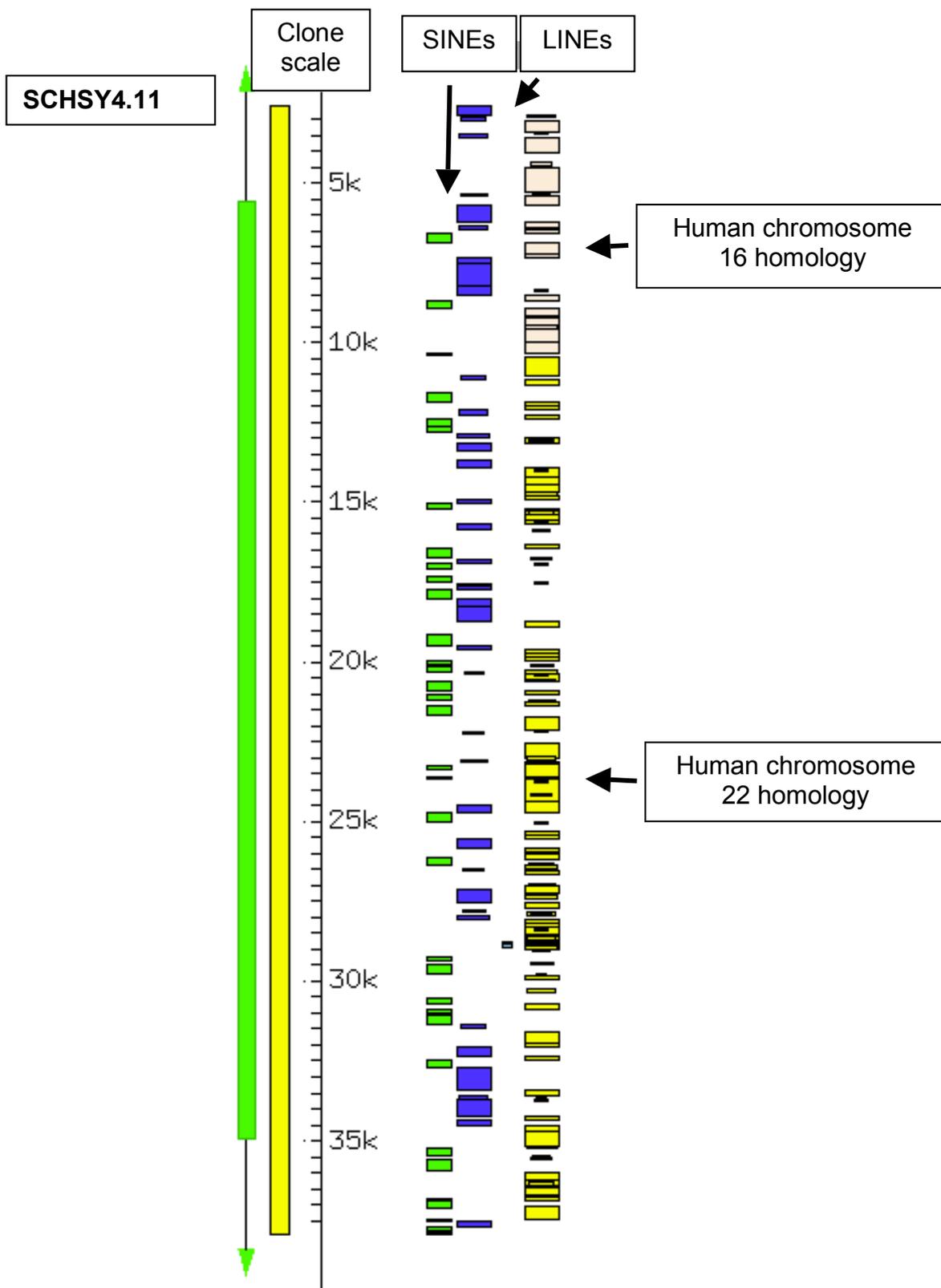
In order to describe the rearrangement points in detail (and generate further evidence for a possible mechanism) a full sequence analysis of clones SCHSY4.11 and SCHSY4.21 was carried out.

6.3   Sequence analysis of SCHSY4.11 and SCHSY4.21 cosmids

DNA from the two cosmids was prepared using a standard alkaline lysis procedure and was submitted to the Sanger Centre sequencing pipeline (overseen by Matt Jones, David Willey, and Kirsten McClay). Highly accurate, finished sequence was generated for each clone, and the sequences were submitted to the EMBL database. The sequence data were analysed for repeat elements (RepeatMasker), DNA homologies (BLASTN) and gene predictions (FgenesH) by Sarah Hunt (Sanger Institute Informatics Group). Along with other annotation information, the results of the analysis were stored in Gibbonace (created by Carol Scott and Sarah Hunt), which is an implementation of ACeDB.

The DNA insert of clone SCHSY4.11 is 40,516 bp long and the insert of clone SCHSY4.21 is 34,056 bp long. The sequences from both junction point cosmids support the lower-resolution analyses, described previously, which placed the HSA 16 and HSA 22 homologous blocks adjacent to each other on either side of the rearrangement junctions on HSY 18p and 18q. The gross features of each clone are summarised in Figure 6.2 and 6.3.

Figure 6.2 and 6.3 (next two pages) Screen capture from Gibbonace illustrating major sequence landmarks of clones SCHSY4.11 and 4.21.

SCHSY4.11

Clone scale

SINEs

LINEs

Human chromosome 16 homology

Human chromosome 22 homology

SCHSY4.21

Clone scale

SINEs

LINEs

SCHSYG4_21

Human chromosome 16 homology

Human chromosome 22 homology

5k

10k

15k

20k

25k

30k

SCHSY4.11 has 10.3 kb of homology to human chromosome 16 and the homology immediately switches to human chromosome 22 to the end of the clone (25 kb). The homology switches at the site of a partial AluJo element, which originates from the human chromosome 22 homologous material. The chromosome 16 homology has two locations on human chromosome 16 separated by a distance of 5 Mb.

SCHSY4.21 has 7 kb of homology to human chromosome 16, and 25 kb of homology to human chromosome 22 to the end of the clone. The homology to human chromosome 22 starts at the site of a partial AluJo element (bases 21 to 148), which originates from the human chromosome 22 homologous material. There is a stretch of 5 kb of sequence between the chromosome 16 and 22 homologies, which has no human homology. An AluJo element is located at the end of the main part of HSA16 homology, but there is a short section (250 bp) of inverted duplicated HSA16 homologous material after the AluJo.

Dotter analyses were carried out to illustrate graphically the homology between the HSA22 breakpoint region (in human chromosome clone HSE81G9), the HSA16 breakpoint region (in sequence AC126763) and the sequences of clones SCHSY4.11 and 4.21.

As can be seen in Figure 6.4, there is almost continuous homology between HSE81G9 (from 8 kb to the end) and SCHSY4.11 (from 10 kb to the end). There is also almost continuous homology over 9 kb between the HSA22-homologous section of SCHSY4.21 and HSE81G9 (Figure 6.5). There is one region of 200 bp, which is found in the gibbon sequence and not in the human, illustrated by a gap in the diagonal line.

The Dotter output from the comparison of 0-15 kb of AC126763 versus SCHSY4.11 shows continuous homology up to the AluJo element at the breakpoint (Figure 6.6). The analysis from the comparison of AC126763 versus SCHSY4.21shows continuous homology up to the breakpoint (Figure 6.7). After the AluJo in 4.21, there is a small region of inverted homology to an earlier section (2.4-2.6 kb) of AC126763. Other than repeat elements, there is no homology between AC126763 and HSE81G9, or between 4_21 and 4_11.

207

*Figures 6.4, 6.5, 6.6 and 6.*7 (on the following four pages) illustrate the Dotter analysis outputs from sequence comparisons between HSE81G9 versus SCHSY4.11, HSE81G9 versus SCHSY4.21, AC126763 versus SCHSY4.11 and AC126763 versus SCHSY4.21, respectively.

HSE81G9 (horizontal) vs. SCHSYG4_11 (vertical)

HSE81G9 (horizontal) vs. SCHSYG4_21 (vertical)

AC126763 (horizontal) vs. SCHSYG4_11 (vertical)

AC126763 (horizontal) vs. SCHSYG4_21 (vertical)

8449, 5249
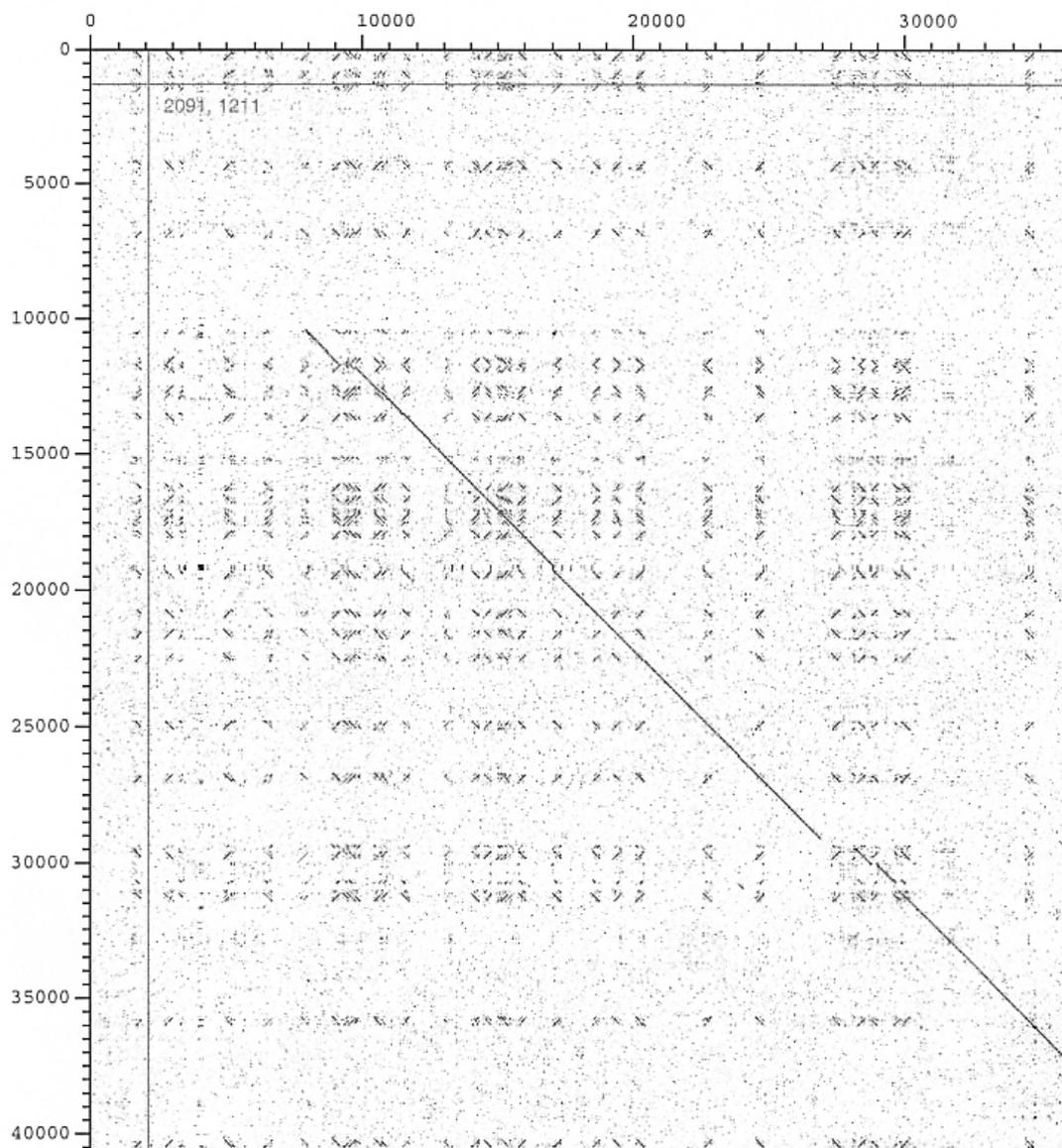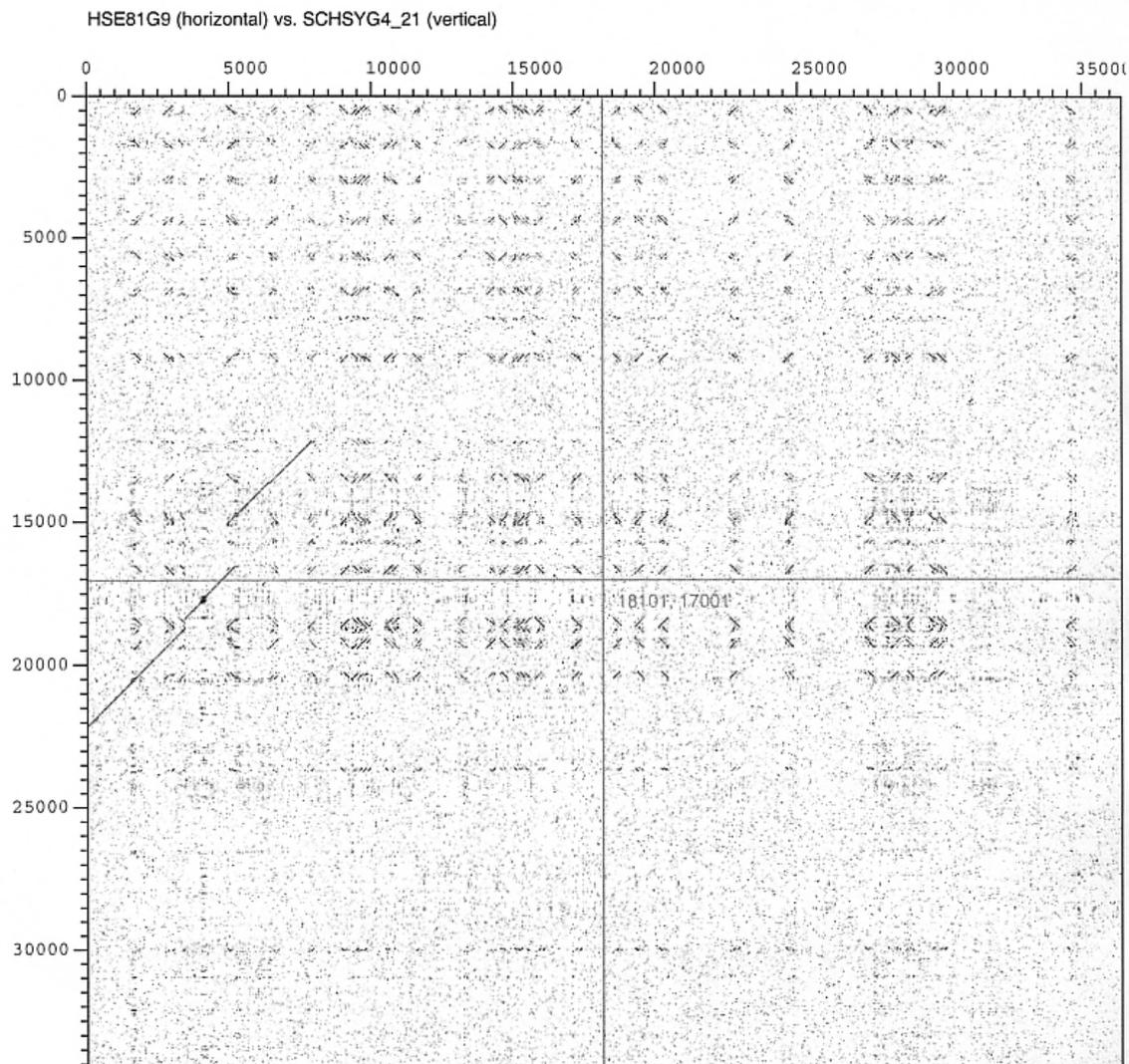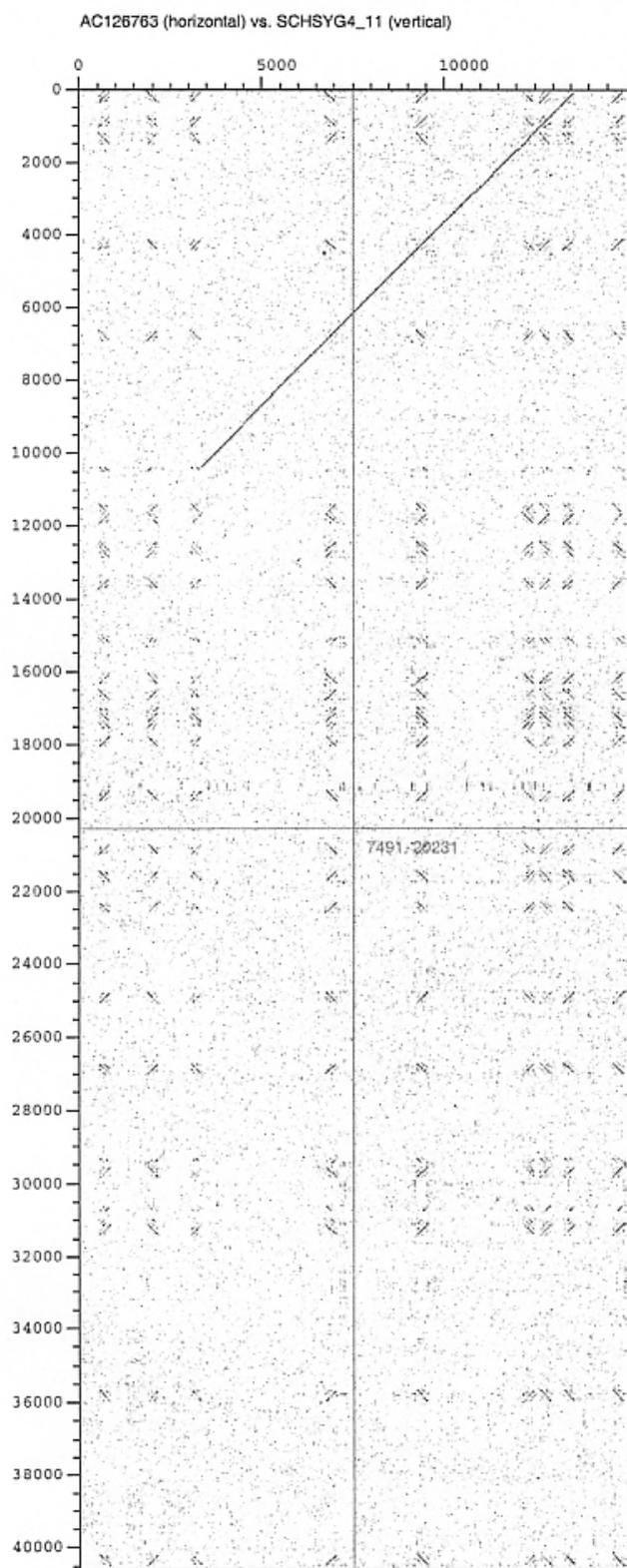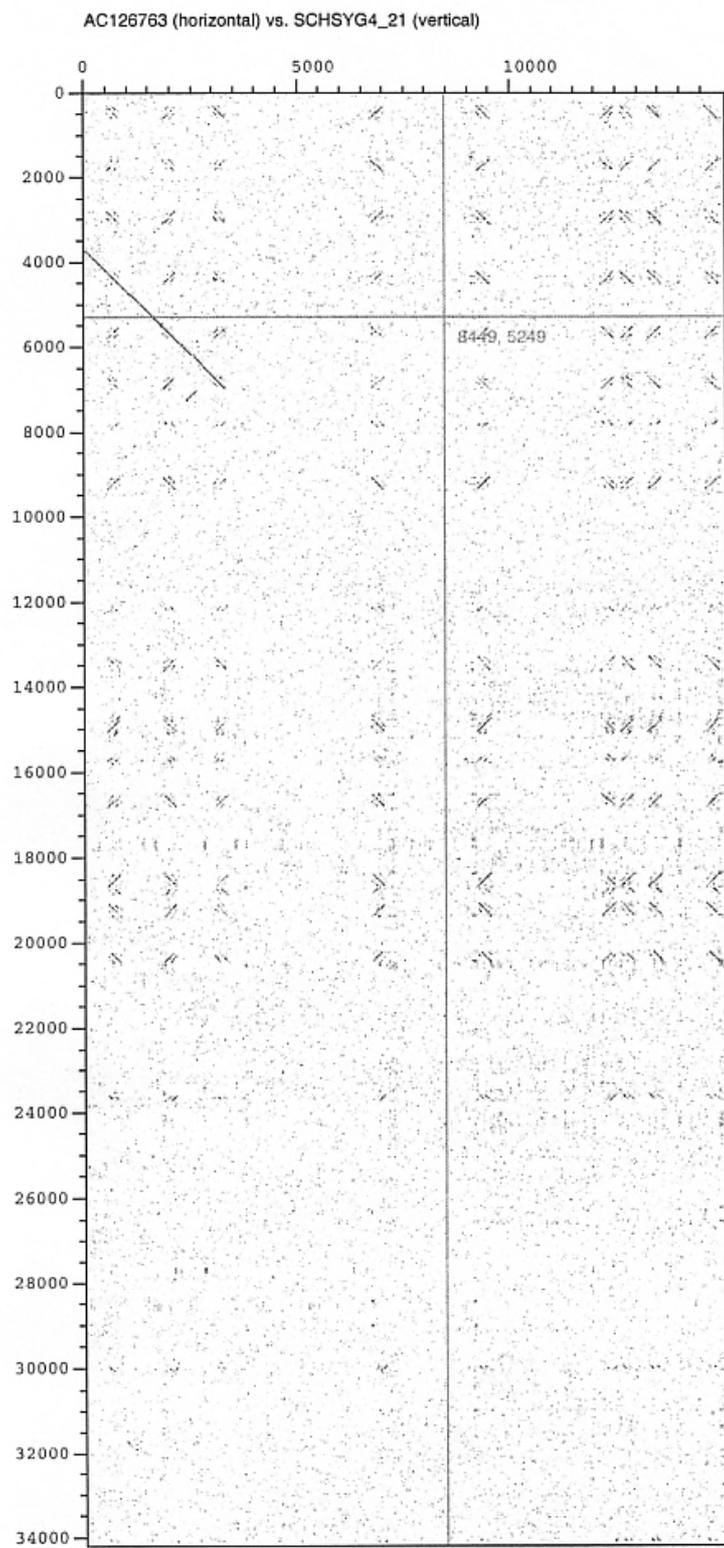
The partial AluJo elements cloned in 4.11 and 4.21 were blasted against non-redundant human sequence and both were found to be homologous to HSA22. The AluJo element adjacent to the inverted duplicated HSA16 material in 4.21 is homologous to HSA16. The repeatmasker co-ordinates of the partial AluJo elements appear to follow on from each other, that is, the two partial Alu elements represent the same region of an AluJo on the human chromosome 22 sequence. This suggests that the two halves of the AluJo elements (one on 4.11, the other in 4.21) have been derived from a breakage and inversion in human chromosome 22 homologous material.

As a consequence of where the breakage occurred relative to the SINEs, one of the products (4_11) has only 0.5 AluJo and the other (4_21) has 1.5 AluJo elements. This is consistent with an homologous recombination event in which the Alu elements had aligned out of register by one sub-repeat unit. In this situation, each of the elements would be partly derived from HSA22 and partly from HSA16 homologous material. If this model were correct, the unknown 5 kb segment would perhaps have inserted and separated the 1.5 AluJo element into the 1 and the 0.5 seen in 4_21 either side of the 5 kb block.

The model proposed for the rearrangement is of a pericentric inversion of the ancestral chromosome, on which the regions homologous to HSA22 and HSA16 were originally on separate arms. The breakpoint in the HSA22 material was within an AluJo element, and the break in the HSA16 material occurred just after an AluJo element. The incorporation of 5 kb of mostly repetitive sequence and the small duplication of material homologous to HSA16 on one of the inversion products is most likely to have occurred during the rearrangement event. The reasoning for this is that if the section were inserted after the rearrangement event, then the insertion would have had to occur precisely at the point of fusion. Another possibility is that the 5 kb of sequence was in fact part of the ancestral chromosome homologous to human 16, that was lost after the lesser apes diverged. This also seems unlikely because one end of the deletion in the human lineage would have to coincide precisely with the inversion point on the gibbon chromosome 16-homologous material. Because of the above reasons, it

seems likely that the rearrangement occurred during a non-homologous end joining event, rather than an homologous recombination.

6.4   Discussion

For isolating the gibbon rearrangement point junctions, the vectorette PCR approach was rapid and successful, and has not been previously reported for use in cloning evolutionary rearrangement breakpoints. However, the sequence reads generated from the PCR products were short and, although it was possible to identify the homologous block junction points, it was impossible to interpret the genomic environment around the rearrangement event. In particular, the insertion of approximately 5 kb of DNA between the chromosome 22 and chromosome 16 homologous regions in one of the clones would have precluded the region's clear description using the vectorette approach. Therefore, a full sequence analysis was carried out on the breakpoint clones.

From the full sequence analysis of the gibbon rearrangement point clones and the HSA22 and HSA16 sequences homologous to the regions, which were broken in the gibbon lineage, it is not at this time possible to identify a specific rearrangement mechanism, although the evidence points to a non-homologous end-joining (NHEJ) event. During double strand break (DSB) repair via homologous recombination, the broken DNA sequence interacts with a homologous donor sequence and genetic information is exchanged between identical or nearly identical DNA sequences. NHEJ is accomplished by the joining of DNA ends without interaction between the broken molecule and a donor sequence. In NHEJ there is no requirement for homology at the DNA termini being joined, although NHEJ may be facilitated by short terminal homologies (Lin and Waldman, 2001). NHEJ may be accompanied by the deletion or gain of genetic material (for example, retrotransposon sequences) prior to healing of the DSB.

To take this work further, a detailed analysis of the relative orientation and degree of sequence similarity of the Alu elements flanking the rearrangement junction in 4_21 could be

carried out by screening the sequences in RepeatMasker. This might indicate whether, due to the orientation of the elements, there was any possibility that these regions could have been involved in stabilising a homologous recombination event.

CHAPTER 7

Discussion

7.1 Overview

The aim of the work described in this thesis was to describe two mammalian evolutionary chromosome rearrangements. In the course of the work, it was possible to start with a low-resolution analysis and, by a series of steps, increase the resolution of the study to narrow down the rearrangement breakpoint locations. Ultimately, one of the rearrangement events was studied at the highest possible resolution, that is, at the sequence level. Initially, a cytogenetic analysis of canine and Siamang gibbon metaphase chromosomes was carried out using chromosome paints, and then the resolution was increased by the use of bacterial clones. After the construction and screening of a gibbon genomic library, sequence was generated from the regions spanning the evolutionary rearrangement junctions between HSA22- and HSA16-homologous blocks on gibbon chromosome 18p and q. The strategy to proceed step-wise from low-resolution cytogenetic mapping to the cloning and sequence analysis of a breakpoint was successful because of the resources already available from the human genome sequencing project, as well as the resources generated for this thesis using established technologies at the Sanger Institute.

Because the technique of chromosome flow sorting was already established at the Sanger Institute during the work for this thesis, it was possible to take the standard protocols for chromosome isolation, analysis and sorting and to develop them for their application to the dog and the gibbon. Thus, the canine and Siamang flow karyotypes were established and chromosomes sorted for the generation of paints for the low-resolution reciprocal zoo-FISH studies described in chapter 3. The panel of dog paints (Langford *et al.,* 1996*)* were made available to the canine karyotype research community, which also benefited from the proposed standard DAPI-banded karyotype (Breen, *et al.,* 1999a). The generation of flow karyotypes, the flow sorting of chromosomes and the production of chromosome-specific paints is now a routine procedure for laboratories focussing on the identification of ECCSs between species on a global scale. Although it is less widely used, access to micro-dissection equipment enables investigators to generate

paint probes for sub-regions of chromosomes to increase the power of reciprocal zoo-FISH analysis.

The current availability of an overlapping tile-path of BAC (and other) clones for the whole human genome is a powerful resource for high-resolution cross species FISH studies not only to identify the boundaries of conserved synteny between the karyotypes of other mammals, but also to generate information about the retention or loss of sequence orientation within ECCSs. The density of the human clone map should enable the approach described in chapter 4 to be applied to the analysis of other mammalian species' chromosomes. Thus by successive rounds of cross-species FISH, the location of other rearrangement breakpoints could be narrowed down until clones containing sequences homologous to regions spanning breakpoints are identified. For cross-species FISH to be successful, the genome sequence of the mammal being analysed should not be so diverged that BAC clones from human are prevented from forming specific duplexes during hybridisation. In this thesis, the approach has been applied to the dog, which represents a period of 70 million years of divergent evolution.

If this work were to be repeated today, one approach could be to utilise DNA microarray technology. Genomic arrays with probes representing overlapping human tile-path clones could be interrogated with genomic samples from individual chromosomes of the species under analysis. If successful, this approach would have the potential to localise multiple ECCS boundaries in a single hybridisation experiment. This approach would require access to the techniques of chromosome flow sorting or micro-dissection in order to generate the material for hybridisation. With the advancement of microarray technology, it may soon be possible to resolve junctions between ECCSs to within 1 kb of human sequence in a single experiment. However, as with zoo-FISH, without chromosome micro-dissection this approach would not be so useful for the analysis of ECCSs (such as HSA 22q homology in HSY 18) where the entire chromosome under analysis is present in one or more blocks on a single chromosome of the animal for comparison.

The generation of sequence information for the work described in this thesis was approached in three ways: using STS PCR, vectorette PCR and shotgun sequencing of the cosmids. As the HSA 22q clone fF45C1sequence was available, it was possible to design primers for the STS PCR analysis of gibbon genomic DNA. The principle of this approach was to attempt to identify a pair of primers that would fail to amplify a PCR product from the gibbon genome. Unfortunately, the results of this approach were inconclusive. Once the gibbon junction point clones had been isolated and analysed for STS content, the location of the breakpoint in the HSA22-homologous material was identified within 1 kb of sequence between the markers B1 and B2. With hindsight, it was possible to reassess the long-range PCR analysis of gibbon genomic DNA and confirm that one of the regions of failed amplification did in fact contain the breakpoint. However, without other supporting evidence, there was insufficient information generated by long-range PCR to deduce that the location of the breakpoint was definitely between B1 and B2 rather than between the other markers. For other studies, if the sequence of a "breakpoint" clone is known, STSs could similarly be identified and primers designed for PCR assays. But, as described in chapter 5, it seems unlikely that long-range PCR will be an efficient approach for further narrowing down the location of a breakpoint, particularly in a more distantly related mammalian genome. This is because the more distantly related a species is to human, the more sequence divergence is likely to have occurred.

Even if STS PCR worked well in a particular species, subsequent isolation of the DNA spanning rearrangement junctions, for example by vectorette PCR, may not generate fragments of sufficient length to enable a full sequence analysis of the rearrangement products. The use of vectorette PCR to isolate short sequences spanning the gibbon synteny block junctions (described in chapter 6) was a rapid way of generating sequence information. From this, it was possible to establish that the breakpoint in the HSA22-homologous material lay within an Alu element. However, if the vectorette fragments generated in this study had been the only source of sequence information for this thesis, it would not have been possible to establish any other information about the genome

architecture in the regions flanking the junction points nor to speculate on the underlying mechanism for the rearrangement.

Generating longer-range sequence data across genomic regions that represent the ancestral or the rearranged states provides the ultimate tool for attempting to understand the rearrangement mechanism. But despite the possibility of analysing the local sequence structures, motifs and homologies, it is still possible that, as in chapter 6, a definition of the precise mechanism is problematic. This is perhaps particularly likely when the mechanism is non-homologous end joining of DNA molecules. It is also difficult to speculate about the reason why the lesser apes have undergone such accelerated karyotype evolution compared to other old world primates and hominids. However, with the analysis of other gibbon rearrangement breakpoints and junctions, it might be possible to build up evidence supporting a specific mechanism or mechanisms. If the Siamang genome had been sequenced, the analysis of the junctions between blocks of synteny could have been carried out by comparative sequence analysis. However, there is no current proposal to sequence the Siamang genome, and so a targeted approach would still be required to analyse more rearrangement breakpoints.

In light of the findings of this thesis about the two possible mechanisms of the Siamang rearrangement, a review of some other breakpoints (evolutionary and pathological) involving Alu-mediated homologous recombination or non-homologous end joining should be considered. Homologous recombination involving Alu elements have been characterised in various diseases, such as haemophilia (Vidal, *et al.,* 2002) and glycogen storage disease (Huie, *et al.,* 1999). However, NHEJ is the main pathway for repairing double-stranded DNA breaks (Lieber, *et al.,* 2003).

7.2 Future work

To take this work further, it would be important to study other syntenic block junctions within the Siamang genome and the genomes of other lesser apes. The work could also be extended to the great apes as well as more distantly related new world monkeys. As more information is obtained, evidence may accrue for the specific mechanisms, which

have caused evolutionary rearrangements and shaped primate karyotypes. Furthermore, as more sequence-level studies are carried out in cases of human disease, it might be possible to assess whether there is any correlation between evolutionary and pathological rearrangement breakpoints in the human genome.

Chapter 8

References

Abdurashidova, G., Danailov, M. B., Ochem, A., Triolo, G., *et al.* (2003) Localization of proteins bound to a replication origin of human DNA along the cell cycle. *EMBO J* 22(16):4294-303

Albertson, D. G., Sherrington, P., and Vaudin, M. (1991) Mapping nonisotopically labeled DNA probes to human chromosome bands by confocal microscopy. *Genomics* 10: 143-150

Alexandrov, I., Kazakov, A., Tumeneva, I., Shepelev, V. and Yurov, Y. (2001) Alpha-satellite DNA of primates: old and new families *Chromosoma* 110(4):253-66.

Ambros, P., Matzke, M., and Matzke, A. (1986) Detection of a 17kb unique sequence (T-DNA) in plant chromosomes. *Chromosoma* 94: 11-18

Andersson, L., Archibald, A., Ashburner, M., Audun, S., *et al*. (1996) Comparative genome organization of vertebrates. The First International Workshop on Comparative Genome Organization. *Mamm. Genome* 7: 717-734

Arnason, U. (1977) The relationship between the four principal pinniped karyotypes. *Hereditas* 87(2):227-42

Arnold, N., Stanyon, R., Jauch, A., O'Brien, P., and Wienberg, J. (1996) Identification of complex chromosome rearrangements in the gibbon by fluorescent in situ hybridization (FISH) of a human chromosome 2q specific microlibrary, yeast artificial chromosomes, and reciprocal chromosome painting. *Cytogenet. Cell. Genet.* 74: 80-85

Baldini, A., Ried, T., Shridhar, V., Ogura, K., D'Aiuto, L., Rocchi, M., Ward, D. C. (1993) An alphoid DNA sequence conserved in all human and great ape chromosomes: evidence for ancient centromeric sequences at human chromosomal regions 2q21 and 9q13. *Hum Genet* 90: 577-583

Bates, P. F. and Swift, R. A. (1983) Double cos site vectors: simplified cosmid cloning. *Gene*. 26(2-3):137-46

Bigoni, F., Koehler, U., Stanyon, R., Ishida, T. and Wienberg, J. (1997) Fluorescence in situ hybridization establishes homology between human and silvered leaf monkey chromosomes, reveals reciprocal translocations between chromosomes

homologous to human Y/5, 1/9, and 6/16, and delineates an X1X2Y1Y2/X1X X2X2 sex-chromosome system. *Am. J. Phys. Anthropol.* 102: 315-327

Breen, M., Bullerdiek, J., Langford, C. F. (1999a) The DAPI banded karyotype of the domestic dog *Canis familiaris* generated using chromosome-specific paint probes, *Chr Res* 7:401-406

Breen, M., Thomas, R., Binns, M. M., Carter, N. P., Langford, C. F. (1999b) Reciprocal chromosome painting reveals detailed regions of conserved synteny between the karyotypes of the domestic dog (Canis familiaris) and human. *Genomics.* 61(2):145-55

Buckland, R. A. and Evans, H. J. (1978) Cytogenetic aspects of phylogeny in the Bovidae. I. G-banding. *Cytogenet Cell Genet* 21(1-2):42-63

Cannizzaro, L. A. (1996) Chromosome microdissection: a brief overview. *Cytogenet. Cell Genet.* 74: 157-160

Carter, N. P. (1994) Cytogenetic analysis by chromosome painting. *Cytometry* 18: 2-10

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. *Mamm. Genome* 7: 297-302

Chowdhary, B. P. (1998). Comparative genome analysis in farm animals using cross-species chromosome painting. *Ag. Biotech.*

Chowdhary, B. P., Raudsepp, T., Frönicke, L. and Scherthan, H. (1998) Emerging patterns of comparative genome organization in some mammalian species as revealed by Zoo-FISH. *Genome Research* 8: 577-589

Colbert & Morales (1991) Fossil evidence of mammalian divergence in *Evolution of the Vertebrates*. New York: Wiley-Liss

Collins, J. and Hohn, B. (1978) Cosmids: a type of plasmid gene-cloning vector that is packageable in vitro in bacteriophage lambda heads. *Proc Natl Acad Sci U S A.* Sep;75(9):4242-6.

Collins, C., Kuo, W.-L.., Segraves, R., Pinkel, D., Fuscoe, J. and Gray, J. W. (1991) Construction and characterization of plasmid libraries enriched in sequences from single human chromosomes. *Genomics* 11: 997-1004

Collins, J. *et al.* (1995) A high density YAC contig map of Human chromosome 22. *Nature* Supplement to Sep; 377: 367

Consigliere, S., Stanyon, R., Koehler, U., Agoramoorthy, G. and Wienberg, J. (1996) Chromosome painting defines genomic rearrangements between red howler monkey subspecies. *Chrom. Res.* 4: 264-270

Craig, J. M. and Bickmore, W. A. (1993) Chromosome bands--flavours to savour. *Bioessays* 15(5):349-54

Darwin, C. (1859) On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life. *London: John Murray, Albermarle Street*

Deloukas, P., McCarthy, L. C., Terrett, J., Davis, M. E., *et al.,* (1997) A first-generation whole genome-radiation hybrid map spanning the mouse genome. *Genome Res* 7: 1153-1161

Demattei MV, Auge-Gouillou C, Pollet N, Hamelin MH, Meunier-Rotival M, Bigot Y (2000) Features of the mammal mar1 transposons in the human, sheep, cow, and mouse genomes and implications for their evolution. *Mamm Genome* 11(12):1111-6

Dunham, I., Shimizu, N., Roe, B. A., Chissoe, S., Hunt, A. R., Collins, J. E., *et al.* (1999) The DNA sequence of human chromosome 22. *Nature* vol 402: 489-495

Dutrillaux, B. (1979) Chromosomal evolution in primates: tentative phylogeny from Microcebus murinus (Prosimian) to man. *Hum Genet* 48(3): 251-314

Eichler, E. E. (2001) Segmental duplications: what's missing, misassigned, and misassembled-and should we care? *Genome Res* 11: 653-656

Ferguson-Smith, M. A., Yang, F. and O'Brien, P. C. M. (1998) Comparative mapping using chromosome sorting and painting. *ILAR J.* 39: 68-76

Franke, U. (1981) High-resolution ideograms of trypsin-Giemsa banded human chromosomes. *Cytogenet Cell Genet* 31(1):24-32

Franke, U. (1994) Digitized and differentially shaded human chromosome ideograms for genomic applications. *Cytogenet Cell Genet* 65(3):206-18

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. *Mamm. Genome* 7: 285-290

Frönicke, L., Müller-Navia, J., Romanakis, K. and Scherthan, H. (1997) Zoo-FISH maps of the harbour seal (Phoca vitulina) and the putative ancestral carnivore karyotype. *Chromosoma* 106: 108-113

Frönicke, L. and Scherthan, H. (1997) Zoo-FISH analysis of human and Indian muntjac karyotypes (*Muntiacus muntjak vaginalis*) reveals satellite-DNA clusters at the margins of conserved syntenic segments. *Chromosome Res* 5: 254-261

Gallagher and Womack (1992) Chromosome conservation in the Bovidae. *J Hered* 83(4):287-98

Gallagher, D. S., Derr, J. N., Womack, J. E. (1994) Chromosome conservation among the advanced pecorans and determination of the primitive bovid karyotype. *J Hered* 85(3):204-10

Gardiner, K. (1995) Human genome organization. *Curr Opin Genet Dev* 5(3):315-22

Glas, R., Wakefield, M.J., Toder, R., Graves, J. A. (1998) Comparative Chromosome Painting (poster). *ILAR J* 39(2-3)

Gottgens, B., Gilbert, J. G., Barton, L. M., Grafham, D., *et al* (2001) Long-range comparison of human and mouse SCL loci: localized regions of sensitivity to restriction endonucleases correspond precisely with peaks of conserved noncoding sequences. *Genome Res* 11: 87-97

Goureau, A., Yerle, M., Schmitz, A., Riquet, J., Milan, D., *et al.* (1996) Human and porcine correspondence of chromosome segments using bidirectional chromosome painting. *Genomics* 36: 252-262

Gross, S. J. and Harris, H. (1975) New methods for mapping genes in human chromosomes. *Nature* 255: 680-684

Haaf, T. and Bray-Ward, P. (1996) Region-specific YAC banding and painting probes for comparative genome mapping: implications for the evolution of human chromosome 2. *Chromosoma* 104: 537-544

Haig, D. (1999) A brief history of human autosomes. *Philos Trans R Soc Lond B Biol Sci* 354: 1447-70

Haldane, J. B. S. (1927) The comparative genetics of colour in rodents and carnivora. *Biol. Rev. Camb. Philos. Soc.* 2: 199-212

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 Res* 5: 5-11

Hayes, H. (1995) Chromosome painting with human chromosome-specific DNA libraries reveals the extent and distribution of conserved segments in bovine chromosomes. *Cytogenet. Cell Genet* 71: 168-174

Henthorn, P. S., Somberg, R.L., Fimiani, V. M., Puck, J. M., *et al.* (1994) IL-2R gamma gene microdeletion demonstrates that canine X-linked severe combined immunodeficiency is a homologue of the human disease. *Genomics* 23(1):69-74

Hoheisel, J. D., Nizetic, D. and Lehrach, H. (1989) Control of partial digestion combining the enzymes dam methylase and MboI. *Nucleic Acids Res* 17(23):9571-82

Huie, M. L., Shanske, A. L. Kasper, J. S., Marion, R. W. and Hirschhorn. R. (1999) A large Alu-mediated deletion, identified by PCR, as the molecular basis for glycogen storage disease type II (GSDII). *Hum Genet.* 104(1):94-8

Hutt, F. B., Rickard, G. C. and Field, R. A. (1948) Sex-linked hemophilia in dogs. *J. Hered* 39: 2-9

Inoue, K., Dewar, K., Katsanis, *et al.* (2001) The 1.4 Mb CMT1A duplication/HNPP deletion genomic region reveals unique genome architectural features and provides insights into the recent evolution of new genes. *Genome Res* 11: 1018-1033

International Human Genome Sequencing Consortium (IHGSC). (2001) Initial sequencing and analysis of the human genome. *Nature* 409(6822):860-921

Jauch, A., Wienberg, J., Stanyon, R., Arnold, N., *et al.* (1992) Reconstruction of genomic rearrangements in great apes and gibbons by chromosome painting. *Proc. Natl. Acad. Sci* 89: 8611-8615

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

Jurka, J. and Milosavljevic, A. (1991) Reconstruction and analysis of human Alu genes. *J Mol Evol* 32(2):105-21

Jurka, J., Zietkiewicz, E. and Labuda, D. (1995) Ubiquitous mammalian-wide interspersed repeats (MIRs) are molecular fossils from the mesozoic era. *Nucleic Acids Res* 23(1):170-5

Kadotani, H., Faraco, J. and Mignot, E. (1998) Genetic studies in the sleep disorder narcolepsy. *Genome Res* 8(5):427-34

Koehler, U., Arnold, N., Wienberg, J., Tofanelli, S. and Stanyon, R. (1995a) Genomic reorganization and disrupted chromosomal synteny in the Siamang (Hylobates syndactylus) revealed by fluorescence in situ hybridization. *Am J Phys Anthropol* 97: 37-47

Koehler, U., Bigoni, F., Wienberg, J. and Stanyon, R. (1995b) Genomic reoganization in the Concolor Gibbon (Hylobates concolor) revealed by chromosome painting. *Genomics* 30: 287-292

Landegent, J. E., Jansen in de Wal, N., van Ommen, G. J., *et al.* (1985) Chromosomal localization of a unique gene by non-autoradiographic in situ hybridization. *Nature* 317: 175-177

Landegent, J. E., Jansen in de Wal, N., Dirks, R. W., *et al.* (1987) Use of whole cosmid cloned genomic sequences for chromosomal localization by non-radioactive in situ hybridization. *Hum Genet* 77: 366-370

Langer-Safer, P. R., Levine, M., and Ward, D. C. (1982) Immunological method for mapping genes on Drosophila polytene chromosomes. *Proc Natl Acad Sci* USA 79: 4381-4385

Langford, C. F., Fischer, P. E., Binns, M.M., *et al.* (1996) Chromosome-specific paints from a high-resolution flow karyotype of the dog. *Chromosome Res* 4(2):115-23

Langford, C. F. and Breen, M. (2003) Comparative cytogenetics. *Nature Encylclopedia of the Human Genome*

Lear, T.L. and Bailey, E. (1997) Localization of the U2 linkage group of horses to ECA3 using chromosome painting. *J Hered* 88: 162-164

Lichter, P., Chang Tang, C.-J., Call, K., Hermanson, G., *et al.* (1990) High-resolution mapping of human chromosome 11 by in situ hybridization with cosmid clones. *Science* 247: 64-69

Lichter, P., Boyle, A. L., Cremer, T. and Ward, D.C. (1991) Analysis of genes and chromosomes by nonisotopic in situ hybridisation. *GATA* 8 (1): 24-35

Lieber, M., Ma, Y, Pannicke, U. and Schwarz, K. (2003) Mechanism and regulation of human non-homologous DNA end-joining. *Nature Rev MolCellBio* 4: 712-720

Lin, Y. and Waldman, A .S. (2001) Promiscuous patching of broken chromosomes in mammalian cells with extrachromosomal DNA. *Nucleic Acids Res.* 29: 3975-3981

Lyons, L. A., Laughlin, T. F., Copeland, N. G., Jenkins, *et al.* (1997) Comparative anchor tagged sequences (CATS) for integrative mapping of mammalian genomes. *Nature Genet* 15: 47-56

Maestre, J., Tchenio, T., Dhellin, O. and Heidmann, T. (1995) mRNA retroposition in human cells: processed pseudogene formation. *EMBO J* 14(24):6333-8

Marshall Graves, J. (1998) Mammalian chromosome evolution *ILAR J.* 39: 12-20

Marshall Graves, J. (1999) Comparative Genomics and Mammalian Radiations. *Science*

Mayall, B. H., Carrano, A. V., Moore, D. H. 2nd, Ashworth, L. K., *et al.* (1984) The DNA-based human karyotype. *Cytometry* 5(4):376-85

Mighell, A. J., Markham, A. F., Robinson, P.A. (1997) Alu sequences. *FEBS Lett* 417(1):1-5

Mighell, A. J., Smith, N. R., Robinson, P. A. and Markham, A. F. (2000) Vertebrate pseudogenes. *FEBS Lett* 468(2-3):109-14

Milan, D., Riquet, J., Yerle, M., Goureau, A., Schmitz, A., *et al.* (1996) Homologous and heterologous FISH painting with PARM-PCR chromosome-specific probes in mammals. *Mamm Genome* 7: 194-199

Morescalchi, M .A., Schempp, W., Consigliere, S., Bigoni, F., *et al.* (1997) Mapping chromosomal homology between humans and the black-handed spider monkey by fluorescence *in situ* hybridization. *Chrom Res* 5: 527-536

Muller, S., O'Brien, P. C., Ferguson-Smith, M. A., Wienberg, J. (1997) Reciprocal chromosome painting between human and prosimians (Eulemur macaco macaco and E. fulvus mayottensis). *Cytogenet Cell Genet* 78(3-4):260-71

Nash, W. G., Wienberg, J., Ferguson-Smith, M. A., Menninger, J. C., O'Brien, S.J. (1998) Comparative genomics: tracking chromosome evolution in the family ursidae using reciprocal chromosome painting. *Cytogenet Cell Genet.* 83(3-4):182-92

Nickerson, E. and Nelson, D.L. (1998) Molecular definition of pericentric inversion breakpoints occurring during the evolution of humans and chimpanzees. *Genomics.* 15;50(3):368-72

Novacek, M. J. (1992) Mammalian phylogeny: shaking the tree. *Nature.* 356(6365):121-5

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

O'Brien, S. J., Cevario, S. J. Martenson, J. S. Thompson, M. A., *et al.* (1997) Comparative gene mapping in the domestic cat (*Felis catus) J Hered* 88: 408-414

Ohno, S., Becak, W. and Becak, M. L. (1964) X-autosome ratio and the behaviour pattern of individual X-chromosomes in placental mammals. *Chromosoma* 15: 14-30

Pardo-Manuel de Villena, F. and Sapienza, C. (2001) Female meiosis drives karyotypic evolution in mammals. *Genetics* 159 (3): 1179-1189

Pathak, S., Multani, A. S., Furlong, C. L. and Sohn, S. H. (2002) Telomere dynamics, aneuploidy, stem cells, and cancer (review). *Int J Oncol* 20(3):637-41

Pinkel, D., Straume, T. and Gray, J. W. (1986) Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. *Proc Natl Acad Sci* USA 83: 2934-2938

Prak, E. T. and Kazazian, H. H. Jr. (2000) Mobile elements and the human genome. *Nat Rev Genet* 1(2):134-44

Prakash, B., Kuosku, V., Olsaker, I., Gustavsson, I. and Chowdhary, B. P. (1996) Comparative FISH mapping of bovine cosmids to reindeer chromosomes demonstrates conservation of the X-chromosome. *Chromosome Res* 4: 214-217

Qumsiyeh, M. B. (1994) Evolution of number and morphology of mammalian chromosomes. *J Hered* 85(6):455-65

Raudsepp, T., Frönicke, L., Scherthan, H., Gustavsson, I. and Chowdhary, B.P. (1996) Zoo-FISH delineates conserved chromosomal segments between horse and man. *Chromosome Res* 4: 218-225

Raudsepp, T., Otte, K., Rozell, B. and Chowdhary, B.P. (1997) FISH mapping of the IGF2 gene in horse and donkey-detection of homoeology with HSA11. *Mamm Genome* 8: 569-572

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

Rettenberger, G., Klett, C., Zechner, U., Bruch, J., Just, W., *et al.* (1995b) Zoo-FISH analysis: Cat and human karyotypes closely resemble the putative ancestral mammalian karyotype. *Chromosome Res* 3: 479-486

Rettenberger, G., Abdo, G. and Stranzinger, G. (1996) ZOO-FISH analysis in the horse, Equus caballus, detects regions homologous to human chromosomes 3 and 14. *J Anim Breed Genet* 113: 145-148

Richard, F., Lombard, M. and Dutrillaux, B. (1996) ZOO-FISH suggests a complete homology between human and Capuchin monkey (Platyrrhini) euchromatin. *Genomics* 36: 417-423

Riley, J., Butler, R., Ogilvie, D., Finniear, R., Jenner, D., Powell, S., *et al.* (1990) A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones. *Nucleic Acids Res* 18: 2887-90

Rogozin, I. B., Mayorov, V. I., Lavrentieva, M. V., Milanesi, L., Adkison, L.R. (2000) Prediction and phylogenetic analysis of mammalian short interspersed elements (SINEs). *Brief Bioinform.* 1(3):260-74

Ross, M. T. and Langford, C. F. (1997) The use of flow-sorted chromosomes in genome mapping. In *Genome Mapping: A Practical Approach*. Paul H. Dear Ed. Oxford University Press

Ryder, O. A., Epel, N. C. and Benirschke, K. (1978) Chromosome banding studies of the Equidae. *Cytogenet Cell Genet* 20(1-6):332-50

Samonte, R. V. and Eichler, E. E. (2002) Segmental duplications and the evolution of the primate genome. *Nat Rev Genet* 3: 65-72

Sawyer, J. R. and Hozier, J. C. (1986) High resolution of mouse chromosomes: banding conservation between man and mouse. *Science* 232: 1632-1635

Scherthan, H., Cremer, T., Arnason, U., Weier, H.-U., *et al.* (1994) Comparative chromosome painting discloses homologous segments in distantly related mammals. *Nature Genet* 6: 342-347

Selden, J. R., Moorhead, P. S., Oehlert, M. L., Patterson, D. F. (1975) The Giemsa banding pattern of the canine karyotype. *Cytogenet Cell Genet* 15(6):380-7

Schatzberg, S. J., Olby, N. J., Breen, M., Anderson, L. V. B. *et al.* (1999) Molecular analysis of a spontaneous dystrophin 'knockout' dog. *Neuromuscular Disorders* 9: 289-295

Sherlock, J. K., Griffin, D. K., Delhanty, J. D. A. and Parrington, J. M. (1996) Homologies between human and Marmoset (Callithrix jacchus) chromosomes revealed by comparative chromosome painting. *Genomics* 33: 214-219

Sillar, R. and Young, B. (1981) A new method for the preparation of metaphase chromosomes for flow analysis. *J Histochem Cytochem* 29(1):74-8

Smit, A. F. (1996) The origin of interspersed repeats in the human genome. *Curr Opin Genet Dev* 6(6):743-8

Smit, A. F. and Riggs, A. D. (1996) Tiggers and DNA transposon fossils in the human genome. *Proc Natl Acad Sci U S A* 93(4):1443-8

Smit, A. F. (1999) Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr Opin Genet Dev* 9(6):657-63

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

Spriggs HF, Holmes NG, Breen MG, Deloukas PG *et al.,*(2003) Construction and integration of radiation-hybrid and cytogenetic maps of dog Chromosome X. *Mamm Genome* 14(3):214-21

Stankiewicz, P. Park, S. S., Inoue, K. and Lupski, J. R. (2001) The evolutionary chromosome translocation 4;19 in Gorilla gorilla is associated with microduplication of the chromosome fragment syntenic to sequences surrounding the human proximal CMT1A-REP. *Genome Res* 11(7): 1205-1210

Switonski, M., Reimann, N., Bosma, A. A., *et al.* (1996) Report on the progress of standardisation of the G-banded canine *(Canis familiaris)* karyotype. *Chromosome Res* 4:306-309

Telenius, H., Carter, N.P., Bebb, C.E., *et al.* (1992a) Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. *Genomics* 13: 718-725

Telenius, H., Pelmear, A.H., Tunnacliffe, A., *et al.* (1992b) Cytogenetic analysis by chromosome painting using DOP-PCR amplified flow-sorted chromosomes. *Genes Chrom. Cancer* 4: 257-263

Thomas, R., Breen, M., Langford, C. F. and Binns, M. M. (1999) Zoo-FISH analysis of dog chromosome 5: identification of conserved synteny with human and cat chromosomes. *Cytogenet Cell Genet* 87(1-2):4-10

Thomas R, Breen M, Deloukas P, Holmes NG, Binns MM (2001) An integrated cytogenetic, radiation-hybrid, and comparative map of dog chromosome 5. *Mamm Genome* 12(5):371-5

Tuddenham, E. G., Schwaab, R., Seehafer, J., Millar, D. S., *et al.* (1994) Haemophilia A: database of nucleotide substitutions, deletions, insertions and rearrangements of the factor VIII gene, second edition *Nucleic Acids Res* 22(17):3511-33

van den Engh, G. J., Trask, B. J., Gray, J. W., Langlois, R. G. and Yu, L. C. (1985) Preparation and bivariate analysis of suspensions of human chromosomes. *Cytometry* 6: 92-100

VanDevanter, D. R., Choongkittaworn, N. M., Dyer, K. A., *et al.* (1994) Pure chromosome-specific PCR libraries from single sorted chromosomes. *Proc Natl Acad Sci U S A* 91(13):5858-62

Van Tuinen, P. and Ledbetter, D. H. (1983) Cytogenetic comparison and phylogeny of three species of Hylobatidae. *Am J Phys Anthropol* 61(4):453-66

Vidal, F., Farssac, E., Tusell, J., Puig, L. and Gallardo, D. (2002) First molecular characterization of an unequal homologous alu-mediated recombination event responsible for hemophilia *Thromb Haemost* 88(1): 3-4

Viegas-Péquignot, E., Dutrillaux, B., Magdelenat, H., and Coppey-Moisan, M. (1989) Mapping of single-copy DNA sequences on human chromosomes by in situ hybridization with biotinylated probes: Enhancement of detection sensitivity by intensified-fluorescence digital-imaging microscopy. *Proc Natl Acad Sci* USA 86: 582-586

Vooijs, M., Yu, L.C., Tkachuk, D., Pinkel, D., Johnson, D. and Gray, J.W. (1993) Libraries for each human chromosome, constructed from sorter-enriched chromosomes by using linker-adaptor PCR. *Am J Hum Genet* 52: 586-597

Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., et al., (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature* 420(6915):520-62

Wienberg, J., Jauch, A., Stanyon, R. and Cremer, T. (1990) Molecular cytotaxonomy of primates by chromosomal in situ suppression hybiridzation. *Genomics* 8: 347-350

Wienberg, J., Stanyon, R., Jauch, A. and Cremer, T. (1992) Homologies in human and *Macaca fuscata* chromosomes revealed by in situ suppression hybridization with human chromosome specific DNA libraries. *Chromosoma* 101: 265-270

Wienberg, J., Jauch, A., Lÿdecke, H.-J., Senger, G., *et al.* (1994) The origin of human chromosome 2 analyzed by comparative chromosome mapping with a DNA microlibrary. *Chromosome Res* 2: 405-410

Wienberg, J. and Stanyon, R. (1995) Chromosome painting in mammals as an approach to comparative genomics. *Curr Opin Genet Dev* 5: 792-797

Wienberg, J., Stanyon, R., Nash, W.G., O'Brien, P.C.M., *et al.* (1997) Conservation of human vs. feline genome organization revealed by reciprocal chromosome painting. *Cytogenet Cell Genet* 77: 211-217

Wienberg, J. and Stanyon, R. (1997) Comparative painting of mammalian chromosomes. *Curr Opin Genet Dev* 7: 784-791

Yang, F., Muller, S., Just, R., Ferguson-Smith, M.A. and Wienberg, J. (1997) Comparative chromosome painting in mammals: Human and the Indian muntjac (*Muntiacus muntjak vaginalis*). *Genomics* 39: 396-401

Yang, Y. -P., Rexroad, III C. E., Schlopfer, J., Womack, J. E. (1998) An integrated radiation hybrid map of bovine chromosome 19 and ordered comparative mapping with human chromosome 17. *Genomics* 48: 93-99

Yerle, M., Pinton, P., Robic, A., Delcros, C., Hawken, R., *et al.* (1998) Construction of a whole genome radiation hybrid panel for high-resolution gene mapping in pigs. *Proc XXVI Internat Conf On Anim Genet* August 9-14, Auckland, New Zealand, p.60

Yokobata, K., Trenchak, B. and de Jong, P. J. (1991) Rescue of unstable cosmids by in vitro packaging. *Nucleic Acids Res* 19(2):403-4

Yunis, J. J. and Prakash, O. (1982) The origin of man: a chromosomal pictorial legacy *Science* 215: 1525-1529

STS	Primer Sequences (5' → 3')	Size (bp)
stfF45C1A1	ATGGGGTCTTGCTATGTTGC ACGTGTCCCTTGGATGACTC	171
stfF45C1A2	CACCTAGGCACAGCATACGA CGCCCTCTGAGAATGAGC	121
stfF45C1A3	CATCCCGTGGCTAACAATG TAAGCCAGGCTGACAGAGGT	149
stfF45C1A4	GATGCCCACTGATGGGAG GGGCAAGAAGAATCCATTGA	131
stfF45C1A5	GAATGCAGCTTCAGTCTTTGG ACAGGCTAAATTGCTTTTTTGC	143
stfF45C1A6	AACAAACAGAGGGAACACGC ACCTGGCAGGAGGAATTTG	137
stfF45C1B1	AGCATCTGTAATTGCCTGGG TATGGTCAGAAAAATCCCGC	125
stfF45C1B2	CCAGGGCAGAAGGAACATAA CAGCACCCAGTGAGCTATGA	127
stfF45C1B3	GCTTTGGAAACCCACGTG CTACATGTCTGCTGGCATTCA	120
stfF45C1B4	TCAAATGCATGTGATACATTGC GCAATTATCCACTGAGGTTTCA	164
stfF45C1B5	AACGAACAAGACTTCGAGAAGG GCTTGCTTGCTTGCTTTCTT	151
stfF45C1B6	CCAGAAAAGCTTGCCTTGAC AGCAGGCAAGTGGAATTCC	125
stfF45C1B7	TCCTCTGTTCCAGCGAGG ATCCAAATCTGCCTCCCAG	144
stfF45C1B8	CTCAGTTGTCAGAGGCCACA TGGAGGCTCAGCTCCATC	166
stfF45C1B9	AAGGTGGATACTTGATTGGGG TCCCTCTCACATTCCTCACC	157
stfF45C1B10	AGCATTTCAGATGCAATTTGC TTCAATTGACTTCCGTTTTTCC	184
stfF45C1B11	AATATCCTTGGATGGCTCCC AGTGCAGGGCTCTGGATG	76
stfF45C1B12	ACAGTTTTGAGTCTCTGGAGGG GTTCAAGAGGATTCCCAAAGC	114
stfF45C1B13	AAGCCCAAATGTTCAAGCAC CCAGAGGCTGAGATGGTGAT	115
stfF45C1B14	GGGTACTGAGCTCTTGCCAC CAGGAAGCTGCATCTTATTGC	150

STS	Primer Sequences (5' → 3')	Size (bp)
stfF45C1B15	CCATGCATTCATTCCAATACC CCTGGAGTGCTCAGTGTTCA	161
stfF45C1B16	GTTGGTGACTGCTGCATTGT AATGATGAGACTTGGGTGGG	126
stfF45C1B18	AGAATGGCCCTCTCCTGG GTCTGGGAGATGTTCAAGTTCC	131
stfF45C1B19	ACTTTGATGCAAGTCAGGGG TTCCTACACCCCACTTCCAC	175
stfF45C1B20	GATCCACACTAGGTGGAGAAGG CATCCTACCCCAGGTCTTGA	154
stfF45C1B21	AACCGTCCTGCTGCAGAC ACGTCACAGACTCATTCACCC	134
stfF45C1B22	CCACCCCACACTCACACTC AGCAACAGGGCCAGACTG	120
stfF45C1B23	TTCCTTTCCGACTGCAGC ACAGACTGAGGTGGGAATGTG	186
stfF45C1B25	CCTAAATCCAAAGGTCATCAGC GCTTGTCTCTGGTCAGGGAG	120
stfF45C1B27	AAGGAAGAGTTGCACACATCTC ATTTTCCCAAGAGGCCAAGT	102
stfF45C1B28	TACAGTTTGGCTCTGTGTCCC TATGGAGGAAACTGTCCCCA	120
stfF45C1B30	ATAAGGCAATCGCCAATCTG CAATTCGTGAATTGGGCAG	165
stfF45C1B31	GCGAGGTAAAAGGGCTCAG CTTCTTGGTCACTTTGCAAGC	132
stfF45C1B32	CAACTCTGGATGGCCAATG TCATTGAGGTGAGCAAGGC	129
stfF45C1B33	TGGCAACCTTTGATATGCAA TTTAATATGCAAATGCAGGGC	165
stfF45C1B34	AAAAGGCACCTCAGTCTCACA AGGTCCTCCCATTGCAAAG	132
stfF45C1B35	TGAGCCTCACTGAGCGTG CTTATCAGACTGAGATGCGGC	183
stfF45C1B36	AGGACATAAGCTGGGGGTCT ACTGAGGACCCTGGAGTGG	135
stfF45C1B37	TCCTTCAGCACCAGGCTC ACACTGCAGAGGCCTCATG	168
stfF45C1B38	ACCACTGGGACCCTCAAAG GCCCAGTGGAGAGAACTGAG	149