CHAPTER 1

Introduction

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.

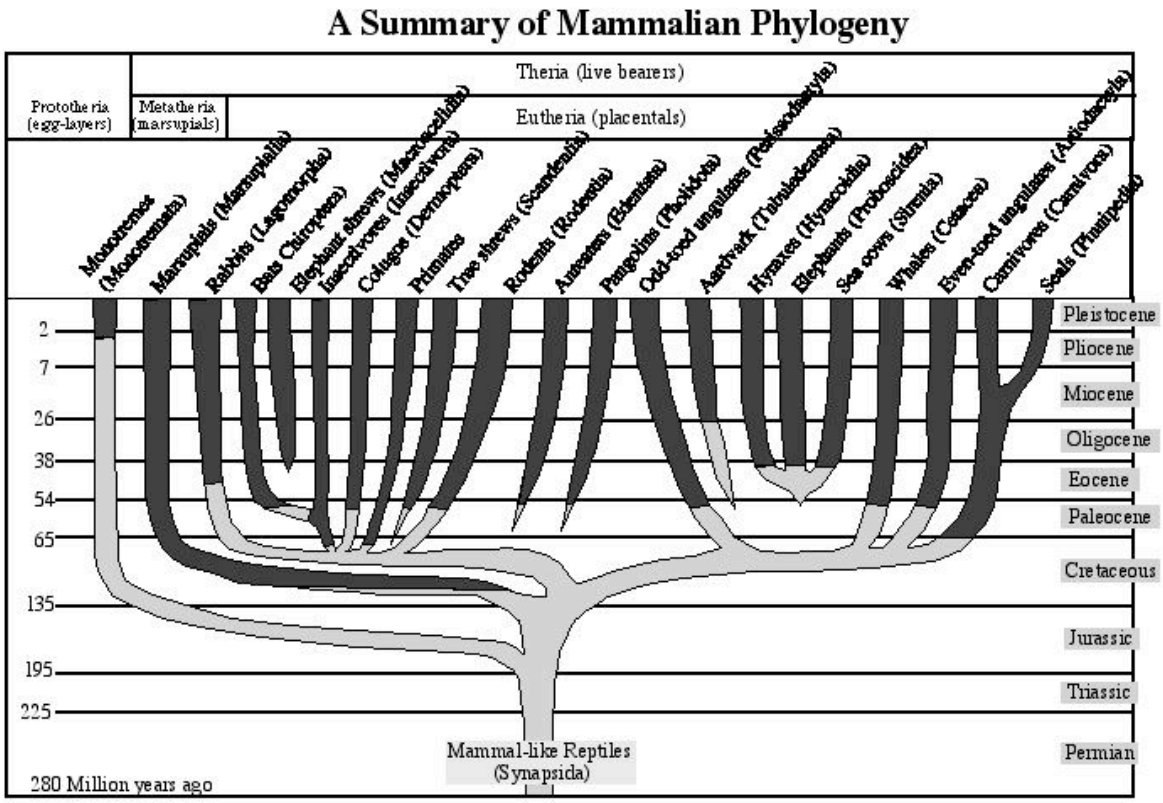


Figure 1.1 The divergent relationship between the Protheria, the Metatheria and the Eutheria is shown along the horizontal axis in the context of geological era and timescale (on the vertical). Reproduced from http://www.qmw.ac.uk/~ugbt991/mammals/week6slides/sld002.htm

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.

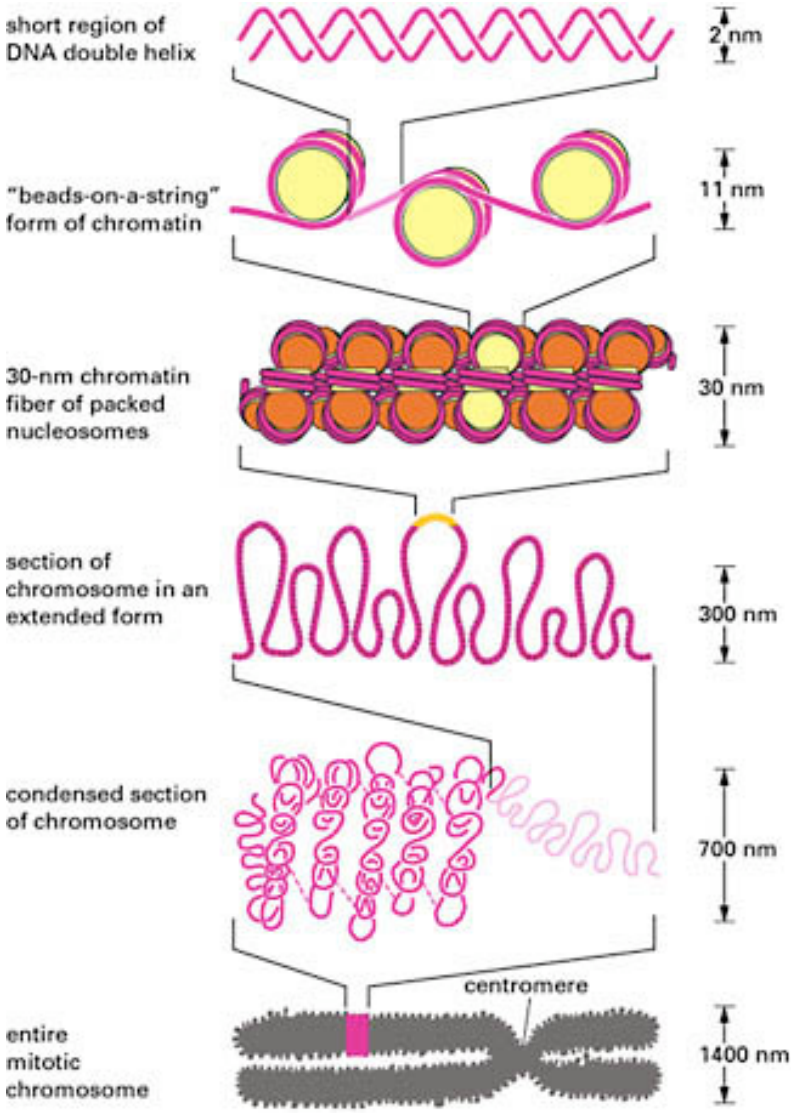


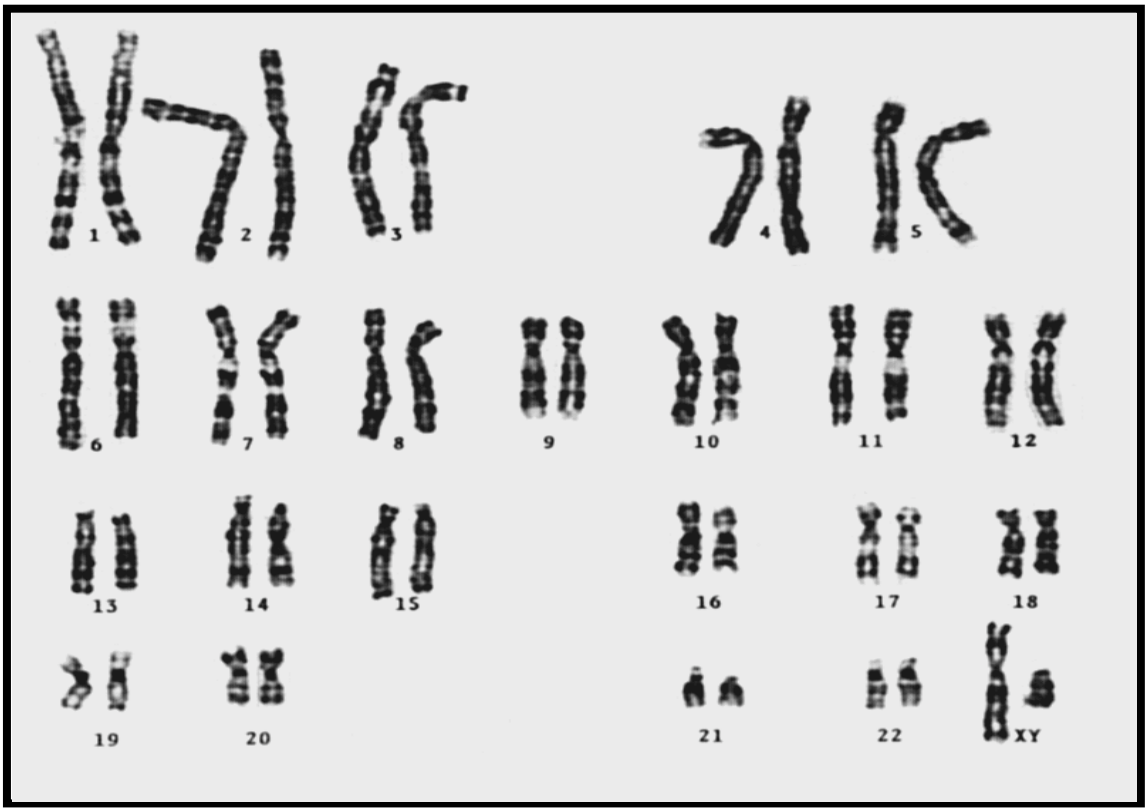
Figure 1.2 Schematic illustrating some of the orders of chromatin packing thought to give rise to the highly condensed mitotic chromosome. Reproduced from Alberts, Bray, Johnson, Lewis, Raff, Roberts and Walter, 1998 © Garland Publishing http://www.garlandscience.com/ECB/about.html

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. Transposon-derived interspersed repeats

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).

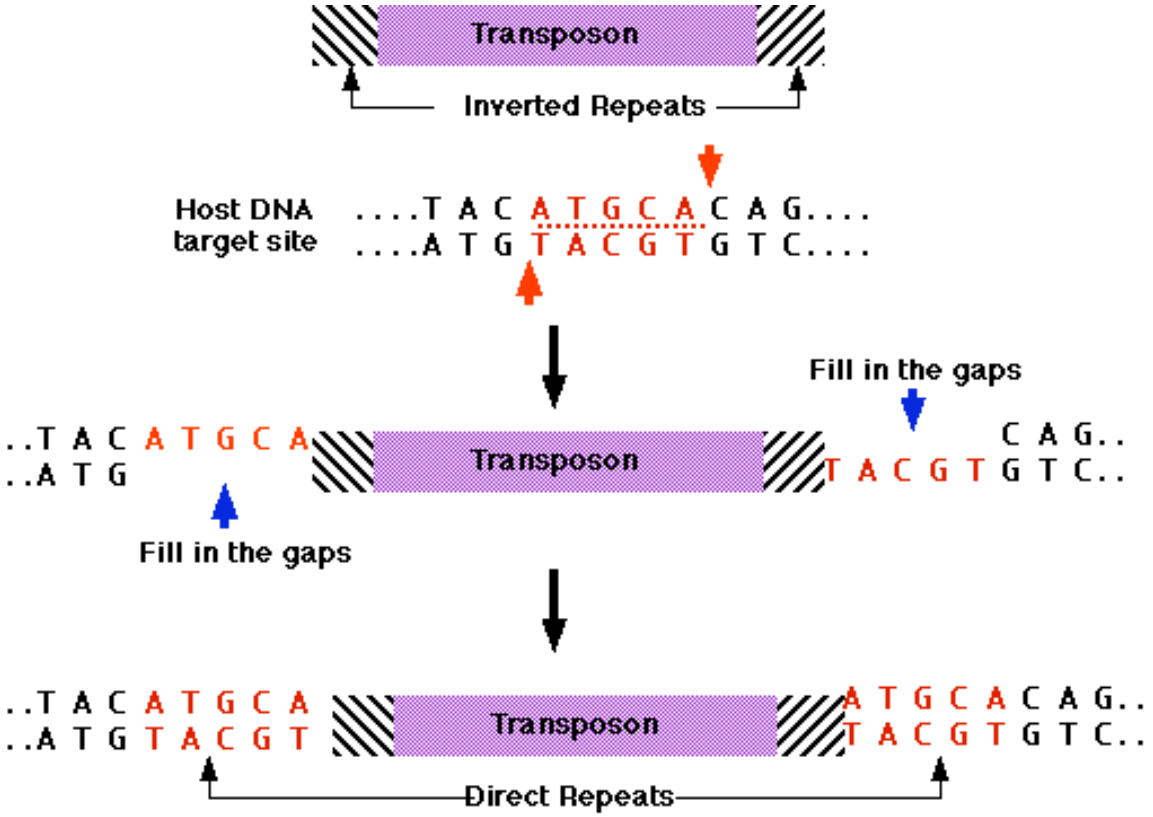


Figure 1.4 Illustration of the mechanism by which a transposon integrates into its target site. Reproduced from http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/T/.

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. Processed pseudogenes

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. Simple sequence repeats

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. Segmental duplications (LCRs)

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. Blocks of localised tandem repeats

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

Inversions

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.

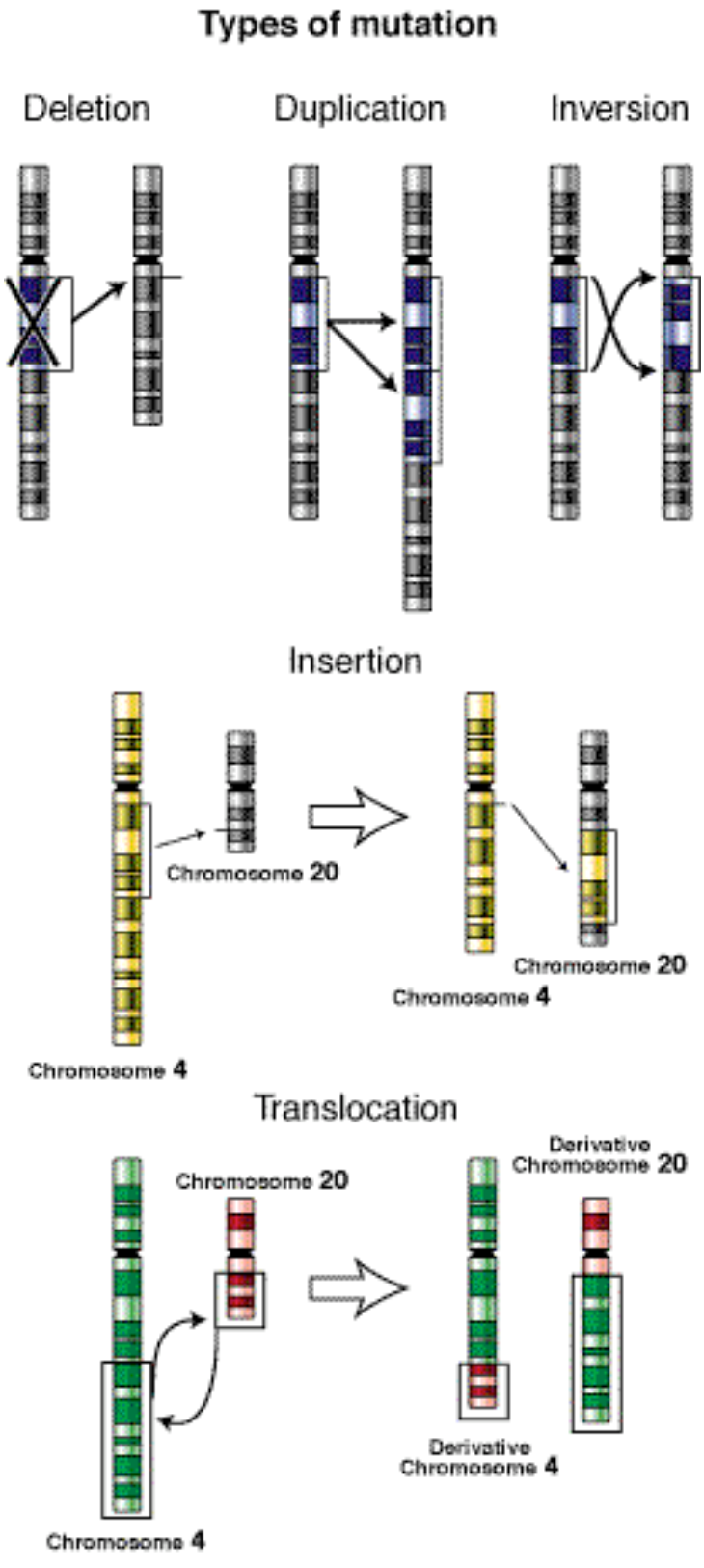


Figure 1.5 Schemtaic illustration of chromosome rearrangements and mutations

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.

Fusions

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 Comparative FISH mapping

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 Comparative Chromosome Painting

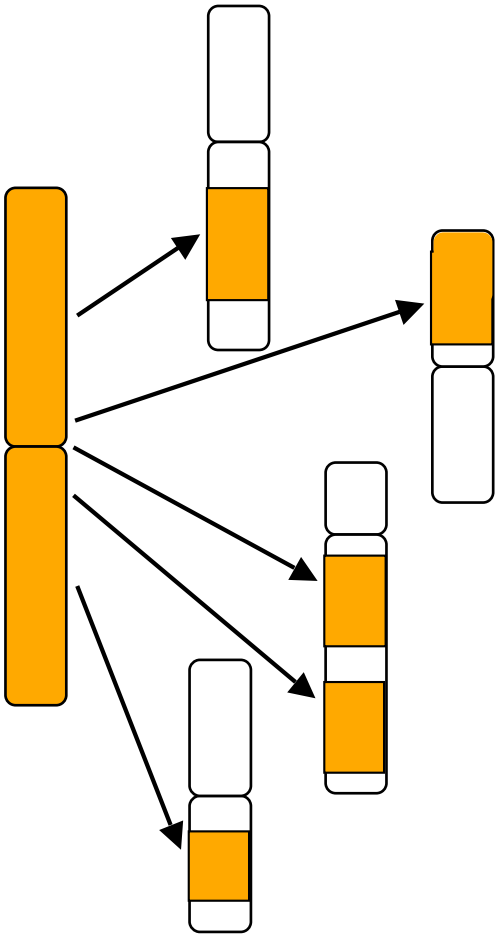
The FISH mapping of individual genes for comparative purposes is time consuming and gives only patchy information on chromosome homology between species. However, this problem can be overcome if chromosome paints are used for FISH. Chromosome paints are complex mixtures of probes, which can be synthesised 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).

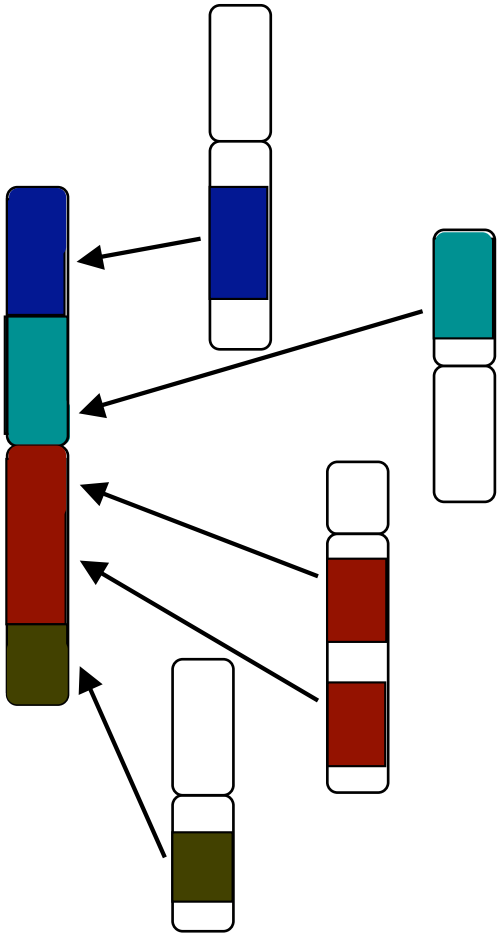
Figure 1.6 (next page) illustrates forward and reciprocal chromosome painting schematically. In a standard forward painting experiment, a whole-chromosome paint from one species (species A) highlights homologous segments in the chromosomes of another species (species B). But the sub-regional origin of 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



Species A Species B

REVERSE



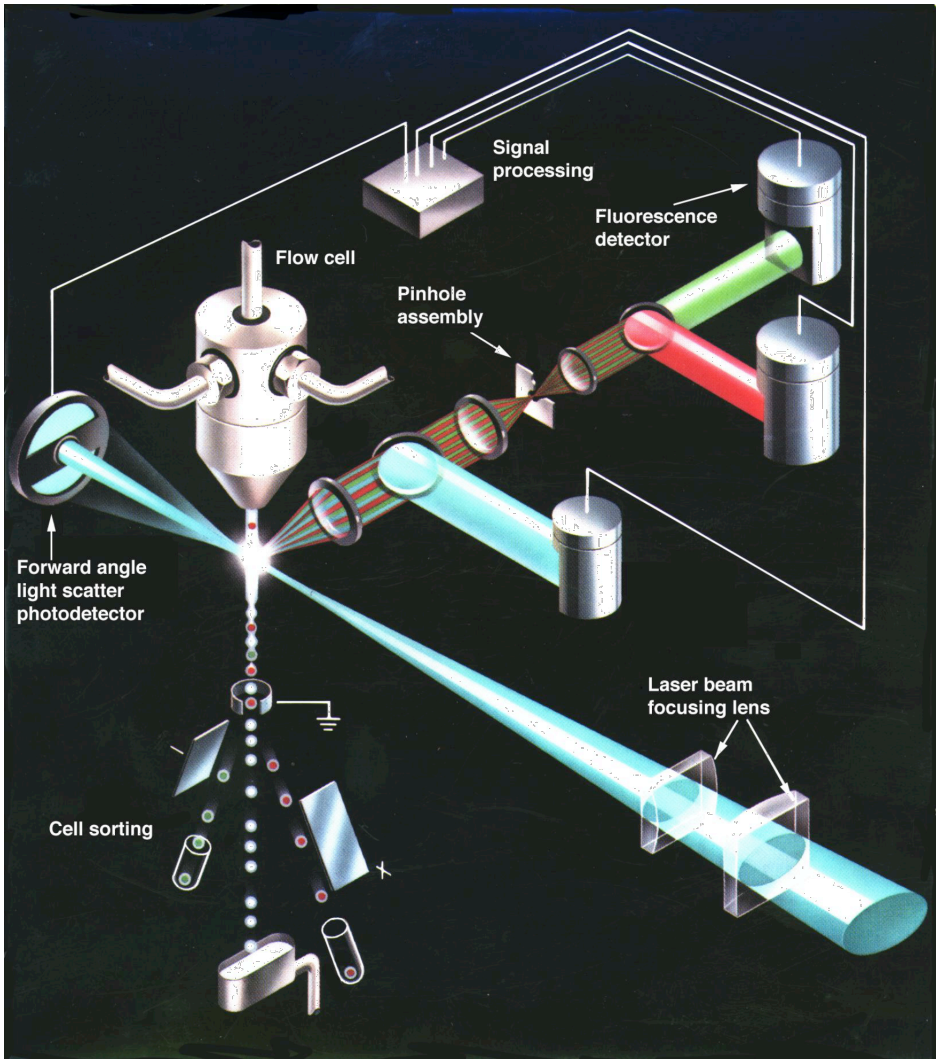
Species A Species B

1.5.4.2.1 Chromosome flow sorting

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.

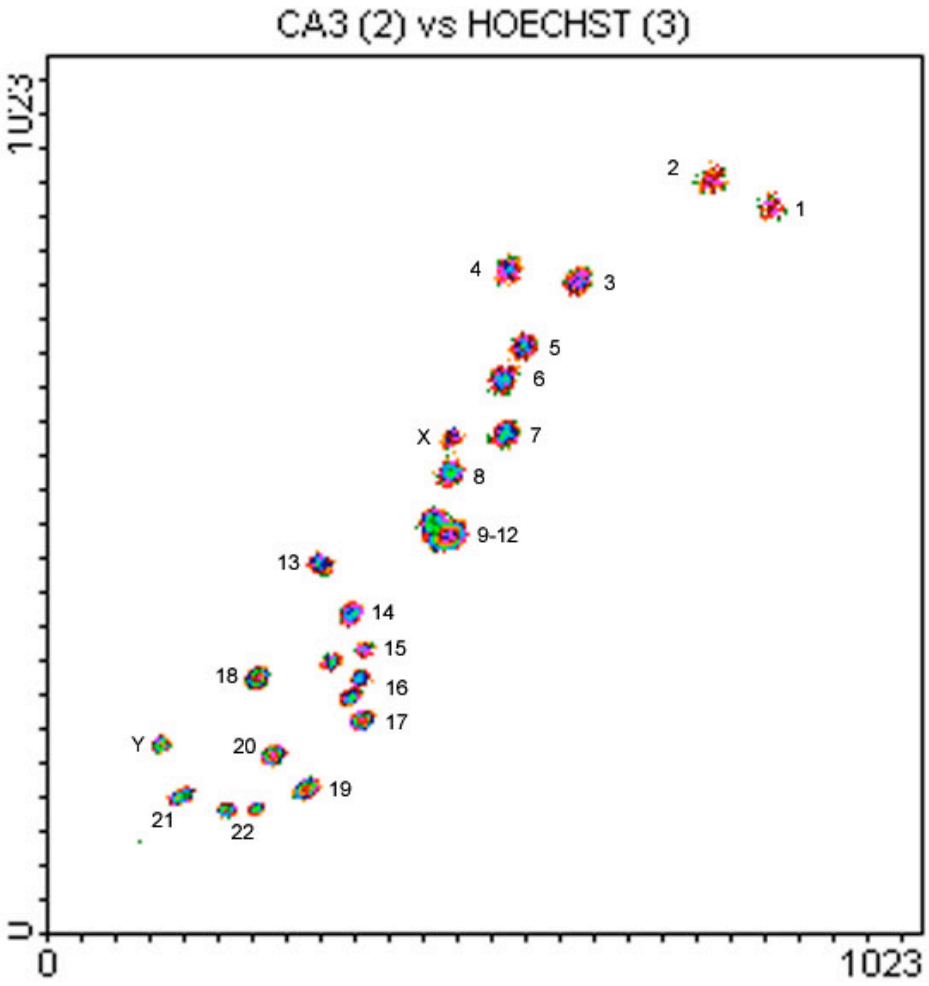
The lay out of a typical commercially available dual-laser flow cytometer is shown in Figure 1.7

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.

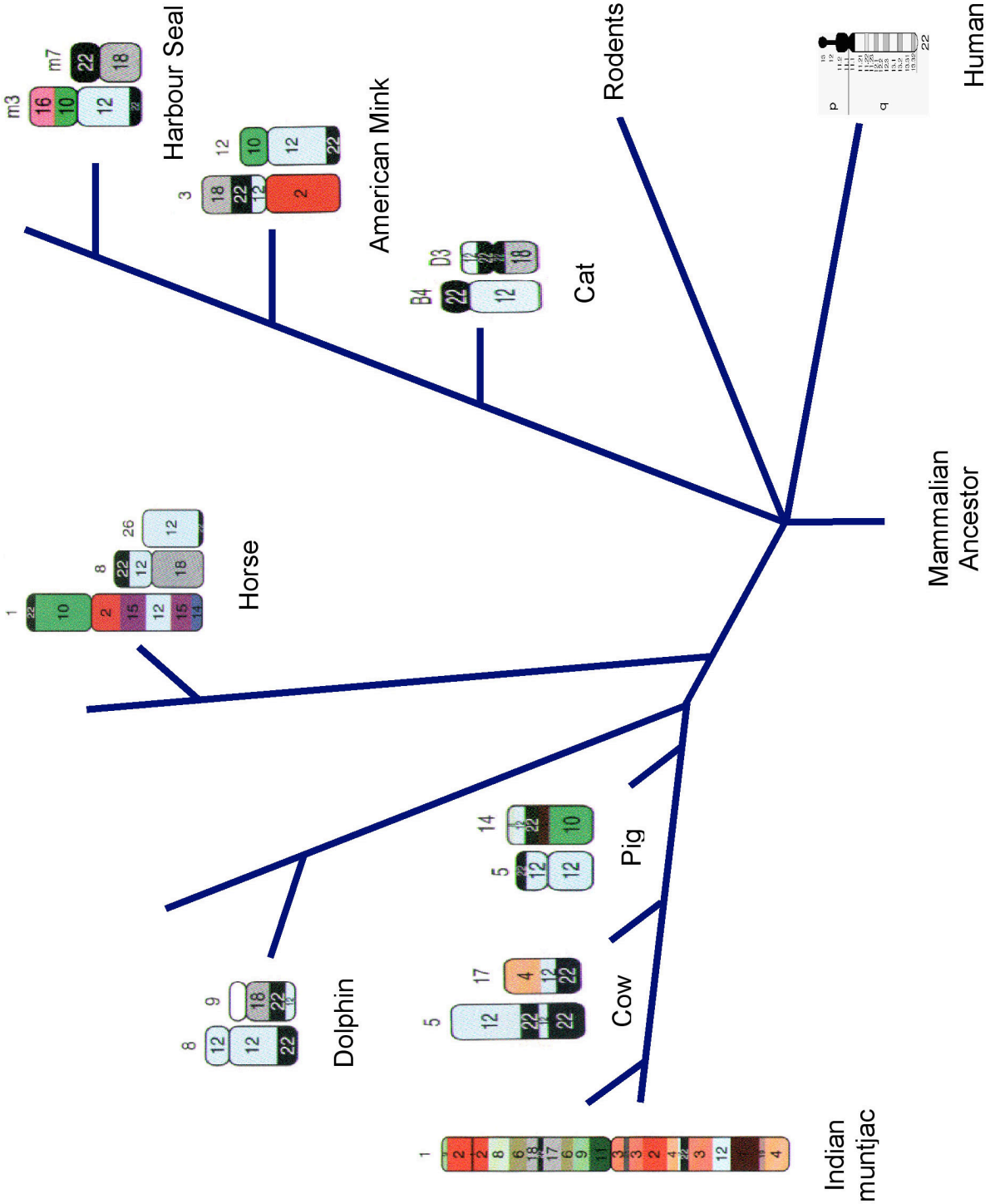
Human chromosome 22

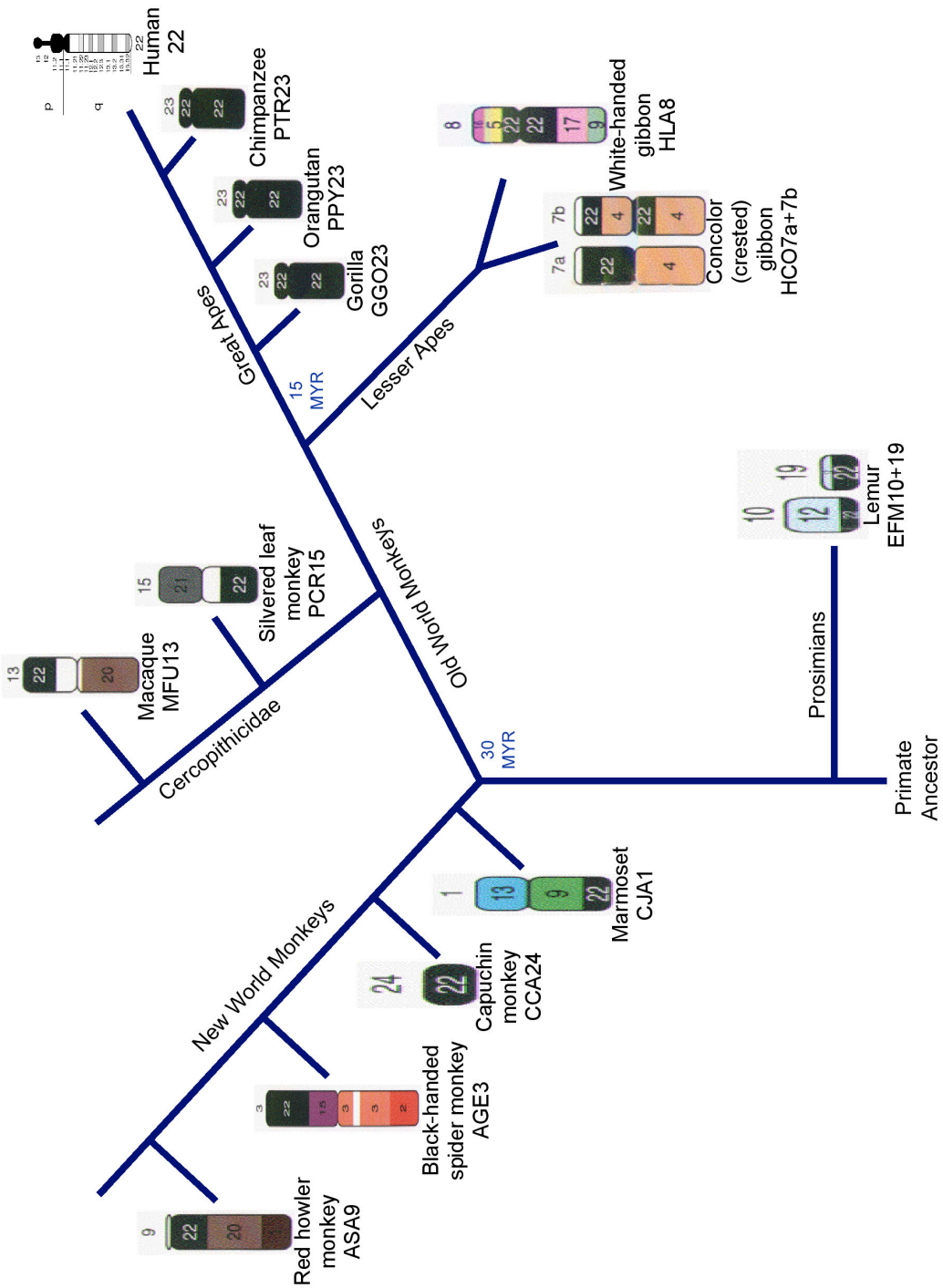
Human chromosome 22 is the second smallest of the human autosomes, being 48 to 54

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.


Harbour Seal
American Mink
Rodents
Human
Horse
Cat
Mammalian Ancestor
Dolphin
Pig
Cow
Indian muntjac



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.