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.