University of Ulm
Department of Human Genetics
Prof. Dr. Walther Vogel



Identification of Large-Scale DNA Copy Number Differences Between Human and Non-Human Primate Genomes and their Role in Mediating Evolutionary Rearrangements

Dissertation

To obtain the doctoral degree of Human Biology

At the Faculty of Medicine, University of Ulm

Presented by

Violaine GOIDTS

from Namur, Belgium

Ulm, 2006

Amtierender Dekan: **Prof. Dr. Klaus-Michael Debatin**

1. Berichterstatter **Prof. Dr. Horst Hameister**

2. Berichterstatter **Prof. Dr. Werner Schempp**

Tag der Promotion: **05.05.2006**

Erklärung

Ich versichere hiermit, dass ich die vorliegende Arbeit selbstständig angefertigt und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie wörtlich oder inhaltlich übernommene Textpassagen als solche gekennzeichnet habe.

Ulm, 27.01.2006

Violaine GOIDTS

“Most scientific theories either deal with questions that only interest specialists or require a forbidding set of technical skills before you can understand them. The theory of evolution is scientifically unique in its combination of universal interests and accessibility.”

Mark Ridley, Evolution 2nd Edition.

Table of content

1. INTRODUCTION ____ *10*

1.1. Importance to do the human-chimpanzee comparison ____ 10

1.2. Differences between human and non-human primate genomes ____ 12

1.2.1. Nucleotide divergence ____ 12

1.2.2. Karyotypic differences ____ 13

1.2.3. Sub-microscopic differences ____ 14

1.2.4. Copy number differences: a class of micro-differences in itself ____ 16

1.3. The array comparative genomic hybridization (aCGH) technique ____ 24

1.4. Objective of the study ____ 26

2. MATERIALS AND METHODS ____ *27*

2.1. Materials ____ 27

2.1.1. Oligonucleotides ____ 27

2.1.2. DNA samples ____ 29

2.1.3. Cell lines ____ 30

2.1.4. BAC clones ____ 31

2.1.5. Chemicals and reagents ____ 32

2.1.6. Enzymes ____ 34

2.1.7. Antibodies ____ 34

2.1.8. Equipment ____ 34

2.1.9. Kits ____ 35

2.2. Methods ____ 36

2.2.1. DNA technology ____ 36

2.2.2. RNA technology ____ 46

2.2.3. Cell technology ____ 47

2.2.4. Chromosome technology ____ 48

2.2.5. *In silico* analysis ____ 52

3. RESULTS ____ *54*

3.1. Study of micro-differences between human and primate genomes. ____ 54

3.1.1. Number of differences detected by aCGH in primate genomes ____ 54

3.1.2. High number of species-specific rearrangements in the gorilla lineage ____ 59

3.1.3. Human-specific rearrangements ____ 61

3.1.4. Ancestral loci of the HLS-CND-detecting DCs ____ 71

3.1.5. Boundaries of putative HLS-CNDs and their respective homologous DCs ___ 72

3.2. Study of two macro-differences between human and primate genomes: examples of evolutionary rearrangements driven by LCRs ___ 74

3.2.1. The pericentric inversion of chimpanzee and gorilla chromosomes homologous to human chromosome 16 ___ 74

3.2.2. The human specific pericentric inversion of chromosome 18 ___ 86

4. DISCUSSION ___ 92

4.1. Characterization of DNA copy number differences between human and non-human primates ___ 92

4.2. Human-specific copy number differences ___ 94

4.3. Genes contained in the putative HLS-CNDs ___ 96

4.4. Polymorphisms vs. differences fixed in the human population ___ 97

4.5. Presence of evolutionary young repeat sequences at the boundaries of the DCs identified ___ 99

4.6. Non-random distribution of CNDs ___ 100

4.6.1. CNDs localized in regions of chromosomal instability ___ 100

4.6.2. CNDs localized in the vicinity of evolutionary chromosomal breakpoints ___ 101

4.6.3. Pericentric inversions of chromosomes 16 and 18, further examples of rearrangements driven by LCRs ___ 102

4.7. Final conclusion ___ 108

5. REFERENCES ___ 109

6. SUMMARY ___ 131

7. APPENDIX ___ 132

Abbreviations

aCGH	Array Comparative Genomic Hybridization
AIDS	Acquired Immunodeficiency Syndrome
ANAPC1	Anaphase Promoting Complex Subunit 1
BAC	Bacterial Artificial Chromosome
BLAST	Basic Local Alignment Search Tool
Bp	Base Pair
cDNA	Complementary DNA
CND	Copy Number Difference
CTD	California Institute of Technology library D
DC	Duplication Cluster
DNA	Desoxyribonucleic Acid
dNTP	Deoxynucleoside Triphosphate
EBV	Epstein Barr Virus
ECACC	European Collection of Cell Cultures
EST	Expressed Sequence Tag
FISH	Fluorescence In Situ Hybridization
FMRPD2	FERM and PDZ Domain Containing 2
FOXP2	Forkhead Box P2
GGO	*Gorilla gorilla*
HLS	Human Lineage Specific
HPRT	Hypoxanthine-Guanine Phosphoribosyltransferase
HSA	*Homo sapiens*
Kb	Kilobase
Mb	Megabase
MFA	*Macaca fascicularis*
MFU	*Macaca fuscata*
mRNA	messenger RNA
Mya	Million years ago
Myrs	Million years
NHEJ	Non Homologous End Joining
PCR	Polymerase Chain Reaction

PPA	*Pan paniscus*
PPY	*Pongo pygmaeus*
PTR	*Pan troglodytes*
RNA	Ribonucleic Acid
ROCK1	Rho-Associated Coiled-Coil Containing Protein Kinase 1
RP	Roswell Park
SHCBP1	Shc SH2-Domain Binding Protein 1
SNP	Single Nucleotide Polymorphism
TAR	Telomere Associated Repeat
THOC1	THO Complex Subunit 1
USP14	Ubiquitin Specific Protease 14
UTR	Untranslated Region
µM	Micromolar
µg	Microgramm
µl	Microliter

1. INTRODUCTION

1.1. Importance to do the human-chimpanzee comparison

“What sets us apart from other species?” is what interests the most the humans. One of the important ways to answer this question is to compare the human genome with the genomes of other species. Up until the end of year 2003, the limited availability of complete genome sequences reduced such comparisons to species such as mouse, fish, and worm, which are distantly related to humans. These comparisons are relevant to understand what is common to all mammals, all vertebrates, or all animals, but uninformative with regard to what is unique to primates (Enard and Pääbo, 2004).

This situation changed radically with the completion of the draft sequence of the chimpanzee genome (Chimpanzee Sequencing and Analysis Consortium, 2005). Humans and chimpanzees diverged about 5 - 6 million years ago, making the chimpanzee the closest relative to human (Goodman et al., 1999; Glazko and Nei, 2003) (Figure 1.1).

Since the chimpanzee genome shows remarkable similarity to that of humans, it will help us to understand the hominines evolution and to determine the genetic basis of human uniqueness.

Moreover, human-chimpanzee comparisons also have more practical implications. Indentifying genes that are divergent between humans and chimpanzees may be extremely helpful in understanding disease susceptibility. For example, even though the chimpanzee DNA sequence is 98.77 % identical to that of the human (Chen and Li, 2001; Ebersberger et al., 2002; Chimpanzee Sequencing and Analysis Consortium, 2005), chimpanzees do not suffer from some diseases that strike humans, such as AIDS, *Plasmodium falciparum* malaria, epithelial cancers, etc (McClure, 1973; Novembre et al., 1997; Olson and Varki, 2003; Martin et al., 2005).

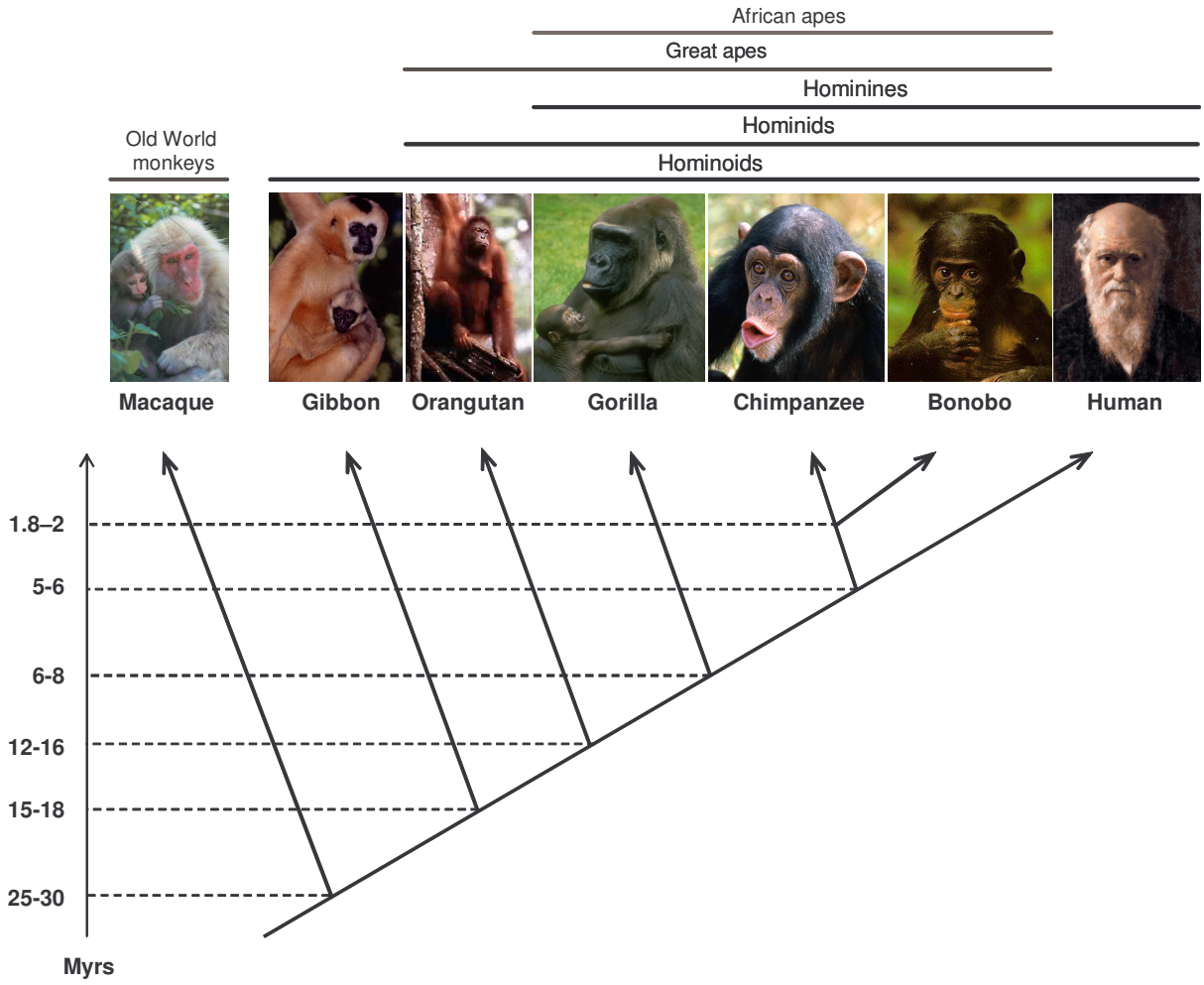


Figure 1.1: Sequence-based phylogenetic tree for the human, the great apes and an Old World monkey (Miyamoto et al., 1988). The numbers indicate estimated millions of years (Myrs) from the last branch point to the present for the respective lineage. The times of divergence are not drawn to scale.

1.2. Differences between human and non-human primate genomes

1.2.1. Nucleotide divergence

Nucleotide substitutions are the most frequent type of difference that exist between human and chimpanzee genome. The recent genome-wide analysis of the sequence of the chimpanzee confirmed earlier observations (Chen and Li, 2001) and set up single-nucleotide substitutions at a mean rate of 1.23 % between human and chimpanzee genomes (Chimpanzee Sequencing and Analysis Consortium, 2005). However, single nucleotide polymorphisms (SNPs) should be taken into account in both species. 14-22 % of the sequence divergence observed is estimated to be polymorphisms in one or the other species. This estimation brings down the number of fixed divergence to ~1.06 % or less (Chimpanzee Sequencing and Analysis Consortium, 2005). It should be noted that the average sequence divergences between humans and gorillas, humans and orangutans, and humans and macaques are 1.6 %, 3.1 %, and 4.9 %, respectively (Chen and Li 2001; Magness et al., 2005). The study of sequence divergence between human and other primates coding regions allows the calculation of the rate of evolution for the genes studied. By applying the ratio of the number of nonsynonymous substitutions (Ka) to the number of synonymous substitutions (Ks) (Ka/Ks), it is possible to determine whether a gene has been positively selected or not during evolution. Thus, genes showing a Ka/Ks > 1 implies adaptive or positive selection, while a Ka/Ks < 1 indicates that changes of amino acids have been eliminated by negative selection, also called purifying selection (Hurst, 2002). However, the statistical power in the Ka/Ks test might be low since both Ka and Ks are estimated from the coding region of a gene. It should be considered that positive selection might act at some sites and negative on others, resulting in a Ka/Ks lower than 1, even though positive selection has occurred. Moreover, a single amino acid change can take place in the active sites of proteins and therefore have a significant impact on their function (Ellegren, 2005). Recently, the sequences of 13,731 human genes have been compared with their chimpanzee orthologs (Nielsen et al., 2005). Many of the genes that seemed to be positively

selected tend to be involved in sensory perception or immune defence. A well-known example of human-specific positive selection is the selection of the FOXP2 gene, where two amino acid exchanges have occurred in the human lineage since it separated from chimpanzee, about 5 million years ago (Enard et al., 2002a). Since FOXP2 is the first gene known to affect specifically the development of speech and language in humans (Lai et al., 2001), it is likely that the selected change, giving rise to a selected phenotype, was related to speech and language development in humans, among other factors and genes involved (Fisher and Marcus, 2006).

1.2.2. Karyotypic differences

The human genome differs to that of the hominoids at different level. In 1982, Yunis and Prakash have highlighted a few disparities that exist between those genomes. During the last ten years, those differences have been studied using cytogenetic and molecular techniques, as well as sequence data, which allowed for several changes a characterization at the base pair level. The most obvious structural difference is the fusion of two acrocentric chromosomes giving rise to human chromosome 2 (Fan et al., 2002). This gives rise to a lower total chromosome number in humans (22, XY vs. 23, XY). This discrepancy in the number of chromosomes between human and the great apes results in a different numbering of the human chromosome and its homologue in the apes. This problem led the cytogeneticists to adopt a phylogenetic nomenclature using Roman numerals for the homologous chromosomes in apes (McConkey, 2004). This nomenclature, listed in Table 1.1, is the one I used in this study.

In addition to the fusion of chromosome 2, there are several additions of heterochromatin (van Geel et al., 2002) and nine visible pericentric inversions, affecting chromosomes 1, 4, 5, 9, 12, 15, 16, 17, and 18 (Yunis and Prakash, 1982; Nickerson and Nelson, 1998; de Pontbriand et al., 2002). Of these inversions, only those on chromosomes 1 and 18 are specific to the human lineage (Goidts et al., 2004; Szamalek et al., submitted), while the remaining changes have occurred in the chimpanzee and gorilla lineages (Montefalcone et al., 1999; Kehrer-Sawatzki et al., 2002; Locke et al., 2003a; Dennehey et al., 2004; Goidts et al., 2005; Kehrer-

Table 1.1: Primates phylogenetic nomenclature used in this study.

HSA (***Homo sapiens***)	**Homologous chromosome in primates**
1	I
2	II (IIa, IIb)
3	III
4	IV
5	V
6	VI
7	VII
8	VIII
9	IX
10	X
11	XI
12	XII
13	XIII
14	XIV
15	XV
16	XVI
17	XVII
18	XVIII
19	XIX
20	XX
21	XXI
22	XXII
X, Y	X, Y

Sawatzki et al., 2005a; Kehrer-Sawatzki et al., 2005b; Kehrer-Sawatzki et al., 2005c; Szamalek et al., 2005; Shimada et al., 2005).The differences in chromosome structure kept the attention of numerous scientists since those changes are hypothesized to be a cause of reproductive isolation. Indeed recombination is reduced in chromosomes heterozygous for the inversions. Because recombination is required for gene flow, the rearrangements create a semi-permeable reproductive barrier, while gene flow continues for genes on collinear chromosomes. Consequently, selected differences would accumulate more easily in rearranged than in collinear chromosomes, driving the isolation of the species, and, therefore, speciation (Noor et al., 2001; Hey, 2003; Navarro and Barton, 2003; Rieseberg and Livingstone, 2003). This hypothesis, though very attractive, was not reproducible by other studies (Lu et al., 2003; Vallender and Lahn, 2004a; Zhang et al., 2004).

1.2.3. Sub-microscopic differences

Early comparative studies between the genomes of humans and other primates have focussed mainly on the characterization of karyotypically visible chromosomal rearrangements (Yunis and Prakash 1982; Mueller and Wienberg 2001). However, the characterization of smaller changes between human and chimpanzee genomes is essential for the understanding of the molecular basis of both species.

In addition to the divergence due to single nucleotide substitutions, a variety of micro-insertions and micro-deletions, also called indels, have been identified (Britten 2002; Britten et al., 2003; Frazer et al., 2003; Liu et al., 2003). These indels have an average length of 36-nucleotides, including indels up to 10-kb. The overall difference due to indel events between the human and chimpanzee genomes corresponds to ~3 % of both genomes.

Recently, by comparing human and chimpanzee genome assemblies, two studies advanced the existence of small inversions (also called micro-inversions), ranging 23-base pairs to several mega bases of DNA (Feuk et al., 2005; Szamalek et al., 2006). However, it should be noted that such assemblies' comparison is limited. Indeed, several of those micro-inversions could not be verified by independent molecular and cyotgenetic methods, most probably because of the currently low-coverage draft sequence of the chimpanzee assembly.

Lately, a number of studies have been performed with the goal of characterizing loss and gain of submicroscopic regions of DNA between human and chimpanzee genomes (Locke et al., 2003b; Fortna et al., 2004; Cheng et al., 2005; Newman et al., 2005; Wilson et al., 2006). Using comparative genomic hybridization and detailed sequence comparison of the human and chimpanzee genome assemblies, duplications or deletions from 20-kb of up to hundreds of kilobases in size were identified. The results revealed that copy number differences (CNDs) are abundant between the human and the chimpanzee genome. A summary of the differences between human and chimpanzee genomes is represented on Figure 1.2.

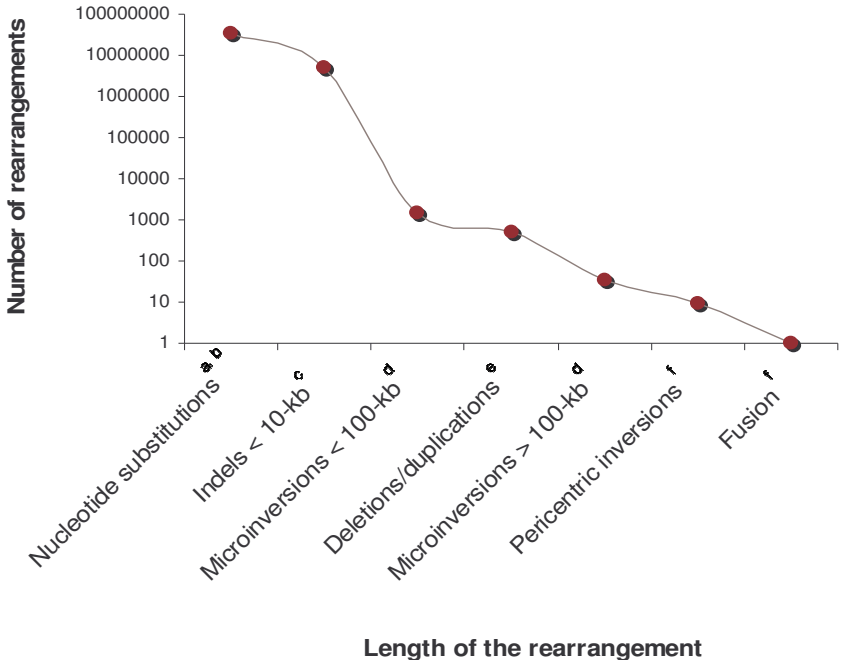


Figure 1.2: Graphical representation of the differences that exist between human and chimpanzee genomes. The rearrangements are ordered by size (x-axis). The number of rearrangements is represented on a logarithmic scale (y-axis). [a]Chimpanzee Sequencing and Analysis Consortium (2005); [b]Chen and Li (2001); [c]Britten (2002); [d]Feuk et al. (2005); [e]Cheng et al. (2005); [f]Yunis and Prakash (1982).

1.2.4. Copy number differences: a class of micro-differences in itself

The study of large-scale copy number differences (CNDs), such as deletions or duplications, between different genomes is of great importance in the understanding of the evolution of organisms. First, genes embedded within duplicated sequences often identify regions of adaptive evolution within species (Johnson et al., 2001; Fortna et al., 2004). Moreover, it has been shown that duplications occurred preferentially in gene-rich regions (Bailey et al., 2002a). Second, at a structural level, regions of highly homologous duplications are preferential sites of chromosomal instability (Stanckiewicz et al., 2001; Armengol et al., 2003; Locke et al., 2003b; Feuk et al., 2005).

About 5.3 % of the human genome is duplicated, representing ~150.8-Mb (She et al, 2004; IHGSC, 2001). Various hypothesis about the mechanism by which genomic

DNA is duplicated exist. As shown in Figure 1.3A, “retrotransposition” describes the integration of reverse transcribed RNAs at random sites in the genome (Hurles, 2004). Another mechanism is “unequal crossing over” due to homologous recombination between paralogous sequences, which could give rise to a tandem duplication of a genomic segment (Figure 1.3B).

Recent studies have shown that the frequent occurrence of segmental duplications during primate evolution has been partly due to a burst of activity of *Alu* retroelements (Bailey et al., 2003).

Alu repeats are a subclass of the SINEs (Short INterspersed Elements) transposable elements. They are about 300-bp in size and are the most frequent repeats found in the human genome (10^6 copies). They encompass about 10 % of the entire human genome. The striking enrichment of *Alu* repeats at the borders of duplicated sequences suggests unequal crossing over between these repeats in the generation of segmental duplications.

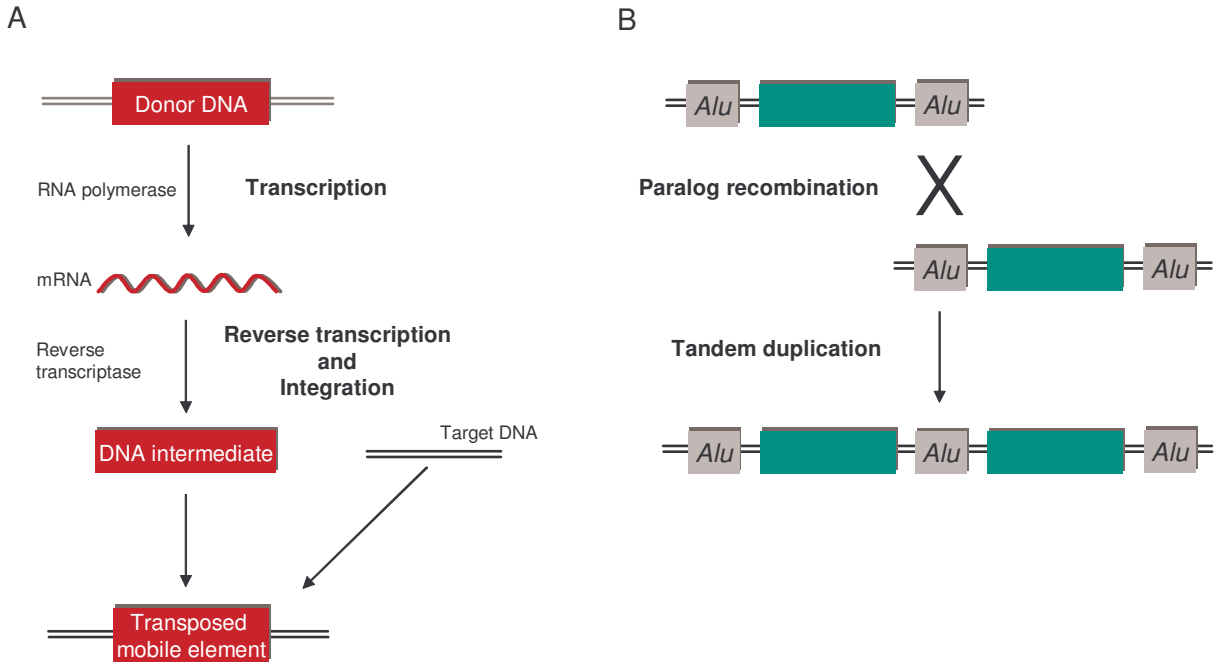


Figure 1.3: Schematic representation of two mechanisms mediating segmental duplications. (A) During retrotransposition, at least two enzymatic activities are used. One is an integrase, which is an endonuclease that cleaves at the site of integration. The other is an RNA-dependent DNA polymerase, also called reverse transcriptase. The RNA transcript of the donor site is copied into cDNA. That cDNA copy of the retrotransposon is then inserted into the integration site. (**B**) Model of *Alu-Alu* mediated duplication. The duplication (green) is due to unequal crossing over that occurred intrachromosomaly between *Alu* elements.

These are two examples of mechanism thought to be responsible for duplications. It should be kept in mind that further models of the spread of segmental duplications exist such as nonhomologous end-joining (NHEJ) (Bailey et al., 2003).
However, no mechanism can fully explain the frequency, the organization, and the distribution of all the segmental duplications in the human genome.

1.2.4.1. Gene duplication: a driving force of evolutionary change

An important line of thinking about why duplicated genes might arise goes back 30 years to Susumo Ohno. He reasoned that gene duplications allow new gene functions to evolve. One copy of a gene can carry out the original function while the duplicon becomes free to accumulate mutations, possibly developing new functions and allowing the big steps in evolution to occur (Ohno, 1970). With the advent of whole genome sequences, it has been shown that duplicate genes are abundant in most genomes such as bacteria, *Arabidopsis*, rice, fly, worm, and primates (Ochman et al., 2000; Vision et al., 2000; Samonte and Eichler, 2002; Yu et al., 2005).
The primary evidence that duplications have played a role in the evolution of new gene functions is the existence of gene families (Nahon, 2003). Members of a gene family that share a common ancestor as a result of duplication events followed by positive selection, are called paralogous, distinguishing them from orthologous genes in different genomes, which share a common ancestor as a result of a speciation event (Hurles, 2004) (Figure 1.4).
Paralogous genes can often be found clustered within a genome. For example, the *morpheus* gene family is the result of the recent proliferation, transposition and selection of a 20-kb duplicated segment throughout 15-Mb of the short arm of human chromosome 16 (Johnson et al., 2001).
Recently, the RGP gene family that arose after the separation of the primates from the rodents was described (Ciccarelli et al., 2005). The eight genes composing that family originated from the highly conserved nucleoporin RanBP2 by numerous complex genomic rearrangements such as inversions, translocations, exon losses, and duplications. The eight segments containing the *RanBP2*-related genes are spread over human chromosome 2. Two of the gene copies are located in the pericentromeric region of the short arm and the other seven on the long arm.

Although all of the eight copies are highly similar to RanBP2 (> 95 % of homology), the gene structure is modified, suggesting a novel function for the RGP family.

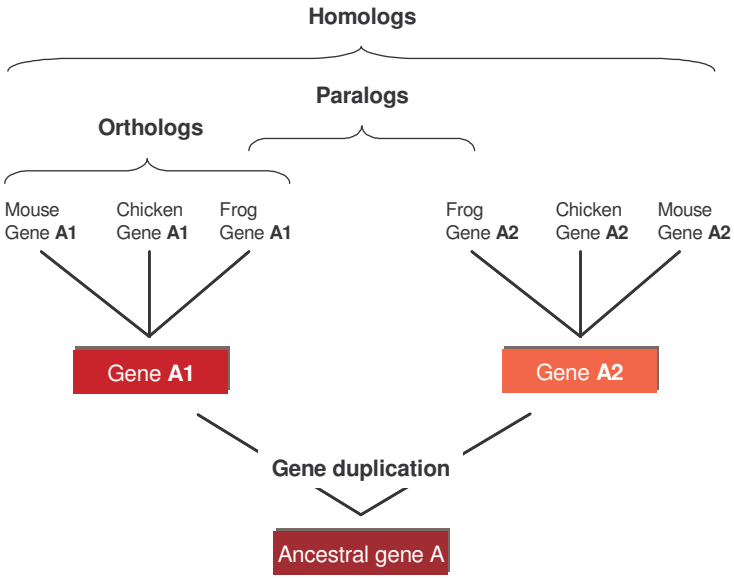


Figure 1.4: Orthologs and paralogs are two types of homologous sequences. Orthology describes genes in different species that derive from a common ancestor. Paralogy describes homologous genes within a single species that diverged by gene duplication (scheme adapted from the NCBI Education web page: “The concepts of Orthology and Paralogy”).

Dispersed paralogous genes within the genome are also common. A major example are the olfactory receptor (OR) genes which represent the largest gene superfamily in the human genome of about 1000 genes and pseudogenes (Glusman et al., 2001). ORs constitute 17 gene families, four of which contain more than 100 members each. They are localized on all the chromosomes except 20 and Y, and about 80 % are found in about two dozen clusters ranging from 6 to 138 genes and encompassing about 30-Mb (~1 %) of the human genome (International Human Genome Sequencing Consortium, 2001).

Another example of a gene superfamily is the very large set of genes that contain immunoglobulin-like (IG) domains and function as cell surface protein or receptors involved in immune function or other aspects of cell-cell interactions. There are dispersed gene and gene families, small clusters, large clusters, and clusters within clusters. Most IG superfamily proteins in invertebrates are cell-surface proteins. In vertebrates, the IG repertoire includes immune functions. The large expansion of IG

domains in vertebrates shows the versatility of a single gene family in responding rapidly and effectively to infection (International Human Genome Sequencing Consortium, 2001).

Given the apparent importance of gene duplication for the evolution of new biological functions, it is of great interest to list the differences due to duplication that exist between our genome and the one of our closest relatives, the great apes. Using interspecies cDNA array-based comparative genomic hybridization, Fortna et al. (2004) identified ~3 % of the 29,000 genes represented on their array as having undergone lineage-specific copy number changes among human and the great apes (bonobo, common chimpanzee, gorilla and orangutan). They observed that genes showing copy number expansions were most pronounced in human (134 genes) and include a number of genes thought to be involved in the structure and function of the brain, as *ROCK1, USP10, ARHGEF5,* etc... Another study has shown the non-random distribution of segmental duplications within the human proteome. Indeed, in 2002, Bailey et al. (2002a) observed that genes associated with immunity and defence (natural killer receptors, defensins, interferons, serine proteases, cytokines), membrane surface interactions (galectins, HLA, lipocalins, carcinoembryonic antigens), drug detoxification (cytochrom P450), and growth/development (somatotropins, chorionic gonadotropins, pregnancy-specific glycoproteins) were particularly enriched in the human genome. Those segmental duplications having arisen during the last 40 million years, the genes embedded in those segments may be considered excellent candidates for adaptation specific to primate evolution.

1.2.4.2. Highly homologous duplications: a source of instability

About 4.0 % of the human genome has been duplicated in recent time (Zhang et al., 2005). Because of their repeat nature, duplications, also called low copy repeats (LCRs), have been a real problem for the Human Genome Project and many of the remaining gaps are associated with LCRs (Eichler, 2001; Cheung et al., 2003a). The distribution of segmental duplications among the chromosomes seems non-uniform, with some chromosomes showing an unusual enrichment for these segments (Figure 1.5). For example, about 25 % of the euchromatin of the Y chromosome is composed of segmental duplications, including blocks as large as ~1.45-Mb with sequence identity of ~99.97 % (Skaletsky et al., 2003).

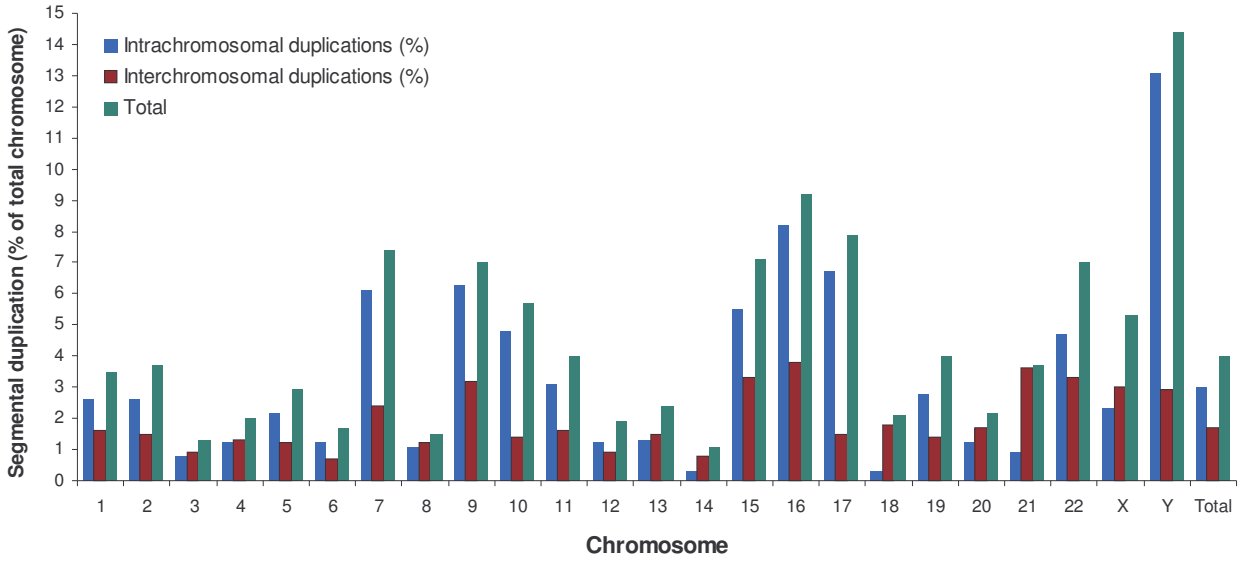


Figure 1.5: Percentage of segmental duplications by chromosome. A distinction was made between inter- (the red bars) and intrachromosomal (the blue bars) duplications ≥ 1-kb and sequence identity ≥ 90 %. The green bars show the ratio of the total number of duplicated bases to the length of the respective chromosome. Figure adapted from the results of Zhang et al. (2005).

The distribution of segmental duplications shows great variation among chromosomal regions. On average, pericentromeric and subtelomeric regions have 2.9-fold and 4.1-fold duplication enrichment, respectively (Bailey et al., 2001; International Human Genome Sequence Consortium, 2001; Zhang et al., 2005). Detailed analyses of pericentromeric regions have revealed that these regions of the genome are often enriched in segmental duplications (Horvath et al., 2000, 2001, 2005; Jackson, 2003). Recently, it has been shown that human subtelomeres are also patchworks of segmental duplications. The amplification of such homologous sequences arose at extremely high frequency during primate evolution through NHEJ (Linardopoulou et al., 2005).

Highly homologous segmental duplications are also found enriched at chromosomal rearrangement breakpoints, suggesting their implications in driving genomic rearrangements (Shaw et al., 2004). Misalignments, followed by recombination between non-allelic segmental duplications, also called non-allelic homologous recombination (NAHR), has been proposed to give rise to many genomic disorders

such as deletions, duplications, translocations and inversions, some of which are listed in Table 1.2.

Table 1.2: Genomic disorders mediated by segmental duplications.

Genomic disorder	Rearrangement	Chr. position	Size (Mb)	References
Charcot-Marie-Tooth disease type1A (CMT1A)	Interstitial duplication	17p12	1.5	Lupski (1998) Chance et al. (1994)
Hereditary neuropathy with pressure palsies (HNPP)	Deletion	17p12	1.5	Chance et al. (1994)
Smith-Magenis syndrome (SMS)	Deletion	17p11.2	5	Shaw et al. (2004)
Duplication 17p11.2	Interstitial duplication	17p11.2	5	Potocki et al. (2000)
Neurofibromatosis type1 (NF1)	Deletion	17p11.2	1.5	Jenne et al. (2003) Dorschner et al. (2000)
Prader-Willi syndrome (PWS)	Deletion	15q11-15q13	4	Pujana et al. (2002)
Angelman syndrome (AS)	Deletion	15q11-15q13	4	Amos-Landgraf et al. (1999)
Inverted duplication 15 (inv dup (15))	Supernumerary marker chromosome	15q11-15q14	4	Huang et al. (1997) Wandstrat et al. (1998)
Williams-Beuren syndrome (WBS)	Deletion	7q11.23	1.6	Perez-Jurado et al. (1998)
DiGeorge and velocardiofacial syndromes (DGS/VCFS)	Deletion	22q11.2	3	Edelmann et al. (1999)
Cat eye syndrome (CES)	Supernumerary marker chromosome	22q11.2	3	McTaggart et al. (1998)
Haemophilia A	Inversion	Xq28	0.5	Naylor et al. (1996)
Deletion of AZFc region	Deletion	Yq	3.5	Repping et al. (2002) Kuroda-Kawaguchi (2001)
Sotos Syndrome (SoS)	Microdeletion	5q35	1.9	Visser et al. (2005)

Segmental duplications have evolved recently in the primate lineage and their potential implication in evolutionary rearrangements is extensively studied (Bailey et al., 2002a; Cheung et al., 2003a; Zhang et al., 2005). Recent studies have indicated

that segmental duplications correlate with breaks of synteny between the human and the mouse genomes, suggesting a non-random distribution of low copy repeats and their involvement in evolutionary rearrangements (Armengol et al., 2003).
Several studies have highlighted the presence of LCRs at breakpoints of large-scale chromosomal rearrangements that distinguish human from primate chromosomes, suggesting their implication in the rearrangements and in genome evolution (Shaw and Lupski, 2004; Samonte and Eichler, 2002). For example, the telomeric fusion of two acrocentric chromosomes that gave rise to *HSA* 2 appears to have been mediated by a duplication-driven mechanism (Fan et al. 2002). Similarly, the breakpoints associated with the gorilla-specific translocation t(4;19) occurred within an LCR-rich region and has been shown to be associated with a gorilla-specific duplication of ~250-kb (Stankiewicz et al. 2001, 2004). In addition, the breakpoints of the pericentric inversion that distinguishes *HSA* 15 from its orthologous *PTR* counterpart map to segmental duplication clusters (Locke et al. 2003a). Moreover, recent studies have highlighted the presence of segmental duplications at breakpoints of micro-inversions (up to 4-Mb) that exist between human and chimpanzee genomes (Feuk et al., 2005; Szamalek et al., 2006).

1.2.4.3. The spreading of segmental duplications: an ongoing process

The spreading of segmental duplications occurred at different rate and at different time during evolution. Although the fraction of segmentally duplicated regions amount to only 1.2% and 2.9% of the mouse and rat genomes, respectively (Cheung et al. 2003; Tuzun et al. 2004), a tremendous increase of segmental duplications occurred during evolution in the primate lineage (Samonte and Eichler, 2001; Bailey et al., 2002a). Duplicated segments comprise about 5.3 % of the human genome (~150.8-Mb) while the chimpanzee genome is composed of about 5 % of duplicated segments (~136.7-Mb) (Bailey et al. 2001, 2002a, IHGSC, 2001; She et al., 2004; Cheng et al., 2005). Thus, the amplification of segmental duplications appears to be an ongoing process, which has occurred throughout primate evolution, with duplicated regions exhibiting significant evolutionary instability (Bailey et al. 2002b). Moreover, several studies have shown that large-scale DNA copy number polymorphisms (CNPs) exist within the human population. Many of these CNPs map

within regions rich in segmental duplications. Thus, it has been hypothesized that many CNPs may be due to the presence or absence of recent segmental duplication events that have not yet been fixed within the human genome. This provides evidence that the process of duplicative transposition is ongoing within the human population (Iafrate et al., 2004; Sebat et al., 2004; Sharp et al., 2005; Tuzun et al., 2005).

1.3. The array comparative genomic hybridization (aCGH) technique

Until recently, genome-wide experimental analyses at a submicroscopic level have been impossible due to the lack of efficient methods of comparing unsequenced genomes with accuracy and high level of resolution. With the completion of the human genome reference sequence, new approaches have emerged that automate and facilitate the simultaneous detection and high-resolution of mapping of large-scale DNA variation across the entire genome (Snijders et al., 2001). One such method is the array comparative genomic hybridization (aCGH) (Figure 1.6).
It has been demonstrated that arrays containing BAC clones have the ability to detect reliably DNA copy-number changes between genomic DNA samples with the resolution of a single BAC. Following the first report of array-CGH in 1997 (Solinas-Toldo et al., 1997), several applications and methodological advances have been reported in the literature. Genomic microarrays have been used to study the profiling of DNA dosage alterations, mainly in tumors (Albertson and Pinkel, 2003), to characterize normal and disease-related genomic variation (Iafrate et al., 2004; Sebat et al., 2004; Sharp et al., 2005), for epigenetics (Adorjan et al., 2002; Mantripragada et al., 2004) and for evolutionary analysis of human and ape genomes (Locke et al., 2003b). The use of cDNA clones as array-CGH targets for detection of DNA dosage changes has also been reported (Pollack et al., 1999).
To date, two genome-wide studies comparing the human and primates' genomes have been made. They give a clue to the ascertainment of the frequency, extent and distribution of large-scale sequence gain or loss events, defined as copy number altering events. The first one used BAC based aCGH and could detect DNA

sequences > 40-kb (Locke et al., 2003b). The second one used a cDNA-based aCGH (Fortna et al., 2004). The aCGH has proven in both studies to be a powerful approach to do such large-scale genome comparisons.

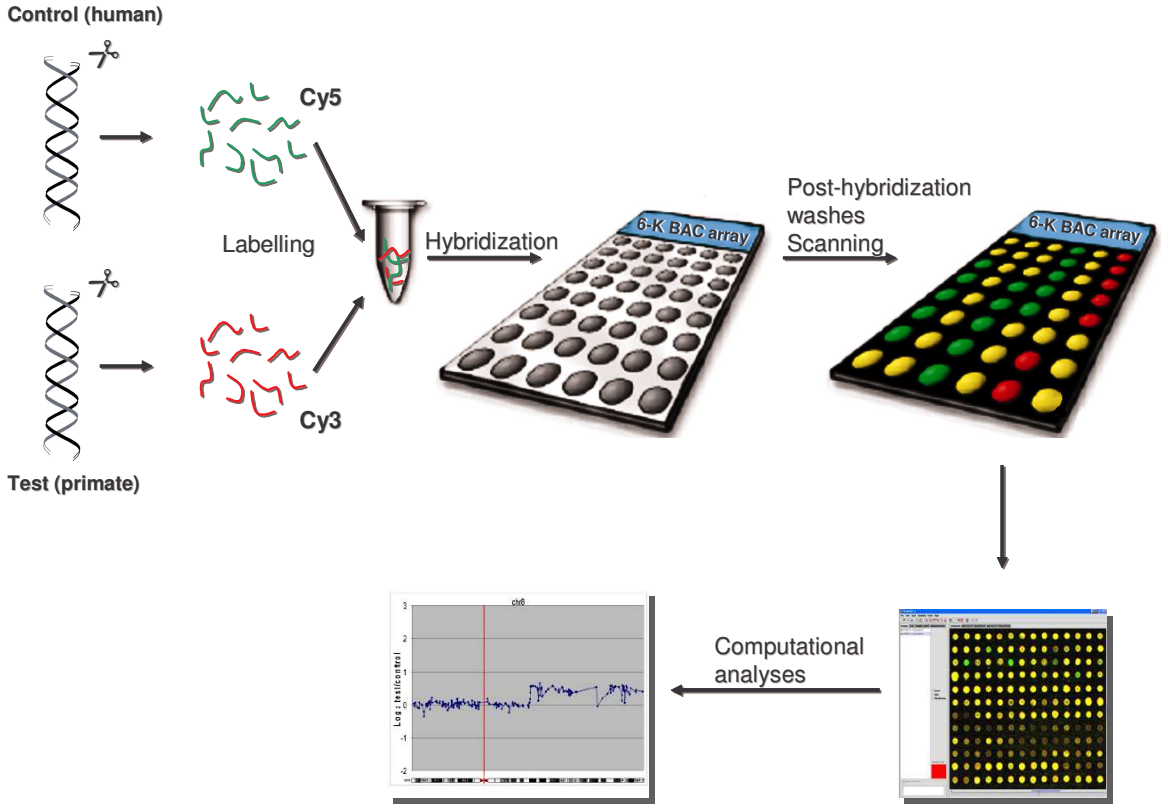


Figure 1.6: Schematic representation of the array CGH technique for a focussed analysis of copy number differences between human and primate genomes. Array CGH is performed by hybridizing prepared and labelled human (Cy5) and primate (Cy3) genomic DNA to the microarray. After hybridization for 24-72 hours, the slides are washed and scanned. Digital data are extracted from the scanned images using image analysis software and the ratios of fluorescence (Cy3 vs. Cy5) for each clone are calculated. The differences of DNA copy number between both species tested are represented by the colour of the spots. A yellow spot stands for a similar DNA copy number between both species. A green spot indicates a gain in the human genome (or a loss in the primate genome tested), while a red spot signifies a loss in the human genome (or a gain in the primate genome tested). The results are transferred in an Excel sheet, allowing their graphical representation.

1.4. Objective of the study

Recent studies demonstrated the evolutionary importance of copy number differences (such as deletions or duplications) that exist between human and primates genomes. However, the studies have been reduced for the most to human-chimpanzee comparisons (Cheng et al., 2005; Newman et al., 2005). Although Fortna et al. (2004) included other apes in their study, they focussed on gene copy-number differences. In their initial study, Locke et al. (2003b) included non-genic regions using an array bearing 2,460 human BAC clones. This clone-set encompassed ~12 % (370-Mb) of the human genome with a resolution of one BAC every 1.4-Mb of DNA sequence.

In this study, I have employed aCGH to extend these inter-specific comparisons of primate genomes using a 6-K array encompassing about 30 % of the human genome. The 6-K array includes 6,094 BAC clones (from the RPCI-11 library) that have been either completely sequenced or are connected to established human contigs via STS or end-sequences, and which have been annotated for gene content (Snijders et al., 2001; Cowell and Nowak, 2003). On average, the human 6-K BAC genomic array provides a resolution of one clone per 0.5-Mb of the human genomic sequence. My goal was to find out the importance of copy number differences in the genomes of macaque, orang-utan, gorilla, bonobo and of the common chimpanzee when compared to human. Even though such comparisons are of great interest, special attention was paid to those changes that occurred in the human genome, since these could have contributed to the development of some characteristics that are unique to the human species. Moreover, DNA copy number changes could have mediated large-scale rearrangements, such as inversions, which could have driven speciation. In order to verify the implication of LCRs in the occurrence of evolutionary rearrangements, I have characterized the breakpoints, at the base-pair level, of two pericentric inversions reported by Yunis and Prakash (1982). I have studied the duplication content of the areas around the breakpoints of the inversions that occurred in the gorilla and chimpanzee chromosomes XVI and of the pericentric inversion of *HSA* 18.

2. MATERIALS AND METHODS

2.1. Materials

2.1.1. Oligonucleotides

2.1.1.1. Oligonucleotides used for PCR

Table 2.1: List of the primers used for PCR to characterize the breakpoints of the inversions on chromosome 16 and 18 and to verify the authenticity of gorilla's genomic DNAs (P23f-P23r).

Name of the primer	Sequence (5' > 3')	Position on the respective BAC (bp)
		RP11-696P19
P1f	TACCCATTAGCATCGCCTTC	104781
P1r	CTGCTCCACGGCTTAAATTC	104962
P2f	GCAGTTGTCATCGAATGGAA	114667
P2r	TCGATGACCAATTTCTGGAAG	114943
P3f	ACACCTCAGAGTTCCGGTTG	115032
P3r	GCCTTGCATCACATTTCAGA	115224
P8f	GCTCTGGATTCCCTCAAAAA	71842
P8r	CAGCATCCCCATCTCTTTGT	72044
P9f	GGAAAACAAGGGGCCTAAAA	119256
P9r	TCTATCCAGGGCAGGAACAG	119451
		CTD-2144E22
P4f	AGAAAGGCAAGGGGTTTCAT	139229
P4r	CAGGCAGAGTCACAGTGGAA	139405
P5f	CAGTGAGCAAGCCAAATTCA	144168
P5r	TCCCCACCGGACTGTTATAG	144360
P6f	TACTCACCCTCCCTCATTGC	162183
P6r	CTCCATTACGGAGCACA	162365
P7f	TGAGAGCAACTGAGGTTTGC	166997
P7r	GCCAGGTATGCGATGAAAAT	167170
		CTD-2522B17
P10f	ACGGCTTGGTTTTCACTGTT	56162
P10r	CTCCCCCATAAGAACGATCA	56364
		RP11-683L23
P11f	GGAAGCCTTTTCCTCACCTT	57331
P11r	AAGGCGCTGATTTTCAGAGA	57658
P12f	CAGAGCTGTCCCTGAACCTC	63502
P12r	GACCACCTGGTGTGACCTCT	63924
P13f	GAAAACTGGACCCAGATGGA	79964

Continuation of Table 2.1

P13r	GCGTGGAATTGACACAGTTG	80310
P14f	CAGCAACCCACACTGAGAGA	85372
P14r	GCACCCCTTTGATTCACTGT	85785
P15f	GCATTTAGCTTTCTGGAACCTG	88905
P15r	CAGTTTCTCCTCATGGAAAGG	89118
P16f	AGTTCTCCGCTGTGTGGAGT	96290
P16r	GCTCTGCTGGAGGAGCTTTA	96492
P17f	AGGTCAAAGAGGCCTATAATATGGT	97555
P17r	AAGGTCCTTAATGATGGAATTTACTC	97687
P18f	ATTCAAAGTCAAGTCTGGTAACTGC	97733
P18r	TTTAAATTTTGACAACTAAAGCTGTTT	97926
P19f	TGCTGCTGTTAGCATGTTCTC	102428
P19r	AAAAATGATGCCCTCTGTGG	102736
P20f	AATGCTCAGCTTGAAAAGGG	126541
P20r	GGGTCTCCTGCTCCATGTT	126841
		RP11-705O1
P21f	TCCAGGGAACTCAAGAATGG	7508
P21r	GCAAGGAGAGAGGTGGTTTG	7846
P22f	CATCCTCATGTGCCTTGAAA	73238
P22r	TGGGTTCATGCACCAGATAA	73647
		CH255-73E2
P23f	GGTGAGACTTGCCCAGTGTT	
P23r	CAGATGGTCCTTCAGCCTCT	

2.1.1.2. Oligonucleotides used for Real Time PCR

Table 2.2: List of the primers used for Real Time PCR (Light Cycler analysis) to study variation of gene expression among the apes and to verify the results obtained by aCGH.

Name of the primer	Sequence (5' > 3')
LOC401848f	GCGGATCCAGAAGGACATAA
LOC401848r	GTTGCCCTCGATTTCAATGT
FLJ43980f	TGCAAGAGAAACGTGAAACA
FLJ43980r	CTGCTATCTCCTGCCTTGCT
SHCBP1f	TTGGAGGAGTTGGGATGAAG
SHCBP1r	TCCTGCTCGGAATCAGAGTT
MGC34800af	TCACCACTAAGGGCAACCTC
MGC34800ar	GGTCGCCATTCATTATTTGG
MGC34800bf	GGATAGGAGTCGGGAAGAGC
MGC34800br	ACAGCTCATGCAGCAGATCC
MGC34800cf	CTTTCAGCCTGCAGTTCCTC
MGC34800cr	ATTCCACATGTGAGTGCCAA
HPRTf	GCTGGTGAAAAGGACCTCT
HPRTr	CACAGGACTAGAACACCTGC

Continuation of Table 2.2

FMRPD2f	CGTGCAGATGGAGAAAGAGA
FMRPD2r	GGCCAGGATAATGTCACCAG
ANAPC1f	GAGTTTTTCAGCTCCCGAAG
ANAPC1r	AGCCACTGGTCCAGGGTATTA
FOXP2f	TTACCTCAAACCCCTACCACAC
FOXP2r	GTCTGAATGTCGCCTTCGTATG

2.1.2. DNA samples

Table 2.3: List of the DNA samples used for array Comparative Genomic Hybridization (aCGH)

DNA#	Species	Gender	Origin	Tissue
181	*Macaca fuscata*	Female	16,3/92 (zoo Wilhelma)	Blood
182	*Macaca fuscata*	Female	4;3/92 (zoo Wilhelma)	Blood
189	*Macaca fuscata*	Female	W9/1;2/92 (zoo Wilhelma)	Blood
190	*Macaca fuscata*	Female	15;2/92 (zoo Wilhelma)	Blood
191	*Macaca fuscata*	Female	3;2/92 (zoo Wilhelma)	Blood
279	*Macaca fuscata*	Female	Nr.10 (zoo Wilhelma)	Blood
280	*Macaca fuscata*	Female	W55 (zoo Wilhelma)	Blood
483	*Macaca fuscata*	Female	Nr.14 (zoo Wilhelma)	Blood
484	*Macaca fuscata*	Female	Nr.7 (zoo Wilhelma)	Blood
485	*Macaca fuscata*	Female	W9.2 (zoo Wilhelma)	Blood
49	*Pongo pygmaeus*	Female	Elsi (Basel, zoo)	Blood
50	*Pongo pygmaeus*	Female	Atjeh (Basel, zoo)	Blood
596	*Pongo pygmaeus*	Female	Katai (zoo Wilhelma)	Blood
606	*Pongo pygmaeus*	Female	Conny (zoo Wilhelma)	Blood
44	*Pongo pygmaeus*	Female	Kwesida (zoo Wilhelma)	Blood
648	*Pongo pygmaeus*	Female	Moni (zoo Wilhelma)	Blood
78	*Pongo pygmaeus*	Female	Dunja (Leipzig, zoo)	Blood
404	*Pongo pygmaeus*	Female	Vicki (Nürnberg, zoo)	Blood
735	*Pongo pygmaeus*	Female	Sirih (Zürich, zoo)	Blood
487	*Pongo pygmaeus*	Female	Suma (Nürnberg, zoo)	Blood
56	*Gorilla gorilla*	Female	Mosseka (zoo Wilhelma)	Blood
102	*Gorilla gorilla*	Female	Ndijo; Atlanta USA	Blood
38	*Gorilla gorilla*	Female	Momo (zoo Wilhelma)	Blood
330	*Gorilla gorilla*	Female	Dafina (zoo Wilhelma)	Blood
338	*Gorilla gorilla*	Female	Sandra (Rotterdam, zoo)	Blood
355	*Gorilla gorilla*	Female	Mimi (zoo Wilhelma)	Blood
417	*Gorilla gorilla*	Female	Neema (zoo Wilhelma)	Blood
425	*Gorilla gorilla*	Female	Dura (zoo Wilhelma)	Blood
461	*Gorilla gorilla*	Female	Delphi (Nürnberg, zoo)	Blood
475	*Gorilla gorilla*	Female	Virunga (Leipzig, zoo)	Blood

Continuation of Table 2.3

828	*Pan troglodytes*	Female	Exota (zoo Wilhelma)	Blood
8	*Pan troglodytes*	Female	Chita II (zoo Wilhelma)	Blood
46	*Pan troglodytes*	Female	Mixi (Neunkirchen, zoo)	Blood
90	*Pan troglodytes*	Female	Luise (Welzheim, zoo)	Blood
96	*Pan troglodytes*	Female	Charlotte (Welzheim, zoo)	Blood
133	*Pan troglodytes*	Female	Katache (Neunkirchen, zoo)	Blood
519	*Pan troglodytes*	Female	Babsy (zoo Wilhelma)	Blood
582	*Pan troglodytes*	Female	Epona (Neunkirchen, zoo)	Blood
694	*Pan troglodytes*	Female	Josi (Düsseldorf, zoo)	Blood
835	*Pan troglodytes*	Female	Susi (zoo Wilhelma)	Blood
2	*Pan paniscus*	Female	Kamiti (zoo Wilhelma)	Blood
47	*Pan paniscus*	Female	Eja (zoo Wilhelma)	Blood
97	*Pan paniscus*	Female	Lina (San Diego, zoo)	Blood
168	*Pan paniscus*	Female	Catherine (zoo Wilhelma)	Fibroblast
556	*Pan paniscus*	Female	Kosana (Antwerpen, zoo)	Blood
295	*Pan paniscus*	Female	Lisala (zoo Wilhelma)	Blood
784	*Pan paniscus*	Female	Unga (Antwerpen, zoo)	Blood
409	*Pan paniscus*	Female	Kombote (zoo Wilhelma)	Blood
447	*Pan paniscus*	Female	Diatou (zoo Wilhelma)	Blood
554	*Pan paniscus*	Female	Hortense (Antwerpen, zoo)	Blood
/	*Homo sapiens*	Female	Marlies (german)	Blood
/	*Homo sapiens*	Female	Nicole (german)	Blood
/	*Homo sapiens*	Female	Susan (german)	Blood
/	*Homo sapiens*	Female	Elke (german)	Blood
113	*Homo sapiens*	Female	(german)	Blood
110	*Homo sapiens*	Female	(german)	Blood
9/9	*Homo sapiens*	Female	(german)	Blood
9/5	*Homo sapiens*	Female	(german)	Blood
9/2	*Homo sapiens*	Female	(german)	Blood
59	*Homo sapiens*	Female	(german)	Blood

2.1.3. Cell lines

Table 2.4: List of the cell lines used for Fluorescence In Situ Hybridization (FISH) and array Comparative Genomic Hybridization (aCGH)

Cell #	Species	Type of cell	Origin
91012416	*Pan troglodytes*	Fibroblast	ECACC
89072705	*Pongo pygmaeus*	Lymphoblast	ECACC
89072703	*Gorilla gorilla*	Lymphoblast	ECACC
/	*Pan paniscus*	Lymphoblast	University of Mainz, Germany
GM03446	*Macaca fascicularis*	Lymphoblast	Coriell Cell Repository

2.1.4. BAC clones

Table 2.5: List of BAC clones used for FISH and PCR

BAC clone	Accession number	Species	Chromosomal position
RP11-426C22	AC009093	*Homo sapiens*	16p11.2
RP11-341P6	AC139476	*Homo sapiens*	16p11.2
RP11-347N4	AC141283	*Homo sapiens*	16p11.2
RP11-598D12	AC133561	*Homo sapiens*	16p11.2
CTD-2522B17	AC136932	*Homo sapiens*	16p11.2
CTD-2144E22	AC135776	*Homo sapiens*	16p11.2
RP11-20D16	AC092133	*Homo sapiens*	16p11.2
RP11-14K3	AC092325	*Homo sapiens*	16p11.2
RP11-65B9	AC109490	*Homo sapiens*	16p11.2
RP11-696P19	AC106819	*Homo sapiens*	16q11.2
RP11-46D6	AC092368	*Homo sapiens*	16q11.2
RP11-93O14	AC012186	*Homo sapiens*	16q11.2
RP11-523L20	AC007857	*Homo sapiens*	16q12.1
RP11-44I10	AC026470	*Homo sapiens*	16q12.1
RP11-683L23	AP001005	*Homo sapiens*	18p11.32
RP11-705O1	AP000845	*Homo sapiens*	18p11.32
RP11-720L2	AP000915	*Homo sapiens*	18p11.32
RP11-476K15	AC008109	*Homo sapiens*	18p11.32
RP11-291G24	AC019183	*Homo sapiens*	18p11.32
RP11-161I6	AP005481	*Homo sapiens*	18p11.32
RP11-870F11	AP000945	*Homo sapiens*	18p11.32
RP11-419P8	AP005061	*Homo sapiens*	18p11.32
RP11-183C12	AP005403	*Homo sapiens*	18p11.31
RP11-865B13	AP001166	*Homo sapiens*	18p11.31
RP11-674N23	AP000864	*Homo sapiens*	18p11.22
RP11-820I16	AP001269	*Homo sapiens*	18p11.21
RP11-666N19	AC036178	*Homo sapiens*	18q11.2
RP11-254G11	AC022795	*Homo sapiens*	18q11.2
RP11-326M20	AC019306	*Homo sapiens*	18q11.2
RP11-784B15	AC117569	*Homo sapiens*	18q12.1
RP11-489I20	AC022174	*Homo sapiens*	18q12.1
RP11-549B18	AC017100	*Homo sapiens*	18q12.1
RP11-82D16	CG734216-CG734215	*Homo sapiens*	1p36.33
RP11-708D7	AQ613607-AQ432767	*Homo sapiens*	2q11.1
RP11-95L24	AQ316687-AQ316684	*Homo sapiens*	3p12.3
RP11-101N17	AQ320386-AQ320389	*Homo sapiens*	4q26
RP11-143K14	AQ389853-AQ389851	*Homo sapiens*	5q11.2
RP11-140K14	AQ386725-AQ386711	*Homo sapiens*	8p23.2
RP11-151D14	AQ379180-AQ379200	*Homo sapiens*	9q21.13
RP11-91A4	AQ283463-AQ283466	*Homo sapiens*	14q21.2
RP11-795L14	AQ492777-AQ468013	*Homo sapiens*	21p11.2-p11.1
RP11-402P6	AZ254497-AZ254498	*Homo sapiens*	Xq13.1

Continuation of Table 2.5

RP11-439A17	AL357493	*Homo sapiens*	1p11.2
RP11-35B4	AL359093	*Homo sapiens*	1q21.1
RP11-91G11	AL691471	*Homo sapiens*	1q21.1
RP11-89F3	BX546486	*Homo sapiens*	1q21.1
RP11-196G18	AL591493	*Homo sapiens*	1q21.1
RP11-68E19	AC017002	*Homo sapiens*	2q13
RP11-397E13	AC069171	*Homo sapiens*	2q13
RP11-79A9	AQ282179-AQ282181	*Homo sapiens*	9q12-13
RP11-81J3	AQ281651-AQ281654	*Homo sapiens*	10q11.1
RP11-80B18	AQ283261-AQ283264	*Homo sapiens*	10q11.22
RP11-760D22	AQ451723-AQ452912	*Homo sapiens*	14q11.2
RP11-499D5	AC007908	*Homo sapiens*	16p11.2
RP11-915P6	AQ563767-AQ622907	*Homo sapiens*	22q11.1
RP11-81H21	AZ521616-AZ521618	*Homo sapiens*	22q11.1
RP43-007E19	AG145870-AG145869	*Pan troglodytes*	
RP43-001I03	AG141422-AG141421	*Pan troglodytes*	
RP43-056O01	AG183403-AG183404	*Pan troglodytes*	
CH255-551C4	AY822675	*Gorilla gorilla*	
CH255-39D12	AY822676-AY822677	*Gorilla gorilla*	

2.1.5. Chemicals and reagents

2-β-Mercaptoethanol	Roche, Mannheim
^{32}P-dCTP	Amersham, Braunschweig
Agarose	Roth, Karlsruhe
Ammoniumacetat	Merck, Darmstadt
6K human BAC-array	Roswell Park Cancer Institute, Buffalo
Betaine	Sigma, München
Biotin-16-dUTP	Roche, Mannheim
Boric Acid	Merck
Bromophenol	Sigma, München
BSA	Sigma, München
Chloramphenicol	Sigma, München
Cot-DNA	Invitrogen
CHORI 255 BAC-Library	Children's Hospital Oakland Research Institute, Oakland

Dextransulfate	Pharmacia, Freiburg
Digoxigenin-11-dUTP	Roche, Mannheim
DMEM	Biochrom, Berlin
1kb PLUS DNA ladder	Invitrogen
dNTP	Fermentas
EDTA	Serva, Heidelberg
Ethanol 100%	Merck, Darmstadt
Ethidiumbromide	Sigma, München
FCS	Roche, Mannheim
Fixogum	Marabuwerke, Tamm
Formaldehyde	AppliChem, Darmstadt
Formamid	Fluka, Neu-Ulm
Gentamycin	Biochrom AG, Berlin
Glycerin	Sigma, München
Glycerol	Sigma, München
HCl, 1M	Merck, Darmstadt
Isopropanol	Merck, Darmstadt
Kanamycin	Sigma, München
KCl	Merck, Darmstadt
L-Glutamin	Sigma, München
$MgCl_2$	Roche, Mannheim
NaAc	Merck, Darmstadt
Na_3Citrate	Merck, Darmstadt
NaCl	Merck, Darmstadt
NaPyruvate	Sigma, München
PBS	Biochrom, Berlin
Pepsin	Biochrom, Berlin
RPMI	Gibco BRL, Eggenstein
SSC	Merck, Darmstadt
Tris-HCl	Serva, Heidelberg
Tryptone	Bector, Dickinson and Co, USA
Tween 20	Sigma, München
Vectashield	Vector, Burlingame, USA
Yeast extract	Roth, Karlsruhe

Xylencyanol	Sigma, München

2.1.6. Enzymes

DNA Polymerase I	Roche, Mannheim
DNase I	Boehringer, Mannheim
RNase A	Sigma, Munich
Proteinase K	Sigma, Munich
Taq DNA Polymerase	Qiagen, Hilden

2.1.7. Antibodies

Anti-Digoxigenin	Roche, Mannheim
Anti-Mouse Texas Red	Dianova, Hamburg
Biotin-16-dUTP	Boehringer, Mannheim
Anti-Avidin biotinylated	Vector, Burlingame, USA
FITC-Avidin	Vector, Burlingame, USA
Anti-Rabbit Texas Red	Dianova, Hamburg

2.1.8. Equipment

ABI PRISM 7700 Sequence Detection System	Applied Biosystems, USA
ISIS (v 5.0)	Metasystems, Germany
IKAROS (v 5.0)	Metasystems, Germany
CO_2 incubator B 6220 CU	Heraeus, Germany
Electrophorese Power PAC 300	Bio Rad, Germany
PTC100 Programmable Thermal Controller (PCR machine)	MJ Research Inc., USA
AxioPlan microscope	Zeiss, Germany
AxioCam MRm camera	Zeiss, Germany
GeneQuant, RNA/DNA calculator	Pharmacia, UK

LightCycler 1.5 Instrument	Roche, Mannheim

2.1.9. Kits

DNeasy Tissue Kit	QIAGEN, Germany
HiSpeed Plasmid Midi Kit	QIAGEN, Germany
GFX PCR DNA and Gel Band Purification Kit	Amersham, UK
ABI PRISM Big Dye Terminator Cycle Sequencing Ready Reaction Kit	Applied Biosystems, USA
RNeasy Kit	QIAGEN, Germany
Superscript First Strand Synthesis System for RT-PCR	Invitrogen, Germany
Quantitect SYBR Green PCR Kit	QIAGEN, Germany

2.2. Methods

2.2.1. DNA technology

2.2.1.1. DNA purification from cell lines

In order to obtain optimum DNA yield and quality, the amount of starting material for the purification of DNA was 4 to 5 x 10^6 cells.
DNA was extracted from cell lines and was purified using the DNeasy Tissue Kit (QIAGEN), according to the DNeasy protocol for Cultured Animal Cells. To ensure that the genomic DNA extracted was RNA-free, after resuspension of the pellet in PBS, 4 µl of RNase A (100 mg/ml) was added.

2.2.1.2. Determination of the DNA concentration

The DNA solutions were diluted to 1:10 (readings should be taken at the wavelengths of 260 nm and 280 nm). The reading at 260 nm allowed the calculation of the concentration of nucleic acids in the samples.
The concentration of DNA was calculated as follows:
[DNA] =OD260 x dilution factor x 50 µg/ml
The ratio between the reading at 260 nm and 280 nm (OD260/ OD280) provides an estimate of the purity of the nucleic acid. Pure preparations of DNA have OD260/ OD280 values of 1.8. If there is a contamination with protein, the OD260/ OD280 will be significantly less than the value given above. If there is contamination with RNA, the OD260/OD280 is close to 2.0.

2.2.1.3. DNA preparation from BAC clones

Chloramphenicol solution (34 mg/ml)

680 mg Chloramphenicol

Add ddH_2O to 20 ml, sterilize by filtration, store at -20°C

Chloramphenicol stock solution (34 mg/ml)

680 mg Chloramphenicol

Add glycerol to 20 ml, vortex and store at -20°C

Kanamycin solution (30 mg/ml)

600 mg Kanamycin

Add ddH2O to 20 ml, sterilize by filtration, store at -20°C

Kanamycin stock solution (30 mg/ml)

600 mg Kanamycin

Add glycerol to 20 ml, vortex and store at -20°C

LB-Agar

10 gr Tryptone

5 gr Yeast extract

10 gr NaCl

20 gr Agarose

Adjust pH to 7.0 with 5 N NaOH

Add ddH_2O to final volume of 1 L, sterilize by autoclaving, and pour into Petri dishes (~25 ml/100-mm plate)

LB-Medium

10 gr Tryptone

5 gr Yeast extract

10 gr NaCl

Add ddH_2O to final volume of 1 L, sterilize by autoclaving

Store at 4°C for several months

Tris

1.211 gr Tris

Add ddH_2O to volume of 850 ml and adjust pH to 8.0

Add ddH_2O to final volume of 1000 ml, sterilize by autoclaving

- **Striking of bacteria on agar plate**

Clones derived from libraries constructed in the PAC vectors (pCYPAC2 or pPAC4) grow on medium with 25 µg/ml Kanamycin. These libraries are RPCI-1, 3, 4, 5, 6, 21, 31, 91 and 92.

Clones derived from all other libraries are constructed in the BAC vectors (pBACe3.6 or pTARBAC2), and grow on medium with 20 µg/ml Chloramphenicol.

The plates were placed in the 37°C room, overnight.

- **Preparation of bacterial culture**

The plates were taken out from the 37°C room the day after.

One colony was picked from the plate and let grow in 5 ml of liquid medium with the right antibiotic. The tubes were placed in the 37°C room overnight, in the shaker.

The next day, one drop of each culture was poured into a stock solution containing the right antibiotic and frozen at -70°C. The rest of the cultures were poured in bottles containing 200 ml of medium and placed on the shaker in the 37°C room, overnight.

- **Preparation of DNA**

The bottles were taken out from the 37°C room. The cultures were transferred into smaller plastic bottles for centrifugation during 15 min at 5000 rpm. After centrifugation, the medium was discarded. The DNA was prepared using the HiSpeed Plasmid Midi Kit (QIAGEN), according to the manufacture's instruction. The elution step was made with 500 µl of Tris.

2.2.1.4. DNA precipitation with ethanol

TE

10 mM	Tris
1 mM	EDTA
Adjust the pH to 8.3	

Most nucleic acids may be precipitated by addition of monovalent cations, recovered by centrifugation and dissolved in an appropriate buffer at a desired concentration. 2.5 volumes of ethanol and 1/10 volume 3 M sodium acetate were added to the DNA sample contained in a 1.5 ml microcentrifuge tube. The mixture was inverted several times and incubated at -75°C for 20 min. The mixture was centrifuged at 4°C, at full speed for 15 minutes. The supernatant was removed using a double tip. 70% ethanol (corresponding to about two volumes of the original sample) was added, then the mixture was incubated at room temperature for 5-10 minutes and finally it was centrifuged again for 10 minutes. The supernatant was removed as previously. The DNA pellet was dried and dissolved in TE.

2.2.1.5. Agarose gel electrophoresis

Ladder

100 µl	ddH_2O
100 µl	TE Buffer
25 µl	Loading Buffer
10 µl	1 Kb Ladder

Loading Buffer

11.5 ml	Glycerin
50 mg	Bromophenol
50 mg	Xylencyanol
400 µl	EDTA (10mM)
20 ml	ddH_2O

TBE Buffer

89 mM	Tris
89 mM	Boric Acid
2 mM	EDTA

Add ddH2O to final volume of 1 L

Adjust pH to 8.2

Agarose gel electrophoresis is employed for example to check the progression of a restriction enzyme digestion, as well as to determine quickly the yield and purity of DNA isolations or PCR reactions. Electrophoresis is used to separate molecules based on their size and charge. DNA has a negative charge in an appropriate buffer solution, so it migrates to the positive pole in an electric field. In an agarose gel electrophoresis, the DNA is forced to move through a a gel matrix, that is made of agarose, acting as a sieve. The result is that the large pieces of DNA move slower than small pieces of DNA. Ethidium bromide is included in the gel matrix to enable fluorescent visualization of the DNA fragments under UV light.

Dried agarose was dissolved in the appropriate volume of 1 x TBE buffer by heating and 1/10000 volume of ethidium bromide (10 mg/ml) was added to the warm gel solution (60°C). Then the gel was poured into a mould, which was fitted with a well-forming comb. The agarose gel was submerged in electrophoresis buffer within a horizontal electrophoresis apparatus. The DNA samples were mixed with loading buffer and loaded into the sample wells. Size markers (ladder) were also loaded with DNA samples to aid in fragment size determination.

After electrophoresis, the gel was placed on an UV light box and pictures of the fluorescent ethidium bromide-stained DNA separation pattern were taken with a video camera.

2.2.1.6. Polymerase-Chain-Reaction (PCR)

dNTPs

Mix equal volumes of 100 mM stock solutions of each dNTP to a final concentration of 25 mM of each dNTP, dispense into aliquots and store at -20°C

P20 Buffer

50 ml	500 mM KCl
10 ml	100 mM Tris HCl (pH 8.3)
2 ml	20 mM $MgCl_2$

PCR enhancer buffer

6.76 gr	Betaine
Add 4 ml ddH_2O	
500 µl	Tris saturated (pH 9.0)
Add ddH_2O to final volume of 30 ml	

PCR is an *in vitro* technique for the amplification of a region of DNA that lies between two regions of known sequence (Mullis *et al.*, 1986). PCR amplification is achieved by using oligonucleotide primers complementary to the outer regions of the known sequence. The denatured strands of the large DNA fragment serve as the template. This results in the synthesis of new DNA strands, which are complementary to the parent template strands. The oligonucleotide directed synthesis of daughter DNA strands can be repeated. Template denaturation, primer annealing and primer extensions comprise a single "cycle" in the PCR amplification methodology.

The preparation of 25 µl PCR reaction was as following:

PCR enhancer buffer	15 µl
dNTPs	4 µl
Reverse primer	1.25 µl
Forward primer	1.25 µl
P20	2.5 µl
Taq Polymerase	0.15 µl
DNA template	1-2 µl

The tubes were placed in a thermal cycler preheated to 94°C. The amplification was processed as follow, denaturation at 94°C for 35 seconds, annealing at 45-65°C

(depending on the primers used) for 30 seconds, elongation at 72°C for 1 minute. After 34 cycles, the reaction was held at 72°C for 5 minutes to increase the yield of completely elongated products.

2.2.1.7. Elution of DNA fragments from agarose gels

The GFX PCR DNA and Gel Band Purification Kit (AMERSHAM, Biosciences) procedure removes primers, nucleotides, enzymes, mineral oil, salts, agarose, polyacrylamide, ethidium bromide, dyes, detergents and other impurities from DNA samples.
DNA fragments were purified using low-melting temperature agarose gels and the band of interest was excised with a scalpel under UV illumination. The gel slices were mixed with the capture buffer and then applied to the GFX columns where the DNA bound to the glass fiber matrix. The impurities were washed away and the pure DNA was eluted in low ionic strength buffer.
The purified DNA was ready for use in any subsequent application.

2.2.1.8. Sequencing

The sequencing consists of three main steps: PCR product purification, sequencing reaction, and running.
The purification of the PCR product was made using the Microcon Centrifugal Filter Units (Millipore) according to the manufacture's protocol.
The sequence reaction was performed with the ABI PRISM Big Dye terminator cycle sequencing ready reaction kit (PE Biosystems). In the ready reaction format, thermally stable AmpliTaq DNA polymerase, modified deoxynucleoside triphosphates (dNTP) and a set of dye terminators labelled with high-sensitivity dyes are provided. Once a sequence-specific primer is designed, the sequencing can be carried out using this kit. The cycle-sequencing reaction system was set up as follows: 10 ng template DNA (less than 1 µg), 1 µl PreMix (DNA Sequencing Kit, Big Dye v. 3.0); 0.5 µl (10 µM) primer; add ddH_2O to a final volume of 10 µl.
The sequencing program was as follows: 96°C for 1 minute, 25 cycles with: 96°C for 10 seconds, 55°C for 5 seconds, 60°C for 4 minutes.

The reaction mixture was transferred to a microcentrifuge tube, 10 µl 3M NaAc (pH 4.6) and 250 µl 100% ethanol were added. The solution was kept at room temperature for 15 minutes and centrifuged at 14000 rpm for 20 minutes. The supernatant was discarded, 300 µl 70% ethanol were added, the mixture was centrifuged at full speed for 15 minutes, and the DNA pellet was dried. The product was dissolved in 30 µl formamide (Hi-Di FA).
28 µl of the mixture were transferred to each well of the sequencing plate. Shortly before sequencing the samples were denatured for 2 minutes at 90°C and placed on ice. Data collections were performed on the ABI PRISM 310 Genetic Analyzer, software: Version 1.6 (PE Biosystems).

2.2.1.9. Array Comparative Genomic Hybridization (aCGH)

The aCGH procedure was established and made by the aCGH facility from the Roswell Park Cancer Institute (Buffalo, USA).

- **Printing of the BAC array**

The human 6K RPCI-11 BAC genomic array was provided, hybridized and analyzed by the Roswell Park Cancer Institute, Buffalo, NY, USA (http://microarrays.roswellpark.org). It was generated essentially as described by Snijders et al. (2001). The BAC DNAs, used as templates for PCR representations, were prepared using a Qiagen BioRobot 3000 using R.E.A.L Prep 96 BioRobots kits. The BAC DNAs serve as templates for restriction digestion and PCR amplification as described (Snijders et al. 2001). To generate DNA for printing, a second amplification was performed in a 100 µl reaction containing 1 µl primary PCR product, 4 µM Mse-21 primer, Platinum *Taq* buffer (1x, Invitrogen), 0.2 mM dNTP mix, 5.5 mM $MgCl_2$ (Invitrogen), 2.5 U Platinum *Taq* DNA Polymerase (Invitrogen) and H_2O. PCR products (100 µl) were prepared for printing. Following ethanol precipitation, the pellets were air-dried and re-suspended in 20 µl 25% DMSO in H_2O (final DNA concentration ~0.8 µg/µl). Slides were printed using 10K Microspot pins and a MicroGrid II TAS arrayer (BioRobotics) in duplicate to create an array of ~12,000 elements on amino-silanated glass slides (Schott Nexterion type A). The BAC array used in this study comprised 6094 RPCI-11 BAC clones of an average length of 150-

kb, providing an approximate resolution of one clone per 0.5 Mb of human genomic sequence.

- **Hybridization on the BAC array**

Hybridization to the BAC arrays involves the differential labelling of test and control DNA. The control for the interspecies comparison consisted of pooled samples from ten cytogenetically normal and unrelated human females. The test samples, for each primates tested, consisted of pooled genomic DNAs from ten unrelated female. Genomic DNA from reference and test samples were fluorescently labelled by random priming in a 100 µl reaction containing 1 µg DNA, 1x random primer solution (BioPrime DNA Labeling System, Invitrogen), 1 mM Tris, pH 7.6, 0.1 mM EDTA, 0.2 mM each dATP, dTTP and dGTP, 0.05 mM dCTP, 0.4 mM Cy3- or Cy5-dCTP (PE Life and Analytical Sciences) and 160 U Klenow fragment (BioPrime DNA Labeling System, Invitrogen). Prior to hybridization, the test and reference probes were resuspended in 110µl SlideHyb Buffer #3 (Ambion) containing 5 µl 20 µg/µl Cot-1 and 5 µl 100 µg/µl yeast tRNA (Invitrogen), heated to 95°C for 5 minutes and incubated for 20 minutes at 37°C. During this incubation, the arrayed slides were placed in a GeneTAC hybridization station (Genomic Solutions) and preheated to 50°C. The entire hybridization mixture was added and the arrays were hybridized for 16 hours at 55°C, with occasional pulsation of the hybridization buffer. After hybridization, the slides were automatically washed in the GeneTAC station with reducing concentrations of SSC and SDS.
The hybridized slides were scanned using a GenePix 4200A Scanner (Axon, Inc.) (Figure 2.1).

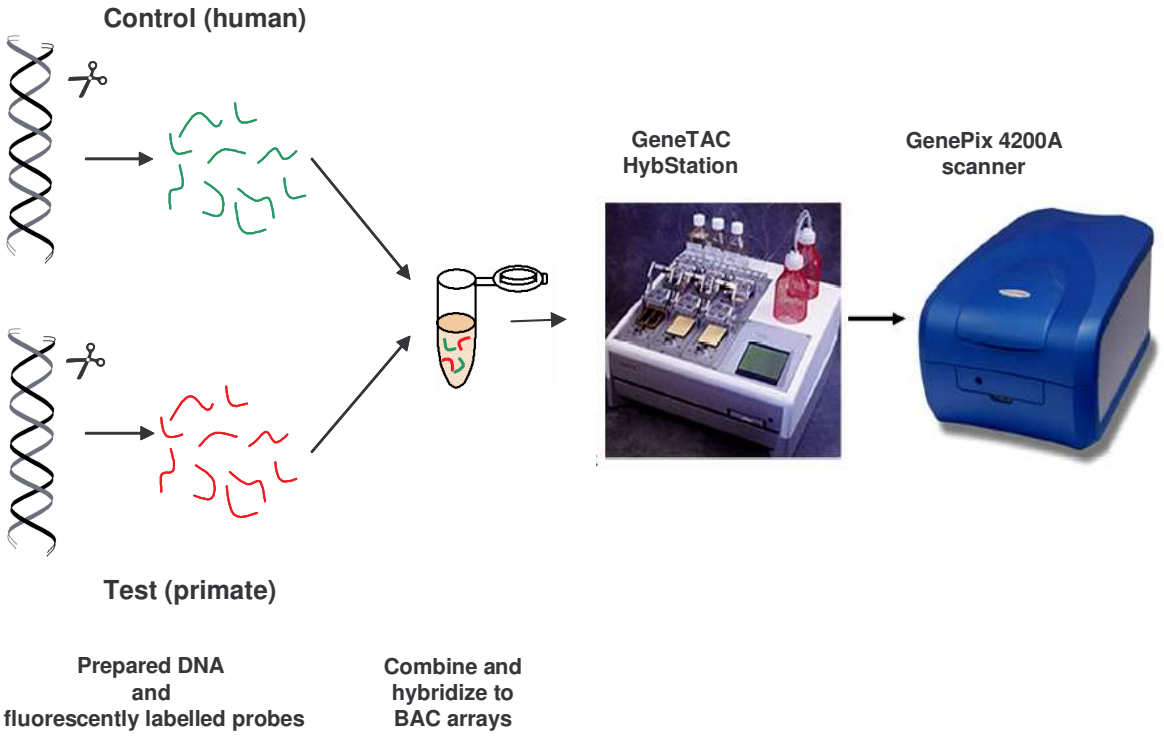


Figure 2.1: Flowchart of the BAC Array-based Comparative Genomic Hybridization procedure. Control and test DNA samples are differently labelled using Cy3 and Cy5, respectively. Both samples are combined and hybridized on the BAC array using the GeneTAC Station. After hybridization, the slides are scanned using the GenePix 4200A scanner.

- **Analysis of the Results**

The goal of image analysis is to identify the spots and quantify the raw fluorescent value for each element on the array.

Image analysis was performed on the raw image files using the ImaGene (version 4.1) software (BioDiscovery, Inc). The spots were identified using an optimised segmentation algorithm. Spots of poor quality, as well as spots with signal levels indistinguishable from the background, were excluded from further analysis. The high rate of “bad” spots in the present study might be caused by sequence divergence between human and great ape DNA samples. The extracted image data were used for data analysis, including background subtraction, data normalization, calculation of ratios, and statistical analysis of replicate spots and slides. These calculations were performed using ImaGene (BioDiscovery) and GeneTraffic (Iobion) softwares, the statistical package R, and PERL scripts.

A grid is aligned over the arrayed spots and a corresponding BAC ID file is integrated. The BAC ID file is generated by the arrayer and when merged with the image analysis software, identifies what clone is spotted in each location. Mapping information from UCSC's genome browser (http://genome.ucsc.edu/) is added to each of the BAC clones. The data are then imported to Excel for visualization (Figure 2.2).

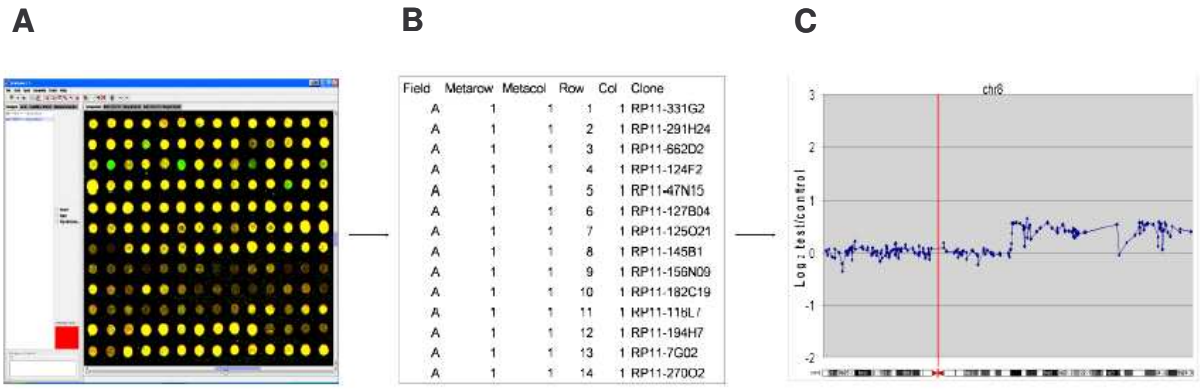


Figure 2.2: BAC array-based comparative genomic hybridization image and data analysis.
After scanning, each spot is identified as a BAC clone (A). The results for each BAC clone are imported to Excel (B). Mapping information from UCSC's genome browser added to each BAC clone allows a graphical overview of the results obtained (C), here shown for human chromosome 8.

In order to facilitate computational evaluation, I applied the most widely used alternative transformation of the ratio, the logarithm base 2. It has the advantage of producing a continuous spectrum of values and treating higher and lower DNA copy number in a similar fashion. Sites with average Log_2 ratios > 0.5 and < -0.5 were selected as representing putative copy number differences. Locke et al. (2003b) recommended a threshold of 0.5 as the most appropriate for interspecies comparison, to minimize the selection of false positives.
In order to avoid redundantly detected loci due to arrayed BACs, which cover the same locus, I considered that concordantly variant neighbouring clones separated by < 250-kb represent the same CND. However, some neighbouring BACs exhibiting concordant differences in one species (and which were therefore taken as a single CND), manifested discordant differences in another species and were thus considered, for that species, to be a different CND from the same locus.

2.2.1.10. Genomic DNA library screening by PCR

In order to identify BACs from the gorilla genome that map to the breakpoint regions, colony-filter hybridizations of the CHORI-255 library (*Gorilla gorilla*) were performed with 40-bp synthetic probes included in the PCR products P8, P9, and P10. PCR primers (listed in Table 2.1) were designed according to the sequences of human BACs. The filter hybridization was made by the group of P. J. de Jong at the Children's Hospital Oakland Research Institute, US. Radiolabelled 40-bp overgo probes were created from 3'-overlapping 24-mers by extending the primers with Klenow DNA polymerase using P^{32}-dCTP as described by Ross et al. (1999). Hybridization-positive BAC clones were first confirmed by PCR analysis with primer pair P8 and three BAC clones were shown to be true positives. Further PCR reactions were performed with primer pairs P9 and P10, thereby potentiating the identification of the breakpoint-spanning gorilla BACs CH255-39D12 and CH255-551C4.

2.2.2. RNA technology

2.2.2.1. RNA extraction and cDNA preparation

Cytoplasmic RNA was isolated from lymphoblastoid and fibroblast cell cultures derived from three chimpanzees and four humans, as well as lymphoblastoid cells from one bonobo, one gorilla, one orangutan and one macaque, using the RNeasy kit (Qiagen, Hilden, Germany). First strand cDNA was prepared with 2 µg total RNA (as determined by absorbance), random hexamers and the SuperScript Preamplification System (Invitrogen, Karlsruhe, Germany).

2.2.2.2. Real time PCR

To measure potential gene expression differences between humans and primates, Real-time RT-PCR was performed using the DNA-binding dye SYBR Green on a Lightcycler with the Quantitect SYBR Green PCR kit (Qiagen). Relative quantification

was used to normalize the target genes (*LOC401848*, *FLJ43980* and *SHCBP1*) to an internal standard (*HPRT*, hypoxanthine-guanine phosphoribosyltransferase). To determine mRNA expression, a standard curve for each of the two fragments was generated. The change in gene expression was then given as the ratio of the target gene to *HPRT* expression. For each analysis, two *Pan troglodytes* lymphoblastoid cell lines were used, together with one *Pan troglodytes* primary fibroblast culture, and single lymphoblastoid cell lines from *Pan paniscus*, *Gorilla gorilla*, *Pongo pygmaeus* and *Macaca fascicularis*. Two foreskin fibroblast cell lines and two EBV-transformed lymphoblastoid cell lines from human donors were also used. Each comparison was performed in triplicate.

Relative copy number differences in the *FMRPD2* and *ANAPC1* genes were also analyzed with real-time quantitative PCR. For each analysis, genomic DNAs from one chimpanzee, as well as from one human, were used. Standard curves were created using dilutions of known amounts of PCR product. All samples were run in triplicate. Results were analyzed using the standard curve method according to the manufacturer's instructions (Roche Applied Science).

2.2.3. Cell technology

2.2.3.1. Cell culture

Growth medium for fibroblasts

50 ml	10x DMEM
50 ml	FCS
2.5 ml	L-Glutamin
14.6 ml	Sodium pyruvate
2.5 ml	Gentamycin

Growth medium for lymphoblasts

500ml	RPMI
75 ml	FCS (15%)
3 ml	Gentamycin

The lymphoblastoid and fibroblast cell lines were cultured in an incubator at 37°C with 5% CO_2 gassing.

Every third day, the cells were supplied with new fresh growth medium.

2.2.4. Chromosome technology

2.2.4.1. Fluorescence in Situ Hybridization technique (FISH)

10% Bovine Serum Albumine (BSA)

10 gr	BSA

100 ml ddH2O, vortex, store at 4°C for several weeks

0.5 M EDTA (pH 8.0)

50 ml	ddH_2O
18.6 gr	Na_2-EDTA

Adjust pH to 7.0 with 5 N NaOH

Add ddH2O to 100 ml, sterilize by autoclaving

1% Formaldehyde solution

68 ml	PBS + $MgCl_2$
2 ml	formaldehyde

MM (Master Mix)

400 µl	50% dextransulfat
200 µl	20% SSC (pH 7.0)
400 µl	ddH_2O

10 x NT Buffer

0.5 M	Tris / HCl (pH 7,5)
50 mM	$MgCl_2$
0.5 gr/ml	BSA

Pepsin solution

99 ml	ddH_2O
30 µl	pepsin
1 ml	HCl

RNase solution

1 µl	RNase A (100 µg/ml)
100 µl	2 x SSC

20 x SSC (1 liter)

175.3 gr	NaCl (3 M)
88.2 gr	Na_3 Citrate (0.3 M)

Adjust pH to 7.0 with HCl

Fluorescence in Situ Hybridization is a cytogenetic technique, which is used to detect and localize DNA sequences on chromosomes. It uses fluorescent probes, which bind only to those parts of a chromosome with which they show a high degree of sequence similarity.

At first, the DNA probe is labeled (Nick Translation). The probe has to be long enough to hybridize specifically to its target, and should be tagged with fluorophores, with targets for antibodies or with biotin. Then, a chromosome preparation is produced. The chromosomes are firmly attached to slides. After preparation, the probe is applied to the chromosome DNA and hybridization is started. In several wash steps, all unhybridized or partially hybridized probes are washed away.

- **Chromosome preparation**

Cells were centrifuged for 10 minutes at 1000 rpm. The supernatant was discarded and the cells were treated with a hypotonic solution as KCl (to let cells swell). The mixture was incubated 20 minutes at 37°C. The cells were centrifuged for 10 min at 1000 rpm and the supernatant was removed. 3 ml of Fix solution was added. The centrifugation and the fixing steps were repeated once more. The mixture was incubated overnight at 4°C. The fixing step was repeated twice. The mixture was dropped on slides and air dried overnight.

- **Nick Translation**

The Nick Translation is a tagging technique in which a DNaseI is used to remove some of the nucleotides of a DNA sequence. Then, DNA polymerase is applied to repair the sequence using fluorescently tagged nucleotides, thus creating a tagged DNA sequence, which can be used as a probe in Fluorescent in Situ Hybridization. At least 1 µg BAC-DNA was labelled with biotin-16-dUTP (Roche-Diagnostics, Mannheim, Germany). To perform co-hybridisation with a differentially labelled BAC, DNA was labelled with digoxigenin-11-dUTP (Roche-Diagnostics, Mannheim, Germany)
The preparation of the mix is as follow:

x µl	DNA template (1-1.5 µg)
10 µl	10x NT Buffer
10 µl	0.1 M β-Mercaptoethanol
10 µl	dNTP-Mix
4 µl	Biotin-16-dUTP or Digoxigenin-11-dUTP
2 µl	DNA Polymerase I
2-3 µl	DNase I (1:1000)

ddH_2O was added to a final volume of 100 µl.

The mixture was incubated at 15°C in a water bath for 40-60 minutes. The length of the fragments was checked by agarose gel electrophoresis.
To inactivate the enzyme, 1 ml of 0.5 M EDTA was added and the mix was incubated for 10 minutes at 65°C.

- **Preparation of the Probe**

2.5 volumes of ethanol, 40 µl of Cot DNA, 3 µl of herrings sperm DNA and 1/10 volume 3 M sodium acetate were added to the marked probe (from the Nick Translation). The mixture was inverted several times and incubated at -70°C for 20 min. The mixture was centrifuged at 4°C, at full speed for 20 minutes. The pellet was washed with 150 µl of 70% ethanol. The mixture was centrifuged again for 15 minutes and then dried on a 37°C thermoblock for about 5 minutes. The DNA pellet was dissolved in 25 µl formamid and incubated 1,5 - 2 hours on a 37°C thermoblock. Finally, 25 µl of master mix was added and the denatured probe was incubated at 76°C for 6-10 min.
Before the hybridization of the probe on the chromosomes, a pre-annealing step was made to mask the repeat sequences by incubating the probe at 37°C for 20-40 minutes.

- **Slide Pre-Treatment**

The slides were treated with RNase and pepsin to reduce an unspecific binding of the probes to RNA and to allow a better access of the probe to the chromosomes. Therefore, the slides were incubated for 1 hour at 37°C with an RNase solution. The slides were then washed 3 x 5 minutes in baths of 2 x SSC.
For the digestion of proteins, the slides were incubated for 8 minutes at 37°C in a pepsin solution. The slides were washed 2 x 5 minutes in 1 x PBS and 1 x 5 minutes in 1 x PBS/50 mM $MgCl_2$. A 15 minutes fixation step in 1% formaldehyde was made to conserve the morphology of the chromosomes. Finally, the slides were washed 2 x 5 minutes in 1 x PBS and dehydrated in a serial of ethanol solutions (70%, 80%, 90%, 99%).

- **Denaturation of the chromosomes and hybridization**

The chromosomes were denaturated in a 70 % formamid solution for 1 minute at 70°C. The slides were washed in 4°C 2X SSC and for 3 minutes in 70% ethanol.
A few µl of the probe was dropped on the dried, denaturated chromosomes and the slides were covered with a coverslip. Per 24 x 24 mm coverslip, 10 µl of the hybridization mix was poured, while per 24 x 32 mm coverslip, 20 µl of the hybridization mix was poured. To avoid the mix to dry out, the coverslip was sealed with Fixogum. The slides were incubated for 36 hours at 37°C.

- **Washing and detection**

After the hybridization, the Fixogum was carefully removed. The coverslip was washed away in 2X SSC at 45°C.
The slides were washed in the following steps: 3x 5 minutes in 2x SSC at 45°C, 3x 10 minutes in 50% FA/2x SSC at 45°C, 2x 5 minutes in 2x SSC at 45°C, and 2x 5minutes in 0.2x SSC at 60°C.
To reduce the unspecific binding of the antibody, the slides were incubated for 30 minutes in 5% BSA/4x SSC/0.1% Tween 20 solution, at 37°C.
The slides were washed shortly in 2x 4X SSC + 0.1% Tween 20 solution.
The detection was made with the following combination of antibodies:

Probes marked with Biotin

30 min incubation with FITC-Avidin (10 µg/ml)

45 min incubation with biotinylated Anti-Avidin (5 µg/ml)

30 min incubation with FITC-Avidin (10 µg/ml)

Probes marked with Digoxigenin

30 min incubation with Anti-Digoxigenin from mouse (0.2 µg/ml)

45 min incubation with Texas Red Anti-mouse from rabbit (14 µg/ml)

30 min incubation with Texas Red Anti-rabbit (14 µg/ml)

All antibodies were diluted in 150 ml of 1% BSA/4x SSC/0.1% Tween 20. The slides were covered with a coverslip and incubated for at least 30 minutes in 37°C.
After each detection, the slides were washed 3x 5 minutes in 4x SSC/0.1% Tween 20 at 45°C, in the darkness.

Finally, slides were counterstained with diamidinophenylindole (DAPI) and mounted with Vectashield antifade solution.

- **Analysis of the results**

The slides were evaluated on an AxioPlan microscope (Zeiss, Germany) using a AxioCam MRm camera (Zeiss, Germany) and the picture analyzer software ISIS v5 from MetaSystem. For each fluorochrome (DAPI, FITC and Texas Red), a specific filter was used.

2.2.5. *In silico* analysis

2.2.5.1. Computer programs

Alignments of BAC sequences were performed by BLAST searches against genome databases (http://www.ncbi.nlm.nih.gov/blast/) and FASTA analyses using the Wisconsin package, version 10.2 (Genetics Computer Group). Repeat masking was carried out using the Repeat Masker server from EMBL, Heidelberg (http://woody.embl-heidelberg.de/repeatmask/).
The search for human-specific duplications was performed using the Human Genome Segmental Duplication Database (http://projects.tcag.ca/humandup) based on the May 2004 Human Genome Assembly (hg17).
Human and chimpanzee genomic sequences were compared using Multicontig View from the Ensembl database (www.ensembl.org). The gene content of the duplications was analyzed by investigating the UCSC EST database (http://genome.ucsc.edu/; Kent et al., 2002) and the NCBI Map Viewer.
In order to identify the ancestral donor locus for each duplication cluster (DC) >5-kb, I applied mouse-human sequence alignment methodology. Using a cross-species megaBLAST, which compares distantly related nucleotide sequence, I was able to identify the “best match” within the mouse genome. Once a mouse locus was identified, the mouse sequence was searched against the genome assembly to identify a putative ancestral human locus.

2.2.5.2. Characterization of the boundaries of the DCs

Screening the Human Genome Segmental Duplication Database, I identified DCs (duplication clusters) encompassing the arrayed BACs detecting human lineage-specific copy number differences (HLS-CNDs) and their respective homologous duplication clusters (DCs) (≥ 5-kb, ≥ 90% identity). Neighbouring DCs that are separated by ≤ 50-kb were grouped in a larger cluster of which the boundaries were analyzed. In order to define precisely the boundaries of each DC and its homologue, I performed pairwise alignments by FASTA analyses including at least 1-kb of flanking sequences.

A total of 500 bp within and flanking each DC at the left and the right boundaries were compared with respect to their repeat content. Left and right sides were assigned according to the orientation of the DCs as indicated in the Human Genome Segmental Duplication Database. For DCs on a given chromosome, the amount of repeats in each of the four regions inspected (flanking and internal, on the left and on the right sites) was averaged. To assess whether the repeat content of the flanking regions of the DCs highlighted in my study was different from that of other genomic regions, I compared each DC boundary separately to 500 randomly chosen regions of equal size and calculated the overall fraction of repeat sequences in randomly chosen genomic sequences. A repeat sequence fraction of randomly chosen genomic sequences that matched or exceeded the base-pair fraction of the boundaries of the DCs was held to be suggestive of a chance occurrence.

Information about repeats was obtained using several perl and awk scripts parsing the information from the chr*_rmsk.txt files available at the UCSC (http://genome.ucsc.edu/).

3. RESULTS

3.1. Study of micro-differences between human and primate genomes.

Recently, Locke et al. (2003b) have shown the presence of large-scale copy number differences (CNDs) between human and the great apes. In their initial study, Locke et al. (2003b) used an array bearing 2,460 human BAC clones. This clone-set encompassed ~12% (370-Mb) of the human genome, with a resolution of one BAC every 1.4-Mb DNA. In total, Locke et al. identified 63 sites that exhibited large-scale CND. Thus, aCGH with BAC clones constitutes a sensitive high-resolution method to assess lineage-specific CND between primate genomes.
In this study, I extended these inter-specific comparisons of primate genomes using a 6-K array encompassing about 30% of the human genome. On average, it provides a resolution of one clone per 0.5 Mb of the human genomic sequence.
The aCGH was performed by hybridization of human genomic DNA as reference to differentially labelled DNA from five primate genomes: macaque (*MFU*) orangutan (*PPY*), gorilla (*GGO*), chimpanzee (*PTR*) and bonobo (*PPA*). I have investigated the DNA copy number changes in each species and I have paid special attention to the identification and the verification of human-specific CNDs since these could have contributed in principle to the development of some of the characteristics that are unique to the human species.

3.1.1. Number of differences detected by aCGH in primate genomes

For each of the primate species investigated (*PTR, PPA, GGO, PPY* and *MFU*), DNA from ten unrelated females was pooled, labelled and hybridized together with the differently labelled human reference DNA pool, also consisting of ten unrelated (female) individuals. In order to validate the aCGH procedure, a control experiment was performed using genomic DNA from one individual per species that had not

been included in the genomic DNA pool. It should be noted that *MFU* (*Macaca fuscata*) was replaced by *MFA* (*Macaca fascicularis*) in the control experiment. No significant differences were observed between the results of both types of experiments (viz. the pooled DNA procedure and the control experiment with one DNA sample per species), except for the gorilla and the macaque (Table 3.1A, B). This might be explained by the use of differently sourced DNA samples in the pool and control experiments (in the latter, lymphoblastoid cell lines were used instead of blood samples). Although these cell lines do not exhibit any obvious karyotypic anomalies, the presence of small-scale rearrangements, that might have introduced between-experiment differences, cannot be excluded. Furthermore, the differences observed between the two experiments using DNA samples from macaque could be explicable in terms of the high DNA sequence divergence of 4.90% (Magness et al., 2005) noted between macaques and humans.

Table 3.1 (A): Comparison of the aCGH experiments using genomic DNAs from a single individual per species with the experiments using pooled DNAs from ten individuals per species

Species	Number of BACs[a]	Number of BACs showing a				Percentage similarity between both experiments[e]
		"bad" spot[b]	difference only in the control experiment[c]	difference only in the pool experiment[d]	difference in both experiments	
PTR	5857	1898	16	10	34	99.34%
PPA	5857	1546	25	14	32	99.10%
GGO	5857	1490	44	37	41	98.15%
PPY	5857	1400	12	19	61	99.30%
MFU	5857	1350	23	111	114	97.03%
Total	29285	7684	120	191	282	98.56%

[a]Excluding the BACs from chromosome X and Y (co-hybridization in the control experiment was performed with genomic DNA from individuals of the opposite sex).

[b]Sum total of the BACs showing a spot of poor quality in at least one of the experiments and which were therefore not taken into account.

[e]$100-\{[(^{c}+^{d})/(^{a}-^{b})]x100\}$.

All the BACs displaying a significant copy number difference have been taken into account.

Correlation coefficient R = 0.978

Table 3.1 (B): Comparison of the 347 sites detected by aCGH using pooled DNAs with the results of the experiment using genomic DNAs from a single individual per species.

Species	Number of BACs showing				Percentage of similarity between both experiments[d]	P-value (df = 1)[e]
	a difference in the pool experiment[a]	a "bad" spot in the control experiment[b]	a difference in the control experiment[c]	no difference in the control experiment		
PTR	59	12	39	8	83%	p>0.2
PPA	55	7	36	12	75%	p>0.05
GGO	94	7	46	41	53%	p<0.0005[f]
PPY	89	15	62	12	84%	p>0.1
MFU	247	27	137	83	63%	p<0.0005[g]

[b]BACs showing a spot of poor quality in one of the experiments and which were therefore not taken into account.

[d] = $[(^{c}/(^{a}-^{b})]x100$

[e]Pearson χ^2 test.

[f,g]Gorilla and macaque exhibited a significant difference in the results of both experiments (p<0.0005).

The difference observed for macaque might be explicable by the DNA sequence divergence between macaques and humans.

In order to exclude an artefactual difference observed in the gorilla, the experiment was repeated using the pooled DNAs from that species: these 94 CND sites were confirmed.

Only the BACs displaying a significant difference in at least one of the primates tested and no "bad" spot in any of the species examined were taken into account.

Correlation coefficient R = 0.979

In order to exclude the influence of experimental artefacts from the human-gorilla comparison, I repeated the human-gorilla pooled-DNA pair-wise hybridization. The 94 CND sites observed in the first experiment using pooled DNAs were confirmed by means of this replication.

Slightly more CNDs were detected by aCGH using single DNA samples, possibly due to an enhanced detection rate of intra-species CNDs. In the final evaluation, however, only those CNDs that were detected in the experiments using pooled DNAs were taken into account.

347 BACs displayed significant difference in the intensity ratio between the human reference pool and at least one of the examined primate species (Appendix 1). Assuming that concordantly variant neighbouring clones separated by <250-kb detect the same CND, the 347 variant BACs correspond to 322 non-redundant loci. Although these include lineage-specific differences and differences shared by two or

more species, the lineage-specific changes clearly predominated. The distribution of these CNDs and their lineage specificity was plotted as a Venn diagram (Figure 3.1). This diagram describes a total of 345 rearrangements. Twenty-three rearranged sites displayed duplication in one species and deletion in another, thereby reducing the total number of rearranged sites from 345 to 322. Interestingly, the majority of the CNDs that I detected are localized close to pericentromeric and subtelomeric regions in the human genome (Figure 3.2).

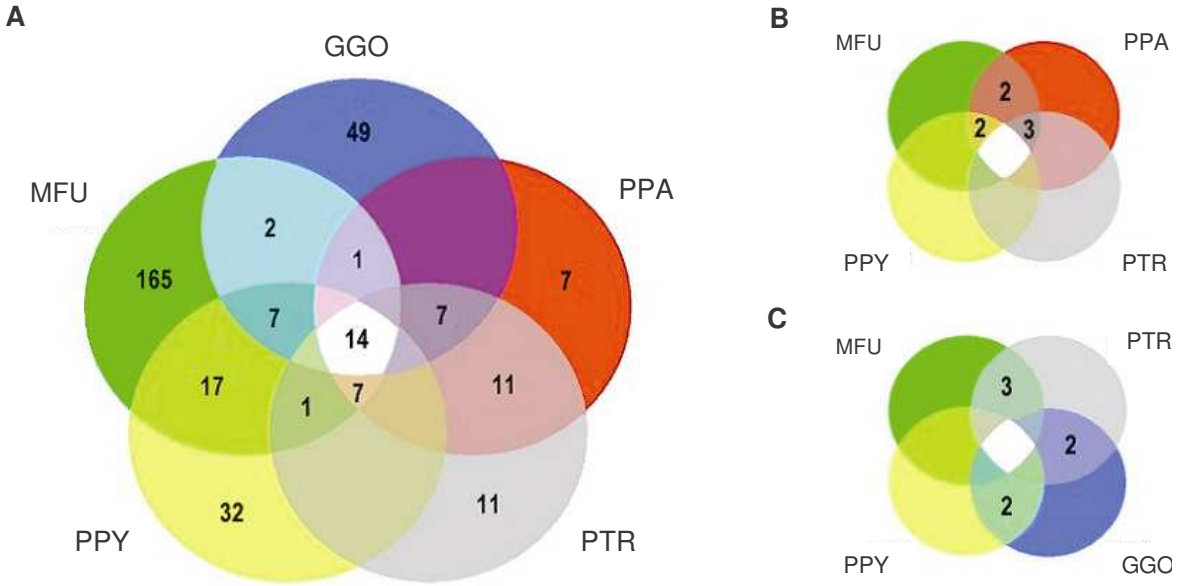


Figure 3.1: Venn diagram depicting sites of species-specific and shared copy-number differences (CNDs) detected by aCGH. Although the Venn diagram describes 345 CNDs, a total of 322 variant sites were detected in my analysis. This disparity is due to some rearrangements presenting a duplication in one species and a deletion in another; these were considered as being distinct and independent events. In the middle of the main diagram, the intersection indicates the 14 human-lineage specific CNDs. The Venn diagram depicted in **A** includes the results from the five species investigated. For the sake of simplicity, and in order to depict the CNDs shared between species not overlapping in **A**, the smaller Venn diagrams **B** and **C** have also been presented.

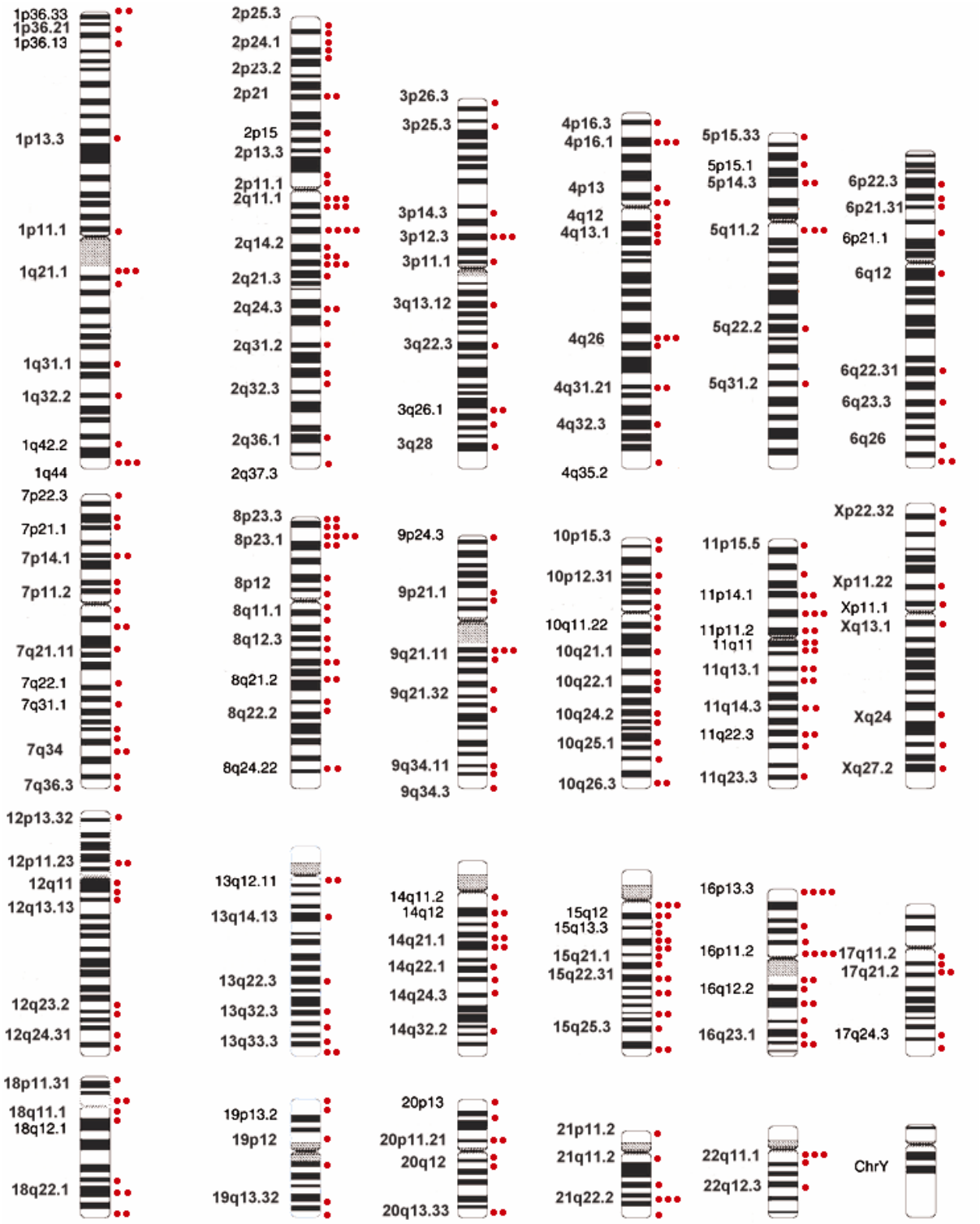


Figure 3.2: Distribution of CNDs in the human genome. Red circles to the right indicate the position of CNDs that exist between human and at least one of the primates tested. No difference has been made between gains or losses. Cytogenetic band positions are shown to the left.

3.1.2. High number of species-specific rearrangements in the gorilla lineage

The majority (93%) of detected lineage-specific CNDs correspond to increases in copy number. Since increases are in practice no easier to detect than decreases, duplications would appear to have been more readily tolerated than deletions during genome evolution. This is consistent with the observation that monosomies are generally less tolerated than trisomies (Brewer et al., 1999). Since I used pooled DNAs from ten human individuals as a reference, I surmise that these copy number gains are *bona fide* and do not simply reflect losses in the reference DNA sample. However, the prevalence of observed rearrangements did not correlate with the evolutionary age of the species investigated (Figure 3.3). Indeed, the gorilla genome was found to have a significantly higher number ($p<0.005$) of lineage-specific variant sites (N = 49) than the genome of orangutan (N = 32), even although the gorilla diverged from the common hominoid ancestral lineage at least 6 million years later than the orangutan (Glazko and Nei, 2003).

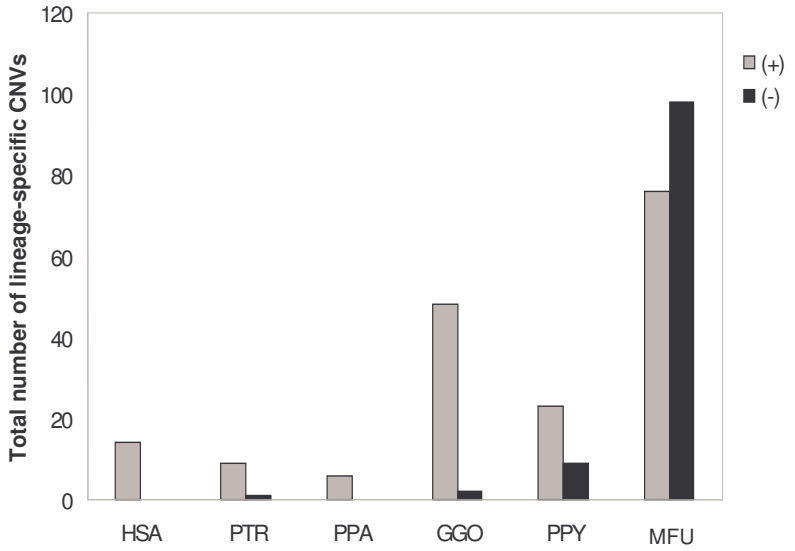


Figure 3.3: Distribution of aCGH-identified lineage-specific CNDs, showing both increases (+) and decreases (-) of copy number among the six species examined. The macaque lineage-specific decreases of copy number might also be interpreted in terms of duplications occurring after the separation of the great apes from the Old World monkeys.

The repetition of the human-gorilla pair-wise comparison by an independent aCGH experiment confirmed all 49 sites of gorilla-specific rearrangements. Sequence analyses of the relevant PCR products also confirmed the authenticity of the gorilla DNA samples. To further verify the gorilla-specific rearrangements, FISH was performed with 10 of the 49 arrayed BAC clones that showed gorilla-specific CNDs (Table 3.3). A clear gain of signal in gorilla as compared to human was detected for six of these BACs. In the four remaining cases in which even interphase FISH did not reveal any obvious gains in gorilla, smaller duplications below the level of FISH resolution may have occurred.

The macaque displayed a high number of putative lineage-specific deletions, which might be due to the 4.9% sequence divergence noted between macaque and human (Magness et al., 2005). Although this would serve to reduce the number of macaque-specific CNDs, *MFU* still exhibited the highest number of duplications among the primate species studied (Figure 3.3). The high rate of CND in macaque is nevertheless quite consistent with the measured evolutionary distance of macaques from the great apes.

Table 3.2: Comparative FISH analysis with BACs that detected putative gorilla-specific CNDs

Clone name	Results of comparative FISH analysis	aCGH Log2 ratio				
		PTR	***PPA***	***GGO***	***PPY***	***MFU***
RP11-82D16	No difference in *GGO* as compared to *HSA*	0,05	0,03	**0,71**	0,02	0,15
RP11-708D7	No difference in *GGO* as compared to *HSA*	0,08	0,05	**0,74**	0,13	-0,24
RP11-95L24	No difference in *GGO* as compared to *HSA*	0,08	0,09	**0,56**	0,07	-0,13
RP11-101N17	Gain in *GGO* as determined by metaphase analysis	0,21	-0,05	**0,62**	0,11	**-0,93**
RP11-143K14	Gain in *GGO* as determined by interphase analysis	-0,14	0	**0,6**	-0,02	-0,05
RP11-140K14	Gain in *GGO* as determined by interphase analysis	-0,01	-0,32	**0,85**	0,07	-0,33
RP11-151D14	Gain in *GGO* as determined by interphase analysis	-0,08	-0,12	**0,58**	0,01	-0,14
RP11-91A4	Loss in *GGO* as determined by metaphase analysis	0,16	0,16	**-0,64**	0,18	0,12
RP11-795L14	Gain in *GGO* as determined by metaphase analysis	**-0,88**	**-1,22**	**1,6**	**-0,87**	**-1,2**
RP11-402P6	No difference in *GGO* as compared to *HSA*	0,03	-0,08	**-0,95**	-0,26	-0,08

3.1.3. Human-specific rearrangements

BACs that detect CNDs in all tested primate species as compared to human are indicative of potential human-specific rearrangements (Figure 3.4). Fourteen BAC clones corresponding to sites of putative human lineage-specific CNDs (HLS-CNDs) were identified (Table 3.3).

Table 3.3: Summary of the 14 BACs that detect human-specific copy number differences (CNDs)

Clone name	Chromosomal position[a]	BAC-specific duplication cluster (DC)[b]	Length of duplication cluster (bp)[b]
RP11-439A17	1p11.2	DC0154	189.538
RP11-35B4[c]	1q21.1		
RP11-91G11	1q21.1	DC0203	43.030
		DC0204	116.988
RP11-89F3	1q21.1	DC0206	131.160
		DC0207	220.312
RP11-196G18	1q21.1	DC0218	99.344
RP11-68E19	2q13	DC0445	266.142
		DC0447	77.357
RP11-397E13	2q13	DC0448	162.865
RP11-79A9	9q12-13	DC2828	371.445
RP11-81J3[d]	10q11.1		
RP11-80B18	10q11.22	DC3093	237.236
RP11-760D22	14q11.2	DC4092	470.695
RP11-499D5	16p11.2	DC3591	584.250
RP11-915P6	22q11.1	DC1786	234.225
RP11-81H21	22q11.1	DC1787	216.815

[a]According to the hg17 assembly.

[b]According to the Human Genome Segmental Duplication Database, based on the May 2004 Human Genome Assembly (hg17).

[c]BAC whose position is as yet unclear in the human genome sequence (unfinished sequencing status).

[d]BAC containing human-specific centromeric satellites.

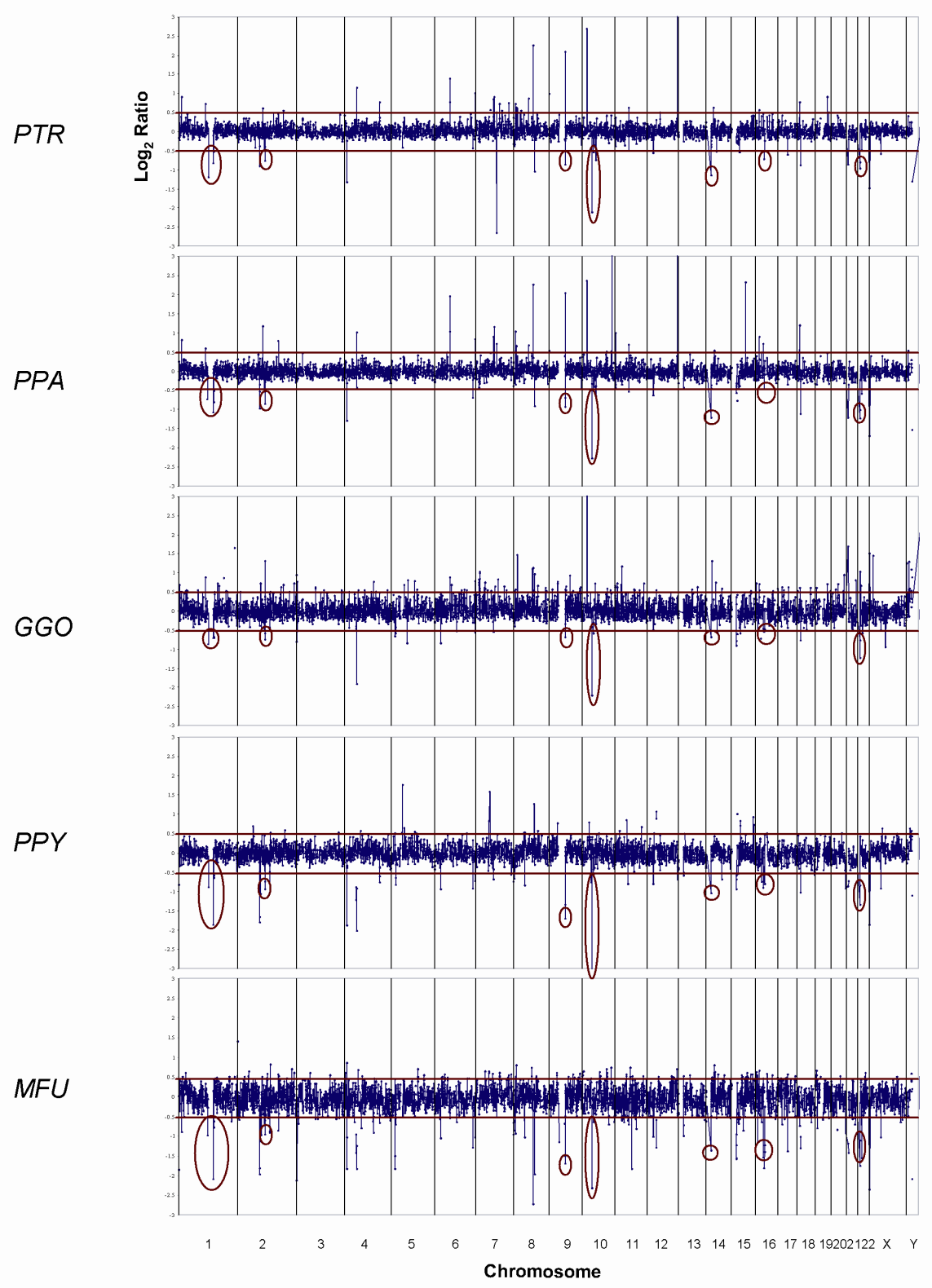


Figure 3.4: Graphs depicting data for ~6000 array loci are presented for each species. (*PTR*) *Pan troglodytes*, (*PPA*), *Pan paniscus*, (*GGO*) *Gorilla gorilla*, (*PPY*) *Pongo pygmaeus*, (*MFU*) *Macaca fuscata*. The sites showing a putative difference in all the primates examined are circled in red. Vertical lines separate the loci from each chromosome, with the p-arm telomere oriented toward the left of each interval and the q-arm telomere toward the right. The horizontal red lines represent the +0.5 and -0.5 $\log_2$ ratio thresholds.

All these changes represent copy number gains in the human lineage, and most of the regions involved contain known genes (Table 3.4).

Table 3.4: Genes contained within the putative HLS-CNDs

Arrayed BAC clone	Genes located within the DC	Gene/protein description	Accession number
RP11-439A17	LOC440748	Hypothetical gene, function unknown	XM_498841
	MGC57827	Hypothetical gene, function unknown	NM_207418
	LOC440686	Similar to histone H3	XM_496408
	LOC440607	Fc-gamma receptor I B2	NM_001017986[e]
RP11-35B4[a]	*VDUP1*	Thioredoxin interacting protein	BC093704[e]
RP11-91G11	FLJ39739[b]	FLJ39739 protein	NM_207400
	FLJ39739 [b]	FLJ39739 protein	NM_207400
	COAS2	Cyclophilin-LC	CAH71953[e]
	MGC8902	Hypothetical gene, function unknown	AAH23087
RP11-89F3	MGC8902 [c]	Hypothetical gene, function unknown	AAH23087
	FLJ20719 [c]	Hypothetical gene, function unknown	BC086308[e]
	MGC8902 [c]	Hypothetical gene, function unknown	AAH23087
	FLJ20719 [c]	Hypothetical gene, function unknown	BC086308[e]
	LOC440682	Hypothetical gene, function unknown	XM_498816
RP11-196G18	*FCGR1A*	Fc fragment of IgG, high affinity Ia, receptor (CD64)	NM_000566[e]
	LOC440607	Fc-gamma receptor I B2	NM_001017986
	HIST2H4	Histone 2, H4	NM_003548[e]
	HIST2H3C	Histone 2, H3c	NM_021059[e]
	HIST2H2AA	Histone 2, H2aa	NM_003516[e]
RP11-68E19	LOC541471	Hypothetical gene, function unknown	NM_001013744
	FLJ41825	Hypothetical gene, function unknown	AK123819
RP11-397E13	*ANAPC1*	Anaphase promoting complex subunit 1	NM_022662[e]
RP11-79A9	LOC375719	Aquaporin 7-like	BC070322
RP11-81J3	-	-	-
RP11-80B18	*FRMPD2*	FERM and PDZ domain containing 2	NM_152428[e]
	PTPN20	Protein tyrosine phosphatase, non-receptor type 20	BC093750[e]
RP11-760D22	LOC440160	Hypothetical gene, function unknown	XM_498571
	ACTBL1	Actin, beta-like 1	NM_001004053
	OR4Q3	Olfactory receptor, family 4, subfamily Q, member 3	NM_172194
	OR4M1	Olfactory receptor, family 4, subfamily M, member 1	NM_001005500

Continuation of Table 3.4

RP11-499D5	*TP53TG3*[d]	TP53TG3 protein	NM_016212[e]
	LOC440362[d]	Similar to hect domain and RLD 2	XM_496146
	FLJ43855	Similar to sodium- and chloride-dependent creatine transporter	NM_198857[e]
	LOC440362[d]	Similar to hect domain and RLD 2	XM_496146
	TP53TG3[d]	TP53TG3 protein	NM_016212[e]
RP11-915P6	LOC400879	Hypothetical gene, function unknown	XM_379029
	LOC440160	Hypothetical gene, function unknown	XM_498571
RP11-81H21	*OR11H1*	Olfactory receptor, family 11, subfamily H, member 1	NM_001005239

[a]BAC RP11- 35B4 contains segmental duplications according to BLAST and FISH analyses but these duplication clusters are not included in the Human Genome Segmental Duplication Database.

[b]Parts of intron 1 and exon 1 of the FLJ39739 gene are included in DC203 and DC204 respectively.

[c]Both DC0206 and DC0207 contain a copy of MGC8902 and FLJ20719.

[d]The duplication cluster DC3591, spanning BAC RP11-499D5, contains local segmental duplications.

[e]Genes represented on the Affymetrix array U95

Human lineage-specific decreases in signal intensity, which could represent HLS-CND losses or gains in the primate genomes tested, were not observed.

To characterize these 14 sites of putative human-specific gains, FISH was performed on human and other primate chromosomes using as probes those human BACs that showed changes by aCGH. It should be noted that FISH was performed on *Macaca fascicularis* chromosomes.

Twelve of the 14 BACs displayed more than two signals on different sites on human metaphase chromosomes, indicating that they represent duplicated sequences (Table 3.5). One exception was BAC RP11-81J3, which contains centromeric satellite sequences specific to human chromosome 10 (Table 3.3), but no signals were detected on other primate chromosomes (data not shown). The other exception was BAC RP11-80B18, which yielded a single FISH signal on human chromosome 10 and its primate orthologues, but BLAST analysis revealed multiple local duplications on 10q11.2-q22.3 (hg17, May 2004 release).

FISH analyses confirmed the human-specific gains for 6 of the 14 BACs and indicated the precise locations of these HLS-CNDs (category 1, Table 3.5).

By way of example, the hybridization patterns generated by BACs RP11-439A17 and RP11-81H21 are shown in Figures 3.5 and 3.6, respectively. On human

chromosomes, BAC RP11-439A17 hybridized to 1p11.2, 1q11.2 and 1q21 but in the chimpanzee, gorilla, orangutan and macaque, hybridization was confined to Iq11.2 and Iq21. Thus, the signal at 1p11.2 probably represents the human-specific duplication. BAC RP11-81H21 yielded very intense FISH signals on 14q11 and 22q11 in humans but not in the other primates tested, suggesting that the human-specific gains are indeed located at these sites. FISH analyses also confirmed the human-specific gain of signal for BAC RP11-915P6 (category 2, Table 3.5), but the unequivocal mapping of the human-specific duplication was found not to be possible by FISH.

For 7 of the 14 BACs detecting putative HLS-CND (category 3, Table 3.5), FISH did not demonstrate an increase in signal on human metaphase chromosomes as compared to the other primates. It would thus appear that the resolution of FISH is insufficient to detect the human-specific amplification events, which may correspond to highly localized copy number duplications. To validate the aCGH results by a second independent method, quantitative PCR was performed for BACs RP11-397E13 and RP11-80B18; this served to confirm the copy number gains at these loci (category 3, Table 3.5).

As shown by FISH, BLAST analysis and screening of the Human Genome Segmental Duplication Database (http://projects.tcag.ca/humandup), 12 of the BAC clones that detected HLS-CNDs contain segmental duplications that are grouped in duplication clusters (DCs) of >40-kb. This is the minimum size range detected by the BAC aCGH technique (Locke et al. 2003b). The acronyms for each DC and the respective CND-detecting arrayed BACs are listed in Table 3.3. Screening the Human Genome Segmental Duplication Database also served to identify DCs highly homologous to those found by aCGH to represent human-specific gains (Appendix 2). The sequence homology between the specific DCs examined in this study and their homologous counterparts is in general >90% and in most instances >98%. I have grouped them together in duplication super-families as shown in Appendix 2. As an example, DC0154 and DC0218, on BACs RP11-439A17 and RP11-196G18 respectively, contain a 46-kb segment with >90% sequence identity. This may represent a human lineage-specific gain of this segment since it was detected twice by BAC aCGH. However, since both DC0154 and DC0218 have homologous DCs (listed in Appendix 2) which were not represented on the array, I am unable to ascertain which of these duplications is human-specific.

Table 3.5: FISH analysis of BACs detecting copy number differences (CND) in humans as compared to other primates

Category[a]	BAC name	*HSA*	*PTR*	*GGO*	*PPY*	*MFA*[e]
1	RP11-439A17	**1p12**, 1q21, 1q32	Iq21, Iq32	Iq21, Iq32	Iq21, Iq32	Iq21, Iq32
1	RP11-196G18	**1p11.2**, **1q21**[b]	Iq21	Iq21	Iq21	Iq21
1	RP11-68E19	**2p11.2**, 2q13	IIq13	IIq13	IIq13	IIq13
1	RP11-81J3[c]	**10cen**	no signals	no signals	no signals	no signals
1	RP11-760D22	2q21, 14q11[b], **15q11**, **22q11**	IIq21, XIVq11	IIq21, XIVq11	XIVq11	XIVq11
1	RP11-81H21	2q21, **14q11**[b], **22q11** [b]	IIq21, XIVq11	IIq21, XIVq11	IIq21, XIVq11, XVq11, XXIq11	XIIq11
2	RP11-915P6	2q21, 13q11, 14q11[b], 15q11, 22q11	IIq21, XVq11[b]	IIq21, XIIIq11, XIVq11, XVq11	IIq21, XVq11, XXIq11	XVq11, XXq11
3	RP11-397E13	2p11.2, 2q13	IIp11, IIq13	IIq13	IIq13	IIq13
3	RP11-35B4	1p36, 1p12, 1q21[b]	Ip36, Ip12[b,d], Iq21	Ip36, Ip12, Iq21	Ip36, Ip12, Iq21, XVq21	Ip36, Ip12
3	RP11-91G11	1p36, 1p12, 1q21[b]	Ip36, Ip12, Iq21	Ip36, Ip12, Iq21	Ip36, Ip12, Iq21, XVq21	Ip36, Ip12
3	RP11-89F3	1p36, 1p12, 1q21[b]	Ip36, Ip12, Iq21	Ip36, Ip12, Iq21	Ip36, Ip12, Iq21	Ip36, Ip12
3	RP11-79A9	2q12, 4p11, 9p11.2[b], 9q12, 14q11	IIq12, IXp11, IXq12, XIVq11	IIq12, IIq21, IXq12, XIVq11, XVq11, XVIIIp11	IXp22, XIVq11	XIVq11
3	RP11-499D5	1q21[b], 2p11, 7p11, 14q11, 15q12, 16cen, 17q11, 22p11	Iq21[b], IIp11, IXp11, Xq11, XIIq11, XIVq11, XVp11, XVIcen[b], XVIIp11	IIp11, XIIIq11, XVq12, XVIcen[b]	XVq12, XVIcen	no signals
3	RP11-80B18	10q11.2-q21	Xq11.2-q21	Xq11.2-q21	Xq11.2-q21	Xq11.2-q21

[a]Category 1: In these cases, FISH indicates the site of the human-specific CND (in bold). Category 2: FISH indicates a gain of FISH signal in human. Category 3: FISH does not show a human-specific gain of signal. [b]High signal intensity is suggestive of multiple local hybridization targets. [c]BAC RP11-81J3 contains human-specific centromeric satellites. [d]*HSA*1 and *PTR*I are distinguished by a pericentric inversion that occurred in the human lineage. [e]Chromosomes from *Macaca fascicularis* were used for the hybridization.

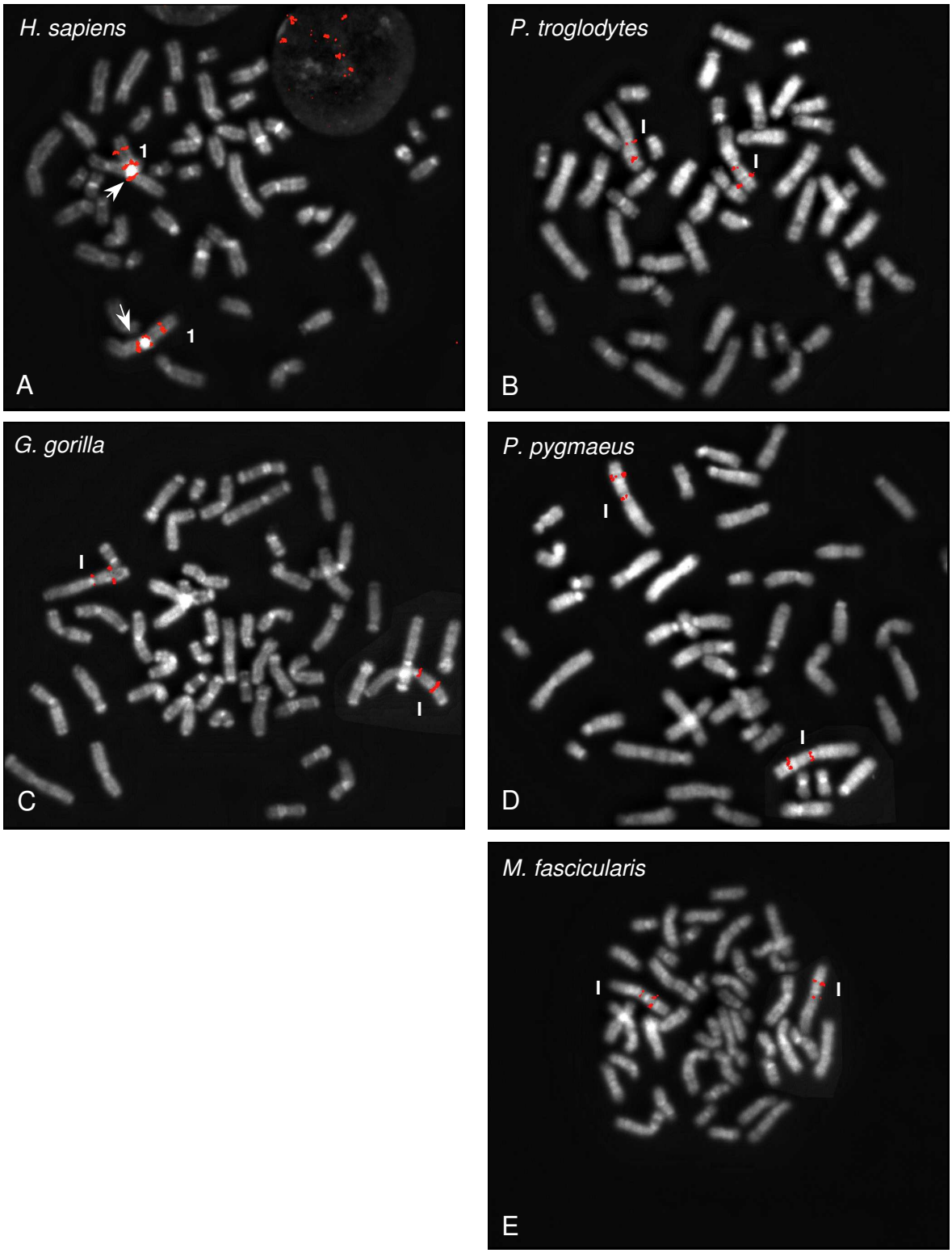


Figure 3.5: Comparative FISH analysis of human BAC RP11-439A17, which detects a human-specific copy number gain, compared to the other primates. On human chromosomes, RP11-439A17 hybridized to three different loci on chromosome 1 (**A**), whereas in *Pan troglodytes* (**B**), *Gorilla gorilla* (**C**), *Pongo pygmaeus* (**D**), and *Macaca fascicularis* (**E**) only two signals were observed.

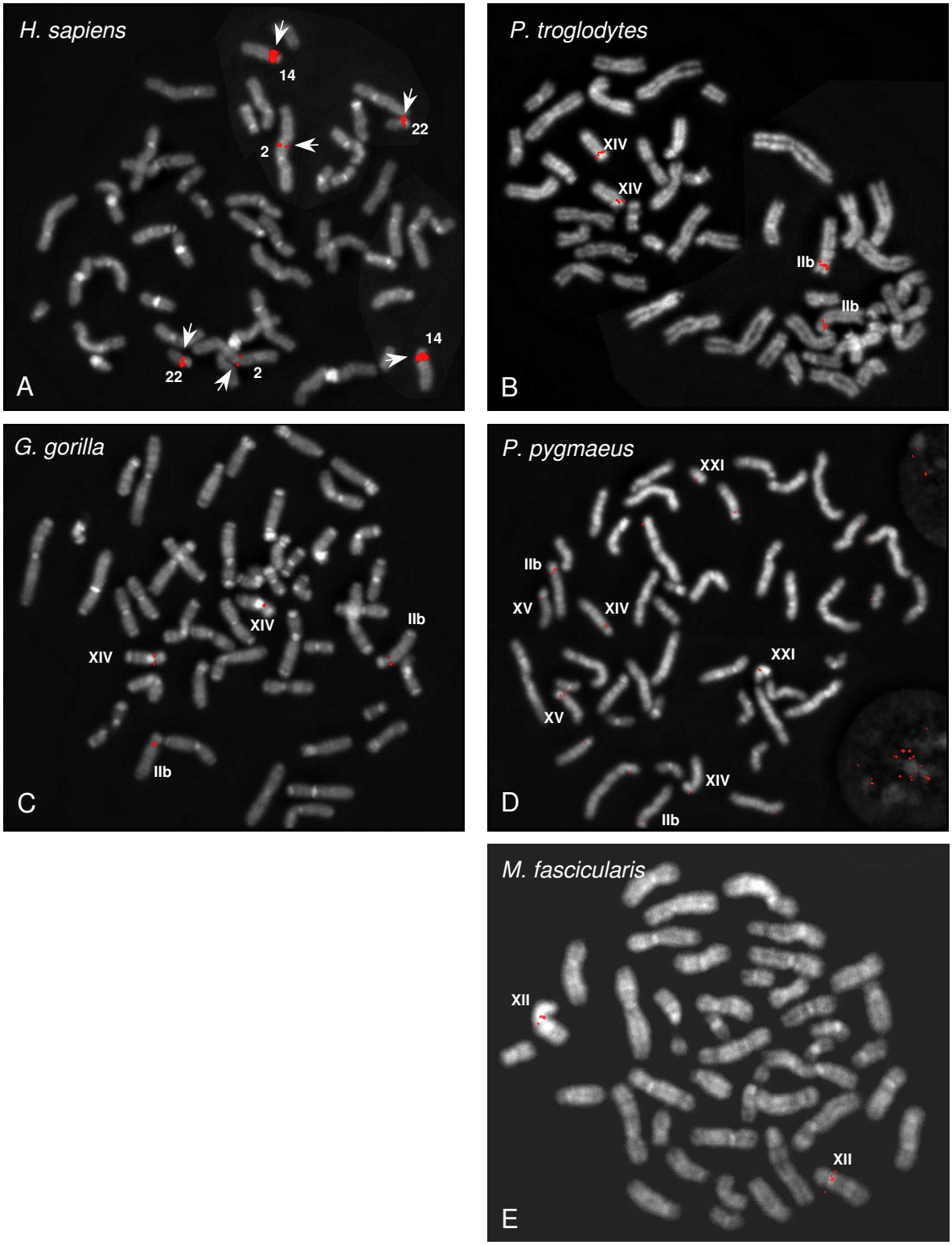


Figure 3.6: Comparative FISH analysis of BAC RP11-81H21 detecting a human-specific copy number gain on human chromosomes 14 and 22 (**A**), *Pan troglodytes* (**B**), *Gorilla gorilla* (**C**), *Pongo pygmaeus* (**D**), *Macaca fascicularis* (**E**) metaphases.

Despite the local homology of the arrayed clones that detect human-specific gains, the FISH and BLAST patterns are specific for each of them (Table 3.5). Therefore, the CNDs detected by these BACs were considered independent events.
In order to assess the polymorphic status of the 12 putative HLS-CNDs and their respective homologous DCs, I checked if these loci were covered by the arrays used by Sebat et al. (2004), Iafrate et al. (2004) and Sharp et al. (2005). As indicated in Appendix 2, 7 of the 12 (58%) putative HLS-CND sites and their homologous DCs are represented on at least one of the arrays used in these three studies. In addition, I compared the localization of the CNDs reported by Tuzun et al. (2005) with my 12 putative HLS-CNDs. I also screened the Database of Genomic Variants (http://projects.tcag.ca/variation) that provides access to the results from these four studies. As summarized in Table 3.6, three of the 12 BACs were found to contain copy number gains corresponding to known CNDs in humans (BACs RP11-68E19, RP11-499D5, RP11-915P6). Moreover, two of the 12 BACs, RP11-89F3 and RP11-760D22, exhibited a high degree of similarity to human DCs known to contain polymorphic loci (Table 3.6). As an example, the duplication cluster specific to BAC RP11-89F3, termed DC0207 and mapping to 1q21.1, is 99% identical to the polymorphic DC0397 sequence (part of locus DC0031) located close to the centromere on chromosome 2, a locus identified by Sebat et al. (2004) as a CND in 15 healthy humans.

Table 3.6: CNDs between humans and other primates associated with copy number polymorphism in the human genome

CND detecting BAC	BAC-specific DC	CND in human[a]	Chromosomal position of the CND	Gain[b]	Loss	DC associated with CND[c] (% identity to the specific DC)	Reference
RP11-89F3	DC0207	Locus0031	2cen	15		DC0397 (98.99)	[e]
RP11-68E19	DC0445	Locus0033	2q13	2		DC0445 (100)	[e]
		Locus0030	2p11.2	1		DC0374 (99.24)	[e]
RP11-760D22	DC4092	Locus0233	15q11.2		18	DC3797 (92.51)	[e, f]
						DC3805 (95.37)	[e]
		Locus0294	22cen		1	DC1786 (99.25)	[e, f]
		Locus0036	2q21.1		1	[DC0458, DC0459, DC0460] (94.5)	[d, e]
		Locus0623	21q11.2		7	DC2077	[f]
		Locus0037	2q21.1		2	DC0472 (93.88)	[d]
						DC0473 (94.85)	[d]
RP11-499D5	DC3591	Locus0251	16p11.2	1	1	DC3586 (99.07)	[d, e]
						DC3591 (99.31[g])	[d, e, f]
						DC3592 (99.46)	[d, e]
		Locus0232	14q32.33	14		[DC4134, DC4135, DC4136] (95.53)	[d, e]
RP11-915P6	DC1786	Locus0294	22cen		1	DC1786 (100)	[e, f]
		Locus0623	21q11.2		7	DC2077	[f]
		Locus0233	15q11.2		18	DC3797 (92.61)	[e, f]
		Locus0037	2q21.1		2	DC0473 (94.8)	[d]

[a]According to the Database of Genomic Variants (http://projects.tcag.ca/variation/). [b]The number of individuals with no obvious disease phenotype showing a gain/loss of the respective locus. [c]Duplications covered by the respective locus. [d]Iafrate et al., 2004, [e]Sebat et al., 2004, [f]Sharp et al., 2005. [g]Local duplication within the duplication cluster

3.1.4. Ancestral loci of the HLS-CND-detecting DCs

In order to identify the ancestral loci of the HLS-CNDs, I aligned the human sequence against mouse using megaBLAST. The numbers of the homologous sequences in the mouse genome are clearly reduced as compared to human (Table 3.7). These CNDs must be both large (>40-kb) and comparatively recent events (>90% identity) that have taken place since the divergence of human and mouse from their common ancestor ~80 Myrs ago. Furthermore, FISH analyses on primate metaphases demonstrated that the macaque possesses the lowest number of CNDs. These results reflect the selective expansion of certain DC families during primate evolution.

Table 3.7: Refinement of the putative ancestral locus for each HLS-CND by mouse-human sequence comparisons using megaBLAST

DC detecting CND by CGH-array	Position	Homology in human	Homology in mouse
DC0154	1p11.2	**1q12**, 1q21.1, **1q21.2, 1q32.1**	1q12, 1q21.2, 1q32.1
DC0203	1q21.1	1q21.1	
DC0204	1q21.1	1p36.13, 1q21.1, **5q11.1**	5q11.1
DC0206	1q21.1	1p36.13-12, 1p12, 1q21.1, 2q21.2, 5q11.1, **7p12.1**, 7q11.21, 9q13, 10q11.21, 13q11, 15q11.2, 16p11.2, 17q11, 18p11.21, 21q11.2	7p12.1
DC0207	1q21.1	1p36.13-12, 1p12, 1p11.2, 1q21.1, 2p11.1, 2q21.2, 4p16.1, 5q11.1, **7p12.1**, 7q11.21, 7q36.1, 9q13, 10q11.21, 13q11, 15q11.2, 16p11.2, 16q11.2, 17q11, 18p11.21, 21q11.2	7p12.1
DC0218	**1q21.2**	1p11.2, 1q21.1, **1q21.2**	1q21.2
DC0445	**2q13**	2p11.2, 2q11.2	2q13
DC0447	**2q13**	2p11.2	2q13
DC0448	**2q13**	2p11.2, 2q12.2	2q13
DC2828	9q12-13	1q21.1, 2p11.2, 2q21.2, 4q11, 6q25.1, **7p12.1**, 7q11.21, 9p11.2, 9q12-13, **10q11.21**, 11q14.3, 13q11, **14q12**, 15q11.2, 16p11.2, 16q11.2, 17q11, 18p11.21, 21p11.2, 21q11.2, 22q11.22	7p12.1, 10q11.21, 14q12
DC3093	10q11.22	**10q11.22**, 10q11.23, 10q21.1, 10q22.3	10q11.22

Continuation of Table 3.7

DC4092	14q11.2	2q21.1, 8q11.21, 9p11.2, 9q12, 10q22.3, 10q11.21, 13q13.2**, 13q13.3**, 14q11, **15q11.2, 17q11.2,** 18p11.21, 21q11.2, 22q11	13q13.3, 15q11.2, 17q11.2
DC3591	16p11.2	1p11.2, 1q21.1, 2q21.2, 5p15.31, **7p12.1**, 7q11.21, 9q13, 10q11.21, **13q31.3**, 14q32.33, 15q11.2, **15q13.1**, 16p11.2, 16q11.2, **17q12**, 18p11.21, 21p11.2, 21q11.2, **Xq28**	7p12.1, 13q31.3, 15q13.1, 17q12, Xq28
DC1786	22q11.1	2q21.2-1, 8q11.21, 9p11.2, 9q12, 10q11.21, 13q13.2, **13q13.3**, 14q11, 15q11.2, 18p11.21, 21q11.2, 22q11	13q13.3
DC1787	22q11.1	2q21.1-2, 10p12.1, 10q22.3, 14q11.2, 15q11.2, **17q11.2**, 22q11	17q11.2

The ancestral loci in human are shown in bold.

3.1.5. Boundaries of putative HLS-CNDs and their respective homologous DCs

In an attempt to identify the mechanism underlying the spread of the duplication clusters (DCs) identified in this study, I examined the repeat sequence content at the boundaries of all DCs overlapping the 12 arrayed BACs that detected HLS-CNDs as well as all their homologues. Neighbouring DCs separated by <50-kb were grouped into a larger cluster. The boundaries of each DC were then determined and those DCs that were not embedded in high-quality sequence contigs were excluded, as were those whose borders could not be precisely demarcated. A total of 138 alignments, met these criteria. I analyzed 500-bp within (internal boundary) and flanking each DC (external boundary) at its beginning (left) and at its end (right), for their repeat sequence content.

In order to ascertain whether certain repeats were specific to the external boundaries of the DCs highlighted in my study, two analyses were performed. Firstly, the repeat sequence fraction of the external boundaries was compared with that of the internal boundaries. Secondly, the level of repeat sequence enrichment of the external boundaries was compared with that of randomly chosen genomic sequences. As shown in Figure 3.7A, only *Alu*Y repeats, a subfamily of *Alu* retroelements, exhibited

significant enrichment in the external boundaries of the DCs. However, the pattern of repeat enrichment was found to be dependent upon the position of the boundaries (left or right). Indeed, the DCs highlighted in my study proved to be significantly ($p<0.0005$) enriched in terms of their *AluY* content on the external right side (16.2%) (Figure 3.7B) but not ($p<0.2$) on the external left side (5.6%) (Figure 3.7C).

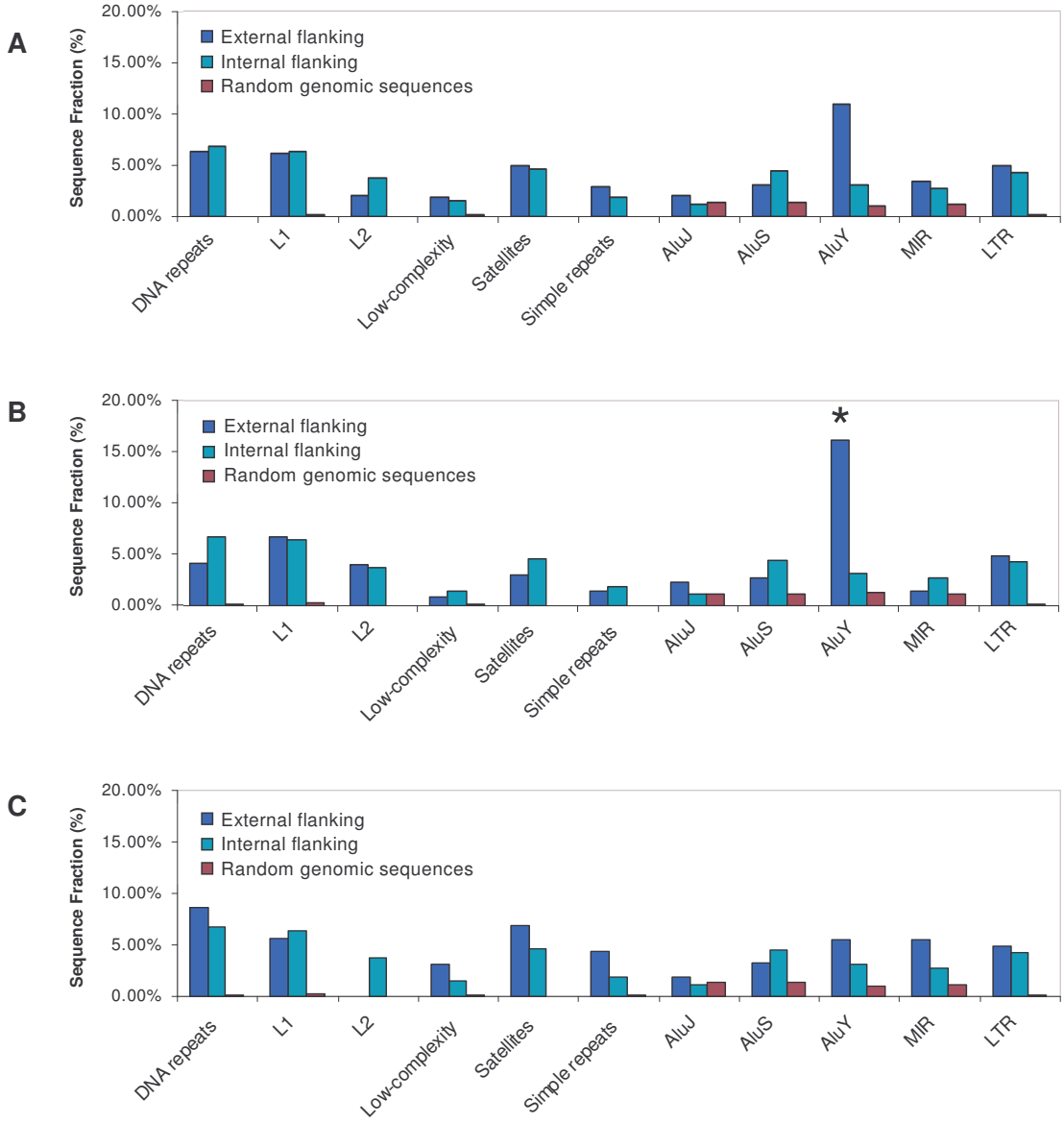


Figure 3.7: Histogram comparing the repeat content of the external flanking regions of the DCs identified in my study with the internal flanking regions of these DCs, as well as the repeat content in randomly chosen genomic sequences (**A**). Repeat content was measured as a total fraction of analyzed base-pairs. A distinction was made between the right (**B**) and left boundaries (**C**) of the DCs. A significant enrichment ($p<0.0005$) of *AluY* repeats (*) was observed in the external flanking regions of the right DCs boundaries.

3.2. Study of two macro-differences between human and primate genomes: examples of evolutionary rearrangements driven by LCRs

It has been shown that DNA copy number changes are implicated in large-scale rearrangements, such as inversions, which could have driven speciation (Navarro and Barton, 2003; Marques-Bonet et al. 2004; Vallender and Lahn 2004a; Zhang et al. 2004). In order to verify the implication of LCRs in evolutionary rearrangements, I have studied two of the nine pericentric inversions reported by Yunis and Prakash (1982). By FISH, PCR and homology searches, I was able to characterize the breakpoints, at the base-pair level, of the inversions that occurred in the gorilla and chimpanzee chromosomes XVI and of the pericentric inversion of *HSA* 18. Such a precise characterization of the breakpoints position allowed me to study the duplication content of the regions around the breakpoints and to verify the putative role of these LCRs in mediating the respective rearrangement.

3.2.1. The pericentric inversion of chimpanzee and gorilla chromosomes homologous to human chromosome 16

3.2.1.1. Molecular description of the inversion breakpoints

In order to characterize the breakpoints of the inversion that distinguishes *PTR* XVI from *HSA* 16, I performed FISH with human BAC clones that map to 16p11-p12 and 16q11-q12. The results of these FISH experiments are presented in Table 3.8. BAC RP11-696P19, the proximal clone from the most centromeric contig (NT_010498) on 16q11.2, exhibited a split signal on *PTR* XVI (Figure 3.8A, B). Although BLAST analyses of BAC RP11-696P19 indicated significant homology to scaffold_32947 of the chimpanzee draft genome sequence, a 4.8-Mb segment from this scaffold was found to be homologous to sequences on *HSA*16q11.2, whereas a 1.3-Mb segment was homologous to proximal *HSA* 16p (Figure 3.9A, B). Moreover, the comparison between the *PTR* scaffold_32947, which is the most proximal scaffold on *PTR* XVIq,

and the human chromosome 16 sequences clearly indicated the presence of an inversion and confirmed that BAC RP11-696P19 spans the evolutionary breakpoint.

Table 3.8: Human BACs used as FISH probes to identify the breakpoint regions of the pericentric inversion of chimpanzee (*PTR*) and gorilla (*GGO*) chromosome XVI.

BAC	Contig	Position on *HSA* 16 in UCSC, May 2004 release	Position in *HSA*	Position in *PTR*	Position in *GGO*
RP11-426C22	NT_010393	28,987,481-29,195,477	p	p	-
RP11-341P6	NT_010393	33,186,123-33,372,826	p	p	-
RP11-347N4	NT_010393	33,556,175-33,754,470	p	p	p
RP11-598D12	NT_010393	33,597,610-33,767,476	p	p	p
CTD-2522B17	NT_010393	33,758,770-33,930,651	p	p	p
CTD-2144E22	NT_024773	34,030,652-34,199,197	p	p/q	p/q
RP11-20D16	NT_024773	34,067,271-34,224,164	p	q	q
RP11-14K3	NT_024773	34,620,966-34,774,594	p	q	-
RP11-65B9	NT_024773	34,982,278-35,143,302	p	q	q
RP11-696P19	NT_010498	44,943,303-45,066,095	q	p/q	p/q
RP11-46D6	NT_010498	45,058,799-45,229,361	q	q	q
RP11-93O14	NT_010498	45,114,469-45,275,991	q	q	q
RP11-523L20	NT_010498	46,457,936-46,605,357	q	q	-
RP11-44I10	NT_010498	47,019,209-47,188,774	q	q	q

-: not performed

To identify a human BAC clone spanning the inversion breakpoint on the p-arm, I analysed the 1.3-Mb fragment of scaffold_32947 that exhibited homology to *HSA* 16p. The distal part of this 1.3-Mb fragment was found to be homologous to a 144-kb stretch of human BAC CTD-2144E22 sequence (Figure 3.9A). FISH analyses confirmed that this BAC spans the inversion breakpoint, since hybridization signals were detected on both *PTR* XVIp and XVIq (Figure 3.8C, D). The breakpoint is located within the very distal portion of BAC CTD-2144E22; most of the corresponding sequence has been inverted and is located on *PTR* XVIq (Figure 3.8D).

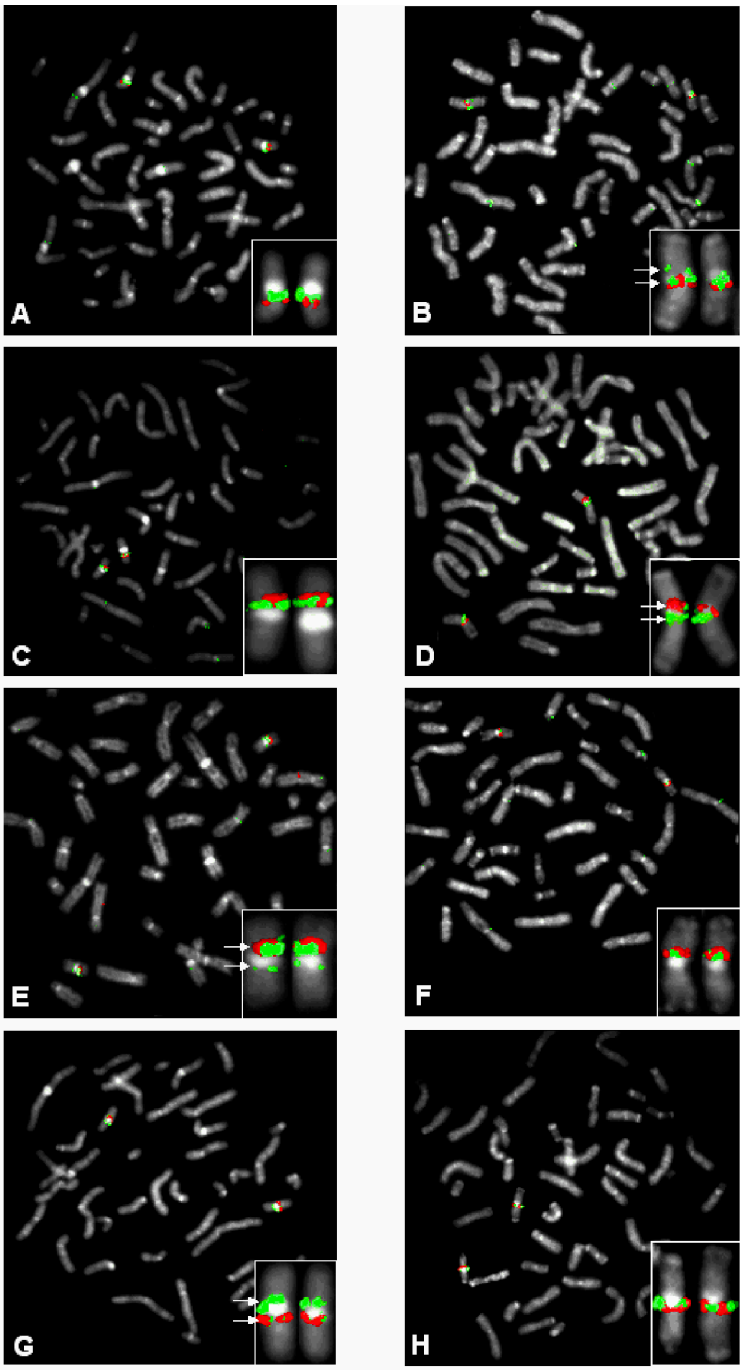


Figure 3.8: Comparative FISH analysis of breakpoint-spanning BACs (arrows indicate split signals). **A, C, E, G:** human metaphases. **B, D, F, H:** chimpanzee metaphases. **A, B**: FISH performed with *HSA* BACs RP11-696P19 (green) from 16q11.2 and RP11-44I10 (red) from 16q12.1. **C, D**: *HSA* BACs CTD-2144E22 (green) and RP11-598D12 (red), both from *HSA*16p. In *PTR*, BAC CTD-2144E22 is split by the inversion. **E, F**: *PTR* BAC RP43-007E19 (green) and *HSA* BAC RP11-347N4 (red) from *HSA* 16p12. BAC RP43-007E19 is breakpoint-spanning since signals are evident at *HSA* 16p11.2 and 16q11 (**E**). On *PTR* XVIp, a single signal of RP43-007E19 is present (**F**). **G, H**: *PTR* BAC RP43-001I03 (green) and *HSA* BAC RP11-46D6 (red). BAC RP43-001I03 hybridized to a single location on *PTR* XVIq (**H**) but was split by the inversion on *HSA* 16 (**G**).

BLAST analysis with BAC RP11-696P19 was then used to identify breakpoint-spanning BACs from the chimpanzee genome: RP43-007E19 and RP43-001I03. The map positions of these clones are indicated in Figure 3.9. FISH analyses using these chimpanzee BACs revealed split signals on *HSA* 16p and 16q (Figure 3.8E-H) confirming that the BACs do indeed span the inversion breakpoints.

To validate the mapping of the inversion breakpoints, I performed FISH on chromosomes of the bonobo (*Pan paniscus*, *PPA*), the other chimpanzee species. Human BACs RP11-696P19 and CTD-2144E22 were also found to be breakpoint-spanning in the bonobo whilst BAC RP11-20D16 is inverted on *PPA* XVIq (data not shown).

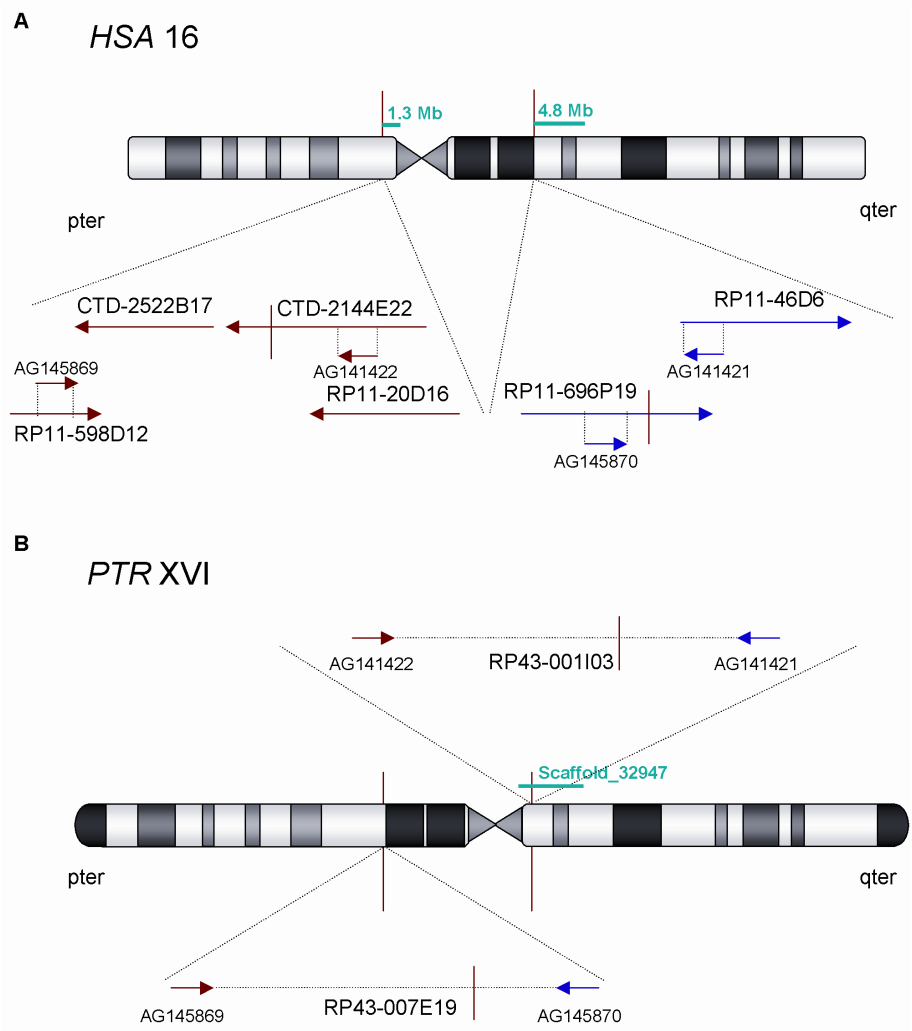


Figure 3.9: Ideogram of human chromosome 16 (*HSA* 16) (**A**) and its chimpanzee homologue (*PTR* XVI) (**A**). BACs used for FISH analyses are indicated in red (*HSA* 16p) and blue (*HSA* 16q). Inversion breakpoints are denoted by vertical red lines. The green lines represent the chimpanzee scaffold_32947, which is split on human chromosome 16.

To narrow down the exact positions of the inversion in relation to the human sequence, various PCR analyses were performed.
Primers were designed by reference to the sequences of the human BACs (summarized in Table 3.9).

Table 3.9: PCR analysis of human and chimpanzee BACs used to identify the breakpoint regions. Primers were designed according to the sequences of the human BACs.

Primer	Position on *HSA* BAC CTD-2144E22	Position on *HSA* BAC RP11-696P19	Present on *PTR* BAC RP43-001I03	Present on *PTR* BAC RP43-007E19
P1f		104781	-	+
P1r		104962		
P2f		114667	-	-
P2r		114943		
P3f		115032	+	-
P3r		115224		
P4f	139229		+	-
P4r	139405			
P5f	144168		+	-
P5r	144360			
P6f	162183		-	+
P6r	162365			
P7f	166997		-	+
P7r	167170			

In relation to the human sequence, the 16q breakpoint maps to between positions 114667-114943 of BAC RP11-696P19 whilst the 16p breakpoint occurred between 144-kb and 162-kb of BAC CTD-2144E22.
The *PTR* XVIq breakpoint was then characterized at the nucleotide level by comparing the sequence of the chimpanzee scaffold_32947 with that of the human BACs, RP11-696P19 and CTD-2144E22.The break in homology between *HSA* 16q and *PTR* XVIq was found to occur at position 114711 of BAC RP11-696P19 (Figure 3.10) whilst the breakpoint on the p-arm was mapped to a short single copy segment of 756-bp located between an L1-element and a stretch of HSATII satellites (Figures 3.10 and 3.11A). Careful analysis of the chimpanzee sequence of *PTR* XVIq revealed the addition of two nucleotides (AC) at the inversion junction. Under the assumption that *HSA* 16 represents the status of the ancestral chromosome, the

breakpoint on *HSA* 16p occurred between two copies of an octanucleotide repeat (TGTGAAAG) whereas the breakpoint on *HSA* 16q occurred within one copy of a juxtaposed pentanucleotide repeat (AATGA) (Figure 3.10). It can also be seen that the core of the juxtaposed pentanucleotide repeat (AATGAAATGA) on 16q corresponds exactly to the core TGAAA motif of both 16p octanucleotide repeat sequences. When q- and p-arm breakpoint sites were further compared, a 22-bp direct repeat was also identified. This local homology between 16p and 16q is indicated in Figure 3.10 by a wavy line.

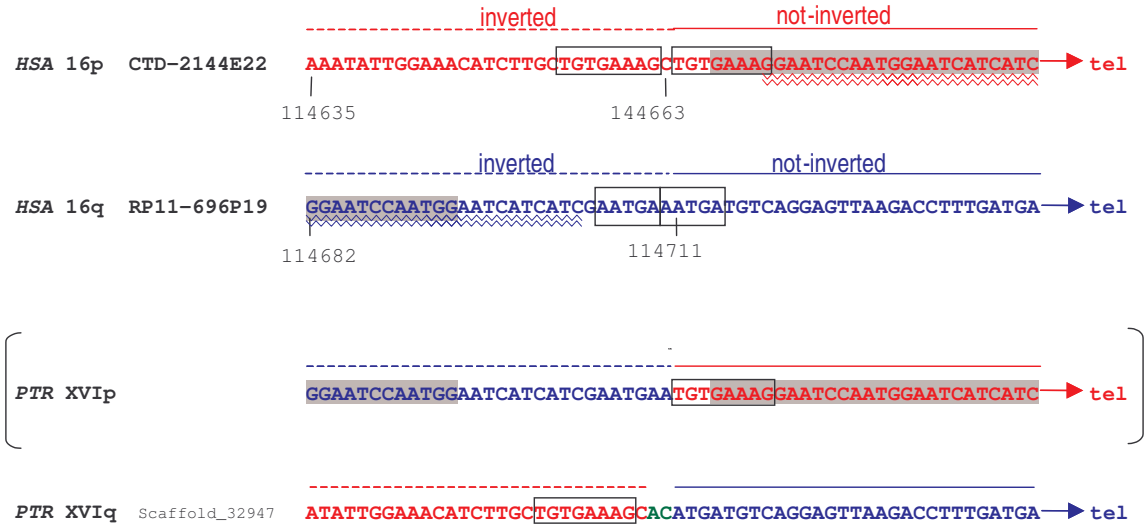


Figure 3.10: Positions of the breakpoints at the nucleotide level. For the sake of simplicity, the orientation of the p-arm has been inverted and the sequence is orientated from the centromere (left) to the telomere (right) and not from the telomere to the centromere. The sequence in blue represents a portion of BAC RP11-696P19 (16q). The sequence in red is derived from BAC CTD-2144E22 (*HSA* 16p). The boxes indicate the relative positions of the octanucleotide repeat that flanks the *HSA*16p breakpoint and the pentanucleotide repeat at 16q. A 22-bp direct repeat found at the breakpoints on chromosomes 16p and 16q is underlined by wavy lines. The HSATII satellite sequences are underlined in grey. The analysis of the chimpanzee sequence in *PTR* XVIq revealed the presence of two additional nucleotides (AC, in green). Parentheses serve to indicate that the breakpoint region at *PTR* XVIp, within the HSATII satellite DNA, was inferred from the human sequence, and was not sequenced in *PTR* owing to the abundance of repetitive satellite sequences.

3.2.1.2. Putative mechanism underlying the inversion

Repeatmasker analysis indicated the presence of long stretches of HSATII satellites in human BACs RP11-696P19 and CTD-2144E22, spanning 70-kb and 20-kb, respectively. Pair-wise sequence comparisons of both BACs further indicated the presence of two inverted repeats of about 23-kb with 97% sequence similarity between them (Figure 3.11A).

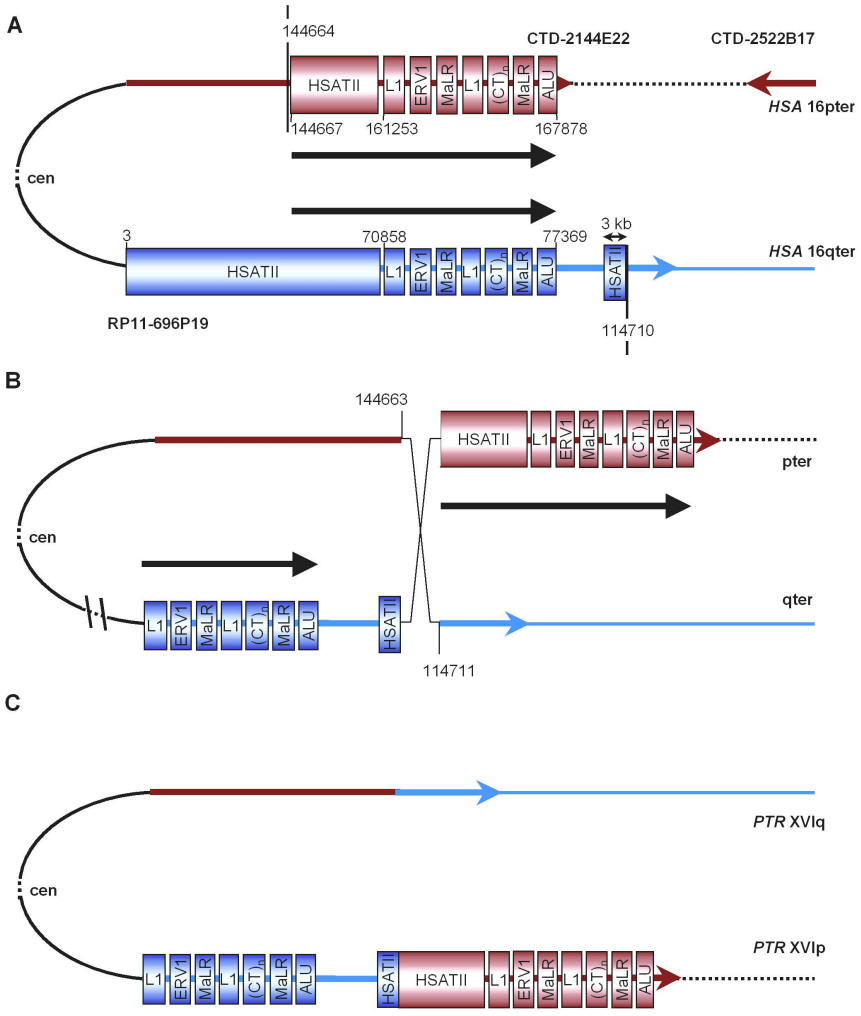


Figure 3.11: Hypothetical model to explain the pericentric inversion of *PTR* XVI homologous to *HSA* 16. (**A**) Breakpoint-spanning BACs RP11-696P19 and CTD-2144E22 are indicated by blue and red lines respectively. The black arrows denote the 23-kb inverted repeat comprising repetitive elements and HSATII satellite DNA. The numbers indicate the nucleotide positions according to the numbering of BAC clones RP11-696P19 and CTD-2144E22. The high degree of similarity conferred by the inverted repeats may have brought the chromosomal regions 16p11-p12 and 16q11-q12 together, thereby mediating an intra-chromosomal recombination event (**B**). (**C**) Schematic of the rearrangement on *PTR* XVI resulting from the recombination between the p- and q-arms. cen: centromere

Given that *HSA* 16 represents the ancestral form of the chromosome, it appears likely that these inverted repeats, together with their flanking HSATII satellite sequences, facilitated the inversion, even although the breakpoints did not occur directly within the repeats (Figure 3.11B, C).

3.2.1.3. Evolutionary history of the breakpoint regions

FISH analysis revealed that *HSA* BAC RP11-696P19 hybridized not only to *HSA* 16, but also to *HSA* 1. To investigate the evolutionary history of the sequences represented by this BAC, I performed FISH analyses on chromosomes of other primates. RP11-696P19 signals were detected on chromosomes homologous to *HSA* 16 and *HSA* 1 in chimpanzee, bonobo and gorilla (Figure 3.12B).
Interestingly, with metaphases from orangutan (*PPY*) and macaque (*MFA*), BAC RP11-696P19 did not hybridize to the *HSA* 16 homologues, but rather hybridized exclusively to the equivalents of *HSA* 1.
The *MFA* I and *PPY* I chromosomes may be distinguished by a pericentric inversion which occurred in the Old World monkey lineage about 25 Mya, after their divergence from the hominoid lineage (Dutrillaux et al. 1979; Maresco et al. 1998). After the separation of the orangutan lineage from that of the other great apes, a duplicational inter-chromosomal transposition to the ancestral chromosome XVIq occurred, which involved the region covered by BAC RP11-696P19. This was followed by the pericentric inversion on *PTR* XVI. The region corresponding to BAC RP11-696P19 was also split by a pericentric inversion on *GGO* XVI, homologous to *HSA* 16 (Figure 3.12A). Human BAC CTD-2144E22 hybridized to the p-arm of the *HSA* 16 homologues of both macaque and orangutan. However, on *PTR* XVI, *PPA* XVI and *GGO* XVI, this BAC displayed a split signal on the p- and q-arms (data not shown). Taken on their own, these findings are not inconsistent with the view that gorilla and chimpanzee share the same pericentric inversion.

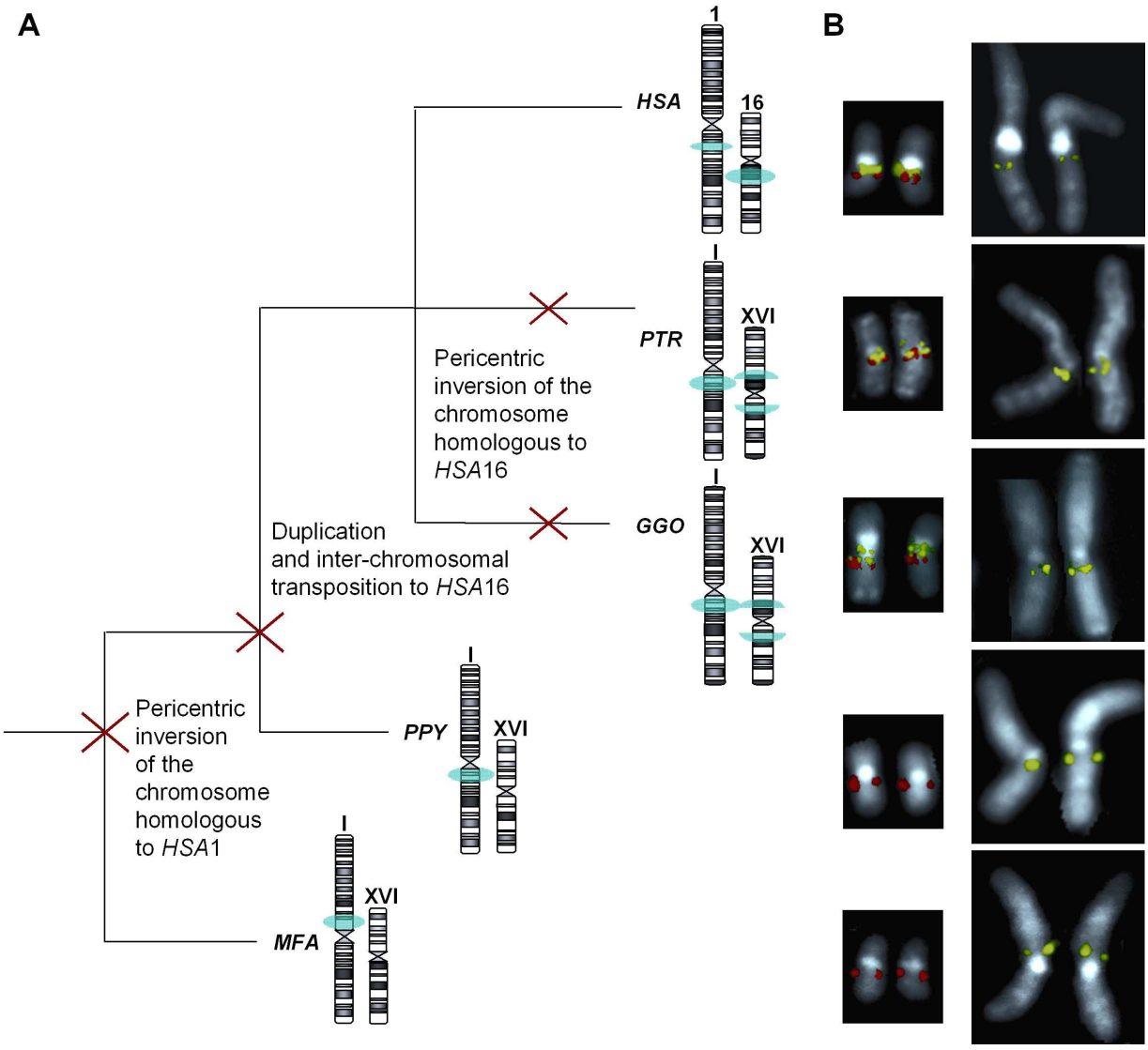


Figure 3.12: FISH analyses designed to follow the evolutionary history of the region represented by BAC RP11-696P19. **(**A) Ideogram of chromosomes I and XVI homologues in *MFA*, *PPY*, *GGO*, *PTR* and *HSA*. The positions of the regions homologous to BAC RP11-696P19, as determined by FISH, are schematically indicated in green. The red crosses indicate the occurrence of the respective rearrangements. (**B**) FISH hybridization pattern of BACs RP11-696P19 (green) and RP11-44I10 (red) from 16q11-12 on chromosome 16 (left) and 1 (right) homologues. In metaphase spreads of *MFA* and *PPY*, signals from RP11-696P19 were found exclusively on chromosome I.

3.2.1.4. Breakpoint position of the GGO XVI inversion

FISH experiments with the human BACs used to narrow down the inversion breakpoint regions in *PTR* XVI yielded hybridization patterns in the gorilla that were comparable to those observed in the chimpanzee (summarized in Table 3.8). This

indicates that gorilla does indeed have an inversion, and that the breakpoints are either identical or at least reside in the same genomic region as in the common chimpanzee.

To determine whether the inversion breakpoints of *PTR* XVI and *GGO* XVI occurred at identical sites, FISH was performed with *PTR* BACs that span the inversion breakpoints on *PTR* XVI. These chimpanzee BACs, RP43-007E19 and RP43-001I03, yielded split signals on *GGO* XVI (data not shown). This finding implies that it is extremely unlikely that the *PTR* and *GGO* inversion breakpoints are identical in terms of their location. As previously mentioned, human BACs CTD-2144E22 and RP11-696P19 are split by the inversion in both *PTR* XVI and *GGO* XVI. It may therefore be concluded that whilst the inversion breakpoints in *PTR* XVI and *GGO* XVI are in close proximity to each other, they are nevertheless non-identical.

To investigate these breakpoints at the nucleotide level, I performed breakpoint-spanning PCR with primer pair P2 (listed in Table 2.1). The primers were designed according to the sequence of *HSA* BAC RP11-696P19 and flank the inversion breakpoint on *HSA* 16q (Figure 3.13). PCR P2 was negative in *PTR*, but positive in *HSA* and *GGO* (data not shown). Thus, the sequence at the q-arm breakpoint must differ in *GGO* XVI from that in *PTR* XVI. By screening a gorilla BAC library with PCR probes (P4-P10), I identified breakpoint-spanning BACs CH255-551C4 (AY822675) and CH255-39D12 (AY822676, AY822677) from the gorilla genome. The positions of these gorilla BAC clones with respect to the breakpoints were confirmed by PCR analysis (P1-P10) [schematically indicated in Figure 3.13] and by sequencing the BAC ends. *GGO* BAC CH255-39D12 mapped distal to the inversion breakpoint in *GGO* XVIq and was positive for PCR markers P9, P3, P1 and P8 (Figure 3.13). This served to confirm the inter-species heterogeneity of the q-arm breakpoints in *PTR* XVI and *GGO* XVI. Gorilla BAC CH255-551C4 was positive for PCR markers P6, P7, P10 and mapped to *GGO* XVIp. Since PCR marker P6 is directly flanked by HSATII repetitive sequences, further localization of the gorilla p-arm breakpoint by PCR was not possible.

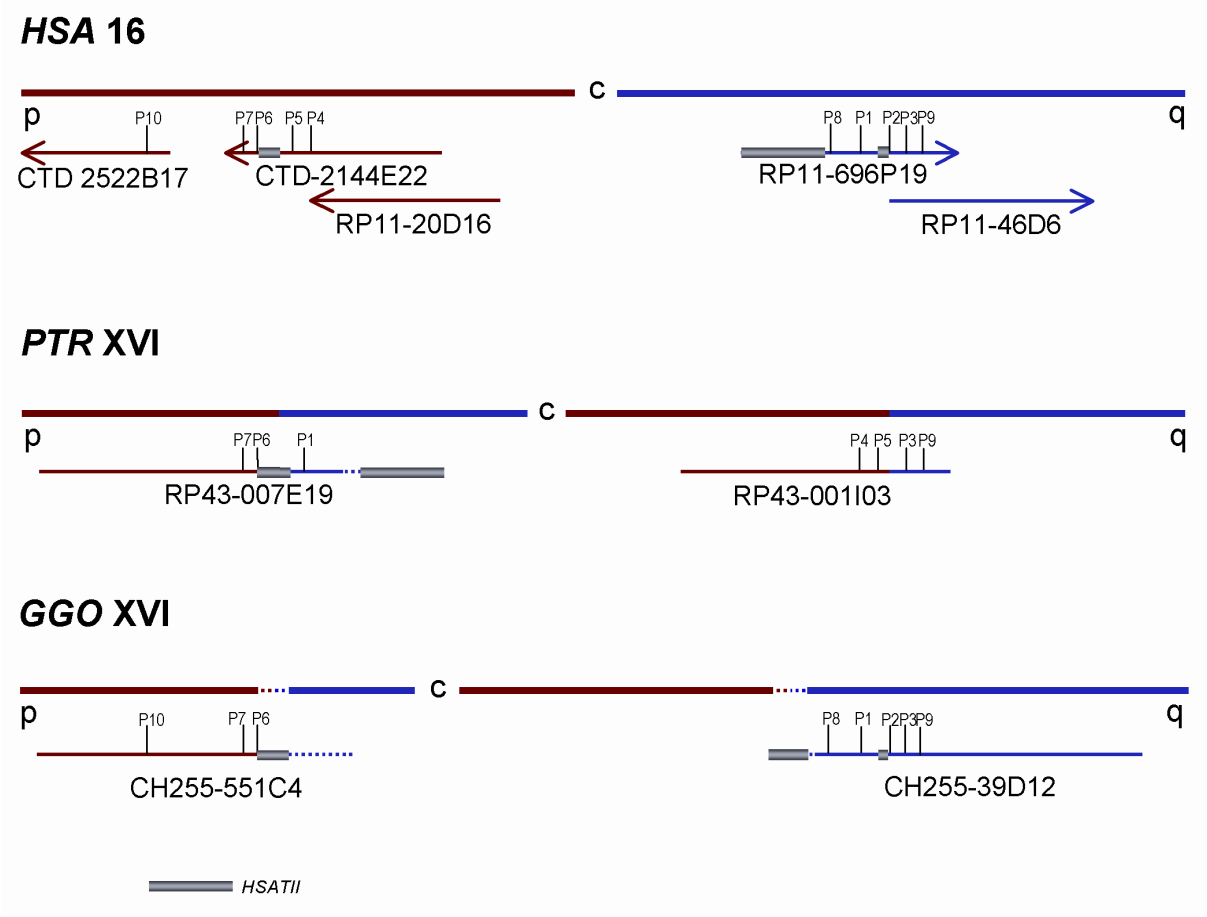


Figure 3.13: Scheme indicating the breakpoint regions of the inversions on *PTR* XVI and *GGO* XVI by comparison with *HSA* 16. The positions of the primer pairs (P1-P10) used to compare the regions in these species are indicated. The dotted lines indicate the regions where the breakpoints of the *GGO* XVI-specific inversion were assigned. HSATII satellites are represented as grey bars. Sequences from the q-arm are indicated in blue and p-arm sequences in red.

I nevertheless conclude that the inversion breakpoint on *GGO* XVIp lies close to or within the HSATII satellites, as found for the breakpoint on *PTR* XVIp. It is likely that the *GGO* XVI inversion was mediated by intra-chromosomal recombination within HSATII, as schematically drawn in Figure 3.14.

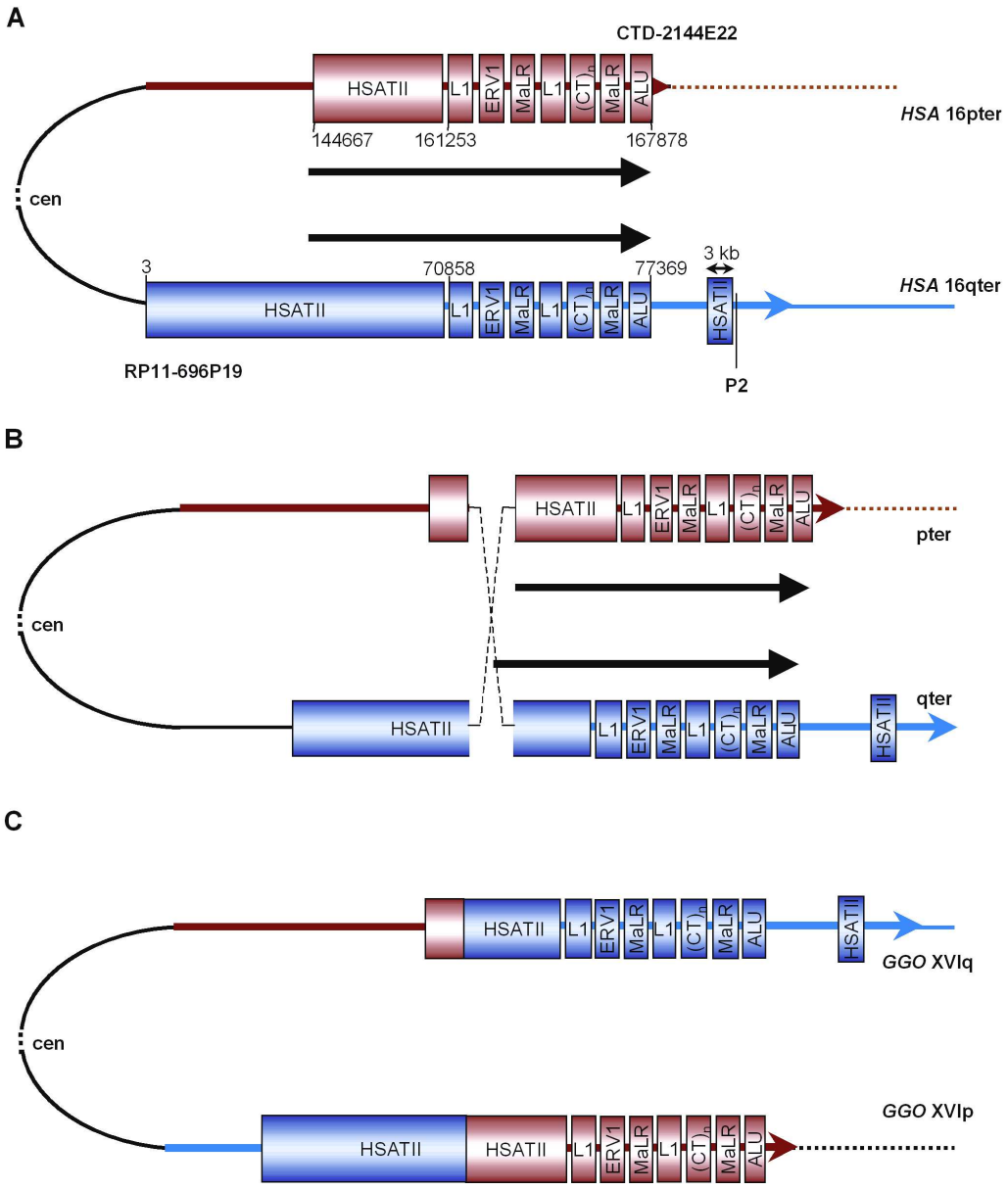


Figure 3.14: Model to explain the pericentric inversion of the gorilla chromosome (*GGO* XVI) homologous to *HSA* 16. P2 indicates the position of the inversion breakpoint on *PTR* XVIq. The numbers indicate the nucleotide positions of BAC clones RP11-696P19 and CTD-2144E22.
(A) Breakpoint-spanning *HSA* BACs RP11-696P19 and CTD-2144E22 are indicated by blue and red lines respectively. The black arrows denote the 23-kb inverted repeats at the breakpoints, which are postulated to have mediated intra-chromosomal recombination underlying the inversion (**B**). The result of this recombination is the inversion on *GGO* XVI as shown in (**C**). cen: centromere.

3.2.2. The human specific pericentric inversion of chromosome 18

3.2.2.1. Analysis of breakpoint spanning BACs by FISH and PCR

FISH analysis with BACs from the very distal contig in 18p11.3 (NT_010859) showed that they hybridise to the q-arm of the chimpanzee chromosome XVIII, homologous to human chromosome 18 (Figures 3.15 and 3.16; Table 3.10). BAC RP11-683L23, which is the most distal BAC of the contig in 18p11.3 displays a complex hybridisation pattern in humans. In addition to a strong signal in 18p11.3, a weaker signal is noticed in 18q11.

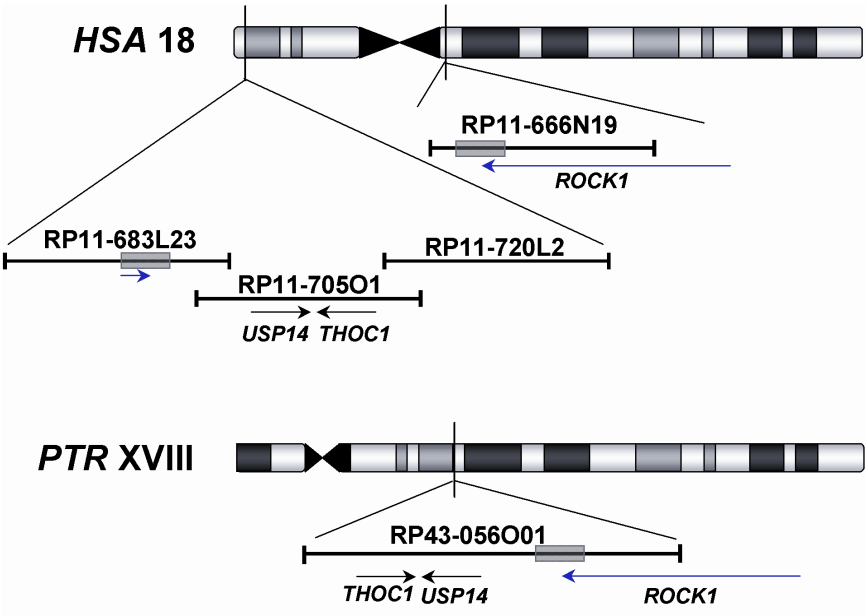


Figure 3.15: Ideogram of human chromosome 18 *(HSA* 18) and the chimpanzee homolog (*PTR* XVIII). The breakpoint spanning human BACs RP11-666N19 from 18q and RP11-683L23 from 18p11.3 are indicated, together with BACs RP11-705O1 and RP11-720L2 which extend the contig NT_010859 proximal to RP11-683L23. Chimpanzee BAC RP43-056O01 is spanning the breakpoint in PTR XVIIIq. The 19 -kb segment, which is duplicated in humans, is shown as a grey rectangle. Genes flanking the breakpoint regions, *ROCK1*, *USP14* and *THOC1,* are indicated by arrows.

Further, cross hybridisations of RP11-683L23 are seen in the pericentromeric region of chromosomes 1, 9, and the acrocentric chromosomes (Figure 3.16A). On

chimpanzee chromosomes, BAC RP11-683L23 hybridises to XVIIIq11 and to the telomeric region of the q-arm (Figure 3.16B).

To find a breakpoint spanning chimpanzee BAC, I performed BLAST analyses with the sequence of BAC RP11-720L2, which extends the contig to the centromeric direction in 18p (Figure 3.15). By this, BAC RP43-056O01 was identified, for which only end sequences were available. The sequence of the telomeric end of BAC RP43-056O01 (AG183404) is matching human BAC RP11-666N19. This BAC is the most proximal one in the Ensembl contig (NT_010966) of 18q11 and contains alpha satellites.

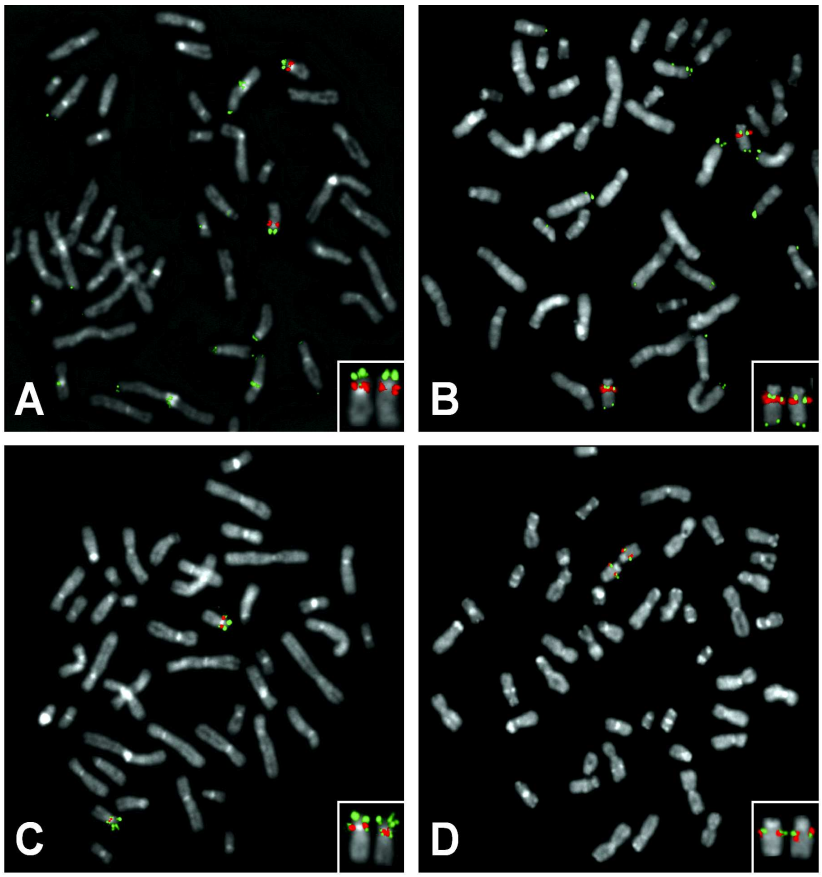


Figure 3.16: Comparative FISH analysis of breakpoint spanning BACs. (**A, B**) Co-hybridisation of BACs RP11-683L23 (green) from 18p11.3 and RP11-666N19 (red) from 18q11 to human (**A**) and to chimpanzee (*PTR* XVIII) (**B**). RP11-683L23 is inverted since it hybridises to 18p11.3 in humans but in chimpanzee to *PTR* XVIIIq11. Weak cross-hybridisation of RP11-683L23 is noted in 18q11 in humans and XVIIIqter in chimpanzee. (**C, D**) Co-hybridisation of chimpanzee BAC RP43-056O01 (green) and human BAC RP11-666N19 (red). RP43-056O01 is spanning the pericentric inversion breakpoint in human, since signals are noticed on 18p11.3 and 18q11 (**C**). On the chimpanzee chromosome XVIII, a single signal of RP43-056O01 is seen in the q-arm (**D**).

Table 3.10: BACs used as FISH probes to identify the breakpoint regions of the pericentric inversion of human chromosome 18 and the homologous chromosome in the chimp

BAC designation	Contig	Position of BAC according to FISH on	
		***HSA* 18**	***PTR* XVIII**
RP11-683L23	NT_010859	p	q
RP11-705O1	NT_010859	p	q
RP11-720L2	NT_010859	p	q
RP11-476K15	NT_010859	p	q
RP11-291G24	NT_010859	p	q
RP11-161I6	NT_010859	p	q
RP11-870F11	NT_010859	p	q
RP11-419P8	NT_010859	p	q
RP11-183C12	NT_010859	p	q
RP11-865B13	NT_010859	p	q
RP11-674N23	NT_010859	p	q
RP11-820I16	NT_010859	p	q
RP11-666N19	NT_010966	q	q
RP11-254G11	NT_010966	q	q
RP11-326M20	NT_010966	q	q
RP11-784B15	NT_010966	q	q
RP11-489I20	NT_010966	q	q
RP11-549B18	NT_010966	q	q

To determine the coverage of chimpanzee BAC RP43-056O01 with respect to the human BACs, PCR analysis with primer pairs P11-P22 was performed (Figure 3.17). The primers were designed from the sequence of human BACs RP11-683L23 and RP11-705O1. PCR-products P11-P16 were amplified from genomic DNA of the chimpanzee, but not from BAC RP43-056O01 (not shown). This indicates that the corresponding region is not covered by BAC RP43-056O01. However, the segment between PCR markers P17-P22 is represented by this BAC.

FISH analysis showed that RP43-056O01 is spanning the breakpoint in *PTR* XVIIIq. While a single signal of this BAC is seen in *PTR* XVIIIq (Figure 3.16D), BAC RP43-056O01 gives signals in *HSA* 18q11 and 18p11.3 (Figure 3.16C).

FISH analysis did not reveal a split signal of BAC RP11-666N19 on chimpanzee chromosome XVIII, suggesting that the breakpoint maps to the most proximal part of this BAC and thus escapes the detection by FISH or is located in the pericentromeric heterochromatin (Figure 3.16D). On human chromosome 18, RP11-666N19

hybridises to q11 (Figure 3.16A, C) and weak cross hybridisation is noticed in 18p11.3 (not shown).

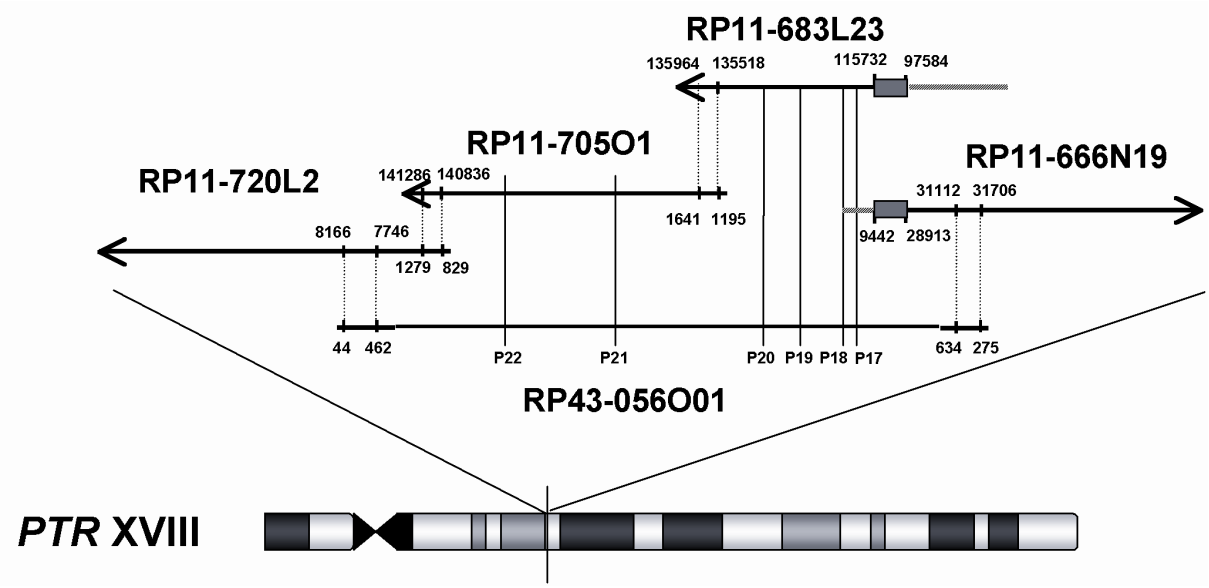


Figure 3.17: Coverage of chimpanzee BAC RP43-056O01, for which end sequences were available, with respect to the human BACs RP11-666N19 from 18q11 and RP11-683L23, RP11-705O1, and RP11-720L2 from 18p11.3. Numbers indicate the positions of overlap, P17 to P22 mark the position of PCR products, which were amplified from RP43-056O01 and the respective human BAC. The dotted horizontal lines represent the regions of BAC RP11-683L23 and RP11-666N19, which are not homologous to the segment covered by the chimpanzee BAC.

3.2.2.2. Sequence comparisons of breakpoint spanning BACs and characterization of the duplicon

As already suggested by the weak co-hybridisation of BAC RP11-683L23 in 18q11 and that of BAC RP11-666N19 in 18p11.3, sequence alignments confirmed a 19-kb duplicated segment. BLAST analysis of RP43-056O01 end sequences against the draft sequence of the chimpanzee genome (Ensembl browser released in December 2003) showed that the sequence of RP43-056O01 matches to position 11057822 - 11240960 in scaffold 37707. Comparisons of these 183-kb of chimpanzee sequence to human BACs RP11-683L23 and RP11-666N19 revealed that the 19-kb segment between scaffold positions 11218740 and 11237850 is duplicated in the human

genome as homologous sequences are found in RP11-683L23 from 18p11.3 and RP11-666N19 from 18q11. Further duplicates of this segment on chromosome 18 or other chromosomes were not detected by BLAST analyses against human and chimpanzee sequences.
The 19-kb segment, which has been duplicated during human evolution (nucleotide positions 96584 to 115732 of BAC RP11-683L23 and 9442-28913 of BAC RP11-666N19 contains 50% of repetitive elements, mainly of the SINE class, and 9.6-kb of the 3' part (exons 29 to 32 and 3' UTR) of the *ROCK1* gene encoding for the Rho-associated, coiled-coil containing protein kinase 1 (NM_005406). The CG-content of the duplicon amounts to 39%-40%.

3.2.2.3. Proposed mechanism underlying the inversion

Since these 19-kb duplicates in the human genome are inverted to one another, the most plausible mechanism underlying the inversion is intrachromosomal homologous recombination between inverted repeats (Figure 3.18). This model presupposes that during human evolution the 19-kb segment was duplicated and one of the duplicates has been transposed to 18p11.3 (Figure 3.18B). Most probably, non-allelic homologous recombination between these inverted duplicates, which show an average identity of 97%, facilitated the inversion. Genome wide BLAST analysis with the sequence of BAC RP11-683L23 distal to the duplicon, from nucleotide positions 1- 96584 (according to AP001005), yielded matches of segments of 1-kb to 10-kb on many other chromosomes (data not shown). This is also confirmed by FISH analysis as BAC RP11-683L23 shows co-hybridisation with pericentromeric regions of chromosomes 1, 9 and the acrocentric chromosomes (Figure 3.16A).

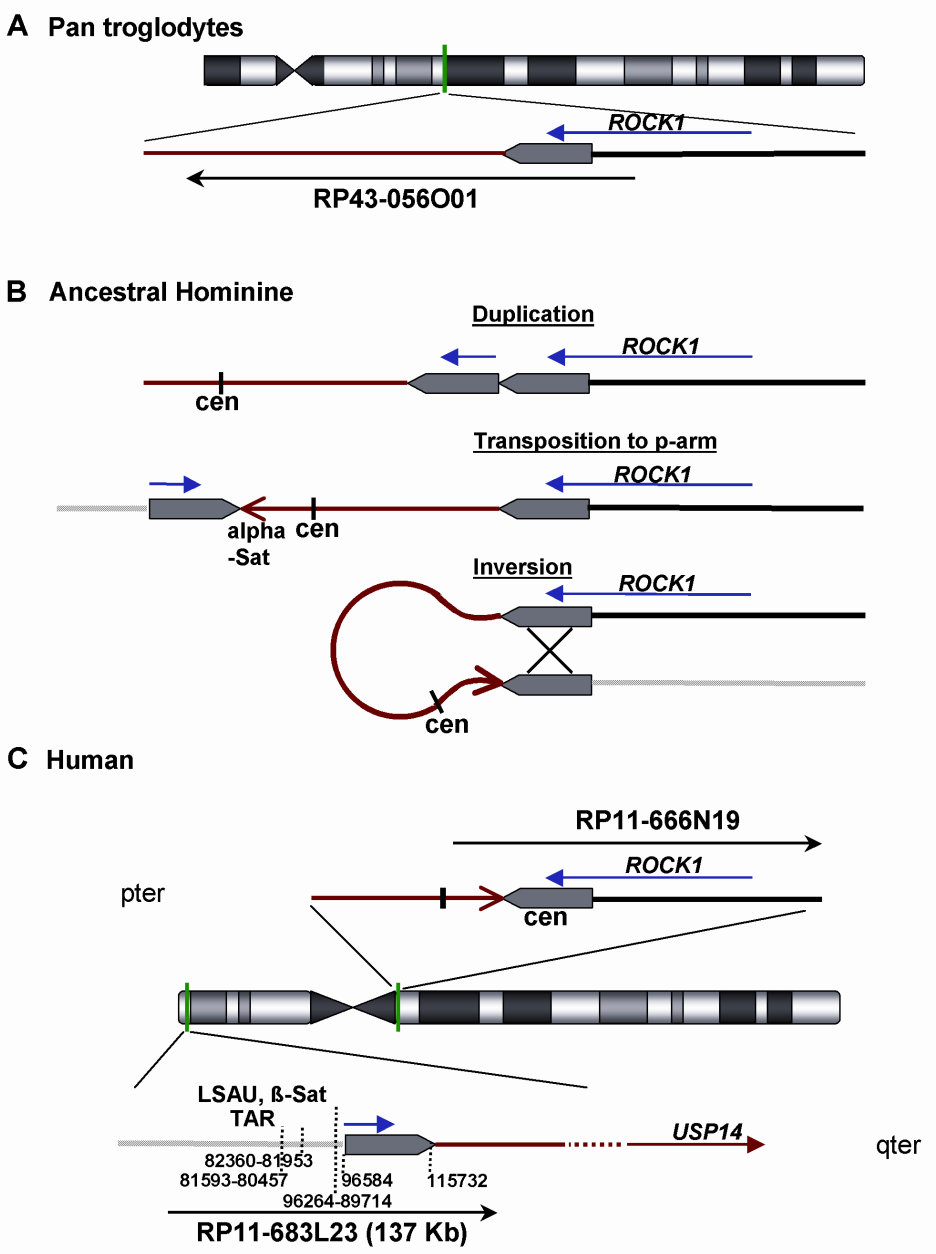


Figure 3.18: Hypothetical model to explain the pericentric inversion of human chromosome 18. (**A**) Chimpanzee BAC RP43-056O01 spans the breakpoint in the long arm of the chimpanzee chromosome homologous to human chromosome 18. The 19-kb segment (grey arrow) has been locally duplicated during human evolution, after the separation of the chimpanzee and human lineage (**B**). The duplicate contains 10-kb of the 3' part of the *ROCK1* gene marked by a blue arrow. Subsequently, the duplicate has been transposed to the p-arm of the ancestral chromosome 18. Homologous recombination between the inverted duplicons is assumed to underlie the inversion of the red segment between the duplicates. The red arrow head indicates alpha-satellites. (**C**) Scheme of the pericentric inversion of human chromosome 18 as a result of recombination between the duplicons. The numbers above the arrow representing BAC RP11-683L23 indicate the position of the 19-kb duplicon, LSAU- and ß-satellites and TAR (telomere associated repeats) according to the base pair numbering of AP001005.

4. DISCUSSION

DNA copy number gains or losses have been shown to play an important role in genomic instability and disease (Ji et al., 2000a). However, besides being implicated in human disorders, segmental aneusomy is also a driving force for evolution. Gene duplications have been suggested to be the key mechanism for the evolution of new biological functions (Ohno, 1970). It should be noted that losses of genes have also undergone positive selection during evolution (Olson et al., 1999; Olson and Varki, 2003; Stedman et al., 2004; Vallender and Lahn, 2004b). Moreover, several studies have highlighted the presence of low copy repeats (LCRs) at breakpoints of large-scale chromosomal rearrangements that distinguish human from primate chromosomes, suggesting their implication in the rearrangements and therefore in the genome evolution (Shaw and Lupski, 2004; Samonte and Eichler, 2002). In this study, I describe the importance of such large-scale losses and gains of DNA within the genome of human as well as of five non-human primates (macaque, orangutan, gorilla, chimpanzee and bonobo), emphasizing those changes that occurred specifically in the human lineage. I also characterized two evolutionary pericentric inversions described previously by Yunis and Prakash (1982). I studied the duplication content of the breakpoint regions in order to verify the implication of LCRs in the occurrence of these two rearrangements.

4.1. Characterization of DNA copy number differences between human and non-human primates

Large-scale copy number differences between hominoid species have been detected using comparative microarray technologies, both cDNA and BAC-based (Locke et al., 2003b; Fortna et al., 2004). In the present study, I have extended these initial comparisons by using a 6-K BAC array that covers about 30% of the human genome sequence with one BAC clone every 500-kb. In pairwise comparisons between humans and five different primate species, a total of 322 sites showing a copy number difference (CND) were detected (Figure 3.1). The disparity between the

number of genomic rearrangements depicted in Figure 3.1 and the number of variant sites is because 23 sites displayed duplications in one species and deletions in another. In such cases, the rearrangements were considered distinct and independent events. Conspicuously, 19 of these 23 loci displayed a deletion in macaque, which could be a duplication that occurred in the common ancestor of the hominoids. I cannot rule out, however, that some of these deletions are artefacts caused by the sequence divergence of 4.9% between macaque and human. Indeed, a recent study has shown that sequence divergences have an impact on the results when doing interspecies gene expression comparisons (Gilad et al., 2005a). Using a new multiprimate cDNA array containing probes from human, chimpanzee, orang-utan and rhesus, they have found an effect of sequence divergence on the hybridization signal, even in the closest pair of species (human and chimpanzee). Interestingly, four of these 23 “CND hotspots” overlap with known human chromosomal polymorphisms (Iafrate et al., 2004; Sebat et al., 2004; Sharp et al., 2005; Tuzun et al., 2005). These features reflect the great plasticity of the corresponding loci during primate evolution and, for some of them, this plasticity is clearly still evident in the human lineage.

As already observed in the initial inter-species aCGH study performed by Locke et al. (2003b), lineage-specific CNDs clearly predominate with respect to the shared gains and losses. Interestingly, 17 CNDs are common to the macaque and orangutan genomes in comparison to the African apes and to the human species, members of the subfamily “homininae” (Figure 1.1 and Figure 3.1A). These 17 CNDs probably originated in the common ancestor of the hominines.

The frequency of CNDs among the great ape species studied did not correlate with the estimated time of divergence of the respective species from the common ancestral lineage. Indeed, the highest number of copy number gains was observed in gorilla (Figure 3.3). A lower rate of difference was found in orangutan (as compared to human) than in gorilla, despite the orangutan having separated 6 million years earlier than gorilla from the common ancestral lineage. At first sight, this finding would seem to contradict the results of Fortna et al. (2004), who noted that gene copy number gains were more frequent in humans than in the great apes. In their study, gorillas did not exhibit a disproportionately high rate of expansion in gene number. However, Fortna et al. (2004) performed inter-hominoid aCGH with cDNA arrays, whereas in my analysis, genomic clones were investigated. In order to assess

whether this bias could be associated with gene density, I analyzed the gene content of 20 arrayed BACs that detected gorilla-specific gains. The resulting gene content was then compared with those of 60 randomly chosen BACs distributed throughout the human genome (Appendix 3). BACs that detected gorilla-specific gains were found to be significantly less gene-rich than the randomly chosen BACs. I may therefore conclude that the gorilla-specific gains identified here occurred preferentially in gene-poor regions. The disproportionately high level of gorilla lineage-specific duplications may reflect a characteristic of the gorilla genome that has been unidentified until now. Consistent with this view are the findings of Fan et al. (2002) who characterized the ancestral fusion region on human chromosome 2q14, a region that is rich in segmental duplications. BAC RP11-432G15 maps to this region and its copy number was highly elevated in gorilla as compared to the other primates tested. In my study, this BAC also displayed a gorilla-specific copy number gain. The higher rate of copy number increases in gorilla as compared to orangutan correlates with the frequency of gross level chromosomal rearrangements in these species. Whereas the gorilla karyotype manifests a variety of derived chromosomal changes, the orangutan karyotype appears to be very similar to the putative ancestral organization (Müller and Wienberg, 2001).

4.2. Human-specific copy number differences

In this study, I have focussed on those CNDs, which occurred specifically in the human lineage (Human Lineage-Specific HLS-CNDs). In contrast to Locke et al. (2003b), who identified a single example of an HLS-CND, the higher resolution array used in my study was successful in detecting a total of 14 putative HLS-CNDs that are present in fewer copies in the other primate genomes examined (Table 3.2). It is quite unlikely that these 14 CNDs would represent false positives since losses of the respective sites were detected in all five primate species investigated. It should however be noted that only human genomic sequences are represented on my 6K-array. Consequently, if some sequences were to have been lost specifically in the human lineage, they would not have been evident from my analysis.

One of the 14 putative HLS-CND sites contains human-specific centromeric satellite sequences from chromosome 10 that are absent in the other primate species examined. It has been shown that centromeric regions consist of DNA rich in α and β satellites, classical satellites (Sat I-Sat IV), and less abundant repeats such as CAGGG (Eichler et al., 1999; Jackson, 2003). Numerous studies have shown that a subset of α satellites has undergone rapid evolution. Indeed, they have highlighted that alphoid DNA is not only chromosome-specific but also exhibit greater sequence identity within a species than between species (Willard and Waye, 1987; Archidiacono et al., 1995; Rudd et al., 2004, 2006). The BAC RP11-81J3 showing only signals on the centromere of *HSA* 10, my findings are consistent with these observations.

Another arrayed BAC, which detects HLS-CND, the clone RP11-35B4, was not included in the Human Genome Segmental Duplication Database since it has not yet been fully sequenced. However, the remaining 12 HLS-CNDs overlap with known segmental duplications and are grouped in duplication clusters (DCs) larger than 40 kb in size. A total of 15 DC families containing putative HLS-CNDs were identified (Table 3.2). The DCs detected by this aCGH analysis exhibit high sequence similarity (>90%) to homologous DCs located at other sites in the human genome (Appendix 2). Recently, Cheng et al. (2005) compared the human genome reference sequence (build 34) with the draft sequence of the chimpanzee genome and classified segmental duplications as (i) detected only in chimpanzee, (ii) detected only in human or (iii) shared between both species. I traced, within build 34, the HLS-specific DCs detected by aCGH and mapped them within the improved build 35 to investigate their status according to the classification of Cheng et al. (2005). Five HLS DCs were classified as detected only in human (Table 4.1). My aCGH analyses using gorilla, orangutan and macaque as outgroups support the conclusion that these are indeed sites of human lineage-specific gains. Two other HLS DCs overlap with sequences showing duplications in human only and also duplications shared by both lineages. The remaining eight HLS-CNDs were found to be shared by chimpanzee and human in the study of Cheng et al. (2005). The human-specific gains are probably duplications of regions that are highly homologous to those represented on my array.

Table 4.1: Comparison of the 14 sites of putative HLS-CND identified in my study with the results of Cheng et al. (2005)

HLS duplications identified in my study	Position in the human genome (Mb)[b]	Compared to the results of Cheng et al. (2005)
DC0154	120.4 - 120.6	Human only
DC0203	144.9 – 145.0	Duplication shared by PTR and HSA[c]
DC0204	145.0 – 145.1	Duplication shared by PTR and HSA[c]
DC0206	145.3 – 145.4	Duplication shared by PTR and HSA[c]
DC0207	145.5 – 145.7	Duplication shared by PTR and HSA[c]
DC0218	146.5 – 146.6	Human only
DC0445[a]	111.7 – 111.9	Human only
DC0447[a]	112.0 – 112.1	Human only
DC0448	112.2 – 112.3	Human only and duplication shared by PTR and HSA[d]
DC2828	67.3 – 67.7	Duplication shared by PTR and HSA[c]
DC3093	45.9 – 46.1	Human only
DC4092	18.8 – 19.3	Human only and duplication shared by PTR and HSA[d]
DC3591	32.6 – 33.1	Duplication shared by PTR and HSA[c]
DC1786	14.4 – 14.6	Duplication shared by PTR and HSA[c]
DC1787	14.6 – 14.8	Duplication shared by PTR and HSA[c]

[a]DC0445 and DC0447 overlap each other in Build 34. However there are distinct loci in Build 35.

[b]According to the hg17 assembly

[c]Since these DCs are shared duplications, the human lineage specific gains are likely to affect one of the homologous DCs listed in Appendix 2.

[d]The DCs overlap with sequences showing duplications in human only and also duplications shared by both lineages.

4.3. Genes contained in the putative HLS-CNDs

Most of the DCs that detect HLS-CNDs contain genes that consequently vary with respect to their dosage in humans as compared to the other primates examined (Table 3.3). Although many of these genes remain uncharacterized, with no information available as to their biological function, some correspond to known genes. In this latter category are members of the histone and the olfactory receptor (OR) gene families. In hominoids, the OR gene family contains several hundred functional members (Gilad et al., 2003) and for certain subsets of these genes, lineage-specific positive selection has been demonstrated (Gilad et al., 2005b). A

human-specific increase in copy number was also evident for the *ANAPC1* gene which encodes the anaphase promoting complex subunit 1, an important meiotic checkpoint regulator (Jorgensen et al., 2001) (Table 3.3).

My study has also identified a human-specific increase in the copy number of the aquaporin 7 gene (*AQP7*) which encodes a protein involved in water transport across membranes. BAC RP11-79A9, recognizing an aCGH-identified HLS-CND and localized on HSA9q, contains the *AQP7*-like gene, which is a functional copy of the *AQP7* gene located on HSA9p13. Human-specific copy number increases of the *AQP7* gene have previously been observed by both Bailey et al. (2002a) and Fortna et al. (2004). Indeed, Fortna et al. (2004) identified a total of seven copies of the *AQP7* gene in the human genome; these are highly similar to each other with at least four appearing to be potentially functional.

Among the genes included within the HLS-CNDs identified here are *VDUP1* and *FCGRIA* which encode proteins involved in the response to oxidative stress and in the immune response, respectively. The expression of these genes is known to be triggered by environmental stimuli. Interestingly, genes whose expression is responsive to environmental stimuli have previously been found to be over-represented among genes associated with copy number polymorphism (Tuzun et al., 2005). Thus, these genes might encode proteins that facilitate adaptation to new environmental conditions and could therefore have been subject to selection during evolution.

It should be noted that only *FLJ20719*, one of the genes represented on the Affmetrix array U95 (Table 3.10) showed an up-regulation in human cortex when compared to chimpanzee cortex (Enard et al., 2002b; Karaman et al., 2003; Caceres et al., 2003; Uddin et al., 2004). No gene expression differences between human and chimpanzee were observed for the remaining genes.

4.4. Polymorphisms vs. differences fixed in the human population

Several recent studies have identified large-scale polymorphisms (inversions, deletions, duplications) in the human population, showing that the arising of

inversions or segmental aneusomy during evolution is still going on in the human population (Iafrate et al., 2004; Sebat et al., 2004; Feuk et al., 2005; Sharp et al., 2005; Tuzun et al., 2005; Szamalek et al., 2006). In order to determine whether the HLS-CNDs identified in my study are fixed in the human population, I compared my results with those from different groups that investigated large-scale polymorphisms in the human population.

Recently, Tuzun et al. (2005) established a fine-scale higher resolution variation map of the human genome by comparing the human genome reference sequence with a second genome assembly based on fosmid ends. They identified 297 sites that displayed structural variation (insertions, deletions and inversions). I have compared the chromosomal localization of these 297 reported variations with those of the 14 HLS-CNDs identified here. However, no overlap was observed between the sites of the HLS-CND that I describe and those sites identified by Tuzun et al. (2005) as being deleted or inserted.

In their landmark aCGH-based studies, Iafrate et al. (2004), Sebat et al. (2004), and Sharp et al. (2005) demonstrated the presence of large-scale CNDs in humans. They detected 255, 76, and 119 such sites, respectively. The gains and losses identified by these studies are summarized in the Database of Genomic Variants (http://projects.tcag.ca/variation/). According to this database, three of the 14 HLS-CNDs identified in my study represent polymorphic CNDs in the human genome. A further two HLS-CNDs are very similar to DCs that contain polymorphic loci (Table 3.6). Thus, at least five of the 14 HLS-CNDs identified here may represent polymorphic variants in the human genome. However a clear distinction between polymorphic variants and lineage specific changes is not possible based on the data available since both types might be found among the families of highly homologous DCs.

Studies comparing the genome diversity of different humans as well as comparisons with an improved version of the chimpanzee genome would be necessary to address this.

However, it should be kept in mind that not much is known about the structural variants within the human genome. Indeed, three recent studies have shown that our knowledge about the genomic variation that exists in the human genome is only the tip of the iceberg (Conrad et al., 2006; Hinds et al., 2006; McCaroll et al., 2006).

By using a new method that uses SNP genotype data from parent-offspring trios from the International HapMap Project, Conrad et al. (2006) have identified polymorphic deletions of 5-kb and larger. Moreover, Hinds et al. (2006) characterized 215 putative polymorphic deletions ranging from 70-bp to 10-kb, using a high-density oligonucleotide array encompassing about 600-Mb of human DNA sequence. Finally, McCaroll et al. (2006) have developed a method to screen deletion variants using the SNP genotyping data from the HapMap Project, in a similar manner to the study made by Conrad et al. (2006). Finding clustered sites that show Hardy-Weinberg disequilibrium and other genotyping errors, they identified 541 deletion variants, with a median lengths of 7-kb. In summary, these three studies, using SNP genotyping platforms, have observed polymorphic losses in the human genome, with median lengths ranging from 500-bp to 10.5-kb (Conrad et al., 2006; Hinds et al., 2006; McCaroll et al., 2006). Surprisingly the sites of deletion identified in these three studies do not overlap with each other, and only ~10 % of these deletion sites are shared with the ones identified by Iafrate et al. (2004), Sebat et al. (2004), Sharp et al. (2005), and Tuzun et al. (2005). This adds another ~1000 sites of polymorphic deletions to the catalogue of structural variants within the human genome (Eichler, 2006).

4.5. Presence of evolutionary young repeat sequences at the boundaries of the DCs identified

In order to obtain insight into the mechanism of spreading of the DCs identified in my study, I examined their flanking sequences for common sequence features. These borders turned out to be enriched for different repeat sequences (Figure 3.4). This is consistent with the results of Bailey et al. (2003) and Horvath et al. (2003), who showed that several types of repetitive sequence elements are abundant at the junctions of human segmental duplications.

The boundaries of the DCs identified in my own study were enriched for one of the three major subfamilies of *Alu* elements. Indeed, *Alu*Y repeats, the youngest *Alu* subfamily (Shen et al., 1991; Kapitonov and Jurka, 1996), occur disproportionately at the junctions of the DCs analyzed. *Alu* elements are members of the SINE (Short

Interspersed Element) family retroposons that amplify in primate genomes through a process called retropositions. It has been shown that some young *Alu* subfamilies have amplified so recently that they are absent from the genomes of non-human primates (Batzer and Deininger, 2002). The sequencing of the chimpanzee genome highlighted that *Alu* elements have been threefold more active in humans than in chimpanzees (7082 human lineage-specific Alu insertions compared to 2340 which are chimpanzee lineage-specific) (Chimpanzee sequencing and alnalysis consortium, 2005). According to the FISH data and the degree of sequence homology, the DCs identified here appear to have originated fairly recently during primate evolution. There is therefore some concordance between the age of these duplications and the timing of the burst in *Alu*Y formation after the separation of the great apes from the Old World monkeys (Jurka, 2004).
The amplification of this retroelement is therefore consistent with homology-based recombination having played a role in the spread of the segmental duplications identified in my analysis.

4.6. Non-random distribution of CNDs

4.6.1. CNDs localized in regions of chromosomal instability

The majority of HLS-CNDs identified in this study are significantly enriched for copy number gains of sequences wich are already duplicated in the human genome (χ^2 = 4.501, 1 df, $p<0.05$). This phenomenon was also observed in a recent study using a full-coverage array and comparing the genomes of human, chimpanzee and gorilla (Wilson et al., 2006). The non-random association of inter-hominoid copy number differences and segmental duplication has already been observed by Locke et al. (2003b).
Although the fraction of segmentally duplicated regions amount to only 1.2% and 2.9% of the mouse and rat genomes, respectively (Cheung et al., 2003; Tuzun et al., 2004), duplicated segments comprise ~5 % of the human genome (Bailey et al., 2001, 2002a). Thus, segmental duplication would appear to be an ongoing process which has occurred throughout primate evolution, with duplicated regions exhibiting

significant evolutionary instability (Bailey et al., 2002b). Previous reports have shown that recent segmental duplications are positionally biased in that they tend to be found more frequently in pericentromeric or subtelomeric regions (Bailey et al., 2001; Samonte and Eichler, 2002) and appear to correlate with breaks of synteny between the mouse and human genomes (Armengol et al., 2003). Consistent with these observations, ~40 % of the CNDs that I identified in my study mapped to the subtelomeric and pericentromeric regions (Figure 3.2).
Some of the putative HLS-CNDs detected here occurred in the vicinity (from a few hundred kb up to 2- 3 Mb) of genomic regions where rearrangements are causes of developmental disorders, e.g. familial hypertrophic cardiomyopathy 1 (CMH) on *HSA* 14, Angelman and Prader Willi syndromes on *HSA* 15, cat eye syndrome and DiGeorge/velocardiofacial syndrome on *HSA* 22. Although I would not go so far as to claim that HLS-CNDs are directly implicated in the genesis of these conditions, their presence may nevertheless be a reflection of the intrinsic instability of these genomic regions (Shaw and Lupski, 2004).

4.6.2. CNDs localized in the vicinity of evolutionary chromosomal breakpoints

Several examples exist which demonstrate the important role that segmental duplications played during genome evolution in higher primates. Indeed, the fusion site of two acrocentric chromosomes, which gave rise to human chromosome 2, is surrounded by segmental duplications, also termed low-copy repeats (LCRs) (Fan et al., 2002). Additionally, the gorilla unique translocation t(4;19) occurred in regions of LCRs and is associated with a duplication of a genomic fragment homologous to the human chromosomal region surrounding the proximal CMT1A (Charcot-Marie-Tooth)-LCR (Stankiewicz et al., 2001). Further evidence that evolutionary rearrangements have been influenced by genomic architecture is provided by the pericentric inversion between human chromosome 15 and the homologous chromosome in the chimpanzee. One breakpoint of this inversion has been mapped to a segmental duplication cluster in 15q11-q13 (Locke et al., 2003a). The pericentric inversion of *PTR* XII was also associated with segmental duplications (Kehrer-

Sawatzki et al., 2005a; Shimada et al., 2005) and the pericentric inversion of *PTR* IV was mediated by inexact inverted repeats (Kehrer-Sawatzki et al., 2005b).
Interestingly, four of the HLS CND-detecting BACs identified in this study map close to breakpoints of human-specific karyotypic changes, e.g. the inversion inv(2)(p11.2q13), which is polymorphic in humans. HLS CND-detecting BACs RP11-68E19 and RP11-397E13 are located at the respective inversion breakpoints as determined by FISH analyses on metaphase chromosomes of four clinically unaffected individuals with this inversion in the heterozygous state (data not shown). Duplicated regions of 266 kb and 77 kb (covered by the arrayed BAC RP11-68E19) occur on *HSA* 2q13 and exhibit a high degree of sequence similarity to BAC RP11-397E13 on 2p11.2. I propose that the high similarity between these segmental duplications at 2q13 and 2p11.2 could have mediated the inversion of chromosome 2. Interestingly, the duplications at 2q13 and 2p11.2, detected by BAC RP11-68E19, are polymorphic in human populations (Sebat et al. 2004). It is thus possible that the presence of these duplications in some individuals will serve to promote the occurrence of inv(2)(p11.2 q13).
Further, CND-detecting BACs RP11-439A17 and RP11-196G18 are located near the breakpoints of the human-specific pericentric inversion of chromosome 1 and both BACs contain highly similar inverted segmental duplications (unpublished results). The human-specific duplications could have mediated, or at least facilitated, the chromosomal rearrangements of *HSA* 1 and *HSA* 2.

4.6.3. Pericentric inversions of chromosomes 16 and 18, further examples of rearrangements driven by LCRs

It has been suggested that the abundance of duplicated sequences gives rise to genomic instability via unequal crossing over during meiosis and may thus serve to drive evolution (Eichler, 1998; Samonte and Eichler, 2002).
Owing to the results obtained from the interspecies aCGH, which indicate the presence of LCRs close to large-scale chromosomal rearrangements, I studied for their duplication content, the breakpoints regions of two pericentric inversions that occurred during hominoids evolution.

By FISH and comparative sequence analyses of breakpoint-spanning BAC clones from the human, chimpanzee and gorilla genomes, I have characterized the breakpoints of the pericentric inversion of *PTR* XVI and *GGO* XVI, homologous to *HSA* 16, and of *HSA* 18.

4.6.3.1. *HSA* 16, a chromosome propitious to instability

The high concentration of paralogous segments within the pericentromeric region of chromosome 16 and its evident plasticity are reflections of its likely importance in an evolutionary context. Further evidence for the instability of proximal 16p and 16q comes from reports of pathological breakpoints in these regions. Thus, the 16p11 and 16q11-12 regions are known to be involved in recurring rearrangements in acute myeloid leukaemia (Panagopoulos et al., 1994; Sharma et al., 1999; Yamamoto et al., 2001), Ewing's tumour (Shing et al., 2002), lymphoma (Ueda et al., 2002) and liposarcoma (Willeke et al., 1998).
Human chromosome 16 is known to manifest one of the highest levels of segmentally duplicated sequences among the autosomes (Martin et al., 2004). Approximately 10% of chromosome 16 consists of segmental duplications whilst nearly 9% of genome-wide human duplication alignments map to this chromosome. It should also be noted that as much as 7% of the mass of human chromosome 16 has been added by segmental duplication events occurring within the last 10 Myrs (Martin et al., 2004). In addition, the 16p11 pericentromeric region harbours a high concentration of inter-chromosomal duplications, accounting for 55% of all chromosome 16 inter-chromosomal alignments (Martin et al., 2004). Finally, large tracts of interstitial alpha-satellite DNA have been identified within proximal 16p11 and it is possible that such sequences have played an important role in the frequent evolutionary inter-chromosomal exchange of pericentromeric DNA (Martin et al., 2004).

4.6.3.2. The pericentric inversion of *HSA* 18, a human-specific evolutionary event

In this study, I also investigated the pericentric inversion of human chromosome 18, which occurred after the separation of chimpanzee and human from a common ancestor, 5-6 myrs ago. Since the inversion is not found in gorilla neither in

orangutan (Yunis and Prakash, 1982), the rearrangement is specific for the human lineage. The pygmy chimpanzees or bonobos (*Pan paniscus*), which separated about 1.8 myr ago from the chimpanzee (*Pan troglodytes*) (Yoder and Yang, 2000; Yu et al., 2003), equally do not have the inversion as shown by FISH analysis with a breakpoint spanning YAC (McConkey, 1997). Therefore, breakpoint analysis of the pericentric inversion of chromosome 18 promises to identify genomic regions, which were rearranged specifically during human evolution.

4.6.3.3. LCRs at the boundaries of both pericentric inversions

Detailed analyses of pericentromeric regions have revealed that paralogous sequences are often located in close proximity to satellite sequences (Horvath et al., 2000, 2001; Jackson, 2003). In the case of chromosome 16p, paralogous sequences extend over several megabases in the centromeric region (Eichler et al., 1996; Martin et al., 2004). My findings are consistent with the pericentric inversions of *PTR* XVI and *GGO* XVI having been mediated by recombination between paralogous sequence blocks and stretches of HSATII satellite. An inverted repeat of 23-kb, comprising satellites, LINE elements, *Alu* sequences and other repetitive elements, was identified in the breakpoint regions in 16p11-12 and 16q11-12 (Figures 3.11 and 3.14). It is likely that the high degree of sequence similarity manifested by these inverted repeats served to mediate the formation of the intra-chromosomal inversion loop, thereby facilitating the respective regions coming into close proximity with each other. The various local sequence homologies apparent within and between the inversion breakpoints [viz. the octanucleotide repeat (TGTGAAAG) at the 16p breakpoint, the pentanucleotide repeat (AATGA) at the 16q breakpoint, the TGAAA motif common to the 16p and 16q breakpoints, and the 22-bp direct repeat that flanks the 16p and 16q breakpoints (Figure 3.10)] could then account for the inversion occurring at the sites that it did. Direct repeats are known to form ‘slipped structures’ and these non-B DNA conformations could have contributed to the formation of local breaks that then led to the inversion in *PTR*. Non-B DNA structures have been shown to play a role in numerous genomic rearrangements associated with human disease (reviewed by Bacolla and Wells, 2004).

In my study, I have shown that the human-specific pericentric inversion of chromosome 18 was also mediated by LCRs. The convincing hypothesis to explain

the mechanism leading to this inversion is that the duplication of the 19 kb segment was followed by an intrachromosomal transposition of the duplicon to the p-arm of an ancestral chromosome 18 (Figure 3.18). The inverted orientation and the high homology of the duplicates enabled intrachromosomal non-allelic homologous recombination and subsequent inversion of the segment between the duplicons. The transposition of the 19 kb duplicate to distal 18p in an ancestral hominoid is quite plausible. The distal part of BAC RP11-683L23 shows homology to many other chromosomes especially in subtelomeric and pericentromeric regions, as determined by FISH and sequence comparisons (Figure 3.16A). This indicates that the region covered by RP11-683L23 showed considerable plasticity, as it has accepted or produced additional duplications during primate genome evolution. However, the evolutionary rearrangements that gave rise to distal human 18p must be even more complicated as suggested by the FISH pattern of BAC RP11-683L23 on chimpanzee chromosomes. RP11-683L23 hybridises to regions homologous to 18q11 and 18qter in the chimpanzee, but not to 18pter (Figure 3.16B). This suggests that during human evolution, in addition to the 19 kb duplicon, a segment from the distal long arm of the ancestral chromosome 18 has been transposed to the 18p11.3 region. Therefore, as for chromosome 16, remarkable plasticity has to be ascribed to this region during genome evolution of higher primates.
It should be noted that none of the BACs containing the duplications above-mentioned where represented on the 6-K BAC array and therefore these LCRs could not have been detected using that method.

4.6.3.4. Evolutionary history of the pericentric inversions of gorilla and chimpanzee chromosomes XVI

The chromosome 16 is known for containing a high amount of LCRs. In order to verify the impact of these LCRs on the evolution of *HSA* 16, I performed comparative FISH analysis in other primates. The q-arm breakpoint region (Figure 3.12), represented by *HSA* BAC RP11-696P19, is derived from an ancestral primate chromosome homologous to *HSA* 1. After the separation of the orangutan from the common ancestral lineage shared with the African hominoids, a duplication of the region in question occurred followed by the inter-chromosomal transposition of the duplicated copy to the ancestral chromosome 16q. The most parsimonious

explanation for my findings is that among the hominoids, humans retained the ancestral form of chromosome 16, whereas *GGO* and *PTR* acquired pericentric inversions of their respective homologues. The inversions of *GGO* XVI and *PTR* XVI were not detected during the molecular cytogenetic studies previously performed by Misceo et al. (2003) because the probes that they used are located distally to the respective breakpoints. Furthermore, it has neither been observed by Martin et al. (2004) in their comparison of the *HSA* 16 sequence with homologous sequences available in other vertebrates, including the chimpanzee, nor by the Chimpanzee Sequencing and Analysis Consortium (2005), which made a genome-wide comparison of the chimpanzee genome with the one of the human. One possible reason for this might be that the scaffolds of the draft chimpanzee genome sequence from the breakpoint regions had not been assembled into larger contigs and, in the absence of FISH data, this could have hampered the assignment of the orientation of these scaffolds.

4.6.3.5. Pericentric inversion of chromosome XVI occurred twice independently during evolution

Comparative analysis of the inversion breakpoint regions in *GGO* XVI and *PTR* XVI indicated that whilst the p-arm breakpoints were located within or very close to HSATII satellite DNA, the q-arm breakpoints occurred at non-identical locations (Figure 3.13). Owing to the repetitive nature of the HSATII satellites, it was not possible to compare the p-arm breakpoints in *GGO* XVI and *PTR* XVI. This notwithstanding, q-arm breakpoint heterogeneity is supportive of the view that the pericentric inversions of *GGO* XVI and *PTR* XVI occurred independently in the two lineages. PCR with primer pair P2, designed according to the sequence of *HSA* BAC RP11-696P19 and flanking the inversion breakpoint on *HSA* 16q, was negative in *PTR*, but positive in *HSA* and *GGO* (data not shown). Sequence analysis of the PCR products also indicated an identical sequence in *HSA* and *GGO* at the site where the inversion/fusion took place in *PTR*. The sequence amplified by primer pair P2 has not been deleted in *PTR*, but rather was split by the inversion (Figures 3.10 and 3.13). This provides unequivocal confirmation that the inversions of *GGO* XVI and *PTR* XVI occurred at different sites. It should be noted that the inversion breakpoints of *PTR* XVI are the same as of *PPA* XVI, the other chimpanzee species, which separated

from the common chimpanzee, *Pan troglodytes*, 0.86-2 Mya (Yoder and Yang, 2000; Won and Hey, 2005).

The fact that the inversions of *GGO* XVI and *PTR* XVI are not identical is quite consistent with what is known of primate phylogeny. While the gorilla branched off 6.2 to 8 Mya from the main hominoid lineage, the chimpanzee separated later, some 4.6 to 6.2 Mya (Chen and Li, 2001). Had the *GGO* and *PTR* inversions been identical by descent, then either humans would have possessed this inversion as well, or the human chromosome would have been the derived one.

Whilst the *GGO* XVI and *PTR* XVI inversions appear to have been independent events, the breakpoints nevertheless occurred in the same chromosomal regions. Although this finding is, as far as I am aware, unprecedented during vertebrate evolution, multiple independent yet cytologically identical inversions have been demonstrated by DNA sequence analysis in the *Anopheles gambiae* complex (Caccone et al., 1998). It might be argued that the parallel evolution of the cytologically identical inversion could have been selectively favoured. Indeed, since chromosome 16 is very rich in duplicated sequences, the probability that inversions with breakpoints in the same regions occur twice quite independently is likely to be rather low. However, despite the fact that *HSA* 16 contains the highest level of duplications among the autosomes, these duplicated regions tend to be clustered, particularly along the p-arm. In contrast to 16p11-12, the 16q11-12 region is not exceptionally rich in duplications. Since the pericentric inversions of both *PTR* XVI and *GGO* XVI appear to have been mediated by recombination between paralogous sequence blocks and stretches of HSATII satellite found in 16p11 and 16q11, these regions may have specifically facilitated inversions.

Bailey et al. (2004) have put forward a non-random model of chromosomal evolution, which proposes that specific regions in mammalian genomes are predisposed both to recurrent small-scale duplications and to large-scale evolutionary rearrangements. The occurrence of two independent inversions in *GGO* and *PTR* with breakpoints in the same regions provides clear support for this non-random model.

4.7. Final conclusion

My data highlight the importance of LCRs for genome evolution and particularly for the one of the human. By interspecies aCGH, I was able to show that DNA copy number losses and gains are not only important for gene dosage but that they were also responsible for numerous large-scale rearrangements that occurred during human genome evolution. Indeed, several BACs showing human-specific DNA copy number gains in my analysis were located in the breakpoint regions of the human-specific pericentric inversion of chromosome 1 as well as of the human polymorphic inversion of chromosome 2. The presence of LCRs at the borders of such inversions suggests their implication in mediating the respective rearrangements by allowing intrachromosomal homologous recombination between these repeats. However, it should be noted that complete genome sequences of new primate species would be necessary to determine sequences that have been completely deleted in the human genome and to increase the number of CNDs that are specific to each primates. Intra-species genome comparisons should also be considered in order to find out the time of fixation of the CNDs identified.

The precise characterization of the breakpoints of the pericentric inversions of HSA 18 and of *PTR* XVI and *GGO* XVI allowed me to study the duplication content of the breakpoints regions of these rearrangements. I found that LCRs bordered each breakpoint, confirming the importance of segmental duplications in mediating evolutionary rearrangements. Moreover, I have determined that the pericentric inversions of *PTR* XVI and *GGO* XVI occurred twice independently during evolution verifying the instability of that region suggested previously by Martin et al. (2004).

In conclusion, my study showed that segmental duplications are frequently assigned to unstable genomic regions and to chromosomal breakpoints, emphasizing their role in determining both karyotypic evolution and genomic diversity in humans.

5. REFERENCES

Adorjan, P., Distler, J., Lipscher, E., Model, F., Muller, J., Pelet, C., Braun, A., Florl, A.R., Gutig, D., Grabs, G., Howe, A., Kursar, M., Lesche, R., Leu, E., Lewin, A., Maier, S., Muller, V., Otto, T., Scholz, C., Schulz, W.A., Seifert, H.H., Schwope, I., Ziebarth, H., Berlin, K., Piepenbrock, C., Olek, A.: Tumour class prediction and discovery by microarray-based DNA methylation analysis. *Nucleic Acid Res* **30**: e21 (2002)

Albertson, D.G. and Pinkel, D.: Genomic microarrays in human genetic disease and cancer. *Hum Mol Genet* **12**: 145-152 (2003)

Amos-Landgraf, J.M., Ji, Y., Gottlieb, W., Depinet, T., Wandstrat, A.E., Cassidy, S.B., Driscoll, D.J., Rogan, P.K., Schwartz, S., Nicholls, R.D.: Chromosome breakage in the Prader-Willi and Angelman syndromes involves recombination between large, transcribed repeats at proximal and distal breakpoints. *Am J Hum Genet* **65**: 370-386 (1999)

Archidiacono, N., Antonacci, R., Marzella, R., Finelli, P., Lonoce, A., Rocchi, M.: Comparative mapping of human alphoid sequences in great apes using fluorescence in situ hybridization. *Genomics* **25**: 477-484 (1995)

Armengol, L., Pujana, M.A., Cheung, J., Scherer, S.W., Estivill, X.: Enrichment of segmental duplications in regions of breaks of synteny between the human and mouse genomes suggest their involvement in evolutionary rearrangements. *Hum Mol Genet* **12**:2201-2208 (2003)

Bacolla, A. and Wells, R.D.: Non-B DNA conformations, genomic rearrangements, and human disease. *J Biol Chem* **279**: 47411-47414 (2004)

Bailey, J.A., Yavor, A.M., Massa, H.F., Trask, B.J., Eichler, E.E.: Segmental duplications: Organization and impact within the current human genome project assembly. *Genome Res* **11**:1005–1017 (2001)

Bailey, J.A., Gu, Z., Clark, R.A., Reinert, K., Samonte, R.V., Schwartz, S., Adams, M.D., Myers, E.W., Li, P.W., Eichler, E.E.: Recent segmental duplications in the human genome. *Science* **297**:1003–1007 (2002a)

Bailey, J.A., Yavor, A.M., Viggiano, L., Misceo, D., Horvath, J.E., Archidiacono, N., Schwartz, S., Rocchi, M., Eichler, E.E.: Human-specific duplication and mosaic transcripts: the recent paralogous structure of chromosome 22. *Am J Hum Genet* **70**:83-100 (2002b)

Bailey, J.A., Liu, G., Eichler, E.E.: An *Alu* transposition model for the origin and expansion of human segmental duplications. *Am J Hum Genet* **73**:823-834 (2003)

Bailey, J.A., Baertsch, R., Kent, W.J., Haussler, D., Eichler, E.E.: Hotspots of mammalian chromosomal evolution. *Genome Biol* **5**: R23 (2004)

Batzer, M.A., Deininger, P.L.: Alu repeats and human genomic diversity. *Nat Rev Genet* **3**: 370-379 (2002)

Brewer, C., Holloway, S., Zawalnyski, P., Schinzel, A., FitzPatrick, D.: A chromosomal duplication map of malformations: regions of suspected haplo- and triplolethality--and tolerance of segmental aneuploidy--in humans. *Am J Hum Genet* **64**:1702-1708 (1999)

Britten, R.J.: Divergence between samples of chimpanzee and human DNA sequences is 5%, counting indels. *Proc Natl Acad Sci USA* **99**:13633-13635 (2002)

Britten, R.J., Rowen, L., Williams, J., Cameron, R.A.: Majority of divergence between closely related DNA samples is due to indels. *Proc Natl Acad Sci USA* **100**:14661-14665 (2003)

Cáceres, M., Lachuer, J., Zapala, M.A., Redmond, J.C., Kudo, L., Geschwind, D.H., Lockhart, D.J., Preuss, T.M., Barlow, C.: Elevated gene expression levels distinguish human from non-human primate brains. *Proc. Natl. Acad. Sci. USA* **100**: 13030-13035 (2003)

Caccone, A., Min, G.S., Powell, J.R.: Multiple origins of cytologically identical chromosome inversions in the *Anopheles gambiae* complex. *Genetics* **150**: 807-814 (1998)

Chance, P.F., Abbas, N., Lensch, M.W., Pentao, L., Roa, B.B., Patel, P.I., Lupski, J.R.: Two autosomal neuropathies result from reciprocal DNA duplication/deletion of a region on chromosome 17. *Hum Mol Genet* **3**: 223-228 (1994)

Chen, F.C. and Li, W.H.: Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. *Am J Hum Genet* **68**:444-456 (2001)

Cheng, Z., Ventura, M., She, X., Khaitovich, P., Graves, T., Osoegawa, K., Church, D., DeJong, P., Wilson, R.K., Pääbo, S., Rocchi, M., Eichler, E.E.: A genome-wide comparison of recent chimpanzee and human segmental duplications. *Nature* **437**:88-93 (2005)

Cheung, J., Estivill, X., Khaja, R., MacDonald, J.R., Lau, K., Tsui, L.-C., Scherer, W.: Genome-wide detection of segmental duplications and potential assembly errors in the human genome sequence. *Genome Biol* **4**: R25 (2003a)

Cheung, J., Wilson, M.D., Zhang, J., Khaja, R., MacDonald, J.R., Heng, H.H., Koop, B.F., Scherer, S.W.: Recent segmental and gene duplications in the mouse genome. *Genome Biol* **4**:R47 (2003b)

Chimpanzee Sequencing and Analysis Consortium: Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**: 69-87 (2005)

Cicarelli, F.D., von Mering, C., Suyama, M., Harrington, E.D., Izaurralde, E., Bork, E.: Complex genomic rearrangements lead to novel primate gene function. *Genome Res* **15**: 343-351 (2005)

Conrad, D.F., Andrews, T.D., Carter, N.P., Hurles, M.E., Pritchard, J.K.: A high-resolution survey of deletion polymorphism in the human genome. *Nat Genet* **38**: 75-81 (2006)

Cowell, J.K. and Nowak, N.J.: High resolution analysis of genetic events in cancer cells using BAC arrays and CGHa. *Adv Cancer Res* **90**:91-125 **(**2003)

Dennehey, B.K., Gutches, D.G., McConkey, E.H., Krauter, K.S.: Inversion, duplication, and changes in gene context are associated with human chromosome 18 evolution. *Genomics* **83**:493-501 (2004)

de Pontbriand, A., Wang, X.P., Cavaloc, Y., Mattei, M.G., Galibert, F.: Synteny comparison between apes and human using fine-mapping of the genome. *Genomics* **80**: 395-401 (2002)

Dorschner, M.O., Sybert, V.P., Weaver, M., Pletcher, B.A., Stephens, K.: *NF1* microdeletion breakpoints are clustered at flanking repetitive sequences. *Hum Mol Genet* **9**: 35-46 (2000)

Dutrillaux, B., Biemont, M.C., Viegas-Pequinot, E., Laurent, C.: Comparison of the karyotypes of four Cercopithecoidae: *Papio papio, P. anubis, Macaca mulatta*, and *M. fascicularis*. *Cytogenet Cell Genet* **23**: 77-83 (1979)

Edelmann, L., Pandita, R.K., Morrow, B.E.: Low-copy repeats mediate the common 3-Mb deletion in patients with velo-cardio-facial syndrome. *Am J Hum Genet* **64**: 1076-1086 (1999)

Eichler, E.E., Lu, F., Shen, Y., Antonacci, R., Jurecic, V., Doggett, N.A., Moyzis, R.K., Baldini, A., Gibbs, R.A., Nelson, D.L.: Duplication of a gene-rich cluster between 16p11.1 and Xq28: a novel pericentromeric-directed mechanism for paralogous genome evolution. *Hum Mol Genet* **5**: 899-912 (1996)

Eichler, E.E.: Masquerading repeats: paralogous pitfalls of the human genome. *Genome Res* **8**: 758-762 (1998)

Eichler, E.E., Archidiacono, N., Rocchi, M.: CAGGG repeats and the pericentromeric duplication of the hominoid genome. *Genome Res* **9**: 1048-1058 (1999)

Eichler, E.E.: Segmental duplications: what's missing, misassigned, and misassembled-and should we care? *Genome Res* **11**: 653-656 (2001)

Eichler, E.E.: Widening the spectrum of human genetic variation. *Nat Genet* **38**: 9-11 (2006)

Ellegren, H.: Evolution: Natural selection in the evolution of humans and chimps. *Curr Biol* R919-R922 (2005)

Enard, W., Przeworski, M., Fisher, S.E., Lai, C.S., Wiebe, V., Kitano, T., Monaco, A.P., Pääbo, S.: Molecular evolution of FOXP2, a gene involved in speech and language. *Nature* **418**: 869-872 (2002a)

Enard, W., Khaitovich, P., Klose, J., Zöllner, S., Heissig, F., Giavalisco, P., Nieselt-Struwe, K., Muchmore, E., Varki, A., Ravid, R., Doxiadis, G.M., Bontrop, R.E., Pääbo, S.: Intra- and interspecific variation in primate gene expression patterns. *Science* **296**: 340-343 (2002b)

Enard, W. and Pääbo, S.: Comparative primate genomics. *Annu Rev Genomics Hum Genet* **5**: 351-378 (2004)

Fan, Y., Linardopoulou, E., Friedman, C., Williams, E., Trask, B.J.: Genomic structure and evolution of the ancestral chromosome fusion site in 2q13-2q14.1 and paralogous regions on other human chromosomes. *Genome Res* **12**:1651-1662 (2002)

Feuk, L., MacDonald, J.R., Tang, T., Carson, A.R., Li, M., Rao, G., Khaja, R., Scherer, S.W.: Discovery of human inversion polymorphisms by comparative analysis of human and chimpanzee sequence assemblies. *PLOS Genetics* 1: 489-498 (2005)

Fisher, S.E. and Marcus, G.F.: The eloquent ape: genes, brains and the evolution of language. *Nat Rev Genet* **7**:9-20 (2006)

Fortna, A., Kim, Y., MacLaren, E., Marshall, K., Hahn, G., Meltesen, L., Brenton, M., Hink, R., Burgers, S., Hernandez-Boussard, T., Karimpour-Fard, A., Glueck, D., McGavran, L., Berry, R., Pollack, J., Sikela, J.M.: Lineage-specific gene duplication and loss in human and great ape evolution. *PLoS Biol* **2**:E207 (2004)

Frazer, K.A., Chen, X., Hinds, D.A., Pant, P.V., Patil, N., Cox, D.R.: Genomic DNA insertions and deletions occur frequently between humans and nonhuman primates. *Genome Res* **13**:341–346 (2003)

Gilad, Y., Bustamante, C.D., Lancet, D., Pääbo, S.: Natural selection on the olfactory receptor gene family in humans and chimpanzees. *Am J Hum Genet* **73**: 489–501 (2003)

Gilad, Y., Rifkin, S.A., Bertone, P., Gerstein, M., White, K.P.: Multi-species microarrays reveal the effect of sequence divergence on gene expression profiles. *Genome Res* **15**: 674-680 (2005a)

Gilad, Y., Man, O., Glusman, G.: A comparison of the human and chimpanzee olfactory receptor gene repertoires. *Genome Res* **15**:224-230 (2005b)

Glazko, G.V. and Nei, M.: Estimation of divergence times for major lineages of primate species. *Mol Biol Evol* **20**:424-434 (2003)

Glusman, G., Yanai, I., Rubin, I., Lancet, D.: The complete human olfactory subgenome. *Genome Res* **11**: 685-702 (2001)

Goidts, V., Szamalek, J.M., Hameister, H., Kehrer-Sawatzki, H.: Segmental duplication associated with the human-specific inversion of chromosome 18: further example of the impact of segmental duplications on karyotype and genome evolution in primates. *Hum Genet* **115**: 116-122 (2004)

Goidts, V., Szamalek, J.M., de Jong, P., Cooper, D.N., Chuzhanova, N., Hameister, H., Kehrer-Sawatzki, H.: Independent intrachromosomal recombination events underlie the pericentric inversion of chimpanzee and gorilla chromosomes homologous to human chromosome 16. *Genome Res* **15**: 1232-1242 (2005)

Goodman, M.: The genomic record of humankind's evolutionary roots. *Am J Hum Genet* **64**: 31-39 (1999)

Hey, J.: Speciation and inversions: chimps and humans. *Bioessays* **25**: 825-828 (2003)

Hinds, D.A., Kloek, A.P., Jen, M., Chen, X., Frazer, K.A.: Common deletions and SNPs are in linkage desiquilibrium in the human genome. *Nat Genet* **38**: 82-85 (2006)

Horvath, J.E., Viggiano, L., Loftus, B.J., Adams, M.D., Archidiacono, N., Rocchi, M., Eichler, E.E.: Molecular structure and evolution of an alpha satellite/non-alpha satellite junction at 16p11. *Hum Mol Genet* **9**: 113-123 (2000)

Horvath, J.E., Bailey, J.A., Locke, D.P., Eichler, E.E.: Lessons from the human genome: transitions between euchromatin and heterochromatin. *Hum Mol Genet* **10**: 2215-2223 (2001)

Horvath, J.E., Gulden, C.L., Bailey, J.A., Yohn, C., McPherson, J.D., Prescott, A., Roe, B.A., de Jong, P.J., Ventura, M., Misceo, D., Archidiacono, N., Zhao, S., Schwartz, S., Rocchi, M., Eichler, E.E.: Using a pericentromeric interspersed repeat to recapitulate the phylogeny and expansion of human centromeric segmental duplications. *Mol Biol Evol* **20**:1463-1479 (2003)

Horvath, J.E., Gulden, C.L., Vallente, R.U., Eichler, M.Y., Ventura, M., McPherson, J.D., Graves, T.A., Wilson, R.K., Schwartz,S., Rocchi, M., Eichler, E.E.: Punctuated duplication seeding events during the evolution of human chromosome 2p11. *Genome Res* **15**: 914-927 (2005)

Huang, B., Crolla, J.A., Wolf-Ledbetter, M.E., Macha, M.E., Papenhausen, P.N., Ledbetter, D.H.: Refined molecular characterization of the breakpoints in small inv dup(15) chromosomes. *Hum Genet* **99**: 11-17 (1997)

Hurles, M.: Gene duplication: the genomic trade in spare parts. *PLOS Biology* **2**: 900-904 (2004)

Hurst, L.D.:The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends in Genet* **18**: 486-487 (2002)

Iafrate, A.J., Feuk, L., Rivera, M.N., Listewnik, M.L., Donahoe, P.K., Qi, Y., Scherer, S.W., Lee, C.: Detection of large-scale variation in the human genome. *Nat Genet* **36**:949-951 (2004)

International Human Genome Sequencing Consortium: Initial sequencing and analysis of the human genome. *Nature* **409**: 860-921 (2001)

Jackson, M.: Duplicate, decouple, disperse: the evolutionary transience of human centromeric regions. *Curr Opin Genet Devel* **13**: 629-635 (2003)

Jenne, D.E., Tinschert, S., Dorschner, M.O., Hameister, H., Stephens, K., Kehrer-Sawatzki, H.: Complete physical map and gene content of the human NF1 tumor suppressor region in human and mouse. *Genes Chromosomes Cancer* **37**: 111-120 (2003)

Ji, Y., Eichler, E.E., Schwartz, S., Nicholls, R.D.: Structure of chromosomal duplicons and their role in mediating human genomic disorders. *Genome Res* **10**: 597-610 (2000a)

Ji, Y., Rebert, N.A., Joslin, J.M., Higgins, M.J., Schultz, R.A., Nicholls, R.D.: Structure of the highly conserved *HERC2* gene and of multiple partially duplicated paralogs in human. *Genome Res* **10**: 319-329 (2000b)

Jorgensen, P.M., Graslund, S., Betz, R., Stahl, S., Larsson, C., Hoog, C.: Characterisation of the human APC1, the largest subunit of the anaphase-promoting complex. *Gene* **262**:51-59 (2001)

Johnson, M.E., Viggiano, L., Bailey, J.A., Abdul-Rauf, M., Goodwin, G., Rocchi, M., Eichler, E.E.: Positive selection of a gene family during the emergence of humans and African apes. *Nature* **413**: 514-519 (2001)

Jurka, J.: Evolutionary impact of human Alu repetitive elements. *Curr Opin Genet Dev* **14**: 603-608 (2004)

Kapitonov, V. and Jurka, J.: The age of *Alu* subfamilies. *J Mol Evol* **42**:59-65 (1996)

Karaman, M.W., Houck, M.L., Chemnick, L.G., Nagpal, S., Chawannakul, D., Sudano, D., Pike, B.L., Ho, V.V., Ryder, O.A., Hacia, J.G.: Comparative analysis of gene-expression patterns in human and African great ape cultured fibroblasts. *Genome Res* **13**: 1619-1630 (2003)

Kehrer-Sawatzki, H., Schreiner, B., Taenzer, S., Platzer, M., Muller, S., Hameister, H.: Molecular characterization of the pericentric inversion that causes differences between chimpanzee chromosome 19 and human chromosome 17. *Am J Hum Genet* **71**: 375-388 (2002)

Kehrer-Sawatzki, H., Sandig, C.A., Goidts, V., Hameister, H.: Breakpoint analysis of the pericentric inversion between chimpanzee chromosome 10 and the homologous chromosome 12 in humans. *Cytogenet Genome Res* **108**: 91-97 (2005a)

Kehrer-Sawatzki, H., Sandig, C., Chuzhanova, N., Goidts, V., Szamalek, J.M., Tanzer, S., Muller, S., Platzer, M., Cooper, D.N., Hameister, H.: Breakpoint analysis of the pericentric inversion distinguishing human chromosome 4 from the homologous chromosome in the chimpanzee (*Pan troglodytes*). *Hum Mutation* **25**: 45-55 (2005b)

Kehrer-Sawatzki, H., Szamalek, J.M., Tanzer, S., Platzer, M., Hameister, H.: Molecular characterization of the pericentric inversion of chimpanzee chromosome 11 homologous to human chromosome 9. *Genomics* **85**: 542-550 (2005c)

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., Haussler, D.: The Human Genome Browser at UCSC. *Genome Res* **12**:996-1006 (2002)

Kuroda-Kawaguchi, T., Skaletsky, H., Brown, L.G., Minx, P.J., Cordum, H.S., Waterston, R.H., Wilson, R.K., Silber, S., Oates, R., Rozen, S., Page, DC.: The AZFc region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. *Nat Genet* **29**: 379-386 (2001)

Lai, C.S.L., Fisher, S.E., Hurst, J.A., Vargha-Khadem, F., Monaco, A.P.: A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* **413**: 519-523 (2001)

Linardopoulou, E.V., Williams, E.M., Fan, Y., Friedman, C., Young, J.M., Trask, B.J.: Human subtelomeres are hot spots of interchromosomal recombination and segmental duplication. *Nature* **437**: 94-100 (2005)

Liu, G., Zhao, S., Bailey, J.A., Sahinalp, S.C., Alkan, C., Tuzun, E., Green, E.D., Eichler, E.E.: Analysis of primate genomic variation reveals a repeat-driven expansion of the human genome. *Genome Res* **13**:358-368 (2003)

Locke, D.P., Segraves, R., Carbone, L., Archidiacono, N., Albertson, D.G., Pinkel, D., Eichler, E.E.: Large-scale variation among human and great ape genomes determined by array comparative genomic hybridization. *Genome Res* **13**:347-357 (2003a)

Locke, D.P., Archidiacono, N., Misceo, D., Cardone, M.F., Deschamps, S., Roe, B., Rocchi, M., Eichler, E.E.: Refinement of a chimpanzee pericentric inversion breakpoint to a segmental duplication cluster. *Genome Biol* **4**: R50 (2003b)

Lupski, J.R.: Charcot-Marie-Tooth disease: lessons in genetic mechanisms. *Mol Med* **4**: 3-11 (1998)

Lu, J., Li, W.-H., Wu, C.-I.: Comment on „chromosomal speciation and molecular divergence-accelerated evolution in rearranged chromosomes. *Science* **302**: 988b (2003)

Magness, C.L., Campion Fellin, P., Thomas, M.J., Korth, M.J., Agy, M.B., Proll, S.C., Fitzgibbon, M., Scherer, C.A., Miner, D.G., Katze, M.G., Iadonato, S.P.: Analysis of the *Macaca Mulatta* transcriptome and the sequence divergence between *Macaca* and human. *Genome Biol* **6**:R60 (2005)

Mantripragada, K.K., Buckley, P.G., de Stahl, T.D., Dumanski, J.P.: Genomic microarrays in the spotlight. *Trends in Genet* **2**: 87-94 (2004)

Maresco, D.L., Blue, L.E., Culley, L.L., Kimberly, R.P., Anderson, C.L., Theil, K.S.: Localization of *FCGR1* encoding Fcgamma receptor class I in primates: molecular evidence for two pericentric inversions during the evolution of human chromosome 1. *Cytogenet Cell Genet* **82**: 71-74 (1998)

Marques-Bonet, T., Caceres, M., Bertranpetit, J., Preuss, T.M., Thomas, J.Y., Navarro, A.: Chromosomal rearrangements and the genomic distribution of gene-expression divergence in humans and chimpanzees. *Trends Genet* **20**: 524-529 (2004)

Martin, J., Han, C., Gordon, L.A., Terry, A., Prabhakar, S., She, X., Xie, G., Hellsten, U., Chan, Y.M., Altherr, M., Couronne, O., Aerts, A., Bajorek, E., Black, S., Blumer, H., Branscomb, E., Brown, N.C., Bruno, W.J., Buckingham, J.M., Callen, D.F., Campbell, C.S., Campbell, M.L., Campbell, E.W., Caoile, C., Challacombe, J.F., Chasteen, L.A., Chertkov, O., Chi, H.C., Christensen, M., Clark, L.M., Cohn, J.D., Denys, M., Detter, J.C., Dickson, M., Dimitrijevic-Bussod, M., Escobar, J., Fawcett, J.J., Flowers, D., Fotopulos, D., Glavina, T., Gomez, M., Gonzales, E., Goodstein, D., Goodwin, L.A., Grady, D.L., Grigoriev, I., Groza, M., Hammon, N., Hawkins, T., Haydu, L., Hildebrand, C.E., Huang, W., Israni, S., Jett, J., Jewett, P.B., Kadner, K., Kimball, H., Kobayashi, A., Krawczyk, M.C., Leyba, T., Longmire, J.L., Lopez, F., Lou, Y., Lowry, S., Ludeman, T., Manohar, C.F., Mark, G.A., McMurray, K.L., Meincke, L.J., Morgan, J., Moyzis, R.K., Mundt, M.O., Munk, A.C., Nandkeshwar, R.D., Pitluck, S., Pollard, M., Predki, P., Parson-Quintana, B., Ramirez, L., Rash, S., Retterer, J., Ricke, D.O., Robinson, D.L., Rodriguez, A., Salamov, A., Saunders, E.H., Scott, D., Shough, T., Stallings, R.L., Stalvey, M., Sutherland, R.D., Tapia, R., Tesmer, J.G., Thayer, N., Thompson, L.S., Tice, H., Torney, D.C., Tran-Gyamfi, M., Tsai, M., Ulanovsky, L.E., Ustaszewska, A., Vo, N., White, P.S., Williams, A.L., Wills, P.L., Wu, J.R., Wu, K., Yang, J., Dejong, P., Bruce, D., Doggett, N.A., Deaven, L., Schmutz, J., Grimwood, J., Richardson, P., Rokhsar, D.S., Eichler, E.E., Gilna, P., Lucas, S.M., Myers, R.M., Rubin, E.M., Pennacchio, L.A.: The sequence and analysis of duplication-rich human chromosome 16. *Nature* **432**: 988-994 (2004)

Martin, M.J., Rayner, J.C., Gagneux, P., Barnwell, J.W., Varki, A.: Evolution of human-chimpanzee differences in malaria susceptibility: relationship to human genetic loss of N-glycolyneuraminic acid. *Proc Natl Acad Sci USA* **102**: 12819-12824 (2005)

McCarroll, S.A., Hadnott, T.N., Perry, G.H., Sabeti, P.C., Zody, M.C., Barrett, J.C., Dallaire, S., Gabriel, S.B., Lee, C., Daly, M.J., Altshuler, D.M.: Common deletion polymorphisms in the human genome. *Nat Genet* **38**: 86-92 (2006)

McClure, H.M.: Tumors in nonhuman primates : observations during a six-yaer period in the Yerkes primate center colony. *Am J Phys Anthropol* **38**: 425-429 (1973)

McConkey, E.H.: The origin of human chromosome 18 from a human/ape ancestor. *Cytogenet Cell Genet* **76**: 189-191 (1997)

McConkey, E.H.: Orthologous numbering of great ape and human chromosomes is essential for comparative genomics. *Cytogenet Genome Res* **105**: 157-158 (2004)

McTaggart, K.E., Budarf, M.L., Drscoll, D.A., Emanuel, B.S., Ferreira, P., McDermid H.E.: Cat eye syndrome chromosome breakpoint clustering: identification of two intervals also associated with 22q11 deletion syndrome breakpoints. *Cytogenet Cell Genet* **81**: 222-228 (1998)

Miyamoto, M.M., Koop, B.F., Slightom, J.L., Goodman, M., Tennant, M.R.: Molecular systematics of higher primates: genealogical relations and classification. *Proc Natl Acad Sci USA* **85**: 7627-7631 (1988)

Misceo, D., Ventura, M., Eder, V., Rocchi, M., Archidiacono, N.: Human chromosome 16 conservation in primates. *Chromosome Res* **11**: 323-326 (2003)

Montefalcone, G., Tempesta, S., Rocchi, M., Archidiacono, N.: Centromere repositioning. *Genome Res* **9**: 1184-1188 (1999)

Müller, S. and Wienberg, J.: "Bar-coding" primate chromosomes: molecular cytogenetic screening for the ancestral hominoid karyotype. *Hum Genet* **109**:85-94 (2001)

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., Erlich, H.: Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol* **51**: 263-273 (1986)

Nahon, J.L.: Birth of 'human-specific' genes during primate evolution. *Genetica* **118**: 193-208 (2003)

Navarro, A. and Barton, N.H.: Chromosomal speciation and molecular divergence-accelerated evolution in rearranged chromosomes. *Science* **300**: 321-324 (2003)

Naylor, J.A., Nicholson, P., Goodeve, A., Hassock, S., Peake, I., Giannelli, F.: A novel DNA inversion causing severe hemophilia A. *Blood* **87**: 3255-3261 (1996)

Newman, T.L., Tuzun, E., Morrison, V.A., Hayden, K.E., Ventura, M., McGrath, S.D., Rocchi, M., Eichler, E.E.: A genome-wide survey of structural variation between human and chimpanzee. *Genome Res* **15**: 1344-1356 (2005)

Nickerson, E. and Nelson, D.L.: Molecular definition of pericentric inversion breakpoints occurring during the evolution of humans and chimpanzees. *Genomics* **50**: 368-372 (1998)

Nielsen, R., Bustamante, C., Clark, A.G., Glanowski, S., Sackton, T.B., Hubisz, M.J., Fledel-Alon, A., Tanenbaum, D.M., Civello, D., White, T.J., Sninsky, J.J., Adams, M.D., Cargill, M.: A scan for positively selected genes in the genome of humans and chimpanzees. *PLOS Biology* **3**: 976-985 (2005)

Noor, M.A., Grams, K.L., Bertucci, L.A., Reiland, J.: Chromosomal inversions and the reproductive isolation of species. *Proc Natl Acad Sci USA* **98**: 12084-12088 (2001)

Novembre, F.J., Saucier, M., Anderson, D.C., Klumpp, S.A., O'Neil, S.P., Brown, N.R. 2nd, Hart, C.E., Guenthner, P.C., Swenson, R.B., McClure, H.M. : Development of AIDS in chimpanzee infected with human immunodefficiency virus type 1. *J Virol* **244**: 261-272 (1997)

Ochman, H., Lawrence, J.G., Groisman, E.A.: Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**: 299-304 (2000)

Ohno, S.: Evolution by gene duplication. *Springer, Berlin Heidelberg New York* (1970)

Olson, M.V.: When less is more: gene loss as an engine of evolutionary change. *Am J Hum Genet* **64**: 18-23 (1999)

Olson, M.V. and Varki, A.: Sequencing the chimpanzee genome: insight into human evolution and disease. *Nat Rev Genet* **4**:20-28 (2003)

Panagopoulos, I., Aman, P., Fioretos, T., Hoglund, M., Johansson, B., Mandahl, N., Heim, S., Behrendtz, M., Mitelman, F.: Fusion of the *FUS* gene with *ERG* in acute myeloid leukemia with t(16;21)(p11;q22). *Genes Chrom. Cancer* **11**: 256-262 (1994)

Pérez Jurardo, L.A., Wang, Y.-K., Peoples, R., Coloma, A., Cruces, J., Francke, U.: A duplicated gene in the breakpoint regions of the 7q11.23 Williams-Beuren syndrome deletion encodes the initiator binding protein TFII-I and BAP-135, a phosphorylation target of BTK. *Hum Mol Genet* **7**: 325-334 (1998)

Pollack, J.R., Perou, C.M., Alizadeh, A.A., Eisen, M.BM, Pergamenschikov, A., Williams, C.F., Jeffrey, S.S., Botstein, D., Brown, P.O.: Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat Genet* **23**: 41-46 (1999)

Potocki, L., Chen, K.S., Park, S.S., Osterholm, D.E., Withers, M.A., Kimonis, V., Summers, A.M., Meschino, W.S., Anyane-Yeboa, K., Kashork, C.D., Shaffer, L.G., Lupski, J.R.: Molecular mechanism for duplication 17p11.2-the homologous recombination reciprocal of the Smith-Magenis microdeletion. *Nat Genet* **24**: 84-87 (2000)

Pujana, M.A., Nadal, M., Guitart, M., Armengol, L., Gratacos, M., Estivill, X.: Human chromosome 15q11-q14 regions of rearrangements contain clusters of LCR15 duplicons. *Eur J Hum Genet* **10**: 26-35 (2002)

Repping, S., Skaletsky, H., Lange, J., Silber, S., Van Der Veen, F., Oates, R.D., Page, D.C., Rozen, S.: Recombination between palindromes P5 and P1 on the human Y chromosome causes massive deletions and spermatogenic failure. *Am J Hum Genet* **71**: 906-922 (2002)

Rieseberg, L.H. and Livingstone, K.: Evolution. Chromosomal speciation in primates. *Science* **300**: 267-268 (2003)

Ross, M.T., LaBrie, S., Mc Pherson, J., Stanton, V.P.: In *Current Protocols in Human Genetics.* (Eds. Dracopoli, N.C., Haines, J.L., Korf, B.R., Moir, D.T., Morton, C.C., Seidman, C.E., Seidman, J.G., and Smith, D.R.) 5.6.1-5.6.5. Wiley, New York (1999)

Rudd, M.K. and Willard, H.F.: Analysis of the centromeric regions of the human genome assembly. *Trends in Genet* **20**: 529-533 (2004)

Rudd, M.K., Wray, G.A., Willard, H.F.: The evolutionary dynamics of α-satellite. *Genome Res* **16**: 88-96 (2006)

Samonte, R.V. and Eichler, E.E.: Segmental duplications and the evolution of the primate genome. *Nat Rev Genet* **3**:65-72 (2002)

Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Maner, S., Massa, H., Walker, M., Chi, M., Navin, N., Lucito, R., Healy, J., Hicks, J., Ye, K., Reiner, A., Gilliam, T.C., Trask, B., Patterson, N., Zetterberg, A., Wigler, M.: Large-scale copy number polymorphism in the human genome. *Science* **305**:525-528 (2004)

Sharp, A.J., Locke, D.P., McGrath, S.D., Cheng, Z., Bailey, J.A., Vallente, R.U., Pertz, L.M., Clark, R.A., Schwartz, S., Segraves, R., Oseroff, V.V., Albertson, D.G., Pinkel, D., Eichler, E.E.: Segmental duplications and copy-number variation in the human genome. *Am J Hum Genet* **77**:78-88 (2005)

Sharma, P., Watson, N., Robson, L., Gallo, J., Smith, A.: Novel chromosome 16 abnormality--der(16)del(16) (q13)t(16;21)(p11.2;q22)--associated with acute myeloid leukemia. *Cancer Genet Cytogenet* **113**: 25-28 (1999)

Shaw, C.J., Withers, M.A., Lupski, J.R.: Uncommon deletions of the Smith-Magenis syndrome region can be recurrent when alternate low-copy repeats act as homologous recombination substrates. *Am J Hum Genet* **75**: 75-81 (2004)

Shaw, C.J. and Lupski, J.R.: Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease. *Hum Mol Genet* **13**:R57-64 (2004)

She, X., Jiang, Z., Clark, R.A., Liu, G., Cheng, Z., Tuzun, E., Church, D.M., Sutton, G., Halpern, A.L., Eichler, E.E.: Shotgun sequence assembly and recent segmental duplications within the human genome. *Nature* **431**: 927-930 (2004)

Shen, M.R., Batzer, M., Deininger, P.: Evolution of the master *Alu* gene(s). *J Mol Evol* **33**:311-320 (1991)

Shimada, M.K., Kim, C.G., Kitano, T., Ferrell, R.E., Kohara, Y., Saitou, N.: Nucleotide sequence comparison of a chromosome rearrangement on human chromosome 12 and the corresponding ape chromosomes. Cytogenet. *Genome Res* **108**: 83-90 (2005)

Shing, D.C., Morley-Jacob, C.A., Roberts, I., Nacheva, E., Coleman, N.: Ewing's tumour: novel recurrent chromosomal abnormalities demonstrated by molecular cytogenetic analysis of seven cell lines and one primary culture. Cytogenet. *Genome Res* **97**: 20-27 (2002)

Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P.J., Cordum, H.S., Hillier, L., Brown, L.G., Repping, S., Pyntikova, T., Ali, J., Bieri, T., Chinwalla, A., Delehaunty, A., Delehaunty, K., Du, H., Fewell, G., Fulton, L., Fulton, R., Graves, T., Hou, S.F., Latrielle, P., Leonard, S., Mardis, E., McPherson, J., Miner, T., Nash, W., Nguyen, C.,Ozersky, P., Pepin, K., Rock, S., Rohlfing, T., Scott, K., Schultz, B., Strong, C., Tin-Wollam, A., Yang, S.-P., Waterston, R.H., Wilson, R.K., Rozen, S., Page, D.C.: The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **19**: 825-837 (2003)

Snijders, A.M., Nowak, N., Segraves, R., Blackwood, S., Brown, N., Conroy, J., Hamilton, G., Hindle, A.K., Huey, B., Kimura, K., Law, S., Myambo, K., Palmer, J., Ylstra, B., Yue, J.P., Gray, J.W., Jain, A.N., Pinkel, D., Albertson, D.G. Assembly of microarrays for genome-wide measurement of DNA copy number. *Nat Genet* **29**:263–264 (2001)

Solinas-Toldo, S., Lampel, S., Stilgenbauer, S., Nickolenko, J., Benner, A., Dohner, H., Cremer, T., Lichter, P.: Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Genes Chromosomes Cancer* **20**: 399-407 (1997)

Stankiewicz, P., Park, S.S., Inoue, K., Lupski, J.R.: 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**: 1205-1210 (2001)

Stankiewicz, P., Shaw, C.J., Withers, M., Inoue, K., Lupski, J.R.: Serial segmental duplications during primate evolution result in complex human genome architecture. *Genome Res* **14**: 2209-2220 (2004)

Stedman, H.H., Kozyak, B.W., Nelson, A., Thesier, D.M., Su, L.T., Low, D.W., Bridges, C.R., Shrager, J.B., Minugh-Purvis, N., Mitchelll, M.A.: Myosin gene mutation correlates with anatomical changes in the human lineage. *Nature* **428**: 415-418 (2004)

Szamalek, J.M., Goidts, V., Chuzhanova, N., Hameister, H., Cooper, D.N., Kehrer-Sawatzki, H.: Molecular characterization of the pericentric inversion that distinguishes human chromosome 5 from the homologous chimpanzee chromosome. *Hum Genet* **117**: 168-176 (2005)

Szamalek, J.M., Cooper, D.N., Schempp, W., Minich, P., Kohn, M., Högel, J., Goidts, V., Hameister, H., Kehrer-Sawatzki, H.: Polymorphic micro-inversions contribute to the genomic variability of humans and chimpanzees. *Hum Genet* (2006)

Szamalek, J.M., Goidts, V., Cooper, D.N., Hameister, H., Kehrer-Sawatzki, H.: Characterization of the human specific pericentric inversion that discriminates human chromosome 1 from its homologous in great apes. (Submitted)

Tuzun, E., Bailey, J.A., Eichler, E.E.: Recent segmental duplications in the working draft assembly of the Brown Norway rat. *Genome Res* **14**:493-506 (2004)

Tuzun, E., Sharp, A.J., Bailey, J.A., Kaul, R., Morrison, V.A., Pertz, L.M., Haugen, E., Hayden, H., Albertson, D., Pinkel, D., Olson, M.V., Eichler, E.E.: Fine-scale structural variation of the human genome. *Nat Genet* **37**:727-732 (2005)

Uddin, M., Wildman, D.E., Liu, G., Xu, W., Johnson, R.M, Hof, P.R., Kapatos, G., Grossman, L.I., Goodman, M.: Sister grouping of chimpanzees and humans as revealed by genome-wide phylogenetic analysis of brain gene expression profiles. *Proc Natl Acad Sci USA* **101**: 2957-2962 (2004)

Ueda, C., Akasaka, T., Kurata, M., Maesako, Y., Nishikori, M., Ichinohasama, R., Imada, K., Uchiyama, T., Ohno, H.: The gene for interleukin-21 receptor is the partner of BCL6 in t(3;16)(q27;p11), which is recurrently observed in diffuse large B-cell lymphoma. *Oncogene* **21**: 368-376 (2002)

Vallender, E.J. and Lahn, B.T.: Effects of chromosomal rearrangements on human-chimpanzee molecular evolution. *Genomics* **84**: 757-761 (2004a)

Vallender, E.J. and Lahn, B.T.: Positive selection on the human genome. *Hum Mol Genet* **13**: 245-254 (2004b)

van Geel, M., Eichler, E.E., Beck, A.F., Shan, Z., Haaf, T., van der Maarel, S.M., Frants, R.R., de Jong, P.J.: A cascade of complex subtelomeric duplications during the evolution of the hominoid and Old World monkey genomes. *Am J Hum Genet* **70**: 269-278 (2002)

Vision, T.J., Brown, D.G., Tanksley, S.D.: The origins of genomic duplications in *Arabidopsis*. *Science* **290**: 2114-2117 (2000)

Visser, R., Shimokawa, O., Harada, N., Kinoshita, A., Ohta, T., Niikawa, N., Matsumoto, N.: Identification of a 3.0-kb major recombination hotspot in patients with Sotos syndrome who carry a common 1.9-Mb microdeletion. *Am J Hum Genet* **76**: 52-67 (2005)

Wandstrat, A.E., Leana-Cox, J., Jenkins, L., Schwartz, S.: Molecular cytogenetic evidence for a common breakpoint in the largest inverted duplications of chromosome 15. *Am J Hum Genet* **62**: 925-936 (1998)

Willard, H.F. and Waye, J.S.: Hierarchical order in chromosome-specific human α satellite DNA. *Trends Genet* **3**: 192-198 (1987)

Willeke, F., Ridder, R., Mechtersheimer, G., Schwarzbach, M., Duwe, A., Weitz, J., Lehnert, T., Herfarth, C., von Knebel Doeberitz, M.: Analysis of FUS-CHOP fusion transcripts in different types of soft tissue liposarcoma and their diagnostic implications. *Clin Cancer Res* **4**: 1779-1784 (1998)

Wilson, G.M., Flibotte, S., Missirlis, P.I., Marra, M.A., Jones, S., Thornton, K., Clark, A.G., Holt, R.A.: Identification by full-coverage array CGH of human DNA copy number increases relative to chimpanzee and gorilla. *Genome Res* Epub (2006)

Won, Y.J. and Hey, J.: Divergence population genetics of chimpanzees. *Mol Biol Evol* **22:** 297-307 (2005)

Yamamoto, K., Nagata, K., Kida, A., Hamaguchi, H.: Deletion of 16q11 is a recurrent cytogenetic aberration in acute myeloblastic leukemia during disease progression. *Cancer Genet Cytogenet* **131**: 65-68 (2001)

Yoder, A.D. and Yang, Z.: Estimation of primate speciation dates using local molecular clocks. *Mol Biol Evol* **17**: 1081-1090 (2000)

Yu, J., Whang, J., Lin, W., Li, S., Li, H. et al.: The genomes of *Oryza sativa*: a history of duplications. *PLoS Biol* **3(2)**: 266-281 (2005)

Yu, N., Jensen-Seaman, M.I., Chemnick, L., Kidd, J.R., Deinard, A.S., Ryder, O., Kidd, K.K., Li, W.H.: Low nucleotide diversity in chimpanzees and bonobos. *Genetics* **164**: 1511-1518 (2003)

Yunis, J.J. and Prakash, O.: The origin of man: a chromosomal pictorial legacy. *Science* **215**:1525-1530 (1982)

Zhang, J., Wang, X., Podlaha, O.: Testing the chromosomal speciation hypothesis for humans and chimpanzees. *Genome Res* **14**: 845-851 (2004)

Zhang, L., Lu, H.H.S., Chung, W.-Y., Yang, J., Li, W.H.: Patterns of segmental duplication in the human genome. *Mol Biol Evol* **22**: 135-141 (2005)

6. SUMMARY

Comparative analysis of human and great ape genomes expose the full spectrum of genomic changes that accompanied human evolution. In particular the concomitant evaluation of divergence and diversity has the power to identify those genes or genomic regions that evolved under selective constraints during evolution and might be associated with human specialization. Recent analyses showed that copy number differences (CNDs), also called low copy repeats (LCRs), of a few kb up to several 100-kb have contributed significantly to the genomic divergence between human and other primates. In order to identify CNDs that occurred specifically in the human lineage, we performed interspecies aCGH including five non-human primate species (macaque, orang-utan, gorilla, chimpanzee and bonobo). The human BAC array used covered about 30 % of the human genome. By this, I was able to identify 14 sites of putative human specific CNDs and I could show that DNA copy number gains are not only important for gene dosage changes but that they are localized at sites of numerous large-scale rearrangements that occurred during human genome evolution. Indeed, several BACs showing human-specific DNA copy number gains in my analysis were located in the breakpoints regions of the human-specific pericentric inversion of chromosome 1 as well as of the human polymorphic inversion of chromosome 2. The presence of LCRs at the borders of such inversions suggests their implication in mediating the respective rearrangements by allowing intrachromosomal homologous recombination between these repeats. Furthermore, the precise characterization of the breakpoints of the pericentric inversions of HSA 18 and of *PTR* XVI and *GGO* XVI allowed me to study the LCR content of the breakpoint regions of these rearrangements. I found that LCRs bordered each breakpoint, confirming the importance of segmental duplications in mediating evolutionary rearrangements. Moreover, I have determined that the pericentric inversions of *PTR* XVI and *GGO* XVI occurred twice independently during evolution verifying the instability of that region suggested previously by Martin et al. (2004).
In conclusion, my study showed that segmental duplications are frequently assigned to unstable genomic regions and to chromosomal breakpoints, emphasizing their role in determining both karyotypic evolution and genomic diversity in humans.

7. APPENDIX

Appendix 1: Summary of aCGH-identified variant sites among five different primate species examined

Chr.	Clone ID	Average Log2 Ratio					Chr.	Clone ID	Average Log2 Ratio				
		PTR	*PPA*	*GGO*	*PPY*	*MFU*			*PTR*	*PPA*	*GGO*	*PPY*	*MFU*
1	RP11-430E19				-0,83	-1,85	**2**	**RP11-397E13**	**-0,57**	**-0,53**	**-0,59**	**-0,75**	**-0,61**
1	RP11-82D16			0,71			2	RP11-432G15			1,4		
1	RP11-90D15	0,91	0,82			-0,9	2	RP11-91A12					0,58
1	RP11-265F14					0,51	2	RP11-91G8					-0,5
1	RP11-297O4	0,72	0,6	0,87			2	RP11-125E20					-0,55
1	**RP11-439A17**	**-1,18**	**-0,7**	**-0,9**	**-0,92**	**-0,99**	2	RP11-91K13			-0,62	-1,02	-0,87
1	**RP11-35B4**	**-0,84**	**-1,1**	**-0,73**	**-1,5**	**-1,73**	2	RP11-89B17					-0,91
1	**RP11-91G11**	**-0,6**	**-0,7**	**-0,65**	**-0,59**	**-0,71**	2	RP11-89C8					-0,75
1	**RP11-89F3**	**-0,73**	**-1,2**	**-0,7**	**-1,32**	**-1,38**	2	RP11-341H1			0,58		
1	**RP11-196G18**	**-0,53**	**-0,5**	**-0,51**	**-0,57**	**-0,56**	2	RP11-79I24					0,57
1	RP11-3B17			0,65			2	RP11-81J13		0,79			-0,87
1	RP11-123O6					0,51	2	RP11-567N19					0,53
1	RP11-123I5			1,25			2	RP11-96M11			0,63		
1	RP11-88H4					-0,52	2	RP11-655A7	0,55		1,16		
1	RP11-927N13					1,41	2	RP11-90C17				0,6	
1	RP11-462C5					-0,56	2	RP11-183N7					0,71
2	RP11-125K7					-0,55	2	RP11-118M12			0,73		
2	RP11-36C8					-0,58	3	RP11-15K13			-0,98		-2,13
2	RP11-797E8					-0,51	3	RP11-202A21					-0,56
2	RP11-499P9					0,51	3	RP11-128A5					-0,68
2	RP11-57F9					0,5	3	RP11-229A12					0,61
2	RP11-299C5					0,54	3	RP11-594C8					-0,57
2	RP11-295P2					0,64	3	RP11-95L24			0,56		
2	RP11-595F15				0,79		3	RP11-161M8			0,57		
2	RP11-568N1		0,51				3	RP11-91A15					-0,57
2	RP11-91K3	-0,92	-0,98		-1,8	-1,97	3	RP11-37D12					-0,51
2	RP11-90H13	-0,5	-0,53		-1,67	-1,81	3	RP11-112O24					0,56
2	RP11-134N21				-1,16	-1,2	3	RP11-90M7					-0,57
2	RP11-708D7			0,74			3	RP11-91F17					-0,54
2	RP11-34G16			-0,69		-0,79	3	RP11-89J17				0,52	
2	RP11-89J19					-0,97	3	RP11-965H19					-0,54
2	RP11-448J8					0,51	3	RP11-11K22					-0,52
2	RP11-315O22	0,61	1,17				4	RP11-478A6					0,6
2	RP11-322F4					-0,67	4	RP11-117J13					0,87
2	**RP11-68E19**	**-0,78**	**-0,86**	**-0,75**	**-0,94**	**-0,96**	4	RP11-427A14	-1,32	-1,29		-1,88	-1,83

Continuation of Appendix 1

4	RP11-751L19					-1,03	7	RP11-366M14			1,05		
4	RP11-91F19					-0,64	7	RP11-160E4			0,67		
4	RP11-89C20			-0,5	-0,89	-0,96	7	RP11-353A8			0,95		
4	RP11-89L19			-0,5	-1,22	-1,42	7	RP11-64I2			0,71		
4	RP11-89D14			-1,85	-2,03	-1,84	7	RP11-112L4			0,63		
4	RP11-11N11	1,16	1,02		-0,92	-1,26	7	RP11-3L22					0,64
4	RP11-89B18					0,68	7	RP11-832D10			0,77		
4	RP11-1021A7				0,58		7	RP11-80L24					0,65
4	RP11-79C8					-0,8	7	RP11-15K19					0,61
4	RP11-92H22					0,52	7	RP11-45N18	0,56				
4	RP11-91O14					-0,54	7	RP11-41F22					0,59
4	RP11-79C14					-0,63	7	RP11-506L7	0,86	0,9		-0,69	
4	RP11-101N17			0,62		-0,93	7	RP11-89L18	0,91	1,16			
4	RP11-96A1					-0,52	7	RP11-336D7					0,52
4	RP11-82L8				-0,61	-0,53	7	RP11-126L15					0,61
4	RP11-33C6	0,78			-0,77	-0,67	7	RP11-213E22	0,73				0,52
4	RP11-134L22					-0,5	7	RP11-92J13	0,55				
4	RP11-199O10					-0,5	7	RP11-140O21			0,76		
4	RP11-194A21					-0,61	7	RP11-144D22			0,59		
5	RP11-773M18					-0,54	7	RP11-137O4					0,56
5	RP11-91D21			-0,62	0,51	-1,51	7	RP11-172L23	0,74				
5	RP11-88L18					-1,83	7	RP11-81I21					0,57
5	RP11-79I8			-0,51	-0,65	-0,7	7	RP11-58F7				-0,52	-0,67
5	RP11-89D24			-0,5	-0,82	-0,54	8	RP11-91J19				-0,75	-0,95
5	RP11-92M7					-0,53	8	RP11-86A14					-0,87
5	RP11-143K14			0,6			8	RP11-240A17					-0,6
5	RP11-190C5				0,66		8	RP11-82K8					-0,64
5	RP11-89L24					0,5	8	RP11-79I19				-0,68	-0,53
5	RP11-233E5					0,55	8	RP11-140K14			0,85		
6	RP11-224B15					-0,56	8	RP11-287P18	0,72	1,03			0,5
6	RP11-91P9			-0,81	-0,95	-1,05	8	RP11-91M20		0,53			
6	RP11-513I15					0,51	8	RP11-122N11					-0,94
6	RP11-25I3					0,57	8	RP11-218N24		0,66			
6	RP11-220P17	0,77	1,03				8	RP11-92C1	0,63				0,8
6	RP11-280O15	1,39	1,95	0,92		-0,54	8	RP11-274K12	0,6				
6	RP11-231F3					-0,51	8	RP11-5J20	0,55				
6	RP11-104D9			0,62	0,5		8	RP11-65O11				0,53	
6	RP11-43B19		-0,69		-0,92	-1,29	8	RP11-91K21				-0,86	-1,03
6	RP11-91O16					-0,51	8	RP11-217N16				0,53	
6	RP11-794H3				-0,54		8	RP11-429B3	0,86	0,68	1,27		
7	RP11-90P13	1,01	0,83	0,84	-0,72	-1,03	8	RP11-117N14					0,74

Continuation of appendix 1

8	RP11-367E12					0,54
8	RP11-92K15			0,68		
8	RP11-257P3	2,25	2,25	1,19		-2,77
8	RP11-219B4				1,27	
8	RP11-90G23	-1,04	-0,91	1,19		-1,96
8	RP11-498C11					0,53
8	RP11-142F22			0,93		
8	RP11-208E21			0,84	0,57	
8	RP11-94M13			0,55		
8	RP11-629O1			0,59		
9	RP11-130C19	1	0,54			
9	RP11-118C13					0,61
9	RP11-113O24			0,85	0,77	
9	RP11-79H1		-0,69		-1,35	-1,5
9	**RP11-79A9**	**-0,88**	**-0,94**	**-0,77**	**-1,7**	**-1,69**
9	RP11-88I18	2,1	2,04	0,72		-0,5
9	RP11-151D14			0,58		
9	RP11-8D10					0,51
9	RP11-65B23			0,82		
9	RP11-228B15			0,71		
9	RP11-17O4					0,59
9	RP11-216L13					0,54
10	RP11-89K18					0,55
10	RP11-289E9					0,63
10	RP11-91D9	2,79	2,22	3,93		
10	RP11-89D1				0,79	
10	RP11-20F24				-0,77	
10	**RP11-81J3**	**-2,12**	**-2,29**	**-1,88**	**-3,19**	**-2,33**
10	**RP11-80B18**	**-0,55**	**-0,51**	**-0,56**	**-0,55**	**-0,64**
10	RP11-91H19	-0,74	-0,66			
10	RP11-80F12				0,51	
10	RP11-69I18					0,63
10	RP11-345K20					0,53
10	RP11-80H11					0,71
10	RP11-105N15					0,6
10	RP11-91A6					0,51
10	RP11-90J11					-0,71
10	RP11-77E11		0,52			0,74
10	RP11-81N3					-0,62
10	RP11-142I8					-0,64
11	RP11-652O18					-0,53
11	RP11-6K5					0,56	
11	RP11-236N9						-0,57
11	RP11-49K17						-0,61
11	RP11-113M18				0,98		
11	RP11-72A10						-0,5
11	RP11-794D23						-0,64
11	RP11-150D18						-0,51
11	RP11-750A9						0,52
11	RP11-125F14						0,57
11	RP11-79A4					0,86	
11	RP11-79G14						-0,6
11	RP11-846G12					-0,8	
11	RP11-684M15	0,62	0,7	0,9			
11	RP11-75H24	-0,51	-0,53				-0,7
11	RP11-80B24						0,74
11	RP11-772K10						0,52
11	RP11-91F5						-1,84
11	RP11-89G10						0,53
11	RP11-91E22						-0,57
11	RP11-141H6						-0,81
11	RP11-179B7						-0,51
11	RP11-51M23						-0,53
11	RP11-356J5					0,67	
11	RP11-215H18						0,53
12	RP11-272L6						-0,53
12	RP11-89L14	-0,56	-0,64			-0,82	-1,01
12	RP11-25D22					0,52	
12	RP11-5J10					0,88	-1,3
12	RP11-496H24					1,07	-0,58
12	RP11-302B13						0,5
12	RP11-793H13	0,5					
12	RP11-51K17						0,53
12	RP11-490H24						0,5
12	RP11-282O18						0,7
12	RP11-809H14	3,54	3,21			-0,68	
13	RP11-77P19					-0,51	-0,73
13	RP11-79H3					-0,58	-0,99
13	RP11-149B7						0,52
13	RP11-20K19						0,55
13	RP11-173B14				0,55		
13	RP11-14C20						-0,52

Continuation of appendix 1

13	RP11-111L17					-0,5
13	RP11-45E4					0,58
13	RP11-3P8					-0,63
13	RP11-19O15					-0,54
14	**RP11-760D22**	**-1,16**	**-1,22**	**-0,87**	**-1,05**	**-1,36**
14	RP11-22O13	0,63				
14	RP11-2K1		0,54			0,81
14	RP11-26M6					0,62
14	RP11-88N14			0,69		
14	RP11-89D19	-0,52				
14	RP11-79M10					-0,56
14	RP11-91A4			-0,64		
14	RP11-1006C22					0,57
14	RP11-364G22			0,62	-0,69	
14	RP11-89H23					0,56
14	RP11-88L4					-0,54
15	RP11-79A23[a]			-0,71	-0,68	-1,53
15	RP11-79C23			-0,86	-0,59	-1,59
15	RP11-90P3[a]		-0,57		-0,95	-1,23
15	RP11-90F19				-0,56	-0,95
15	RP11-313K16		-0,78	-0,57	1,015	-0,56
15	RP11-88P10				-0,55	
15	RP11-256M2					-0,65
15	RP11-94G7					-0,69
15	RP11-115G22	-0,54				-0,57
15	RP11-15C9					-0,51
15	RP11-81J11				0,83	
15	RP11-96M16					0,5
15	RP11-2D17					0,55
15	RP11-90M9					0,5
15	RP11-90J19					-0,87
15	RP11-213H12		2,32			
15	RP11-30K9					0,5
15	RP11-53H4					0,61
15	RP11-54P3					0,5
15	RP11-282P15			0,54		
15	RP11-81A1					0,57
15	RP11-532F12					0,57
15	RP11-429B14				0,93	
15	RP11-49M6				0,73	
16	RP11-243K18					0,62
16	RP11-728H8				1,1	
16	RP11-89M4					0,54
16	RP11-113H7					-1,15
16	RP11-520B23				0,55	0,66
16	RP11-486I11				0,72	
16	RP11-166B2	0,57	0,89	0,56		
16	RP11-89B7				-0,75	-0,64
16	RP11-28A6		0,72		-0,92	-1,54
16	**RP11-499D5**	**-0,72**	**-0,5**	**-0,55**	**-0,72**	**-1,89**
16	RP11-728G11				-0,81	-1,23
16	RP11-80F22					-1,3
16	RP11-91A22				0,51	
16	RP11-294A3					0,58
16	RP11-61I15					-0,53
16	RP11-246M14					-0,51
16	RP11-50C5					-0,51
16	RP11-252A24				-0,52	-0,8
16	RP11-803N24					-0,54
16	RP11-1B11					0,57
16	RP11-178P16					-0,51
17	RP11-79K15					-0,57
17	RP11-29C11					-0,53
17	RP11-52N13					0,56
17	RP11-403G3	0,6				-1,38
17	RP11-201E5					0,52
17	RP11-61B11					0,53
18	RP11-1011E7				0,52	
18	RP11-933M19	0,78	1,19			-0,74
18	RP11-91K22				-0,77	-0,51	-1,31
18	RP11-91A8	-0,9	-1,12		-1,03	-1,01
18	RP11-63N12					-0,52
18	RP11-75O12					-0,64
18	RP11-88B2					-0,51
18	RP11-79A24					-0,54
18	RP11-1094F6				-0,56	
18	RP11-7H17				-0,52	
19	RP11-330I7					0,55
19	RP11-79F15					-0,79
19	RP11-89N18					0,51
19	RP11-206O6					-0,57
19	RP11-152P7					-0,59

Continuation of appendix 1

19	RP11-687M15	0,91		0,71			21	RP11-114H1							-0,52
19	RP11-423F16					-0,55	21	RP11-88N2				0,68			
20	RP11-530N10			0,6			**22**	**RP11-915P6**	**-0,8**	**-1,24**	**-0,85**	**-1,36**	**-1,76**
20	RP11-40J24					-0,53	**22**	**RP11-81H21**	**-0,97**	**-1,02**	**-1,11**	**-0,85**	**-1,12**
20	RP11-7F10			0,56			22	RP11-36P14			0,93		
20	RP11-32N12					-0,51	22	RP11-562F10			0,75		
20	RP11-57G7			0,7			22	RP11-89J4					0,52
20	RP11-69I10					0,54	X	RP11-399D12	-1,49	-1,7	1,03	-1,87	-2,36
20	RP11-256F8			0,52			X	RP11-180F16			1,45		
20	RP11-1105M4			1,02			X	RP11-528A24	-0,58			-0,92	-1,03
20	RP11-266K16				-0,92	-0,91	X	RP11-431N15				0,62	
21	RP11-795L14	-0,88	-1,22	1,6	-0,87	-1,2	X	RP11-402P6			-0,95		
21	RP11-47B13				-0,72	-1,43	X	RP11-561J7					0,5
21	RP11-74D19					-0,57	X	RP11-357N12					-0,65
21	RP11-66C14					-0,55	X	RP11-388L20					-0,59
21	RP11-1P3					-0,71							

[a]BACs RP11-79A23 and RP11-79C23 represent the same CND whereas BACs RP11-90P3 and RP11-90F19 correspond to another CND. The arrayed clones showing HLS-CNDs are shown in bold. Concordantly variant neighbouring clones separated by <250-kb are highlighted in grey

Appendix 2: Summary of the duplication clusters (DCs) exhibiting high similarity (>90 %; >40-kb) to the DCs specific for the arrayed BAC showing human-specific copy number differences

Arrayed BAC clone	DC specific to this BAC[a]	Chromosomal position of the specific DC	Length of DC (bp) [a]	Homologous DCs	Approximate length of the homologous DC	Chromosomal position of the homologous DC
RP11-439A17	DC0154	1p11.2	189.538	DC0165	74 kb	1q12
				DC0215	91 kb	1q21.1
				DC0218	46 kb	1q21.2
				DC0573	125 kb	R[b]
RP11-196G18	DC0218	1q21.1	99.344	DC0154	46 kb	1p11.2
				DC0215[f]	84 kb	1q21.1
RP11-91G11	DC0203	1q21.1	43.030	DC0216	43 kb	1q21.1
	DC0204	1q21.1	116.988	DC0168	81 kb	1q21.1
				DC0206	81 kb	1q21.1
				DC0207[f]	81 kb	1q21.1
				DC0216	86 kb	1q21.1
				[DC1707, DC1708, DC1709] [c]	82 kb	5cen
RP11-89F3	DC0206	1q21.1	131.160	DC0168	54 kb	1q21.1
				DC0204	81 kb	1q21.1
				DC0207	112 kb	1q21.1
				DC0216	60 kb	1q21.1
				DC1708	46 kb	5cen
	DC0207	1q21.1	220.312	DC0168	54 kb	1q21.1
				DC0204	81 kb	1q21.1
				DC0206	112 kb	1q21.1
				DC0216	60 kb	1q21.1
				DC0397[e]	95 kb	2cen
				[DC1707, DC1708] [c]	53 kb	5cen
RP11-68E19	DC0445[e, f]	2q13	266.142	DC0374[e]	266 kb	2p11.2
	DC0447	2q13	77.357	DC0371[f]	77 kb	2p11.2
RP11-397E13	DC0448	2q13	162.865	DC0367[d]	44 kb	2p11.2
				DC0371[f]	119 kb	2p11.2
RP11-79A9	DC2828	9q12-13	371.445	DC0398[f]	58 kb	2cen
				DC0400	84 kb	2cen
				DC0401	105 kb	2cen
				DC1800[f]	40 kb	22cen
				DC2078	41 kb	21q11.2
				DC2812	193 kb	9p11.2
				DC2813	87 kb	9p11.2
				DC2821[d, f]	193 kb	9q12
				DC2822[f]	181 kb	9q12
				DC2830	136 kb	9q13
				DC3059[f]	45 kb	10cen
				DC3061	60 kb	10q11.21
				[DC4100, DC4101] [c]	117 kb	14q12
RP11-80B18	DC3093	10q11.22	237.236	DC3106[f]	236 kb	10q11.22
				DC3108[f]	122 kb	10q11.22

Continuation of Appendix 2

RP11-499D5	DC3591[f]	16p11.2	584.250	[DC3287, DC3288][c, d]	53 kb	17q12
				[DC3585, DC3586][c, d, f]	330 kb	16p11.2
				DC3591[d, f]	180 kb	16p11.2
				DC3592[d]	49 kb	16p11.2
				DC3598[f]	123 kb	16p11.2
				DC3600[f]	102 kb	16p11.2
				[DC4134, DC4135, DC4136][c, e]	67 kb	14q32.33
RP11-760D22	DC4092[f]	14q11.2	470.695	[DC0458, DC0459, DC0460][c, d, e, f]	62 kb	2q21.1
				DC0463[f]	49 kb	2q21.1
				[DC0472, DC0473][c, d]	140 kb	2q21.1
				DC0474	59 kb	2q21.2
				DC1786[e, f]	199 kb	22_cen
				DC1787[d, f]	167 kb	22_cen
				DC2077[f]	43 kb	21q11.2
				DC2916	44 kb	18p11.21
				DC3797[e, f]	44 kb	15q11.2
				DC3805[e]	66 kb	15q11.2
				[DC4090, DC4091][c, f]	375 kb	14_cen
RP11-81H21	DC1787[d, f]	22q11.1	216.815	[DC0465, DC0466][c]	45 kb	2q21.1
				DC4090	132 kb	14cen
				DC4091[d]	78 kb	14cen
				DC4092	167 kb	14q11.2
RP11-915P6	DC1786[e, f]	22q11.1	234.225	DC0463[f]	44 kb	2q21.1
				DC0465[f]	44 kb	2q21.1
				[DC0472, DC0473, DC0474][c, d]	224 kb	2q21.1
				DC2077[f]	90 kb	21q11.2
				DC2916	90 kb	18p11.21
				DC3797[e, f]	90 kb	15q11.2
				DC4091[f]	234 kb	14cen
				DC4092	199 kb	14cen

[a]According to the Human Genome Segmental Duplication Database, based on the May 2004 Human Genome Assembly (hg17)

[b]R = random chromosome, not assigned to a specific locus

[c]Duplication clusters comprising a larger cluster

[d]Covered by the array used by Iafrate et al. (2004)

[e]Identified as CNP by Sebat et al. (2004)

[f]Covered by the array used by Sharp et al. (2005)

DCs exhibiting HLS-CNDs that are highly similar to one another are grouped in a super-family and are highlighted in grey.

Appendix 3: Comparative investigation of the gene content of human BACs using the Pearson χ^2 test ($p < 0.005$)

Number and category of BACs used for the Pearson χ^2 test	BAC clone	Accession number	Position[a]	Number of UniGenes[a]
20 BACs showing a gorilla-specific CND	RP11-3B17	AC012663	1q31	1
	RP11-432G15	AC017074	2q13	12
	RP11-341H1	AC016725	2q21	11
	RP11-95L24	AQ316684-AQ316687	3p12	5
	RP11-161M8	AQ379219-AQ379221	3p12	0
	RP11-101N17	AQ320386-AQ320389	4q26	11
	RP11-143K14	AQ389851-AQ389853	5q11.2	4
	RP11-160E4	AC073629	7p21	1
	RP11-353A8	AQ532989-AQ532986	7p21	1
	RP11-140K14	AQ386711-AQ386725	8p23	1
	RP11-90G23	AQ281632-AQ281629	8q21.3	1
	RP11-151D14	AL591031	9q21	3
	RP11-113M18	AQ341786-AQ372515	11p14	0
	RP11-173B14	AL137244	13q22	7
	RP11-88N14	AQ281568-AQ281565	14q21	0
	RP11-282P15	AC010672	15q25	8
	RP11-728H8	AC067772	16p13.3	9
	RP11-1011E7	AQ668764-AQ679338	18p11.3	2
	RP11-530N10	AC011857	20p13	6
	RP11-36P14	AQ045891-AQ045894	22q11.1	4
				87
1st group of 20 BACs chosen randomly	RP11-186E11	AC104691	4q31	4
	RP11-434P1	Z97056	19q13.3	3
	RP11-326K13	AC009831	18q12	2
	RP11-645N11	AC073127	7q22	6
	RP11-361I14	AL087242	12p12	4
	RP11-193D23	AL590733	6q22	5
	RP11-1079K10	AC091180	17q22	12
	RP11-85G20	AL132838	14q32	11
	RP11-473E2	AL669970	9q34	5
	RP11-867G2	AP000943	11q21	8
	RP13-44L19	AL671884	6q16	3
	RP11-707P17	AC066613	15q21	4
	RP11-490A22	AC107216	4q34	4
	RP11-23B15	AL499604	9q22	9
	RP11-526G5	AC069025	10p13	3
	RP11-11K9	AC016465	8q13	2
	CTB-33G10	AC011495	19q13.3	3
	RP11-750I4	AL355074	14q32	20
	RP11-48O9	AC107939	11p13	6
	RP11-480I12	AC098934	1q32	14
				128
2nd group of 20 BACs chosen randomly	RP4-802A10	AC096947	1p32	13
	RP11-681B22	AC093684	2q24	2
	RP11-839L21	AC098482	3p14	9
	RP11-446F17	AL591242	6p12	7
	RP11-373A6	AL162388	9p21	4
	RP1-166H1	AC003982	12q24.2	5

Continuation of Appendix 3

	RP11-452J21	AC108063	4p15.3	6
	RP11-65N13	AL354710	9q33	5
	RP11-104E19	AL136018	14q11.2	4
	RP11-404G1	AC025335	7p13	5
	RP11-249M12	AC126767	5q23	3
	RP11-637F16	AC021636	8q24.3	7
	RP11-384P18	AL390920	10q26	7
	RP11-318C2	AC026894	11p15	10
	RP11-14P21	AC011248	12q23	2
	RP11-351N4	AL596092	13q12.3	2
	RP11-315D16	AC107871	15q23	8
	RP11-462D18	AL132765	20p12	7
	RP11-529H20	AL049872	14q32	7
	CTB-187L3	AC011492	19p13	5
				118
3rd group of 20 BACs chosen randomly	RP11-51I15	AL391595	6q14	2
	RP11-736P16	AL358292	14q31	1
	RP11-392P7	AC007688	12p13	11
	RP11-936I5	AC111182	17q25	9
	RP11-191H13	AC022721	8q22	4
	RP11-90B9	AC022558	15q25	9
	RP11-549E17	AL451045	9p21	0
	RP11-327P2	AL162377	13q14	15
	CTD-2396E7	AC010503	19p13.3	5
	RP11-251A15	AL359707	10q23	5
	RP11-157G15	AL354659	1q41	3
	CTD-2530N21	AC105206	8p21	17
	RP11-541G9	AC007437	12q22	1
	RP1-101G11	AL021877	22q12	0
	RP11-640A1	AC100843	18q21	4
	RP11-233C13	AC018988	15q26	7
	RP3-351K20	AL109939	6q22	5
	RP11-147H23	AL136525	13q14	7
	RP5-1111F22	AC004967	7q21	4
	CTD-3074O7	AP002748	11q13	13
				122

[a]according to NCBI Map Viewer (*Homo sapiens* Build 35.1)

The test has been made comparing the number of genes on the human BACs showing *GGO*-specific CNDs with each group of 20 BACs, separately. The test used is the Pearson χ^2 test, using the formula:

$$\chi^2 = \frac{(\text{Observed frequency} - \text{Expected frequency})^2}{\text{Expected frequency}}.$$

The test was significant with dl=1 and p<0,005

Acknowledgements

First of all, I would like to thank Prof Dr. W. Vogel, Department of Human Genetics for allowing me to join his department.

I am also grateful to Prof. Dr. H. Hameister for the opportunity to work in his lab and all the rewarding experiences I have acquired during my Ph.D. thesis.

Specials thanks to Dr. H. Kehrer-Sawatzki for her expert advices and critical reading of this manuscript.
Hilde, I am very thankful to you for your patience in answering my endless questions and for your support throughout the whole project. Your enthusiasm, your motivation, your sense of organization and your encouragement has been the basis of this work.

I am grateful to Prof. Dr. P.J. de Jong, Prof. Dr. W. Schempp, Prof. Dr. D.N. Cooper, Prof. Dr. X. Estivill, Dr. J. Conroy, Dr. W. Enard, Dr. S. Müller, and Dr. N. Chuzhanova for their critical discussion and helpful advices during the preparation of the different manuscripts we have written. And of course a very special thank to Lluis Armengol. Thank you for being such a ready-to-help person. Thanks also for the fun evenings in Bari and for showing me around the CGR. It was a real pleasure to work with you!!

Thanks to all the members of the lab for being such a great team: Matthias Kohn, Justyna Szamalek, and particularly Steffi Raith, Antje Kollack, and Helene Spöri (I did not forget to save...) for their expert technical advises and their help in performing the experiments. Thanks to all of you for your support, the discussions, and for your friendship beyond professional work!

Last but not least many thanks to my family (my parents, my brothers Sam and David, and my sister Claire) and my friends (particularly Aurore, Béné and Rob) for remembering me that there is something in life beyond work. Special thanks to my husband Andreas, who has lived with me through all the ups and downs of this thesis. Thanks a lot!!!

MERCI ...

Publications

Publications

Goidts V., Szamalek J., Hameister H., Kehrer-Sawatzki H.
Segmental duplication associated with the human specific inversion of chromosome 18: a further example of the impact of segmental duplications on karyotype and genome evolution in primates
Hum Genet. 2004 Jul; 115: 116-122 (IF-2004: 4.328)

Kehrer-Sawatzki H., Sandig C., **Goidts V.**, Hameister H.
Breakpoint analysis of the pericentric inversion between chimpanzee chromosome 10 and the homologous chromosome 12 in humans
Cytogenet Genome Res. 2005; 108: 91-97 (IF-2004: 1.341)

Kehrer-Sawatzki H., Sandig C., Chuzhanova N., **Goidts V.**, Szamalek J., Tänzer S., Müller S., Platzer M., Cooper D., Hameister H.
Breakpoint analysis of the pericentric inversion distinguishing human chromosome 4 from the homologous chromosome in the chimpanzee (*Pan troglodytes*)
Hum Mutat. 2005 Jan; 25: 45-55 (IF-2004: 6.845)

Szamalek J., **Goidts V.**, Chuzhanova N., Hameister H., Cooper D., Kehrer-Sawatzki H.
Molecular characterization of the pericentric inversion that distinguishes human chromosome 5 from the homologous chimpanzee chromosome
Hum Genet. 2005 Jul; 117: 168-176 (IF-2004: 4.328)

Goidts V., Szamalek J., de Jong P., Cooper D., Chuzhanova N., Hameister H., Kehrer-Sawatzki H.
Independent intrachromosomal recombination events underlie the pericentric inversion of chimpanzee and gorilla chromosomes homologous to human chromosome 16
Genome Res. 2005 Sep; 15: 1232-1242 (IF-2004: 10.382)

Szamalek J., **Goidts V.**, Searle J., Cooper D., Hameister H., Kehrer-Sawatzki H.
The chimpanzee-specific pericentric inversions that distinguish humans and chimpanzees have identical breakpoints in *Pan troglodytes* and *Pan paniscus*
Genomics 2006 Jan; 87: 39-45 (IF-2004: 3.840)

Szamalek J., Cooper D., Schempp W., Minich P., Kohn M., Högel J., **Goidts V.**, Hameister H., Kehrer-Sawatzki H.
Polymorphic micro-inversions contribute to the genomic variability of humans and chimpanzees
Hum Genet. 2006 (IF-2004: 4.328)

Goidts V., Armengol L., Schempp W., Conroy J., Nowak N., Müller S., Cooper D., Estivill X., Enard W., Szamalek J., Hameister H., Kehrer-Sawatzki H.
Identification of large-scale human-specific copy-number variations by inter-species array comparative genomic hybridization
Hum Genet. 2006 (IF-2004: 4.328)

Manuscript in preparation

Szamalek J., **Goidts V.**, Cooper D., Hameister H., Kehrer-Sawatzki H.
Characterization of the human specific pericentric inversion that discriminates human chromosome 1 from its homologous in great apes.

Posters

The human genome: a conserved gene arrangement allows development of the most complex traits.
Hameister H., Kohn M., **Goidts V.**, Szamalek J., Kehrer-Sawatzki H.
Gordon Research Conference: Molecular evolution, California, USA
February 2004.

Segmental duplication associated with the human specific inversion of chromosome 18: further example of the impact of segmental duplications on karyotype and genome evolution in primates
Goidts V., Szamalek J., Hameister H., Kehrer-Sawatzki H.
European Human Genetics Conference, Munich, GERMANY
June 2004

Intrachromosomal recombination within pericentromeric HSATII repeats, the mechanism underlying the pericentric inversion of the chimpanzee chromosome 18, homologous to human chromosome 16
Goidts V., Szamalek J., Hameister H., Kehrer-Sawatzki H.
European Human Genetics Conference, Munich, GERMANY
June 2004

The nine pericentric inversions that distinguish the karyotype of human and chimpanzee predate the separation of *Pan paniscus* (bonobo) and *Pan troglodytes* lineages
Szamalek J., **Goidts V.**, Hameister H., Kehrer-Sawatzki H.
European Human Genetics Conference, Munich, GERMANY
June 2004

Precise characterization of pericentric inversions that distinguish human and chimpanzee genomes
Goidts V., Szamalek J., Sandig C., Chuzhanova N., Tänzer S., Müller S., Platzer M., Cooper D., Hameister H., Kehrer-Sawatzki H.
Gordon Research Conference: Molecular cytogenetics, Oxford, UK
July 2004

Evolution in the light of genome conservation: sex chromosomes direct morphologic and functional diversification
Kohn M., **Goidts V.**, Szamalek J., Kehrer-Sawatzki H., Hameister H.
3rd Conference on Neurogenetics, Weimar, GERMANY
September 2004.

Identification of human-specific DNA copy-number differences compared to non-human primates as determined by BAC CGH-array
Goidts V., Szamalek J., Armengol L., Schempp W., Conroy J., Enard W., Müller S., Hameister H., Kehrer-Sawatzki H.
Tagung der Deutschen Gesellschaft für Humangenetik, Halle/Saale, GERMANY
March 2005 (Poster Award)

The contribution of multiple pericentric inversions to human/chimpanzee speciation
Szamalek J., **Goidts V.**, Assum G., Högel J., Hameister H., Kehrer-Sawatzki H.
Tagung der Deutschen Gesellschaft für Humangenetik, Halle/Saale, GERMANY
March 2005

Identification of human-specific DNA copy-number differences compared to non-human primates as determined by BAC CGH-array
Goidts V., Szamalek J., Armengol L., Schempp W., Conroy J., Enard W., Müller S., Hameister H., Kehrer-Sawatzki H.
18th European course in Medical Genetics, Betinoro, ITALY
May 2005

Identification of lineage specific and polymorphic copy-number variations in the human genome by inter- and intra-species array comparative genomic hybridization (aCGH)
Goidts V., Schempp W., Armengol L., Conroy J., Hameister H., Kehrer-Sawatzki H.
Tagung der Deutschen Gesellschaft für Humangenetik, Heidelberg, GERMANY
March 2006

Characterization of human specific copy number gains by comparative genomic hybridization using a customized duplication cluster array.
Goidts V., Cooper D., Conroy J., Schempp W., Hameister H., Kehrer-Sawatzki H.
American Society for Biochemistry and Molecular Biology: DNA Structure, Genomic Rearrangements, and Human Disease meeting, Houston, Texas, USA

Characterization of the human-specific pericentric inversion that discriminates human chromosome 1 from the homologous chromosome in great apes.
Szamalek J., **Goidts V.**, Hameister H., Kehrer-Sawatzki H.
Tagung der Deutschen Gesellschaft für Humangenetik, Heidelberg, GERMANY
March 2006

Polymorphic microinversions contribute to the genomic variability of humans and chimpanzees.
Szamalek J., Cooper D., Schempp W., Minich P., Kohn M., Högel J., **Goidts V.**, Hameister H., Kehrer-Sawatzki H.
Tagung der Deutschen Gesellschaft für Humangenetik, Heidelberg, GERMANY
March 2006

Talks

Breakpoint analysis of pericentric inversions distinguishing human from chimpanzee chromosomes
Kehrer-Sawatzki H., **Goidts V.**, Szamalek J., Hameister H.
Human Genome Meeting (HUGO), Berlin, GERMANY
April 2004

Precise characterization of pericentric inversions that distinguish human and chimpanzee genomes
Szamalek J., **Goidts V.**, Sandig C., Chuzhanova N., Tänzer S., Müller S., Platzer M., Cooper D., Hameister H., Kehrer-Sawatzki H.
15th International Chromosome Conference, London, UK
September 2004

Looking back: From the human karyotype to the ancestral vertebrate proto-genome
Hameister H., Szamalek J., **Goidts V.**, Kohn M., Assum G., Högel J., Kehrer-Sawatzki H.
Tagung der Deutschen Gesellschaft für Humangenetik, Halle/Saale, GERMANY
March 2005

Identification of large-scale human-specific copy-number variations by inter-species aCGH
Goidts V., Armengol L., Schempp W., Conroy J., Müller S., Enard W., Szamalek J., Cooper D., Estivill X., Hameister H., Kehrer-Sawatzki H.
2nd Marie Curie Conference on arrayCGH and Molecular Cytogenetics, Bari, ITALY
October 2005

Award

Preis des Jahres 2005 für das beste Poster aus dem Themengebiet “Human Genetics”
Tagung der Deutschen Gesellschaft für Humangenetik, Halle/Saale, GERMANY
March 2005

Curriculum Vitae

PERSONAL DETAILS

Name	Violaine GOIDTS
Citizenship	Belgian
Date of birth	06.08.1981
Address	University of Ulm Department of Human Genetics Albert Einstein Allee, 11 89081 Ulm, GERMANY Phone: 49 (0) 731 500 23386 Fax: 49 (0) 731 500 23438

EDUCATION

2003 – 2006 University of Ulm, GERMANY
Department of Human Genetics
Prof. Dr. W. Vogel

Candidate for Ph.D. degree in Human Genetics (Dr. Hum. Biol.). Thesis under Prof. Dr. H. Hameister on "Identification of Large-Scale DNA Copy Number Differences Between Human and Non-Human Primate Genomes and their Role in Mediating Evolutionary Rearrangements".

02/2003 – 05/2003 Dana Farber Cancer Institute, Cancer Biology, Harvard Medical School, Boston, MA, USA
Department of Genetics
Prof. M. Vidal

Diploma thesis under Prof. Dr. J. Vandenhaute on "Participation to the realisation of the RNAiome library and its use for a screening on *Caenorhabditis elegans* in order to understand the TGF-β pathway."

09/2001 – 02/2002 Erasmus Exchange program with the University of Uppsala, SWEDEN.

Developmental Biology at Prof. R. Ohlsson's lab.
Molecular Biology at Prof. S. Dasgupta's lab.

1999 – 2003 Facultés Universitaires Notre Dame de la Paix, Namur, BELGIUM.
Undergraduate studies in biology (Dpt. of Molecular Biology)

Human Genetics: Prof. Vekemans, Necker Hospital-Enfants Malades, Paris.
Functional genomics: Prof. Vidal, Dana Farber Cancer Institute, Boston.
Parasitology: Prof. Dujardin, Antwerpen.
Transduction of the eukaryotic message: Prof. Renard, Namur.
Prokaryote's genomics: Prof. De Bolle, Namur.

1993 – 1999 Institut Saint Louis, Namur, BELGIUM.

Orientation: Greek, Latin, Dutch and Mathematics.