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

Cytogenetic and Molecular Characterization of the Macro- and Micro-inversions, which Distinguish the Human and the Chimpanzee Karyotypes - from Speciation to Polymorphism

Thesis
Applying for the Degree of Doctor of Human Biology (Dr. hum. biol.)
Faculty of Medicine
University of Ulm

Presented by
Justyna Monika Szamalek
from Września in Poland

2006

Amtierender Dekan: Prof. Dr. Klaus-Michael Debatin

1. Berichterstatter: Prof. Dr. med. Horst Hameister

2. Berichterstatter: Prof. Dr. med. Konstanze Döhner

Tag der Promotion: 28.07.2006

Content

1. Introduction 7
1.1. Primate phylogeny 7
1.2. Africa as the place of human origin and the living area of the present-day chimpanzee populations 9
1.3. Cytogenetic and molecular differences between human and chimpanzee genomes 10
1.4. Cytogenetic and molecular differences between common chimpanzee and bonobo genomes 17
1.5. Theory of speciation 18
1.6. Theory of selection 20
1.7. Aims of the work presented in this thesis 21

2. Materials and Methods 22
2.1. Materials 22
2.2. Methods 36

3. Results 47
3.1. Characterization of the chimpanzee lineage-specific pericentric inversion of chromosome 5 47
3.2. Characterization of the human lineage-specific pericentric inversion of chromosome 1 54
3.3. Verification of identity of the chimpanzee-specific pericentric inversions between *Pan troglodytes* and *Pan paniscus* 66
3.4. Determination of the human-chimpanzee DNA divergence rate within inverted and collinear regions 71
3.5. Genome-wide search for micro-inversions by gene order comparison between human and chimpanzee 75

4. Discussion 84
4.1. Comparison of the pericentric inversions of chromosomes 1 and 5 84
4.2. Determination of the time interval within which the chimpanzee lineage-specific inversions became introduced 89
4.3. Testing of the chromosomal speciation theory 90

4.4. Micro-inversions as a newly discovered class of human-chimpanzee divergence 94
4.5. Inversions and their influence on neighbouring genes 97
4.6. Conclusions 99

5. Summary 101

6. References 103

7. Appendix 118

8. Acknowledgements 131

9. Curriculum vitae 132

Abbreviations

aCGH	Array based comparative genomic hybridization
AIDS	Acquired immune deficiency syndrome
Alu	Short stretch of DNA characterized by the action of the Alu restriction endonuclease isolated from *Arthrobacter luteus*
BAC	F-factor-based bacterial artificial chromosome
bp	Base pair
BSA	Bovine serum albumin
CCD	Charge-coupled device
CND	Copy number difference
CNV	Copy number variation
CNP	Copy number polymorphism
DAPI	4,6 Diamidino-2-phenylindole
DC	Duplication cluster
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
dUTP	Deoxyuridine triphosphate
EBV	Epstein-Barr virus
EDTA	Ethylene diaminetetraacetic acid
ERV	Endogenous retroviral sequences
F1	First filial generation, in Mendelian genetics means the hybrid offspring of purebred parental generation
FCS	Fetal calf serum
FISH	Florescence *in situ* hybridization
FITC	Fluorescein isothiocyanate
GGO	*Gorilla gorilla* (Gorilla)
HSA	*Homo sapiens* (Human)
HERV-K	Human endogenous retrovirus K
ISCN	International System for Human Cytogenetic Nomenclature
kb	Kilobase
L1	Family of long transposable sequences belonging to LINEs
LCR	Low copy repeat
LINE	Long interspersed nuclear elements

LTR	Long terminal repeat
Mb	Megabase
MFA	*Macaca fascicularis* (Macaque)
Mya	Million years ago
NAHR	Non-allelic homologous recombination
NHEJ	Non-homologous end joining
PAC	P1-derived artificial chromosome
PBS	Phosphate buffered saline solution
PCR	Polymerase chain reaction
PHA	Phytohemagglutinin, an initiator of mitosis
PPA	*Pan paniscus* (Bonobo)
PPY	Pongo pygmaeus (Orangutan)
PtERV	*Pan troglodytes* endogenous retrovirus
PTR	*Pan troglodytes* (Common chimpanzee)
RT	Room temperature
SD	Segmental duplication
SDS	Sodium dodecyl sulfate
SINE	Short interspersed nuclear elements
SSC	Saline sodium citrate solution
TE	Tris-EDTA
Tris	Tris (hydroxymethyl) aminomethane

1. Introduction

It was postulated by Goodman in 1999 that, if the taxonomic classification were based exclusively upon genomic DNA sequence similarity, the nearly 99% identity of human, chimpanzee and bonobo genomes would require a reclassification of the latter two into the genus *Homo*. However, whereas there are many similarities in the biology, life history, and the behaviour of humans and great apes there are also many striking variations that need to be explained. The mostly recognized and often mentioned differences are the bipedalism and the developed speech, which are exclusive to humans, but also differences in susceptibility to diseases such as AIDS or malaria, and above all the large brain size and extended cognitive capacities that unique human characteristics (reviewed in Olson and Varki 2003; Carroll 2003; Varki and Altheide 2006).

1.1. Primate phylogeny

More than 300 species of over 60 genera were classified as extant members of the order Primates. The data obtained from comparative sequencing approaches suggest that Primates divide into the *Strepsirrhini* (e.g. lemurs, lorises), the *Tarsiiformes* (tarsiers) and the *Anthropoidea*. The letter ones split into the *Platyrrhini*, which are represented by New World monkeys, and the *Catarrhini* that consist of Old World monkeys and apes (*Hominoidea*) (Goodman et al. 2005; Figures 1 and 2). The story of the hominoid taxonomy is one of gradual demotion of humans from a special position in the taxonomy to being one branch among many. Although according to the modern taxonomy, based on the genetic similarities, humans are included in the group of apes, in this work I will refer to humans and great apes or African great apes separately, in the now colloquial sense (Figure 1). The close evolutionary relationship of humans with the African great apes i.e. chimpanzees and gorillas was posited already more than 100 years ago (Darwin 1871; Huxley 1863). In the traditional theory based on anatomical and mental differences between these species, the chimpanzees and the gorillas were grouped closer to the orangutans rather then to humans. In fact, there is already convincing DNA sequence evidence, that not only are chimpanzees and gorillas more closely related to humans than to orangutans but also chimpanzees and humans are more closely related to each other than either is to gorillas (Goodman et al. 2005). Despite the cladistics nuisances, there is general agreement that human (*Homo sapiens*) shared a common ancestor with chimpanzee (*Pan troglodytes*) and bonobo (*Pan paniscus*) ~5-7 million years ago (Mya), with gorilla

(*Gorilla gorilla*) ~7-8 Mya, and with orangutan (*Pongo pygmeus*) ~12-13 Mya (Yoder and Yang 2000; Chen and Li 2001; Brunet et al. 2002).

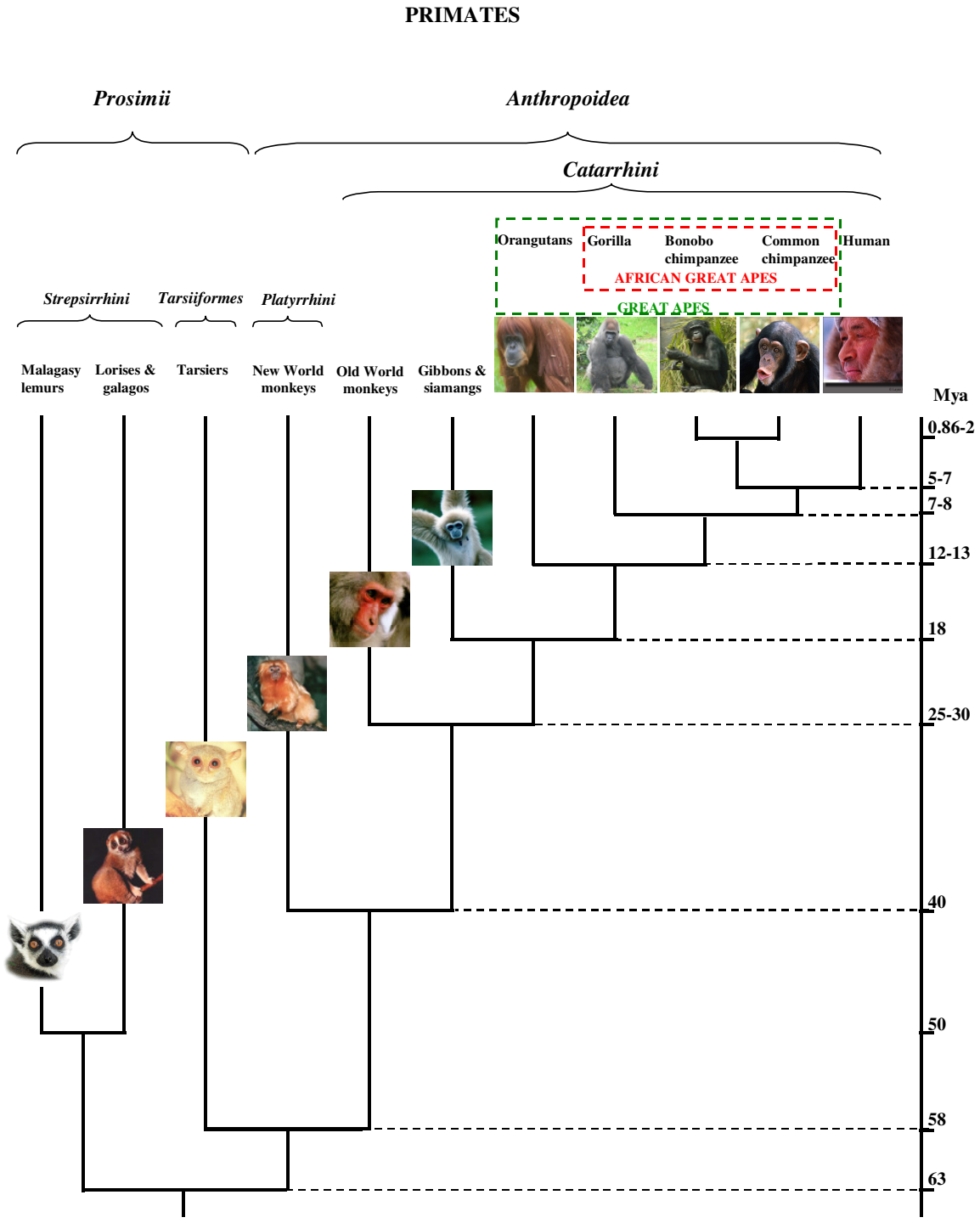

Figure 1: Primates phylogenetic tree. On the right axis the approximate time of divergence between the lineages is given (Goodman et al. 2005 and references therein).

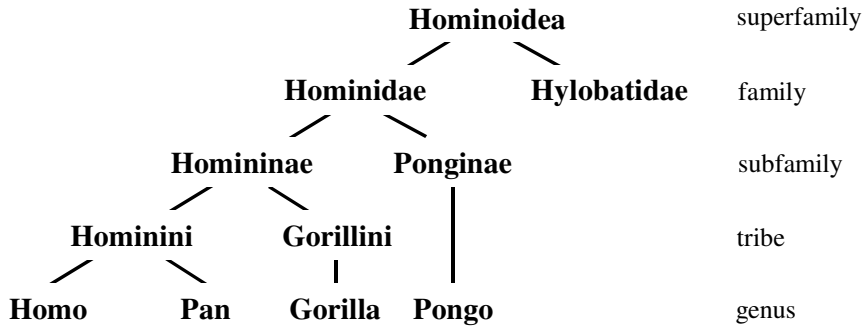

Figure 2: Hominoids taxonomy based on DNA comparisons (Goodman et al. 1990, 2005).

1.2. Africa as the place of human origin and the living area of the present-day chimpanzee populations

Bio-geographic and anthropological evidences suggest that early forms of humans and chimpanzees are likely to have co-inhabited the same region of East Africa. The forested tropical regions of Africa have been always a geographical environment of *Pan troglodytes*, the common chimpanzee. Whereas the humid condition of wooded tropical areas did not favour the preservation of chimpanzee fossils, most human fossil sites occur in the semi-arid East African Rift Valley. This situation has caused speculations regarding reasons for the divergence of the human and the chimpanzee lineages. Some investigators have invoked a shift from wooded to savannah vegetation in East Africa, to explain the apparent separation between chimpanzee and human ancestral populations and the origin of the bipedalism (Darwin 1871; Washburn 1968; Kortlandt 1972; Pilbeam and Young 2004). Recently McBrearty and Jablonski (2005) reported the first chimpanzee fossil, which was found in the Kapthurin Formation in Kenya. This discovery shows that representatives of *Pan* were present in the East African Rift Valley during the Middle Pleistocene, where they were contemporary with an extinct species of *Homo*.

Whereas according to the Out-of-Africa theory the human populations dispersed from Africa less than 200,000 years ago and progressively colonized other continents, the wild chimpanzee populations are still to find exclusively in their place of origin. Although in the last decades of the 20th century their numbers and living area shrunk dramatically (Gagneux et al. 2001), the chimpanzee populations are still distributed in Equatorial Africa between 13° N and 7° S. Population studies of the common chimpanzee have revealed the existence of several distinct subspecies: central - *Pan troglodytes troglodytes* and *Pan troglodytes vellerosus*, western - *Pan troglodytes verus* and eastern - *Pan troglodytes*

schweinfurthii (Morin et al. 1994; Gonder et al. 1997; Gagneux et al. 2001, 2002). The range of their sister species, the bonobos which are also called pygmy chimpanzees, do not overlap with ranges of the common chimpanzees. Bonobos distribution in the wild is limited to the Democratic Republic of Congo, south of the River Congo (Figure 3). As the time divergence between bonobos and common chimpanzees was estimated at 0.86 – 2 Mya (Yoder and Yang 2000; Won and Hey 2005), and the River Congo creates a natural geographical barrier between their territories, it is not surprising that they exhibit many distinct behavioural, morphological and genetic traits (Gagneux et al. 2001).

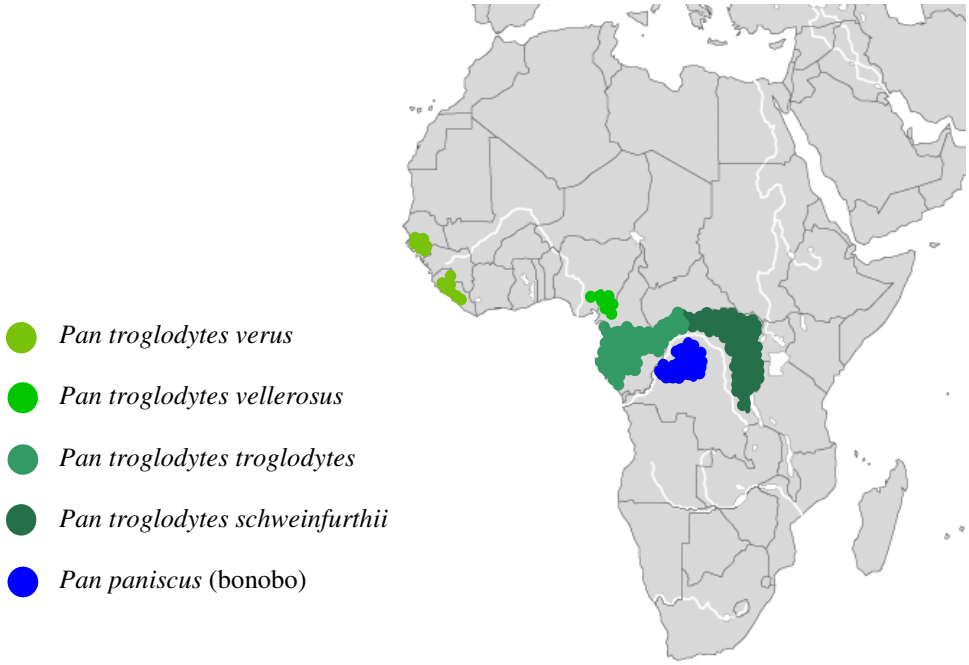

Figure 3: Geographical distribution of the common chimpanzee subspecies and bonobo populations in the wild (Gagneux et al. 2001, 2002).

1.3. Cytogenetic and molecular differences between human and chimpanzee genomes

The development of many genetic techniques, especially in the last three - two decades, brought an impressive amount of publications on the genetic differences between humans and their closest relatives, the chimpanzees. Despite the very high similarity between the DNA sequence of these two species, there are many reports on differences in cytogenetics, in the type and number of repetitive genomic DNA and transposable elements, abundance and distribution of endogenous retroviruses, gene sequence and expression, presence and extend of allelic polymorphism, messenger RNA splicing variations, gene inactivation events, gene duplications, single nucleotide polymorphisms etc. (Gagneux and Varki 2001; Varki and Altheide 2006 and the references therein).

Although most of these genetic dissimilarities are rather modest in size, in combination with many environmental factors they have large effects on the phenotype of the species. The structural variations between human and chimpanzee genomes are divided into large and microscopically visible macro-rearrangements, smaller micro-rearrangements and single nucleotide differences.

1.3.1. Macro-rearrangements

On the level of cytogenetic resolution, the most conspicuous dissimilarities between human and chimpanzee karyotypes are the chromosome number difference (humans have 46, while all great apes have 48 chromosomes), the presence of multiple pericentric inversions and heterochromatic additions (Dutrillaux 1979, Yunis and Prakash 1982, Wienberg et al. 1990).

The human chromosome 2 is the result of the head-to-head telomeric fusion of two ancestral submetacentric chromosomes followed by a subsequent inactivation of one original centromere (IJdo et al. 1991, Fan et al. 2002). The change in the number of chromosomes caused by this fusion resulted in confusions with respect to the chromosome nomenclature. According to ISCN (1978), the chromosomes were numbered from largest to smallest with the use of Arabic numerals (e.g. *HSA* 5 = *PTR* 4) but some authors used to apply a system with Roman numerals in the apes, which corresponds to the human ones (e.g. *HSA* 5 = *PTR* V). Therefore McConkey (2004) proposed a new chromosome numbering system (Table 1) and for simplification I will follow these rules in my thesis.

Table 1: Proposal of the new ape chromosome numbering system with former numbers from ISCN 1978 in brackets (McConkey 2004)

HSA [a]	***PTR*** [b] **and** ***PPA*** [c]	***GGO*** [d]	***PPY*** [e]
1	1 [1]	1 [1]	1 [1]
2pter-q13	2A [12]	2A [12]	2A [12]
2q14-qter	2B [13]	2B [11]	2B [11]
3	3 [2]	3 [2]	3 [2]
4	4 [3]	4 [3]	4 [3]
5	5 [4]	5 [4]	5 [4]
6	6 [5]	6 [5]	6 [5]
7	7 [6]	7 [6]	7 [10]

HSA [a]	***PTR*** [b] **and** ***PPA*** [c]	***GGO*** [d]	***PPY*** [e]
8	8 [7]	8 [7]	8 [6]
9	9 [11]	9 [13]	9 [13]
10	10 [8]	10 [8]	10 [7]
11	11 [9]	11 [9]	11 [8]
12	12 [10]	12 [10]	12 [9]
13	13 [14]	13 [14]	13 [14]
14	14 [15]	14 [18]	14 [15]
15	15 [16]	15 [15]	15 [16]
16	16 [18]	16 [17]	16 [18]
17	17 [19]	17 [19]	17 [19]
18	18 [17]	18 [16]	18 [17]
19	19 [20]	19 [20]	19 [20]
20	20 [21]	20 [21]	20 [21]
21	21 [22]	21 [22]	21 [22]
22	22 [23]	22 [23]	22 [23]

[a] *HSA – Homo sapiens* (human); [b] *PTR – Pan troglodytes* (common chimpanzee); [c] *PPA – Pan paniscus* (bonobo chimpanzee); [d] *GGO – Gorilla gorilla* (gorilla); [e] *PPY – Pongo pygmaeus* (orangutan).

Nine large pericentric inversions distinguish human chromosomes 1, 4, 5, 9, 12, 15, 16, 17, and 18 from their respective chimpanzee orthologues. Comparative analyses using other primate species as out-groups have indicated that the inversions of chromosomes 1 and 18 are human-specific, whereas the inversions of chromosomes 4, 5, 9, 12, 15, 16 and 17 occurred in the chimpanzee lineage (Dutrillaux 1979; Yunis and Prakash 1982; Nickerson and Nelson 1998; Montefalcone et al. 1999; Marzella et al. 2000; Müller and Wienberg 2001; Kehrer-Sawatzki et al. 2002; Locke et al. 2003a; Goidts et al. 2004; Kehrer-Sawatzki et al. 2005a, b, c; Weise et al. 2005; Szamalek et al. 2005; Goidts et al. 2005).

Differences in the amount of constitutive heterochromatin are caused by additions of heterochromatic blocks in the pericentromeric regions of human chromosomes 1, 9 and 16, and in the short arms of acrocentric chromosomes 13 - 15, 21 and 22 in chimpanzee. Due to a heterochromatic addition the Y chromosome of humans and great apes varies notably in size. Further, the telomeres of many chimpanzee and gorilla chromosomes acquired

heterochromatic “caps”, and an additional heterochromatic band was intercalated in the q-arm of *PTR* 13. Conspicuously, many of these additions are absent in orangutan, and those which are present in humans and African great apes are often polymorphic (Yunis and Prakash 1982).

1.3.2. Micro-rearrangements

The extent of variations affecting sequences, which are smaller than ~5 Mb in size has been difficult to resolve strictly with the use of classical cytogenetic methods. These intermediate-sized insertions, deletions, duplications and inversions have become apparent only recently by application of new molecular genetic techniques and have a great potential to affect the structure and genic complement in each species (Snijders et al. 2001; Locke et al. 2003b; Stankiewicz et al. 2004; Cheng et al. 2005; Feuk et al. 2005; Newman et al. 2005; Sharp et al. 2005; Tuzun et al. 2005; Conrad et al. 2006; McCarroll et al. 2005).

1.3.2.1. Micro-insertions and micro-deletions

Insertions and deletions (indels) account for a noticeably large proportion of the difference between the human and the chimpanzee genomes. Aligning five chimpanzee BAC clones with the human orthologous sequences, Britten (2002) detected that 3.4% of human-chimpanzee sequence difference is due to the presence of indels. Frazer et al. (2003) used a larger data set from human chromosome 21 and via comparison with the orthologous chimpanzee sequence identified 57 indels ranging in size from 0.2 to 8.0 kb what corresponds to ~0.6% base pair differences between human and chimpanzee syntenic sequences. The both aforementioned studies were however preliminary with respect to the amount of compared nucleotides: 779 kb and ~27 Mb respectively, and their genomic distributions. Applying the newly developed method of chimpanzee fosmid paired-end sequence mapping against the human reference genome, Newman et al. (2005) detected 393 putative sites of chimpanzee deletions and 184 insertions, which were larger than 12 kb. According to the genome-wide study performed by The Chimpanzee Sequencing and Analysis Consortium (2005) insertion and deletion events result in approximately 3% of sequence dissimilarity between these two species (~1.5% of the euchromatic sequence in each species being lineage-specific; Table 2).

Table 2: Number and size of insertion and deletion events, which distinguish human and chimpanzee sequences (The Chimpanzee Sequencing and Analysis Consortium 2005)

Size of indels	Human-specific	Chimpanzee-specific
Modest-sized indels (1 bp - 5 kb)	~32 Mb ~5 million events	~35 Mb ~5 million events
Large indels (>15 kb)	~8.3 Mb 163 events	Lack of direct measure in chimpanzee, but likely similar as in human

It appears that more than a third of the >80 bp indels are due to repeated sequences (microsatellites and satellites), and about a quarter to transposable elements (The Chimpanzee Sequencing and Analysis Consortium, 2005). The components of the latter, the Alu elements (short transposable sequences of about 300 bp in length) and L1 elements (long transposable sequences), can move to different genomic regions. The considerable difference between the amount of Alus in the human and in the chimpanzee genomes indicates that these elements have been more active in the human lineage (The Chimpanzee Sequencing and Analysis Consortium 2005). Another class of transposable elements consists of endogenous retroviruses (ERVs). The only active human endogenous retrovirus K (HERV-K) was found to be responsible for 73 human-specific and 45 chimpanzee-specific insertions. The chimpanzee genome has still two other active retroviral elements (PtERV1 and PtERV2) which are not present in humans (Yohn et al. 2005; The Chimpanzee Sequencing and Analysis Consortium 2005)

1.3.2.2. Recent segmental duplications

It was shown that segmental duplications (SDs, also called LCRs – low copy repeats), that is duplicated blocks of genomic DNA of over 90% identity between copies and more than 1 kb in length, have an important role in the evolution of the genome's architecture (Armengol et al. 2003; Bailey et al. 2004). They are implicated in the emergence of new genes (Johnson et al. 2001; Courseaux et al. 2003) and differences in gene expression (Cheng et al. 2005), contribute to the large-scale structural polymorphism (Sebat et al. 2004; Iafrate et al. 2004; Tuzun et al. 2005; Sharp et al. 2005) and to the

generation of structural rearrangements leading to specific diseases (Stankiewicz et al. 2002).

In several studies the extent and character of segmental duplications, which contribute to the divergence of the recent primates were investigated. Using an array of BAC clones covering approximately 12% of the human genome, Locke et al. (2003b) detected 63 sites of putative copy number differences (CNDs) between human and the great apes, which ranged in size from 40 kb to 175 kb. Interestingly, these large-scale genomic variations occur in actively transcribed euchromatin rather than in the gene-poor heterochromatic regions. Recently, such array-CGH comparisons were performed in a significantly larger scale by Goidts et al. (2006), who identified a total 322 sites of CNDs between the genomes of human, common chimpanzee, bonobo, gorilla, orangutan and macaque. While in the data set of Locke et al. (2003b) only one copy number difference turned out to be a human-specific gain, Goidts et al. (2006) detected fourteen such sites.

In general, the recent genome-wide comparisons determined that 33% of human duplications are not duplicated in chimpanzee, including some disease-causing LCRs. These CNDs result finally in 2.7% of euchromatic sequence difference between these two species (Cheng et al. 2005). Moreover, interspecies cDNA array-based CGH revealed that gene copy number expansion was more pronounced in humans than in the great apes (Fortna et al. 2004).

1.3.2.3. Micro-inversions

Although modern technologies like array based comparative genomic hybridizations allow the detection of most types of genomic variations, limitations in available methodology have prevented the genome-wide discovery of balanced rearrangements such as inversions (Feuk et al. 2005). Also standard cytogenetic assays are insensitive to submicroscopic changes in sequence orientation (Stefansson et al. 2005, Newman et al. 2005). In two recent studies however, Newman et al. (2005) and Feuk et al. (2005) were able to detect a significant amount of regions of inverted orientation between human and chimpanzee genomes. The first group identified 174 inversion events with the newly developed tool of chimpanzee fosmid paired-end sequence mapping against the human genome, and the second team found a total of 1,576 putative inversion regions analyzing direct sequence alignments. Because the authors verified experimentally only a very small fraction of these rearrangements and several of them corresponded to the regions of known

large pericentric inversions, the exact amount of differences due to micro-inversions between human and chimpanzee genomes remains unknown.

1.3.3. Sequence divergence

The studies performed by Chen and Li (2001) proved that the average sequence divergence, i.e. proportion of nucleotide differences between species, between human and chimpanzee is 1.24% ± 0.07%. The new data confirm this finding and show that the human and chimpanzee genomes differ by only 1.23% in terms of single-nucleotide substitutions, with 1.06% or less corresponding to fixed divergence between the species. The sequence divergence varies among genomic regions, most likely because of regional variations in mutation rate, selective constraints and the rate of recombination between homologous chromosomes during meiosis (The Chimpanzee Sequencing and Analysis Consortium 2005). The highest divergence was found for the Y chromosome. It amounts to 1.9% when the chimpanzee draft sequence of the Y chromosome is compared to the human genomic sequence (The Chimpanzee Sequencing and Analysis Consortium 2005) and to 1.72% - 1.78% in the completely sequenced X-degenerate region (Kuroki et al. 2006 and Hughes et al. 2005, respectively). The lowest divergence was detected for the X chromosome (0.94% for the chimpanzee draft sequence). In comparison to the mouse and rat genomes, the nucleotide substitution rates are much more elevated in the subtelomeric regions of the hominids chromosomes. One-quarter of all observed substitutions occurred at CpG dinucleotides and took place at more similar rates in male and female germ lines than non-CpG substitutions (The Chimpanzee Sequencing and Analysis Consortium 2005).

Taken together, there is an inverse logarithmic relationship between the size and the number of rearrangements when human and chimpanzee are compared (Figure 4).

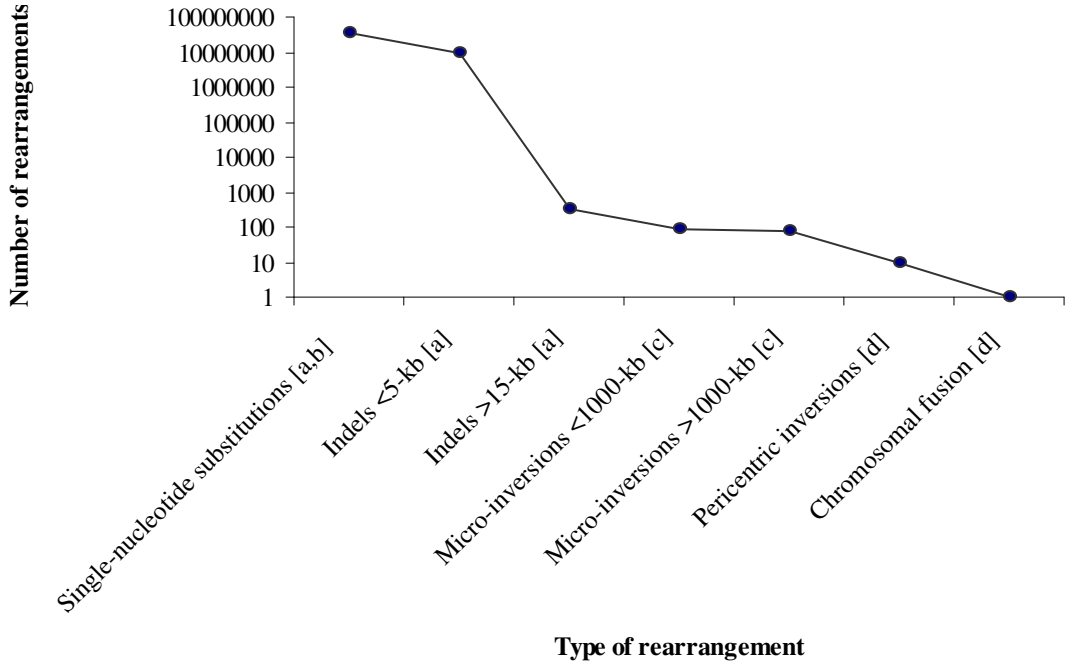

Figure 4: Diagram illustrating the number of rearrangements of different types which are arranged according to their lengths. [a] Chimpanzee Sequencing and Analysis Consortium (2005), [b] Chen and Li (2001); [c] Newman et al. (2005); [d] Yunis and Prakash (1982).

1.4. Cytogenetic and molecular differences between common chimpanzee and bonobo genomes

The karyotypes of the common chimpanzee and the bonobo are very similar but a few differences have been reported. Using chromosome banding techniques, Dutrillaux et al. (1975) identified a bonobo lineage-specific pericentric inversion in *PPA* 2B and an insertion in the long arm of *PPA* 7 chromosome. Further, a large heterochromatin insertion is present in *PTR* 13q, which is absent in human and bonobo and the chromosome 22 is acrocentric in both *HSA* and *PTR*, but metacentric in *PPA* owing to the addition of heterochromatin.

The nucleotide divergence between the bonobo and the common chimpanzee genomes amounts to 0.373%, what is almost three times higher than 0.132%, which corresponds to the substitution rate between common chimpanzee subspecies: *Pan troglodytes troglodytes*, *Pan troglodytes vellerosus*, *Pan troglodytes verus* and *Pan troglodytes schweinfurthii* (Yu et al. 2003). This difference reflects the longer separation time between the bonobo and the common chimpanzee than among the common chimpanzee subspecies.

1.5. Theory of speciation

Speciation is defined as the evolutionary formation of new biological species, usually by the division of a single species into two or more genetically distinct ones. Allopatric speciation occurs when a given population splits because of a natural geographical barrier. In contrast, when different species evolve in the same area by finding different ecological niches scientists use the term sympatric speciation. Another mode of speciation is called parapatric and is used to describe a population at the periphery of a species that adapts to different environments but remains contiguous with its parent (Strickberger 2000). In such situation the gene flow is still possible between these two populations until the point when accumulation of genetic, and as a consequence phenotypic changes will make further hybridization impossible. Consistent with this speciation model two categories of isolating mechanisms (Mayr 1963) play a role: these are premating and postmating mechanisms. The former one would be important in the first phase of the speciation process when the parapatric species can still meet and hybridise, but certain combinations are more favoured. The second mechanism would operate after sufficient separation of the subspecies, which results either in gametic and zygotic mortality or in inviability and sterility of progeny.

Among the extant great apes, the orangutan arose in East Asia and diverged under allopatric conditions with respect to the African great apes. By contrast, the gorillas, chimpanzees and humans arose parapatrically in East Africa (Chen and Li 2001; Hedges and Kumar 2003), therefore it must have existed a period of time when these individuals could hybridize, before the speciation appeared completed.

In the last years a new chromosomal speciation model was proposed, which underlines the role of chromosomal inversions in the formation of new species in the absence of geographical barriers (Rieseberg 2001; Noor et al. 2001; Navarro and Barton 2003a; Rieseberg and Livingstone 2003). According to the traditional chromosomal speciation theory, the chromosomal rearrangements such as inversions might cause low fitness of the heterozygous carriers (hybrids), for example in the form of male sterility or lower hybrid survival, owing to heterozygosity for chromosomal rearrangements, which cause meiotic disturbances (White 1978). This model, however has been criticized, because if a rearrangement causes a substantial reduction in fitness of heterozygotes, it cannot be fixed in a population unless the population is very small (Walsh 1982, Lande 1985, Rieseberg 2001, and references therein). In contrast, the new chromosomal speciation model concentrates on the effect that inversions have on local recombination rates. As the gene flow requires recombination, the reduction in crossing over associated

with chromosomal rearrangements results in a lower gene flow in rearranged chromosomal regions, compared with those which remain collinear (Figure 5). This would facilitate the accumulation of incompatible alleles within rearranged regions, what with the time would lead to the complete reproductive species isolation. Both models, the old one as well as the newly developed one underline the role of postmating isolation mechanisms prior to the absolute species separation.

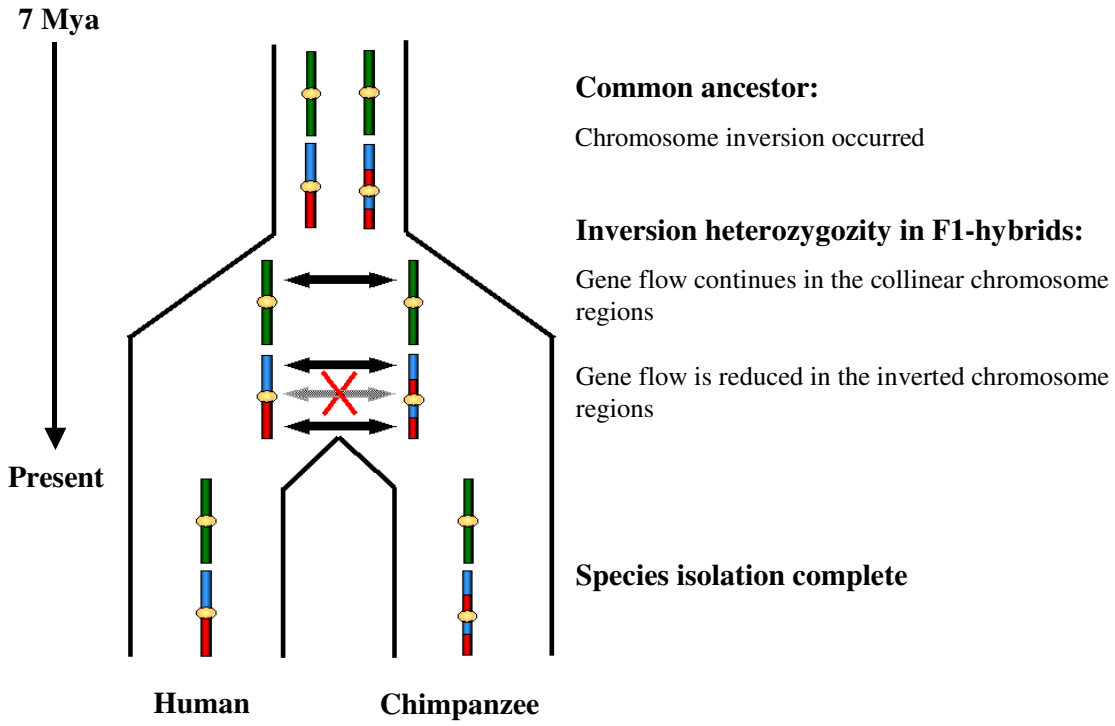

Figure 5: Schematic representation of a new chromosomal speciation model, which relies upon the suppression effect that inversions have on recombination within inverted regions in heterozygotes (F1-hybrids) (Rieseberg 2001; Noor et al. 2001; Navarro and Barton 2003a; Rieseberg and Livingstone 2003).

1.6. Theory of selection

Natural selection operates on certain traits, phenotypic or behavioural ones, i.e. on the gene sequence level. For this reason, nucleotide changes in protein-coding regions are classified into two groups: synonymous changes, which do not cause a change in amino acid, and non-synonymous changes, which change the amino acid. If the coding region is subject to strong selective constraints, then K_A (number of nonsynonymous substitutions per nonsynonymous site) will be lower than K_S (number of synonymous changes per synonymous site), this means that the K_A/K_S ratio will be less than 1. On the other hand, if a gene is subject to positive selection the K_A/K_S ratio is higher than 1.

According to the new model of chromosomal speciation, positive selection should operate more intensively on rearranged than on collinear chromosomal regions. This is because positive selection drives the fixation of genomic incompatibilities inherent during speciation. Several studies have addressed this hypothesis by searching for evidence of accelerated positive selection between humans and chimpanzees in rearranged chromosomes. The obtained results were however often discordant (Navarro and Barton 2003b; Zhang et al. 2004; Vallender and Lahn 2004) thus it must be stated that the influence of the chromosomal rearrangements on the rate of genetic divergence between human and chimpanzees is at best questionable.

Without questioning the importance of chromosomal inversions during human/chimpanzee speciation and focussing on the genome-wide comparison performed by the Chimpanzee Sequencing and Analysis Consortium (2005) we find that the average K_A/K_S ratio of 13,454 human-chimpanzee orthologous gene pairs is 0.23. A total of 585 genes, however, display a higher K_A than the substitution rate in non-coding sequences (K_I). The highest K_A/K_I values were identified for several genes, which encode for proteins involved in immune response and in reproduction. These are genes encoding glycophorin C, granulysin, protamine and semenogelin. In an independent study, Nielsen et al. (2005), by comparison of human-chimpanzee orthologous gene pairs, identified that in addition to genes involved in immune defence and spermatogenesis, also genes implicated in sensory perception, tumour suppression and apoptosis exhibit evidence of positive selection.

There are at least some more speculations, founded in the literature, concerning genes which may have contributed to the human/chimpanzee speciation process. Among them are the *MCPH1* (Microcephalin) and the *ASPM* (abnormal spindle-like microcephaly associated) genes, which are involved in regulation of the brain size. It was shown that the evolution of both these genes was driven by strong positive selection in the lineage leading

to humans (Mekel-Bobrov et al. 2005 and Evans et al. 2005, respectively). However, the availability of the human genome reference sequence and of the first chimpanzee genome draft sequence has even complicated searches for the lineage-specific events, which played a key role in the speciation process. This is because a hitherto unknown class of sequence divergence due to deletions, insertions, duplications and inversions became apparent. These differences raised the total DNA sequence divergence between human and chimpanzee near to 5%.

1.7. Aims of the work presented in this thesis

Besides providing a contribution to the comparative analysis of the human and chimpanzee genomes, the specific aims of my thesis are:

1) Precise molecular characterization of the two not yet analysed pericentric inversions of chromosomes 1 and 5.
2) Comparison of chimpanzee lineage-specific pericentric inversions in the two chimpanzee species: *Pan troglodytes* (common chimpanzee) and *Pan paniscus* (bonobo).
3) Re-examination of the new chromosomal speciation theory taking advantage of the precise breakpoint knowledge.
4) Genome-wide screening for gene-containing micro-inversions.
5) Detailed molecular characterization of these putative micro-inversions.

2. Materials and Methods

2.1. Materials

2.1.1. Chemicals

β-Mercaptoethanol	Roche, Mannheim, Germany
4,6 Diamidino-2-phenylindole: DAPI	Serva, Heidelberg, Germany
Acetic acid, glacial 100%	Merck, Darmstadt, Germany
Agarose	Roth, Karlsruhe, Germany
Agarose, NEEO	Roth, Karlsruhe, Germany
Bacto yeast extract	Gibco BRL, Eggenstein, Germany
Bacto tryptone	Gibco BRL, Eggenstein, Germany
Betaine	Sigma, München, Germany
Biotin-16-dUTP	Roche, Mannheim, Germany
Boric acid	Merck, Darmstadt, Germany
Bromophenol blue	Sigma, München, Germany
BSA	Sigma, München, Germany
Chloramphenicol	Sigma, München, Germany
Colcemid	BiochromAG, Berlin, Germany
Chromosome medium 1A with PHA	Invitrogen, Karlsruhe, Germany
Dextran sulphate	Pharmacia, Freiburg, Germany
Digoxygenin-11-dUTP	Roche, Mannheim, Germany
Dimethyl sulfoxide: DMSO	Serva, Heidelberg, Germany
dNTPs	Roche, Mannheim, Germany
EDTA	Serva, Heidelberg, Germany
Ethanol absolute	Merck, Darmstadt, Germany
Ethidium bromide	Sigma, München, Germany
FCS (fetal calf serum)	Roche, Mannheim, Germany
Fixogum (rubber gum)	Marabuwerke, Tamm, Germany
Formaldehyde	AppliChem, Darmstadt, Germany
Formamide	Fluka, Neu-Ulm, Germany
Gentamycin	Biochrom AG, Berlin, Germany
Glycerin	Sigma, München, Germany
Glycerol	Sigma, München, Germany

HCl	Merck, Darmstadt, Germany
Isopropanol	Merck, Darmstadt, Germany
Kanamycin	Sigma, München, Germany
KCl	Merck, Darmstadt, Germany
Methanol	Roth, Karlsruhe, Germany
$MgCl_2$	Roche, Mannheim, Germany
NaAcetate	Merck, Darmstadt, Germany
Na_3Citrate	Merck, Darmstadt, Germany
NaCl	Merck, Darmstadt, Germany
Pepsin	Sigma, München, Germany
PBS	Biochrom, Berlin, Germany
Phytohemagglutinin, PHA	Invitrogen, Karlsruhe, Germany
RPMI	Gibco BRL, Eggenstein, Germany
RPMI 1640 Medium with GlutaMAX	Invitrogen, Karlsruhe, Germany
SDS	Merck, Darmstadt, Germany
Tris-HCl	Serva, Heidelberg, Germany
Tween 20	Sigma, München, Germany
Vectashield antifade solution	Vector, Burlingame, USA
Yeast extract	Roth, Karlsruhe, Germany
Xylencyanol	Sigma, München, Germany

2.1.2. Enzymes

DNA Polymerase I	Roche, Mannheim, Germany
DNase I	Sigma, München, Germany
Proteinase K	Sigma, München, Germany
RNase A	Sigma, München, Germany
Taq DNA Polymerase	Qiagen, Hilden, Germany

2.1.3. Antibodies

Anti-digoxigenin	Roche, Mannheim, Germany
Anti-mouse Texas-Red	Dianova, Hamburg, Germany
Anti-avidin biotinylated	Vector, Burlingame, USA
Anti-rabbit Texas-Red	Dianova, Hamburg, Germany

FITC-avidin	Vector, Burlingame, USA

2.1.4. DNA probes

Cot-1 DNA	Invitrogen, Karlsruhe, Germany
1 kb Plus DNA Ladder	Invitrogen, Karlsruhe, Germany

2.1.5. Solutions, Buffers, Culture medium

0.1 M β-Mercaptoethanol

- 34.7 µl β-Mercaptoethanol 99% (14.4 M)
- add ddH_2O to final volume of 5 ml, dispense into aliquots and store at –20°C

Agar medium

- 10 g Bacto tryptone
- 5 g Bacto yeast extract
- 10 g NaCl
- 20 g Agarose
- adjust pH to 7.0 with 5 N NaOH
- add ddH_2O to final volume of 1 l, sterilize by autoclaving, pour into Petri dishes (~25 ml/100 mm plate) and store at 4°C for several months

Chloramphenicol stock solution (34 mg/ml)

- 680 mg Chloramphenicol
- add absolute ethanol to final volume of 20 ml, sterilize by filtration, store at –20°C

Chloramphenicol stock solution for deep freezing (for storage cultures) (34 mg/ml)

- 680 mg Chloramphenicol
- add glycerol to final volume of 20 ml, sterilize by filtration, store at –20°C

Culture medium for lymphoblasts

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

4,6 Diamidino-2-phenylindole: DAPI

- 90 ml 4x SSC
- 90 µl DAPI

dNTPs for standard PCR

- mix 12.5 µl of 100 mM stock solution of each dNTP
- add ddH_2O to final volume of 1 ml and store at –20°C

dNTPs for expand PCR

- mix 100 µl of 100 mM stock solution of each dNTP
- add ddH_2O to final volume of 500 ml and store at –20°C

dNTPs for Nick-translation

- mix 5 µl of 100 mM stock solution of ATP, CTP and GTP, and 1 µl of TTP
- add ddH_2O to final volume of 1 ml and store at –20°C

0.5 M EDTA (pH 8.0)

- 18.6 g Na_2-EDTA
- 50 ml ddH_2O
- adjust pH to 7.6 – 8.0 with 5 N NaOH
- add ddH_2O to final volume of 100 ml, sterilize by autoclaving

1% Formaldehyde solution

- 68 ml PBS+$MgCl_2$
- 2 ml Formaldehyde 35%

Kanamycin stock solution (25 mg/ml)

- 500 mg Kanamycin
- add ddH_2O to final volume of 20 ml, sterilize by filtration, store at –20°C

Kanamycin stock solution for deep freezing (for storage cultures) (25 mg/ml)

- 500 mg Kanamycin
- add glycerol to final volume of 20 ml, sterilize by filtration, store at –20°C

Ladder

- 100 µl ddH_2O
- 100 µl TE buffer
- 25 µl Loading buffer
- 10 µl 1 kb ladder

LB medium

- 10 g Tryptone
- 5 g Yeast extract
- 10 g NaCl
- adjust pH to 7.0 with 5 N NaOH
- add ddH_2O to final volume of 1 l, sterilize by autoclaving and store at 4°C for several months

Lysis buffer

- 0.0188 g Na_2EDTA
- 0.5 g $KHCO_3$
- 4.15 g NH_4Cl
- adjust pH to pH 7.4

Loading buffer

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

MM (Master Mix)

- 400 µl 50% Dextran sulphate
- 200µl 20% SSC (pH 7.0)
- 400 µl ddH_2O

NaAcetate 3M

- 12.3 g NaAcetate
- add ddH_2O to final volume of 50 ml, adjust pH to 5.2 with acetic acid

10x NT buffer

- 5 ml 1 M Tris-HCl (pH 7.5)
- 0.5 ml 1 M $MgCl_2$
- 0.5 ml BSA 10 mg/ml
- add ddH_2O to final volume of 10 ml, dispense into aliquots and store at –20°C

Pepsin stock solution (100mg/ml)

- dissolve 1 g lyophilised powder (3,200-4,500 units/mg protein) in pure ddH_2O at 37ºC
- dispense into aliquots of 30 µl and store at –20°C

Pepsin solution for FISH

- add 30µl of pepsin stock solution to 99 ml ddH_2O
- adjust pH with 1 ml of 1M HCl

10x PBS

- 95 g PBS in 1 l ddH_2O

1x PBS with 50 mM $MgCl_2$

- 100 ml 10x PBS
- 100 ml $MgCl_2$ (0.5M)
- 800 ml ddH_2O

P20 buffer

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

Pepsin solution

- 99 ml ddH_2O
- 30 µl pepsin
- 1 ml HCl

Pre-incubation buffer (FISH)

- 3g BSA in 80 ml of 4x SSC + 0.1 % Tween 20
- dissolve by heating and mixing, filter and incubate in 37°C

RNase solution

- 1µl RNase A (10 mg/ml)
- add 99 µl 2x SSC

SE buffer

- 2.19 g NaCl
- 4.65 g Na_2EDTA
- add ddH_2O to final volume of 0.5 l, adjust pH with NaOH to 8.0, sterilize by autoclaving

10% SDS

- 100 g SDS
- add ddH_2O to final volume of 1 l, adjust pH to 7.2 with HCl

20x SSC

- 175.3 g NaCl (3M)
- 88.2 g Na_3Citrate (0.3 M)
- add ddH_2O to final volume of 1 l, adjust pH to 7.0 with HCl

SQ buffer

- 6.76 g Betaine
- add 4 ml ddH_2O
- 500 µl Tris saturated (pH 9.0)
- add ddH_2O to final volume of 30 ml

50x TAE

- 242 g Trsi base
- 57.1 ml Acetic acid
- 100 ml 0.5 M EDTA
- add ddH_2O to final volume of 1 l

TE

- 10 ml Tris-HCl (1 M)
- 2 ml EDTA (0.5 M)
- add ddH_2O to final volume of 1 l, adjust pH to 8.3, sterilize by autoclaving

10x TBE buffer

- 540 g Tris base
- 275 g Boric acid
- 200 ml 0.5 M EDTA
- add ddH_2O to final volume of 5 l, adjust pH to 8.2

Tris-HCl (1M)

- 121.1 g Tris base
- add ddH_2O to volume of 850 ml and adjust pH with HCl to 7.5
- add ddH_2O to final volume of 1 l, sterilize by autoclaving

Washing buffer (FISH) - 4x SSC + 0.1% Tween 20

- 400 ml 20x SSC
- add ddH_2O to final volume of 2 l, adjust pH with HCl to 7.0
- add 2 ml Tween 20

2.1.6. Kits

ABI PRISM Big Dye Terminator Cycle Sequencing Ready Reaction Kit	Applied Biosystems, Foster City, USA
DNeasy Tissue Kit	Qiagen, Hilden, Germany
Expand Long Template PCR System	Roche, Mannheim, Germany
GFX PCR DNA and Gel Band Purification Kit	Amersham Biosciences, Buckinghamshire, UK
HiSpeed Plasmid Midi Kit	Qiagen, Hilden, Germany

2.1.7. Equipment and Software

ABI PRISM 3100 Genetic Analyzer	Applied Biosystems, Foster City, USA
ABI PRISM 3100 Genetic Analyzer Data Collection Software v1.1	Applied Biosystems, Foster City, USA
AxioCam MRm camera	Zeiss, Göttingen, Germany
AxioPlan microscope	Zeiss, Göttingen, Germany
CO_2 incubator BB6220 CU	Heraeus, Hanau, Germany
Electrophorese Power PAC 300	Bio Rad, München, Germany
Eppendorf Mastercycler	Eppendorf North America, Westbury USA
GeneQuant, RNA/DNA Calculator	Pharmacia, Cambridge, UK
ISIS FISH Imaging System v5.0	Metasystems, Altlussheim, Germany
PTC-100 Programmable Thermal Controller	MJ Research Inc, Minnesota, USA

2.1.8. Cell lines

The Epstein-Barr virus (EBV) - transformed lymphoblastoid cell lines of four great ape species were used for the studies. The common chimpanzee cell line *PTR*-L2008 was a generous gift of Prof. Dr. W. Schempp from Freiburg, Germany, while the two bonobo (*Pan paniscus*, *PPA*) cell lines were obtained from Dr. U. Zechner from Mainz, Germany. Further cell lines of *Pan troglodytes* (*PTR*, common chimpanzee), *Gorilla gorilla* (*GGO*, gorilla) and *Pongo pygmaeus* (*PPY*, orangutan) were purchased from the European Collection of Cell Cultures (www.ecacc.org.uk), (Table 3).

Table 3: The origins and designations of the great ape cell lines

Species	Gender	Name/designation (origin)
Pan troglodytes	female	*PTR*-L2008 (Stuttgart ZOO)
Pan troglodytes	male	EB176 JC (ECACC No. 89072704)
Pan paniscus	female	*PPA*-Cheka (Frankfurt ZOO)
Pan paniscus	female	*PPA*-Binti (Frankfurt ZOO)
Gorilla gorilla	female	EB JC (ECACC No. 89072703)
Pongo pygmaeus	male	EB185 JC (ECACC No. 89072705)

2.1.9. Blood and DNA samples

Human (*Homo sapiens*) chromosomal slides for FISH analyses and DNA samples were prepared from the blood samples of normal human donors. Further, the DNA samples isolated from peripheral blood of nine additional common chimpanzees (*Pan troglodytes*) were obtained from Dr. W. Schempp from Freiburg, Germany (Table 4).

Table 4: The origins of the great ape blood samples

Species	Gender	Name/designation (origin)
Pan troglodytes	female	Chita II (Stuttgard ZOO)
Pan troglodytes	female	Mixi (Neunkirchen)
Pan troglodytes	male	Hans (TNO Holland)
Pan troglodytes	female	Jana (TNO Holland)
Pan troglodytes	female	Corry (TNO Holland)
Pan troglodytes	female	Luise (Welzheim, Schwabenpark)

Species	Gender	Name/designation (origin)
Pan troglodytes	male	Zeef (TNO Holland)
Pan troglodytes	male	Walter (TNO Holland)
Pan troglodytes	female	Chita (Welzheim, Schwabenpark)

2.1.10. FISH probes

BAC/PAC clones derived from *Homo sapiens* genomic library (Tables 5, 6, 7, 9) were selected according to their position in the human genome reference sequence (NCBI Build 35) annotated by the Ensembl Genome Browser (http://www.ensembl.org/). Chimpanzee (*Pan troglodytes*) BACs (Table 8) were chosen according to their homology with respective human BAC/PACs. All BAC/PAC clones were purchased from the BACPAC Resources Center (http://www.chori.org/bacpac).

Table 5: Human BAC/PAC clones used for characterization of the chromosome 1 pericentric inversion

Chromosome	BAC/PAC	Accession number
1p12	RP11-192J8	BZ749068; BZ749069
1p12	RP4-712E4	AL139420
1p12	RP4-683H9	AL589734
1p12	RP5-1042I8	AL359752
1p12-11.2	RP11-114O18	AL596222
1p11.2	RP11-439A17	AL357493
1p11.2	RP11-343N15	AL358175
1p11.2-11.1	RP11-435B6	BX248407
1q21.1	RP11-30I17	B87817; B87818
1q21.1	RP11-337C18	AL356378
1q21.1	RP11-314N2	AL445591
1q21.1	RP3-328E19	AL022240
1q21.1	RP11-763B22	AL592492
1q21.1	RP11-14N7	AL513526
1q21.1	RP11-744H18	AL732363
1q21.1	RP11-403I13	AL356957
1q21.1	RP5-998N21	AL109948
1q21.1	RP11-277L2	AL451058
1q21.1-2	RP11-353N4	AL358813
1q21.2	RP11-196G18	AL591493
1q21.2	RP11-21J4	AC018785
1q21.2	RP11-458I7	AL358073
1q21.2	RP11-363I22	AL356292

Table 6: Human BAC/PAC clones used for characterization of the chromosome 5 pericentric inversion

Chromosome	BAC/PAC	Accession number
5p15.1	RP1-137K24	AC016648
5p15.1	RP11-19O2	AC092335
5p15.1	RP11-55N10	AC113415
5p15.1	RP11-211K15	AC091911
5p15.1	RP11–35A11	AC113389
5p14.3	RP11-270H9	AC106768
5p14.3	RP11-60F22	AC118463
5p14.3	RP11-414I19	AC109474
5p14.3	RP11-774D14	AC140168
5q13.2	RP11-118I21	AC116340
5q13.2	RP11-60A8	AC093302
5q13.3	RP11-229C3	AC093259
5q13.3	RP11-524L6	AC108120
5q13.3	RP11-206N2	AC114962
5q14.1	RP11-107N7	AC108482
5q14.1	RP11-21K15	AC114963
5q15	RP11-432G16	AC104125

Table 7: Human BAC/PAC breakpoint-spanning clones used for the pericentric inversion breakpoint comparison between *Pan troglodytes* and *Pan paniscus*

Chromosome	BAC/PAC	Accession number
4p13	RP11-779N22	AC110811
4q21.23	RP11-8N8	AC108021
5p14.3-p15.1	RP11-35A11	AC113389
5q15	RP11-432G16	AC104125
9p12	RP11-259A5	BX664724
9q21.33	RP11-507D14	AL137849
12p12.2	RP11-80N2	AC011604
12q15	RP3-491B7	AL606524
16p11.2	CTD-2144E22	AC135776
16q11.2	RP11-696P19	AC106819
17p13.1	RP1-179H24	AF503578 [a]
17q21.33	RP5-1029K10	AC006487

[a] The accession number refers to the sequenced breakpoint-spanning fragment of PAC RP1-179H24.

Table 8: Common chimpanzee BAC/PAC clones used for the pericentric inversion breakpoint comparison between *Pan troglodytes* and *Pan paniscus*

Chromosome	BAC/PAC	Accession number
4p	RP43-59P20	AY335550
4q	RP43-41D24	AY335551
5p	RP43-001C16	AG141159; AG141160
5q	RP43-023F12	AG158191; AG158192
9p	RP43-163C1	AY569337
9q	RP43-156H15	CW691359; CW691360
12q	RP43-77C18	AC006582
12q	RP43-135M19	AC007214
16p	RP43-007E19	AG145869; AG145870
16q	RP43-001I03	AG141421; AG141422
17p	RP43-141Q35	AF503579
17q	RP43-134L13	AY117035

Table 9: Human BAC/PAC clones used for the experimental verification of the micro-inversions

Chromosome	BAC/PAC	Accession number
1p13.2-13.3 [a]	RP4-773N10 (inversion probe)	AL160006
	RP4-671G15 (inversion probe)	AL354760
	RP5-1037B23 (reference probe)	AL162594
5p13.2-13.3 [a, b, c]	RP11-784O19 (reference probe)	AQ520286; AQ661755
	RP11-779D3 (inversion probe)	AQ467685; AQ467607
	RP11-727F19 (inversion probe)	AC109491
	RP1-137K24 (reference probe)	AC016648
	RP11-43F23 (inversion probe)	AC025189
	RP11-727F19 (inversion probe)	AC109491
	RP11-94N8 (spanning putative breakpoint)	AQ311892; AQ311893
7p22.1 [a]	RP11-1275H24 (reference probe)	AC092171
	RP4-810E6 (inversion probe)	AC004895
	RP11-425P5 (inversion probe)	AC009412
7p13-14.1 [a]	RP4-806A17 (reference probe)	AC005483
	RP4-570D2 (inversion probe)	AC004837
	RP5-1032B10 (inversion probe)	AC004455
	RP11-630O5 (reference probe)	AC073345
	RP4-570D2 (inversion probe)	AC004837
	RP11-36H20 (inversion probe)	AC011738
17q21.31-24.1 [b]	RP11-613C6 (inversion probe)	AC015795
	RP11-188B1 (inversion probe)	AC129531
	RP11-28G8 (reference probe)	AC016245
18p11.21-11.22 [a]	RP11-243E13 (reference probe)	AP006219
	RP11-21G15 (inversion probe)	AP005117
	RP11-78A19 (inversion probe)	AP005137

Chromosome	BAC/PAC	Accession number
19q13.12 [a]	RP11-45P2 (inversion probe)	AC009756
	RP11-44B13 (inversion probe)	AC009194
	RP11-9B17 (reference probe)	AC007676

[a] interphase FISH; [b] metaphase FISH; [c] to verify the presence of the inverted transposition on *PTR* 5 (*HSA* 5p13.2-13.3) two probe combinations were applied, the first for interphase FISH and the second for metaphase FISH, where only BAC RP11-727F19 is located within the inverted region.

2.1.11. Primers used for PCR analyses

PCR primers were designed either according to the reference sequence of the human genome (NCBI Build 35) and chimpanzee draft sequence (The Chimpanzee Sequencing and Analysis Consortium, 2005) or according to the sequence of the BAC/PAC clones. Care was taken that the primers were located in non-repetitive sequences (Tabeles 10 and 11).

Table 10: Primers used for the junction-PCR analyses of the breakpoint-spanning fragments

PTR chromosome [a]	Primer sequence 5' to 3'	Designed on the sequence of	Primer position	Product size (bp)
4p (3p)	TCTTTTGTCTTTGTCTGTCTGGA	RP43-59P20 (AY335550)	F: 2,108	1,393
	CCAGAAATAGGGACAGGTGAA		R: 3,501	
5p (4p)	AAACAAATATCCACCCACAACC	Scaffold *PTR* Chr:4	F: 19,219,378	993
	CATTCCCACTCACAATGCTC		R: 19,220,371	
9p (11q)	CAAATACATTTTTGTCACTATTTGTCA	RP43-163C1 (AY569337)	F: 145,549	1,921
	TGCTCCAAATGTCTCCCAGT		R: 147,468	
12q (10q)	CCTTGACTGGCTCTTCCACT	RP43-135M19 (AC007214)	F: 147,897	226
	GGACACTGGATATCTCACATGG		R: 148,123	
16q (18q)	TGGGTGAGAGTTTTCCAAGC	Scaffold *PTR* Un_random	F: 8,823,489	197
	CTACTGGCACTGGACTTCAGC		R: 8,823,686	
17q (19q)	ACCCTTTAGATCAGAAGAATCCTG	RP43-134L13 (AY117035)	F: 7,895	340
	AAAAACCTGATCCCGTAGGC		R: 8,235	

[a] According to the Ensembl version: 27.1a.1 based on the draft sequence of the chimpanzee genome, in parenthesis the former chimpanzee chromosome numbers from ISCN 1978 are given.

Table 11: Primers used for the PCR investigations of the micro-inversions in humans and chimpanzees

Chromosomal region [a]	Primer sequence 5' to 3'	Designed on the sequence of	Primer position	Product size (bp)
***HSA* 5p13.2-13.3**	GAGCCTGAAAGCATCTTTGG CAAGCAAGCGTCCCTAAAAG	CTD-2241P19 (AC010449)	F: 36,855 R: 39,138	2,283
***HSA* 7p13-14.1**	TCATTATGCCCACCTCAGAAG CTTCCTGCCTTAGTTTGCTGA	RP11-64I2 (AC011290)	F: 65,558 R: 60,631	4,927
	GATGCCACTCCTTTCTGCTC CTGCCCTTGACCTGATTCTC	RP5-1165K10 (AC004985)	F: 113,396 R: 108,337	5,059
***HSA* 19p13.11**	GAGAAGGCTCGCTGCTTAGA CCTTTCTGAGGCCTCTTCCT	CTD-2528A14 (AC020908)	F: 102,030 R: 106,816	4,786
***PTR* 19**	CACACTCTGGCCTGTGACAT CAGAGAGCGGTTTCTTCAGC	*PTR* 20 (Scaffold Chr_20)	F: 18,240,003 R: 18,240,713	710

[a] According to the Ensembl version 28.35a.1 based on the NCBI Build 35 assembly and Ensembl version 28.1a.1 based on the draft sequence of the chimpanzee genome; *PTR* 19 (20) – number in in parenthesis indicates the former number of *PTR* 19 according to ISCN 1978.

2.2. Methods

2.2.1. Lymphocyte culture

Peripheral blood samples were obtained from normal human donors and dispensed into 10 ml sterile vials containing sodium heparin. For each person several Falcon tubes containing 10 ml of Chromosome Medium 1A with phytohemagglutinin (PHA) supplemented with additional PHA (1 ml of PHA per 500 ml of Chromosome Medium 1A) were inoculated with 1 ml of blood. The samples were incubated at 37°C in a 5% CO_2 atmosphere for 72 h. During the last 30 min of incubation, 125 µl of colcemid was added and the lymphocyte cultures were centrifuged at 1,000 rpm for 10 min at RT. After removing of supernatant, the hypotonic treatment was performed by resuspending of the pellets in 10 ml of 0.4% KCl, which was pre-warmed to 37° C, and by subsequent incubation for 30 min at 37°C. The tubes were centrifuged again at 1,000 rpm for 10 min at RT and the supernatant was removed. In the next step a freshly prepared ice-cold fixative (methanol: glacial acetic acid, 3:1) was added by slow dropping. During fixation the cells were carefully resuspended by pipetting up and down. Afterwards, the fixation step was repeated by rounds of centrifugation, removal of supernatant, resuspension of pellets and addition of new fixative. The cell suspensions in fixative were then incubated at 4°C overnight, allowing further lysis of cellular components. On the next day the last round of fixation was performed and then the cells were resuspended in a small volume of fixative and dropped onto slides.

2.2.2. Lymphoblastoid cell culture

Each 500 ml bottle with RPMI 1640 Medium with GlutaMAX was supplied with 15% FCS (fetal calf serum) and 2.5 ml Gentamycin. The lymphoblastoid cell lines were cultivated in T25, T75 or T80 flasks and every second day supplied with fresh RPMI 1640 Medium described above. For chromosome preparations 10 ml of each cell culture were transferred into separate Falcon tubes and 100 µl of colcemid were added.

2.2.3. Slide preparation

Microscope slides of 26 x 76 mm were cleaned by 2-days incubation in 100% ethanol and subsequent rinsing for several minutes with running deionised H_2O. After that, the slides were stored in ddH_2O at 4°C. Using a Pasteur pipette a few drops of the fixed

cell suspensions were dropped from about 10 cm of distance onto cleaned, cold and wet slides. Slides were left for drying at RT and then the concentration of mitotic spreads and interphase nuclei was verified under a phase contrast microscope. On the next day the dried slides were frozen in boxes containing silica gel and stored until use.

2.2.4. Growth of bacterial cultures

Agar medium was prepared with autoclaved LB medium. After cooling to 60°C, the medium was supplemented with kanamycin or chloramphenicol and about 30 ml were poured in each Petri dish, and allowed to gelify at RT. Petri dishes were than stored at 4°C until use. Bacteria containing BAC or PAC vectors were streaked on the agar plate with a loop under sterile conditions and grown up overnight at 37°C. Clones derived from libraries constructed in PAC vectors (pCYPAC2 or pPAC4) grow on medium supplied with kanamycin. These libraries are RPCI-1, 3, 4, 5, 6, 21, 31, 91 and 92. Clones derived from all other libraries are constructed in BAC vectors (pBACe3.6 or pTARBAC2) and grow on medium supplied with chloramphenicol.

On the next day single colonies were picked with a sterile toothpick and inoculated in Falcon tubes containing 5 ml of LB medium supplemented with the corresponding antibiotic. These pre-cultures were shaken at 250 rpm for several hours at 37°C until they appeared turbid, indicating that the bacteria had grown to saturation. Each pre-culture was transferred into a 500 ml bottle containing 180 ml of LB medium with addition of the antibiotic, and shaken overnight at 37°C.

2.2.5. Qiagen BACs/PACs midi-preparation

After incubation, the culture volumes were transferred into plastic bottles and centrifuged at 5,000 rpm for 15 min. After centrifugation the medium was discarded and the BAC/PAC preparations were carried out using the HiSpeed Plasmid Midi Kit (Qiagen) according to the protocol of the manufacturer. In the last step the BAC/PAC DNA was eluted with 500 µl of 10 mM Tris-HCl (pH 8.3) and stored at 4°C. The DNA yield from a ~180 ml culture amounted from 20 to 200 ng/µl.

2.2.6. Probe labelling for FISH

2.2.6.1. Nick-translation

The probes were labelled using the Nick-translation technique. The protocol is based upon the activities of the DNase I and of the E.coli DNA polymerase I. The DNase I introduces nicks into the DNA molecule and the exonuclease activity of polymerase I acts at the sites of nicks and removes nucleotides in the 5´→3´ direction. At the same time, using the 3´-hydroxyl group of the terminal nucleotide of the nick as a primer, the polymerase activity of polymerase I replaces the pre-existing unlabelled nucleotides with labelled nucleotides. In a standard reaction, 1 µg of template DNA was labelled with biotin-16-dUTP or digoxigenin-11-dUTP in 100 µl of final volume.

A standard Nick-translation reaction was prepared as follows:

ddH_2O	x µl
10 x NT buffer	10 µl
0.1 M β-Mercaptoethanol	10 µl
dNTPs	10 µl
Biotin-16-dUTP/	
Digoxigenin-11-dUTP	4 µl
Probe DNA	x µl
Polymerase I (5 U/µl)	2 µl
DNase I (10 U/µl; 1:1000)	2 µl
	100 µl

This mixture was incubated for 45 - 60 min in a 15°C water bath and after that time the efficiency of DNase I was checked on the agarose gel. If the size of the DNA probe ranged between 1,000 and 100 nucleotides the labelling reaction was stopped by adding of 1 µl of 0.5 M EDTA followed by heating to 65°C for 10 min.
Advantages of this technique are an uniform labelling through the DNA molecule and the possibility to adjust the fragment size by varying the amount of DNase I. The correct size of the probe after the labelling reaction is in fact a requirement for successful FISH.

2.2.6.2. Purification of the labelled DNA

After the Nick-translation and subsequent enzyme inactivation, 40 µl of Cot-1 DNA (40 µg) was added to each probe. The Cot-1 DNA is a DNA probe highly enriched for repetitive sequences and is used for background suppression, which can be caused by interspersed repetitive elements present within large insert clones such as BACs and PACs. To the total volume of the Nick-translation reaction mixed with Cot-1 DNA, 10% of 3 M NaAcetate (pH 5.2) was added. The NaAcetate salt molecules help the following precipitation of the probe DNA. Subsequently 2.5 reaction volume of absolute cold ethanol were added and after mixing the probe was placed in –20°C overnight or in –80°C for 30 min. After that time probes were centrifuged at 14,000 rpm for 30 min at 4°C and the supernatant was carefully discarded. 200 µl of cold 70% ethanol were added in order to wash out rests of salt, and centrifugation was repeated under the same conditions for 10 min. The supernatant was again discarded and probe DNA was dried at 37°C on the thermo block for 5 min. Consequently the probes were dissolved in 25 µl of 100% deionised formamide and incubated at 42°C for two hours. During that time probe DNA was mixed thoroughly in order to help dissolving, and centrifuged. Finally 25 µl of Master Mix was added to each probe following by mixing, and the labelled probes were stored at 4°C until use.

2.2.7. Fluorescence *in situ* hybridisation (FISH)

2.2.7.1. Slides pre-treatment and hybridisation with labelled probes

Prior to denaturation the slides were treated with 150 µl of RNase A (100 µg/ml in 2x SSC) under a coverslip for 1 h at 37°C. Thereafter, they were washed 3 x 5 min in 2x SSC and incubated in pepsin solution (0.005% enzyme in 0.01 M HCl) for 5 - 8 min at 37°C. Immediately after that, the slides were washed in 1x PBS, once shortly and 2 times for 5 min. Next washing was performed in 1x PBS with 50 mM $MgCl_2$ for another 5 min. Subsequently slides were fixed for 10 - 15 min in 1x PBS with 50 mM $MgCl_2$ and 1% of formaldehyde, washed in 1x PBS and dehydrated in a series of 70%, 80%, 90% and absolute ethanol, 3 min each at RT.

Each single probe, or a combination of two to three differentially labelled probes were denaturated at 76°C for 6 - 10 min, depending on the probe amount, and reannealed for 40 min at 37°C. During this procedure the repetitive sequences contained in the

genomic insert of the FISH probe reanneal with Cot-1 DNA. After reannealing the probe DNA was applied onto slides under coverslips. For the middle-size coverslips (24 x 24 mm) I used 8-12 µl of each probe, what corresponds to ~150 - 230 ng of DNA probe. For larger size coverslips (50 x 24 mm) I used 20-24 µl of each probe, that is around 380 - 460 ng DNA. If a combination of two or three differentially labelled probes was used, I tried to apply smaller amounts of each probe than for single hybridizations. After that, the coverslips were sealed with rubber gum and the hybridization was carried out for two up to three nights at 37°C in a moist chamber.

2.2.7.2. Signal detection

After hybridization, the coverslips were removed in the first wash in 2x SSC and next the slides were washed two times in 50% formamide dilution with 2x SSC at 45°C, 15 min each. Then subsequent washing steps were performed, three times in 2x SSC at 45°C, and 2 times in 0.2x SSC at 60°C. After washing the slides were shortly immersed in 4x SSC with 0.1% of Tween20 to equilibrate them to a lower stringency. Immediately after that, the slides were pre-incubated in 5% BSA solution in 4x SSC and 0.1% of Tween20 for 20 - 30 min at 37°C. Next the detection of the hybridization signals was performed. In case of probes labelled with biotin-16-dUTP the signals were detected with FITC-avidin, biotinylated anti-avidin and FITC-avidin in a first, second and third step, respectively. Probes labelled with digoxygenin-11-dUTP were detected by mouse anti-digoxygenin in a first step. In a second round, anti-mouse antibodies coupled with Texas-Red and produced in rabbit were used, followed by anti-rabbit antibodies conjugated with Texas-Red in the third round of detection. The concentrated antibodies were centrifuged for 5 min at 13,000 rpm before their application. Directly before use, the antibodies were diluted with an appropriate amount of 5% BSA in 4x SSC + 0.1% Tween 20 as follows:

FITC-Avidin (1:200)
Anti-Avidin (1:100)
Mouse anti-digoxygenin (1:500)
Anti-mouse antibody from rabbit coupled with Texas-Red (1:25)
Anti-rabbit antibody from goat coupled with Texas-Red (1:50)

In each round of detection, 150 µl of diluted antibodies were applied onto slides under coverslips, following by 45 min incubation in a moist chamber at 37°C. After each

incubation the slides were washed 3 x 5 min in 4x SSC + 0.1% Tween20 at 45°C in the darkness to prevent a possible decrease of fluorescence. After last washing the slides were collected in 4x SSC, counterstained with diamidinophenylindole (DAPI), rinsed with ddH_2O, and left for drying in the darkness. When the slides were dry, I mounted them with Vectashield antifade solution.

2.2.7.3. Microscopic evaluation

After FISH the slides were evaluated under the Axioplan microscope (Zeiss, Germany) with the use of specific filters corresponding to the used fluorochromes (DAPI, FITC, Texas-Red). Image capture and processing were performed with a CCD camera AxioCam MRm (Zeiss, Germany) driven by the computer aided software ISIS (FISH Imaging System v5.0 from MetaSystems GmbH, Germany).

2.2.7.4. Breakpoint-spanning BAC/PACs

The principle of the search for breakpoint-spanning BAC/PAC clone is that the probe represents a continuous DNA sequence which corresponds to a single FISH signal on a chromosome of a species from which the clone is derived. This DNA sequence is however broken in species or individuals which carry an inversion with a breakpoint contained within this given clone, what is observed as a split signal on a chromosome. This rule applies only to single copy clones, i.e. those which do not contain segmental duplications.

2.2.8. DNA isolation

2.2.8.1. DNA isolation from cultured cell lines

In order to obtain the optimal yield and quality of DNA, the amount of starting material for the DNA extraction was 4 - 5 x 10^6 cultured cells. DNA was isolated with the use of the DNeasy Tissue Kit (Qiagen), according to the DNeasy protocol for cultured animal cells. Briefly, after the cell lysis with 20 µl of proteinase K (20 mg/ml) and treatment with 4 µl of RNase A (100 mg/ml), the DNA was purified on DNeasy mini columns containing silica-gel-membranes. Eluted DNA was ready for use in subsequent analyses.

2.2.8.2. DNA isolation from blood

For DNA isolation from whole peripheral blood the NaCl-method was used. 10 ml of blood collected into sterile vials containing EDTA were incubated on ice with 30 ml of lysis buffer for 30 min and centrifuged at 1,000 rpm for 10 min at 4°C. Obtained cells pellets were then resuspended in lysis buffer and incubated for 10 min on ice. After a second centrifugation step pellets were dissolved in 5 ml of SE buffer supplied with 500 µl of 10% SDS and 30 µl of proteinase K (10 mg/ml). Next the samples were incubated in a 55°C water bath and 1.5 ml of 6 M NaCl was added. After vortexing and centrifugation at 4,500 rpm for 15 min at 20°C the supernatant was transferred into clean Falcon tubes and precipitated with 2 volumes of absolute ethanol at RT. The precipitated DNA was washed with 70% ethanol, dried and dissolved in TE buffer.

2.2.9. Determination of the DNA concentration

To determine the concentration and purity of DNA, samples were diluted 20 times with ddH_2O and the absorbency of these dilutions was measured at the wave lengths of 260 nm and 280 nm in the spectrophotometer. Since a solution with an OD_{260} of 1 contains approximately 50 µg DNA per millilitre, the DNA concentration in a sample was as follows: [DNA concentration] = OD_{260} x dilution factor x 50 µg/ml.
The ratio between the reading at 260 nm and 280 nm (OD_{260}/OD_{280}) provides an estimate of the DNA purity. DNA isolates which are protein and RNA free have OD_{260}/OD_{280} values of 1.8.

2.2.10. Polymerase Chain Reaction (PCR)

Polymerase chain reaction (PCR) is a basic molecular biology technique which allows for fast enzymatic replication of any DNA fragment determined by two oligonucleotide primers, which are complementary to the DNA template. DNA amplification was performed either in the standard PCR reaction with thermostable Taq DNA polymerase or with the Expand Long Template PCR System (Roche). The second method allows for amplification of genomic DNA fragments up to 22 kb and is based on the activity of thermostable Taq DNA polymerase and Tgo DNA polymerase with "proofreading" activity.

A standard PCR reaction was prepared as follows:

dNTPs (1.25 mM)	4.0 µl
P20 buffer (20 mM $MgCl_2$)	2.5 µl
Forward primer (5 pM)	1.25 µl
Reverse primer (5 pM)	1.25 µl
Taq Polymerase (5 U/µl)	0.15 µl
SQ buffer	x µl
DNA template	x µl (100-200 ng)
	25 µl

An expand PCR reaction was prepared as follows:

dNTPs (20 mM)	0.5 µl
10x PCR buffer 1-3 (17.5 mM - 27.5 mM $MgCl_2$)	2.5 µl
Forward primer (10 pM)	1.5µl
Reverse primer (10 pM)	1.5 µl
Taq + Tgo polymerase (5 U/µl)	0.5 µl
H_2O	x µl
DNA template	x µl (300-400 ng)
	25 µl

The amplification profile of a standard PCR consisted of a pre-denaturation step at 94°C for 2 min, 34 cycles with 30 sec denaturation at 93°C, 1 min annealing at primer specific temperatures (45 - 65°C) and 1 min elongation at 72°C. The amplification was completed with a final elongation at 72°C for 5 min. In case of expand PCR reaction, I also used 34 cycles but with 15 sec denaturation at 93°C, 1 min annealing at primer specific temperatures (45 - 65°C) and an elongation step at 68°C with duration depending on the size of the expected PCR fragment (2 min for fragments up to 3 kb and 4 min for fragments up to 6 kb).

2.2.11. Agarose gel electrophoresis

Agarose gel electrophoresis is used for separation of molecules based on their size and charge. Different fluorescent dyes, which incorporate into the DNA strand, can be applied for visualization of the DNA fragments in UV light. For an agarose gel preparation

the appropriate amount of dry agarose powder was dissolved by heating in the proper volume of 1x TBE buffer. The fluorescent dye, ethidium bromide (10 mg/ml) was added to the warm gel solution (60°C) in an amount of 10 µl per each 100 ml of dissolved agarose. Then the liquid agarose was poured into a module, which was fitted with a well-forming comb, and left in RT to allow the gelification. 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 gel wells. Size markers (ladders) were also loaded next to DNA samples to aid in fragment size determination. After electrophoresis, the gel was placed on an UV light box and pictures of the florescent ethidium bromide-stained DNA separation pattern were taken with a video camera.

2.2.12. Isolation of the DNA fragments from agarose gels

With the use of the GFX PCR DNA and Gel Band Purification Kit (Amersham, Biosciences) variant substances such as primers, nucleotides, enzymes, mineral oils, salts, agarose, polyacrylamide, ethidium bromide, dyes, detergents and other impurities can be removed from the DNA samples. After PCR amplification DNA fragments were separated in low-melting temperature agarose gels and the specific bands were cut out of the gel with a scalpel under UV illumination. These gel slices were then mixed with the capture buffer and then applied to the GFX columns where the DNA is bound to the glass fiber matrix. The impurities were removed during washing steps and the pure DNA was eluted with 10 mM Tris-HCl (pH 8.3) or with sterile ddH_2O.

2.2.13. DNA sequencing

The PCR products were first purified using the Microcon Centrifugal Filter Units (Millipore) or extracted from agarose gel with the use of GFX PCR DNA and Gel Band Purification Kit (Amersham, Biosciences). The sequencing reactions were performed with the BigDye Terminator v.3.1 Cycle Sequencing Kit which contains thermally stable AmpliTaq DNA polymerase, modified dNTPs and a set of dye terminators labelled with high-senility dyes.

The sequencing reaction was prepared as follows:

BigDye sequencing mix	1 µl
Specific primer (10 pM)	0.5 µl
DNA template	x µl (10 ng)
ddH_2O	x µl
	10 µl

The sequencing program consisted of a pre-denaturation step at 96°C for 1 min, followed by 25 cycles of 10 sec denaturation at 96°C, 5 sec annealing at 55°C and 4 min elongation at 60°C. The finished sequencing reactions were filled up to 90 µl with ddH_2O and precipitated with 10 µl of 3 M NaAcetate (pH 5.2) and 250 µl absolute ethanol. After mixing, the samples were incubated at RT for 15 min and centrifuged at 13,000 rpm for 15 min. The pellets were washed with 200 µl of 70% ethanol, centrifuged at 13,000 rpm for 15 min and dried at RT. Then the DNA was resuspended at 37°C for 10 min in 20 µl of formamide (Hi-Di) and transferred into a sequencing plate. After denaturation of samples for 2 min at 95°C, automated DNA sequencing was performed with the ABI Prism 3100 sequencer (Applied Biosystems). Data collections were performed on the ABI PRISM 3100 Genetic Analyzer Data Collection Software v1.1 (Applied Biosystems).

2.2.14. *In silico* analysis

The sequence alignments and homology searches were performed with the BLAST program at the NCBI server (http://www.ncbi.nlm.nih.gov/BLAST/) and FASTA analyses using the Wisconsin Package Version 10.2, Genetics Computer Group. Further, the GenAlyzer software (Choudhuri et al. 2004) and the Miropeat software (Parsons 1995) were also used for the sequence comparisons. Inversion breakpoint regions were analyzed by reference to the Ensembl database based on NCBI Build 35 and on the chimpanzee draft sequence (Chimpanzee Sequencing and Analysis Consortium 2005). The search for segmental duplications was performed using the Human Genome Segmental Duplication Database (http://projects.tcag.ca/humandup) based on the NCBI Build 35 and the Non-Human Segmental Duplication Database (http://projects.tcag.ca/xenodup). Repeat masking was carried out using the RepeatMasker server from EMBL, Heidelberg (http://www.woody.embl-heidelberg.de/repeatmask/). The complexity analysis, based on text-compression algorithms (Gusev et al. 1999), of the chromosome 5 pericentric

inversion breakpoint regions was performed by Nadia Chuzhanova in the Biostatistics and Bioinformatics Unit in Cardiff University, United Kingdom.

The comparisons of the human-chimpanzee DNA sequence divergence of non-coding genomic regions were based on alignments of the chimpanzee draft sequence (Ensembl v27.1a.1) with sequenced human DNA clones (Appendix 1). I analysed those alignments, in which the best hit and overlap were given for at least 90% of the query sequence. Orthologous sequences were aligned after the masking of repetitive sequences and gap removal. The mean length of aligned sequence blocks was 22,768 bp for non-inverted regions on rearranged chromosomes, 20,443 bp for inverted regions, and 22,202 bp for collinear chromosomes.

The gene order comparison between human and chimpanzee was performed using the EnsMART search tool (http://www.ensembl.org/Multi/martview; Kasprzyk et al. 2004). The relative position of human genes mapped in the human reference sequence (Ensembl v28.35a.1 based on the NCBI 35 assembly of the human genome) were compared with those of their orthologous counterparts in the chimpanzee genome (Ensembl v28.1a.1; The Chimpanzee Sequencing and Analysis Consortium, 2005). All available human/chimpanzee orthologous gene pairs were extracted from this database (but only those human genes with confirmed status, as indicated by an NM_ accession number, were considered). For each gene included in the comparison, the gene ID, accession number, chromosomal location and starting position were retrieved. This initial dataset was however intrinsically redundant, since isoforms of the same gene often appeared as distinct entries. In order to avoid multiple representations of the same gene, the gene ID with the lowest numerical starting position was therefore selected. Care was taken to ensure that if more than one accession number was available for a given gene, only one example was included in the comparison. Furthermore, in order to avoid inclusion of false positive rearrangements, the genes to be selected were confined to those that were present within extended chimpanzee contigs and which were well anchored on both sides. The final dataset consisted of 11,518 orthologous gene pairs, and the list of the most proximal and distal genes found within the identified inversions is contained in Appendix 2. The gene functional analysis was performed according to the PANTHER Classification of Biological Processes system (Thomas et al. 2003a; 2003b; https://panther.appliedbiosystems.com) (Appendix 3).

3. Results

3.1. Characterization of the chimpanzee lineage-specific pericentric inversion of chromosome 5

3.1.1. Identification of the breakpoint-spanning BAC clones

From the comparative ideogram proposed by Yunis and Prakash (1982), the breakpoint positions of the chimpanzee-specific pericentric inversion of chromosome 5 should be located within the chromosomal bands 5p13.3 and 5q13.2, respectively. To identify the DNA clones, which are split by the inversion on the homologous chromosome of the chimpanzee, FISH with BAC/PACs from *HSA* 5p14.3-p15.1 and 5q13.2-q14.1 was performed (Table 12). First, *HSA* BAC RP11-35A11 was identified as spanning the inversion breakpoint at 5p15.1 (Figure 6A, B). This BAC is however located distally from the putative breakpoint region assigned by Yunis and Prakash (1982) with the use of chromosome banding technique. Probes located distal to the *HSA* BAC RP11-35A11 hybridized to *PTR* 5p (were not inverted), whereas those located more proximally hybridized to the inverted region and yielded signals on *PTR* 5q. All investigated probes from *HSA* 5q13.2-5q14.1 were however inverted to the p arm in chimpanzee. During the time of these experiments, the initial version of chimpanzee draft sequence became available what facilitated BLAST searches for the homology between *HSA* BAC RP11-35A11 and the published chimpanzee sequence. By this means the *HSA* BAC RP11-432G16 from 5q15 was identified that showed split signal on the homologous chimpanzee chromosome (Figure 6C, D).

Further BLAST sequence comparisons led to the identification of two chimpanzee BACs: RP43-023F21 and RP43-001C16, which exhibited sequence similarities to both human BACs RP11-432G16 and RP11-35A11 (Figure 7). As was to be expected from breakpoint-spanning clones, the *PTR* BACs displayed split FISH signals on *HSA* 5 (Figure 6). *PTR* BAC RP43-001C16 exhibited a very weak FISH signal on *HSA* 5p and a much stronger signal on *HSA* 5q, suggesting that the breakpoint was acentrically located within this BAC (Figure 6A).

Table 12: FISH analysis with BAC/PAC clones from *HSA* 5 hybridized to human and chimpanzee metaphase spreads

BAC name	Chromosome band	BAC position in the *HSA* 5 contig [a]	Localization on	
			HSA	*PTR*
RP1-137K24	5p15.1	15,191,411-15,315,356	p	p
RP11-19O2	5p15.1	16,193,635-16,358,453	p	p
RP11-55N10	5p15.1	17,583,652-17,754,449	p	p
RP11-211K15	5p15.1	18,298,331-18,465,883	p	P
RP11–35A11 [b]	5p15.1	18,434,022-18,604,727	p	p+q
RP11-270H9	5p14.3	19,044,235-19,238,504	p	q
RP11-60F22	5p14.3	19,727,845-19,825,926	p	q
RP11-414I19	5p14.3	20,414,929-20,595,578	p	q
RP11-774D14	5p14.3	20,856,715-21,047,822	p	q
RP11-118I21	5q13.2	71,141,615-71,301,758	q	p
RP11-60A8	5q13.2	72,529,826-72,709,210	q	p
RP11-229C3	5q13.3	74,286,601-74,469,980	q	p
RP11-524L6	5q13.3	75,103,518-75,316,272	q	p
RP11-206N2	5q13.3	76,144,573-76,251,850	q	p
RP11-107N7	5q14.1	77,156,054-77,322,769	q	p
RP11-21K15	5q14.1	78,203,458-78,284,614	q	p
RP11-432G16 [c]	5q15	95,843,232-96,023,456	q	p+q

[a] Ensembl version 26.35.1 based on the NCBI Build 35 assembly; [b] p-arm breakpoint-spanning BAC; [c] q-arm breakpoint-spanning BAC.

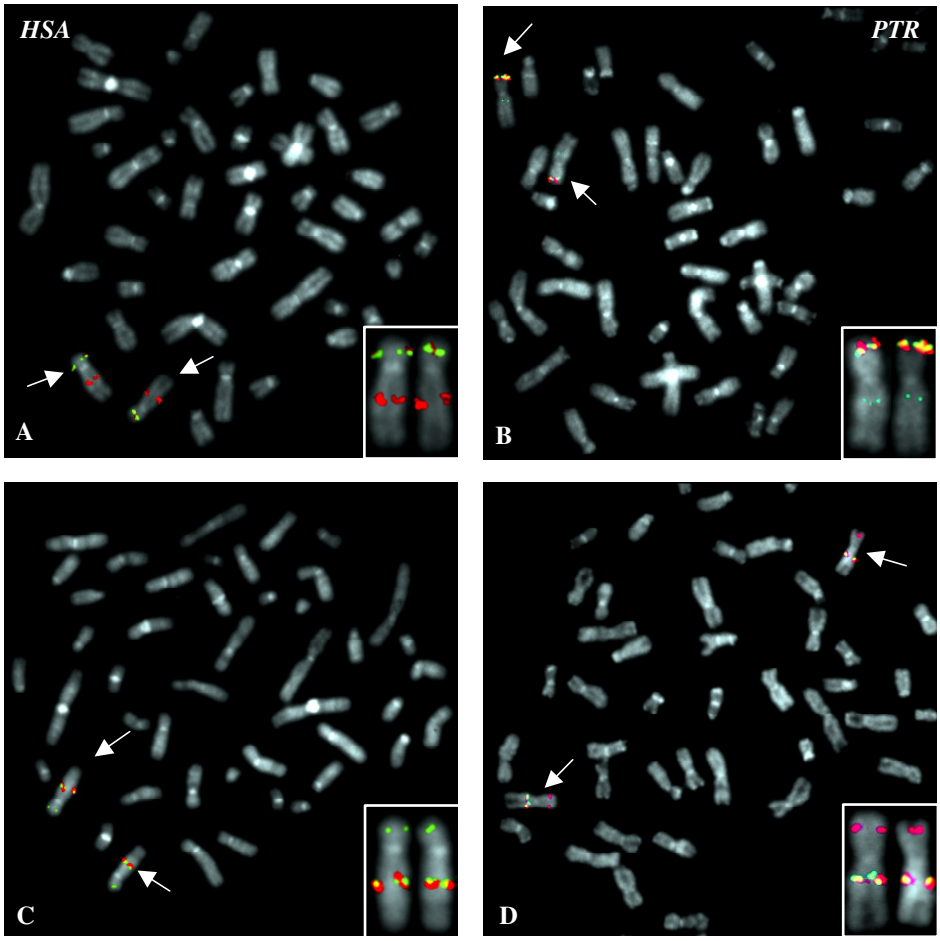

Figure 6: **(A,B)** FISH-analysis using breakpoint-spanning BAC clones. Arrows indicate the chromosomes involved in the inversion. *HSA* BAC RP11-35A11 (green) from 5p14.3-15.1 co-hybridized with *PTR* 5p BAC RP43-001C16 (red) to human (**A**) and chimpanzee (**B**) metaphases. The split signal of *PTR* BAC RP43-001C16 on *HSA* 5p is only faintly visible. (**C, D**) FISH signals of *HSA* BAC RP11-432G16 (red) from 5q15 and *PTR* 5q BAC RP43-023F21 (green) on human (**C**), and chimpanzee (**D**) chromosomes.

3.1.2. Characterization of the breakpoint regions architecture

The coverage of the *PTR* 5/*HSA* 5 breakpoint-spanning BACs from the human and chimpanzee genomes was determined by BLAST analyses and their relative positions are schematically depicted in Figure 7.

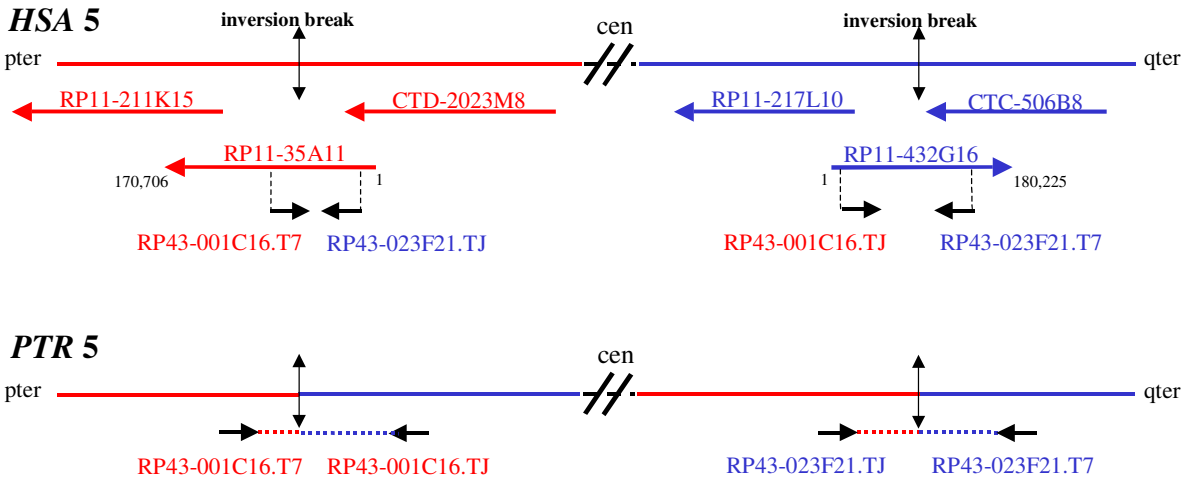

Figure 7: Schematic representation of the pericentric inversion in chimpanzee chromosome 5 (*PTR 5*) orthologous to human chromosome 5 (*HSA* 5). The relative position of the breakpoint-spanning BACs from the human (RP11-35A11 and RP11-432G16) and chimpanzee (RP43-001C16 and RP43-023F21) genomes are indicated. For *PTR* BAC RP43-001C16 (AG141160, AG141159) and *PTR* BAC RP43-023F21 (AG158210, AG158209), end sequences are available which served to anchor their map positions. Cen: centromere.

The homology between the chimpanzee chromosome 5 scaffold and the human sequence was found to be disrupted at positions 15,650 and 15,658 of *HSA* BAC RP11-35A11. As indicated in Figure 8, the inversion appears to be associated with a 7 bp deletion in the chimpanzee genome (corresponding to positions 15,651-15,657 of RP11-35A11). Comparison of the chimpanzee genome with *HSA* BAC RP11-432G16 indicated the existence of two other small chimpanzee-specific rearrangements in the breakpoint regions; an 11 bp duplication and a 9 bp deletion (Figure 8A). The DNA sequence in the immediate vicinity of the inversion breakpoints was examined for the presence of mutation-prone motifs known to play a role in recombination (Abeysinghe et al. 2003). Polypyrimidine and polypurine tracts of 7 bp and 16 bp respectively, were found adjacent to the 5p breakpoint in *HSA* BAC RP11-35A11, whilst the breakpoint on 5q is preceded by two alternating purine-pyrimidine tracts of length 10 bp and 6 bp, respectively (Figure 8B). Using complexity analysis (Gusev et al. 1999), a method based on text-compression

algorithms, the local DNA sequences in the immediate vicinity of the breakpoints were searched for direct and inverted repeats, symmetric elements, and inversions of inverted repeats. Such repeat sequences are capable of forming slipped structures, triplexes, tetraplexes and cruciforms that can promote the occurrence of double-strand breaks. Two short direct repeats, AAAAA and TATAC, shown in bold in Figure 8B, are capable of slipped structure formation that could have contributed to DNA breakage.

Repeatmasker analyses of the extended DNA sequences flanking the breakpoints indicated that the break on *PTR* 5p disrupted an LTR element, whereas the breakpoint in *PTR* 5q occurred within a single copy segment. Complexity analysis was also used to examine the sequences in the vicinity of the breakpoints. Although several inverted and direct repeats were found in the region covered by BACs RP11-35A11 and RP11-432G16, none of these repeats were located adjacent to both breakpoints. Nevertheless, they could still have been responsible for bringing DNA ends into close juxtaposition thereby potentiating illegitimate end joining, a first step toward the inversion.

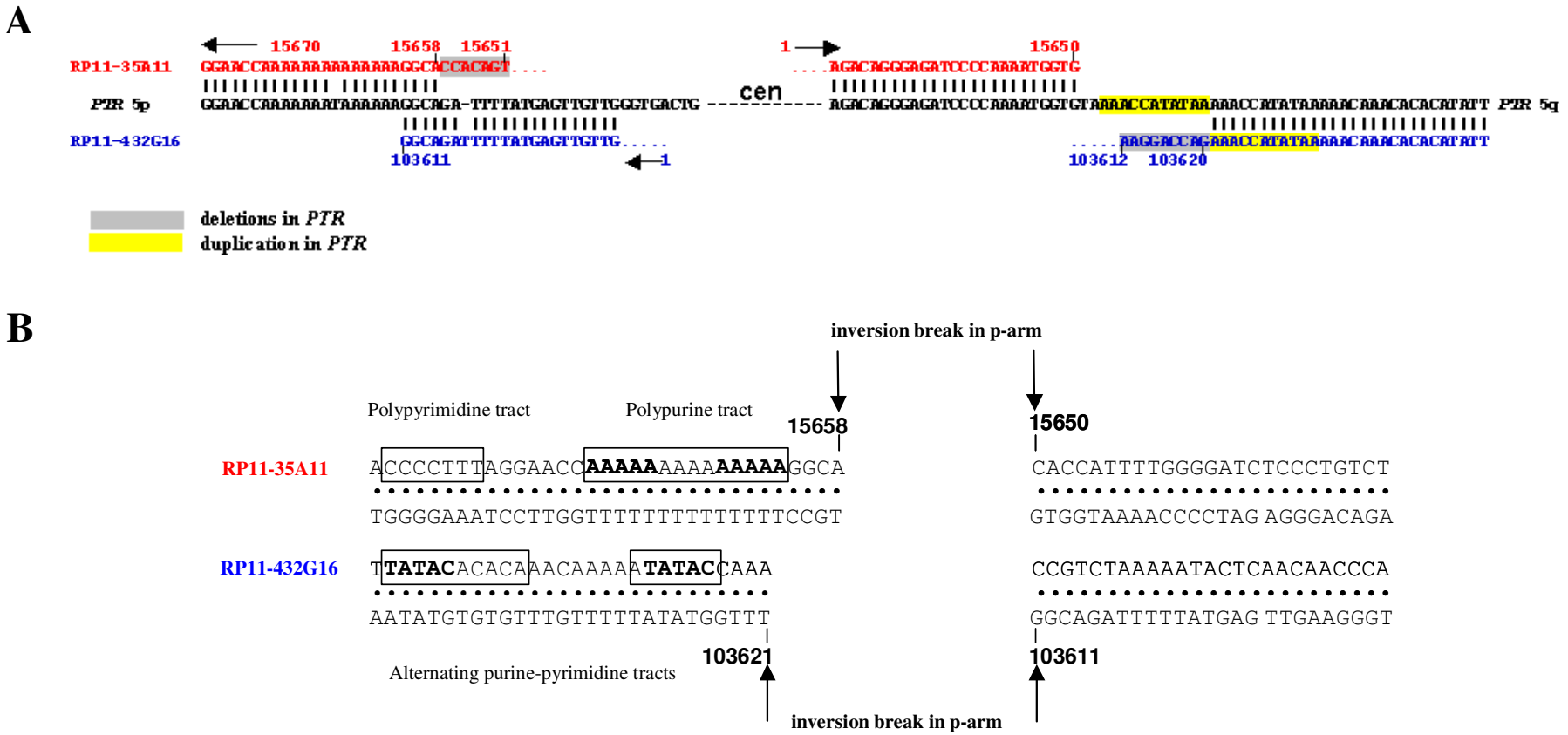

Figure 8: **(A)** Sequence comparison of *HSA* breakpoint-spanning BACs (RP11-35A11 and RP11-432G16) with the available sequence of the appropriate portion in the chimpanzee genome. The alignments indicate the sites of homology interruption between human and chimpanzee sequences (denoted by vertical lines). The two chimpanzee-specific deletions are indicated by grey background shading. The 11 bp duplication in PTR 5q is marked by yellow background shading.

(B) Pyrimidine, polypurine and alternating purine-pyrimidine (RY) tracts are located close to the breakpoint sites (denoted by rectangles). Short direct repeats in the immediate vicinity of the breakpoints are shown in bold.

3.1.3. Examination of the gene content in the breakpoints vicinity

Comparison of the breakpoint-spanning human BACs with the draft sequence of the chimpanzee genome revealed that the breakpoints of this ~ 80 Mb pericentric inversion do not disrupt the coding sequence of any gene. In the context of *HSA* 5, which represents the ancestral form of this chromosome, the q-arm breakpoint maps between the *PCSK1* (proprotein convertase subtilisin/kexin type 1) and *CAST* (calpastatin) genes, and is located at least 150 kb from the former and 76 kb from the latter one. The inversion breakpoint on the p-arm is located in an even more gene-poor area, about 1.26 Mb distant from the *BASP1* (brain abundant, membrane attached signal protein 1) gene and 920 kb from the *CDH18* (cadherin 18, type 2) gene (Figure 9).

Previously performed comparative expression studies (Caceres et al. 2003; Enard et al. 2002; Karaman et al. 2003; Khaitovich et al. 2004; Uddin et al. 2004) have failed to provide any evidence that this inversion affected the expression of genes residing in the vicinity of the breakpoints.

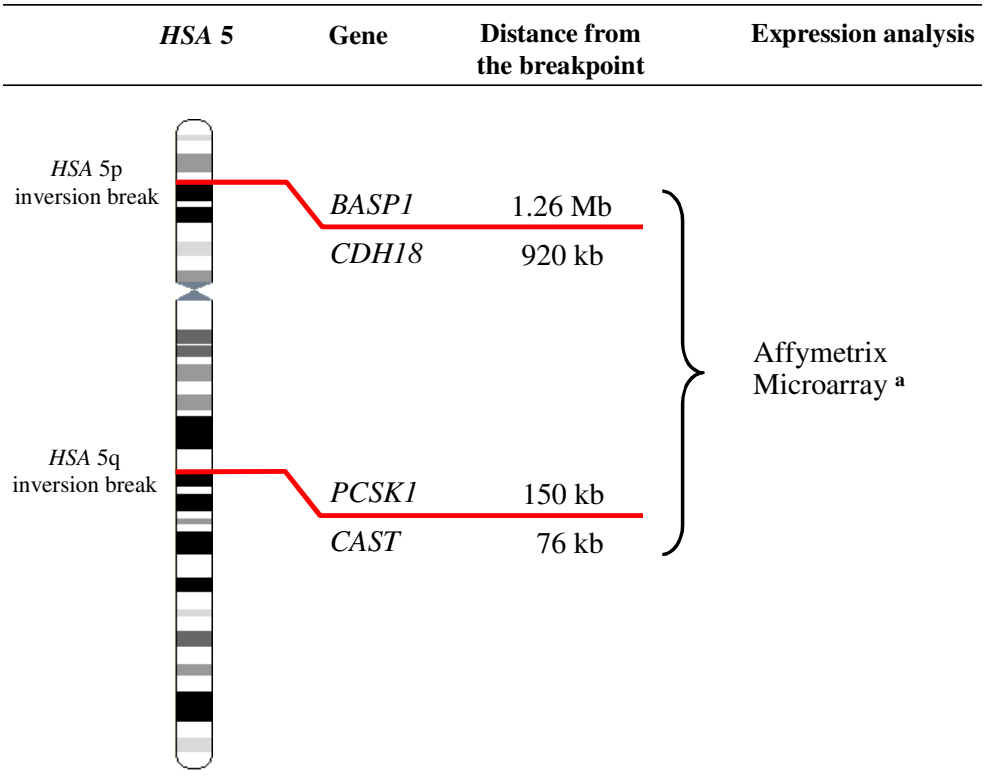

Figure 9: Schematic representation of HSA 5. The genes located next to the inversion breakpoints are indicated. None of these genes had been previously identified as being differentially expressed between *PTR* and *HSA* in microarray-based expression analyses using the Affymetrix Gene Chips (Caceres et al. 2003, Enard et al. 2002, Karaman et al. 2003, Uddin et al. 2004, Khaitovich et al. 2004).

3.2. Characterization of the human lineage-specific pericentric inversion of chromosome 1

Chromosome 1 constitutes the largest physical element of the human and great ape karyotypes and as such was shown to represent an intact, ancestral chromosome, maintained as a single unit in phylogentically distinct taxa (Murphy et al. 2003). On the other hand, it is still prone to various intra-chromosomal rearrangements and instabilities, reported in research on human polymorphism (Tuzun et al. 2005; Sebat et al. 2004; Conrad et al. 2005; McCarroll et al. 2005), pathology (Barbi et al. 2005; Faas et al. 2003; Mitelman et al. 2006) and mammalian evolution (Haig 2005). The below presented results give an insight into the complex structure of the pericentromeric region of this chromosome.

3.2.1. Computational comparison of the pericentromeric region of *HSA* 1 and *PTR* 1

The pericentromeric region of *HSA* 1 (ranging from 118 Mb to 147 Mb according to the NCBI Build 35) was compared with the orthologous sequences of the chimpanzee draft sequence (The Chimpanzee Sequencing and Analysis Consortium 2005) using the GenAlyzer software (Choudhuri et al. 2004). This analysis revealed the presence of numerous duplicated sequences with homologues on both 1p and 1q (Figure 10A). To distinguish between inverted and intra-chromosomally duplicated sequences, the latter were excluded from the alignments; only unduplicated sequences larger than 1 kb were considered. According to these comparisons, the inversion breakpoint on *HSA* 1p maps to a ~538 kb region between 120,240,975 bp and 120,778,956 bp at 1p11.2-12 whereas the q-arm breakpoint is located within a ~4 Mb region at 1q21.1-21.2 between positions 142,657,563 bp and 146,690,980 bp (Figure 10B).

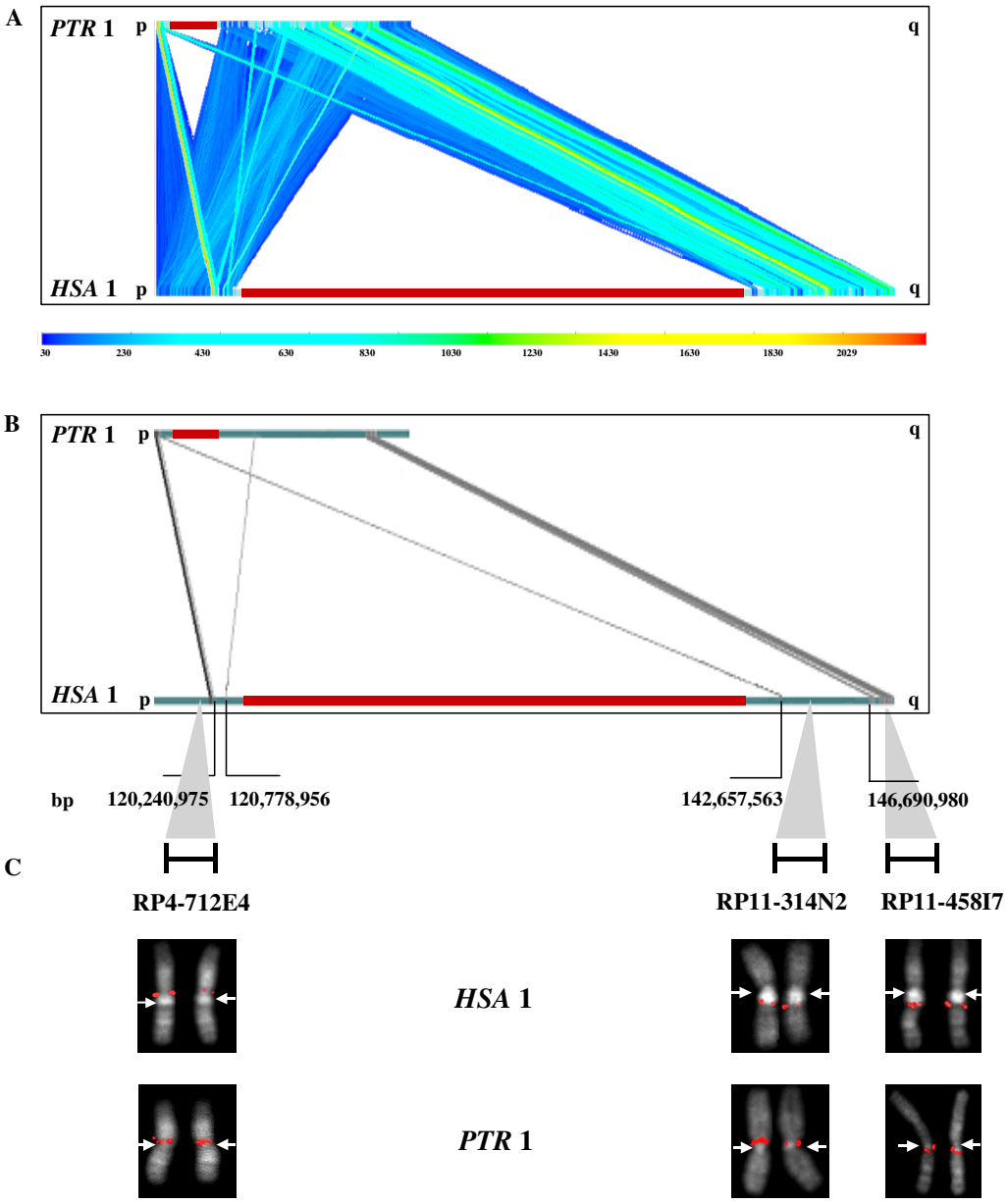

Figure 10: GenAlyzer display of homologous sequences between human and chimpanzee chromosome 1 in the pericentromeric regions (118 - 147Mb according to the NCBI Build 35). The red horizontal bars represent the centromeres and heterochromatin. Homologous sequences between *PTR* and *HSA* are indicated as vertical or diagonal lines. **(A)** Homologous sequences between *PTR* 1 and *HSA* 1 ranging from 30 to 2,029 bp according to the colour scale below, which indicates the length of the respective segments. **(B)** Homologous sequences ranging from 1,000 to 2,029 bp after exclusion of the duplicated sequences. According to these analyses, the p-arm inversion breakpoint maps to the region between 120,240,975 bp and 120,778,956 bp, whereas the q-arm breakpoint is located between 142,657,563 bp and 146,690,980 bp. **(C)** Clones RP4-712E4 and RP11-458I7 localize outside the putative inversion region on *HSA* 1 and hybridised to the same chromosomal arm in *HSA* and *PTR*. BAC RP11-314N2 maps within the inverted region since it hybridized at *HSA* 1q21 but in chimpanzee at 1p11-12. The white arrows indicate the position of the centromere.

3.2.2. FISH analysis of the inversion using BAC/PACs that do not span segmental duplications

In order to confirm the results obtained by sequence alignment, comparative FISH was performed on human and chimpanzee chromosomes (Tables 13 and 14). In accordance with the results of the *in silico* analyses, *HSA* PAC RP4-712E4 from 1p12 and BAC RP11-458I7 from 1q21.2 were shown not to be inverted (Figure 10C). The ~22 Mb inverted region (between 120,778,956 bp and 142,657,563 bp) contains numerous segmental duplications. This notwithstanding, the respective regions were screened to identify interspersed unduplicated sequences that would facilitate unambiguous FISH-based breakpoint mapping. By these means, I identified BACs RP11-30I17, RP11-337C18, and RP11-314N2 which form part of contigs NT_004434 and NT_034400 in *HSA* 1q21.1. FISH analysis then revealed that these BACs are inverted since they hybridize at *HSA* 1q21 but at *PTR* 1p11-12 (BAC RP11-314N2, Figure 10C; Table 14).

Taken together, these FISH analyses allowed to narrow down the q-arm breakpoint to a ~2 Mb region between 144,630,713 bp (map position of BAC RP11-314N2) and 146,690,980 bp. All BAC/PAC clones located within this ~2 Mb region on *HSA* 1q21.1-21.2, as well as within the ~538 kb region between 120,240,975 bp and 120,778,956 bp on *HSA* 1p11.2-12, within which the inversion breakpoints must have occurred, contain segmental duplications.

3.2.3. Inversion breakpoint mapping by FISH and sequence comparisons using BAC/PACs that span the segmental duplications

To refine further the breakpoint regions, I investigated several BAC/PACs that map to the above mentioned breakpoint intervals and which contain segmental duplications.

3.2.3.1. Breakpoint on *HSA* 1p

PACs RP4-683H9, RP5-1042I8 and BAC RP11-114O18 map to *HSA* 1p12 in Ensembl (v37.35j; NCBI Build 35) but FISH revealed additional signals on chromosome 1 due to segmental duplications detected by these clones (summarized in Table 13). GenAlyzer and BLAST analyses indicated that the original locations of these clones mapping to *HSA* 1p12 are on the p-arm of *PTR* 1 and therefore cannot be inverted. Accordingly, all three clones displayed FISH signals only on the p-arm of *PTR* I. From the

map position of BAC RP11-114O18 (Table 13; Figure 11A, B) on *HSA* 1p and the results of the sequence alignments, the p-arm breakpoint was narrowed down to the 382 kb region on 1p11.2, from positions 120,395,759 bp to 120,778,956 bp.

Table 13: FISH analyses designed to narrow down the inversion breakpoint on *HSA* 1p

BAC/PAC	Chromosomal position	Map position in bp [a]	FISH signals on		Position with respect to the inversion [b]
			HSA	*PTR*	
RP11-192J8	1p12	118,079,902-118,243,631	**p12**	**p12**	not inverted
RP4-712E4	1p12	119,194,364-119,316,485	**p12**	**p12**	not inverted
RP4-683H9	1p12	119,979,355-120,045,236	p36, **p12**, q21	p36, **p12**	not inverted
RP5-1042I8	1p12	120,045,137-120,183,091	p36, **p12**, q21	p36, **p12**	not inverted
RP11-114O18	1p11.2-12	120,288,704-120,395,759	**p11-12**, q21	**p11-12**	not inverted
RP11-439A17 [c]	1p11.2	120,459,199-120,648,737	**p11**, **q21**, q32	**q21**, q32	in breakpoint region
RP11-343N15	1p11.2	120,698,738-120,794,484	**p11**, **q21**, q32	**p11**, **q21**, q32	inverted
RP11-435B6	1p11.1-11.2	120,956,393-121,097,476	**p11**	centromere	inverted

[a] Ensembl version 37.35j based on the NCBI Build 35 assembly; the main signals of the respective probes are indicated in bold; [b] According to the combined results of FISH analyses and sequence alignments; [c] On the basis of the ~91 kb sequence homology of BAC RP11-439A17 to the q-arm PAC RP5-998N21, it was concluded that non-allelic homologous recombination between these inverted repeats probably mediated the inversion.

BAC RP11-343N15 is located on the proximal boundary of this 382 kb region and hybridizes both to *HSA* 1p11, 1q21, 1q32 and to orthologous regions on *PTR* 1. Owing to the segmental duplications detected by this BAC, FISH analyses *per se* would not have been informative. However, taken together with detailed sequence analysis, RP11-343N15 turned out to be informative with respect to breakpoint mapping. In addition to the intra-chromosomally duplicated sequences, RP11-343N15 also spans a 19 kb segment that has not been duplicated on chromosome 1 (1,789–20,655 bp of AL358175). Sequence alignments of this 19 kb region indicated that it is inverted in chimpanzee as compared to human. Thus, RP11-343N15 must be located within the boundaries of the *HSA* 1 inversion.

The completely sequenced BAC RP11-435B6 from *HSA* 1p is located very close to the centromere in the Ensembl database and contains 92.94% alpha satellite DNA. Employing FISH, only faint signals from this BAC were noted in the vicinity of the centromere on *PTR* 1, indicative of the weak conservation of chromosome 1-specific alpha-satellite sequence between humans and chimpanzees.

BAC RP11-439A17 played an important role with respect to the mapping of the *HSA* 1p breakpoint which is localized between inverted and non-inverted clones (Table 13). Unfortunately, this BAC is unconnected to flanking sequence contigs and is instead surrounded by sequence gaps. RP11-439A17 detects a human-specific segmental duplication as determined by FISH performed in this study (Figure 11C, D) and array CGH as previously described (Goidts et al. 2006). Three signals were observed on human chromosome 1, at p11, q21 and q32 but only two signals were noted on *PTR* 1q in regions orthologous to *HSA* 1q21 and 1q32. It was not possible solely on the basis of this FISH result to categorize the BAC as mapping either inside or outside the inverted region. However, since this BAC is directly flanked by inverted and non-inverted regions, and exhibits high sequence homology to PAC RP5-998N21 (which maps to the breakpoint region on *HSA* 1q), it is concluded that the inversion breakpoint on *HSA* 1p is located either within BAC RP11-439A17 or within the adjacent unsequenced gaps.

3.2.3.2. Breakpoint on *HSA* 1q

FISH and sequence alignments were employed to investigate several BAC/PAC clones mapping to the ~2 Mb breakpoint interval on *HSA* 1q. Although clones RP3-328E19 and RP11-763B22 span segmental duplications, FISH was informative since the signal patterns differed between human and chimpanzee. Both clones displayed strong hybridization signals on *PTR* 1p11-12 but hybridized to both chromosomal arms in human (1p11-12 and 1q21; RP11-763B22, Figure 11E, F). I therefore conclude that the loci detected by these clones on 1q21 are inverted by comparison with *PTR* 1.

Table 14: FISH analyses designed to narrow down the inversion breakpoint in *HSA* 1q

BAC/PAC	Chromosomal position	Map position in bp [a]	FISH signals on *HSA*	FISH signals on *PTR*	Position with respect to the inversion [b]
RP11-30I17	1q21.1	142,951,739-143,139,578	**q21**	**p11-12**	inverted
RP11-337C18	1q21.1	143,826,832-144,029,786	**q21**	**p11-12**	inverted
RP11-314N2	1q21.1	144,490,622-144,630,713	**q21**	**p11-12**	inverted
RP3-328E19	1q21.1	145,043,332-145,140,950	p36, p11-12, **q21**	p36, **p11-12**	inverted
RP11-763B22	1q21.1	145,513,740-145,671,828	p36, p11-12, **q21**	p36, **p11-12**	inverted
RP11-14N7	1q21.1	145,669,829-145,727,426	p11-12, **q21**	**p11-12**, q21	in breakpoint region
RP11-744H18	1q21.1	145,784,053-145,890,590	p36, p11-12, **q21**	p36, **p11-12**, q21	in breakpoint region
RP11-403I13	1q21.1	145,888,591-146,093,292	p36, p11-12, **q21**	p36, **p11-12, q21**	in breakpoint region
RP5-998N21 [c]	1q21.1	146,143,293-146,272,718	**p11**, **q21**, q32	**q21**, q32	in breakpoint region
RP11-277L2	1q21.1	146,322,719-146,401,747	p36, p11-12, **q21**	p36, p11-12, **q21**	in breakpoint region
RP11-353N4	1q21.1-2	146,399,748-146,524,913	p36, p11-12, **q21**	p36, p11-12, **q21**	in breakpoint region
RP11-196G18	1q21.2	146,522,914-146,711,869	p11-12, **q21**	**q21**	not inverted
RP11-21J4	1q21.2	146,606,778-146,787,659	p11-12, **q21**	**q21**	not inverted
RP11-458I7	1q21.2	146,785,068-146,973,278	**q21**	**q21**	not inverted
RP11-363I22	1q21.2	147,432,838-147,586,905	p36, **q21**	p36, **q21**	not inverted

[a] Ensembl version 37.35j based on the NCBI Build 35 assembly; the main signals of the respective probes are indicated in bold; [b] According to the combined results of FISH analyses and sequence alignment; [c] On the basis of the ~91 kb sequence homology of PAC RP5-998N21 to the p-arm BAC RP11-439A17, it was concluded that non-allelic homologous recombination between these inverted repeats probably mediated the inversion.

BAC RP11-196G18 was also informative with respect to breakpoint mapping, exhibiting 46 kb of sequence homology with BAC RP11-439A17 from *HSA* 1p11.2. This 46 kb region was specifically duplicated in the human lineage as evidenced by array CGH (Goidts et al. 2006). FISH analysis with RP11-196G18 and the overlapping BAC RP11-21J4 confirmed the duplication, since in addition to a strong signal on *HSA* 1q21, a signal at 1p11-12 was noted (RP11-196G18, Figure 11G). Since this duplication is absent from *PTR*, BACs RP11-196G18 and RP11-21J4 hybridized exclusively to *PTR* 1q21 (RP11-196G18, Figure 11H). Since the 'original' locus is located on the q-arm in both human and chimpanzee, I conclude that the q-arm inversion breakpoint must map proximal to the region covered by BAC RP11-196G18 on *HSA* 1q21 (Table 14). In summary, the FISH patterns of BACs RP11-196G18 and RP11-763B22 delimit the inversion breakpoint on *HSA* 1q to the 851 kb interval between positions 145,671,828 bp and 146,522,914 bp.

Subsequently, I investigated BAC/PAC clones from this interval and noted that, with the exception of PAC RP5-998N21, all of them hybridize to both p and q arms in *HSA* and *PTR* (Table 14). Surprisingly, PAC RP5-998N21, which is not anchored on either side to any sequence contig, exhibits an identical hybridization pattern to BAC RP11-439A17 located within the p-arm breakpoint region. The similar FISH patterns displayed by RP11-439A17 and RP5-998N21 result from a 91 kb segmental duplication. According to the Human Genome Segmental Duplication Database, the duplication covered by RP5-998N21 is termed DC0215 and shares >99% sequence homology with DC0154 covered by RP11-439A17.

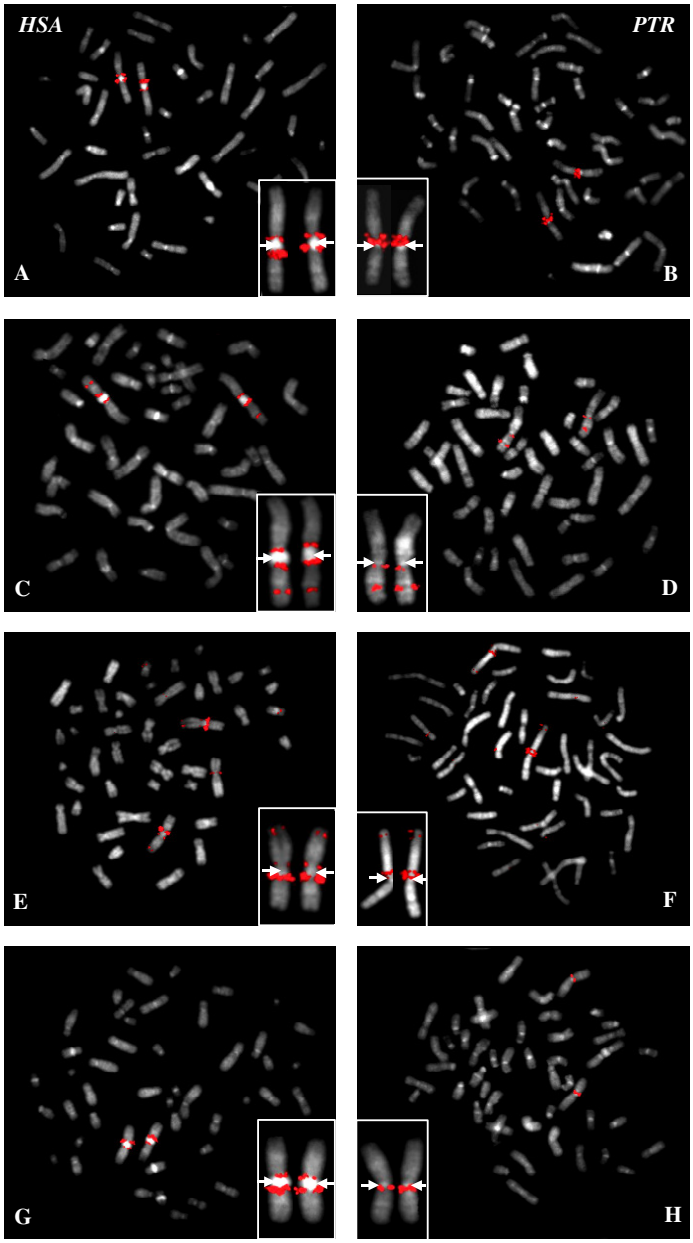

Figure 11: FISH analysis performed with human BACs from the pericentromeric region containing segmental duplications. White arrows indicate the position of the centromere. **(A,B)** BAC RP11-114O18 detects a duplication on 1q21 in human, but hybridizes only on the p-arm of *PTR* 1. **(C,D)** BAC RP11-439A17 from *HSA* 1p covers a human-specific segmental duplication located at *HSA* 1p11, and paralogous sequences in 1q21 and 1q32. In *PTR*, no signal was observed on the p-arm. **(E,F)** BAC RP11-763B22 maps to 1q21 according to Ensembl (v37.35j based on NCBI Build 35) and detects segmental duplications on 1p36 and 1p11-12. This BAC displayed signals only on the p-arm of *PTR* 1 therefore it must be within the inverted region. **(G,H)** The unduplicated portion of BAC RP11-196G18 hybridized at 1q in humans and chimpanzees and is thus not inverted in humans.

3.2.4. Proposed mechanism underlying the inversion

Miropeat analysis (Parsons 1995) indicated many regions of high sequence homology between the pericentromeric regions of *HSA* 1p and 1q (Figure 12).

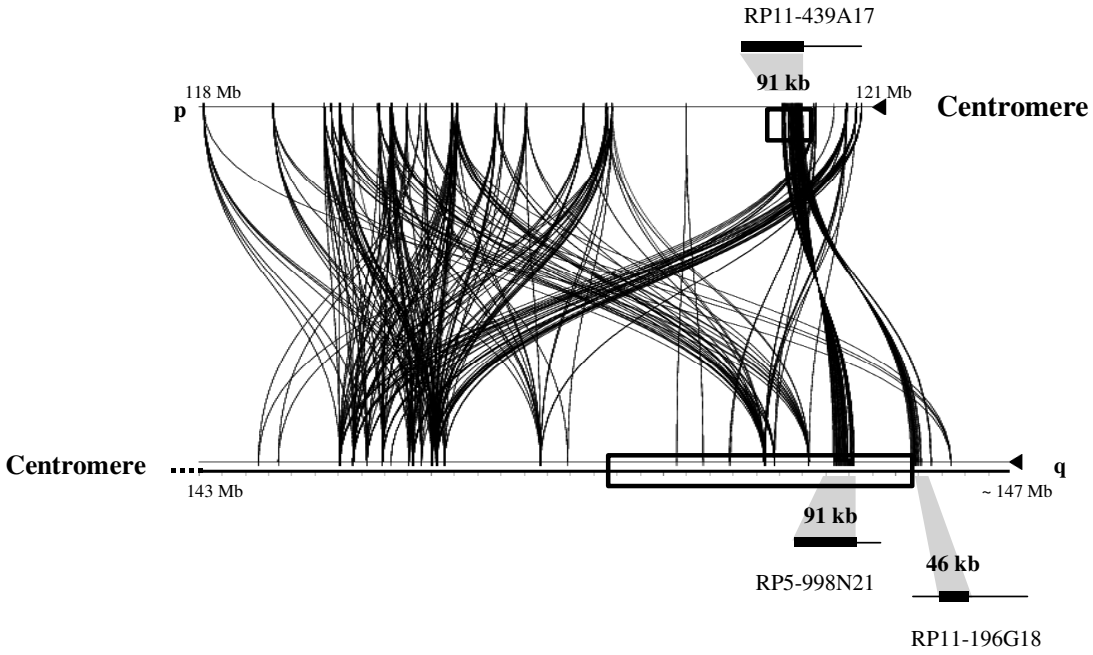

Figure 12: Comparison of the pericentromeric region of *HSA* 1 using the Miropeat software (Parsons 1995). The input sequences from *HSA* 1p and *HSA* 1q are represented as horizontal lines. ~3 Mb from the p-arm are compared to ~4 Mb of the q-arm. The homology between the p-arm and the q-arm are depicted as connectors between both input sequences. Each connector symbolizes at least 2 kb. BAC RP11-439A17 from 1p shows sequence homology with PAC RP5-998N21 and BAC RP11-196G18 of 91 kb and 46 kb, respectively. PAC RP5-998N21 is located within the region of the putative q-arm breakpoint (indicated with horizontal rectangle), while RP11-196G18 maps outside the inversion, as shown by FISH analysis.

As mentioned above, segments of high homology (>99%) between the p- and q-arm breakpoint-spanning regions are covered by clones RP11-439A17 and RP5-998N21. It is highly likely that these 91 kb segmental duplications mediated, or at least facilitated, the inversion (Figure 12). These 91 kb segmental duplications are absent in the chimpanzee and therefore arose after the divergence of the human and chimpanzee lineages. A model to explain how these segmental duplications could have arisen during human evolution is presented in Figure 13. In the chimpanzee, sequences with homology to portions of human clones RP5-998N21 and RP11-439A17 map to *PTR* 1q21 and 1q32, respectively (Figure 13A). Two duplicative transposition events occurred in the ancestral forms of *HSA* 1 (Figure 13B, C) thereby giving rise to the highly homologous 91 kb segmental duplications

located in inverted orientation on the p- and q-arms of the ancestral chromosome 1 (Figure 13C). It would appear likely that the high degree of sequence homology between these inverted repeats brought chromosomal regions 1p11 and 1q21 together, thereby mediating the intra-chromosomal recombination event (Figure 13D). According to this model, based both on FISH data and sequence comparisons, the inversion breakpoints map to between 120,395,759 bp and 120,549,995 bp on *HSA* 1p and between 145,671,828 bp and 146,234,395 bp on *HSA* 1q.

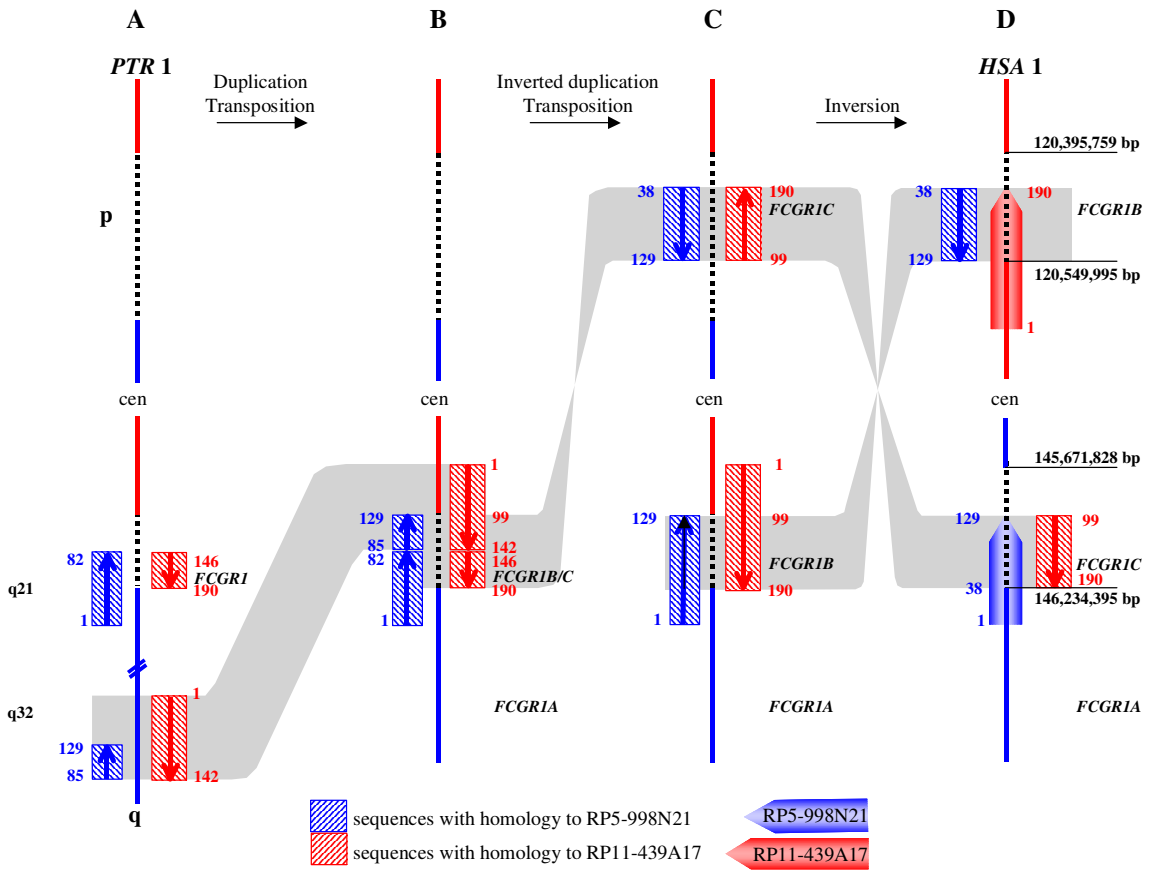

Figure 13: Proposed model of the mechanism which underlies the pericentric inversion of *HSA* 1 (detailed explanations are given in the text above). **(A)** Schematic representation of *PTR* 1 homologous to *HSA* 1. The red and blue striped boxes indicate sequences homologous to the human BAC RP11-439A17 and PAC RP5-998N21, respectively. Their orientation is indicated by arrows. **(B, C)** Ancestral form of *HSA* 1. **(D)** Extant human chromosome 1. The red and blue vertical lines represent the p-and q-arms respectively. Red and blue numbers indicate the positions within a BAC/PAC, in kb, whilst the black numbers refer to the positions within the Human Genome Assembly (NCBI Build 35). The dotted lines denote the region of the putative position of the inversion breakpoints. According to this inversion model, clones RP11-439A17 and RP5-998N21 are located in inverted orientation by comparison with the data given by Ensembl Database (v 37.35j; based on NCBI Build 35), where neither clone is anchored to any contig within the chromosome 1 assembly. The position of *FCGR1* genes is schematically depicted.

3.2.5. Analysis of the gene content in the inversion breakpoint regions

The inversion breakpoint regions were assessed in terms of their gene content. Owing to the 91 kb sequence homology between BAC RP11-439A17 and PAC RP5-998N21, the genes encompassed by these clones represent duplicated, human-specific copies. Among them are genes encoding the human high-affinity receptors for immunoglobulin G (*FCGR1B*, *FCGR1C*, the histone 2 pseudogenes (*HIST2H2B*, *HIST2H2BA*) and the *FAM72A* and *FAM72C* genes (family with sequence similarity 72, member A and C). In the region covered by BAC clones RP11-14N7, RP11-744H18 and RP11-403I13 on the q-arm, mainly tRNA pseudogenes and genes encoding spliceosomal RNA were noted. Since the exact nucleotide positions of the breakpoints are unknown, I cannot establish whether any coding sequence was disrupted by the inversion.

3.2.6. Evolutionary history of the *HSA* 1 pericentric inversion

In addition to the pericentric inversion of *HSA* 1, a further large pericentric inversion of this chromosome occurred during primate evolution. Comparative FISH analyses indicated that the chromosome 1 homologue of the Old World monkeys differs from the analogous chromosome in great apes by a second pericentric inversion (Buckland et al. 1992; Maresco et al. 1998; Ruiz-Herrera et al. 2002). To verify these findings and to investigate the question of breakpoint position similarity or heterogeneity, I performed FISH with BACs RP11-314N2 and RP11-458I17. In the macaque (*MFA*), an Old World monkey, both BACs map to *MFA* 1p (Figure 14A). The first pericentric inversion occurred after the separation of the Old World monkeys from the lineage leading to the great apes and humans. As a consequence, BAC RP11-458I7 is inverted in *PTR* 1q (Figure 14B). The second pericentric inversion is confined to the human lineage and was introduced after the separation from chimpanzees. This second inversion includes the region covered by BAC RP11-314N2, which maps to the q-arm in humans. Thus, I may conclude that the two inversions that have occurred during the evolutionary history of chromosome 1 in primates have different breakpoints (Figure 14D). As a consequence of the first inversion, the *FCGR1* gene became inverted in great apes by comparison with Old World monkeys (Maresco et al. 1998). During human evolution, two duplications gave rise to the *FCGR1A, FCGR1B* and *FCGR1C* genes, which are separated by the second inversion that occurred in the human lineage. From these comparisons, it may be inferred that the breakpoints of both inversions are separated by a minimum of several hundred kb.

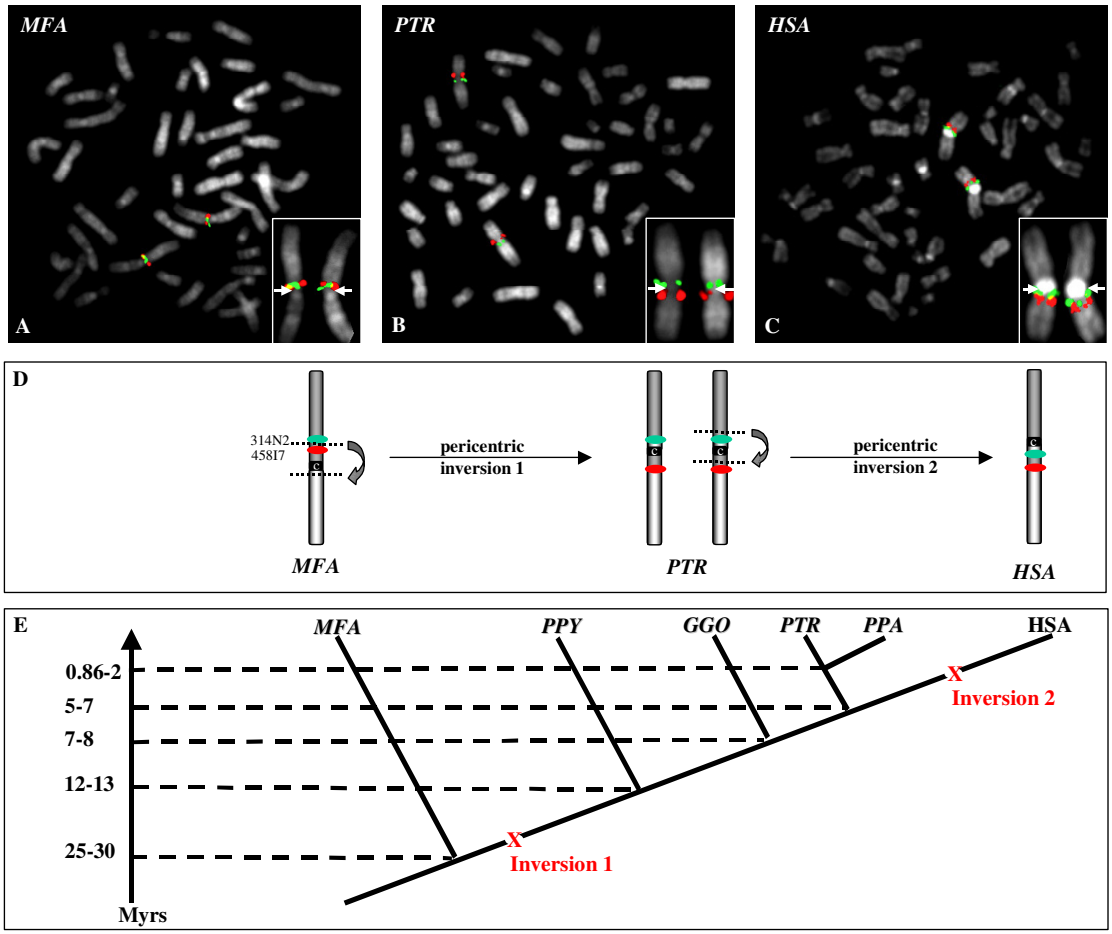

Figure 14: FISH with BACs RP11-314N2 (green) and RP11-458I7 (red) on the metaphase chromosomes of macaque (*MFA*) (**A**), chimpanzee (*PTR*) (**B**) and human (*HSA*) (**C**). As schematically indicated under (**D**) and according to the FISH pattern, both BACs are located on *MFA* 1p. The first pericentric inversion event transposed BAC RP11-458I7 to the q arm in *PTR* (**B**). The second pericentric inversion occurred specifically in the human lineage and included the region covered by BAC RP11-314N2, which maps to the q-arm in humans (**C**). Dotted lines in (**D**) indicate breakpoint regions for the two inversions that are located at different locations. (**E**) The time scale indicates the estimated times in millions of years (Myrs) for the separations of the different primate lineages from their common ancestor as well as the occurrence of the two pericentric inversions.

3.3. Verification of identity of the chimpanzee-specific pericentric inversions between *Pan troglodytes* and *Pan paniscus*

The information if the two chimpanzee species: *Pan troglodytes* (common chimpanzee) and *Pan paniscus* (bonobo) share exactly the same breakpoints of the chimpanzee-specific pericentric inversions is important for several reasons. Knowing if these large chromosomal rearrangements have become fixed before the separation of these two species will not only give an information about the upper time boundary of the inversion's occurrence, but also support the suggestion that these inversions have been important for the human-chimpanzee speciation. To verify these questions I investigated the identity of the breakpoint regions in common chimpanzee and bonobo at the cytogenetic and molecular level.

3.3.1. FISH with human and common chimpanzee breakpoint-spanning BAC/PACs

Initially the FISH analyses were performed with BAC/PAC clones isolated from the human genome library. As it is known from our earlier studies (Kehrer-Sawatzki et al. 2002, 2005a, b, c; Goidts et al. 2005; Szamalek et al. 2005) these clones span the inversion breakpoints of common chimpanzee (*PTR*) chromosomes homologous to human (*HSA*) chromosomes 4, 5, 9, 12, 16 and 17. The pericentric inversion of *PTR* 15 was not included in the current study because the exact breakpoint positions were not identified at the base-pair level (Locke at al. 2003a). For each breakpoint-spanning probe split hybridisation signals were observed in bonobo (*PPA*) chromosomes (Figure 15 and data not shown; Table 15). These findings imply that the inversion breakpoints in common chimpanzee as well as in bonobo map to the same genomic regions.

To verify this conclusion, additional FISH analyses with breakpoint-spanning BACs from the genome of the common chimpanzee were performed (Table 15). With *PTR* BACs that detect the inversion breakpoints in *HSA* 4p, 4q, 5p, 5q, 9q, 17p and 17q, I observed only single signals in the orthologous regions of *PPA*. In contrast, the breaks in the *PTR* chromosomes homologous to *HSA* 9p, 12p and 12q mapped to duplicated sequences as was previously shown by Kehrer-Sawatzki et al. (2005b, c). Therefore, the *PTR* breakpoint-spanning BACs from these regions yielded signals on both the p- and q-arms of the homologous chromosomes in *PPA*, *PTR* and *HSA* (data not shown). These results indicate that in the bonobo, as well as in common chimpanzee, duplicated sequences are present within the orthologous breakpoint regions. Finally, hybridisation of *PTR* BACs (RP43-

007E19 and RP43-001I03) spanning the inversion breakpoints of *PTR* 16, yielded split signals on the homologous chromosomes of bonobo but not of common chimpanzee. This is indicative of a bonobo-specific duplication in addition to the pericentric inversion (Figures 16 and 17).

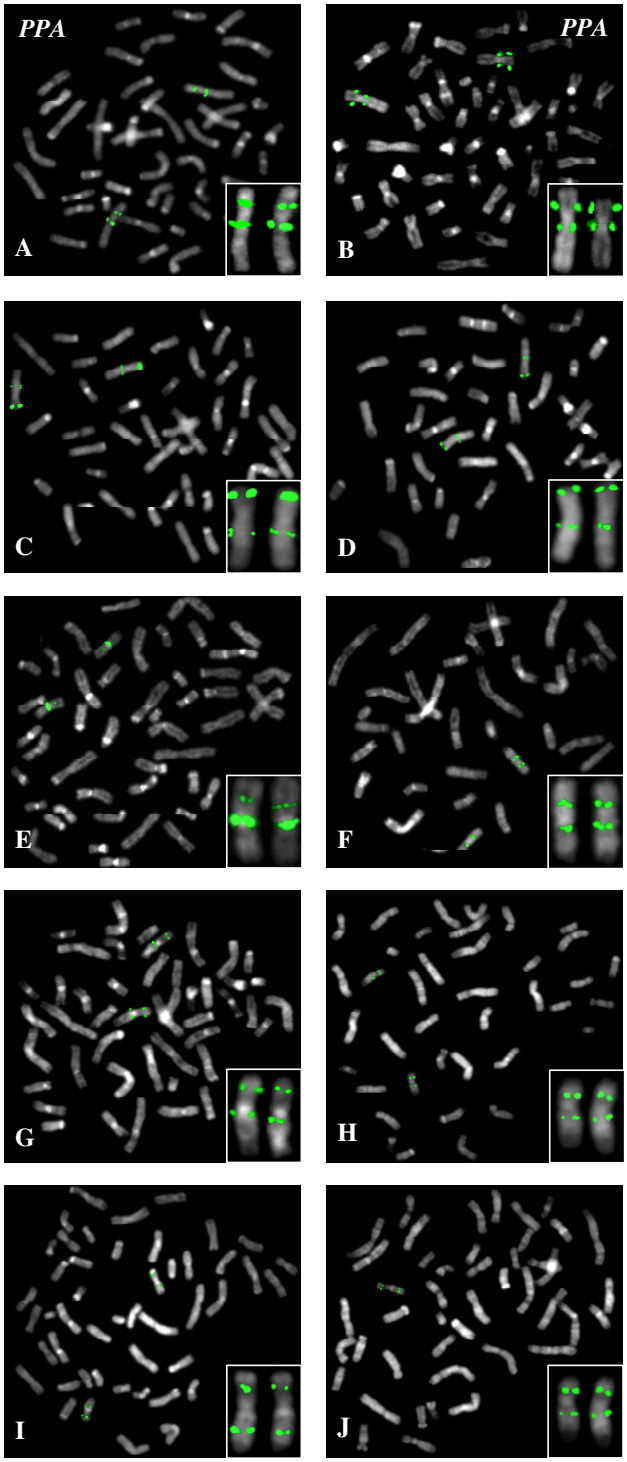

Figure 15: Bonobo (*PPA*) metaphases after hybridization with human breakpoint-spanning BAC/PACs: **(A)** RP11-779N22 from *HSA* 4p; **(B)** RP11-8N8 from *HSA* 4q; **(C)** RP11-35A11 from *HSA* 5p; **(D)** RP11-432G16 from *HSA* 5q; **(E)** RP11-259A5 from *HSA* 9p; **(F)** RP11-507D14 from *HSA* 9q; **(G)** RP11-80N2 from *HSA* 12p; **(H)** RP3-491B7 from *HSA* 12q; **(I)** RP1-179H24 from *HSA* 17p; **(J)** RP5-1029K10 from *HSA* 17q.

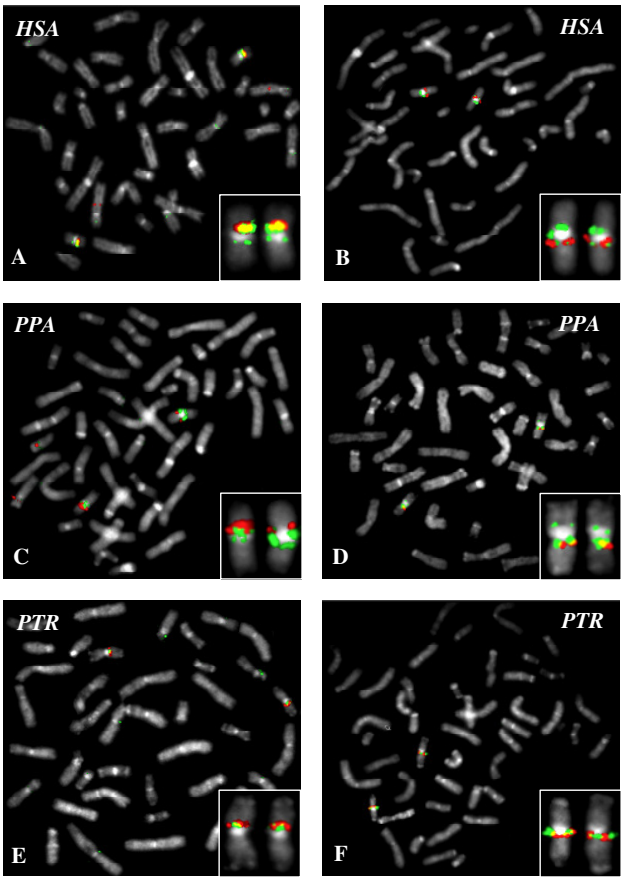

Figure 16: *HSA*, *PPA* and *PTR* metaphases after hybridization with BACs from chromosome homologous to *HSA* 16. **(A,C,E)** Co-hybridization of *PTR* p-arm breakpoint-spanning BAC RP43-007E19 (green) and *HSA* p-arm indicating BAC RP11-347N4 (red) to **(A)** *HSA*, **(C)** *PPA* and **(E)** *PTR* chromosomes. **(B,D,F)** Co-hybridization of *PTR* q-arm breakpoint-spanning BAC RP43-00I03 (green) and *HSA* q-arm indicating BAC RP11-46D6 (red) to (B) *HSA*, (D) *PPA* and (F) *PTR* chromosomes. The split signals of *PTR* BACs RP43-007E19 and RP43-00I03 on *PPA* 16 are indicative of a duplication that must have occurred after the separation of the *PTR* and *PPA* lineages.

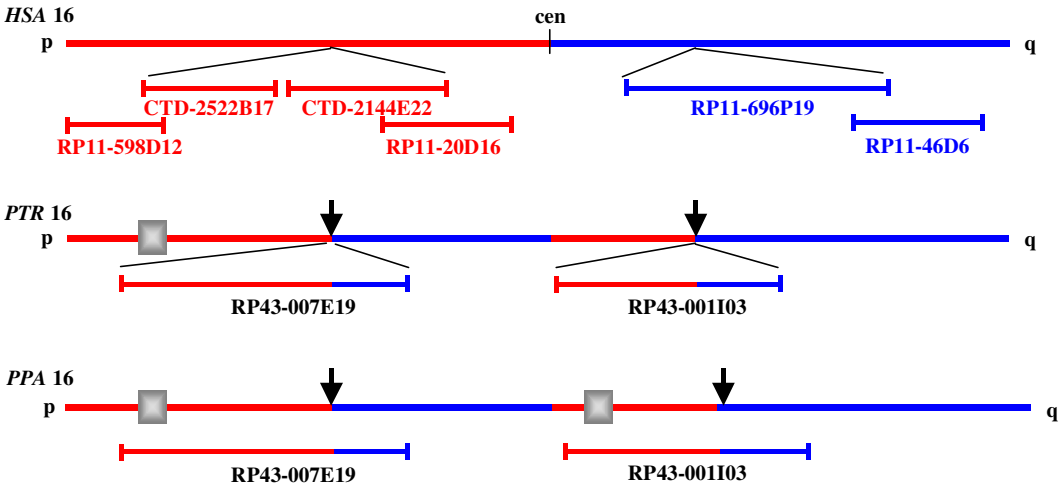

Figure 17: Illustration of the pericentric inversion distinguishing *HSA* 16 and the homologous *PTR* and *PPA* chromosomes. The breakpoint-spanning and neighbouring BACs are indicated by horizontal bars. Duplicated signals of *PTR* BACs RP43-007E19 and RP43-001I03 on bonobo chromosome 16 (*PPA* 16) are due to the duplication of a segment (grey rectangle) that occurred exclusively in the bonobo lineage.

Table 15: Summary of FISH results with human and common chimpanzee breakpoint-spanning BAC/PACs

Human chromosome	Breakpoint spanning clone	Chromosomal coordinates (bp) [a]	Localization in			Chimpanzee chromosome	Breakpoint spanning clone	Localization in		
			HSA	*PTR*	*PPA*			*HSA*	*PTR*	*PPA*
4p13	RP11-779N22 [b]	44,656,078-44,677,181	p	p + q	p + q	**4p**	RP43-59P20 [b]	p + q	p	p
4q21.23	RP11-8N8 [b]	86,174,177-86,360,148	q	p + q	p + q	**4q**	RP43-41D24 [b]	p + q	q	q
5p14.3-p15.1	RP11-35A11 [c]	18,434,022-18,604,727	p	p + q	p + q	**5p**	RP43-001C16 [c]	p + q	p	p
5q15	RP11-432G16 [c]	95,843,232-96,023,456	q	p + q	p + q	**5q**	RP43-023F12 [c]	p + q	q	q
9p12	RP11-259A5 [d]	42,421,610-42,550,187	p	p + q	p + q	**9p**	RP43-163C1 [d]	p + q	p + q	p + q
9q21.33	RP11-507D14 [d]	86,028,571-86,160,254	q	p + q	p + q	**9q**	RP43-156H15 [d]	p + q	q	q
12p12.2	RP11-80N2 [e]	20,833,482-21,009,087	p	p + q	p + q	**12q**	RP43-77C18 [e]	p + q	p + q	p + q
12q15	RP3-491B7 [e]	66,593,873-66,695,640	q	p + q	p + q	**12q**	RP43-135M19 [e]	p + q	p + q	p + q
16p11.2	CTD-2144E22 [f]	34,030,652-34,199,197	p	p + q	p + q	**16p**	RP43-007E19 [f]	p + q	p	p + q
16q11.2	RP11-696P19 [f]	44,943,303-45,066,095	q	p + q	p + q	**16q**	RP43-001I03 [f]	p + q	q	p + q
17p13.1	RP1-179H24 [g]	7,868,375-7,878,854	p	p + q	p + q	**17p**	RP43-141Q35 [g]	p + q	p	p
17q21.33	RP5-1029K10 [g]	44,918,026-45,104,642	q	p + q	p + q	**17q**	RP43-134L13 [g]	p + q	q	q

[a] Ensembl Genome Browser version 30.35c based on the NCBI Build35 assembly; Breakpoint spanning clones identified in the studies of: [b] Kehrer-Sawatzki et al. (2005a) ; [c] Szamalek et al. (2005); [d] Kehrer-Sawatzki et al. (2005b) ; [e] Kehrer-Sawatzki et al. (2005c), Shimada et al. (2005); [f] Goidts et al. (2005); [g] Kehrer-Sawatzki et al. (2002).

3.3.2. PCR analysis of the breakpoints

At last, to compare the breakpoints at the nucleotide sequence level, I performed PCR experiments that amplified a DNA sequence across the inversion breakpoints. Primer design utilized *PTR* sequences that flank the respective breakpoints. Identical PCR products of the expected lengths were obtained with genomic DNA from both *PTR* and *PPA*. However, this was not the case using human genomic DNA as a template (Figure 18). Next, the authenticity of the respective PCR products was verified by sequence analyses and the breakpoint-spanning sequences were submitted to the NCBI GenBank under accession numbers: DQ000175 - DQ000195, and DQ002477 - DQ002479.

Since the pericentric inversion breakpoints of the chimpanzee and bonobo chromosomes homologous to *HSA* 16 (Figure 18E) appear to be identical, the FISH pattern of the breakpoint-spanning BACs RP43-007E19 and RP43-001I03 may be explicable in terms of a duplication that occurred exclusively in the bonobo lineage.

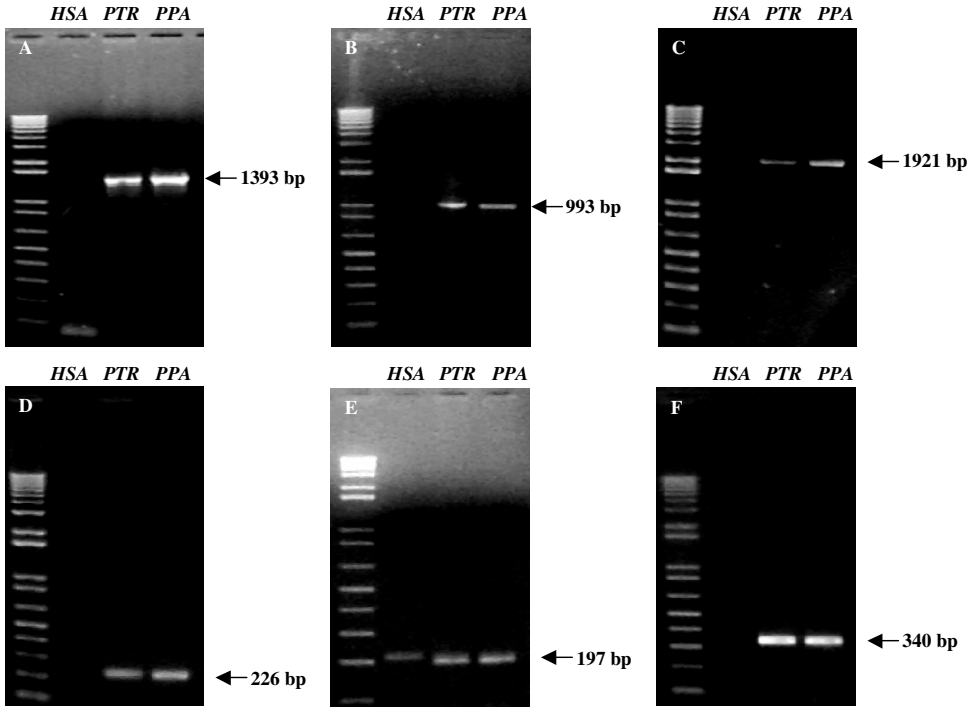

Figure 18: Junction-PCR analysis of the breakpoints. Breakpoint-spanning fragments of regions homologous to **(A)** *PTR* 4, **(B)** *PTR* 5, **(C)** *PTR* 9, **(D)** *PTR* 12, **(E)** *PTR* 16 and **(F)** *PTR* 17. Positive junction-PCR products were obtained in *PTR* and *PPA* but not in *HSA*. The only exception was the product amplified from the inversion junction of *PTR* 16, were a PCR product of atypical length was obtained using human genomic DNA as a template **(E)** but was subsequently found by sequence analysis to be a false positive. The sizes of the respective PCR products are indicated on the right.

3.4. Determination of the human-chimpanzee DNA divergence rate within inverted and collinear regions

From the first part of the results it becomes clear that the six chimpanzee-specific pericentric inversions are shared by two distinct chimpanzee species, *Pan troglodytes* and *Pan paniscus*. The experiment performed in the next step aimed in verification of the hypothesis that the pericentric inversions acted as potential drivers of the speciation event that permanently separated humans and chimpanzees. As it was explained already in the introduction, the new chromosomal speciation model requires that due to suppressed recombination between the inverted and non-inverted regions of the homologous chromosomes in heterozygous inversion carriers, the accumulation of genetic incompatibilities within these non-collinear regions is possible (Rieseberg 2001; Noor et al. 2001; Rieseberg and Livingstone 2003; Navarro and Barton 2003a, b).

In order to determine whether the divergence differs between inverted and non-inverted genomic regions of the common chimpanzee and human chromosomes, nine chromosomes were chosen for analyses. Chromosomes 4, 5, 12 and 17 were good candidates for this study, because the pericentric inversions, which characterize these chromosomes encompass large genomic segments. In addition to those chromosomes bearing chimpanzee-specific pericentric inversions, one of the two chromosomes that acquired a pericentric inversion in the human lineage, *HSA* 18, was also examined. As a reference, four chromosomes which are collinear between human and chimpanzee, *HSA* 3, 6, 7 and 10, were included. The analysed sequences were non-coding since they may be regarded as likely to be evolving under a neutralist model (Venter et al. 2001; Zhang et al. 2004), and may therefore be less prone to variation in substitution rates than the gene sequences (Shi et al. 2003). I compared the publicly available genomic sequences from the common chimpanzee (The Chimpanzee Sequencing and Analysis Consortium, 2005) with the completely sequenced DNA clones from the human genome.

After removal of masked repetitive sequences and gaps, the analysed dataset contained a total of 1,290,141 bp. Knowledge of the exact breakpoint locations enabled me to discriminate precisely between the inverted and non-inverted regions. The divergence rates of three classes of sequences were then compared, and these sequences were categorized with respect to their location: inverted regions (I), regions on rearranged chromosomes that were not within the inverted segments (NI) and collinear chromosomes without inversion (C). The results are summarized in Table 16. When the non-inverted regions were compared to inverted regions on rearranged chromosomes, the percentage

sequence divergence was found to be lower in inverted regions although this difference was not significant (P = 0.108, two-sided Mann-Whitney-U-test). The comparison of inverted regions (I) versus all collinear regions (NI+C) also revealed lower sequence divergence in inverted regions, albeit with borderline statistical significance (P = 0.054). No difference was however found in the degree of sequence divergence between rearranged chromosomes (I+NI) and collinear chromosomes (C) (P = 0.353) (Table 16).

To investigate whether the divergence rates were correlated with rates of recombination, I utilized data from the high-resolution recombination map of the human genome (Sex-averaged rates of recombination based on the deCODE genetic map (http://genome.ucsc.edu). A small positive correlation between recombination and divergence was noted (Table 16; Figure 19A, B) and this turned out to be significantly different from zero (Spearman Correlation Coefficient: R = 0.26, P = 0.05).

Table 16: Genomic DNA sequence divergence between human and chimpanzee

	Inverted regions (I)	**Non-inverted regions (NI)**	**Collinear chromosomes (C)**
Total number of compared regions	18	21	20
Total number of nucleotides (bp)	367,973	478,126	444,042
% Divergence	1.19 ± 0.21	1.34 ± 0.23	1.36 ± 0.27
Average recombination rate ± SD	1.10 ± 1.04	1.16 ± 0.80	1.20 ± 0.54

When the divergence rates of individual chromosomes with inversions were analysed separately, chromosome 5 turned out to a have significantly lower divergence rate within inverted regions as compared to non-inverted regions (P = 0.016, Exact Mann-Whitney-U-test; Table 17).

Table 17: Divergence between human and chimpanzee sequences on individual rearranged chromosomes

Rearranged chromosome	Non-inverted regions (NI)		Inverted regions (I)	
	Total number of compared regions	Average divergence %	Total number of compared regions	Average divergence %
4	7	1.32	5	1.14
5	5	1.40	4	1.06
12	3	1.19	3	1.17
17	3	1.42	3	1.28
18	3	1.34	3	1.37

From these findings it becomes evident that no remarkable difference in the human-chimpanzee divergence of non-coding sequences between rearranged and collinear chromosomes was observed. However, it must be emphasized that these analyses alone do not provide sufficient evidence to assert that the multiple pericentric inversions did not play any role in the human-chimpanzee speciation process.

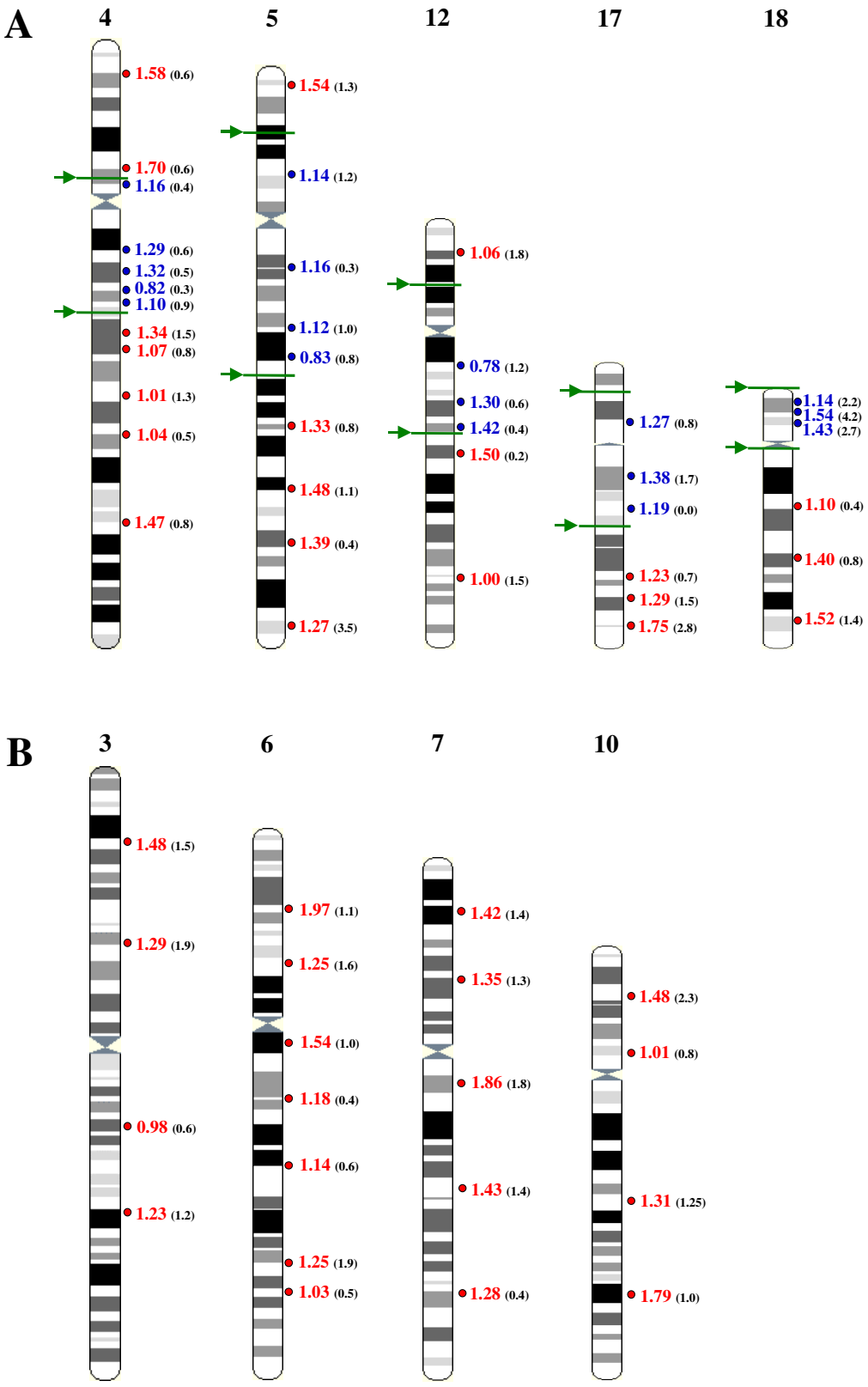

Figure 19: (A) Ideograms of rearranged (*HSA* 4, 5, 12, 17, 18) and **(B)** collinear chromosomes (*HSA* 3, 6, 7, 10) between human and chimpanzee. The divergence rates of the measure points within non-inverted and inverted regions are given in red and in blue, respectively. The recombination rates of the respective sequences are shown in parentheses, as extracted from the UCSC-database (http://genome.ucsc.edu/).

3.5. Genome-wide search for micro-inversions by gene order comparison between human and chimpanzee

In the two previous sections I concentrated on the analyses of large, microscopically visible pericentric inversions and their putative role in the human-chimpanzee speciation process. While cytogenetic comparisons of human and chimpanzee karyotypes have been effective in detecting these large-scale structural variations, the abundance of intermediate-sized rearrangements remained undetected.

3.5.1. Identification of the gene-containing micro-inversions

To identify potential gene-containing, sub-microscopic inversions, the relative positions of 11,518 human genes mapped in the human reference sequence (NCBI Build 35, Ensembl v28.35a.1) were compared with those of their orthologous counterparts in the chimpanzee genome using the EnsMART search tool. These analyses were based on the sequence of the chimpanzee genome, publicly accessible in Ensembl v28.1a.1 (The Chimpanzee Sequencing and Analysis Consortium 2005). All available human/chimpanzee orthologous gene pairs were extracted from this database and restrictive filtering (described in the Methods section) was applied. In this manner, a non-redundant set of 11,518 orthologous gene pairs, displaying a one-to-one correspondence between human and chimpanzee, was obtained. For the purposes of this study, a 'syntenic block' of genes was defined as a human chromosomal region for which a particular order of genes was also apparent in the chimpanzee, either with an identical or a reverse gene order. The human sequence was employed as a reference. In case when a human gene could not be matched to a chimpanzee gene in the list of genes ordered according to their base pair position, then a 'syntenic block' was deemed to have been interrupted. Potential inversions were identified in cases where a break in synteny was noted in addition to a reversal of gene order between humans and chimpanzees.

This comparison revealed a total of 81 such regions. Among these 81 inverted syntenic blocks, 11 were located within the well-characterized microscopically detectable pericentric inversions of *HSA* 4, 5, 15 and 18 and therefore served for their demarcation. The remaining 70 rearrangements were considered as potentially novel micro-inversions or small inverted transpositions of DNA. Whereas, according to the gene dataset, the majority (N= 62) of these putative micro-inversions encompass only two genes, eight contained more than three genes. Out of these eight, two actually demarcated one micro-

rearrangement on *HSA* 1p13.2-3 and therefore were further regarded as one unit. Furthermore, two additional micro-inversions containing more than three genes were detected on *HSA* 5 and 18, which were located within the known pericentric inversions and thus seen as collinear in the gene-order comparison. All inversions identified in this way are listed in Appendix 2. Since the chimpanzee sequence is an initial draft version, and hence still imperfect and incomplete, some of these putative inversions are likely to have been caused by errors in assembly.

3.5.2. Experimental validation: identification of inversion polymorphism and lineage-specificity

To confirm the authenticity of these micro-rearrangements, FISH and PCR were used to analyse the putative inversions, ranging from ~400 kb to ~18.8 Mb in size, which encompassed more than 3 known genes (Table 18). For seven out of the nine potential micro-inversions, BAC/PAC probes were identified so as to allow informative interphase or metaphase FISH analyses (the respective FISH probes used in these experiments are summarized in Table 9, see Materials and Methods). The FISH assays were performed with one probe located outside the region of the inversion (fixed probe) and two differentially labelled probes from within the inverted segment. The ~400 kb micro-rearrangement in the region orthologous to *HSA* 19p13.11 was analysed only with the use of PCR. Five out of the nine micro-inversions were confirmed by the FISH analyses, whereas three were excluded by FISH and/or PCR, at least for the individuals tested. For the remaining inversion on human chromosome 16, neither FISH nor PCR analyses were informative. Interestingly, heterozygosity was observed for four confirmed inversions (Table 18; Figures 20 and 21).

3.5.2.1. *HSA* 1p13.2-13.3

The ~3.5 Mb inversion in the region orthologous to *HSA* 1p13.2-13.3 was found by FISH to be heterozygous in one common chimpanzee, whereas another chimpanzee individual, orangutan, bonobo and human all displayed the homozygous non-inverted arrangement, which can be regarded as the ancestral one. Interestingly, according to the BLAST analyses and Segmental Duplication Database, inverted segmental duplications were detected at both breakpoints in chimpanzee but only at one break in human. It is

remarkable that some of these duplications are also associated with copy number differences in human (Table 19).

3.5.2.2. *HSA* 5p13.2-13.3

According to the gene order comparison the ~ 2.5 Mb inverted transposition on *PTR* 5 (orthologous to *HSA* 5p13.2-13.3) is located within the large 80 Mb pericentric inversion on this chromosome (Szamalek et al. 2005). The pattern of FISH with the combinations of diagnostic probes was however incompatible with the proposed rearrangement in all the investigated species. Also the analysis with *HSA* BAC RP11-94N8, which should have spanned the breakpoint according to the draft sequence of the chimpanzee genome, revealed no split signal in *PTR*. As this putative breakpoint does not map to site of segmental duplications, primers were designed by reference to the sequence of the putative breakpoint-spanning human BAC clone CTD-2241P19, which completely overlaps with BAC RP11-94N8 (the primers are listed in Table 11, in the Material and Methods chapter). PCR was performed using DNA from five humans, eleven unrelated common chimpanzees, one bonobo, one gorilla, and one orangutan, and the PCR products were obtained for all these DNA samples. This result suggests that no inversion is present, at least in the heterozygous form, in all four primate species.

3.5.2.3. *HSA* 7p22.1

Interestingly the ~800 kb inversion at 7p22.1 was found to be polymorphic in humans. The heterozygous and the homozygous non-inverted rearrangements were noted in a 2 and 3 individuals, respectively. This inversion was found to be surrounded by segmental duplication clusters of over 100 kb in length and more than 99% identity in both the human and the chimpanzee. Therefore, screening of a larger group of individuals by PCR techniques is impossible.

Table 18: Summary of the eight micro-inversions encompassing more than 3 genes together with the results of subsequent FISH and PCR investigations

Human chromosomal position	**Position and size of the inversion in humans** [a]	**No. of inverted genes**	**Experimental validation**	**FISH/PCR results**				
				HSA	***PTR***	***PPA***	***GGO***	***PPY***
1p13.2-13.3	109.9 – 113.4 ~3.5 Mb	50	Confirmed	not inverted	polymorphic	not inverted	not analysed	not inverted
5p13.2-13.3	32.0 – 34.5 ~2.5 Mb	17	Rejected	not inverted	not inverted	not inverted	not inverted	not inverted
7p22.1	5.8 – 6.6 ~800 kb	14	Confirmed	polymorphic	inverted	inverted	inverted	not analysed
7p13-14.1	39.3 – 43.7 ~4.4 Mb	16	Confirmed	not inverted	inverted	inverted	inverted	not inverted
17q21.31-24.1	41.6 – 60.4 ~18.8 Mb	180	Rejected	not inverted	not inverted	not analysed	not analysed	not analysed
18p11.21-11.22	10.6 – 12.4 ~1.8 Mb	7	Confirmed	not inverted	polymorphic	not inverted	not inverted	inverted
19p13.11 [b]	16.7 – 17.1 ~400 kb	4	Rejected	not inverted	not inverted	not inverted	not inverted	not inverted
19q13.12	41.5 – 42.5 ~1.0 Mb	16	Confirmed	not inverted	polymorphic	not inverted	not inverted	not inverted

[a] Approximately designated according to NCBI Build 35 assembly, (Ensembl v28.35a.1); [b] only PCR analysis were performed.

3.5.2.4. *HSA* 7p13-14.1

The ~4.4 Mb inversion at the region orthologous to *HSA* 7p13-14.1 was detected in two chimpanzee samples, in one bonobo, and one gorilla. In 11 humans, however, the non-inverted arrangement was observed. Interestingly, orangutan also displays the homozygous non-inverted form (Figure 21A). Thus this rearrangement can be regarded as specific for African great apes. The inversion at *HSA* 7p13-14.1 was further investigated by PCR with primers designed according to the sequence of human BAC RP11-64I2 and PAC RP5-1165K10 (the primers are listed in Table 11, in the Material and Methods chapter). These clones span the inversion breakpoints according to alignments with the scaffold sequences of the chimpanzee genome. The PCR was designed to detect the non-inverted rearrangement and should therefore not have been productive in the investigated African great apes, since FISH analyses indicated a homozygous inverted arrangement in these individuals (Table 18). However, contrary to expectation, the PCR reactions were indeed productive, implying either that the inversion breakpoints deduced from the draft sequence of the chimpanzee genome were incorrect or, more likely, that they were located in segmental duplications specific to these species.

3.5.2.5. *HSA* 17q21.31-24.1

The metaphase FISH analysis served to disprove the ~18.8 Mb inversion on *PTR* 17 (orthologous to *HSA* 17q21.31-24.1). The fact that the previously identified pericentric inversion of this chromosome (Kehrer-Sawatzki et al. 2002) is also not present in the correct orientation in the current draft sequence of the chimpanzee genome, suggests that there are several assembly errors in the respective regions of *PTR* 17.

3.5.2.6. *HSA* 18p11.21-11.22

The inversion of *PTR* 18 (orthologous to *HSA* 18p11.21-11.22) was heterozygous in one *PTR* individual, whereas the homozygous non-inverted arrangement was noted in humans, in the second chimpanzee, in bonobo, and in gorilla. The orangutan, however, displayed a homozygous inverted arrangement by comparison with humans. If the orangutan is considered as an outgroup, then the inverted form may be regarded as ancestral and the non-inverted arrangement would have been introduced into the genome of the common ancestor of humans, chimpanzees and gorillas.

3.5.2.7. *HSA* 19p13.11

The inverted transposition of 400 kb on *PTR* 19 (orthologous to *HSA* 19p13.11) was also not confirmed. Since FISH was not informative in the investigation of this small rearrangement, PCR was performed using primers designed according to the human sequence at the distal breakpoint (the primers are listed in Table 11, in the Material and Methods chapter). These PCR reactions were productive using DNA from five humans, 11 different chimpanzees, one bonobo, gorilla, and orangutan as templates, implying that the genomic configuration is the same in all these species (Figure 20, only one human sample is shown). To exclude possible heterozygosity of this event in the investigated chimpanzees, PCR was performed with primers designed according to the chimpanzee draft sequence, to amplify across one of the putative breakpoints, but none of the samples proved to be productive (data not shown).

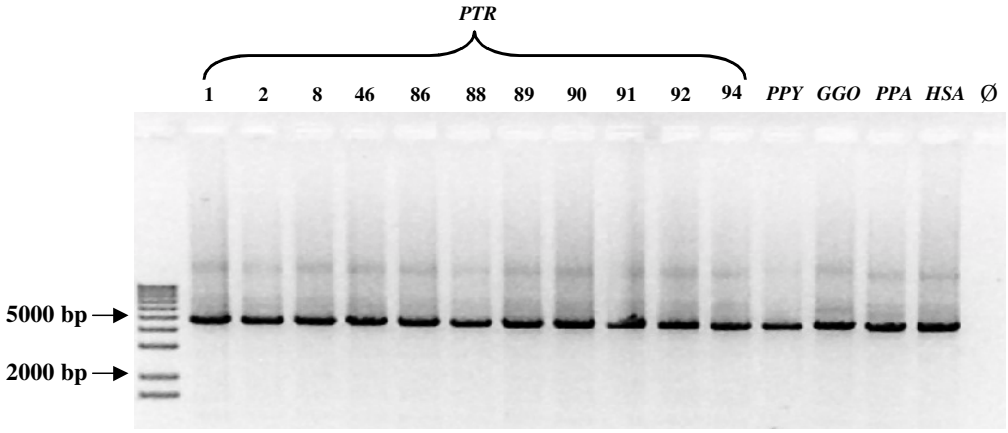

Figure 20: PCR investigation of the inverted transposition in *PTR* 19 orthologous to *HSA* 19p13.11.
The primers used were designed according to the human sequence and flank the putative distal breakpoint of the rearrangement. PCR was performed on genomic DNA of 11 common chimpanzees (*PTR*), one orangutan (*PPY*), one gorilla (*GGO*), one bonobo (PPA), and five humans (*HSA*) (only one human sample is shown). The authenticity of the PCR products was confirmed by sequence analysis. The 4786 bp PCR product was amplified from the non-inverted genomic arrangement found in humans and great apes, and serves to disprove the predicted micro-rearrangement. In the right lane, a negative control was loaded (Ø).

3.5.2.8. *HSA* 19q13.12

The inversion of *PTR* 19 (orthologous to *HSA* 19q13.12) occurred in heterozygous form in one investigated chimpanzee and in homozygous form in the second chimpanzee, but was not detected in the bonobo (*PPA*) (Figure 21B). Since the homozygous non-

inverted arrangement was observed also in humans, in orangutan and in gorilla, I surmise that this inversion occurred specifically in the chimpanzee lineage.

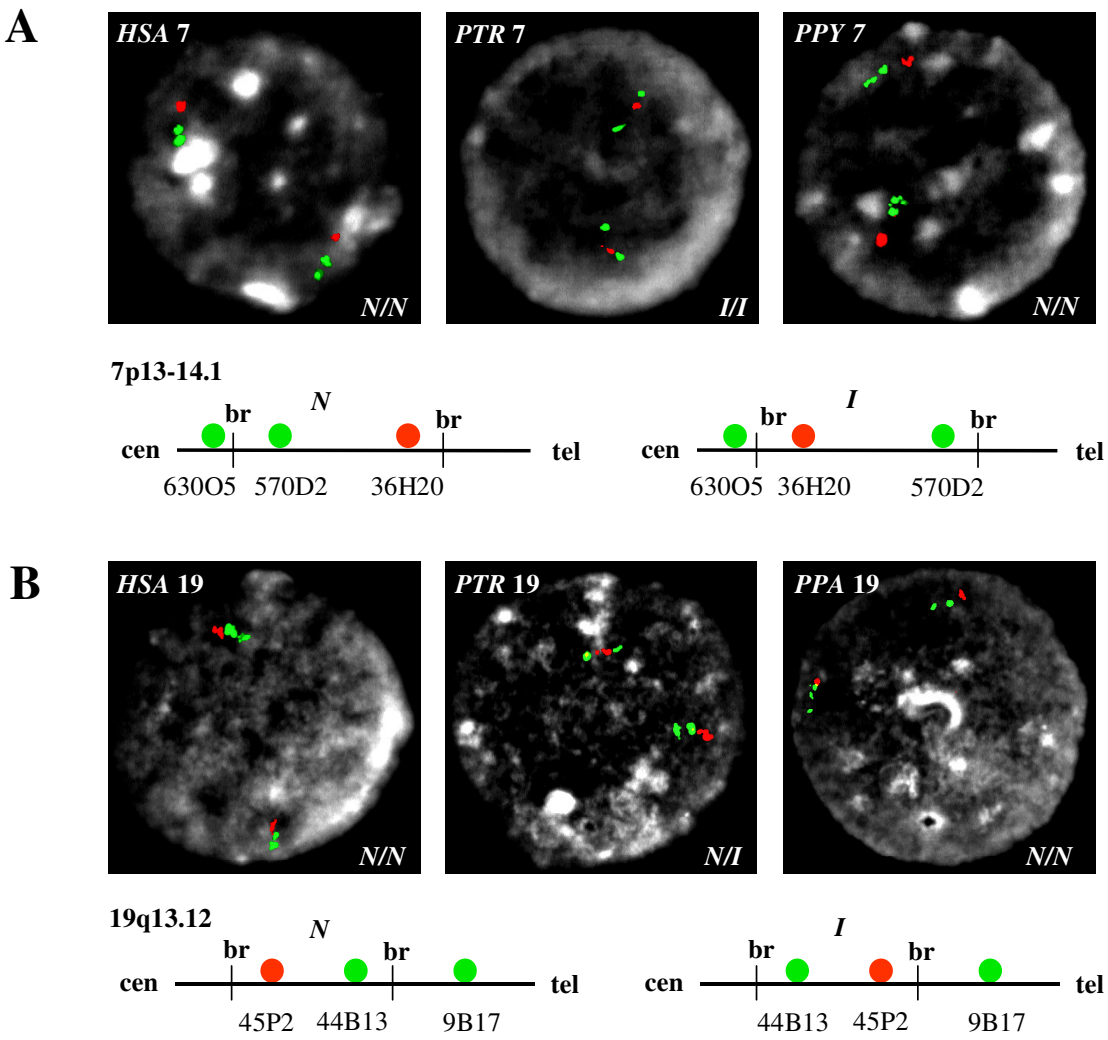

Figure 21: Interphase FISH performed to verify the putative inversions: **(A)** Three *HSA* BAC/PAC clones from *HSA* 7p13-14.1 were co-hybridised: RP11-630O5 (green) is located outside the inverted region whilst RP4-570D2 (green) and RP11-36H20 (red) lie within the inverted region. Signal configurations in human (*HSA*) and orangutan (*PPY*) indicate the ancestral, non-inverted state (N/N). The homozygous occurrence of the inversion (I/I) was observed in the common chimpanzee (*PTR*). Vertical lines in the schemes below the FISH figures indicate the locations of the inversion breakpoints (br) with respect to the position of the FISH probes.

(B) Combination of three *HSA* BACs from 19q13.12 employed as FISH probes. RP11-45P2 (red) and RP11-44B13 (green) are located within the inverted region whereas RP11-9B17 (green) maps outside the inverted region (fixed probe). Signal configurations in human (*HSA*) and bonobo (*PPA*) depict the non-inverted situation in the homozygous state (N/N). The common chimpanzee (*PTR*) is heterozygous for this inversion (N/I).

Table 19: Segmental duplications and genomic variants in the breakpoint regions of the five confirmed micro-inversions

Human chromosomal position	Segmental duplications (SD) [a, b]	Duplication Cluster (DC) ID [a, b]	Human genomic variants in the break regions [c]
1p13.2-3	SD at one break in *HSA*	DC0124, DC0126	Copy number difference [d, e, f]
	Inverted SDs at both breaks in *PTR*	DC0093, DC0096-DC0097, DC0099	
7p22.1	Inverted SD at both breaks in *HSA*	DC2117, DC2120	No
	Inverted SD at both breaks in *PTR*	DC1902-DC1903, DC1906-DC1907	
7p13-14.1	No large SD in *HSA*	No	No
	Inverted SDs at both breaks in *PTR*	DC1953, DC1963	
18p11.21-22	Inverted SD at both breaks in *HSA*	DC2902, DC2903, DC2912, DC2914	No
	SD at one break in *PTR*	DC3556-DC3557	
19q13.12	Inverted SD at both breaks in *HSA*	DC2657, DC2658, DC2661, DC2662	Copy number difference [g]
	Inverted SD at both breaks in *PTR*	DC2780, DC2782-DC2783	

[a] according to the Human Genome Segmental Duplication Database; http://projects.tcag.ca; [b] according to the Non-Human Segmental Duplication Database; http://projects.tcag.ca/xenodup/; [c] according to the Database of Genomic Variants; http://projects.tcag.ca/variation/; [d] Variation_0343, McLellan et al. (1997), [e] Variation_0385, Tuzun et al. (2005); [f] Variation_0013, Iafrate et al. (2004), [g] Variation_0512, Tuzun et al. (2005).

3.5.3. Gene content analyses within the five confirmed micro-inversions

The number of known genes within the inverted segments of the five confirmed inversions ranges from seven to 50 (Appendix 3). Functional analyses of these genes indicated the presence of clusters of genes with related functions. For example 14 genes included in the micro-inversion on *HSA* 19q13.12 belong to the zinc finger protein gene family members (*ZNF*) that are involved in the regulation of mRNA transcription. In case of other micro-inversions, no significant over-representation of functionally related genes was observed, although there were some tendencies for clustering of genes with similar functions. Thus, the micro-inversion on *HSA* 7p22.1 contains 14 genes that are involved in cell cycle control and cellular proliferation, in protein biosynthesis and phosphorylation, and proteolysis and protein trafficking. Of the 50 genes contained within the micro-inversion interval on *HSA* 1p13.2-13.3, several groups of genes were found to be involved in different forms of signal transduction, cation transport, metabolism and developmental processes. Interestingly, the proximal breakpoint of the micro-inversion on *HSA* 1p13.2-13.3 serves to disrupt a cluster of *GSTM 1-5* genes involved in xenobiotic detoxification. Similarly, the micro-inversion on *HSA* 19q13.12 disrupts a cluster of *ZNF* (zinc finger protein) genes.

3.5.4. Evolutionary analyses of the micro-inversion regions

Using the EnsMART management system of the Ensembl database, I compared the gene order in the breakpoint regions of the confirmed micro-inversions to the gene order in the dog, mouse and rat genomes. Interestingly, the distal breakpoint of the *PTR* 7p13-14.1 micro-inversion, which maps between the *BLVRA* and *MRPS24* genes, corresponds also to the syntenic breakpoints in the dog (*CFA* 18/16), mouse (*MMU* 13/2/11) and rat (*RNU* 17/3/14). Moreover, the region orthologous to the region of distal breakpoint in *HSA* 7p22.1 corresponds also to the breaks in synteny in the same three species (*CFA* 6/14, *MMU* 5/6 and *RNU* 12/4). In the vicinity of the proximal breakpoint on *HSA* 18p11.21-11.22, the synteny of the mouse chromosomes *MMU* 17/18 and the rat chromosomes *RNU* 9/18 is disrupted. These findings may be indicative of the general instability of the human-chimpanzee micro-inversion breakpoint regions.

4. Discussion

The results obtained in this work provide a contribution into the evolutionary research on the human and chimpanzee genomes. Starting with the detailed characterization of the already long known macro-inversions, through the testing of their putative role in the human-chimpanzee speciation, I came down to the new class of genomic divergence, the micro-inversions, that contribute to the genomic variability of these two species.

4.1. Comparison of the pericentric inversions of chromosomes 1 and 5

The nine large pericentric inversions involving homologs of human chromosomes 1, 4, 5, 9, 12, 15, 16, 17 and 18, and the telomeric fusion of two ancestral chromosomes into chromosome 2 in humans, represent the major karyotypic differences between humans and chimpanzees (Dutrillaux 1979; Yunis and Prakash 1982). The two pericentric inversions of chromosomes 1 and 5, which were subject of the molecular investigations in the presented thesis, are examples of the macro-rearrangements, which according to several categories fall into two distinct classes.

4.1.1. Lineage specificity

The first important difference is the lineage specificity of these inversions. The pericentric inversion of chromosome 1 occurred only in humans after the divergence from chimpanzees. This is concluded from the comparison with homologous chromosomes of gorilla and orangutan (Dutrillaux 1979, Yunis and Prakash 1982). However, when the Old World monkeys were considered as an outgroup, several authors found evidence of another pericentric inversion in the common ancestor of great apes, that occurred in the homolog of human chromosome 1 (Buckland 1992; Maresco et al. 1998; Ruiz-Herrera et al. 2002). One of the two asparagine transfer RNA gene (tRNAAsn) clusters was mapped to the pericentromeric region of the chromosome arm, which is homologous to human 1p, in two species of Old World monkeys, the macaque (*Macaca fascicularis*) and the African green monkey (*Ceropithicus aethiops*). The same gene cluster was however localized on the long arm of chromosome 1 in human and orangutan (Buckland, 1992). Maresco et al. (1998) by FISH mapping of the *FCGR1* gene family in primates, noticed the hybridisation signal on the chromosome arm, which is homologous to human 1p in baboon and rhesus monkey,

representing Old World monkeys. In chimpanzee and orangutan, however, the *FCGR1* probe hybridised to the long arm of chromosome 1 indicating an ancestral pericentric inversion. During human evolution, two duplications gave rise to the *FCGR1A, FCGR1B* and *FCGR1C* genes, which are separated by the second inversion that occurred in the human lineage. In the current study, by localization of two human BACs on the chromosomes of macaque, chimpanzee and human, I was able to confirm the breakpoint heterogeneity of these two inversions (Figure 14).

In contrast to the human-specific inversion of chromosome 1, the pericentric inversion of chromosome 5 occurred specifically in the chimpanzee lineage. Comparative FISH analyses performed by Müller and Wienberg (2001) indicated that human and orangutan share an ancestral form of chromosome 5 whereas chimpanzee and gorilla both possess the derived forms. While the chimpanzee has an pericentric inversion, chromosome 5 in the gorilla is involved in a t(5;17) translocation. One breakpoint of this reciprocal translocation was mapped to the region syntenic to human chromosome 5q13.3, between the HMG-CoA reductase gene (*HMGCR*) and RAS p21 protein activator 1 gene (*RASA1*) (Stankiewicz et al. 2001), that is approximately 10 Mb proximal to the q-arm inversion breakpoint on *PTR* 5.

4.1.2. Architecture of the breakpoint regions suggests the inversion mechanism

From the comparison of the genomic architectural features in the breakpoint regions of chromosomes 1 and 5 another striking difference between these two rearrangements becomes evident. While the chromosome 1 inversion occurred in the area rich with segmental duplications (Figures 10, 11 and 12), in the breakpoint regions of chromosome 5 no low copy repeats were detected (Figure 8).

FISH and sequence alignments allowed to narrow down the p-arm inversion breakpoint of *HSA* 1 to the region between positions 120,395,759 bp and 120,549,995 bp in the human genome assembly. The q-arm breakpoint was limited to the sequence fragment between 145,671,828 bp and 146,234,395 bp. Since both these areas contain extensive segmental duplications and un-sequenced gaps, it was not possible to determine the exact breakpoint positions at the base pair level. Already in several previous studies the putative breakpoints of this inversion were discussed but never characterized in detail. Notably, it turned out that the *FCGR1B* and *FCGR1C* genes, the chromosomal location of which was investigated by Maresco et al. (1998), are located within the breakpoint regions of the p- as well as the q-arm, as identified in my work. The breakpoint intervals

determined in later studies differ however from my findings (Weise et al. 2005; Newman et al. 2005; Chimpanzee Sequencing and Analysis Consortium 2005). The difficulties in the breakpoint mapping are due to the presence of segments with high sequence similarity between the putative breakpoint sites on both chromosome arms. Inferring from the FISH analyses and sequence alignments, I suggested the most plausible mechanism of the inversion occurrence (Figure 13). According to the proposed model, this human-specific inversion was mediated by the recombination between two ~91 kb human-specific duplications covered by BAC RP11-439A17 and PAC RP5-998N21. Such putative mechanism underlying the rearrangement is described in the literature as non-allelic homologous recombination (NAHR), which occurs between identical sequences located at both breakpoint sites.

Also in the breakpoint regions of five further pericentric inversions which distinguish human and chimpanzee karyotypes, those of the chromosomes homologues to *HSA* 9, 12, 15, 16 and 18, segmental duplications (low copy repeats) have been identified (Table 20). Whereas the proximal breakpoint of the chimpanzee-specific inversion of chromosome 9 maps to low copy repeats near to the centromere (Kehrer-Sawatzki et al. 2005b), in case of the *PTR* 12 inversion, segmental duplications of 86 kb and 23 kb were detected at both breakpoints (Kehrer-Sawatzki et al. 2005c). From comparison with human orthologous sequences it turned out that these two segmental duplications on *PTR* 12 are chimpanzee-specific, and most likely have mediated this chimpanzee-specific inversion. The q-arm breakpoint of the *PTR* 15 pericentric inversion also maps to a low copy repeat cluster. Therefore, this breakpoint was refined only approximately, to a 600 kb interval consisting of entirely duplicated material (Locke et al. 2003a). At the breaks of homology between *HSA* 16 and the homologous *PTR* chromosome, long stretches of HSATII satellites spanning 70 kb and 20 kb as well as high amounts of other repetitive elements were detected. Pairwise comparisons between these sequences indicated the presence of two inverted repeat clusters of ~23 kb on each chromosome arm, with 97% sequence similarity between them. It appears likely that these inverted repeats together with their flanking HSATII satellite sequences facilitated the inversion, although the breakpoints did not occur directly within the repeats (Goidts et al. 2005). Furthermore, the human-specific pericentric inversion of chromosome 18 was most probably mediated by an inverted segmental duplication of 19 kb (Dennehey et al. 2004; Goidts et al. 2004). This duplication encompasses the 3´ of the *ROCK1* gene and occurred in the human lineage since only one copy of this segment was found in the chimpanzee. Not only in the

breakpoint sites of these pericentric inversions but also in the ancestral chromosomal fusion site in *HSA* 2q13-2q14.1 large segmental duplications were reported (Fan et al. 2002).

However, the LCR-mediated NAHR does not explain all cases of evolutionary rearrangements. The breakpoints of the three remaining pericentric inversions, which distinguish the human and chimpanzee chromosomes 4, 5, and 17, do not cluster within LCRs, but lie in regions of abundant interspersed high-copy number repeats such as Alu, LINE and LTR elements (Table 20) (Kehrer-Sawatzki et al. 2002, 2005a; Szamalek et al. 2005). In absence of significant homology between the breakpoint regions, another mechanism described as Non-homologous End Joining (NHEJ) may play an important role in repair of double-strand breaks. Complexity analysis of the breakpoint regions of the chromosome 5 pericentric inversion revealed the presence of several repeats that could have served to mediate this rearrangement. The *PTR* 5 q-arm breakpoint was found to harbour stretches of alternating purine-pyrimidine $(RY)_5$-tracts as well as short direct repeats. There are evidences that the alternating (RY) tracts are capable of forming non-B DNA conformations that can in turn trigger genomic rearrangements (Bacolla et al. 2004; Bacolla and Wells 2004). Since topoisomerase II cleaves within RY repeats (Spitzner et al. 1990), these tracts could possibly have promoted the formation of double-strand breaks during non-homologous pairing of sequences on the long and short arms of a chromosome within an intra-chromosomal inversion loop. Also the breakpoints of the pericentric inversions between *HSA* 4, *HSA* 17 and their chimpanzee homologs are located in genomic segments where no segmental duplications were detected. These breakpoints map within regions rich in repetitive elements from the LINE, LTR and SINE-families. At the breakpoints in *HSA* 4p14 and *HSA* 4q21.3 imperfect inverted repeats of ~5 kb were identified, which could have mediated the intra-chromosomal pairing of the p- and q-arms (Kehrer-Sawatzki et al. 2005a). In case of the inversion breakpoints in *HSA* 17 (Kehrer-Sawatzki et al. 2002), at a position 60 bp distal to the q-arm breakpoint a 26 bp Alu-core sequence was found, which is often observed close to the sites of recombination (Rudiger et al. 1995).

Table 20: Summary of the molecular characterization of the inversions breaks

HSA (*PTR*)	Genes in inverted region [a]	Sequences mediating the inversion breaks		Human lineage	Chimpanzee lineage	Reference
		LCR	Short, high copy repeat			
1	140	+		+		**Szamalek et al. in preparation**
4	250		AT-rich repeats		+	Kehrer-Sawatzki et al. 2005a
5	390		AAAAA and TATAC repeats		+	**Szamalek et al. 2005**
9	180	+			+	Kehrer-Sawatzki et al. 2005b
12	500	+			+	Kehrer-Sawatzki et al. 2005c Shimada et al. 2005
15	130	+			+	Locke et al. 2003
16	15	+			+	Goidts et al. 2005
17	710		AluS element and GGGT repeat		+	Kehrer-Sawatzki et al. 2002
18	120	+		+		Goidts et al. 2004 Dennehey et al. 2004

[a] Approximate number of mapped genes, according to the NCBI Build 35.

Although the analyses of genomic features specific for the breakpoint regions may facilitate the development of models of the inversion mechanism, it must be remembered that these rearrangements occurred in ancestral DNA sequence contexts that do not longer exist, and that the DNA sequences that originally mediated the rearrangement might have decayed to some extent. This notwithstanding, the study of extant orthologues sequences represents the only available means to understand the causes and consequences of primate chromosome evolution and more specifically, to infer the nature of the DNA sequences responsible for genomic instability at these sites. The analyses of the chromosome breakpoints in primate evolution as well as in human genome disorders reveal that non-allelic homologous recombination (NAHR) is a major mechanism in recurrent rearrangements whereas non-homologous end-joining (NHEJ) may be responsible for many of the non-recurrent events. It must however be stressed that even if the LCR mediated rearrangements occur in the same region, it is very unlikely that two evolutionary independent events took place at exactly the same nucleotide positions. There are many

instances of chromosomal rearrangements arising recurrently at approximately the same location during chromosome evolution. For example, the pericentric inversions of chimpanzee and gorilla chromosomes homologous to *HSA* 16 have breakpoints in the same region as determined by identical breakpoint-spanning BACs. However, at the DNA sequence level, the breakpoints turned out to have occurred at different sites what provides evidence that these two rearrangements are evolutionary independent events (Goidts et al. 2005).

4.2. Determination of the time interval within which the chimpanzee lineage-specific inversions became introduced

The knowledge of the exact breakpoint positions of the six chimpanzee-specific pericentric inversions facilitated the molecular comparison of these rearrangements between the common chimpanzee (*Pan troglodytes*) and the bonobo (*Pan paniscus*). The results of FISH analyses with the breakpoint-spanning BAC/PAC clones from human and common chimpanzee genomes gave an evidence that bonobo possesses the same pericentric inversions of the chromosomes 4, 5, 9, 12, 16 and 17. In case of the seventh chimpanzee-specific inversion, which affects the chimpanzee homologue of *HSA* 15, such comparison was not possible since the DNA clone that unambiguously spans the inversion breakpoint were not identified (Locke et al. 2003a).

The PCR amplification and sequencing of the breakpoint spanning fragments further demonstrated that the breakpoints of these six chimpanzee-specific pericentric inversions occurred at the same nucleotide positions in these two extant species of chimpanzee. This indicates that these pericentric inversions are identical by descent rather than recurrent and predate the separation of the common chimpanzee and bonobo. Yoder and Yang (2000) estimated the time of the speciation between these sister species of chimpanzee at approximately 2 Mya. According to this and to our findings the pericentric inversions of chromosomes 4, 5, 9, 12, 16 and 17 should have become fixed between 2-7 Mya, that is after the divergence from the human lineage and before the separation of the two chimpanzee species. Recently Won and Hey (2005) suggested that the chimpanzee-bonobo speciation took place later, around 0.86 Mya (Figure 22). Nevertheless these large karyotypic differences can be regarded as early acquisitions during chimpanzee evolution and it is unlikely that they represent polymorphic variation among the subspecies of the common chimpanzee: *Pan troglodytes troglodytes*, *Pan troglodytes vellerosus*, *Pan troglodytes verus* and *Pan troglodytes schweinfurthii*.

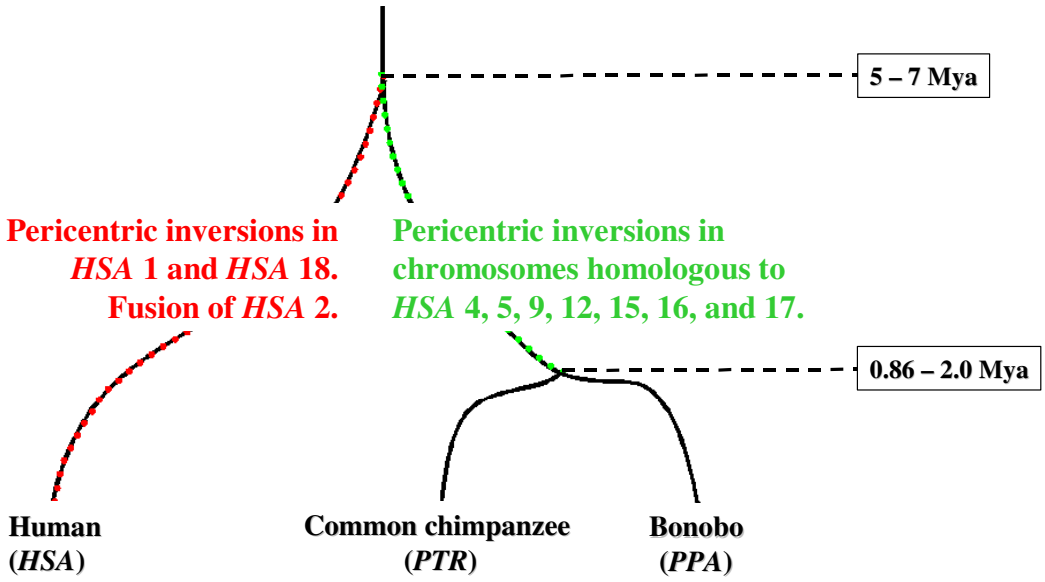

Figure 22: Time scale of occurrence of the nine pericentric inversions and one chromosomal fusion during the evolution of human and chimpanzee karyotypes. On the right site the approximate time of divergence between lineages is given.

4.3. Testing of the chromosomal speciation theory

Furthermore, according to the chromosomal speciation models, these evolutionary old and stable inversions could have been important for the human-chimpanzee speciation.

4.3.1. Evidence from the literature

It is evident that the karyotypes of species, which diverged under parapatric conditions (i.e., human, chimpanzee and gorilla), show more derived changes than do those species that diverged under allopatric conditions (i.e., orangutan and African great apes). This is also reminiscent of chromosomal speciation in *Drosophila*. Those *Drosophila* species that are sympatric with their sister species differ from them by inversions, in contrast to the closely related species that do not co-occur geographically (Noor et al. 2001). The new model of chromosomal speciation predicts that in heterozygotes for the inversions, the chromosomal rearrangements act as barriers to the gene flow by virtue of their ability to suppress recombination. This is then postulated to lead to accumulation of genetic incompatibilities, reproductive isolation and finally speciation. A number of studies

in fungi, plants and animals did indeed support this hypothesis (Rieseberg et al. 2001; Noor et al. 2001; Rieseberg 2001; Machado et al. 2002; Delneri et al. 2003).

The relevance of the pericentric inversions to the process of speciation between the ancestors of extant humans and chimpanzees has been tested by comparing the DNA divergence rates between rearranged and collinear chromosomes. Navarro and Barton (2003b) investigated the coding sequence divergence of 115 orthologous human and chimpanzee genes and observed that the rate of protein sequence divergence (K_A/K_S) was 120% higher for genes that were located on rearranged chromosomes than those located on collinear chromosomes. However, concerns have been raised about their analytical approach and hence the validity of their conclusions (Bowers 2003; Hey 2003; Lu et al. 2003; Vallender and Lahn 2004; Zhang et al. 2004). Vallender and Lahn (2004) examined over 7,000 genes using alternative methods of analysis and did not observe a higher rate of protein evolution on rearranged chromosomes. These authors considered that a combination of the small dataset employed, the non-uniform distribution of genes, and stochastic noise in K_S values may have been responsible for the elevated K_A/K_S ratios on rearranged chromosomes, as was observed by Navarro and Barton (2003b). Zhang et al. (2004) also failed to replicate the findings of Navarro and Barton (2003b). In addition to the analysis of a set of 1,126 human-chimpanzee orthologous gene pairs, they compared the alignments of non-coding sequences represented by chimpanzee BAC clones and the corresponding human genomic sequences, and did not find any sings of accelerated evolution of the sequences on rearranged chromosomes. In response to these contradictory findings, the group of Arcadi Navarro investigated genomic distribution of the gene expression divergence in humans and chimpanzees, and found a trend towards increased gene expression divergence of genes mapping to rearranged chromosomes (Marques-Bonet et al. 2004).

It is still very important to note that in the studies performed to date the authors have investigated differences in the genetic divergence of entire chromosomes, whether rearranged or collinear. However, inherent to the new chromosomal speciation model (Rieseberg 2001; Noor et al. 2001, Navarro and Barton 2003a) is the prediction that recombination suppression is confined to the inverted regions. Thus, when inversions are present in the heterozygous state, the inverted regions are unable to participate in homologous pairing during meiosis, resulting in the suppression of recombination. In the non-inverted yet still homologous segments of the rearranged chromosomes, meiotic pairing and recombination is however still possible. This has not invariably been taken into

account in the aforementioned studies (Navarro and Barton 2003b; Vallender and Lahn 2004; Zhang et al. 2004, Marques-Bonet et al. 2004). For example, of the 36 BACs analyzed by Zhang et al. (2004), only two were actually located within inverted regions.

4.3.2. Evidences from the own study

The detailed characterization of all pericentric inversions in the human/chimpanzee comparison, which was part of our previous studies (Kehrer-Sawatzki et al. 2002; Kehrer-Sawatzki et al. 2005a, b, c; Goidts et al. 2004; Goidts et al. 2005) as well as a subject of the presented thesis (Szamalek et al. 2005, Szamalek et al. 2005b, Szamalek et al. in preparation) served to demarcate the inversions accurately at the DNA sequence level. Taking advantage from the knowledge of the exact breakpoint positions I reinvestigated the human-chimpanzee sequence divergence within the rearranged and collinear chromosomal regions.

For these analyses chromosomes 4, 5, 12, 17 and 18 were chosen since the inverted segments encompass several tens of megabases of euchromatin so that multiple sequences within the inverted regions could be analyzed. The chromosomes homologues to *HSA* 1, 9, 15 and 16 were excluded from this approach since the breakpoints of their pericentric inversions map very close to the centromere or pericentric heterochromatin and often lie within the regions of extensive segmental duplications (Locke et al. 2003; Kehrer-Sawatzki et al. 2005c; Goidts et al. 2005; Szamalek et al. in preparation). Thus it was difficult to compare unique sequences between the inverted and non-inverted regions. In addition to the rearranged chromosomes, four chromosomes that are collinear between humans and chimpanzees (*HSA* 3, 6, 7 and 10) were also included. In this study, I avoided analysing of protein coding sequences that might be subject to either purifying selection associated with minimal divergence, or positive selection which could be characterized by very high divergence rates. Therefore, I included neutrally evolving non-coding sequences that were derived from the following regions: inverted regions (I), non-inverted regions on rearranged chromosomes (NI) and sequences from collinear chromosomes (C). Three categories of divergence rate comparisons were then performed: (1) I vs. NI on rearranged chromosomes; (2) I vs. NI + C; (3) I + NI vs. C.

The inverted regions (I) exhibited a lower degree of DNA sequence divergence as compared to both non-inverted regions (NI) from the same chromosome and collinear chromosomes (C). As these differences were only moderate, my results therefore tend to be in agreement with those of Zhang et al. (2004), who failed to observe any difference in

sequence divergence between rearranged and collinear chromosomes. Consistent with my findings, lower sequence divergence rates for the two analysed BACs derived from the inverted regions were also noted by Zhang et al. (2004). A lower neutral divergence rate in inverted regions was also apparent in the study of Navarro and Barton (2003b): when K_S values were compared between inverted and non-inverted regions, the latter were slightly higher, although not significantly. Under the new model of chromosomal speciation, chromosomal regions involved in the pericentric inversions would be expected to show greater sequence divergence than other regions. In practice, however, the tendency is for the reverse to be apparent. I observed a positive correlation with border-line significance between recombination rates in humans and the human-chimpanzee divergence. This is consistent with the findings of Hellmann et al. (2002) who observed that regions of the human genome with reduced recombination rates exhibit reduced divergence when compared with chimpanzees and baboons. The authors suggested that the variation in divergence levels likely reflects changes in the underlying mutation rate. This observation apply to non-coding sequences, where it is possible to assume that most mutations have no fitness effect. If most mutations were, instead, slightly deleterious, divergence levels would be expected to decrease in regions of high recombination, owing to purifying selection (Li 1987).

I conclude that if non-coding DNA sequences are compared between humans and chimpanzees, the divergence appears to be influenced more by the local recombination rate than by the history of species separation. In summary, the comparison of non-coding sequences of humans and common chimpanzees failed to reveal the patterns of genome evolution predicted by the recombination suppression model of chromosomal speciation as suggested by Navarro and Barton (2003b). Taken together with other recent data (Vallender and Lahn 2004; Zhang et al. 2004), these findings therefore suggest one of the following:

(a) The pericentric inversions that distinguish chimpanzees and humans occurred after the separation of the two lineages, and therefore had no influence on the speciation process itself.

(b) The pericentric inversions were present at the time of speciation, but had no significance for the process of speciation.

(c) The pericentric inversions did indeed play a significant role in the speciation process according to the recombination suppression model, but that the data currently available are inadequate to prove it.

(d) Not all pericentric inversions contributed to the speciation of early humans and chimpanzees. Separate analysis of the chromosomes 4, 5, 12, 17, and 18, revealed that the average divergence rates in inverted and collinear regions varied between individual chromosomes. For example chromosome 5 showed a significantly lower average divergence rate within the inverted region than within the collinear one, whereas for chromosome 18, which acquired the pericentric inversion in the human lineage, a even slightly reverse situation was observed.

(e) Each chromosomal rearrangement could have exerted a small incremental consequence such that the speciation would have been a gradual process rather than a single "event" as often described.

It is remarkable that the pericentric inversions distinguishing chimpanzee and human are not only numerous but their establishment has also been clustered in time. In the gorilla lineage also different inversions and only one translocation have been fixed in the gorilla lineage in approximately the same time period. This phenomenon of repeated and rapid occurrence of the same type of chromosomal rearrangement in a given lineage has been commonly reported in studies of chromosomal evolution and has been termed 'karyotypic orthoselection' (White 1973). The basis of karyotypic orthoselection is however not well understood and further work is needed to elucidate the biological role of the chromosomal inversions during human/chimpanzee speciation.

It can be also possible that in addition to these large pericentric inversions, which are recognizable at the level of classical cytogenetics, a number of smaller, sub-microscopic karyotypic differences did contribute to the process of human-chimpanzee speciation.

4.4. Micro-inversions as a newly discovered class of human-chimpanzee divergence

Whereas studies of karyotypically visible chromosomal rearrangements and their meaning for speciation reach early 70ies, in the last years there has been an expansion of research on more subtle differences which exist between humans and their closest relatives, the chimpanzees. Nowadays, is not possible to neglect this large amount of micro-differences between the genomes of these two species, in an attempt to understand the genetic basis of the human-chimpanzee divergence. Concentrating on the chromosomal inversions as a class of genetic variances which could have been important during evolution of recent primates, I performed a gene order comparison employing the human and chimpanzee reference DNA sequence. These analyses uncovered a total of 71 potential

micro-rearrangements, which ranged in size from a few hundred kilobases up to several megabases, but were still below the detection level of classical cytogenetic techniques such as G-banding analysis. Nine of these micro-inversions encompassed more than 3 genes included in our data set and eight of these could be analysed by FISH and/or PCR (Table 18). Five of the inversions were confirmed, indicating that this type of rearrangement occurs quite frequently. Such rearrangements are therefore likely to make a significant contribution to the genetic divergence between the genomes of humans and chimpanzees. Since the draft sequence of the chimpanzee genome still contains many unordered segments and gaps in alignment, it is not surprising that three of the potential inversions could not be confirmed in this study.

It is noteworthy that four out of the five confirmed micro-inversions were detected in the heterozygous state. The inversions involving *PTR* 1, 18 and 19, were polymorphic in chimpanzees, whereas the inversion in *HSA* 7p22.1 was found to be a polymorphic variant in humans. Interestingly, the micro-inversion of ~400 kb on *HSA* 16p11.2, which was identified in this study by gene order comparison between human and chimpanzee but not validated experimentally due to the small size and extensive segmental duplications at the putative inversion breakpoints, turned out to be also polymorphic in humans (Martin et al. 2004). During their mapping of the most extensively duplicated regions on *HSA* 16, Martin et al. (2004) identified two haplotype variants (EIFvar1 and EIFvar2) that differ by virtue of a 450 kb inversion and are surrounded by a 200 kb palindromic duplication. However, in order to determine unequivocally the polymorphism frequency of these inversions, it would be necessary to investigate additional individuals from each species.

Recently Tuzun et al. (2005) noticed that 55% of the structural variations detected in their study mapped to segmentally duplicated regions. Since SDs represent only 5% of the human genome (Bailey et al. 2002; She et al. 2004), this reflects a ten-fold enrichment of SDs at sites of structural variation as compared to unique sequences (Tuzun et al. 2005). As it was already discussed, the segmental duplications were also found in association with the pericentric inversions of chromosomes 1, 9, 12, 15, 16 and 18. It is therefore not surprising that the breakpoints of all five confirmed micro-inversions from my study map to sites of segmental duplications, which were present either at both breakpoints in human and chimpanzee (micro-inversions in 7p22.1 and 19q13.12) or at both breakpoints in only one species (micro-inversions in 1p13.2-3, 7p13-14.1 and 18p11.21-22; Table 19). Lineage-specific segmental duplications would appear to have contributed significantly to the divergence between human and chimpanzee as determined by genome-wide

comparisons of recent segmental duplications (Cheng et al. 2005) or interspecies array CGH (Locke et al. 2003a; Goidts et al. 2005b).

Interestingly the duplications identified at the breakpoints of micro-inversions in 1p13.2-3 and 19q13.12 are also associated with copy number differences in humans (McLellan et al. 1997; Iafrate et al. 2004; Tuzun et al. 2005), what provides a further indication of the inherent plasticity of these regions. There is a growing body of evidence in support of the view that the frequency of polymorphic small inversions as well as copy number variations (CNV; also known as copy number polymorphisms CNP) in the human genome is higher than previously suspected (Sebat et al. 2004; Iafrate et al. 2004; Tuzun et al. 2005; Sharp et al. 2005, Conrad et al. 2006, McCarroll et al. 2006). With the use of oligonucleotide microarray analysis of 20 individuals, Sebat et al. (2004) showed that on average, individuals differed by 11 CNPs. In a similar study, with the array based comparative genomic hybridisation of the DNA obtained from 55 unrelated individuals, Iafrate et al. (2004) identified 255 loci that contain genomic imbalances among unrelated individuals. Sharp et al. (2005) constructed a specific BAC microarray consisting of 130 potential rearrangement hotspots and identified further 73 CNPs, which were previously unreported. Applying a different technique by anchoring the fosmid paired-end sequences derived from a single individual on the human genome reference sequence, Tuzun et al. (2005) was able to detect intermediate-sized structural variants >8 kb in length. By these means in addition to 241 insertion and deletion events, 56 inversions were identified. Also the SNP genotype data from parent-offspring trios collected by the International HapMap Project are useful for the identification of structural polymorphisms (Conrad et al. 2006, McCarroll et al. 2006). By this way Conrad et al. (2006) estimated that typical individuals are hemizygous for roughly 30-50 deletions larger than 5 kb.

From clinical studies it is known that micro-rearrangements, for example micro-inversions, which are present in heterozygous state in single individuals can cause disturbances of chromosome pairing during meiosis and deleterious deletions in offspring (Osborne et al. 2001; Gimelli et al. 2003; Scherer et al. 2005; Visser et al. 2005). It therefore follows that any inversion heterozygosity that exists and segregates for a longer time in a population may reflect a selective advantage accruing to individual bearers. Such a polymorphism was recently described by Steffanson et al. (2005): a 900 kb inversion polymorphism at *HSA* 17q21.31 within a region that includes the microtubule-associated protein tau (*MAPT*) gene. Chromosomes with the inverted segment in different orientations represent two distinct lineages, termed H1 and H2, that have apparently been diverging for

as many as 3 million years with no evidence of their having recombined. Further, the H2 lineage would appear to have experienced positive selection in the Icelandic population since carrier females have had significantly more children than non-carriers (Steffanson et al. 2005).

The studies addressing the prevalence of micro-rearrangement polymorphisms in the chimpanzee genome are still limited. The micro-inversions on *PTR* 1, 18 and 19 identified in this study (Szamalek et al. 2006) suggest however that polymorphic micro-inversions might be quite frequent also in the chimpanzee genome. Comparison of the nine micro-inversions which were investigated experimentally in my study revealed that seven of them were also detected either in the study of Newman et al. (2005) or Feuk et al. (2005). In these two latter studies not all of the detected variations were validated with molecular methods and in case of those which were confirmed the discrepancies with respect to the lineage specificity and polymorphism were noticed. For example in the study of Feuk et al. (2005) no inversion polymorphisms in chimpanzee was observed, and the inversion in 7p13-14.1 was regarded as specific for the human lineage whereas according to my study the orangutan seems to have the same arrangement like humans. Therefore it becomes clear that these kind of studies are still preliminary and further experiments employing more individuals and additional techniques are required.

4.5. Inversions and their influence on neighbouring genes

It can be suspected that speciation would benefit from the direct effect which inversion breakpoints have on the genes, which are located next to the breaks. It is however remarkable that none of the characterized breakpoints of the large pericentric inversions, which distinguish humans and chimpanzees, do change the coding sequence of any known gene in a irreversible way. Only the breakpoint of the pericentric inversion in *PTR* 12p leads to disruption of the *SLCO1B3* gene (encoding solute carrier organic anion transporter family, member 1B3) in intron 11. However, owing to the duplication, the complete copy of this gene retains in *PTR* 12q and most probably, represents the functional gene in *PTR*.

It nevertheless remains possible that the inversions could have influenced the expression of genes residing in the vicinity of the breakpoints by disruption and relocation of regulatory elements. Indeed, two genes: the *USP17* gene encoding the ubiquitin specific peptidase 14 (tRNA-guanine transglycosylase) and the *ENOSF1* gene coding for enolase superfamily member 1, which are located proximal to the p-arm breakpoint of the human

lineage-specific pericentric inversion of *HSA* 18 were found to be up-regulated in human cortex compared with chimpanzee, as has been identified in microarray-based expression analyses (Caceres et al. 2003). Also the *NTRK2* gene encoding the neurotrophic tyrosine kinase, receptor, type 2 is up-regulated in the chimpanzee anterior cingulated cortex compared to human (Uddin et al. 2004). However this gene is located ~1 Mb from the breakpoint of the chimpanzee-specific pericentric inversion of chromosome 9. Thus, this expression difference may be not due to a position effect of the inversion breakpoin.

From the studies on micro-inversions (Newman et al. 2005; Feuk et al. 2005; Szamalek et al. 2006), it becomes apparent that they could have greater direct influence on the genic component in humans and chimpanzees. This is not surprising since the micro-inversions are found in much higher number than the large pericentric inversions. Newman et al. 2005 noticed that the inversion breakpoints identified by them overlap with the coding region from 14 RefSeq genes encoding e.g. chemokine protein, homeobox protein as well as several zinc fingers and hypothetical proteins. Also Feuk et al. (2005) showed that of the 1,576 computationally detected putative inversion regions, 151 overlap RefSeq genes in the human genome assembly. Moreover, 83 of these inversions were found to be contained within a gene and 29 regions have a breakpoint that intersects a gene. My study of the gene-containing micro-inversions revealed that the proximal breakpoint of the micro-inversion in *PTR* region orthologous to *HSA* 1p13.2-13.3 serves to disrupt a cluster of *GSTM 1-5* genes involved in xenobiotic detoxification, and the micro-inversion in *PTR* 19 corresponding to 19q13.12 in humans, disrupts a cluster of *ZNF* (zinc finger protein) genes. Since there is some evidence for co-regulated genes being syntenic in mammals (Cosseddu et al. 2004), disruption of this synteny caused by micro-inversions could be of functional significance. Moreover the *PMS2* gene, which is located in the breakpoint region of the polymorphic micro-inversion in *HSA* 7p22.1 is known to be involved in colorectal cancer in humans. Further studies would be however needed to understand the real impact of the inversion events on the surrounding genes.

4.6. Conclusions

4.6.1. General observations

From the detailed analyses of the macro- and micro-inversions, which contribute to the karyotype divergence between humans and chimpanzees it becomes evident that most of these events occurred in the chimpanzee lineage. Moreover, it was shown that the gorilla lineage underwent its own chromosomal speciation process, since several gorilla-specific peri- and paracentric inversions were identified (Yunis and Prakash 1982; Nickerson and Nelson 1998; Müller and Wienberg 2001; Goidts et al. 2005). Also the prevalence of observed copy number differences did not correlate with the time divergence of the respective primates from the common ancestral lineage. Among hominoids the highest CND frequency was observed in gorilla, although it diverged several million years later than orangutan from the common hominoid ancestor (Goidts et al. 2006). Micro-inversions and copy number differences have fundamentally contributed to the genomic divergence in primates. The micro-rearrangements form a continuous overlap from species diversity (polymorphism) to lineage-specific divergence.

4.6.2. Summary of the main findings

1) The human lineage-specific pericentric inversion of *HSA* 1 was most likely mediated by non-homologous recombination between two 91 kb segmental duplications, which arose specifically in humans after the divergence from chimpanzee.

2) Chromosomal breakpoints of the chimpanzee lineage-specific pericentric inversion of *PTR* 5 occurred within unique, unduplicated sequences, which harbour mutation-prone motifs known to play a role in recombination. These polypyrimidine, polypurine and alternating purine-pyrimidine tracts may have facilitated these rearrangement. Moreover, no gene sequence was found to be disrupted by the inversion breakpoints.

3) Two species of chimpanzee, *Pan troglodytes* (common chimpanzee) and *Pan paniscus* (bonobo) share exactly the same breakpoints of the six chimpanzee lineage-specific pericentric inversions, which therefore predate the separation of these two species 0.86 - 2 Mya.

4) The signatures of chromosomal speciation driven by the pericentric inversions is not detectable in non-coding sequences of human and chimpanzee. The lower DNA divergence within inverted regions of rearranged chromosomes is probably due to a positive correlation with present-day recombination rates in humans.

5) The newly detected class of rearrangements, the micro-inversions, show ongoing dynamics and contribute to the genomic variability of humans and chimpanzees. These multiple events are of special importance since they have a potential to influence gene coding sequences.

6) The detailed characterization of the macro- and micro-inversions performed in this study can be a basis for further investigations of the genetic backgrounds of the phenotypic differences between humans and their closest relatives, the chimpanzees.

5. Summary

During the last 5 - 7 million years, since humans and chimpanzees shared a common ancestor, numerous cytogenetic and molecular changes have accumulated in their genomes. Regardless of the high similarity between the DNA sequence of human and chimpanzee (98.77%, when the single-nucleotide substitutions are considered), there are still many macro- and micro-differences, which remain to be clarified. One chromosomal fusion, nine large pericentric inversions and numerous micro-rearrangements such as insertions, deletions, duplications and inversions raise the total DNA sequence divergence between human and chimpanzee to near 5%.

In the presented thesis I provided detailed cytogenetic and molecular characterisation of the human-specific pericentric inversion of chromosome 1 and the chimpanzee lineage-specific pericentric inversion of chromosome 5. From these analyses results that the *HSA* 1 inversion is associated with emergence of human-specific segmental duplications, which most likely mediated this rearrangement. Conversely, the breakpoints of the *PTR* 5 inversion occurred within gene poor regions where no low copy repeats were detected, but several mutation-prone motifs were identified, which are known to play a role in recombination.

Subsequently, I confirmed that the two chimpanzee species, *Pan troglodytes* and *Pan paniscus*, share exactly the same breakpoints of the six chimpanzee-specific pericentric inversions, which were precisely characterized in this and previous studies. This finding indicates that these inversions must have occurred prior to separation of common chimpanzee and bonobo lineages, 0.86 - 2 Mya.

I also examined whether, according to the new chromosomal speciation theory, the effect of suppressed recombination within regions, which are inverted between human and chimpanzee, is still detectable in non-coding genomic sequences of these two species. Contrary to the chromosomal speciation hypothesis, which presume that the genetic incompatibilities should have accumulated within the rearranged regions, I found lower sequence divergence in inverted areas, which is in positive correlation with present-day recombination rates in humans.

In the second part of my thesis, a genome-wide search for micro-inversions was performed by gene order comparison between human and chimpanzee. A total of 71 putative micro-inversions were identified, from which the nine largest were verified. Four out of five experimentally confirmed events were found to be polymorphic in chimpanzees

or in humans. Thus it can be concluded, that the micro-inversion differences have fundamentally contributed to the genomic divergence in primates. They form a continuous overlap from species diversity (polymorphism) to lineage-specific divergence. Moreover, these multiple events have a potential to influence gene coding sequences.

6. References

Abeysinghe SS, Chuzhanova N, Krawczak M, Ball EV, Cooper DN: Translocation and gross deletion breakpoints in human inherited disease and cancer I: Nucleotide composition and recombination-associated motifs. Hum Mutat 22: 229-244 (2003)

Armengol L, Pujana MA, Cheung J, Scherer SW, 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, Jaworski A, Larson JE, Jakupciak JP, Chuzhanova N, Abeysinghe SS, O'Connell CD, Cooper DN, Wells RD: Breakpoints of gross deletions coincide with non-B DNA conformations. Proc Natl Acad Sci USA 101: 14162-14167 (2004)

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

Bailey JA, Gu Z, Clark RA, Reinert K, Samonte RV, Schwartz S, Adams MD, Myers EW, Li PW, Eichler EE: Recent segmental duplications in the human genome. Science 297: 1003-1007 (2002)

Bailey JA, Baertsch R, Kent WJ, Haussler D, Eichler EE: Hotspots of mammalian chromosomal evolution. Genome Biol 5: R23 (2004)

Barbi G, Spaich Ch, Adolph S, Rossier E, Kehrer-Sawatzki H: Supernumerary der(1) marker chromosome derived from a ring chromosome 1 which has retained the original centromere and euchromatin from 1q21.1→q21.3 with substantial loss of 1q12 heterochromatin in a female with dysmorphic features and psychomotoric developmental delay. Am J Med Genet A 132: 419-424 (2005)

Bowers EJ: Chromosomal speciation. Science 301: 764-765 (2003)

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

Brunet M, Guy F, Pilbeam D, Mackaye HT, Likius A, Ahounta D, Beauvilain A, Blondel C, Bocherens H, Boisserie JR, De Bonis L, Coppens Y, Dejax J, Denys C, Duringer P, Eisenmann V, Fanone G, Fronty P, Geraads D, Lehmann T, Lihoreau F, Louchart A, Mahamat A, Merceron G, Mouchelin G, Otero O, Pelaez Campomanes P, Ponce De Leon M, Rage JC, Sapanet M, Schuster M, Sudre J, Tassy P, Valentin X, Vignaud P, Viriot L, Zazzo A, Zollikofer C: A new hominid from the Upper Miocene of Chad, Central Africa. Nature 418: 145-51 (2002)

Buckland RA: A primate transfer RNA gene cluster and the evolution of human chromosome 1. Cytogenet Cell Genet 61: 1-4 (1992)

Caceres M, Lachuer J, Zapala MA, Redmond JC, Kudo L, Geschwind DH, Lockhart DJ, Preuss TM, Barlow C: Elevated gene expression levels distinguish human from non-human primate brains. Proc Natl Acad Sci U S A 100: 13030-13035 (2003)

Carroll SB: Genetics and the making of Homo sapiens. Nature 422:849-857 (2003)

Chen FC, Li WH: 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 RK, Paabo S, Rocchi M, Eichler EE: Genome-wide comparison of recent chimpanzee and human segmental duplications. Nature 437: 88-93 (2005)

Choudhuri JV, Schleiermacher C, Kurtz S, Giegerich R: GenAlyzer: interactive visualization of sequence similarities between entire genomes. Bioinformatics 20: 1964-1965 (2004)

Conrad DF, Andrews TD, Carter NP, Hurles ME, Pritchard JK: A high-resolution survey of deletion polymorphism in the human genome. Nat Genet 38: 75-81 (2006)

Cosseddu GM, Perez-Enciso M, Fellous M, Vaiman D: Interspecific chromosome-wide transcription profiles reveal the existence of mammalian-specific and species-specific chromosome domains. J Mol Evol 59: 317-328 (2004)

Courseaux A, Richard F, Grosgeorge J, Ortola C, Viale A, Turc-Carel C, Dutrillaux B, Gaudray P, Nahon JL: Segmental duplications in euchromatic regions of human chromosome 5: a source of evolutionary instability and transcriptional innovation. Genome Res 13: 369-381 (2003)

Cruts M, Rademakers R, Gijselinck I, van der Zee J, Dermaut B, de Pooter T, de Rijk P, Del-Favero J, van Broeckhoven C: Genomic architecture of human 17q21 linked to frontotemporal dementia uncovers a highly homologous family of low-copy repeats in the tau region. Hum Mol Genet 14: 1753-1762 (2005)

Darwin C: The descent of man, and selection in relation to sex. D Appleton and Company, New York (1871)

Delneri D, Colson I, Grammenoudi S, Roberts IN, Louis EJ, Oliver SG: Engineering evolution to study speciation in yeasts. Nature 422: 68-72 (2003)

Dennehey BK, Gutches DG, McConkey EH, Krauter KS: Inversion, duplication, and changes in gene context are associated with human chromosome 18 evolution. Genomics 83: 493-501 (2004)

Dutrillaux B, Rethore MO, Lejeune J: Analysis of the karyotype of *Pan paniscus*. Comparison with other *Pongidae* and man. Humangenetik 28: 113-119 (1975)

Dutrillaux B: Chromosomal evolution in primates: tentative phylogeny from *Microcebus murinus* (Prosimian) to man. Hum Genet 48: 251-314 (1979)

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

Evans PD, Gilbert SL, Mekel-Bobrov N, Vallender EJ, Anderson JR, Vaez-Azizi LM, Tishkoff SA, Hudson RR, Lahn BT: Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans. Science 309: 1717-1720 (2005)

Faas BH, Mieloo H, Van Es-Van Gaal JW, Van Ravenswaaij C: A new case of dup(1)(q21.2q12) in an individual with mild mental retardation. Genet Couns 14: 407-411 (2003)

Fan Y, Linardopoulou E, Friedman C, Williams E, Trask BJ: 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 JR, Tang T, Carson AR, Li M, Rao G, Khaja R, Scherer SW: Discovery of human inversion polymorphisms by comparative analysis of human and chimpanzee DNA sequence assemblies. PLoS Genet 1: e56 (2005)

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 JM: Lineage-specific gene duplication and loss in human and great ape evolution. PLoS Biol 2: E207 (2004)

Frazer KA, Chen X, Hinds DA, Pant PV, Patil N, Cox DR: Genomic DNA insertions and deletions occur frequently between humans and nonhuman primates. Genome Res 13: 341–346 (2003)

Gagneux P, Gonder MK, Goldberg TL, Morin PA: Gene flow in wild chimpanzee populations: what genetic data tell us about chimpanzee movement over space and time. Philos Trans R Soc Lond B 356: 889-897 (2001)

Gagneux P, Varki A: Genetic differences between humans and great apes. Mol Phylogenet Evol 18: 2-13 (2001)

Gagneux P: The genus *Pan*: population genetics of an endangered outgroup. Trends Genet 18: 327-330 (2002)

Gimelli G, Pujana MA, Patricelli MG, Russo S, Giardino D, Larizza L, Cheung J, Armengol L, Schinzel A, Estivill X, Zuffardi O: Genomic inversions of human chromosome 15q11-q13 in mothers of Angelman syndrome patients with class II (BP2/3) deletions. Hum Mol Genet 12: 849-858 (2003)

Goidts V, Szamalek JM, 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. 115:116-122 (2004)

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

Goidts V, Armengol L, Schempp W, Conroy J, Nowak N, Müller S, Cooper DN, Estivill X, Enard W, Szamalek JM, Hameister H, Kehrer-Sawatzki H: Identification of large-scale human-specific copy number variations by inter-species array comparative genomic hybridization. 119:185-198 (2006)

Gonder MK, Oates JF, Disotell TR, Forstner MR, Morales JC, Melnick DJ: A new west African chimpanzee subspecies? Nature 388: 337 (1997)

Goodman M, Tagle DA, Fitch DH, Bailey W, Czelusniak J, Koop BF, Benson P, Slightom JL: Primate evolution at the DNA level and a classification of hominoids. J Mol Evol 30: 260-266 (1990)

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

Goodman M, Grossman LI, Wildman DE: Moving primate genomics beyond the chimpanzee genome. Trends Genet 21: 511-517 (2005)

Gusev VD, Nemytikova LA, Chuzhanova NA: On the complexity measures of genetic sequences. Bioinformatics 15: 994-999 (1999)

Haig D: The complex history of distal human chromosome 1q. Genomics 86: 767-770 (2005)

Hedges SB, Kumar S: Genomic clocks and evolutionary timescales. Trends Genet 19: 200-206 (2003)

Hellmann I, Ebersberger I, Ptak SE, Paabo S, Przeworski M: A neutral explanation for the correlation of diversity with recombination rates in humans. Am J Hum Genet 72: 1527-1535 (2003)

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

Hughes JF, Skaletsky H, Pyntikova T, Minx PJ, Graves T, Rozen S, Wilson RK, Page DC: Conservation of Y-linked genes during human evolution revealed by comparative sequencing in chimpanzee. Nature 437: 100-103 (2005)

Huxley TH: Evidence as to man's place in nature. Williams and Norgate, London (1863)

Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C: Detection of large-scale variation in the human genome. Nat Genet 36: 949-951 (2004)

IJdo JW, Baldini A, Ward DC, Reeders ST, Wells RA: Origin of human chromosome 2 an ancestral telomere-telomere fusion. Proc Natl Acad Sci USA 88: 9051-9055 (1991)

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

Kaessmann H, Wiebe V, Weiss G, Paabo S: Great ape DNA sequences reveal a reduced diversity and an expansion in humans. Nat Genet 27: 155-156 (2001)

Karaman MW, Houck ML, Chemnick LG, Nagpal S, Chawannakul D, Sudano D, Pike BL, Ho VV, Ryder OA, Hacia JG: Comparative analysis of gene-expression patterns in human and African great ape cultured fibroblasts. Genome Res 13: 1619-1630 (2003)

Kasprzyk A, Keefe D, Smedley D, London D, Spooner W, Melsopp C, Hammond M, Rocca-Serra P, Cox T, Birney E: EnsMart: a generic system for fast and flexible access to biological data. Genome Res 14: 160-169 (2004)

Kehrer-Sawatzki H, Schreiner B, Tanzer 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, Chuzhanova N, Goidts V, Szamalek JM, Tanzer S, Muller S, Platzer M, Cooper DN, 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 (2005a)

Kehrer-Sawatzki H, Szamalek JM, 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 (2005b)

Kehrer-Sawatzki H, Sandig CA, 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 (2005c)

Khaitovich P, Muetzel B, She X, Lachmann M, Hellmann I, Dietzsch J, Steigele S, Do HH, Weiss G, Enard W, Heissig F, Arendt T, Nieselt-Struwe K, Eichler EE, Paabo S: Regional patterns of gene expression in human and chimpanzee brains. Genome Res 14: 1462-1473 (2004)

King MC, Wilson AC: Evolution at two levels in humans and chimpanzees. Science 188: 107-116 (1975)

King M: Species evolution: The role of chromosome change. Cambridge University Press, Cambridge (1993)

Koop BF, Tagle DA, Goodman M, Slightom JL: A molecular view of primate phylogeny and important systematic and evolutionary questions. Mol Biol Evol 6: 580-612 (1989)

Kortlandt A: New perspectives on ape and human evolution. University of Amsterdam, Amsterdam (1972)

Kuroki Y, Toyoda A, Noguchi H, Taylor TD, Itoh T, Kim DS, Kim DW, Choi SH, Kim IC, Choi HH, Kim YS, Satta Y, Saitou N, Yamada T, Morishita S, Hattori M, Sakaki Y, Park HS, Fujiyama A: Corrigendum: Comparative analysis of chimpanzee and human Y chromosomes unveils complex evolutionary pathway. Nat Genet 38: 389 (2006)

Lande R: The fixation of chromosomal rearrangements in a subdivided population with local extinction and colonization. Heredity 54: 323-332 (1985)

Li WH: Models of nearly neutral mutations with particular implications for non-random usage of synonymous codons. J Mol Evol 24: 337-345 (1987)

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

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

Lu J, Li WH, Wu CI: Comment on "Chromosomal speciation and molecular divergence-accelerated evolution in rearranged chromosomes". Science 302: 988 (2003)

Machado CA, Kliman RM, Markert JA, Hey J: Inferring the history of speciation from multilocus DNA sequence data: The case of Drosophila pseudoobscura and close relatives. Mol Biol Evol 19: 472-488 (2002)

Maresco DL, Blue LE, Culley LL, Kimberly RP, Anderson CL, Theil KS: 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 TM, Thomas JW, 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 LA, Terry A, Prabhakar S, She X, Xie G, Hellsten U, Chan YM, Altherr M, Couronne O, Aerts A, Bajorek E, Black S, Blumer H, Branscomb E, Brown NC, Bruno WJ, Buckingham JM, Callen DF, Campbell CS, Campbell ML, Campbell EW, Caoile C, Challacombe JF, Chasteen LA, Chertkov O, Chi HC, Christensen M, Clark LM, Cohn JD, Denys M, Detter JC, Dickson M, Dimitrijevic-Bussod M, Escobar J, Fawcett JJ, Flowers D, Fotopulos D, Glavina T, Gomez M, Gonzales E, Goodstein D, Goodwin LA, Grady DL, Grigoriev I, Groza M, Hammon N, Hawkins T, Haydu L, Hildebrand CE, Huang W, Israni S, Jett J, Jewett PB, Kadner K, Kimball H, Kobayashi A, Krawczyk MC, Leyba T, Longmire JL, Lopez F, Lou Y, Lowry S, Ludeman T, Manohar CF, Mark GA, McMurray KL, Meincke LJ, Morgan J, Moyzis RK, Mundt MO, Munk AC, Nandkeshwar RD, Pitluck S, Pollard M, Predki P, Parson-Quintana B, Ramirez L, Rash S, Retterer J, Ricke DO, Robinson DL, Rodriguez A, Salamov A, Saunders EH, Scott D, Shough T,

Stallings RL, Stalvey M, Sutherland RD, Tapia R, Tesmer JG, Thayer N, Thompson LS, Tice H, Torney DC, Tran-Gyamfi M, Tsai M, Ulanovsky LE, Ustaszewska A, Vo N, White PS, Williams AL, Wills PL, Wu JR, Wu K, Yang J, Dejong P, Bruce D, Doggett NA, Deaven L, Schmutz J, Grimwood J, Richardson P, Rokhsar DS, Eichler EE, Gilna P, Lucas SM, Myers RM, Rubin EM, Pennacchio LA: The sequence and analysis of duplication-rich human chromosome 16. Nature 432: 988-994 (2004)

Marzella R, Viggiano L, Miolla V, Storlazzi CT, Ricco A, Gentile E, Roberto R, Surace C, Fratello A, Mancini M, Archidiacono N, Rocchi M: Molecular cytogenetic resources for chromosome 4 and comparative analysis of phylogenetic chromosome IV in great apes. Genomics 63: 307-313 (2000)

Mayr E: Animal species and evolution. Harvard University Press, Cambridge, MA (1963)

McBrearty S, Jablonski NG: First fossil chimpanzee. Nature 437: 105-108 (2005)

McCarroll SA, Hadnott TN, Perry GH, Sabeti PC, Zody MC, Barrett JC, Dallaire S, Gabriel SB, Lee C, Daly MJ, Altshuler DM: International HapMap Consortium Common deletion polymorphisms in the human genome. Nat Genet 38: 86-92 (2006)

McConkey EH: Orthologous numbering of great ape and human chromosomes is essential for comparative genomics. Cytogenet Genome Res 105: 157-8 (2004)

McLellan RA, Oscarson M, Alexandrie AK, Seidegard J, Evans DA, Rannug A, Ingelman-Sundberg M: Characterization of a human glutathione *S*-transferase mu cluster containing a duplicated *GSTM1* gene that causes ultrarapid enzyme activity. Mol Pharmacol 52: 958-965 (1997)

Mekel-Bobrov N, Gilbert SL, Evans PD, Vallender EJ, Anderson JR, Hudson RR, Tishkoff SA, Lahn BT: Ongoing adaptive evolution of ASPM, a brain size determinant in *Homo sapiens*. Science 309: 1720-1722 (2005)

Mitelman F, Johansson B, Mertens F: Mitelman database of chromosome aberrations in cancer. http://cgap.nci.nih.gov/Chromosomes/Mitelman (2006)

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

Müller S, Stanyon R, O'Brien PC, Ferguson-Smith MA, Plesker R, Wienberg J: Defining the ancestral karyotype of all primates by multidirectional chromosome painting between tree shrews, lemurs and humans. Chromosoma 108: 393-400 (1999)

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

Murphy WJ, Fronicke L, O'Brien SJ, Stanyon R: The origin of human chromosome 1 and its homologs in placental mammals. Genome Res 13: 1880-1888 (2003)

Murphy WJ, Larkin DM, Everts-van der Wind A, Bourque G, Tesler G, Auvil L, Beever JE, Chowdhary BP, Galibert F, Gatzke L, Hitte C, Meyers SN, Milan D, Ostrander EA, Pape G, Parker HG, Raudsepp T, Rogatcheva MB, Schook LB, Skow LC, Welge M, Womack JE, O'brien SJ, Pevzner PA, Lewin HA: Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. Science 309: 613-617 (2005)

Navarro A, Barton NH: Accumulating postzygotic isolation genes in parapatry: a new twist on chromosomal speciation. Evolution 57: 447-459 (2003a)

Navarro A, Barton NH: Chromosomal speciation and molecular divergence-accelerated evolution in rearranged chromosomes, Science 300: 321-324 (2003b)

Newman TL, Tuzun E, Morrison VA, Hayden KE, Ventura M, McGrath SD, Rocchi M, Eichler EE: A genome-wide survey of structural variation between human and chimpanzee. Genome Res 15: 1344-1356 (2005)

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

Nielsen R, Bustamante C, Clark AG, Glanowski S, Sackton TB, Hubisz MJ, Fledel-Alon A, Tanenbaum DM, Civello D, White TJ, Sninsky J, Adams MD, Cargill M: A scan for positively selected genes in the genomes of humans and chimpanzees. PLoS Biol 3: e170 (2005)

Noor MA, Grams KL, Bertucci LA, Reiland J: Chromosomal inversions and the reproductive isolation of species. Proc Natl Acad Sci USA 98: 12084-12088 (2001)

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

Ortiz-Barrientos D, Reiland J, Hey J, Noor MA: Recombination and the divergence of hybridizing species. Genetica 116: 167-178 (2002)

Osborne LR, Li M, Pober B, Chitayat D, Bodurtha J, Mandel A, Costa T, Grebe T, Cox S, Tsui LC, Scherer SW: A 1.5 million-base pair inversion polymorphism in families with Williams-Beuren syndrome. Nat Genet 29: 321-325 (2001)

Parsons JD: Miropeats: graphical DNA sequence comparisons. Comput Appl Biosci 11: 615-619 (1995)

Pevzner P, Tesler G: Human and mouse genomic sequences reveal extensive breakpoint reuse in mammalian evolution. Proc Natl Acad Sci USA 100: 7672-7677 (2003)

Pilbeam D, Young N: Hominoid evolution: synthesizing disparate data. C R Palevol 3: 305–321 (2004)

Rieseberg LH: Chromosomal rearrangements and speciation. Trends Ecol Evol 16: 351-358 (2001)

Rieseberg LH, Livingstone K: Evolution. Chromosomal speciation in primates. Science 300: 267-268 (2003)

Rudiger NS, Gregersen N, Kielland-Brandt MC: One short well conserved region of Alu-sequences is involved in human gene rearrangements and has homology with prokaryotic chi. Nucleic Acids Res 23: 256-260 (1995)

Ruiz-Herrera A, Garcia F, Azzalin C, Giulotto E, Egozcue J, Ponsa M, Garcia M: Distribution of intrachromosomal telomeric sequences (ITS) on *Macaca fascicularis* (*Primates*) chromosomes and their implication for chromosome evolution. Hum Genet 110: 578-586 (2002)

Scherer SW, Gripp KW, Lucena J, Nicholson L, Bonnefont J-P, Pérez-Jurado LA, Osborne LR: Observation of a parental inversion variant in a rare Williams-Beuren syndrome family with two affected children. Hum Genet 117: 383-388 (2005)

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 TC, Trask B, Patterson N, Zetterberg A, Wigler M: Large-scale copy number polymorphism in the human genome. Science 305: 525-528 (2004)

Sharp AJ, Locke DP, McGrath SD, Cheng Z, Bailey JA, Vallente RU, Pertz LM, Clark RA, Schwartz S, Segraves R, Oseroff VV, Albertson DG, Pinkel D, Eichler EE: Segmental duplications and copy-number variation in the human genome. Am J Hum Genet 77: 78-88 (2005)

She X, Jiang Z, Clark RA, Liu G, Cheng Z, Tuzun E, Church DM, Sutton G, Halpern AL, Eichler EE: Shotgun sequence assembly and recent segmental duplications within the human genome. Nature 431: 927-930 (2004)

Shimada MK, Kim CG, Kitano T, Ferrell RE, 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)

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

Spitzner JR, Chung IK, Muller MT: Eukaryotic topoisomerase II preferentially cleaves purine-pyrimidine repeats. Nucleic Acids Res 18: 1-11 (1990)

Stankiewicz P, Park SS, Inoue K, Lupski JR: 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, Lupski JR: Genome architecture, rearrangements and genomic disorders. Trends Genet 18: 74-82 (2002)

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

Starke H, Seidel J, Henn W, Reichardt S, Volleth M, Stumm M, Behrend C, Sandig KR, Kelbova C, Senger G, Albrecht B, Hansmann I, Heller A, Claussen U, Liehr T: Homologous sequences at human chromosome 9 bands p12 and q13-21.1 are involved in different patterns of pericentric rearrangements. Eur J Hum Genet 10: 790-800 (2002)

Stefansson H, Helgason A, Thorleifsson G, Steinthorsdottir V, Masson G, Barnard J, Baker A, Jonasdottir A, Ingason A, Gudnadottir VG, Desnica N, Hicks A, Gylfason A, Gudbjartsson DF, Jonsdottir GM, Sainz J, Agnarsson K, Birgisdottir B, Ghosh S, Olafsdottir A, Cazier JB, Kristjansson K, Frigge ML, Thorgeirsson TE, Gulcher JR, Kong A, Stefansson K: A common inversion under selection in Europeans. Nat Genet 37: 129-137 (2005)

Strickberger MW: Evolution. Jones and Bartlett Publishers, Boston Toronto London Singapore, S. 590-593 (2000)

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

Szamalek JM, Goidts V, Searle JB, Cooper DN, 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 87: 39-45 (2006)

Szamalek JM, Cooper DN, Schempp W, Minich P, Kohn M, Hoegel J, Goidts V, Hameister H, Kehrer-Sawatzki H: Polymorphic micro-inversions contribute to the genomic variability of humans and chimpanzees. Hum Genet 119: 103-112 (2006)

Szamalek JM, Goidts V, Cooper DN, Hameister H, Kehrer-Sawatzki H: Characterization of the human lineage-specific pericentric inversion that distinguishes human chromosome 1 from the homologous chromosomes of the great apes. (manuscript in preparation)

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

Thomas PD, Kejariwal A, Campbell MJ, Mi H, Diemer K, Guo N, Ladunga I, Ulitsky-Lazareva B, Muruganujan A, Rabkin S, Vandergriff JA, Doremieux O: PANTHER: A browsable database of gene products organized by biological function, using curated protein family and subfamily classification. Nucleic Acids Res 31: 334–341 (2003a)

Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, Diemer K, Muruganujan A, Narechania A: PANTHER: A library of protein families and subfamilies indexed by function. Genome Res 13: 2129–2141 (2003b)

Tuzun E, Sharp AJ, Bailey JA, Kaul R, Morrison VA, Pertz LM, Haugen E, Hayden H, Albertson D, Pinkel D, Olson MV, Eichler EE: Fine-scale structural variation of the human genome. Nat Genet 37: 727-732 (2005)

Uddin M, Wildman DE, Liu G, Xu W, Johnson RM, Hof PR, Kapatos G, Grossman LI, Goodman M: Sister grouping of chimpanzees and humans as revealed by genome-wide phylogenetic analysis of brain gene expression profiles. Proc Natl Acad Sci U S A 101: 2957-2962 (2004)

Vallender EJ, Lahn BT: Effects of chromosomal rearrangements on human-chimpanzee molecular evolution, Genomics 84: 757-761 (2004)

Varki A, Altheide TK: Comparing the human and chimpanzee genomes: searching for needles in a haystack. Genome Res 15: 1746-1758 (2005)

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)

Walsh JB: Rate of accumulation of reproductive isolation by chromosomal rearrangement. Am Nat 120: 510-532 (1982)

Washburn SL: In Rothblatt B (ed) Changing perspectives on man, University Chicago Press, Chicago, S. 193-201 (1968)

Weise A, Starke H, Mrasek K, Claussen U, Liehr T: New insights into the evolution of chromosome 1. Cytogenet Genome Res 108: 217-222 (2005)

White MJD: Animal cytology and evolution. Cambridge University Press, London (1973)

White MJD: Modes of speciation. Freeman, San Francisco (1978)

Wienberg J, Jauch A, Stanyon R, Cremer T: Molecular cytotaxonomy of primates by chromosomal in situ suppression hybridization. Genomics 8: 347-350 (1990)

Wilson GM, Flibotte S, Missirlis PI, Marra MA, Jones S, Thornton K, Clark AG, Holt RA: Identification by full-coverage array CGH of human DNA copy number increases relative to chimpanzee and gorilla. Genome Res 16:173-181 (2006)

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

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

Yohn CT, Jiang Z, McGrath SD, Hayden KE, Khaitovich P, Johnson ME, Eichler MY, McPherson JD, Zhao S, Paabo S, Eichler EE: Lineage-specific expansions of retroviral insertions within the genomes of African great apes but not humans and orangutans. PLoS Biol 3: 1−11 (2005)

Yu N, Jensen-Seaman MI, Chemnick L, Kidd JR, Deinard AS, Ryder O, Kidd KK, Li WH: Low nucleotide diversity in chimpanzees and bonobos. Genetics 164: 1511-1518 (2003)

Yunis JJ, Sawyer JR, Dunham K: The striking resemblance of high-resolution G-banded chromosomes of man and chimpanzee. Science 208: 1145-1148 (1980)

Yunis JJ, 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)

7. Appendix

Appendix 1: Genomic DNA sequence divergence between humans and chimpanzees

***PTR* contig**	***HSA* chromosome**	**Corresponding human sequence**	**Position of the human sequence on chromosome (Mb) [a]**	**Length (nt) [b]**	**Sequence divergence %**	**Average recombination rate[c]**
Collinear regions on rearranged chromosomes						
AADA01296522	4p15.33	RP11-352E6	12,173,446-12,335,659	39,713	1.58	0.6
AADA01077501	4p13	RP11-16I14	43,443,993-43,606,067	23,909	1.70	0.6
AADA01290361	4q22.1	RP11-115D19	90,893,328-91,062,811	41,810	1.34	1.5
AADA01259474	4q22.3	RP11-416E16	95,268,142-95,437,204	13,729	1.07	0.8
AADA01036409	4q25	RP11-477G18	113,288,164-113,459,908	12,321	1.01	1.3
AADA01283430	4q27	RP11-724L20	123,765,982-123,855,863	19,553	1.04	0.5
AADA01040055	4q31.3	RP11-18H21	153,367,579-153,554,414	26,158	1.47	0.8
AADA01272696	5p15.31	RP11-325M13	8,283,345-8,397,763	26,572	1.54	1.3
AADA01271200	5q22.3	RP11-297K24	113,105,710-113,303,876	19,544	1.33	0.8
AADA01203049	5q31.1	AC005571	133,017,569-133,079,856	25,686	1.48	1.1
AADA01005193	5q33.1	RP11-41I1	152,152,520-152,225,128	22,152	1.39	0.4
AADA01013394	5q35.2	CTC-216O13	173,705,708-173,807,010	27,528	1.27	3.5
AADA01293397	12p13.32-33	RP11-319E16	5,163,922-5,344,309	16,920	1.06	1.8
AADA01181778	12q21.1	RP11-166B14	72,214,269-72,379,961	11,188	1.50	0.2
AADA01190034	12q24.13	RP5-1048I22	111,447,905-111,587,875	12,959	1.00	1.5
AADA01111140	17q24.1	CTD-2535L24	60,755,493-60,998,405	29,804	1.23	0.7
AADA01181217	17q24.2	RP11-118B18	64,178,992-64,303,770	18,039	1.29	1.5
AADA01265928	17q25.3	RP11-13K12	73,020,572-73,183,186	13,943	1.75	2.8
AADA01240602	18q12.2	RP11-956K8	33,980,744-34,174,312	30,541	1.10	0.4
AADA01230686	18q21.2	RP11-202D1	49,331,393-49,510,377	15,690	1.40	0.8
AADA01266539	18q22.3	RP11-454J3	67,703,807-67,916,648	30,367	1.52	1.4
Subtotal				**478,126**	**1.34±0.23**	**1.16±0.80**

PTR contig	*HSA* chromosome	Corresponding human sequence	Position of the human sequence on chromosome (Mb) [a]	Length (nt) [b]	Sequence divergence %	Average recombination rate[c]
Inverted region on rearranged chromosomes						
AADA01063099	4p12	RP11-381K8	46,300,175-46,408,210	30,933	1.16	0.4
AADA01164033	4q13.3	RP11-1366O24	71,305,974-71,509,397	12,440	1.29	0.6
AADA01274022	4q13.3	RP11-545L5	73,039,440-73,216,709	17,180	1.32	0.5
AADA01190640	4q21.22	RP11-689K5	83,060,683-83,254,888	9,551	0.82	0.3
AADA01301758	4q21.23	RP11-176K15	85,303,831-85,451,557	22,902	1.10	0.9
AADA01157755	5p13.2	CTD-3159E21	36,398,426-36,467,397	31,852	1.14	1.2
AADA01082355	5q12.1	RP11-586E1	62,343,310-62,519,576	33,460	1.16	0.3
AADA01124278	5q14.3	RP11-456A14	83,069,191-83,214,237	13,609	1.12	1.0
AADA01320464	5q15	CTC-424O20	92,410,217-92,482,891	35,867	0.83	0.8
AADA01107816	12q13.11	RP11-66E20	45,309,120-45,477,058	25,417	0.78	1.2
AADA01307609	12q14.1	RP11-767I20	57,166,976-57,340,420	26,114	1.30	0.6
AADA01296510	12q14.1	RP11-471N19	59,829,484-60,026,445	11,301	1.42	0.4
AADA01202096	17p11.2	CTD-3073J20	17,202,346-17,343,393	14,589	1.27	0.8
AADA01140350	17q12	RP11-445F12	32,189,537-32,371,527	10,761	1.38	1.7
AADA01081274	17q21.31	CTD-3070M1	41,130,491-41,241,610	15,801	1.19	0.0
AADA01329181	18p11.32	RP11-78F17	1,111,566-1,162,565	9,255	1.14	2.2
AADA 01084978	18p11.32	RP11-620B7	2,285,350-2,461,875	25,070	1.54	4.2
AADA01225804	18p11.23-22	RP11-661O13	8,449,167-8,644,909	21,871	1.43	2.7
Subtotal				**367,973**	**1.19±0.21**	**1.1±1.04**
Collinear chromosomes						
AADA01021456	3p24.2	RP11-711F20	24,882,986-25,031,899	32,202	1.48	1.5
AADA01260852	3p14.2	RP11-285B24	59,159,647-59,345,581	24,834	1.29	1.9
AADA01061604	3q13.31	RP11-190P13	116,398,884-116,554,710	22,666	0.98	0.6
AADA01016714	3q24	RP11-94I16	145,273,728-145,429,781	42,622	1.23	1.2
AADA01167741	6p22.3	RP11-440J4	23,839,560-23,971,554	16,016	1.97	1.1
AADA01011595	6p21.1	RP3-449H6	44,584,484-44,669,096	42,403	1.25	1.6
AADA01268566	6q12	RP11-141E7	67,753,888-67,903,736	16,706	1.54	1.0
AADA01326467	6q14.1	RP11-388H7	81,342,951-81,476,347	22,049	1.18	0.4

***PTR* contig**	***HSA* chromosome**	**Corresponding human sequence**	**Position of the human sequence on chromosome (Mb) [a]**	**Length (nt) [b]**	**Sequence divergence %**	**Average recombination rate[c]**
AADA01285174	6q16.3	RP11-530P14	104,849,074-104,892,393	18,456	1.14	0.6
AADA01062264	6q23.3	RP11-394G3	136,068,957-136,219,028	6,856	1.25	1.9
AADA01231293	6q24.2	RP11-43G16	145,372,442-145,417,827	16,912	1.03	0.5
AADA01103492	7p21.1	RP11-71F18	19,212,957-19,387,283	15,656	1.42	1.4
AADA01140238	7p14.1	RP11-302L6	37,214,812-37,404,464	40,885	1.35	1.3
AADA01090361	7q11.22	RP5-917G4	67,638,484-67,745,310	12,703	1.86	1.8
AADA01088005	7q22.3	CTB-111H14 RP11-17I10	105,555,345-105,635,512 105,635,313-105,682,354	15,765	1.43	1.4
AADA01011335	7q31.33	AC004092	125,062,934-125,106,577	14,230	1.28	0.4
AADA01059687	10p13	RP11-37P5	16,094,699-16,196,799	22,798	1.48	2.3
AADA01087119	10p11.22	RP13-259N13	34,012,154-34,056,134	15,237	1.01	0.8
AADA01246078	10q22.3	RP11-17G2	79,961,462-80,018,162	26,870	1.31	1.25
AADA01224067	10q25.1	RP11-242A15	109,808,875-109,961,772	18,176	1.79	1.0
Subtotal				**444,042**	**1.36±0.27**	**1.20±0.54**

[a] Ensembl v27.35a.1; [b] Length of alignment after the removal of gaps; [c] Sex-averaged rates of recombination based on the deCODE genetic map (http://genome.ucsc.edu)

Appendix 2: The inversions identified by gene order comparison between human and chimpanzee

Gene name	RefSeq (Accession number)	*HSA* Chromosome	*HSA* chromosomal location	*PTR* chromosomal location	Number of genes in inversion
CB45_HUMAN	NM_016176	1	1,180,538	1,077,875	2
TNFRSF4	NM_003327		1,186,630	1,072,138	
NP_660155	NM_145172	1	85,176,851	83,913,256	2
MCOLN3	NM_018298		85,195,787	83,869,673	
GSTM5	NM_000851	1 [b]	109,966,907	112,480,987	3
CSF1	NM_000757		110,165,499	112,262,561	
NP_981957	NM_203412	1 [b]	110,367,110	111,915,302	29
SLC16A1	NM_003051		113,166,896	109,077,736	
PTPN22	NM_012411	1	114,003,500	113,856,379	2
RSBN1	NM_018364		114,019,295	113,806,563	
NOTCH2	NM_024408	1	120,100,879	120,035,183	2
ADAM30	NM_021794		120,148,629	120,017,058	
NP_848516	NM_178429	1	149,461,565	131,237,035	2
SPRL1B	NM_014357		149,471,672	131,236,318	
NP_848130	NM_178353	1	149,571,826	131,369,636	2
NP_848129	NM_178352		149,583,344	131,362,321	
FCER1A	NM_002001	1	156,072,577	138,142,611	2
OAJ5_HUMAN	NM_001004469		156,317,941	138,132,093	
ADIPOR1	NM_015999	1	199,641,611	182,493,708	2
NQO3A2	NM_016243		199,662,661	182,464,430	
OBSCN	NM_024868	1	224,702,589	208,494,706	2
TRIM11	NM_145214		224,888,114	208,434,060	
IL8RB	NM_001557	2	218,816,252	108,979,772	2
IL8RA	NM_000634		218,853,076	108,913,014	
IL17RE	NM_144640	3	9,903,711	10,293,559	2
NP_115881	NM_032492		9,907,272	10,281,226	
NP_775929	NM_173658	3	44,601,460	45,890,231	2
ZNF197	NM_006991		44,645,484	45,860,393	
ZNF27	NM_024903	4	254,978	270,298	2
ZNF141	NM_003441		327,570	44,493	
GABRA2	NM_000807	4 [a]	46,092,512	100,714,244	56
NP_079019	NM_024743		69,974,927	76,888,456	
UGT2B4	NM_021139	4 [a]	70,526,645	76,526,687	17
NP_997722	NM_212557		71,565,058	75,468,131	
ENAM	NM_031889	4 [a]	71,860,253	61,866,264	73
NP_598375	NM_133636		84,685,682	49,002,091	
NP_116106	NM_032717	4 [a]	84,814,440	48,825,399	4
WDFY3	NM_014991		85,947,874	47,435,187	
CCRN4L	NM_012118	4	140,294,548	157,569,938	2
ELF2	NM_006874		140,336,476	157,540,102	
FGA	NM_000508	4	155,826,091	173,226,243	2
FGB	NM_005141		155,841,768	173,205,730	

Gene name	RefSeq (Accession number)	*HSA* Chromosome	*HSA* chromosomal location	*PTR* chromosomal location	Number of genes in inversion
CDH12	NM_004061	5 [a]	21,786,732	95,411,215	7
PDZK3	NM_015022		31,675,274	85,336,076	
GOLPH3	NM_022130	5	32,160,581	97,465,874	11
C1QTNF3	NM_030945		34,053,770	99,399,863	
RAI14	NM_015577	5 [a]	34,692,274	81,989,812	34
NP_699192	NM_153361		43,228,084	73,207,789	
NP_071928	NM_022483	5 [a]	43,480,512	72,999,200	72
NP_689621	NM_152408		75,005,789	40,385,234	
F2R	NM_001992	5 [a]	76,047,542	39,362,626	37
EDIL3	NM_005711		83,272,175	31,441,891	
COX7C	NM_001867	5 [a]	85,949,514	29,286,601	24
PCSK1	NM_000439		95,751,875	19,367,370	
EPB41L4A	NM_022140	5	111,507,433	117,634,053	2
NP_443726	NM_053000		111,524,122	117,629,581	
TXNDC5	NM_022085	6	7,826,751	8,333,468	2
NP_689913	NM_152700		7,931,335	8,305,611	
MCA2_HUMAN	NM_006303	7	5,822,123	6,965,774	7
ZDHHC4	NM_018106		6,390,324	6,396,276	
C7orf36	NM_020192	7	39,379,249	45,134,565	10
BLVRA	NM_000712		43,571,512	40,620,998	
NP_958861	NM_201453	9	67,940,488	52,048,351	2
PGM5	NM_021965		68,201,781	51,975,752	
FCMD	NM_006731	9	105,295,556	89,862,671	2
NP_997530	NM_207647		105,342,495	89,812,555	
COQ4	NM_016035	9	128,164,369	113,012,245	2
SLC27A4	NM_005094		128,182,394	112,996,935	
C9orf9	NM_018956	9	132,783,304	117,676,204	2
TSC1	NM_000368		132,796,291	117,660,216	
RPL7A	NM_000972	9	133,231,074	118,284,242	2
SURF5	NM_006752		133,234,714	118,276,918	
PHYHIPL	NM_032439	10	60,606,353	61,418,106	2
FAM13C1	NM_001001971		60,675,900	61,360,703	
AMID	NM_032797	10	71,343,619	72,710,741	2
H2AFY2	NM_018649		71,482,612	72,651,309	
VAMP1	NM_014231	12	6,425,495	6,975,808	2
TAPBPL	NM_018009		6,431,676	6,965,520	
O9K2_HUMAN	NM_001005243	12	53,809,886	55,080,314	2
OAA7_HUMAN	NM_001005280		53,901,076	55,066,032	
MBD6	NM_052897	12	56,202,892	57,452,850	2
DCTN2	NM_006400		56,210,202	57,451,260	
CCT2	NM_006431	12	68,265,512	69,815,472	2
VMD2L3	NM_032735		68,323,601	69,810,492	
PSME2	NM_002818	14	23,677,961	22,582,923	2
C14orf122	NM_016049		23,678,009	22,578,504	

Gene name	RefSeq (Accession number)	*HSA* Chromosome	*HSA* chromosomal location	*PTR* chromosomal location	Number of genes in inversion
KIAA1409	NM_020818	14	92,869,324	93,449,071	2
COX8C	NM_182971		92,883,290	93,319,557	
TUBGCP5	NM_052903	15 [a]	20,384,869	26,176,149	17
TJP1	NM_003257		27,779,656	19,157,482	
KLF13	NM_015995	15	28,915,746	28,979,078	2
TRPM1	NM_002420		29,080,845	28,652,287	
U235_HUMAN	NM_144597	15	81,464,396	82,263,722	2
BTBD1	NM_025238		81,476,189	82,256,181	
NP_061141	NM_017979	15	89,274,414	90,230,413	2
NP_940929	NM_198527		89,275,161	90,226,044	
NP_055529	NM_014714	16	1,500,432	1,670,295	2
C16orf30	NM_024600		1,518,743	1,598,236	
CLN3	NM_000086	16	28,396,102	29,251,863	4
NP_612423	NM_138414		28,472,758	29,142,962	
NP_112584	NM_031294	17	17,686,706	18,223,069	2
TOM1L2	NM_144678		17,690,865	18,099,559	
HUS7_HUMAN	NM_018096	17	30,471,195	34,004,479	2
NP_060029	NM_017559		30,472,763	33,994,750	
CCL4	NM_002984	17	31,421,234	34,984,132	2
CCL3	NM_002983		31,439,718	34,968,495	
KRTAP9-3	NM_031961	17	36,642,280	39,865,699	2
KRTAP9-5	NM_033191		36,659,465	39,848,707	
NP_055649	NM_014834	17	41,728,274	63,387,660	51
SMURF2	NM_022739		59,971,945	45,102,092	
NP_055775	NM_014960	17	63,766,918	67,110,754	2
SLC16A6	NM_004694		63,775,933	67,070,990	
USP14	NM_005151	18 [a]	148,553	12,600,588	27
NAPG	NM_003826		10,516,031	2,119,482	
GNAL	NM_002071	18	11,679,263	1,504,605	3
IMPA2	NM_014214		11,971,427	1,796,789	
TXNL1	NM_004786	18	52,421,053	49,295,805	2
WDR7	NM_015285		52,490,745	49,175,064	
CSNK1G2	NM_001319	19	1,892,188	2,018,860	2
C19orf34	NM_152771		1,903,530	2,000,182	
ZNF77	NM_021217	19	2,847,754	3,017,633	2
NP_775751	NM_173480		2,851,955	2,983,589	
ZNF563	NM_145276	19	12,289,308	12,631,901	2
ZNF443	NM_005815		12,401,521	12,537,111	
SIN3B	NM_015260	19	16,801,231	18,138,393	3
CPAMD8	NM_015692		16,864,758	17,993,382	
FCHO1	NM_015122	19	17,599,507	18,714,631	2
BPY2IP1	NM_018174		17,691,303	18,682,410	
ZNF566	NM_032838	19	41,630,424	39,105,827	4
ZNF383	NM_152604		42,409,206	38,342,486	
ZNF571	NM_016536	19	42,745,134	39,560,386	2
ZNF540	NM_152606		42,777,571	39,549,022	

Gene name	RefSeq (Accession number)	*HSA* Chromosome	*HSA* chromosomal location	*PTR* chromosomal location	Number of genes in inversion
MIA	NM_006533	19	45,973,218	42,965,061	2
RAB4B	NM_016154		45,976,011	42,955,813	
PSG2	NM_031246	19	48,121,727	45,155,482	2
PSG11	NM_002785		48,205,990	45,096,769	
ZNF235	NM_004234	19	49,444,793	46,619,819	2
NP_861421	NM_181756		49,446,158	46,571,222	
CCDC8	NM_032040	19	51,606,291	48,895,866	2
NP_060685	NM_018215		51,645,244	48,875,424	
NUP62	NM_012346	19	55,064,113	52,448,314	2
TBC1D17	NM_024682		55,072,641	52,436,250	
BIRC8	NM_033341	19	58,484,729	55,905,759	2
NP_001004301	NM_001004301		58,642,274	55,820,296	
ZNF524	NM_153219	19	60,791,991	58,246,024	2
NP_116225	NM_032836		60,794,558	58,235,302	
NP_001002836	NM_001002836	19	61,291,201	58,836,070	2
ZNF444	NM_018337		61,344,368	58,814,643	
SCAND1	NM_016558	20	34,004,960	36,225,941	2
C20orf152	NM_080834		34,019,943	36,056,068	
SLA2	NM_032214	20	34,657,114	36,910,773	2
C20orf24	NM_018840		34,667,581	36,904,060	
WFDC13	NM_172005	20	43,764,069	46,286,363	2
WFDC3	NM_080614		43,836,254	46,260,341	
UBE2C	NM_007019	20	43,874,662	46,404,292	2
TNNC2	NM_003279		43,885,264	46,401,584	
COL18A1	NM_030582	21	45,649,485	46,094,882	2
SLC19A1	NM_003056		45,759,057	45,966,312	
HIC2	NM_015094	22	20,096,247	20,144,766	2
NP_955377	NM_199283		20,151,843	20,112,735	
APOL2	NM_030882	22	34,946,755	35,131,746	2
MYH9	NM_002473		35,001,823	35,083,003	
SA50_HUMAN	NM_015380	22	42,676,202	42,974,103	2
PARVB	NM_001003828		42,719,992	42,909,215	
HDHD1A	NM_012080	X	6,826,697	7,027,836	2
STS	NM_000351		6,997,233	6,956,737	

[a] demarcate the well-characterized, microscopically detectable pericentric inversions of *HSA* 4, 5, 15 and 18;
[b] demarcate one micro-inversion on *HSA* 1p13.2-3 and in further analyses regarded as a one unit..

Appendix 3: List of *HSA* and orthologous *PTR* genes encompassed by the confirmed micro-inversions

Chromosome *HSA* (*PTR*)	**Genes name (Accession number)** [a]	**Gene description**	**PANTHER Biological Process** [a]
***HSA* 1p13.2-13.3 (*PTR* 1)**	GSTM2 (NM_000848)	Glutathione *S*-transferase M2 (muscle)	Detoxification
	GSTM1 (NM_000561)	Glutathione *S*-transferase M1	Detoxification
	ENK13_HUMAN	No description	[b]
	GSTM5 (NM_000851)	1Glutathione *S*-transferase M5	Detoxification
	GSTM3 (NM_000849)	Glutathione *S*-transferase M3 (brain)	Detoxification
	EPS8L3 (NM_024526)	EPS8-like 3	Developmental processes
	CSF1 (NM_000757)	Colony stimulating factor 1 (macrophage)	Ligand-mediated signaling Macrophage-mediated immunity Granulocyte-mediated immunity Hematopoiesis
	AHCYL1 (NM_006621)	*S*-adenosylhomocysteine hydrolase-like 1	Purine metabolism
	FAM40A (NM_033088)	Family with sequence similarity 40, member A	Biological process unclassified
	ALX3 (NM_006492)	Aristaless-like homeobox 3	mRNA transcription regulation Developmental processes
	NP_981957 (NM_203412)	Hypothetical protein LOC164153	Biological process unclassified
	KCNC4 (NM_004978)	Potassium voltage-gated channel, Shaw-related subfamily, member 4	Cell communication Cation transport Synaptic transmission
	RBM15 (NM_022768)	RNA binding motif protein 15	Biological process unclassified
	Q9UHU7_HUMAN	No description	[b]
	SLC16A4 (NM_004696)	Solute carrier family 16 (monocarboxylic acid transporters), member 4	Cation transport
	HBXIP (NM_006402)	Hepatitis B virus x interacting protein	Developmental processes
	PROK1 (NM_032414)	Prokineticin 1	Neuronal activities
	KCNA10 (NM_005549)	Potassium voltage-gated channel, shaker-related subfamily, member 10	Cation transport

Chromosome *HSA* (*PTR*)	Genes name (Accession number) [a]	Gene description	PANTHER Biological Process [a]
	KCNA2 (NM_004974)	Potassium voltage-gated channel, shaker-related subfamily, member 2	Cation transport
	KCNA3 (NM_002232)	Potassium voltage-gated channel, shaker-related subfamily, member 3	Cation transport
	Q8NHC3_HUMAN	No description	[b]
	CD53 (NM_000560)	CD53 antigen	Cell adhesion-mediated signaling Cell adhesion
	NP_001006946 (NM_018372)	Receptor-interacting factor 1 isoform 2	Biological process unclassified
	NP_848549 (NM_178454)	Hypothetical protein LOC128338	Biological process unclassified
	CEPT1 (NM_006090)	Choline/ethanolamine phosphotransferase 1	Regulation of lipid, fatty acid and steroid metabolism Phospholipid metabolism
	NP_079177 (NM_024901)	Hypothetical protein FLJ22457	Signal transduction Tumor suppressor
	CHI3L2 (NM_004000)	Chitinase 3-like 2	Other polysaccharide metabolism
	CHIA_HUMAN (NM_021797)	Acidic mammalian chitinase precursor	Other polysaccharide metabolism
	NP_857594 (NM_181643)	Hypothetical protein LOC128344	Biological process unclassified
	OVGP1 (NM_002557)	Oviductal glycoprotein 1, 120kDa (mucin 9, oviductin)	Other polysaccharide metabolism
	MEP50_HUMAN (NM_024102)	Methylosome protein 50 (MEP50 protein).	Biological process unclassified
	ATP5F1 (NM_0010020)	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit b, isoform 1	Cation transport
	NP_777556 (NM_174896)	Hypothetical protein LOC128346	Biological process unclassified
	ADORA3 (NM_000677)	Adenosine A3 receptor	G-protein mediated signaling
	RAP1A (NM_002884)	RAP1A, member of RAS oncogene family	Signal transduction Regulated exocytosis Receptor mediated endocytosis
	NP_945120 (NM_019099)	Hypothetical protein LOC55924 isoform 2	Biological process unclassified
	DDX20 (NM_007204)	DEAD (Asp-Glu-Ala-Asp) box polypeptide 20	Nucleoside, nucleotide and nucleic acid metabolism

Chromosome *HSA* (*PTR*)	**Genes name (Accession number) [a]**	**Gene description**	**PANTHER Biological Process [a]**
	KCND3_HUMAN (NM_004980)	Potassium voltage-gated channel subfamily D member 3 (Voltage-gated potassium channel subunit Kv4.3)	Cation transport
	NP_061174 (NM_018704)	Hypothetical protein LOC55917	Biological process unclassified
	WNT2B (NM_004185)	Wingless-type MMTV integration site family, member 2B	Ligand-mediated signaling Developmental processes Alzheimer disease-presenilin
	ST7L (NM_138729)	Suppression of tumorigenicity 7 like	Tumor suppressor
	CAPZA1 (NM_006135)	Capping protein (actin filament) muscle Z-line, alpha 1	Cell structure
	MOV10 (NM_020963)	Mov10, Moloney leukemia virus 10, homolog (mouse)	DNA replication Developmental processes DNA replication
	RHOC (NM_175744)	Ras homolog gene family, member C	G-protein mediated signaling Developmental processes Cell proliferation and differentiation Cell structure and motility
	NP_005158 (NM_005167)	Protein phosphatase 1J (PP2C domain containing)	Biological process unclassified
	FAM19A3 (NM_001004440)	Family with sequence similarity 19 (chemokine (C-C motif)-like), member A3	Biological process unclassified
	XP_060943.1	Similar to nuclear transport factor 2 (NTF-2	[b]
	Q8N9X3_HUMAN (NM_207399)	No description	[b]
	SLC16A1 (NM_003051)	Solute carrier family 16 (monocarboxylic acid transporters), member 1	Cation transport
	LRIG2 (NM_014813)	Leucine-rich repeats and immunoglobulin-like domains 2	Cell adhesion Other neuronal activity
***HSA* 7p22.1 (*PTR* 7)**	PMS2 (NM_000535)	PMS2 postmeiotic segregation increased 2 (*S. cerevisiae*)	DNA repair, Meiosis, Cell cycle Cell proliferation and differentiation
	MCA2_HUMAN (NM_006303)	Multisynthetase complex auxiliary component p38	Protein biosynthesis Other metabolism
	EIF2AK1 (NM_014413)	Eukaryotic translation initiation factor 2-alpha kinase 1	Protein biosynthesis Protein phosphorylation -

Chromosome *HSA* (*PTR*)	Genes name (Accession number) [a]	Gene description	PANTHER Biological Process [a]
	USP42 (XM_374396)	Ubiquitin specific protease 42	Proteolysis
	PSCD3 (NM_004227)	Pleckstrin homology, Sec7 and coiled-coil domains 3	Intracellular protein traffic
	MGC12966 (NM_032706)	Hypothetical protein MGC12966	Biological process unclassified
	RAC1_HUMAN (NM_006908)	Ras-related C3 botulinum toxin substrate 1	G-protein mediated signaling Cell structure and motility
	NP_631918.1 (NM_139179)	Hypothetical protein LOC221955	Biological process unclassified
	KDELR2 (NM_006854)	KDEL (Lys-Asp-Glu-Leu) endoplasmic reticulum protein retention receptor 2	Constitutive exocytosis
	Q9NXD8_HUMAN	Hypothetical protein FLJ20306	[b]
	ZDHHC4 (NM_018106)	Zinc finger, DHHC-type containing 4	Biological process unclassified
	C7orf26 (NM_024067)	Chromosome 7 open reading frame 26	Biological process unclassified
	XP_499257.1 (XM_496849)	Hypothetical protein DKFZp434J1015	[b]
	ZNF12 (NM_016265)	Zinc finger protein 12	[b]
***HSA* 7p13-14.1 (*PTR* 7)**	C7orf36 (NM_020192)	Chromosome 7 open reading frame 36	Biological process unclassified
	Q8N8G3_HUMAN	Hypothetical protein FLJ39538	[b]
	RALA (NM_005402)	V-ral simian leukemia viral oncogene homolog A (ras related)	Signal transduction Embryogenesis Cell motility
	CDC2L5 (NM_003718)	Cell division cycle 2-like 5 (cholinesterase-related cell division controller)	Protein phosphorylation Cell cycle control Cell proliferation and differentiation
	C7orf11 (NM_138701)	Chromosome 7 open reading frame 11	Biological process unclassified
	C7orf10 (NM_024728)	Chromosome 7 open reading frame 10	Fatty acid metabolism
	INHBA (NM_002192)	Inhibin, beta A (activin A, activin AB alpha polypeptide)	Receptor protein serine/threonine kinase signaling pathway Ligand-mediated signaling

Chromosome *HSA* (*PTR*)	**Genes name (Accession number) [a]**	**Gene description**	**PANTHER Biological Process [a]**
	GLI3 (NM_000168)	GLI-Kruppel family member GLI3 (Greig cephalopolysyndactyly syndrome)	mRNA transcription regulation Embryogenesis Segment specification Neurogenesis
	C7orf25 (NM_024054)	Chromosome 7 open reading frame 25	Biological process unclassified
	PSMA2 (NM_002787)	Proteasome (prosome, macropain) subunit, alpha type, 2	Proteolysis
	MRPL32 (NM_031903)	Mitochondrial ribosomal protein L32	Biological process unclassified
	Q8NA17_HUMAN	Hypothetical protein FLJ35943	[b]
	HECW1 (NM_015052)	HECT, C2 and WW domain containing E3 ubiquitin protein ligase 1	Proteolysis
	STK17A (NM_004760)	Serine/threonine kinase 17a (apoptosis-inducing)	Protein phosphorylation Apoptosis
	NP_060694.1 (NM_018224)	Hypothetical protein FLJ10803	Biological process unclassified
	BLVRA (NM_000712)	Biliverdin reductase A	Coenzyme metabolism Porphyrin metabolism
***HSA* 18p11.21-11.22 (*PTR* 18)**	FAM38B (NM_022068)	Family with sequence similarity 38, member B	Biological process unclassified
	C18orf30 (NM_207347)	Chromosome 18 open reading frame 30	[b]
	C18orf58 (NM_173817)	Chromosome 18 open reading frame 58	Biological process unclassified
	GNAL (NM_002071)	Guanine nucleotide binding protein (G protein), alpha activating activity polypeptide, olfactory type	G-protein mediated signaling
	MPPE1 (NM_023075)	Metallophosphoesterase 1	Biological process unclassified
	IMPA2 (NM_014214)	Inositol(myo)-1(or 4)-monophosphatase 2	Phospholipid metabolism Phosphate metabolism Other intracellular signaling cascade
	XP_377488.2	Similar to KIAA1074 protein	[b]
***HSA* 19q13.12 (*PTR* 19)**	Y1559_HUMAN (NM_020917)	Hypothetical zinc finger protein KIAA1559	mRNA transcription regulation
	ZNF545 (NM_133466)	Zinc finger protein 545	mRNA transcription regulation
	ZNF566 (NM_032838)	Zinc finger protein 566	mRNA transcription regulation

Chromosome *HSA* (*PTR*)	Genes name (Accession number) [a]	Gene description	PANTHER Biological Process [a]
	ZNF529 (NM_020951)	Zinc finger protein 529	mRNA transcription regulation
	ZNF382 (NM_032825)	Zinc finger protein 382	mRNA transcription regulation
	GIOT1_HUMAN (NM_153257)	Gonadotropin inducible transcription repressor-1	mRNA transcription regulation
	ZNF567 (NM_152603)	Zinc finger protein 567	mRNA transcription regulation
	XP_375606.2	Hypothetical protein DKFZp779O175	[b]
	XM_292740.3	Hypothetical protein LOC342892	[b]
	NP_996777.1 (NM_206894)	Hypothetical protein LOC388536	mRNA transcription regulation
	ZNF345 (NM_003419)	Zinc finger protein 345	mRNA transcription regulation
	ZNF568 (NM_198539)	Zinc finger protein 568	mRNA transcription regulation
	Q96ML5 (NM_144689)	Zinc finger protein 420	mRNA transcription regulation
	ZNF585A (NM_152655)	Zinc finger protein 585A	mRNA transcription regulation Cell proliferation and differentiation
	ZNF585B (NM_152279)	Zinc finger protein 585B	mRNA transcription regulation Cell proliferation and differentiation
	ZNF383 (NM_152604)	Zinc finger protein 383	mRNA transcription regulation

[a] PANTHER Classification System (https://panther.appliedbiosystems.com); [b] Not annotated in PANTHER Classification System.

8. Acknowledgments

I am very grateful to Prof. H. Hameister for giving me opportunity to complete my dissertation in the Department of Human Genetics, University of Ulm, for his support and help, academically and personally, during my Ph.D. studies. I really appreciate that he gave me possibility to attend international meetings.

I would like to thank Dr. Hildegard Kehrer-Sawatzki for being my adviser, for her useful advises, patience and sharing her knowledge in the field of human genetics. I am especially grateful for help in publishing my results.

I am grateful to the referees of my thesis as well as to the co-examiners for their attention and interest in my work.

I would like to thank Violaine Rosenstiel-Goidts and Dr. Matthias Kohn for their valuable help and contribution to this project.

I am very grateful to Antje Kollak, Helene Spöri and Stefanie Raith for technical support and help with experiments.

Many thanks to Manuel Lüdeke for help with PCR experiments.

I also would like to express my gratitude to Prof. Günter Assum and Prof. Walter Just for helpful comments and discussions.

I thank to Prof. W. Schempp from Freiburg and to Dr. Ulrich Zechner from Mainz for providing me with the chimpanzee and bonobo DNA samples, and cell lines.

I am grateful to Herbert Heinz for his help with printing this thesis.

Many thanks to Laimis Kucinskas for his friendship and a nice time, which we had working together in Department of Human Genetics.

My very special thanks to Dr. Josef Högel for his precious support and for help with statistical analyses, and to my beloved family in Poland, for their support during my stay in Germany.

9. Curriculum vitae

Personal Data

Surname	Szamalek
Name	Justyna Monika
Date of birth	15.01.1978
Citizenship	Polish
Work address	Department of Human Genetics, University of Ulm Albert-Einstein-Allee 11, 89081 Ulm, Germany

Education

Jul 2006 – Oct 2003	**Medical Faculty of University of Ulm, Germany** **Department of Human Genetics:** PhD thesis: "Cytogenetic and Molecular Characterization of the Macro- and Micro-inversions, which Distinguish the Human and Chimpanzee Karyotypes - From Speciation to Polymorphism"
Jul 2003 – Sep 2002	**Swiss Federal Institute of Technology, Zürich, Switzerland** **Institute of Animal Sciences and Breeding Biology:** Practical work in the molecular genetics laboratory as a postgraduate student
May 2002	**Agricultural University of Poznań, Poland:** M.Sc. thesis "Chromosomal localization of the keratin gene clusters in genomes of the red fox, arctic fox, Chinese raccoon dog and domestic dog"
May 2002 – Oct 2000	**Agricultural University of Poznań, Department of Genetics and Animal Breeding:** Practical course within the M.Sc. program in the molecular cytogenetics laboratory
May 2002 – Oct 1997	**Agricultural University of Poznań, Poland:** Studies in the Field of Biotechnology
Jun 1997 – Sep 1993	**Secondary School, Poznań, Poland:** Secondary School Certificate

Publications

Szamalek JM, Goidts V, Cooper DN, Hameister H, Kehrer-Sawatzki H: Characterization of the human lineage-specific pericentric inversion that distinguishes human chromosome 1 from the homologous chromosomes of the great apes. Hum Genet 120:126-138 (2006)

Goidts V, Armengol L, Schempp W, Conroy J, Nowak N, Müller S, Cooper DN, Estivill X, Enard W, **Szamalek JM**, Hameister H, Kehrer-Sawatzki H: Identification of large-scale human-specific copy number variations by inter-species array comparative genomic hybridization. Hum Genet 119:185-198 (2006)

Szamalek JM, Cooper DN, Schempp W, Minich P, Kohn M, Hoegel J, Goidts V, Hameister H, Kehrer-Sawatzki H: Polymorphic micro-inversions contribute to the genomic variability of humans and chimpanzees. Hum Genet 119: 103-112 (2006)

Szamalek JM, Goidts V, Searle JB, Cooper DN, 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 87:39-45 (2006)

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

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

Kehrer-Sawatzki H, **Szamalek JM**, 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 (2005)

Kehrer-Sawatzki H, Sandig C, Chuzhanova N, Goidts V, **Szamalek JM**, Tänzer S, Müller S, Platzer M, Cooper DN, Hameister H: Breakpoint analysis of the pericentric inversion

distinguishing human chromosome 4 from the homologous chromosome in the chimpanzee (*Pan troglodytes*) Hum Mutat 25: 45-55 (2005)

Goidts V, **Szamalek JM**, 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 115:116-122 (2004)

Szczerbal I, Switonski M, Michalak E, Rogalska-Niznik N, **Szamalek J**, Schelling C: FISH-mapping of the NOR and two putative keratin gene clusters on the cat (*Felis catus*) chromosomes. Journal of Heredity 94:107-110 (2003)

Szamalek JM, Szczerbal I, Rogalska-Niznik N, Switonski M, Ladon D, Schelling C: Chromosomal localization of two keratin gene families in the karyotype of three species of the family *Canidae*. Animal Genetics 33: 404-405 (2002)

Conference posters

Szamalek JM, Goidts V, Hameister H, Kehrer-Sawatzki H: Precise characterization of macro- and microchromosomal differences that distinguish human and chimpanzee genomes. HUGO Meeting; May 31 - Jun 3, Helsinki, Finland (2006)

Hameister H, **Szamalek JM**, Goidts V, Kehrer-Sawatzki H: Precise characterization of macro- and microchromosomal differences that distinguish human and chimpanzee genomes. The Biology of Genomes: May 10 - 14, Cold Spring Harbor, USA (2006)

Szamalek JM, Cooper DN, Schempp W, Minich P, Kohn M, Hoegel J, Goidts V, Hameister H, Kehrer-Sawatzki H: Polymorphic micro-inversions contribute to the genomic variability of humans and chimpanzees. Tagung der Deutschen Gesellschaft für Humangenetik; Mar 8 – 11, Heidelberg, Germany (2006)

Szamalek JM, Goidts V, Hameister H, Kehrer-Sawatzki H: Characterization of the human-specific pericentric inversion that discriminates human chromosome 1 from the homologous chromosomes in great apes. Tagung der Deutschen Gesellschaft für Humangenetik; Mar 8 – 11, Heidelberg, Germany (2006)

Szamalek JM, Goidts V, Assum G, Hoegel J, Hameister H, Kehrer-Sawatzki H: The contribution of multiple pericentric inversions to the human/chimpanzee speciation. 18th European course in medical genetics; May 15 - 21, Bertinoro, Italy (2005)

Goidts V, **Szamalek JM**, Armengol L, Schempp W, Conroy J, Enard W, Müller S, Hameister H, Kehrer-Sawatzki H: Identification of human-specific DNA copy-number differences compared to non-human primates as detected by BAC CGH-array. 18th European course in medical genetics; May 15 - 21, Bertinoro, Italy (2005)

Goidts V, **Szamalek JM**, Armengol L, Schempp W, Conroy J, Enard W, Müller S, Hameister H, Kehrer-Sawatzki H: Identification of human-specific DNA copy-number differences compared to non-human primates as detected by BAC CGH-array. Tagung der Deutschen Gesellschaft für Humangenetik; Mar 9 - 12, Halle/Saale, Germany (2005)

Szamalek JM, Goidts V, Assum G, Hoegel J, Hameister H, Kehrer-Sawatzki H: The contribution of multiple pericentric inversions to the human/chimpanzee speciation. Tagung der Deutschen Gesellschaft für Humangenetik; Mar 9 - 12, Halle/Saale, Germany (2005)

Goidts V, **Szamalek JM**, Sandig C, Chuzhanova N, Tänzer S, Müller S, Platzer M, Cooper DN, Hameister H, Kehrer-Sawatzki H: Precise characterization of pericentric inversions that distinguish human and chimpanzee genomes. Gordon Research Conference: Molecular cytogenetics; Jul 18 - 23, Oxford, United Kingdom (2004)

Szamalek JM, Goidts V, Hameister H, Kehrer-Sawatzki H: The nine pericentric inversions that distinguish the karyotypes of human and chimpanzee predate the separation of *Pan paniscus* (bonobo) and *Pan troglodytes* lineages. European Human Genetics Conference; Jun 12-15, Munich, Germany (2004)

Hameister H, Goidts V, **Szamalek JM**, Kehrer-Sawatzki H: Intrachromosomal recombination within pericentromeric HSATII repeats, the mechanism underlying the pericentric inversion of the chimpanzee chromosome 18, homologous to human chromosome 16. European Human Genetics Conference; Jun 12-15, Munich, Germany (2004)

Goidts V, **Szamalek JM**, 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. European Human Genetics Conference; Jun 12-15, Munich, Germany (2004)

Hameister H, Kohn M, Goidts V, **Szamalek JM**, Kehrer-Sawatzki H: Evolution in the light of genome conservation: sex chromosomes direct morphologic and functional diversification. Gordon Research Conference: Molecular evolution; Feb 1-6, Ventura, USA (2004)

Switonski M, Michalak E, Szczerbal J, Rogalska-Niżnik N, **Szamalek J**, Schelling C: FISH-mapping of NOR and two putative keratin gene clusters in the cat (*Felis catus*) karyotype. Advances in Canine and Feline Genomics; May 16-19, St. Louis, USA (2002)

Conference oral presentations

Hameister H, **Szamalek JM**, Goidts V, Kohn M, Assum G, Hoegel J, Kehrer-Sawatzki H: Looking back: From the human karyotype to the ancestral vertebrate proto-genome. Tagung der Deutschen Gesellschaft für Humangenetik, Mar 9 - 12, Halle/Saale, Germany (2005)

Szamalek JM., Goidts V, Sandig C, Chuzhanova N, Tänzer S, Müller S, Platzer M, Cooper DN, Hameister H, Kehrer-Sawatzki H: Precise characterization of pericentric inversions that distinguish human and chimpanzee genomes. 15th International Chromosome Conference; Sep 5 - 10, London, United Kingdom (2004)

Kehrer-Sawatzki H, Goidts V, **Szamalek JM**, Hameister H: Breakpoint analysis of pericentric inversions distinguishing human from chimpanzee chromosomes. HUGO Meeting: Primate Genomics Session; Apr 4-7, Berlin, Germany (2004)

Szamalek J, Szczerbal I, Rogalska-Niznik N, Michalak E, Ladon D, Schelling C, Switonski M: Chromosomal localisation of two keratin gene families in the karyotype of four species of the order Carnivora. 15th European Colloquium on Cytogenetics of Domestic Animals and Gene Mapping; Jun 2-4, Sorrento, Italy (2002)

Award

3rd award in the contest organized by The Michał Oczapowski Polish Society of Animal Production, for M.Sc. thesis entitled:
"Chromosomal localization of the keratin gene clusters in genomes of the red fox, arctic fox, Chinese raccoon dog and domestic dog"

Courses

18th Course in Medical Genetics, Bologna University Residential Centre – Bertinoro Di Romagna, May 15-21, 2005

Course in Cancer Epidemiology, 15th International Summer School of Epidemiology, July 5-9 2004, University of Ulm

Course in Genetic Epidemiology, 15th International Summer School of Epidemiology, July 5-9 2004, University of Ulm