Tomàs Marquès i Bonet

Tesi Doctoral 2006

UNIVERSITAT POMPEU FABRA

ON THE ASSOCIATION BETWEEN CHROMOSOMAL REARRANGEMENTS AND GENIC EVOLUTION IN MAMMALS

Memòria presentada per a optar al títol de Doctor per la Universitat Pompeu Fabra (UPF)

Treball dirigit pel Dr. Arcadi Navarro i Cuartiellas

Unitat de Biologia Evolutiva

Departament de Ciències Experimentals i de la Salut (CEXS)

Programa de Doctorat en Ciències de la Salut i de la Vida de la Universitat Pompeu Fabra

ARCADI NAVARRO I CUARTIELLAS TOMÀS MARQUÈS I BONET

Barcelona, desembre de 2006

Al pare, present sempre en cada insecte.

Pyrrhocoris apterus

Table of contents

Introduction

1. Variation in the genome

Genomes are not static, but constantly evolving entities that contain traces of history ready to be reconstructed. The study of the distributions of genomic variables and of the causes of these distributions is an extremely useful source of information that can help shedding light, not only on the evolution of genomes, but in other general biological processes, such as, for example, the genetic architecture of complex disease or mechanisms of speciation. The mammalian genome is a mosaic formed by fragments presenting different values of variables such as GC content, mutation rates, gene expression or even rates of sequence evolution (Paabo 2003). Although all these genomics variables are of different nature, they have in common that they are the result of accumulation of changes over time and, thus, comparative whole-genome analyses are needed to understand the evolution of genomes and, in particular, the dynamics of evolutionary factors affecting these mosaic variables. Here I present several complementary analysis of the genomic distribution of some of those genomic factors and I try to evaluate their individual contributions into the evolution of mammalian genomes, with an emphasis on humans.

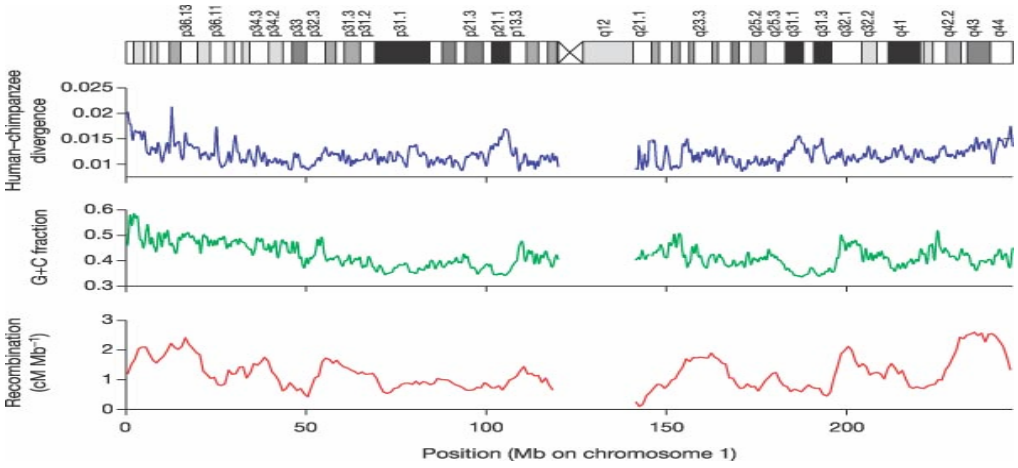


Figure 1. Variation in the Human Genome. This graphs represents divergence with chimpanzees, GC content or Recombination (Mikkelsen et al. 2005)

1.1. Variation in rates of molecular evolution

The controversial molecular clock hypothesis postulated by Zuckerlandl (1962) stated that the rate of evolution in a given protein or DNA sequence is approximately constant over time in all evolutionary lineages. Under the molecular clock hypothesis, all lineages in a phylogenetic tree should have accumulated substitutions at the same rate. Many years of research, however, have proved that this is not the case, and that substitution rates are dependent on many factors, including the metabolic rate in a species, its generation time, bottleneck events and selective pressures. Therefore, in comparative genomics and specially when focusing on short evolutionary times, such as the time separating humans and chimpanzees, the molecular clock does not hold, and great variation in substitution rates is observed.

1.1.1. Variation between lineages

There is no homogeneity in evolutionary rates among lineages, as shown by plenty of examples. For instance, rodents show acceleration in their evolutionary rates compared with primates or carnivores (Mouse Consortium 2002; Gibbs et al. 2004; Li and Wu 1987; Lindblad-Toh et al. 2005). This was first suggested in 1969 (Laird et al. 1969) and although subsequent results were controversial (Kumar and Subramanian 2002), the acceleration if now accepted and the consensus view is that it is related to the biology of organisms. Among suggested causes are generation-time effects, changes in metabolic rates or even correlations with body sizes (or maybe just a unique effect as some of the aforementioned causes are in fact correlated) (Li et al. 1996; Martin and Palumbi 1993).

Another classical example of heterogeneity in lineage specific evolutionary rates comes from arms races in virus. Virus have one of the highest rates of nucleotide

substitutions (Gojobori et al. 1990; Li et al. 1988). Their genomes evolve several million times faster than the eukaryotic and prokaryotic genomes as a result of a higher mutation rate (10^{-3} per site per year in virus compared with 10^{-9} or 10^{-10} per site per year for eukaryotes) and their short generation time (1.2-2.6 days for HIV-1). These high rates are an obvious obstacle for effective antiviral immunity responses.

1.1.2. Variation across the genome

Mutational rates are also variable across the genome (Wolfe et al. 1989). Recently completed mammalian genomes have evidenced, for example, that the rate of substitution varies among genes and is correlated with their base composition and their flanking DNA (Wolfe et al. 1989). The variation in both substitution rate and base composition can be attributed to systematic differences in the rate and pattern of mutation over regions of the genome. It has been proposed that the differences arise because mutation patterns vary with GC content, recombination, duplications, selective effects or the timing of replication of different chromosomal regions. These factors will be discussed with detail below.

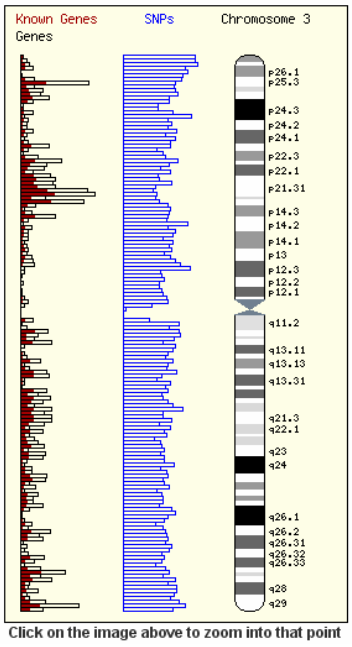


Figure 2. Gene content and SNP variation across Human Chromosome 3 (www.ensembl.org)

1.1.3. Variation among genes

Genes are one of the most striking examples of variation within genomes. The analyses of coding sequences has unveiled great variability among proteins in their rates of non-synonymous substitution (nucleotide substitutions that change the amino acid structure of the encoded protein). Variability among genes is also great, albeit

lower, for synonymous substitution rates (substitutions that, due to the degeneracy of the universal genetical code, do not change the protein sequence). For instance, non-synonymous substitutions can vary as much as three orders of magnitude between genes. Some proteins, for example, are extremely conserved between human and drosophila (e.g. Histones or Ubiquitin), while others differ at intermediate rates (e.g. Erythropoietin) or at very high rates (e.g. Relaxin). To explain this large variation in rates among genes, two main factors are to be considered: mutation rates and intensity of selection. These sources of variation will be also discussed below.

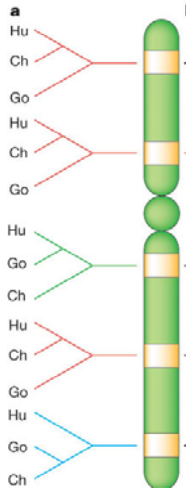


Figure 3. This schematic chromosome shows that different human genes can have different phylogenetic relationships with chimpanzee and gorilla (Paabo 2003).

1.1.4. Variation within genes

Rates of nucleotide substitution differ not only among genes but also in different regions within the same gene. On average, the fastest rate of substitution occurs in the non-coding parts of genes, whereas coding sequences are usually conserved, suggesting various degrees of selective constraints (Hartl and Clark 1998). Moreover, 5' and 3' untranslated regions have lower substitution rates than the rate of synonymous substitutions in coding regions, revealing some functional constraints in these regions. Indeed, flanking untranslated regions are known to contain important signals for the regulation of the translation process (Li 1997). Besides selective

pressures, mutation rates are also different when comparing translated and untranslated parts of the gene (Subramanian and Kumar 2003).

Finally, not all positions within a single codon present the same substitution rates. Because of the structure of the genetic code, synonymous changes occur mainly at the third position of codons. In contrast, all the nucleotide changes in the second position are non-synonymous, and it is the same for the majority of nucleotide changes at the first position (Li 1997). As a result, the three positions within a codon are subject to different pressures and, thus, evolve at different rates.

1.2. Sources of variation in evolutionary rates

The rate of substitutions between two sequences is determined by two factors: rate of mutation and probability of fixation of those mutations. The latter depends on whether the mutation is advantageous, neutral o deleterious. We will briefly discuss some of the factors that model evolutionary rates in mammals.

1.2.1. Variations in evolutionary time

Speciation is a complex subject that will be properly addressed latter on. I will only mention here an issue relevant to variation in rates of substitution: modes of speciation. It is accepted that different kinds of speciation events leave different traces in the genomes. For instance, the allopatric model of speciation predicts that expectation of divergence time will be uniform across genomic regions. Under this model, a geographical barrier would prevent gene flow among different subpopulations within a species. Then, if speciation occurred instantaneously and ignoring ancestral polymorphism, we should expect homogeneous substitutions rates along the genome.

Alternatively, under a genic view of speciation and in models of parapatric speciation (in which nascent species may still exchange genes), reproductive isolation would be initiated by differential adaptation in small parts of the genome (i.e., ''speciation genes,'' Wu and Ting (2004) for a review). Then, those parts would have gene flow restricted while other parts of the genome would freely exchange genetic information (Wu 2001). If so, genetic isolation in speciation genes and their flanking regions would be achieved earlier than in other parts of the genome, making the variance of divergence time across the genome larger than than predicted under the allopatric model. Successful examples of the application of those predictions are seen in *Drosophila* species (Machado et al. 2002; Wang et al. 1997).

1.2.2. Effective Population size

Since not all the individuals in a population take part in reproduction, the population size that matters in evolutionary processes is different from the census size. Wright (1931) introduced the concept of effective population size (Ne), which was defined as the size of an idealized population that would have the same effect of random sampling on allele frequencies as that of the actual population.

The neutral theory of evolution (Kimura 1983) predicts that rates and patterns of molecular evolution will be influenced by effective population size (Ne). In small populations, slightly deleterious mutations are expected to drift to fixation because they are strongly affected by stochastic fluctuations in allele frequencies, so drift can overpower selection for alleles with small selection coefficients. Therefore, the fixation of ''nearly neutral'' alleles by drift is expected to be the greatest in small populations (Kimura 1983; Ohta 1987). If a substantial proportion of mutations are nearly neutral, overall substitution rates should increase in species with small Ne compared to those with larger Ne, provided mutation rates are similar in both

species (Ota 1971). The ratio of nonsynonymous to synonymous substitution rates (ω) is also predicted to increase in these species (Ohta 1993), as many non-synonymous mutations are expected to be nearly neutral, while synonymous mutations are more likely to be neutral and thus fix at a rate unaffected by Ne.

Few empirical studies have directly tested these predictions, partly because of the difficulty of identifying species which differ only in effective population size. However, it has been found that endosymbiotic species of bacteria and fungi with small effective population sizes have significantly higher substitution rates and ω values than their free-living relatives (Woolfit and Bromham 2003). Eusocial species (those with reproductive division of labor) have been found to have faster rates of molecular evolution than their nonsocial relatives because of greatly reduced effective population size (Schmitz and Moritz 1998). Finally, comparisons of rates of evolution in primates and rodents support these predictions (Weinreich 2001; Wu and Li 1985) but these lineages differ in many other aspects of their biology which could affect substitution rates, such as generation time, metabolic rate and DNA repair mechanisms (Bromham et al. 1996).

1.2.3. Duplications

Recently, gene and genome duplications have been recognized as a prominent factor in the evolution of eukaryotes (Ohno; Taylor and Raes 2004). The most obvious contribution of gene duplication to evolution is providing new genetic material for mutation, drift, and selection to act upon, making new evolutionary opportunities possible. It was therefore predicted that evolutionary rates should increase following duplications (Ohno 1970). Indeed, this has been repeatedly observed in both DNA sequence and gene expression levels (Gu et al. 2002; Kondrashov et al. 2002; Lynch

and Conery 2000a; Marques-Bonet et al. 2004; Marques-Bonet and Navarro 2005; Marques-Bonet et al. 2006b). Several factors can account for that alteration in divergence. Gene duplications can lead to a change in gene dosage, which in dosage-sensitive genes will influence the carrier phenotype. On the other hand, duplications can also lead to new gene forms or novel fusion genes, although most of such events are likely to be non-functional unless an open reading frame is maintained (Sharp et al. 2006).

Rate accelerations may be due to either a relaxation of purifying selection in one or both gene duplicates or to the action of positive or diversifying selection between duplicates (Conant and Wagner 2003; Zhang et al. 2003). Interestingly, it has been also suggested that after an initial increase in substitution rates, duplicated genes evolve slower due to their newly adquired functional constraints (Jordan et al. 2004).

Alternatively to strong differentiation, in some cases duplicated genes are organized in gene families. A gene family is a set of genes defined by homology almost certainly formed by duplication of an ancestral gene, and they have a recognizably similar sequence and some times function. They are prone to suffer gene conversions events, which constitute an homogenization force arising the from non-reciprocal transfer of genetic information. When genes with high identity are repeated on the chromosome, gene conversion occurs among them with high frequency in the course of evolution and, therefore, is one of the major mechanisms for the observed homogeneity of sequences in some multigene families (Ohta 1984). This is one of the causes explaining with gene family members would have similar sequences, thus generating underestimations of evolutionary rates.

In summary, duplications have been convincingly related with several forms of evolutionary rates. Given this and the fact that they are non-uniformly distributed all along the genome, heterogeneity in evolutionary rates across the genomes of mammalian species is expected.

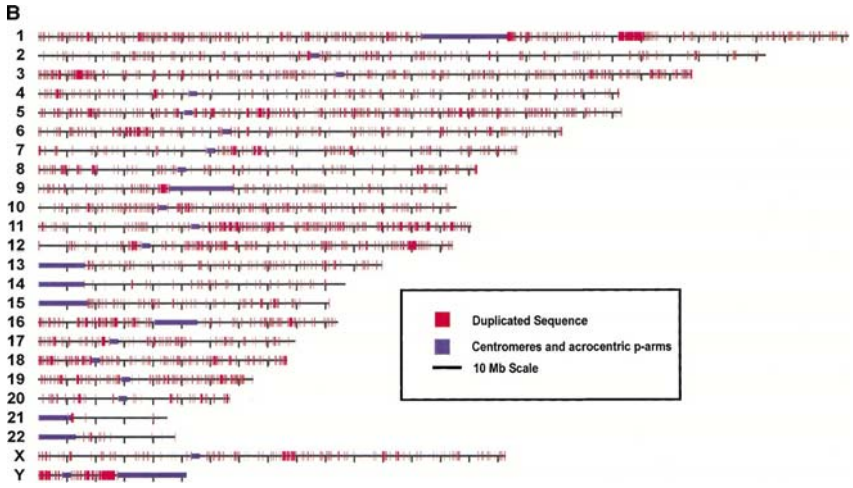


Figure 4. Distribution of segmental duplications in the human genome. (Bailey et al. 2001)

1.2.4. GC content and CpG islands

The analysis of the human genome unveiled long-range variability of GC content (Human Consortium 2001). Contrarily to the view that was predominant until the early 1970s, base composition heterogeneity in the genomes is continuous and not discontinuous. Initially, Bernadi and colleagues proposed a model of organization in mammalian genomes consisting in a mosaic of compositionally homogeneous regions of GC content called isochores (Bernardi 2000). The complete sequence of the human genome modified that initial picture and it is now assumed that the GC content varies continuously (that is, that there is no clear boundaries of that isochores) (Nekrutenko and Li 2000).

In mammalian genomes, the palindromic dinucleotide CpG is usually methylated on the cytosine residue. Methyl-CpG is mutated by deamination to TpG, leading to approximately fivefold underrepresentation of CpG in the human and mouse genomes

comparison (Mouse Consortium 2002). However, in some regions of the genome that have been implicated in gene regulation, CpG dinucleotides are not methylated and thus are not subject to deamination and mutation. Such regions, termed CpG islands, are usually a few hundred nucleotides in length, have high GC content and above average representation of CpG dinucleotides.

Nucleotide substitution rates at neutral sites are greatly affected by this hypermutable regions (Cooper and Youssoufian 1988). These rates are also strongly correlated with the GC fraction (Mouse Consortium 2002; Hardison et al. 2003). This leads to an uneven distribution of mutation rates among genomic regions with different GC content.

1.2.5. Functional constraints and selection

Although some amino acid substitutions may be functionally equivalent or nearly equivalent, there are many more substitutions that are expected to affect protein function to such an extent that they reduce the fitness of the organisms. Similarly, some mutations that are considered neutral substitutions (silent or synonymous) may also be under selective constraints (Hartl and Clark 1998). Some synonymous substitutions alter splicing sites, while others affect the secondary structure of the RNA, specially through codon preferences, which correlate with the relative abundance of tRNA molecules that interact with codons.

One general conclusion of molecular evolution studies is that the stronger the functional constraints, the slower the rate of nucleotide substitution will be. This is supported by numerous comparative studies of both protein and DNA sequences which show an inverse relationship between stringency of functional constraint, or importance, on the one hand and the rate of evolution on the other. Therefore,

molecules or parts of a molecule that functionally less important evolve faster than more important ones. This rule is frequently used to reverse-engineer the genome, and rates of nucleotide substitutions are used to infer the stringency of structural and functional constraints in particular sequences (Mouse Consortium 2002; Gibbs et al. 2004; O'Brien et al. 1999).

1.2.6. Mutational effects of recombination

Recombination is another factor that might cause regional variations in mutation rates because it has been suggested to be mutagenic (Hellmann et al. 2003a; Kong et al. 2002a; Wolfe and Li 2003) and its rate varies along the genome. However, other authors have argued that the finding that recombination rates covary with neutral mutation rates is not caused by mutagenic effects but that the two phenomena are linked to a third and as yet unknown factor (Hardison et al. 2003). Moreover, in primate and rodent genomes, local recombination rates are also positively correlated with the GC fraction (Eyrewalker 1993; Hardison et al. 2003; Lander et al. 2001), and thus this could lead to increases on mutation rate just by means of base composition effects.

It has also been suggested that increased recombination rates drive the elevation of GC fractions (Hardison et al. 2003) maybe as a result of “biased gene conversion” mechanisms (Eyrewalker 1993; Marais 2003). In summary, the relationship among recombination and divergence is a complex issue that has not been fully resolved but, whatever it is, recombination can be an important source of variation for mutation rates.

1.3. Variation in gene expression

Mutations can alter the phenotype not only by changing aminoacids in coding sequences but also by changing regulatory DNA sequences that control transcription, translation or transcript degradation. As a result, gene expression levels are also highly heterogeneous across the genome.

If gene expression was evolving neutrally, expression divergence should increase with time and evolutionary divergence (Khaitovich et al. 2006). Then, differences among different lineages should accumulate approximately linearly with time. This has been observed in various species such as fruitflies or primates (Rifkin et al. 2003; Whitehead and Crawford 2006). Gene expression is also subject to local genomic influences in the same way as DNA sequence. As an example, both gene expression and DNA sequence divergence between humans and chimpanzees, have been found to be higher in sub-telomeric regions (Marques-Bonet et al. 2004).

Besides the aforementioned causes of regional variation in evolutionary rates (that also apply to gene expression divergence), gene expression can also be affected by specific factors such as developmental stage, environment or differences across and within different tissues.

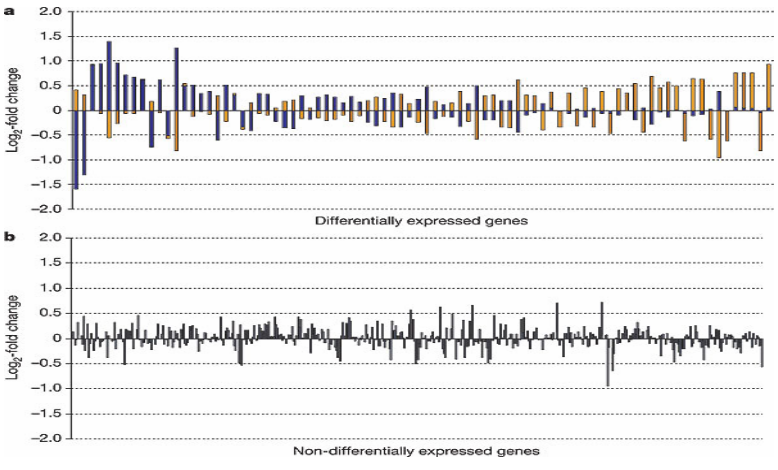


Figure 5. An example of up and down regulation of human and chimpanzee gene expression patterns (Gilad et al. 2006)

2. Inferences of rates of molecular evolution in coding sequences.

DNA sequences do not remain constant over time and, as the process of nucleotide substitutions is usually extremely slow, complex inference methods have been devised to retrieve information about it (Li 1997). This is not an easy task (although it is usually assumed that it is) because when obtaining a reliable estimate of those rates, the degree of divergence among sequences to analyze should be neither too small, to avoid large stochastic effects, nor too large, because the estimation of rates would be unreliable due to the difficulties that correcting for multiple substitutions presents.

In the analyses of coding sequences, it is important to discern between synonymous and non- synonymous substitutions. The synonymous substitution rate (Ks) is the number of synonymous substitutions per synonymous site, while the non-synonymous substitution rate (Ka) refers to nucleotide changes in the coding sequence that modify the amino acid per non-synonymous site. We can use these two rates to infer events of selection in the evolution of those sequences. The ratio of the two rates (ω = Ka/Ks) is a measure of the selective pressure at the protein level. If selection has no effect on fitness, non-synonymous mutations will be fixed approximately at the same rate than synonymous mutations, so we will have Ka ~ Ks, and thus ω will tend to be close to 1. Alternatively, if negative selection is strong over that coding sequence (for instance in housekeeping genes), non-synonymous mutations are likely to be deleterious and see their fixation rates reduced. In that case, Ka would be much lower than Ks and, therefore, ω would be lower than 1. Finally, in some circumstancew, if recurrent positive (or adaptative) selection has affected aminoacid changes along a lineage, non-synonymous mutation will be favored and then Ka may

be greater than Ks and ω greater than 1. A non-synonymous rate significantly higher than the synonymous rate of the same gene is generally accepted as evidence of recurrent adaptative selection at the molecular level (Balding et al. 2003). This criterium, although stringent, has been successfully used to identify positive selection in many cases in different organisms (Bonhoeffer et al. 1995; Hughes and Nei 1988; Messier and Stewart 1997).

The ω ratio is a unique measure for the whole coding sequence because it comes from an average of all codons. It is difficult to detect positive selection using it, unless many consecutive positive selection events would have left very strong signals on the whole sequence (Sharp 1997). Recently, many technical improvements in the basic methodology to detect selection at molecular levels have been achieved. We can now, by means of maximum likelihood models, statistically compare two ratios, and then be able to discern among different evolutionary scenarios (Yang 1997). For example, we can detect recurrent adaptative evolution in a single codon (Yang et al. 2000), in a single branch of a phylogenetic tree (Yang 1998), or even in a specific single codon in a given branch (Yang and Nielsen 2002). In the future, improvements on these statistical methods will greatly help us to detect positive selection in DNA sequences.

3. Chromosomal rearrangements

Chromosomal rearrangements are also relatively common mutations that occur in eukaryotic organisms. They occur when a substantial track of DNA is inverted or repositioned on a chromosome. If a chromosome segment between two breakpoints becomes inverted with the result that the gene order for the segment is reversed relative to its original order, the rearrangement is called an inversion. There are two types of inversions: pericentric and paracentric. In the former, the inverted segment includes the centromere. A translocation occurs when a piece of one chromosome breaks off and attaches elsewhere in the genome, usually to another chromosome.

The development of modern techniques for the study of chromosomes has made it possible to obtain accurate comparisons of chromosomes in various species of primates (Goidts et al. 2004; Kehrer-Sawatzki et al. 2005a; Kehrer-Sawatzki et al. 2005b; Kehrer-Sawatzki et al. 2002; Kehrer-Sawatzki et al. 2005c; Locke et al. 2003), so we now have detailed maps of how chromosomal structure has evolved in out lineage.

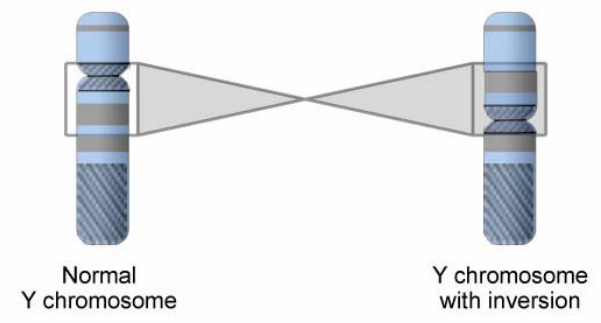


Figure 6. An schematic view of a pericentric inversion in Y chromosome.

3.1 Effects over evolutionary rates

Chromosomal rearrangements are complex to analyze in the context of rates of molecular evolution, because they have been related with many of the factors known to affect substitution rates

3.1.1. Changes in genomic context

Chromosomal rearrangements may alter the genomic context of linked genes. Repositioning fragments of the genome could affect at least three of the sources of variation of substitution rates. If genes trapped within rearrangements are moved from one region to another with a different recombinational landscape, changes in their substitution rates may be triggered (Kong et al. 2002a; Myers et al. 2005).

Flanking base composition can affect evolutionary rates (Hardison et al. 2003; Williams and Hurst 2000). If so, movement of genes into a new environment can result in the nucleotide composition of the genes changing to fit their new location (Kumar and Subramanian 2002). If one gene, for example, is moved from a low GC region to a higher new zone, the base composition context of this new regions could lead to increases in mutation rates, and most probably to increases in substitution rates. Besides regional mutation rates, regional selective pressures can affect genes moving to a new position unless recombination is strong enough to allow for independent evolution of the newly inserted gene. If recombination rates are low, any selective processes affecting a region will reduce variability and may protentially affect substitution rates. Background selection, for example, reduces effective population size and, thus, may modify the efficiency of selection on the newly inserted gene.

3.1.2. Changes in expression patterns

Chromosomal reorganizations have direct effects over gene expression patterns. There is experimental evidence that rearrangements can induce changes in the expression patterns of genes located around breakpoints (Spitz et al. 2003; Tanimoto et al. 1999). Changes of expression could result from a disruption of the enhancer or promoter elements such as seen at the red/green pigment genes (Deeb 2005). Spitz et al. (2003) also found that a large inversion separating the enhancer from the gene cluster, induced a downregulation of Hoxd genes expression in limb development. Chromosomal rearrangements can also disrupt or change the structure of genes themselves by moving genes to a different region of the chromosome in which they could not be transcriptionally active. Finally, translocations of regions next to heterochromatine can lead to changes in expression due to the propagation of the inactive chromatine structure into the genes (Bedell et al. 1996; Kleinjan and van Heyningen 1998). All these processes result on an apparent change of the phenotype, thus modifying the functional constraint of the initial gene conformation. If so, the effect of the rearrangements over gene expression could also modify evolutionary rates.

3.1.3. Implications in speciation

Chromosome rearrangements may play a role in speciation. There are many theoretically possible but untested models under which chromosomal rearrangements accelerate genic diversification between populations and, therefore, facilitate speciation (King 1993). Here I will consider two classes of models, the "hybrid-dysfunction" and "suppressed-recombination" models of speciation (Ayala and Coluzzi 2005).

Most of the classical models of chromosomal speciation have in common that chromosomal differences that have accumulated between the neospecies and its progenitors are assumed to impair the fertility or viability of hybrids, reducing gene flow as a consequence of this impairment (Rieseberg 2001). In fact, hybrid-dysfunction models claim that recombination between rearranged chromosomes generates unbalanced gametes with problematic chromosomal segments, and thus creates a partial reproductive barrier because the heterokaryotypic hybrid exhibits reduced reproductive fitness ("underdominance"). Under these conditions, natural selection will, in both populations, favor mutations that reduce the probability of intercrossing and will lead to complete reproductive isolation.

An ideal and simplified succession of events would be as follows. A chromosomal rearrangement would first become established in a small sub-population, either at the periphery of the distribution area or inside of the ancestral species, by random drift. Individuals carrying this alternative chromosomal form may expand within a certain area and there displace the ancestral form if its members display high fitness in that local environment. If hybrids do show low fitness (because of the unbalanced gamete production or another mechanism), this will keep the two populations separate and facilitate the evolution of prezygotic isolating mechanisms, which will inhibit the formation of hybrids. Usually this model is also named as "stasipatric speciation model" (White 1968; White 1978).

However, these models have been controversial and are subject to a strong paradox (Spirito 2000). If underdominance were strong, it would be very unlikely that different arrangements could fix in different sub-populations, unless they would be established in small inbred populations. On the other hand, if underdominance were weak enough for fixation to be likely, chromosomal rearrangements would be very

poor barriers to gene flow and, thus, unlikely to contribute to speciation. In other words, only rearrangements that are strongly underdominant are considered likely to contribute to speciation, but these kinds of rearrangements are exceedingly difficult to fix in natural populations.

Alternatively, the ''suppressed-recombination'' models of speciation are based on the fact that chromosome rearrangements act as agents delaying fixation of alleles between populations. Mutations associated with the rearranged chromosomes cannot flow from one to another subpopulation, whereas genetic exchange will freely occur between colinear chromosomes. The fundamental difference between these new models and previous ones is that they try to escape the underdominance paradox by changing their focus from semiesterility to suppression of recombination in heterokaryotypes. Suppressed-recombination models of speciation have been recently proposed by Rieseberg (Rieseberg 2001), Noor et al. (Noor et al. 2001a; Noor et al. 2001b), Machado et al. (Machado et al. 2002) and Navarro and Barton (Navarro and Barton 2003a; Navarro and Barton 2003b).

In the model by Navarro and Barton (2003a), the scenario proposed does not require any hybrid infertility. Although some selection must be maintaining different frequencies of the arrangement in different locations, the rearrangements themselves may not be selected. Genetic barriers (in this case chromosomal rearrangements) decrease migration between populations, thus delaying the fixation of alleles and allowing for differentiation between populations. The stronger the barrier, the lower is effective migration and the longer is the delay in fixation of alleles in the whole population. This is the key parameter of the model, rearrangements do not promote speciation by themselves, but they play the key role of delaying the establishment of adaptive alleles located within a given

rearrangement into the whole population. The effects of those genetic barriers would be especially strong if divergence is through the accumulation of incompatible alleles, as proposed by Bateson, Dobzhansky, and Muller (Orr 1997). The spread of new favorable alleles through the whole population will be delayed if they are linked to chromosomal differences that are already established within that species. This gives different alleles time to spread through the rest of the species range, thus after new alleles have been fixed at the first locus in one subpopulation and in the second locus in the other. These new alleles could prove to be incompatible upon secondary contact of the two populations, because they were never tested together before their fixation, and thus even with a secondary contact, the genetic background of those two subpopulations will most likely never be mixed again and thus they will strength reproductive isolation. The accumulation of incompatibilities facilitated by chromosomal differences generates genetic barriers of growing strength that, eventually, produce complete reproductive isolation and, therefore, speciation.

In summary, a simplified version the sequence of events in this new model would be as follows. Chromosomal rearrangements occur as a result of random mutation and spread though a part the population, because they have different fitness in different subpopulations or just by neutral genetic drift. The suppression of recombination in homologous chromosomes with different conformations isolates the genetic material on those chromosomes, and thus, each chromosome evolves independently (and new genetic variants are selected into the new genomic environment) as it were in a separate non-interbreeding species. Over time, more mutations become established differentially in both kinds of chromosomes because the lack of gene flow in the rearranged chromosomes leads to more and more divergent evolution. This will eventually result in sexual incompatibility in secondary contacts among populations that carry and do not carry the rearrangements and speciation will be completed.

Several studies in many organisms (fungi, plants and animals) have provided evidences for the new class of chromosomal speciation models (Delneri et al. 2003; Machado et al. 2002; Noor et al. 2001b; Rieseberg 2001; Rieseberg et al. 1999). For instance, in the study by Noor et al. (2001b) they show that genetic loci responsible for hybrid male sterility in *Drosophila pseudoobscura* and *D. persimilis*, two occasionally hybridizing species, are located in rearranged chromosomal regions. In addition, a multilocus analysis of the same species pair also indicated reduced gene flow in rearranged chromosome regions (Machado et al. 2002). As to plant evolution, it has been shown that the rates of introgression in colinear chromosomes between sunflowers *Helianthus petiolaris* and *H. annuus* are double than introgression rates in rearranged chromosomes (Rieseberg et al. 1999). In the yeast species *Saccharomyces cerevisiae* and *S. mikatae*, hybridization produces sterile offpring, as they have reciprocal translocations involving three chromosomes. Delneri et al. (2003) engineered the chromosomes of *S.cerevisiae* to make them colinear with those of *S. mikatae* and interestingly, the interspecific hybrids are now fertile. In mosquitoes, a suppressed-recombination model of speciation was proponed by Coluzzi to account for the speciation patterns shown in the *A. gambiae* species. The currently available evidence is limited but consistent with the model (Besansky et al. 2003; Coluzzi et al. 2002). Finally, in mammals, it has been found that an european group of *Sorex* populations (Order: Insectivora) are parapatric and form hybrid zones. Variability in this species is related with the chromosome composition of those populations, which are differentiated by fusions of a subset of acrocentric chromosomes. In the work by Basset (2006) about Sorex hybrid zones, a higher degree of genetic structure across rearranged chromosomes than across colinear chromosomes was reported.

Predictions of suppressed-recombination models

If it is true that chromosomal rearrangements occurred before the speciation process was complete (or even triggered it) and that heterozygous hybrids interbred fertilely for some time, these process may have left traces in our genomes.

During the interbreeding period between subpopulations with different chromosome conformations, gene flow can freely occur on colinear chromosomes. However, gene flow would be restricted on those chromosomes that have undergone a rearrangement. Therefore, if there is a significantly long period when interbreeding between individuals with and without the chromosomal rearrangement occur, then it might be possible to see greater degree of neutral divergent evolutionary mutations on those chromosomes that had undergone rearrangements compared with colinear chromosomes. Since gene flow is not restricted in colinear chromosomes, beneficial mutations (or neutral mutations linked to beneficial mutations through a selective sweep) will tend to fix in the entire population to the same extent that they would do in any single species. On the other hand, since there is no gene flow in the rearranged chromosomes, mutations that fix in one version of the chromosome cannot spread to the other version and so every fixed mutation in either case results in a difference (*i.e.*, results in divergence) between original and rearranged chromosomes.

Furthermore, if favorable mutations, which drive the divergence of the species, accumulate on those rearranged chromosomes signatures of the action of positive selection may be found on those chromosomes. This prediction is particular of the model by Navarro and Barton, involving incompatibilities fixed by positive selection. One way to detect positive selection is by means of ω. Then, three potential observations are possible if the model is true. First, overall divergence should be higher in rearranged chromosomes. Second, either higher average ω and/or an

excess of genes with $\omega >> 1$ should be found in rearranged chromosomes relative to colinear chromosomes. Finally, if chromosomal rearrangements have participated in any speciation event by means of the suppressed recombination model, then selective sweeps produced by the fixation of such favorable alleles should also have decreased neutral polymorphism in these chromosomes. I should underline that this effect should be only be detectable if the fixation of the allele is recent enough (Navarro and Barton 2003a; Navarro and Barton 2003b).

However, several questions have been raised against this new model. Ayala and Coluzzi (2005) argued that based on the model and its implication in human-chimpanzee speciation, the period in which chromosomal rearrangements should have participated in the speciation should not last very long because of the snow-ball effect produced by incompatible alleles. Then, as the speciation between human and chimps did happen millions of years ago, it is likely to think that any trace left by the chromosomal speciation would be erased by other processes that would have largely contributed to the differentiation of both genomes in the anagenic process. Moreover, as the participation of all rearranged chromosomes in the speciation event is also quite unlikely, several confounding factors are expected when analyzing all the rearrangements, then masking the hypothetical speciation-related traces. In addition, Hey (2003) and based on the analyses of Navarro and Barton (2003b) underlined several criticism not to the theory of speciation but about the interpretation of the results obtained on that paper. In short, the main point of the criticism was absence of and outstandingly higher Ks associated with rearranged chromosomes as predicted by the model. There have been other papers suggesting other interpretations and reanalyzes of the initial analyses performed by Navarro and Barton (2003b), but they will be broadly commented in the discussion.

As a summary of this introduction, I just want to stress that the relationship among chromosomal rearrangements and genic evolution is a complex issue, most likely the result of a combination of many factors and processes acting simultaneously over time. Here, I present several studies trying to understand the non-uniform distribution of evolutionary rates and gene expression in the mammalian genome, the relative contribution of some of the potential causes will be discussed.

Objectives

The main objectives of this work are:

a) **To test the predictions of suppressed-recombination chromosomal speciation models on two different lineages of mammals: rodents and primates.**

Suppressed-recombination chromosomal speciation is still quite elusive as a mode of speciation in mammals. Experimental results are scarce and the first objective of this work is to analyze whole-genome data looking for traces of events of chromosomal speciation. Rodent and primate lineages were chosen for this search, not just because of their particular biological and cytological characteristics, which make them good candidates to have speciated by this mechanism, but also because they were the first mammalian organisms to be fully sequenced.

b) **To study the effects of chromosomal rearrangements on genic evolutionary rates.**

As have been seen in the introduction, there are many of potential interactions among chromosomal rearrangements and evolutionary rates, so the second goal of this work was to try to understand the impact of chromosomal rearrangements over substitution rates by means of other mechanisms not related with speciation.

c) To distinguish individual contributions of different genomic factors in the potential association among chromosomal rearrangements and evolutionary rates.

The third main goal of this thesis was to discern among the different factors that could be explaining the many associations between chromosomal and genic evolution that were detected in different studies.

Results

Response to Comment on“Chromosomal Speciation and Molecular Divergence—Accelerated Evolution in Rearranged Chromosomes”.

Science, 2003. 302(5647).

Arcadi Navarro,Tomàs Marquès-Bonet, N. H. Barton

Arcadi Navarro, Tomàs Marquès-Bonet, and N. H. Barton
Response to Comment on "Chromosomal Speciation and Molecular Divergence-Accelerated Evolution in Rearranged Chromosomes"
Science 7 November 2003 302: 988
[DOI: 10.1126/science.1090460] (in Technical Comments)

Chromosomal rearrangements and the genomic distribution of gene-expression divergence in humans and chimpanzees.

Trends in Genetics, 2004 Nov;20(11):524-9

Tomàs Marquès-Bonet, Mario Cáceres, Jaume Bertranpetit, Todd M. Preuss, James W. Thomas and Arcadi Navarro

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 in genetics. 2004 Nov;20(11):524-9.

Chromosomal rearrangements are associated with higher rates of molecular evolution in mammals.

Gene 2005 Jul 4;353(2):147-54.

Tomàs Marquès-Bonet, Arcadi Navarro

Marques-Bonet T, Navarro A.
Chromosomal rearrangements are associated with higher rates of molecular evolution in mammals.
Gene. 2005 Jul 4;353(2):147-54

Murine segmental duplications are hot spots for chromosome and gene evolution.

Genomics, Volume 86, Issue 6, 2005, Pages 692-700.

Lluís Armengol, Tomàs Marquès-Bonet, Joseph Cheung, Razi Khaja, Juan R. González, Stephen W. Scherer, Arcadi Navarro, Xavier Estivill

Armengol L, Marques-Bonet T, Cheung J, Khaja R, Gonzalez JR, Scherer SW, Navarro A, Estivill X.
Murine segmental duplications are hot spots for chromosome and gene evolution.
Genomics. 2005 Dec;86(6):692-700. Epub 2005 Oct 26.

On the association between chromosomal rearrangements and genic evolution in humans and chimpanzees.

Genome Biology (submitted to Genome Biology)

Tomàs Marquès-Bonet, Jesús Sànchez-Ruiz, Lluís Armengol, Razi Khaja, Jaume Bertranpetit, Mariano Rocchi, Elodie Gazave, Arcadi Navarro

Title: On the association between chromosomal rearrangements and genic evolution in humans and chimpanzees.

Authors: Tomàs Marquès-Bonet[1], Jesús Sànchez-Ruiz[1], Lluís Armengol[2], Razi Khaja[3], Jaume Bertranpetit[1], Mariano Rocchi[4], Elodie Gazave[1], Arcadi Navarro[5].

Addresses:

[1] Unitat de Biologia Evolutiva Departament de Ciències Experimentals i de la Salut. Universitat Pompeu Fabra. Barcelona, Catalonia, Spain.

[2] Genes and Disease Program, Center for Genomic Regulation, Barcelona Biomedical Research Park, Barcelona, Catalonia, Spain.

[3] The Center for Applied Genomics, Toronto, Canada

[4] Dipartimento di Genetica e Microbiologia. Universita di Bari. Bari, Italy.

[5] Institucio Catalana de Recerca i Estudis Avancats (ICREA) and Unitat de Biologia Evolutiva, Departament de Ciències Experimentals i de la Salut. Universitat Pompeu Fabra. Barcelona, Catalonia, Spain.

Corresponding author:

Arcadi Navarro
Institucio Catalana de Recerca i Estudis Avancats (ICREA)
Unitat de Biologia Evolutiva.
Departament de Ciències Experimentals i de la Salut
Universitat Pompeu Fabra
Doctor Aiguader 80
08003 Barcelona (Spain)
Phone: +34 93 5422841
Fax: +34 93 5422802
E-mail: arcadi.navarro@upf.edu

ABSTRACT

Background: The role that chromosomal rearrangements might have played in the speciation processes that have separated the lineages of humans and chimpanzees has recently come into the spotlight. To date, however, results are contradictory. Here we revisit this issue by making use of the available human and chimpanzee genome sequence to study the relationship between chromosomal rearrangements and rates of DNA sequence evolution.

Results: Contrary to previous findings for this pair of species we show that genes located in the rearranged chromosomes that differenciate the genomes of humans and chimpanzees, specially genes within rearrangements themselves, present lower divergence than genes elsewhere in the genome. Still, there are considerable differences between individual chromosomes. Chromosome 4, in particular, presents higher divergence in genes located within its rearrangement.

Conclusions: A first conclusion of our analysis is that divergence is lower for genes located in rearranged chromosomes than for those in colinear chromosomes. We also report that non-coding regions within rearranged regions tend to have lower divergence than non-coding regions outside them.

These results suggest an association between chromosomal rearrangements and lower non-coding divergence that has not been reported before and suggest that chromosomal speciation has not been common along the human and chimpanzee lineage.

BACKGROUND

Genomic DNA sequences of humans and chimpanzees differ by only 1.23% if considering only point mutations (Chen and Li 2001; Mikkelsen et al. 2005), a figure that grows up to 5% if small insertions and deletions are taken into account (Britten 2002) and up to a yet unknown percentage when segmental duplications are added to the picture (Cheng et al. 2005; Mikkelsen et al. 2005; She et al. 2006) Besides such relatively small-scale changes in their DNA sequences, the two species differ by large-scale rearrangements in their karyotypes. Human chromosome 2 results from the fusion of two acrocentric chromosomes that are independent in the great apes (Yunis and Prakash 1982). In addition, there are at least 7 major (larger than 10 Mb) pericentric inversions (in human chromosomes 4, 5, 9, 12, 15, 17 and 18) that range in size between 16 and 77 Mb and many smaller ones. Breakpoint regions of most of these rearrangements have been well defined both *in silico* (Feuk et al. 2005; Mikkelsen et al. 2005) and experimentaly (Dennehey et al. 2004; Goidts et al. 2005; Goidts et al. 2004; Kehrer-Sawatzki et al. 2005a; Kehrer-Sawatzki et al. 2005b; Kehrer-Sawatzki et al. 2002;

Kehrer-Sawatzki et al. 2005c; Locke et al. 2003; Szamalek et al. 2005; Yunis and Prakash 1982) although the exact location of some of them is still unclear.

Over the last three years, the role that these chromosomal rearrangements might have played in the speciation processes that have separated the lineages of humans and chimpanzees has come into the spotlight. According to models of chromosomal speciation based on the recombination-reducing effects of rearrangements, rearranged genomic regions involved in speciation processes would become isolated earlier compared to the rest of the genome (Navarro and Barton 2003a; Noor et al. 2001b; Rieseberg 2001). Thus, these models predict an association between speciation-related rearrangements and higher rates of divergence of genes and non-coding sequences linked to them. Current evidence for or against such models is contradictory. The first studies, including our own, making use of human and chimpanzee DNA sequence data seemed to support the existence of an association of chromosomal rearrangements with higher rates of protein and DNA sequence evolution (Lu et al. 2003; Navarro and Barton 2003b; Navarro et al. 2003). However, these studies were seriously affected by problems such as small sample size and biases in the data that were available in the Gene Bank at the time (Vallender and Lahn 2004). More recent studies, using larger datasets, have detected opposite trends (Zhang et al. 2004) or no association at all (Navarro et al. 2003; Vallender and Lahn 2004; Zhang et al. 2004). Also, a study based on human-chimpanzee gene expression divergence suggested that some inversions (in particular those in chromosomes 4, 5, 9, 15 and/or 16) could have been involved in the original speciation event separating the human and chimpanzee lineage (Marques-Bonet et al. 2004). Finally, an increasing amount of data coming from other species seem to fit the chromosomal speciation model. This is the case, at the moment, of studies involving such different lineages as drosophila, anopheles, murids, shrew or sunflowers (Armengol et al. 2005; Armengol et al. 2003; Ayala and Coluzzi 2005; Basset et al. 2006; Marques-Bonet and Navarro 2005; Noor et al. 2001b; Rieseberg et al. 1995; Rieseberg et al. 1999). So far, thus, the question remains unsolved: have the human and chimpanzee lineages been separated by processes of chromosomal speciation?

Here we revisit this issue by making use of the recently available chimpanzee genome sequence (Mikkelsen et al. 2005). Our aims are, first, to exhaustively compare rates of pairwise human-chimpanzee sequence divergence in rearranged and in colinear genomic regions and, second, to study lineage-

specific divergence rates in these same regions. To do so, we made use of the sets of measures of divergence between orthologous genes in humans, chimpanzees, rats and mice (including information for coding and non-coding sequences) gathered by the Chimpanzee Genome Consortium (Mikkelsen et al. 2005).

RESULTS

Filtering of factors affecting divergence

Before examining our main hypothesis we endeavoured to sequentially remove the effect of any factors that are known to affect rates of DNA sequence evolution in different genomic regions. First, we considered sex chromosomes. It has long been known that, due to the particular evolutionary dynamics of sex chromosomes (Crow 2000; Hurst and Ellegren 1998; Li et al. 2002; Makova and Li 2002), sequences linked to the X chromosome have lower divergence rates than those linked to autosomes (Li et al. 2002; Marques-Bonet and Navarro 2005; Wolfe and Sharp 1993). These results are confirmed by our analysis of human-chimpanzee parwise divergence. Genes located in the X chromosomes presented lower synonymous substitution rates (K_S) and lower non-coding divergence (K_I) than those in autosomes , whereas non.synonimous divergence rates (K_A) did not differ (Table 1). Lineage-specific substitution rates (obtained from the second dataset, see Matherials and Methods), showed the same trends significance was lost is some comparisons (in Additional data file 1, Table A1). As in previous studies (Marques-Bonet et al. 2004; Marques-Bonet and Navarro 2005), we removed genes linked to sex chromosomes from further analysis.

Next we dealt with segmental duplications (SDs), since they are known to be associated with higher rates of molecular evolution (Lynch and Conery 2000b; Marques-Bonet and Navarro 2005; Zhang et al. 2003). In the pariwise dataset, divergence rates in the non-coding regions of genes involved in SDs (either in the chimpanzee or in the human lineage) are not different from divergence rates of single-copy genes. This is also the case for K_A and the K_A/K_I ratio (Table 1). Surprisingly, however, K_S is significantly lower in genes within SDs. To explore this discrepancy with the previous literature referenced above, we split genes overlapping SDs in three main categories: those genes that overlap SDs shared by the human and the chimpanzee lineages; genes that overlap human SDs but not chimpanzee SDs; and genes that overlap chimpanzee SDs but not human SDs (Table 2). As expected, genes overlapping human SDs showed higher divergence than genes that do not overlap with SDs. On the other hand, genes overlapping chimpanzee SDs show present the opposite pattern, that is, evolutionary

rates are significantly lower for coding evolutionary rates. Finally, for those genes that overlap SDs that are shared by the human and chimpanzee lineages, only synonymous divergence is lower within shared SDs. This suggests that the lower rates of divergence for genes overlapping SDs detected in the overall analysis is an artifact of the preliminary state of the annotation of chimpanzee SDs. At any rate, we excluded from further analysis any gene overlapping SDs.

The chimpanzee genome project unveiled higher human-chimpanzee divergence within 10Mb from the telomeres (Mikkelsen et al. 2005). This effect can be detected in both the pairwise and the lineage-specific datasets (Table 1) and for both exonic and non-coding divergence. This is a particularly important factor, since nine out of the ten major rearrangements separating the two species are pericentric inversions, *i.e.*, they exclude telomeres. Thus, considering genes in telomeres might lead to under-estimation of divergence within rearrangements. To avoid such bias, genes within 10Mb of the telomeres were removed from further analysis.

Recent evidence suggests that, just as telomeres do, centromeric and centromeric transition regions exhibit unique organizational and evolutionary characteristics (Rudd and Willard 2004; She et al. 2004a; She et al. 2004b). In our pairwise dataset, genes located within 5 Mb of pericentromeric regions on each side of centromere showed significantly lower divergence rates than genes elsewhere in the genome (Table 1). In contrast, there are no significant lineage-specific differences in substitution rates between genes located in centromeric regions and genes in other parts of the genome (in Additional data file 1, Table A1). Given these interesting but potentially confusing patterns, genes in centromeric regions were removed from our dataset.

Finally, human chromosome 19 (HSA19) has been reported to present peculiar divergence and nucleotide composition patterns (Castresana 2002). Our results also pinpoint this chromosome as an outlier. All neutral divergence measures in the pairwise are markedly higher in HSA19 (Table 1). Differences in lineage-specific substitution rates are not as striking. Still, significant differences for K_S in the human and chimpanzee lineages and for K_A in the hominid lineage can be found (in Additional data file 1, Table A1). Thus, genes located in this chromosome were also removed from our dataset.

The successive removal of all the genes whose divergence values could be affected by any of the aforementioned confounding factors left 5804 genes for

pairwise analysis (Dataset 1) and 2742 in the lineage-specific analysis (Dataset 2). Such filtered datasets, even if dramatically reducing our sample size, allow for a detailed testing of the hypothesis of an association between chromosomal rearrangements and genic divergence rates. A graphic overview of the regions that were included in the following analysis or excluded from it is presented in Figure 1.

Major rearrangements.

As a rough preliminary test we compared rearranged and colinear chromosomes. Human-chimpanzee pairwise divergence rates are not different neither for synonymous sites (K_S) nor for the K_A/K_I ratio (Table 3). In contrast to these results and to previous literature, average rates of non-coding, K_I, and non-synonymous divergence, K_A, are significantly lower in rearranged chromosomes (Table 3). None of the comparisons performed upon lineage-specific rates are strikingly different. Only non-synonymous divergence for humans and neutral divergence in the hominid branches present marginal differences, being lower in rearranged chromosomes.

We then focused on rearranged chromosomes themselves and compared genes within inversions against genes outside them. In the pairwise dataset, non-coding sequences showed significantly lower divergence within rearrangements than outside them (0.0120 vs 0.0117, P-value < 0.001) whereas no significant divergence differences were detected for KA, KS and the KA/KI ratio (Table 3). No general pattern was detected in the lineage-specific analysis, even if genes within rearrangements show marginally lower rates in some cases (KA in human branch, KS in the chimpanzee branch and both KA and KS in the hominid lineage, in Additional data file 1 Table 2). This suggests that the association between rearranged chromosomes and lower divergence rates reported above is mainly due to genes within the rearrangements themselves. However, when the analysis is repeated removing genes within rearrangements, divergence is still lower in genes located in rearranged chromosomes (but outside rearrangements, Table 3). It is important to stress that these results cannot be biased by the strict filtering applied before our main analysis. Equivalent, only that stronger, trends were obtained before filtering when all genes were included in the analysis (data not shown).

If rearrangements did affect divergence rates due to their recombination-reducing effect (including effects due to speciation-related processes) their effect should be maximum around the rearrangement breakpoints, where recombination between different chromosomal arrangements is most strongly reduced (Andolfatto et al. 2001). To test for this possibility we defined windows of 2 Mb

around each rearrangement breakpoint (1 Mb at each side). Then, we compared genes within these windows against all genes in rearranged chromosomes (Table 4). In the pairwise analysis, we detected lower divergence in non-coding regions surrounding the evolutionary breakpoints. Exons, also show lower K_S and K_A values near breakpoints when compared to the rest of chromosomes, although neither of these results are statistically significant (Table 4). None of these differences can be detected in lineage-specific substitution rates (in Additional data file 1, Table 3).

It would thus seem that evolutionary rates of genes close to breakpoints follow the same trend as genes within rearrangements. To check whether these two trends are independent we removed genes surrounding breakpoints and repeated the main analysis comparing divergence within and outside rearrangements. Results did not change: in the pairwise analysis, genes within rearrangements displayed lower non-coding divergence than the rest of rearranged chromosomes (Table 5), even if reduced sample size limits our power and some results are not significant anymore. (in Additional data file 1, Table A4).

Finally, the accumulation of genes with K_A/K_S >1 in colinear chromosomes reported by Zhang *et al.* (Zhang et al. 2004) can also be detected in our pairwise dataset, although K_A/K_I is used instead of the "standard" K_A/K_S ratio. When focusing on rearranged chromosomes alone, no significant accumulation of genes with $K_A/K_I > 1$ was found either within our outside rearrangements (in Additional data file 1, Table A9)

Simulated rearrangements.

As explained above, genes located near the centromere had lower divergence than genes elsewhere in the genome, (see Table 1). This suggests that a possible explanation for our observation of lower divergence within rearrangements could be related to the fact that all the rearrangements analyzed are pericentric inversions. It is thus possible that removing genes in the centromeres and within a 5 Mb pericentromeric region on each side, as we did, is not enough to control for any potential centromere-related effects.

To test this hypothesis, we defined virtual pericentric inversions in colinear chromosomes, spanning the same average proportion of each chromosome as the real nine major inversions do in rearranged chromosomes. We compared genes within these virtual regions with genes outside them but in the same chromosomes. Table 6 shows that divergence patterns in these virtual rerrangements are similar to those in real

rearranged chromosomes. In the pairwise comparison, non-coding divergence is also lower within virtual inversions (Table 6) and, again, no pattern can be detected in the lineage-specific analysis (in Additional data file 1, Table A5). This suggests that centromere-related effects extending beyond the 5Mb windows we considered may be responsible of some, even if not all, of our observations.

Smaller rearrangements

All the above results referred to the 10 major rearrangements separating humans and chimanzees. More detailed information on the structural changes between the two species has recently become available by means of mapping chimpanzee fosmid paired-end sequences against the human genome (Newman et al. 2005). This analysis unveiled 37 smaller rearrangements (usually <1 Mb) which, in contrast to the major ones, do not include centromeric regions and, thus, allow to exclude any potential bias caused by centromeres. We compared substitution rates of genes overlapping these rearrangements with genes in colinar regions. Pairwise non-coding substitution rates are found marginally higher within these rearrangements (K_I, 0.0121 vs 0.0128, P-value = 0.023.Table 7) whereas other divergence measures do not present significant differences. This observation can not be retrieved in the lineage-specific analysis but in any case, sample size for this kind of approach is really small and should be treated with caution (in Additional Data File 1, Table A6).

Chromosome by chromosome analysis.

So far, all the tests presented here were performed pooling all rearranged chromosomes together. It is clear, however, that no chromosomal speciation model proposes that every single rearrangement ought to have played a relevant role in the speciation processes that separated humans and chimpanzees. In fact, it is reasonable to assume that most rearrangements would have appeared and become fixed along the evolutionary history of lineages (anagenesis) and not during the relatively shorter cladogenic periods (Lu et al. 2003; Navarro et al. 2003). It is thus possible that a majority of speciation-unrelated rearrangements could be masking the molecular signature of chromosomal speciation in the few rearrangements involved in such processes. Provided, of course, that there are any speciation-related rearrangements at all. In fact, a recent comparative gene-expression study hints at some chromosomes (such as HSA4, HSA5, HSA9, HSA15 and HSA16) as the most different in terms of differences in expression pattern (Marques-Bonet et al. 2004).

Thus, we repeated all previous analyses in a chromosome-per-chromosome basis (Table 8 and, in Additional data file 1,

Table A7). In most cases, the small sample size caused by our extremely conservative filtering process precludes the detection of any trend or even the performance of tests (for example, no genes from chromosomes HSA15, HSA16 or HSA 18 are included in our dataset after filtering).

HSA 4 clearly stands out in the pairwise comparison. It presents statistically higher K_A, K_I and K_A/K_I within the inversion (having removed the breakpoints). The centromeric region of HSA4 presents the usual lower divergence thus confirming that the effect of HSA4 was not due to a some special behaviour of its centromere extending beyond 5 Mb. In contrast to other chromosomes, genes outside the inversion in HSA4 also present higher divergence than genes in colinear chromosomes (Data not shown).

The other chromosome that stands out in the analysis is HSA12, that presents lower divergence, both for genes within its inversion relative to those outside it and for genes outside the inversion relative to genes in colinear chromosomes (Data not shown). HSA15 present the same trend, even if with less statistical strength. Together, these two chromosomes are the major contributors to the observation of lower divergence for genes outside rearrangements that for genes in colinear chromosomes.

Recombination rates.

Recombination rates have shown to be correlate positively with divergence (Hellmann et al. 2003a). We first examined the relationship between recombination and the factors we have excluded from our analysis. All figures are given in $cM \cdot Mb^{-1}$. In our data set, recombination rates are higher for genes located in X chromosome than for genes elsewhere in the geome (1.43 *vs.* 1.21, P-value 0.027). This is also the case for genes in telomeric regions (1.09 *vs.* 1.97, P-value < 0.001) and in HSA19 (1.08 *vs.* 1.57, P-value < 0.001). All these results are congruent with previous observations (Kong et al. 2002b). Recombination rates are also lower for genes located in Segmental Duplications (1.28 *vs.* 1.04, P-value < 0.001) and centromeric regions (1.10 *vs*. 0.82, P-value = 0.002).

We then focused on chromosomal rearrangements. Recombination rates on both classes of chromosomes (colinear and rearranged) are very similar (1.06 *vs.* 1.09, P-value n.s.). Within rearranged chromosomes, recombination rates are significantly higher within inversions than in regions outside the inversion, but marginally so (1.07 *vs*. 1.24, P-value = 0.07). Also, regions surrounding breakpoints show higher levels of recombination than the rest of their chromosome (1.91 *vs.* 1.08, P-value = 0.002).

DISCUSSION

In the present whole-genome analysis, several puzzling patterns have been detected that were not reported by previous publications. In particular, Mikkelsen *et al*. (Mikkelsen et al. 2005) performed a full-fledged descriptive analysis of the new sequence of the chimpanzee genome and, among other analysis, they tested for an increase of the rates of protein evolution of genes in rearranged chromosomes relative to genes on colinear chromosomes and of genes within the rearrangements themselves relative to genes outside them. We extended our analysis not only to the ratio of evolutionary rates, but also to individual synonymous and non-synonymous evolutionary rates. Moreover, we carefully screened rearranged and colinear regions together with their breakpoints.

A first conclusion of our analysis is that divergence is lower for genes located in rearranged chromosomes than for those in colinear chromosomes (Tables 2a, 2c). These results contradict all previous observations. They contradict the first analysis, which, based on small datasets, reported a trend to increased divergence in rearranged chromosomes (Lu et al. 2003; Navarro and Barton 2003b; Navarro et al. 2003). They are also contrary to the results by Zhang *et al*. (Zhang et al. 2004) and Vallender *et al*. (Vallender and Lahn 2004), who found no significant association between rearrangements and average genic evolutionary rates using large datasets. Another pattern emerging from our results is that, when focusing on rearranged chromosomes, non-coding regions within rearranged regions tend to have lower divergence than non-coding regions outside them. Again, this result suggests a relationship between chromosomal rearrangements and lower non-coding divergence that has not been reported before.

Why should non-synonymous and non-coding divergence be lower in rearranged chromosomes, particularly within rearrangements? It is tempting to speculate that rearrangements tend to occur in regions with particular sequence features, such as lower recombination and, thus, lower ancestral polymorphism that would translate in lower divergence. Also, it is possible that changes in recombination rates induced by rearrangements could be affecting mutation rates. However, we lack the ancestral recombination data that would be needed to properly test these hypotheses. Extant evidence is not only scarce, but contradictory. For example, in humans there are no differences in rates of recombination between rearranged and colinear chromosomes (see Additional Data File 1,Table A2), but, of course, one would not expect fixed inversions to affect current recombination

rates. Evidence weakly hinting at lower ancestral polymorphism comes from current polymorphism levels in humans. Using intraspecific population data from the 256 genes in SeattleSNP , we found that nucleotide divergence is lower in rearranged chromosomes than in colinear chromosomes ($8.13x10^{-4}$ vs $9.34 x10^{-4}$, P-value = 0.021), but there were no differences between genes outside the rearrangements *versus* genes inside them ($7.45x10^{-4}$ vs $8.26 x10^{-4}$, P-value = 0.42). Still, the last analysis must be taken with care, since the number of genes within inversions was as low as 20.

Another potential explanation comes from the effect of centromeres. The major rearrangements analyzed in this paper are all pericentromeric. Even when removing genes in centromeres and within 5Mb of pericentromeric regions, we can still see lower divergence within rearrangements. This is not the case for small inversions which do present slightly higher non-coding divergence. Taken together, these data suggest that centromeres have a divergence-reducing effect that extends beyond 5 Mb and helps explaining our global observation. However, divergence rates are still lower for genes rearranged chromosomes after removing genes within rearrangements, a result for which, at the moment, we lack an explanation. At any rate this observations should be interpreted carefully, as they are based based on the comparison of only two genomes. As noted by Navarro and Barton (Navarro and Barton 2003b) and Vallender *et al*. (Vallender and Lahn 2004) the genomewide non-uniform distribution of genes and rates of divergence could be at the origin of our observation. Additional analyses involving more species and making use of outgroup sequences are needed to clarify this point .

As to the evolutionary rates of specific lineages, it is not surprising to find almost no significant differences. The murid lineage can not be defined as a “close” brother lineage to the human-chimpanzee speciation, and thus, is giving us an unbalanced tree with long inner and short terminal branches. As a consequence we lack power in the interesting terminal branches (that is, the chimpanzee and human branches). More appropriate species for this sort of comparisons will be shortly available, making it possible to increase the power of this analysis by adding density to the primate tree**.**

A fourth interesting observation is related to the relationship between recombination rates and rearrangements. We report higher recombination rates in regions surrounding evolutionary breakpoints. It is widely admitted that recombination is greatly reduced around

rearrangement breakpoints of heterokariotypic individuals (Andolfatto et al. 2001) and this may seem to contradict our results. However, it is quite clear that measures of recombination reported here correspond to present, and not to ancestral, recombination rates. Because recombination rates change dramatically over time (Ptak et al. 2005) we can not infer any relevant conclusion of this relationship. It is, however, tempting to speculate that rearrangements may tend to take place in regions of high recombination. New primate recombination data from chimpanzees and other primate species (such as Bornean and Sumatran orang-utans, specially since a chromosomal inversion differentiates these two subspecies (Seuanez et al. 1979)) will help to shed some light on this issue.

Our fifth and final observation is that certain chromosomes seem to present some strong individual trends. Blurry results are to be expected in this analysis, since our statistical power was greatly reduced by the conservative approach we choose (outright removal of certain factors instead of trying to control them by, for example, normalisation and correction by multiple regression) and, thus, any putative chromosome-per-chromosome patterns are likely to be overshaded by the great variation of rates of divergence across the genome. Still, chromosome 4 presents significantly higher divergence rates inside its rearrangement. This result is consistent with previous analysis of gene expression and sequence data (Marques-Bonet et al. 2004; Mikkelsen et al. 2005).

A final important issue is the relevance of our observations to the problem of the mode of speciation between humans and chimpanzees and along their respective lineages. Our results clearly show that there is no evidence for recurrent chromosomal speciation along the human or chimpanzee lineages. DNA. The prediction of higher DNA sequence divergence that suppressed-recombination models of chromosomal speciation make is not fulfilled by most rearrangements. Some isolated event of chromosomal speciation can not be ruled out, since there are rearrangements that harbor highly divergent genes, still we lack proper outgroups to test whether this high divergence is due to the presence of the inversion or to, for example, a cluster of weakly constrained genes. The issue of the mode of speciation between humans and chimpanzees remains elusive, as shown by recent works trying to look for signals of parapatric or allopatric speciation in the DNA sequences of both species (Barton 2006; Innan and Watanabe 2006; Osada and Wu 2005; Patterson et al. 2006). More experimental and theoretical knowledge needs to be gathered before the debate can be satisfactorily settled

CONCLUSION.

Based on the observations we report here, chromosomal speciation does not appear to have been common along the human and chimpanzee lineages, although chromosome 4 clearly stands out as the best candidate to have played a role in some particular speciation process. In the future, the detailed study of the interaction of chromosomal rearrangements with some of the factors we removed in the present study, particularly with segmental duplications, will certainly shed light on the issue of the genomic distribution of rates of genic evolution.

MATERIALS AND METHODS

Sequence gathering.and evolutionary rates.

All data analyzed was retrieved from the initial chimpanzee genome sequence (Mikkelsen et al. 2005) and methods there should be consulted. In summary, two databases were used. First, a set of more than 13000 unambiguous human-chimpanzee orthologous genes filtered to avoid overrepresentation of gene families. From that initial dataset, only those genes with unequivocal coordinates in both species were kept. The chromosomal position of the sequences is a key parameter of our analysis, and, thus, genes in random chromosomes were also removed from our analysis leaving a total of 12135 genes.

For every coding sequence (CDS), several conventional indexes of molecular evolution, such as the number of non-synonymous substitutions per non-synonymous site (K_A), the number synonymous substitutions per silent site (K_S), and their ratio (K_A/K_S) were estimated using the maximum likelihood method implemented in the package PAML (Yang 1997). Substitution rates for non-coding sequence were calculated as K_I, the number of substitutions per non-coding nucleotide. A K_I value was obtained for a window of 250kb, centered on each gene. We used K_A/K_I instead of K_A/K_S as the measure of rates of protein evolution, because of the close proximity between human and chimpanzees which results quite often in a K_S equal to 0. The averages for K_A, K_S, K_I, and the ratio K_A/K_I are 0.00317, 0.0142, 0.0126 and 0.2483 respectively. Because of the strict criteria defined to retrieve the set of orthologous genes, the maximum values of each index are not high enough to be suspicious of false orthology or misalignment ($K_S < 0.32$, $K_A < 0.055$ and $K_I < 0.0259$)

A second dataset was used to calculate lineage specific evolutionary rates. More than 7000 unambiguously orthologous genes were recovered for 4 species (human, chimpanzee, rat and mouse).

We applied the same filtering criteria as in the previous dataset and were left with a set of 4905 orthologous genes with coordinates in both species and evolutionary rates for every branch in the non-rooted tree. Finally, the lineage specific evolutionary rates were estimated using a nonrooted tree in PAML

Polymorphism data

Polymorphism data was gathered from from SeattleSNP webpage . Briefly, we downloaded nucleotide diversity measures for 256 genes. These measures have been obtained from full resequenceing of 24 African-American and 23 European (CEPH) subjects.

Recombination.

Human recombination rates, measured in $cM \cdot Mb^{-1}$, were obtained from the fine-resolution recombination map in USCS genome browser by selecting the track SNP Recombination Rates. Estimates are based on the HapMap Phase I data, release 16a, and Perlegen data (Hinds et al. 2005). Fine scale recombination maps are not yet available for chimpanzees. All genes were assigned a recombination rate computed as the average of all SNPs included within them. Any genes for which recombination rates could not be determined were removed from any recombination-based analysis.

Structural information.

Coordinates of telomeres and centromeres of all chromosomes were obtained from Build 34 of the human genome and NCBI Build 1 of the chimpanzee genome . We considered as rearranged chromosomes all those for which major chromosomal rearrangements in either the human or the chimpanzee lineages have been evidenced by recent *in silico* (Feuk et al. 2005; Mikkelsen et al. 2005) or cytological data (Goidts et al. 2004; Kehrer-Sawatzki et al. 2005a; Kehrer-Sawatzki et al. 2005b; Kehrer-Sawatzki et al. 2002; Kehrer-Sawatzki et al. 2005c; Szamalek et al. 2005). This comprised human chromosomes 1, 4, 5, 9,12,15, 16, 17 and 18, which differ by a pericentric inversion, and human chromosome 2, which has been generated by an ancestral telomere-telomere fusion (Yunis and Prakash 1982). For all chromosomes, all *in silico*-estimated coordinates were compared with newly available cytological data in order to confirm inversion coordinates. The most remarkable difference from both methodologies comes from chromosome 1, in which an inversion of about 30 Mb was detected *in silico* that has not been detected by cytological approaches. (in Additional data file 1, Table 8).

Segmental duplications.

Human Segmental Duplication (SD) and Chimpanzee SD coordinates were

downloaded from the Segmental Duplications Database and . As a conservative measure against false orthology, genes in our dataset overlapping the positions of SDs were removed from the analysis related to rearrangements.

Genomic position of genes

Location information was derived from both humans and chimpanzees. When genes located in different genomic regions of interest (such as sex chromosomes, SDs or telomeres) were studied, being in one of such regions in either humans or chimpanzees was enough to classify a gene as located in such regions. Location was established sequentially as shown in the Results section.

Permutation tests

Genes in different categories were compared by means of pairwise permutation tests (based on 1000 permutations). P-values in such tests, are calculated as the proportion of times that the difference of averages between two categories in a permuted dataset is equal or larger than the observed difference.

ADDITIONAL DATA FILE

File name :

"Marques_Bonet_etal_Additional_Data_File_1.doc"

File format: Microsoft Word document.

Description of this dataset: Analysis of lineage-specific evolutionary rates and recombination rates for factors known to affect evolutionary rates and according to their position in relation to rearrangements. Comparison of evolutionary breakpoints between human and chimpanzee**.**

ACKNOWLEDGEMENTS

We thank J. Bertranpetit, O. Lao, O. Fernando, E. Eichler, M. Przeworski and the members of the Evolutionary Biology Unit in UPF for enriching discussions during the preparation of this work. A. N. is a member of the Ramón y Cajal Program (Spain). This research was supported by grants to A.N. from the Ministerio de Ciencia y Tecnologia (Spain, BOS 2003-0870) and the Genome Canada-Genoma España Joint R+D+I Projects in Human Health (JLI/038).

REFERENCES.

1. 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* 2001, **68**(2):444-456.
2. Mikkelsen TS, Hillier LW, Eichler EE, Zody MC, Jaffe DB, Yang SP, Enard W, Hellmann I, Lindblad-Toh K, Altheide TK *et al*: **Initial sequence of the chimpanzee genome and comparison with**

the human genome. *Nature* 2005, **437**(7055):69-87.
3. Britten RJ: **Divergence between samples of chimpanzee and human DNA sequences is 5%, counting indels**. *P Natl Acad Sci USA* 2002, **99**(21):13633-13635.
4. Cheng Z, Ventura M, She XW, Khaitovich P, Graves T, Osoegawa K, Church D, DeJong P, Wilson RK, Paabo S *et al*: **A genome-wide comparison of recent chimpanzee and human segmental duplications**. *Nature* 2005, **437**(7055):88-93.
5. She XW, Liu G, Ventura M, Zhao S, Misceo D, Roberto R, Cardone MF, Rocchi M, Green ED, Archidiacano N *et al*: **A preliminary comparative analysis of primate segmental duplications shows elevated substitution rates and a great-ape expansion of intrachromosomal duplications**. *Genome Res* 2006, **16**(5):576-583.
6. 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* 2005, **1**(4):489-498.
7. 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**. *Cytogenetic and Genome Research* 2005, **108**(1-3):91-97.
8. 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)**. *Human Mutation* 2005, **25**(1):45-55.
9. 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**. *Human Genetics* 2005, **117**(2-3):168-176.
10. 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* 2005, **85**(5):542-550.
11. 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* 2002, **71**(2):375-388.
12. 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**. *Human Genetics* 2004, **115**(2):116-122.
13. Yunis JJ, Prakash O: **The Origin of Man - a Chromosomal Pictorial Legacy**. *Science* 1982, **215**(4539):1525-1530.
14. 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 Biology* 2003, **4**(8):R50.
15. 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* 2005, **15**(9):1232-1242.
16. Dennehey BK, Gutches DG, McConkey EH, Krauter KS: **Inversion, duplication, and changes in gene context are associated with human chromosome 18 evolution**. *Genomics* 2004, **83**(3):493-501.
17. Navarro A, Barton NH: **Accumulating postzygotic isolation genes in parapatry: A new twist on chromosomal speciation**. *Evolution* 2003, **57**(3):447-459.
18. Rieseberg LH: **Chromosomal rearrangements and speciation**. *Trends in Ecology & Evolution* 2001, **16**(7):351-358.
19. Noor MAF, Grams KL, Bertucci LA, Reiland J: **Chromosomal inversions and the reproductive isolation of species**. *P Natl Acad Sci USA* 2001, **98**(21):12084-12088.
20. Navarro A, Barton NH: **Chromosomal speciation and molecular divergence - Accelerated evolution in rearranged chromosomes**. *Science* 2003, **300**(5617):321-324.
21. Lu J, Li WH, Wu CI: **Comment on " Chromosomal speciation and molecular divergence - Accelerated evolution in rearranged chromosomes"**. *Science* 2003, **302**(5647):988.
22. Navarro A, Marques-Bonet T, Barton NH: **Response to comment on " chromosomal speciation and molecular divergence - Accelerated evolution in rearranged chromosomes"**. *Science* 2003, **302**(5647):988.
23. Vallender EJ, Lahn BT: **Effects of chromosomal rearrangements on human-chimpanzee molecular evolution**. *Genomics* 2004, **84**(4):757-761.
24. Zhang JZ, Wang XX, Podlaha O: **Testing the chromosomal speciation hypothesis for humans and chimpanzees**. *Genome Res* 2004, **14**(5):845-851.
25. 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 in Genetics* 2004, **20**(11):524-529.
26. 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**. *Human Molecular Genetics* 2003, **12**(17):2201-2208.
27. Marques-Bonet T, Navarro A: **Chromosomal rearrangements are associated with higher rates of molecular evolution in mammals**. *Gene* 2005, **353**(2):147-154.
28. Rieseberg LH, Whitton J, Gardner K: **Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species**. *Genetics* 1999, **152**(2):713-727.
29. Rieseberg LH, Vanfossen C, Desrochers AM: **Hybrid Speciation Accompanied by Genomic Reorganization in Wild Sunflowers**. *Nature* 1995, **375**(6529):313-316.
30. Armengol L, Marques-Bonet T, Cheung J, Khaja R, Gonzalez JR, Scherer SW, Navarro A, Estivill X: **Murine segmental duplications are hot spots for chromosome and gene evolution**. *Genomics* 2005, **86**(6):692-700.
31. Ayala FJ, Coluzzi M: **Chromosome speciation: Humans, Drosophila, and mosquitoes**. *P Natl Acad Sci USA* 2005, **102**:6535-6542.

32. Basset P, G. Y, H. B, J. H: **Restricted Gene Flow at specific parts of the shrew genome in chromosomal hybrids zones**. *Evolution (In press)* 2006.
33. Crow JF: **A new study challenges the current belief of a high human male : female mutation ratio**. *Trends in Genetics* 2000, **16**(12):525-526.
34. Hurst LD, Ellegren H: **Sex biases in the mutation rate**. *Trends in Genetics* 1998, **14**(11):446-452.
35. Li WH, Yi SJ, Makova K: **Male-driven evolution**. *Current Opinion in Genetics & Development* 2002, **12**(6):650-656.
36. Makova KD, Li WH: **Strong male-driven evolution of DNA sequences in humans and apes**. *Nature* 2002, **416**(6881):624-626.
37. Wolfe KH, Sharp PM: **Mammalian Gene Evolution - Nucleotide-Sequence Divergence between Mouse and Rat**. *Journal of Molecular Evolution* 1993, **37**(4):441-456.
38. Lynch M, Conery JS: **The evolutionary fate and consequences of duplicate genes**. *Science* 2000, **290**(5494):1151-1155.
39. Zhang P, Gu ZL, Li WH: **Different evolutionary patterns between young duplicate genes in the human genome**. *Genome Biology* 2003, **4**(9):R56.
40. Rudd MK, Willard HF: **Analysis of the centromeric regions of the human genome assembly**. *Trends in Genetics* 2004, **20**(11):529-533.
41. She XW, Horvath JE, Jiang ZS, Liu G, Furey TS, Christ L, Clark R, Graves T, Gulden CL, Alkan C *et al*: **The structure and evolution of centromeric transition regions within the human genome**. *Nature* 2004, **430**(7002):857-864.
42. She XW, Jiang ZX, Clark RL, 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* 2004, **431**(7011):927-930.
43. Castresana J: **Genes on human chromosome 19 show extreme divergence from the mouse orthologs and a high GC content**. *Nucleic Acids Research* 2002, **30**(8):1751-1756.
44. Andolfatto P, Depaulis F, Navarro A: **Inversion polymorphisms and nucleotide variability in Drosophila**. *Genetical Research* 2001, **77**(1):1-8.
45. 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* 2005, **15**(10):1344-1356.
46. 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* 2003, **72**(6):1527-1535.
47. Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G *et al*: **A high-resolution recombination map of the human genome**. *Nat Genet* 2002, **31**(3):241-247.
48. **SeattleSNPs (http://pga.mbt.washington.edu/)**.
49. Ptak SE, Hinds DA, Koehler K, Nickel B, Patil N, Ballinger DG, Przeworski M, Frazer KA, Paabo S: **Fine-scale recombination patterns differ between chimpanzees and humans (vol 37, pg 429, 2005)**. *Nat Genet* 2005, **37**(4):445-445.
50. Seuanez HN, Evans HJ, Martin DE, Fletcher J: **Inversion of Chromosome-2 That Distinguishes between Bornean and Sumatran Orangutans**. *Cytogenet Cell Genet* 1979, **23**(1-2):137-140.
51. Innan H, Watanabe H: **The effect of gene flow on the coalescent**

time in the human-chimpanzee ancestral population. *Mol Biol Evol* 2006, **23**(5):1040-1047.
52. Osada N, Wu CI: **Inferring the mode of speciation from genomic data: A study of the great apes**. *Genetics* 2005, **169**(1):259-264.
53. Patterson N, Richter DJ, Gnerre S, Lander ES, Reich D: **Genetic evidence for complex speciation of humans and chimpanzees**. *Nature* 2006, **441**(7097):1103-1108.
54. Barton N: **How did the Human Species Form?** *Current Biology* 2006, **16**(16):647-650.
55. Yang ZH: **PAML: a program package for phylogenetic analysis by maximum likelihood**. *Computer Applications in the Biosciences* 1997, **13**(5):555-556.
56. Hinds DA, Stuve LL, Nilsen GB, Halperin E, Eskin E, Ballinger DG, Frazer KA, Cox DR: **Whole-genome patterns of common DNA variation in three human populations**. *Science* 2005, **307**(5712):1072-1079.
57. **USCS Genome Browser (http://genome.ucsc.edu)**.
58. **Human Segmental Duplications Database (http://humanparalogy.gs.washington.edu/)**.
59. **Chimpanzee Segmental Duplications Database (http://chimpparalogy.gs.washington.edu/)**.

Table 1. Analysis of factors known to affect evolutionary rates. Average divergence measures are compared between genes within and outside genomic regions previously shown to be affected by processes influencing divergence rates. See text for details.

	HSA X *vs.* Autosomes		Segmental Duplications		Telomeres *vs.* rest of genome		Centromeres *vs.* rest of genome		HSA19	
	Genes in autosomes	Genes in HSA X.	Genes outside SDs	Genes within SDs	Genes outside Telomeres	Genes within Telomeres	Genes outside Centromeres	Genes within Centromeres	Genes outside HSA19	Genes within HSA19
N	11691	434	8431	3260	6627	1804	6165	462	5804	361
K_I	0.0127	0.0094	0.0127	0.0127	0.0121	0.0149	0.0121	0.0118	0.0121	0.0132
		< 0.001		0.982		< 0.001		< 0.001		< 0.001
K_A	0.0032	0.0029	0.0031	0.0033	0.0029	0.0040	0.0029	0.0030	0.0029	0.0032
		0.129		0.048		< 0.001		0.687		0.114
K_S	0.0145	0.0088	0.0147	0.0138	0.0129	0.0213	0.0130	0.0118	0.0127	0.0176
		< 0.001		0.002		< 0.001		0.039		< 0.001
K_A/K_I	0.2459	0.2987	0.2434	0.2525	0.2370	0.2669	0.2364	0.2453	0.2360	0.2422
		0.002		0.161		< 0.001		0.537		0.671

Table 2. Comparison of genes overlapping segmental duplications. Genes in Sex Chromosomes, in telomeres, centromeres and Chr 19 were removed before this analysis to avoid known confounding factors.

	Genes overlapping Shared SDs			Genes overlapping Human specific SDs			Genes overlapping Chimp specific SDs		
	Genes outside SDs	Genes within SDs	P-value	Genes outside SDs	Genes within SDs	P-value	Genes outside SDs	Genes within SDs	P-value
N	5804	330		5804	720		5804	1364	
K_I	0.0121	0.0121	0.574	0.0121	0.0122	0.032	0.0121	0.0121	0.127
K_A	0.0029	0.0030	0.502	0.0029	0.0040	< 0.001	0.0029	0.0025	0.001
K_S	0.0127	0.0110	0.009	0.0127	0.0138	0.016	0.0127	0.0118	0.005
K_A/K_I	0.2360	0.2425	0.713	0.2360	0.3126	< 0.001	0.2360	0.2068	0.002

Table 3. Analysis of genes according to their position in relation to rearrangements. Comparison of genes in regions involved in rearrangements vs. genes in colinear chromosomes or regions. Genes in breakpoints are included.

	Genes in Rearranged *vs.* Colinear chromosomes			**Genes within *vs.* outside inversions (excluding HSA2, PTR12, PTR13)**			**Genes Outside inversions *vs.* Genes in Colinear Chromosomes (excluding HSA2, PTR12, PTR13)**		
	Colinear	**Rearranged**	**P-value**	**Outside**	**Inside**	**P-value**	**Colinear**	**Outside**	**P-value**
N	2677	3127		2072	610		2677	2072	
K_I	0.0122	0.0120	0.001	0.0120	0.0117	< 0.001	0.0122	0.0120	0.027
K_A	0.0030	0.0028	0.036	0.0027	0.0028	0.648	0.0030	0.0027	0.014
K_S	0.0131	0.0125	0.122	0.0127	0.0119	0.119	0.0131	0.0127	0.518
K_A/K_I	0.2442	0.2290	0.080	0.2255	0.2346	0.504	0.2442	0.2255	0.038

Table 4. Comparison of genes in breakpoints versus genes in other rearranged chromosomes or regions.

	Genes in Breakpoints vs. inverted chromosomes (excluding HSA2, PTR12, PTR13)		
	Rearranged	**BKP**	**P-value**
N	2610	72	
K_I	0.0120	0.0113	0.001
K_A	0.0028	0.0023	0.260
K_S	0.0126	0.0117	0.427
K_A/K_I	0.2283	0.2001	0.406

Table 5. Comparison of genes in regions involved in rearrangements vs. genes outside inversions. Genes in breakpoints are excluded.

	Genes within vs. outside inversions (excluding breakpoints and HSA2, PTR12, PTR13)		
	Outside	**Inside**	**P-value**
N	2070	540	
K_I	0.0120	0.0118	0.001
K_A	0.0027	0.0029	0.301
K_S	0.0127	0.0119	0.144
K_A/K_I	0.2251	0.2406	0.316

Table 6. Comparison of genes with pericentric inversions simulated in colinear chromosomes vs. genes outside them.

	Genes in simulated pericentric inversions in colinear chromosomes (without HSA2 and without centromere)		
	Outside	**Inside**	**P-value**
N	2237	440	
K_I	0.0122	0.0119	0.009
K_A	0.0030	0.0029	0.562
K_S	0.0129	0.0133	0.551
K_A/K_I	0.2448	0.2410	0.810

Table 7. Comparison of genes overlapping those inversion located “in silico” in Newman TL et al. (Newman et al. 2005).

	Genes overlapping microinversions versus genes in rest of chromosomes		
	Outside	**Inside**	**P-value**
N	5778	26	
K_I	0.0121	0.0128	0.020
K_A	0.0029	0.0026	0.744
K_S	0.0127	0.0090	0.079
K_A/K_I	0.2362	0.2079	0.625

Table 8. Comparison of evolutionary rates of genes within inversions in individual chromosomes vs. genes outside inversions. Genes in breakpoints are excluded.

	HSA1				HSA9				HSA16		
	Genes within vs outside inversion (no BKP 1Mb)				Genes within vs outside inversion (no BKP 1Mb)				Genes within vs outside inversion (no BKP 1Mb)		
	outside	inside	P-value		outside	inside	P-value		outside	inside	P-value
N	774	6		N	197	17		N	219		
K_I	0.0117	0.0111	0.207	K_I	0.0123	0.0117	0.117	K_I	0.0120		
K_A	0.0029	0.0032	0.833	K_A	0.0027	0.0024	0.667	K_A	0.0029		
K_S	0.0134	0.0050	0.049	K_S	0.0127	0.0135	0.750	K_S	0.0158		
K_A/K_I	0.2387	0.2754	0.769	K_A/K_I	0.2221	0.2016	0.777	K_A/K_I	0.2395		

	HSA4				HSA12				HSA17		
	Genes within vs outside inversion (no BKP 1Mb)				Genes within vs outside inversion (no BKP 1Mb)				Genes within vs outside inversion (no BKP 1Mb)		
	outside	inside	P-value		outside	inside	P-value		outside	inside	P-value
N	183	66		N	161	170		N	40	174	
K_I	0.0125	0.0130	0.015	K_I	0.0118	0.0115	0.013	K_I	0.0126	0.0114	< 0.001
K_A	0.0030	0.0047	0.002	K_A	0.0023	0.0022	0.787	K_A	0.0023	0.0030	0.248
K_S	0.0122	0.0120	0.896	K_S	0.0119	0.0105	0.181	K_S	0.0130	0.0148	0.537
K_A/K_I	0.2353	0.3468	0.017	K_A/K_I	0.1946	0.1907	0.891	K_A/K_I	0.1899	0.2533	0.221

	HSA5				HSA15				HSA18		
	Genes within vs outside inversion (no BKP 1Mb)				Genes within vs outside inversion (no BKP 1Mb)				Genes within vs outside inversion (no BKP 1Mb)		
	outside	inside	P-value		outside	inside	P-value		outside	inside	P-value
N	217	105		N	195			N	72		
K_I	0.0120	0.0120	0.950	K_I	0.0122			K_I	0.0131		
K_A	0.0026	0.0029	0.503	K_A	0.0024			K_A	0.0033		
K_S	0.0113	0.0097	0.078	K_S	0.0109			K_S	0.0105		
K_A/K_I	0.2154	0.2420	0.477	K_A/K_I	0.1932			K_A/K_I	0.2476		

Additional Table A1.

Analysis of factors known to affect evolutionary rates. Divergence rates are compared between genes within or outside genomic regions previously shown to be affected by their own evolutionary dynamics. See text for details.

	HSA X . *vs*. Autosomes			Segmental Duplications			Telomeres *vs* rest of Chromosome		
	Genes in Autosomes	Genes in HSA X	P-value	Genes outside SDs	Genes related with SDs	P-value	Genes not in Telomeres	Genes in Telomeres	P-value
N	4768	137		3696	1072		2955	741	
K_A (Human)	0.0012	0.0010	0.091	0.0013	0.0011	0.038	0.0012	0.0016	< 0.001
K_S (Human)	0.0069	0.0063	0.388	0.0072	0.006	< 0.001	0.0065	0.0100	< 0.001
K_A (Chimpanzee)	0.0012	0.0007	0.009	0.0012	0.0010	< 0.001	0.0012	0.0014	0.038
K_S (Chimpanzee)	0.0071	0.004	0.002	0.0072	0.0068	0.271	0.0065	0.0102	< 0.001
K_A (Hominid)	0.0141	0.0098	0.001	0.0145	0.0127	< 0.001	0.0128	0.0215	< 0.001
K_S (Hominid)	0.0025	0.0017	0.004	0.0026	0.0021	< 0.001	0.0024	0.0030	< 0.001

	Centromeres *vs*. rest of genome			HSA19		
	Genes outside Centromeres	Genes within Centromeres	P-value	Genes outside HSA19	Genes within HSA19	P-value
N	2742	213		2620	122	
K_A (Human)	0.0012	0.0011	0.327	0.0012	0.0011	0.737
K_S (Human)	0.0065	0.0065	0.951	0.0064	0.0094	< 0.001
K_A (Chimpanzee)	0.0012	0.0012	0.860	0.0012	0.0013	0.476
K_S (Chimpanzee)	0.0064	0.0067	0.645	0.0063	0.0098	< 0.001
K_A (Hominid)	0.0128	0.0130	0.839	0.0125	0.0197	< 0.001
K_S (Hominid)	0.0024	0.0023	0.501	0.0024	0.0025	0.879

Average human recombination rates for genomic regions of interest. Divergence rates are compared between genes within or outside genomic regions previously shown to be affected by their own evolutionary dynamics. See text for details.

	HSA X *vs.* Autosomes		Segmental Duplications		Telomeres *vs.* rest of genome		Centromeres *vs.* rest of genome		HSA19	
	Genes in autosomes	**Genes in HSA X .**	**Genes outside SDs**	**Genes within SDs**	**Genes outside Telomeres**	**Genes within Telomeres**	**Genes outside Centromeres**	**Genes within Centromeres**	**Genes not in HSA19**	**Genes in HSA19**
N	11440	426	8244	3197	6474	1770	6024	450	5672	352
Recombination	1.2154	1.4353	1.28	1.0496	1.0887	1.9798	1.1081	0.8292	1.0792	1.5733
P-value		0.027		< 0.001		< 0.001		0.002		< 0.001

Additional Table A2.

Analysis of genes according to their position in relation to rearrangements. Comparison of genes in regions involved in rearrangements *vs*. genes in colinear chromosomes or regions. Genes in breakpoints are included.

	Genes in Rearranged *vs.* Colinear chromosomes			**Genes within *vs*. outside inversions (excluding HSA2, PTR12, PTR13)**		
	Colinear	**Rearranged**	**P-value**	**Outside**	**Inside**	**P-value**
N	1170	1450		965	265	
K_A (Human)	0.0013	0.0012	0.077	0.0012	0.0009	0.033
K_S (Human)	0.0063	0.0064	0.606	0.0066	0.0061	0.328
K_A (Chimpanzee)	0.0012	0.0011	0.212	0.0012	0.0009	0.083
K_S (Chimpanzee)	0.0064	0.0062	0.438	0.0064	0.0053	0.021
K_A (Hominid)	0.0127	0.0123	0.410	0.0126	0.0110	0.016
K_S (Hominid)	0.0026	0.0023	0.059	0.0024	0.0019	0.015

Analysis of recombination rates of genes according to their position in relation to rearrangements. Comparison of genes in regions involved in rearrangements *vs*. genes in colinear chromosomes or regions. Genes in breakpoints are included.

	Genes in Rearranged *vs.* Colinear chromosomes			**Genes within *vs.* outside inversions (excluding HSA2, PTR12, PTR13)**		
	Colinear	**Rearranged**	**P-value**	**Outside**	**Inside**	**P-value**
N	2610	3062		2022	601	
Recombination	1.0672	1.0895	0.639	1.0738	1.2418	0.070

Additional Table A3.

Comparison of genes in breakpoints *vs.* genes in other chromosomes or regions.

	Genes in breakpoints *vs.* inverted chromosomes (excluding HSA2, PTR12, PTR13)		
	Rearranged	**BKP**	**P-value**
N	1227	27	
K_A (Human)	0.0011	0.001	0.610
K_S (Human)	0.0065	0.0058	0.644
K_A (Chimpanzee)	0.0011	0.0009	0.673
K_S (Chimpanzee)	0.0062	0.0069	0.633
K_A (Hominid)	0.0123	0.0127	0.821
K_S (Hominid)	0.0023	0.0020	0.624

Comparison of recombination rates of genes in breakpoints *vs.* genes in other chromosomes or regions.

	Genes in breakpoints *vs.* inverted chromosomes (excluding HSA2, PTR12, PTR13)		
	Rearranged	**BKP**	**P-value**
N	2551	72	
Recombination	1.0896	1.9157	0.002

Additional Table A4.

Comparison of genes in regions involved in rearrangements *vs*. genes in colinear chromosomes or regions. Genes in breakpoints are excluded.

	Genes within *vs*. outside inversions (excluding breakpoints and HSA2, PTR12, PTR13)		
	Outside	**Inside**	**P-value**
N	972	255	
K_A (Human)	0.0012	0.001	0.121
K_S (Human)	0.0066	0.0061	0.351
K_A (Chimpanzee)	0.0012	0.0009	0.083
K_S (Chimpanzee)	0.0064	0.0052	0.013
K_A (Hominid)	0.0126	0.0108	0.013
K_S (Hominid)	0.0024	0.0019	0.037

Comparison of recombination rates in regions involved in rearrangements *vs*. genes in colinear chromosomes or regions. Genes in breakpoints are excluded.

	Genes within vs. outside inversions (excluding breakpoints and HSA2, PTR12, PTR13)		
	Outside	**Inside**	**P-value**
N	2020	531	
Recombination	1.0744	1.1473	0.424

Additional Table A5. Comparison of genes within simulated pericentric inversions *vs*. genes outside them.

	Genes in simulated pericentric inversions in colinear chromosomes (without HSA2 and without centromere)		
	Outside	**Inside**	**P-value**
N	989	181	
K_A (Human)	0.0013	0.0012	0.731
K_S (Human)	0.0063	0.0059	0.451
K_A (Chimpanzee)	0.0013	0.0010	0.166
K_S (Chimpanzee)	0.0063	0.0074	0.149
K_A (Hominid)	0.0125	0.0134	0.367
K_S (Hominid)	0.0026	0.0023	0.285

	Genes in simulated pericentric inversions in colinear chromosomes (without HSA2 and without centromere)		
	Outside	**Inside**	**P-value**
N	2182	428	
Recombination	1.0685	1.0607	0.937

Additional Table A6. Comparison of genes overlapping the microinversions described by (Newman et al. 2005).

	Genes overlapping microinversions *vs.* genes in rest of chromosomes		
	Outside	**Inside**	**P-value**
N	2603	17	
K_A (Human)	0.0012	0.0012	0.909
K_S (Human)	0.0064	0.0045	0.301
K_A (Chimpanzee)	0.0012	0.0013	0.766
K_S (Chimpanzee)	0.0063	0.0040	0.244
K_A (Hominid)	0.0125	0.0078	0.068
K_S (Hominid)	0.0024	0.0026	0.822

	Genes overlapping microinversions *vs.* genes in rest of chromosomes		
	Inside	**Outside**	**P-value**
N	5646	26	
Recombination	1.0795	1.015	0.875

Additional Table A7.Comparison of evolutionary rates of genes within inversions *vs*. genes outside inversions in individual chromosomes. Genes in breakpoints are excluded.

HSA1				HSA9				HSA16			
	Inside vs Outside (no BKP 1Mb)				Inside vs Outside (no BKP 1Mb)				Inside vs Outside (no BKP 1Mb)		
	Outside	inside	P-value		Outside	inside	P-value		Outside	inside	P-value
N	347	2		N	102	9		N	114		
K_A (Human)	0.0011	0.0000	0.161	K_A (Human)	0.0011	0.0012	0.874	K_A (Human)	0.0012		
K_S (Human)	0.0079	0.0153	0.220	K_S (Human)	0.0063	0.0076	0.551	K_S (Human)	0.0059		
K_A (Chimpanzee)	0.0013	0.0000	0.289	K_A (Chimpanzee)	0.0012	0.0014	0.849	K_A (Chimpanzee)	0.0012		
K_S (Chimpanzee)	0.0062	0.0000	0.194	K_S (Chimpanzee)	0.0072	0.0076	0.909	K_S (Chimpanzee)	0.0070		
K_A (Hominid)	0.0138	0.0111	0.715	K_A (Hominid)	0.0136	0.0149	0.748	K_A (Hominid)	0.0127		
K_S (Hominid)	0.0025	0.0000	0.235	K_S (Hominid)	0.0023	0.0025	0.848	K_S (Hominid)	0.0025		
HSA4				**HSA12**				**HSA17**			
	Inside vs Outside (no BKP 1Mb)				Inside vs Outside (no BKP 1Mb)				Inside vs Outside (no BKP 1Mb)		
	Outside	inside	P-value		Outside	inside	P-value		Outside	inside	P-value
N	84	28		N	74	71		N	20	91	
K_A (Human)	0.001	0.0016	0.155	K_A (Human)	0.0012	0.0007	0.076	K_A (Human)	0.0007	0.0009	0.513
K_S (Human)	0.0061	0.0048	0.301	K_S (Human)	0.0059	0.0059	0.966	K_S (Human)	0.0044	0.0062	0.366
K_A (Chimpanzee)	0.0011	0.0017	0.108	K_A (Chimpanzee)	0.0009	0.0009	0.882	K_A (Chimpanzee)	0.0007	0.0007	0.947
K_S (Chimpanzee)	0.0053	0.0051	0.837	K_S (Chimpanzee)	0.0069	0.0042	0.013	K_S (Chimpanzee)	0.0088	0.0058	0.105
K_A (Hominid)	0.0107	0.0100	0.659	K_A (Hominid)	0.0121	0.0093	0.072	K_A (Hominid)	0.0121	0.0115	0.823
K_S (Hominid)	0.0022	0.0034	0.064	K_S (Hominid)	0.0021	0.0017	0.318	K_S (Hominid)	0.0014	0.0017	0.658
HSA5				**HSA15**				**HSA18**			
	Inside vs Outside (no BKP 1Mb)				Inside vs Outside (no BKP 1Mb)				Inside vs Outside (no BKP 1Mb)		
	Outside	inside	P-value		Outside	inside	P-value		Outside	inside	P-value
N	109	54		N	86			N	29		
K_A (Human)	0.0016	0.0011	0.118	K_A (Human)	0.0011			K_A (Human)	0.0009		
K_S (Human)	0.0052	0.0062	0.374	K_S (Human)	0.0066			K_S (Human)	0.0053		
K_A (Chimpanzee)	0.0011	0.0008	0.262	K_A (Chimpanzee)	0.0012			K_A (Chimpanzee)	0.0009		
K_S (Chimpanzee)	0.0075	0.0052	0.081	K_S (Chimpanzee)	0.0053			K_S (Chimpanzee)	0.0040		
K_A (Hominid)	0.0124	0.0112	0.449	K_A (Hominid)	0.0113			K_A (Hominid)	0.0088		
K_S (Hominid)	0.0028	0.0020	0.113	K_S (Hominid)	0.0024			K_S (Hominid)	0.0018		

Comparison of Recombination rates of genes within inversions vs. genes outside inversions in individual chromosomes. Genes in breakpoints are excluded.

	HSA1				HSA9				HSA16		
	Inside vs Outside (no BKP 1Mb)				Inside vs Outside (no BKP 1Mb)				Inside vs Outside (no BKP 1Mb)		
	Outside	inside	P-value		Outside	inside	P-value		Outside	inside	P-value
N	763	6		N	191	17		N	215		
Recomb Rate	1.256	0.2597	0.200	Recomb Rate	1.268	1.0998	0.752	Recomb Rate	0.8088		
	HSA4				HSA12				HSA17		
	Inside vs Outside (no BKP 1Mb)				Inside vs Outside (no BKP 1Mb)				Inside vs Outside (no BKP 1Mb)		
	Outside	inside	P-value		Outside	inside	P-value		Outside	inside	P-value
N	182	66		N	159	168		N	39	169	
Recomb Rate	0.9805	1.1491	0.511	Recomb Rate	0.8383	1.1926	0.048	Recomb Rate	1.1714	1.2167	0.908
	HSA5				HSA15				HSA18		
	Inside vs Outside (no BKP 1Mb)				Inside vs Outside (no BKP 1Mb)				Inside vs Outside (no BKP 1Mb)		
	Outside	inside	P-value		Outside	inside	P-value		Outside	inside	P-value
N	195	103		N	192			N	72		
Recomb Rate	0.8698	1.0387	0.369	Recomb Rate	0.9759			Recomb Rate	1.072		

Additional Table 8. Comparison of evolutionary breakpoints between human and chimpanzee based either in cytological or "in silico" approaches (both coordinates are based on Human assembly Build 34).

Cytological Approaches				"in silico" Approach			
Hum. Chr	INV start	INV end	*Reference*	Hum. Chr	INV start	INV end	*Reference*
HSA1				HSA1	112870424	145835091	*10*
HSA2	114347090	114455823	*1*	HSA2			
HSA4	44730692	86461364	*2*	HSA4	44558445	86436221	*10*
HSA5	18443766	96071773	*3*	HSA5	18417476	95998631	*10*
HSA9	40390489	84428949	*4*	HSA9			
HSA12	20833487	66695639	*5*	HSA12	20854309	66688318	*10*
HSA15	17000000	28486050	*6*	HSA15	28637194		
HSA16	35254239	46289682	*7*	HSA16	35278710	46359581	*10*
HSA17	8128215	48224281	*8*	HSA17			
HSA18	5961	16898525	*9*	HSA18	134812	16930430	*10*

1 Fan Y. et al, 2002. Genome Research 12:1651-1662 // Hillier, L.W., et al, 2005. Nature 434:724-731

2 Kehrer-Sawatzki H. Et al, 2005. Human Mutation 25: 45-55

3 Szamalek J.M. et al, 2005. Hum. Genet. 117: 168-176

4 Kehrer-Sawatzki H. et al, 2005. Genomics 85: 542-550

5 Kehrer-Sawatzki H. et al, 2005. Cytogenet. Genome Res. 108: 91-97

6 Locke D.P. et al, 2003. Genome Biol 4: R50

7 Goidts V. et al, 2005. Genome Res. 15: 1232-1242

8 Kehrer-Sawatzki H. et al, 2002. Am. J. Hum. Genet. 71: 375-388

9 Goidts V. et al, 2004. Genomics 83: 493-501

10 Mikkelsen, T.S et al. 2005. Nature 437: 69-87..

Additional Table 9. Chi square test to detect accumulation of genes with high Ka/Ki ratios.

K_A/K_I	Colinear Chr.	Rearrranged Chr.	Total	*Degrees of freedom: 1*
< 1	6052	5666	11718	*Chi-square = 11.126*
> 1	250	167	417	*p is less than or equal to 0.001.*
Total	6302	5833	12135	*The distribution is significant.*

K_A/K_I	Inside	Outside	Total	*Degrees of freedom: 1*
< 1	1083	4583	5666	*Chi-square = 0.579*
> 1	28	139	167	*For significance at the .05 level, chi-square should be greater than or equal to 3.84.*
Total	1111	4722	5833	*The distribution is not significant.*

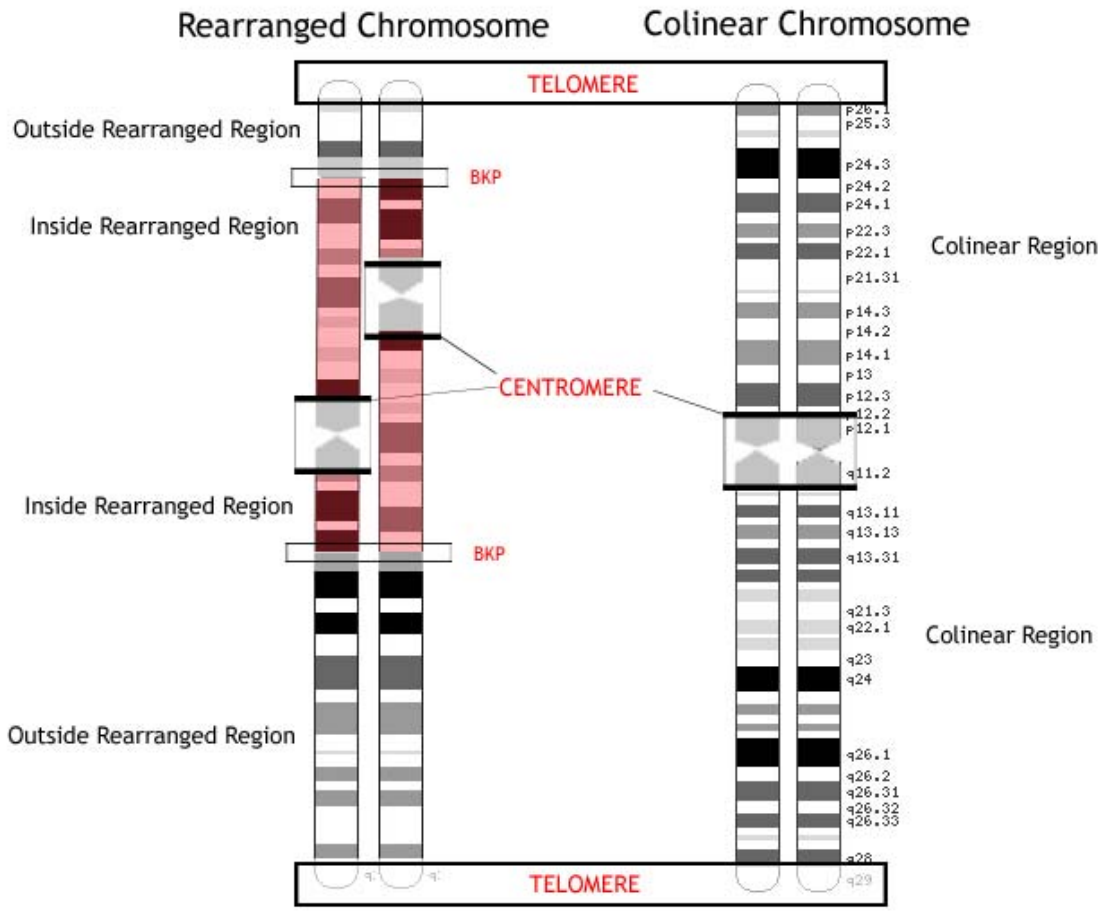


Figure 1. Abstract overview of the regions in a chromosome included and excluded form our analysis. A colinear and an inverted chromosome are presented. The inversion in the rearranged chromosome is highlighted in red. For every chromosome, regions considered in this paper are labelled in black. Regions excluded from the main analysis (telomeres, centromeres and breakpoints (BKP)) are within boxes and labelled in red.

The genomic distribution of intra and inter divergence of human SD.

(In preparation)

Tomàs Marques-Bonet, Ze Cheng, Xinwei She, Evan E. Eichler, Arcadi Navarro

Title: The genomic distribution of intra and inter divergence of human SD.

Authors: Tomàs Marques-Bonet[1], Ze Cheng[2], Xinwei She[2], Evan E. Eichler[2], Arcadi Navarro[1].

Addresses:

[1]Unitat de Biologia Evolutiva Departament de Ciències Experimentals i de la Salut. Universitat Pompeu Fabra. Barcelona, Catalonia, Spain.

[2]Department of Genome Sciences, University of Washington School of Medicine, Seattle, Washington 98195, USA.

Corresponding author:

Arcadi navarro

Unitat de Biologia Evolutiva.

Departament de Ciències Experimentals i de la Salut

Universitat Pompeu Fabra

Doctor Aiguader 80

08003 Barcelona (Spain)

Phone: +34 93 5422841

Fax: +34 93 5422802

E-mail: arcadi.navarro@upf.edu

ABSTRACT

Different studies have unveiled the footprint of biological phenomena such as changes in gene regulation caused by chromosomal rearrangements or increased sequence divergence due to, perhaps, speciation processes triggered by chromosomal rearrangements. So far, all these studies have focused on the relationship between structural changes and the rates of divergence of linked single-copy DNA and have tried to exclude Segmental Du6plications (SDs). Here we take the opposite view and focus on SDs, since they are one of the primary forces driving the evolution of structure and function in our genomes and been linked not only with novel genes acquiring new functions, but also with overall higher DNA sequence divergence and major chromosomal rearrangemen We analyze the whole-genome distribution of intraspecific divergence between paralogous copies of human SD and of interspecific divergence between SDs shared by humans and chimpanzees. We study how these divergence patterns relate to chromosomal rearrangements, while considering other factors. We find that interspecific SD divergence behaves similarly than divergence of single-copy DNA. In contrast, old and recent paralogous copies of SDs do present different patterns of intraspecific divergence. Also, we show that recent SDs accumulate in regions that carry an inversion in sister lineages, thus supporting the idea that some DNA fragments harbor highly active sequences that can lead to both, a burst of SDs and chromosomal rearrangements.

INTRODUCTION

Initial analyses of the human genome sequence have identified a large amount of interspersed segmental duplications (SDs) (Bailey et al., 2001). SDs, defined as blocks of DNA ranging from 1-400kb in length, copies of which can be found in multiple sites and that typically share high sequence similarity (>90%). Studies based in both experimental and computational analyses show that ~5% of the human genome is composed of duplicated sequences. The distribution of these duplications is non-uniform within and among chromosomes with a tendency to cluster in pericentromeric and subtelomeric regions (Bailey et al., 2002; Cheng et al., 2005; Cheung et al., 2003; Linardopoulou et al., 2005; Sharp et al., 2006; She et al., 2006) and in the breakpoints of chromosomal rearrangements (Armengol et al., 2003; Bailey et al., 2004; Ji et al., 2000; Samonte & Eichler, 2002; Sharp et al., 2005)

SDs are an important feature of genome evolution, having both functional and structural effects (Bailey et al., 2004; Bailey et al., 2002; Bailey et al., 2001; Cheng et al., 2005; Coghlan et al., 2005; Eichler & Sankoff, 2003; Sharp et al., 2006; Stankiewicz et al., 2004), By predisposing

chromosomal architectures to be rearranged by non-allelic homologous recombination (Ji et al., 2000; Lupski, 1998; Sharp et al., 2006; Shaw & Lupski, 2004; Stankiewicz & Lupski, 2002) SDs constitute genetic risk factors for many diseases (e.g. Prader-Willi, Williams-Beuren Syndromes, juvenile nephronophtisis or spinal muscular atrophy). Also, SDs are related with genic evolution because they produce partial or full coding sequence duplications that can lead to genes with new functions (Conrad et al., 2006; Courseaux & Nahon, 2001; Eichler, 2001; Johnson et al., 2001; Jordan et al., 2004; Sebat et al., 2004; Sharp et al., 2006; Zhang et al., 2003). Moreover, evolutionary rates of duplicated genes have been shown to be accelerated just after the duplication event. These accelerations could be due to either a higher rate of mutation or to a relaxation of purifying selection due to the duplication of functional genes or to the action of positive diversifying selection on one or both copies (Conant & Wagner, 2003; Kondrashov et al., 2002; Lynch & Conery, 2000; She et al., 2006; Zhang et al., 2003).

Finally, the fact that SDs are associated to rearrangements has led some authors to propose a strong relationship between SDs and large-scale genome structure. Again, the fact that SDs predispose chromosomes to suffer rearrangements suggests that SDs may be the main force driving the evolution of genomic structure along the lineages of man mammalian species (Armengol et al., 2003; Bailey et al., 2004; Ji et al., 2000; Samonte & Eichler, 2002; Sharp et al., 2005). Other studies, however, point to both SDs and chromosomal rearrangements as different manifestations of the intrinsic instability of some particular DNA sequences (Bailey et al., 2004; Coghlan et al., 2005).

Recently, interest on the role of chromosomal rearrangements in speciation processes has been renewed. Models of chromosomal speciation based on the reduction of recombination induced by rearrangements, pose that regions involved in those rearrangements could become isolated earlier compared to the rest of the genome (Navarro & Barton, 2003a; Noor et al., 2001; Rieseberg, 2001). Moreover, these models predict an association between higher evolutionary rates of DNA sequence in those regions involved in any speciation process. Current evidence for or against such models is extremely contradictory. In human-chimpanzee comparisons, higher evolutionary rates were originally linked to chromosomal rearrangements (Lu et al., 2003; Navarro & Barton, 2003b; Navarro et al., 2003), whereas other studies found no effect (Vallender & Lahn, 2004; Zhang et al., 2004) and even the most recent ones have detected lower evolutionary rates within inversions (Marques-Bonet et al., 2007). In other lineages, results are more consistent, with the main finding of higher evolutionary rates associated with chromosomal

rearrangements having been confirmed in several studies (Armengol et al., 2005; Lindblad-Toh et al., 2005; Marques-Bonet & Navarro, 2005).

Other explanations have been proposed to account for the relationship between chromosomal rearrangements and faster or slower evolutionary rates. For example, chromosomal rearrangements can act over divergence of DNA simply by inducing changes in genomic contexts. That is, if some DNA fragments are moved by a chromosomal inversion from a region of low recombination to a region where recombination is high, this could affect mutation rates and, thus, divergence (Cooper & Youssoufian, 1988; Sved & Bird, 1990). Also, it is also possible that rearrangements might tend to occur or to be fixed in regions of relaxed purifying selection and, thus, of faster genic evolution (Lu et al., 2003; She et al., 2006).

Independently of how the issue of the relationship between the evolutionary rates of DNA sequences and chromosomes, it is important to consider the possibility of an association between SDs and chromosomal rearrangements in relation with speciation. If rearranged chromosomes, whose breakpoints are enriched with SDs, take part in speciation processes in which individuals bearing different chromosomal structures become genetically isolated, it is possible that novelties contained in these duplications play some role in these models.

To tackle this issue we must start by understanding the rates and patterns of SD divergence in the primate lineages. Here, we analyze the genomic distribution of divergence among paralogous copies of human SDs and among chimpanzee and human duplicated regions. We take into account not only chromosomal rearrangements, but also other genomic factors that have been shown to affect evolutionary rates of single copy DNA such as, for example, linkage to the X chromosome (HSAX) (Li et al., 2002; Marques-Bonet & Navarro, 2005; Wolfe & Sharp, 1993), or to telomeric and centromeric regions (Marques-Bonet et al., 2007; Rudd & Willard, 2004; She et al., 2004)) (Marques-Bonet et al., 2007; Mikkelsen et al., 2005)).

MATERIALS AND METHODS

Structural information.
Coordinates of telomeres and centromeres of all chromosomes were obtained from Build 35 of the human genome (http://genome.ucsc.edu) and NCBI Build 1 of the chimpanzee genome (http://genome.ucsc.edu). We considered as rearranged chromosomes all those for which major chromosomal rearrangements in either the human or the chimpanzee lineages have been evidenced by recent in

silico (Feuk et al., 2005; Mikkelsen et al., 2005) or cytological structures (Goidts et al., 2004; Kehrer-Sawatzki et al., 2005a; Kehrer-Sawatzki et al., 2005b; Kehrer-Sawatzki et al., 2002; Kehrer-Sawatzki et al., 2005c; Szamalek et al., 2005). This comprised HSA1, HSA4, HSA5, HSA9, HSA12, HSA15, HSA16, HSA17 and HSA18, which differ by a pericentric inversion, and human chromosome 2, which has been generated by an ancestral telomere-telomere fusion (Yunis & Prakash, 1982). For all chromosomes, all in silico-estimated coordinates were compared with newly available cytological data in order to confirm inversion coordinates. When indicated, the mini-inversions detected "in silico" by (Newman et al., 2005) have been used.

Source of SD data

We retrieved information of segmental duplication about Human and Chimpanzee SDs from the Segmental Duplication Database (http://humanparalogy.gs.washington.edu/ and http://chimpanzeeparalogy.gs.washington.edu/). In brief, we used the whole genome assembly comparison (WGAC), composed by SDs that were detected by the Blast-based method (Bailey et al., 2001) to identify pairwise of DNA sequence of high similarity within the human assembly (Built 35).

Three datasets were built for analysis.

1) Dataset 1. Raw dataset. This is the standard dataset as downloaded from the Segmental Duplication Database. It contains pairs of coordinates of fragments of the human genome that fit two criteria: each pair has a minimum overlap size of 1kb and presents >90% identity among copies. (Bailey et al., 2001). A divergence measure was calculated for every pairwise detection as the number of substitutions per site (applying Jukes-Cantor correction). Besides divergence we also recorded the overlapping size (length) of every pair.
2) Dataset 2. Non-overlapping intraspecific dataset. Because of the methodology used in WGAC, most fragments in the raw dataset are repeated in many partially overlapping pairs, thus adding the same information several times especially in SD clusters. To eliminate this redundant information, we constructed a new dataset containing samples of SDs representative of every region of the genome covered by SD. The steps used to construct our new dataset were as follows:

2.a. We constructed a "coverage map of SDs". We recorded the bound coordinates of overlapping SDs thus reporting every region in the human genome in which there are SDs. If two coverage zones were separated by a distance lower than 10 kb we joined them to avoid over-representing some parts of the genome. This procedure is

similar that the one used in (She et al., 2006), when constructing "duplication hubs", that is, regions with an excess of aligned SDs.

2.b. From this coordinate list and for every "covered" region, we kept only one pair of SD as a representative of the region. The criteria to select one SD against the others were (1) Longer SDs were preferred, as measured by percentage of occupancy within that coverage zone and (2) SDs that had both paralogous copies in the same class of regions. That is, if one coverage zone is in, say, a telomere, we kept the longer SD having its paralogous copy also in a telomere. In case of not having copies in comparable regions, we just keep the longest. We considered seven classes of genomic location: sex chromosomes, telomeres, centromeres, HSA19, colinear chromosomes, colinear regions in rearranged chromosomes, rearranged regions (inversions) and rearrangement breakpoints. The goal of these criteria is to retrieve some nonredundant basic information of this portion of the genome. (see supplementary Figure 1 for a schematic view of the process).

3) Dataset 3. Non-overlapping interspecific dataset. This third dataset was designed to recover a sample of divergence between humans and chimpanzees in regions covered by SDs. From the coverage map of human SD we recovered chimpanzee WGS (v1) sequences (She et al., 2006). For every "covered" zone (a slice of coordinates), we split it in non overlapping windows of 5000 bp. For every one of those windows, divergence was calculated as the average of all chimpanzee WGS sequences against the human sequence. Finally the average of all windows was computed as the average divergence of the coverage zone. Divergence was calculated applying Kimura's correction. We also constructed a parallel dataset computing divergence of the chimpanzee SDs from human WGS sequences.

These three datasets were built in order to tackle different questions. To detect clusters of SDs in some parts of the genome we used the raw dataset, which provides a good perspective of the amount of SDs in every region. When we aim to study divergence in different regions of the genome while avoiding some biases such as overlapping SDs or copies in non-comparable regions of the genome, we should use the non overlapping datasets, either for intraspecific divergence (dataset 2) or interspecific divergence (dataset 3).

Filtering.

Previous to every analysis we performed a sequential filtering process to remove the genomic variables that are known to affect evolutionary rates. After getting the result for each one of the categories, those SDs

located in that specific category were removed from the analysis.

Permutation tests

SDs divergence measures in different categories were compared by means of pairwise permutation tests (based on 1000 permutations). Empirical P-values in such tests, are calculated as the proportion of times that the difference of averages between two categories in a permuted dataset is equal or larger than the observed difference.

RESULTS

1) Overrepresentation of young SDs in rearranged regions.

We started using the raw dataset (Dataset 1) to study the general distribution of paralogous copies of human SDs relative to major rearrangements between humans and chimpanzees. We defined "young" SDs as those with a greater percentage of sequence identity among copies (> 98% ID). To avoid noise, we removed SDs associated to factors that are well know to modify evolutionary rates in single-copy DNA (such as sex chromosomes, telomeres (10Mb from the tip of chromosome), centromeres (5Mb) and human chromosome 19 (HSA19)) we also eliminated pairs of SDs that had one copy in rearranged regions and the other copy in colinear regions. After this filtering we observed a higher proportion of young SDs within rearranged regions than outside them: ~40% of SDs located within rearranged regions are young, when this figure is ~12% for SDs outside inversions in rearranged chromosomes. Also, inverted regions showed higher percentage of accumulation of young SDs when compared with colinear chromosomes, where only 11% of SDs are young (Table 1, Figure 1). Most of the rearrangements took place in the chimpanzee lineage (Szamalek et al., 2006) , and here we are analyzing human SDs, and thus this association would not be related with an accumulation of SDs within the inversion itself, but in the homologous chromosome of the sister species in which the rearrangement took place.

To check whether these results were due to an overall phenomenon or were driven by some particular chromosome, we performed a chromosome per chromosome analysis. This allowed us to pinpoint HSA5 and HSA9 as primarily responsible for the reported association, since these chromosomes present not only the largest differences in percentages of identity but also the largest sample size (total number of SD pairs). No other chromosome showed differential accumulation of identity-specific SDs. (Figure 2). Therefore the association above is mainly due to these two chromosomes, which, being inverted in one lineage (chimpanzee) have accumulated an expansion of recent SDs in its sister linage (human).

SDs tend to cluster within pericentromeric and subtelomeric zones (Bailey et al., 2001; She et al., 2004; She et al., 2006). To test whether these accumulations of young SDs in regions that contain pericentric inversions in other lineages are due to regions near centromeres (even if 5Mb around centromeres had already been filtered out), rather than to some other cause, we simulated pericentric inversions in colinear chromosomes. Pseudo-inverted pericentric regions in colinear chromosomes were defined as regions equivalent in length and location to real rearrangements. Given that the average inversion spans 24.98% of its chromosome, we create a virtual inversion of that size in each colinear chromosome, keeping centromere as the middle point of the inversion. The virtual inversions did present a higher proportion of young SDs, but the increase was only 50% of that in real inversions (Table 2, Figure 3). When looking for individual trends in individual colinear chromosomes (Figure 4), only HSA10 and HSA7 seem to be accumulating some local clustering of recent SDs. However, the clustering is not exclusive of the inverted region as in HSA5 and HSA9 but extends all over the chromosome. The rest of colinear chromosomes did not show any particular age distribution of SDs inside vs. outside virtual rearrangements, allowing us to discard that the association of young SDs and certain rearranged chromosomes is due to SDs accumulating near centromeres.. HSA7 was also highlighted by (Zhang et al., 2005) as showing considerable enrichment of SDs.

2) SDs and factors that are known to shape evolutionary rates.

To study how factors such as the location in sex chromosomes or telomeres affect the rates of intraspecific evolution of SD we used our non-overlapping dataset (Dataset 2). We split it in two subsets: young SDs (> 98% ID) and old SDs (< 92% ID). The coverage map to create the non-overlapping dataset (see Methods) was constructed a posteri of the splitting between young and old duplications. This was done to avoid a bias in our algorithm of choosing a sample SDs for every region, because old segmental duplications are shorter than young ones (She et al., 2006) and following our criteria (for instance the higher coverage criterion) they would have had lower probabilities of being selected as a sample of the region of interest. Again, we selected SDs so as to keep as the representative of every covered zone those SDs that had both copies in the same region of interest (see supplementary Figure 2).

To analyze how those factors that affect evolutionary rates in single-copy DNA sequences are affecting divergence among paralogous copies of human SDs (intra-specific divergence), we sequentially analyzed and removed every individual factor (Table 3). We also included the sizes

of SDs in our studies. Length can be informative since, as mentioned above, SDs become shorter with time (She et al., 2006), probably as a result of recombination or subsequent deletion events that breakdown their structure.

As to sex chromosomes, young human SDs located in the HSAX presented less divergence among copies than recent SDs located in autosomes. This is not the case for old SDs. No length differences were detected in SDs located in HSAX. When located in HSAY, young SDs presented lower intra-specific divergence and increased length. Old SDs also showed an increase of length when located in HSAY, but, in contrast, they present higher divergence between paralogous copies.

Regarding position along chromosomes, we first considered telomeres. Only young SDs located in telomeres showed higher divergence between paralogous copies. They also showed shorter alignment sizes. On the contrary, old SDs did not show any divergence differences between telomeres and the rest of the genome. When focusing on centromeres, we found that SDs near them are longer in both subsets (young and old SDs). As to divergence, only old SDs showed a slight decrease of paralogous divergence in pericentromeric regions compared to SDs located elsewhere in the genome.

Although HSA19 has been shown to have specific divergence and nucleotide composition patterns and its SD appear to have a deficit of interspersed (as opposed to tandem) duplication on this chromosome (Castresana, 2002; She et al., 2006). SDs located in this chromosome did not present any differences, neither in sizes or divergence compared with SDs located in the rest of the autosomes.

When we finally compared paralogous copies of human SDs located in rearranged chromosomes versus SDs located in colinear chromosomes, the only detectable patterns were that young SDs are significantly longer and less divergent when located in rearranged chromosomes. However, the observation can not be exclusively attributed to inversions, because when comparing divergence among copies of human SDs within the inverted zones (recall that most rearrangements took place in the chimpanzee lineage) versus outside the inversion in rearranged chromosomes, there were no divergence differences although we SDs where longer within rearranged regions. Since SDs have been shown to be enriched in regions near evolutionary breakpoints in many species (Armengol et al., 2003; Bailey et al., 2004; Ji et al., 2000; Samonte & Eichler, 2002; Sharp et al., 2005), we assessed the sequence features of SDs located at evolutionary BKPs in human chromosomes. Neither the length nor the divergences of those SDs are statistically different from

SDs located elsewhere in the genome. Finally, we considered the inversions detected "in silico" from (Newman et al., 2005), that SDs located within these inversions showed a slight increase in divergence (highly significant for old SDs and marginally for young SDs) although small sample size ascertainment can not be excluded. As expected, only young SD showed a remarkable increase of length within those rearrangements.

3) Human-Chimpanzee divergence.

We retrieved a sample of divergence between human and chimpanzee (interspecific divergence) in the duplicated regions, using the raw non-overlapping dataset (Dataset 3). Basically we had two sets of SDs for which we have measures of chimpanzee divergence from human for non-overlapping human SDs and measures of human divergence from chimpanzees for non-overlapping chimpanzee SDs. Again, we studied the effect of all the factors considered above in the divergence of SDs among species by analyzing and removing every individual factor sequentially (Table 4).

Our first observation was that SDs located in HSAX showed lower divergence than SDs located in autosomes. This effect was consistent for both datasets of inter-specific divergence in SDs (human and chimpanzee). Second, and as previously seen for single copy genes in other studies (Marques-Bonet et al., 2007; Mikkelsen et al., 2005), regions near telomeres also showed higher divergence than the rest of the chromosome. This pattern was observed again in both human and chimpanzee SDs. In contrast, and opposite to other studies (Marques-Bonet & Navarro, 2005; Marques-Bonet et al., 2007) inter-specific divergence in SDs is higher near pericentromeric regions. Finally, HSA19 clearly stood out as a highly divergent chromosome as to human and chimpanzee divergence in SDs, again supporting the aforementioned studies (Table 4).

When studying the effect of rearrangements over divergence in SDs between humans and chimpanzees, and in agreement with the most recent results for single copy genes (Marques-Bonet et al., 2007), divergence in duplicated regions was found significantly lower within rearranged chromosomes than in colinear chromosomes. Opposite to previous results, however, there were no significant differences when comparing divergence in SDs located within vs. outside rearranged regions (basically inversions in the chimpanzee lineage) in neither of the two directions studied (human divergence from chimp SDs and vice versa). Finally, and again differing from results in single copy genes, SDs located within small inversions (Newman et al., 2005), revealed lower divergence rates compared to SDs located elsewhere in the genome.

To unveil any specific individual contributions of chromosomes, we analyzed interspecific-divergence for every inversion (Table 5). There was no clear pattern to be detected as to divergence (neither in one direction nor the other). Only HSA9 had higher human divergence within its inversion and only considering chimpanzee duplicated regions.

DISCUSSION

Several conclusions arise from our whole-genome SDs analysis. First, there is an accumulation of recent human SDs within certain chromosomes that carry an evolutionary rearrangement between human and chimpanzees. Seven of the nine major inversions between human and chimpanzee occur in the chimpanzee lineage (HSA4, HSA5, HSA9, HSA12, HSA15, HSA16 and HAS17), and, thus, there cannot be any direct effect of the inversions causing accumulation of young SDs in humans. It is, however, tempting to speculate that both observations (chromosomal rearrangements and SDs) may be, in fact, consequences of a third factor, perhaps regions of high instability (Johnson et al., 2006), as opposed to the classical explanation that rearrangements and SDs are related only because highly similar regions promote rearrangements by non-allelic recombination (Armengol et al., 2003; Bailey et al., 2004; Ji et al., 2000; Samonte & Eichler, 2002; Sharp et al., 2005). That would be the case of HSA5 and HSA9, in which there is an excess of young human SDs (> 98 % ID) within chimpanzee-specific inverted regions.

HSA5 has also been pointed out as one of the two chromosomes in which there is an excess of human-specific SDs. Several different authors have found that the association among rearrangement breakpoints and segmental duplications is only maintained between different lineages, but not within the same lineage (Bailey et al., 2004; Cheng et al., 2005; Coghlan et al., 2005). For instance, primate segmental duplications occur at specific locations that are enriched for mouse-human synteny and mouse-rat synteny breaks. As the majority of synteny rearrangements have occurred in rodent lineage there is not a cause and consequence relationship but rather primate segmental duplications at the same locations in which rodent chromosomes have rearranged. Instability thus, would seem, a long standing property of these genomes at these locations. In addition, (She et al., 2006) described a non-uniform distribution of intrachromosomal human SDs and highlighted nine autosomal human chromosomes with an excess of representation of young human duplications seven of which presented rearrangements between humans and chimpanzees (5 of which were chimpanzee specific), thus reinforcing the idea of a link between expansions of recent SDs in one lineage and chromosomal rearrangements in the

other. Only deeper analysis in chimpanzee of the two chromosomes that carry human specific rearrangements (HSA1 and HSA2) and also show excess of young human chromosomal SDs will help to clarify any direct relationship among chromosomal rearrangements and expansion of SDs. This analysis, however, is beyond the scope of the present work and will require a high quality sequence assembly of the chimpanzee genome which correctly annotates segmental duplications..

As to why chromosomal rearrangements and (especially inversions) and young SDs should accumulate in sister lineages, several explanations can be discussed. The first one relates with the aforementioned instability regions. A recent change on the understanding of the evolution and behavior of SDs. (Johnson et al., 2006; Zody et al., 2006a; Zody et al., 2006b) poses that there are "core elements" that may act as sources for the dispersal of new SDs, by creating an excess of copies of themselves. Those copies tend to cluster by means of tandem duplications. Thus, an explanation for our results would be that some core elements were present in the ancestral chromosomes to those of humans and chimpanzees that currently harbour inversions and SDs. As inversions decrease recombination between homologous chromosomes (Navarro & Barton, 2003a; Noor et al., 2001; Rieseberg, 2001), core elements would not be able to escape from their source regions while rearrangements are still segregating in the ancestral population and thus, they would accumulate copies of itself locally in one lineage while the rearranged regions would remain in the other. Moreover, this reduction of recombination would also prevent the dispersal of the rest of SDs (not just the "core" elements). SDs trapped within rearrangements would be more similar to the "original" state because they would be prevented to invade other regions or chromosomes that could affect mutation rates that lead to divergent SDs.

A second possibility, related to the previous explanation, is that lower recombination rates within inversions could help explaining our results. As suggested in previous works (Eyrewalker, 1993; Hardison et al., 2003; Hellmann et al., 2003; Meunier & Duret, 2004), there is a positive correlation among low recombination rates, low diversity within species and low divergence that can be explained by a mutagenic effect of recombination. Then, regions within rearrangements should have less divergence (either inter-specific or intra-specific) provided that rearrangements segregated for long enough in the population and, thus, reduced recombination for a significant amount of time..

Finally, some of the pairwise alignments classified as young SDs (high identity SDs), may in fact not be young, but their high

identity may have been maintained by gene conversion events (Cheng et al., 2005). Gene conversion is a homogenizing force that might erase differences among copies leading to underestimation their age of the SD. It is be possible that during the segregation of new rearrangements, the resolving structure of the few recombination events within inversion would be biased towards increased gene conversion instead of the reciprocal exchange of chromatids. This would help explaining and excess of highly similar tracks of SDs in one lineage together with inversions in the other lineage, however, this implies that most conversions ought to have happened before the separation of the two lineages and while the inversions were segregating in the population, which makes this explanation unlikely. Moreover (She et al., 2006) did consider that gene conversion events can not explain most of the identity of SD copies.

The second main conclusion coming out from our analysis is that old SDs do present different trends than more recent SDs as to their relationships with factors known to shape evolutionary rates in single copy DNA sequences and that, in general, young and old SDs are not equally related to such genomic factors, hinting at different evolutionary histories for different SD classes. It might be the case that young SDs are reflecting the history of recent primate evolution that lead to our species, while old SDs may reflect periods of duplication during early primate evolution, and thus they can shed some light on the very beginnings of the evolution of our respective lineages. Our results, for example, could reflect a recent expansion of young SDs (for instance in X chromosome, as has been suggested by (She et al., 2006)) or be the result of the complex interaction among recombination and SDs. That would be the case for telomeres for instance, in which young SDs are marginally more divergent, but very significantly shorter than elsewhere in the genome, maybe as a result of telomeres having higher rates of recombination (Kong et al., 2002; Myers et al., 2005). Also, old SDs do not present this trend, but this is expected since telomeres are likely to have moved during primate evolution. HSAY also presents larger segments of SDs, and this could be well related with the lack of recombination of chromosome. Regarding centromeres and probably as a result of their decreased recombination (Kong et al., 2002; Myers et al., 2005) we obtained larger sizes of pairwise alignments of SDs. However, as centromeres have been reported to be prone to be repositioned during evolution (Murphy et al., 2005), this result could be reflecting some other cause rather than a direct centromere effect. On the other hand, we should keep in mind that telomeres are not included in this comparison because they were removed in the previous analysis, so these latter results can not be biased as a result of differential features of telomeric SDs.

As to major rearrangements between humans and chimpanzees and their relationship with SDs sizes and divergences, our main conclusion is that only young SDs located in rearranged chromosomes are more similar among them -and longer- than SDs located in colinear chromosomes. This could be expected, since all rearrangements are known to be either human or chimp specific and thus old SDs should not be affected by such recent rearrangements. Still, young and old paralogous copies of SDs tend to be larger within rearranged zones compared with SDs of the rest of rearranged chromosomes (this is also the case for smaller rearrangements detected "in silico" (Newman et al., 2005)). This is a puzzling pattern hinting again at some period of decreased recombination in rearranged regions. Finally, we observed higher levels of intraespecific divergence between SDs within smaller inversions (Newman et al., 2005). Altogether, these data suggest that chromosomal rearrangements might affect evolutionary rates among copies of SDs during the recent primate evolution.

Our third and last finding is that interspecific divergence of SDs shows patterns that are roughly equivalent to those of single-copy DNA as to genomic variables such as sex chromosomes, telomeres, centromeres or HSA19 (Marques-Bonet et al., 2007). SDs located in telomeres and in HSA19 show higher levels of interspecific divergence for SDs. Also, SDs located in rearranged chromosomes show less divergence between species. Still there are some discrepancies between single-copy and duplicated DNA, such as the higher divergence in SDs located in centromeres or the lower divergence of SDs within small inversions (Marques-Bonet et al., 2007). Finally, HSAY does not show the previously reported higher degree of divergence (Armengol et al., 2005; Marques-Bonet & Navarro, 2005; Marques-Bonet et al., 2007), perhaps as the result of the recent expansion of young SDs in that chromosome (She et al., 2006) of extensive gene conversion (Rozen et al., 2003).

Regarding the study of individual inversions, we should point out that, in spite of sample size being small, HSA9 stands out the unique chromosome showing significantly higher human-chimpanzee divergence within rearranged zones. This suggests either a burst of interspecific divergence within the inversion due to an undetermined cause or that SDs within the inversion of HSA9 played a role in a chromosomal speciation event separation our two lineages. Right now, therefore, HSA9 is the best candidate to further study any potential relationship among SDs, rearrangements, divergence and speciation event. If chromosomes have played any role in any of the speciation events that leaded to humans and chimpanzees, it is clear that not all the chromosomes would

have the same contribution and thus would not leave the same traces in our genomes. Its is perhaps this help explaining that HSA4, stands out as a highly divergent within its inversion for both single-copy coding and non-coding sequences (Marques-Bonet et al., 2007) and does not present any particular pattern when considering its duplications. Also, certain chromosomes (such as HSA4, HSA5, HSA9, HSA15 and HSA16) have been pinpointed as the most different between humans and chimpanzees in terms the expression patterns of their genes (Marques-Bonet et al., 2004), which is only partially consistent with the results presented here.

In summary, we conclude are that intra and interspecific divergence between SDs are affected by factors that were known to affect divergence of single copy DNA sequences. Although chromosomal rearrangements do affect the evolution and fate of SDs, chromosomal speciation (and its relation with SDs novelties) does not seem to have been a common process along the human and chimpanzee lineages, although we should consider HSA9 as the best possible candidate to have been involved in some complex interaction among rearrangements, SDs and evolutionary novelties contained in these duplications. Studies including more species and focusing on the powerful novelty generating force that duplications represent and on their relationship to recombination and chromosomal structures are needed to increase our knowledge about this exciting topic.

ACKNOWLEDGEMENTS

We want to thank J. Bertranpetit, E. Gazave, O. Fernando, M. Przeworski, T. Newman, A. Sharp and the members of the Evolutionary Biology Unit in UPF and Evan Eichler's lab at University of Washington for enriching discussions and lots of help during the preparation of this work. A. N. is a member of the Ramón y Cajal Program (Spanish Government). This research was supported by grants to A.N. from the Ministerio de Ciencia y Tecnologia (Spain, BOS 2003-0870) and the Genome Canada-Genoma España Joint R+D+I Projects in Human Health (JLI/038) and by BE2006 fellowship to T.M.B from the "Departament d'Educació i Universitats de la Generalitat de Catalunya".

BIBLIOGRAPHY

Armengol L., Marques-Bonet T., Cheung J., Khaja R., Gonzalez J. R., Scherer S. W., Navarro A., and Estivill X. (2005). Murine segmental duplications are hot spots for chromosome and gene evolution. Genomics 86: 692-700.

Armengol L., Pujana M. A., Cheung J., Scherer S. W., and Estivill X. (2003). Enrichment of segmental duplications in regions of breaks of synteny between the human and mouse genomes suggest their involvement in evolutionary

rearrangements. Human Molecular Genetics 12: 2201-2208.
Bailey J. A., Baertsch R., Kent W. J., Haussler D., and Eichler E. E. (2004). Hotspots of mammalian chromosomal evolution. Genome Biology 5: -.
Bailey J. A., Gu Z. P., Clark R. A., Reinert K., Samonte R. V., Schwartz S., Adams M. D., Myers E. W., Li P. W., and Eichler E. E. (2002). Recent segmental duplications in the human genome. Science 297: 1003-1007.
Bailey J. A., Yavor A. M., Massa H. F., Trask B. J., and Eichler E. E. (2001). Segmental duplications: Organization and impact within the current Human Genome Project assembly. Genome Research 11: 1005-1017.
Castresana J. (2002). Genes on human chromosome 19 show extreme divergence from the mouse orthologs and a high GC content. Nucleic Acids Research 30: 1751-1756.
Cheng Z., Ventura M., She X. W., Khaitovich P., Graves T., Osoegawa K., Church D., DeJong P., Wilson R. K., Paabo S., Rocchi M., and Eichler E. E. (2005). A genome-wide comparison of recent chimpanzee and human segmental duplications. Nature 437: 88-93.
Cheung J., Estivill X., Khaja R., MacDonald J. R., Lau K., Tsui L. C., and Scherer S. W. (2003). Genome-wide detection of segmental duplications and potential assembly errors in the human genome sequence. Genome Biol 4: R25.
Coghlan A., Eichler E. E., Oliver S. G., Paterson A. H., and Stein L. (2005). Chromosome evolution in eukaryotes: a multi-kingdom perspective. Trends Genet 21: 673-82.
Conant G. C., and Wagner A. (2003). Asymmetric sequence divergence of duplicate genes. Genome Res 13: 2052-8.
Conrad D. F., Andrews T. D., Carter N. P., Hurles M. E., and Pritchard J. K. (2006). A high-resolution survey of deletion polymorphism in the human genome. Nat Genet 38: 75-81.
Cooper D. N., and Youssoufian H. (1988). The CpG dinucleotide and human genetic disease. Hum Genet 78: 151-5.
Courseaux A., and Nahon J. L. (2001). Birth of two chimeric genes in the Hominidae lineage. Science 291: 1293-7.
Eichler E. E. (2001). Segmental duplications: What's missing, misassigned, and misassembled - and should we care? Genome Research 11: 653-656.
Eichler E. E., and Sankoff D. (2003). Structural dynamics of eukaryotic chromosome evolution. Science 301: 793-797.
Eyrewalker A. (1993). Recombination and Mammalian Genome Evolution. Proceedings of the Royal Society of London Series B-Biological Sciences 252: 237-243.
Feuk L., MacDonald J. R., Tang T., Carson A. R., Li M., Rao G., Khaja R., and Scherer S. W. (2005). Discovery of human inversion polymorphisms by comparative analysis of human and chimpanzee DNA sequence assemblies. Plos Genetics 1: 489-498.
Goidts V., Szamalek J. M., Hameister H., and Kehrer-Sawatzki H. (2004). 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. Human Genetics 115: 116-122.
Hardison R. C., Roskin K. M., Yang S., Diekhans M., Kent W. J., Weber R., Elnitski L., Li J., O'Connor M., Kolbe D., Schwartz S., Furey T. S., Whelan S., Goldman N., Smit A., Miller W., Chiaromonte F., and Haussler D. (2003). Covariation in frequencies of substitution, deletion, transposition, and recombination during eutherian evolution. Genome Res 13: 13-26.
Hellmann I., Ebersberger I., Ptak S. E., Paabo S., and Przeworski M. (2003). A neutral explanation for the

correlation of diversity with recombination rates in humans. American Journal of Human Genetics 72: 1527-1535.
Ji Y., Eichler E. E., Schwartz S., and Nicholls R. D. (2000). Structure of chromosomal duplicons and their role in mediating human genomic disorders. Genome Res 10: 597-610.
Johnson M. E., Cheng Z., Morrison V. A., Scherer S., Ventura M., Gibbs R. A., Green E. D., and Eichler E. E. (2006). Eukaryotic Transposable Elements and Genome Evolution Special Feature: Recurrent duplication-driven transposition of DNA during hominoid evolution. Proc Natl Acad Sci U S A 103: 17626-31.
Johnson M. E., Viggiano L., Bailey J. A., Abdul-Rauf M., Goodwin G., Rocchi M., and Eichler E. E. (2001). Positive selection of a gene family during the emergence of humans and African apes. Nature 413: 514-9.
Jordan I. K., Wolf Y. I., and Koonin E. V. (2004). Duplicated genes evolve slower than singletons despite the initial rate increase. Bmc Evolutionary Biology 4: -.
Kehrer-Sawatzki H., Sandig C., Chuzhanova N., Goidts V., Szamalek J. M., Tanzer S., Muller S., Platzer M., Cooper D. N., and Hameister H. (2005a). Breakpoint analysis of the pericentric inversion distinguishing human chromosome 4 from the homologous chromosome in the chimpanzee (Pan troglodytes). Human Mutation 25: 45-55.
Kehrer-Sawatzki H., Sandig C. A., Goidts V., and Hameister H. (2005b). Breakpoint analysis of the pericentric inversion between chimpanzee chromosome 10 and the homologous chromosome 12 in humans. Cytogenetic and Genome Research 108: 91-97.
Kehrer-Sawatzki H., Schreiner B., Tanzer S., Platzer M., Muller S., and Hameister H. (2002). Molecular Characterization of the Pericentric Inversion That Causes Differences Between Chimpanzee Chromosome 19 and Human Chromosome 17. American Journal of Human Genetics 71: 375-388.
Kehrer-Sawatzki H., Szamalek J. M., Tanzer S., Platzer M., and Hameister H. (2005c). Molecular characterization of the pericentric inversion of chimpanzee chromosome 11 homologous to human chromosome 9. Genomics 85: 542-550.
Kondrashov F. A., Rogozin I. B., Wolf Y. I., and Koonin E. V. (2002). Selection in the evolution of gene duplications. Genome Biol 3: RESEARCH0008.
Kong A., Gudbjartsson D. F., Sainz J., Jonsdottir G. M., Gudjonsson S. A., Richardsson B., Sigurdardottir S., Barnard J., Hallbeck B., Masson G., Shlien A., Palsson S. T., Frigge M. L., Thorgeirsson T. E., Gulcher J. R., and Stefansson K. (2002). A high-resolution recombination map of the human genome. Nature Genetics 31: 241-247.
Li W. H., Yi S. J., and Makova K. (2002). Male-driven evolution. Current Opinion in Genetics & Development 12: 650-656.
Linardopoulou E. V., Williams E. M., Fan Y., Friedman C., Young J. M., and Trask B. J. (2005). Human subtelomeres are hot spots of interchromosomal recombination and segmental duplication. Nature 437: 94-100.
Lindblad-Toh K., Wade C. M., Mikkelsen T. S., Karlsson E. K., Jaffe D. B., Kamal M., Clamp M., Chang J. L., Kulbokas E. J., 3rd, Zody M. C., Mauceli E., Xie X., Breen M., Wayne R. K., Ostrander E. A., Ponting C. P., Galibert F., Smith D. R., DeJong P. J., Kirkness E., Alvarez P., Biagi T., Brockman W., Butler J., Chin C. W., Cook A., Cuff J., Daly M. J., DeCaprio D., Gnerre S., Grabherr M., Kellis M., Kleber M., Bardeleben C., Goodstadt L., Heger A., Hitte C., Kim L., Koepfli K. P., Parker H. G., Pollinger J. P., Searle S. M., Sutter N. B., Thomas R., Webber C.,

Baldwin J., Abebe A., Abouelleil A., Aftuck L., Ait-Zahra M., Aldredge T., Allen N., An P., Anderson S., Antoine C., Arachchi H., Aslam A., Ayotte L., Bachantsang P., Barry A., Bayul T., Benamara M., Berlin A., Bessette D., Blitshteyn B., Bloom T., Blye J., Boguslavskiy L., Bonnet C., Boukhgalter B., Brown A., Cahill P., Calixte N., Camarata J., Cheshatsang Y., Chu J., Citroen M., Collymore A., Cooke P., Dawoe T., Daza R., Decktor K., DeGray S., Dhargay N., Dooley K., Dooley K., Dorje P., Dorjee K., Dorris L., Duffey N., Dupes A., Egbiremolen O., Elong R., Falk J., Farina A., Faro S., Ferguson D., Ferreira P., Fisher S., FitzGerald M., et al. (2005). Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature 438: 803-19.

Lu J., Li W. H., and Wu C. I. (2003). Comment on " Chromosomal speciation and molecular divergence - Accelerated evolution in rearranged chromosomes". Science 302: 988.

Lupski J. R. (1998). Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. Trends Genet 14: 417-22.

Lynch M., and Conery J. S. (2000). The evolutionary fate and consequences of duplicate genes. Science 290: 1151-1155.

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

Marques-Bonet T., and Navarro A. (2005). Chromosomal rearrangements are associated with higher rates of molecular evolution in mammals. Gene 353: 147-154.

Marques-Bonet T., Sànchez-Ruiz J., Armengol L., Khaja R., Bertranpetit J., Rocchi M., Gazave E., and Navarro A. (2007). On the association between chromosomal rearrangements and genic evolution in humans and chimpanzees. Genomics (Submitted).

Meunier J., and Duret L. (2004). Recombination drives the evolution of GC-content in the human genome. Mol Biol Evol 21: 984-90.

Mikkelsen T. S., Hillier L. W., Eichler E. E., Zody M. C., Jaffe D. B., Yang S. P., Enard W., Hellmann I., Lindblad-Toh K., Altheide T. K., Archidiacono N., Bork P., Butler J., Chang J. L., Cheng Z., Chinwalla A. T., deJong P., Delehaunty K. D., Fronick C. C., Fulton L. L., Gilad Y., Glusman G., Gnerre S., Graves T. A., Hayakawa T., Hayden K. E., Huang X. Q., Ji H. K., Kent W. J., King M. C., Kulbokas E. J., Lee M. K., Liu G., Lopez-Otin C., Makova K. D., Man O., Mardis E. R., Mauceli E., Miner T. L., Nash W. E., Nelson J. O., Paabo S., Patterson N. J., Pohl C. S., Pollard K. S., Prufer K., Puente X. S., Reich D., Rocchi M., Rosenbloom K., Ruvolo M., Richter D. J., Schaffner S. F., Smit A. F. A., Smith S. M., Suyama M., Taylor J., Torrents D., Tuzun E., Varki A., Velasco G., Ventura M., Wallis J. W., Wendl M. C., Wilson R. K., Lander E. S., Waterston R. H., and Consortium C. S. A. (2005). Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 437: 69-87.

Murphy W. J., Larkin D. M., Everts-van der Wind A., Bourque G., Tesler G., Auvil L., Beever J. E., Chowdhary B. P., Galibert F., Gatzke L., Hitte C., Meyers S. N., Milan D., Ostrander E. A., Pape G., Parker H. G., Raudsepp T., Rogatcheva M. B., Schook L. B., Skow L. C., Welge M., Womack J. E., O'Brien S J., Pevzner P. A., and Lewin H. A. (2005). Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. Science 309: 613-7.

Myers S., Bottolo L., Freeman C., McVean G., and Donnelly P. (2005). A fine-scale map of recombination rates

and hotspots across the human genome. Science 310: 321-4.
Navarro A., and Barton N. H. (2003a). Accumulating postzygotic isolation genes in parapatry: A new twist on chromosomal speciation. Evolution 57: 447-459.
Navarro A., and Barton N. H. (2003b). Chromosomal speciation and molecular divergence - Accelerated evolution in rearranged chromosomes. Science 300: 321-324.
Navarro A., Marques-Bonet T., and Barton N. H. (2003). Response to comment on " chromosomal speciation and molecular divergence - Accelerated evolution in rearranged chromosomes". Science 302: 988.
Newman T. L., Tuzun E., Morrison V. A., Hayden K. E., Ventura M., McGrath S. D., Rocchi M., and Eichler E. E. (2005). A genome-wide survey of structural variation between human and chimpanzee. Genome Research 15: 1344-1356.
Noor M. A. F., Grams K. L., Bertucci L. A., and Reiland J. (2001). Chromosomal inversions and the reproductive isolation of species. Proceedings of the National Academy of Sciences of the United States of America 98: 12084-12088.
Rieseberg L. H. (2001). Chromosomal rearrangements and speciation. Trends in Ecology & Evolution 16: 351-358.
Rozen S., Skaletsky H., Marszalek J. D., Minx P. J., Cordum H. S., Waterston R. H., Wilson R. K., and Page D. C. (2003). Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. Nature 423: 873-876.
Rudd M. K., and Willard H. F. (2004). Analysis of the centromeric regions of the human genome assembly. Trends in Genetics 20: 529-533.
Samonte R. V., and Eichler E. E. (2002). Segmental duplications and the evolution of the primate genome. Nature Reviews Genetics 3: 65-72.
Sebat J., Lakshmi B., Troge J., Alexander J., Young J., Lundin P., Maner S., Massa H., Walker M., Chi M., Navin N., Lucito R., Healy J., Hicks J., Ye K., Reiner A., Gilliam T. C., Trask B., Patterson N., Zetterberg A., and Wigler M. (2004). Large-scale copy number polymorphism in the human genome. Science 305: 525-8.
Sharp A. J., Cheng Z., and Eichler E. E. (2006). Structural Variation of the Human Genome. Annu Rev Genomics Hum Genet.
Sharp A. J., Locke D. P., McGrath S. D., Cheng Z., Bailey J. A., Vallente R. U., Pertz L. M., Clark R. A., Schwartz S., Segraves R., Oseroff V. V., Albertson D. G., Pinkel D., and Eichler E. E. (2005). Segmental duplications and copy-number variation in the human genome. American Journal of Human Genetics 77: 78-88.
Shaw C. J., and Lupski J. R. (2004). Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease. Human Molecular Genetics 13: R57-R64.
She X. W., Horvath J. E., Jiang Z. S., Liu G., Furey T. S., Christ L., Clark R., Graves T., Gulden C. L., Alkan C., Bailey J. A., Sahinalp C., Rocchi M., Haussler D., Wilson R. K., Miller W., Schwartz S., and Eichler E. E. (2004). The structure and evolution of centromeric transition regions within the human genome. Nature 430: 857-864.
She X. W., Liu G., Ventura M., Zhao S., Misceo D., Roberto R., Cardone M. F., Rocchi M., Green E. D., Archidiacano N., and Eichler E. E. (2006). A preliminary comparative analysis of primate segmental duplications shows elevated substitution rates and a great-ape expansion of intrachromosomal duplications. Genome Research 16: 576-583.
Stankiewicz P., and Lupski J. R. (2002). Genome architecture, rearrangements and genomic disorders. Trends in Genetics 18: 74-82.
Stankiewicz P., Shaw C. J., Withers M., Inoue K., and Lupski J. R. (2004). Serial segmental duplications during

primate evolution result in complex human genome architecture. Genome Research 14: 2209-2220.
Sved J., and Bird A. (1990). The expected equilibrium of the CpG dinucleotide in vertebrate genomes under a mutation model. Proc Natl Acad Sci U S A 87: 4692-6.
Szamalek J. M., Goidts V., Chuzhanova N., Hameister H., Cooper D. N., and Kehrer-Sawatzki H. (2005). Molecular characterisation of the pericentric inversion that distinguishes human chromosome 5 from the homologous chimpanzee chromosome. Human Genetics 117: 168-176.
Szamalek J. M., Goidts V., Searle J. B., Cooper D. N., Hameister H., and Kehrer-Sawatzki H. (2006). The chimpanzee-specific pericentric inversions that distinguish humans and chimpanzees have identical breakpoints in Pan troglodytes and Pan paniscus. Genomics 87: 39-45.
Vallender E. J., and Lahn B. T. (2004). Effects of chromosomal rearrangements on human-chimpanzee molecular evolution. Genomics 84: 757-761.
Wolfe K. H., and Sharp P. M. (1993). Mammalian Gene Evolution - Nucleotide-Sequence Divergence between Mouse and Rat. Journal of Molecular Evolution 37: 441-456.
Yunis J. J., and Prakash O. (1982). The Origin of Man - a Chromosomal Pictorial Legacy. Science 215: 1525-1530.
Zhang J. Z., Wang X. X., and Podlaha O. (2004). Testing the chromosomal speciation hypothesis for humans and chimpanzees. Genome Research 14: 845-851.
Zhang L. Q., Lu H. H. S., Chung W. Y., Yang J., and Li W. H. (2005). Patterns of segmental duplication in the human genome. Molecular Biology and Evolution 22: 135-141.
Zhang P., Gu Z. L., and Li W. H. (2003). Different evolutionary patterns between young duplicate genes in the human genome. Genome Biology 4: R56.
Zody M., Garber M., Sharpe T., Young S., Rowen L., O'Neill K., Whittaker C., Kamal M., Chang J., Cuomo C., Dewar K., Fitzgerald M., Kodira C., Madan A., Qin S., Yang X., Abbasi N., Abouelleil A., Arachchi H., Baradarani L., Birditt B., Bloom S., Bloom T., Borowsky M., Burke J., Butler J., Cook A., Dearellano K., Decaprio D., Dorris L., Dors M., Eichler E., Engels R., Fahey J., Fleetwood P., Friedman C., Gearin G., Hall J., Hensley G., Johnson E., Jones C., Kamat A., Kaur A., Locke D., Madan A., Munson G., Jaffe D., Lui A., Macdonald P., Mauceli E., Naylor J., Nesbitt R., Nicol R., O'Leary S. Â. a., Ratcliffe A., Rounsley S., She X., Sneddon K., Stewart S., Sougnez C., Stone S., Topham K., Vincent D., Wang S., Zimmer A., Birren B., Hood L., Lander E., and Nusbaum C. (2006a). Analysis of the DNA sequence and duplication history of human chromosome 15. Nature 440: 671-675.
Zody M. C., Garber M., Adams D. J., Sharpe T., Harrow J., Lupski J. R., Nicholson C., Searle S. M., Wilming L., Young S. K., Abouelleil A., Allen N. R., Bi W. M., Bloom T., Borowsky M. L., Bugalter B. E., Butler J., Chang J. L., Chen C. K., Cook A., Corum B., Cuomo C. A., de Jong P. J., DeCaprio D., Dewar K., FitzGerald M., Gilbert J., Gibson R., Gnerre S., Goldstein S., Grafham D. V., Grocock R., Hafez N., Hagopian D. S., Hart E., Norman C. H., Humphray S., Jaffe D. B., Jones M., Kamal M., Khodiyar V. K., LaButti K., Laird G., Lehoczky J., Liu X. H., Lokyitsang T., Loveland J., Lui A., Macdonald P., Major J. E., Matthews L., Mauceli E., McCarroll S. A., Mihalev A. H., Mudge J., Nguyen C., Nicol R., O'Leary S. B., Osoegawa K., Schwartz D. C., Shaw-Smith C., Stankiewicz P., Steward C., Swarbreck D., Venkataraman V., Whittaker C. A., Yang X. P., Zimmer A. R., Bradley A., Hubbard T., Birren B. W., Rogers J., Lander E. S., and

Nusbaum C. (2006b). DNA sequence of human chromosome 17 and analysis of rearrangement in the human lineage. Nature 440: 1045-1049.

Table 1. Distribution of SDs identities relative to major genomic rearrangements between humans and chimpanzees. The percentages were calculated as the proportion of pairwise alignments at each percent identity.

	Inside rearranged regions	Outside rearranged regions	Colinear Chromosome
Similarity(ID)	Percentage of age within each cathegory (%)		
90-91 % ID	12.20	17.64	14.16
91-92 % ID	7.76	12.79	16.50
92-93 % ID	8.50	11.25	12.86
93-94 % ID	5.55	8.96	15.24
94-95 % ID	5.91	9.19	10.61
95-96 % ID	6.47	7.51	8.36
96-97 % ID	7.39	7.98	5.67
97-98 % ID	6.47	12.42	6.43
98-99 % ID	17.56	7.31	5.76
99-100 % ID	21.63	4.81	4.41

Table 2. Distribution of SDs identities relative to simulated rearrangements in collinear chromosomes between humans and chimpanzees.The percentages were calculated as the proportion of pairwise alignments at each percent identity.

	Inside rearranged regions	Outside rearranged regions
Similarity(ID)	Percentage of age within each category (%)	
90-91 % ID	10.41	16.35
91-92 % ID	16.99	15.62
92-93 % ID	8.49	14.16
93-94 % ID	16.44	13.43
94-95 % ID	8.22	12.77
95-96 % ID	7.12	7.97
96-97 % ID	4.93	6.27
97-98 % ID	6.85	4.88
98-99 % ID	9.04	4.15
99-100 % ID	11.51	4.39

Table 3. Average of divergences and lengths among paralogous copies of SDs as to genomic factors and rearrangements between humans and chimpanzees. Divergence (K) is calculated as the number of substitution per site between the two duplication alignments. Length (Size) correspond to the aligned basepairs. P-values are calculated by means of permutation test (see Material and Methods).

Sex Chromosomes				ID
	SDs in Autosomes	**SDs in HSAX**	**P-value**	**> 98 %**
N	889	103		
K	0.0107	0.0076	< 0.001	
Size	41,439.50	52,887.93	0.115	
	SDs in Autosomes	**SDs in HSAX**	**P-value**	**< 92 %**
N	3273	261		
K	0.0958	0.0962	0.364	
Size	4,689.73	4,458.48	0.578	

	SDs in Autosomes	**SDs in HSAY**	**P value**	**> 98 %**
N	889	32		
K	0.0107	0.0052	< 0.001	
Size	41,439.50	169,158.75	< 0.001	
	SDs in Autosomes	**SDs in HSAY**	**P value**	**< 92 %**
N	3273	132		
K	0.0958	0.0976	0.001	
Size	4,689.73	12,290.17	< 0.001	

Telomeres (10 Mb)				ID
	SDs not in telomeres	**SDs in Telomeres**	**P-value**	**> 98 %**
N	719	170		
K	0.0105	0.0115	0.052	
Size	44,040.90	30,437.11	0.010	
	SDs not in telomeres	**SDs in Telomeres**	**P-value**	**< 92 %**
N	2831	442		
K	0.0958	0.0958	0.874	
Size	4,746.56	4,325.72	0.224	

Centromere (5Mb)				ID
	SDs not in Centromeres	**SDs in Centromeres**	**P-value**	**> 98 %**
N	572	147		
K	0.0106	0.0100	0.316	
Size	36,111.33	74,896.07	< 0.001	
	SDs not in Centromeres	**SDs in Centromeres**	**P-value**	**< 92 %**
N	2096	735		
K	0.0959	0.0953	0.029	
Size	3,908.13	7,137.51	< 0.001	

HSA19				ID
	SDs in other autosomes	**SDs in HSA19**	**P-value**	**> 98 %**
N	561	11		
K	0.0105	0.0126	0.320	
Size	36,600.01	11,188.90	0.139	
	SDs in other autosomes	**SDs in HSA19**	**P-value**	**< 92 %**
N	2029	67		
K	0.0959	0.0958	0.906	
Size	3,875.29	4,902.67	0.141	

Rearranged Chromosomes				ID
	SDs in Colinear chr	**SDs in Rearranged Chr**	**P-value**	**> 98 %**
N	208	353		
K	0.0114	0.0100	0.009	
Size	25,385.48	43,208.01	< 0.001	
	SDs in Colinear chr	**SDs in Rearranged Chr**	**P-value**	**< 92 %**
N	890	1139		
K	0.0959	0.0960	0.902	
Size	3,791.72	3,940.59	0.534	

Inside rearranged regions versus Outside rearranged regions, without HSA2				ID
	SDs Outside rearranged regions	**SDs Inside rearranged regions**	**P-value**	**> 98 %**
N	216	87		
K	0.0104	0.0096	0.347	
Size	40,400.12	55,156.56	0.058	
	SDs Outside rearranged regions	**SDs Inside rearranged regions**	**P-value**	**< 92 %**
N	715	267		
K	0.096	0.0957	0.586	
Size	3,879.46	4,868.64	0.016	

Inversions detected in (Newman et al 2005) vs rest of chromosomes				ID
	SDs Outside Inversion	**SDs Inside Inversion**	**P-value**	**> 98 %**
N	541	20		
K	0.0104	0.0131	0.063	
Size	35,170.62	75,264.90	0.003	
	SDs Outside Inversion	**SDs Inside Inversion**	**P-value**	**< 92 %**
N	1977	52		
K	0.0959	0.0986	0.004	
Size	3,853.94	4,686.98	0.281	

Breakpoints versus inverted chromosomes (excluding HSA2)				ID
	SDs rest of Chr	**SDs in BKP**	**P-value**	**> 98 %**
N	286	17		
K	0.0103	0.0080	0.135	
Size	44,189.30	52,171.11	0.611	
	SDs rest of Chr	**SDs in BKP**	**P-value**	**< 92 %**
N	953	29		
K	0.0960	0.0941	0.150	
Size	4,159.23	3,792.82	0.748	

Table 4. Average of inter-specific divergences in human SDs and chimpanzee SDs as to genomic factors and rearrangements between humans and chimpanzees. Divergence is calculated as the number of substitution per site between the two duplication alignments.

HUMAN SD	X-Chromosome vs. Autosomes		Y-Chromosome vs. Autosomes		Telomeres vs. rest of genome		Centromeres vs. rest of genome		Chromosome 19	
	SDs in autosomes	SDs in the HSAX.	SDs in autosomes	SDs in the HSAY.	SDs outside Telomeres	SDs within Telomeres	SDs outside Centromeres	SDs within Centromeres	SDs outside HSA19	SDs within HSA19
N	*1303*	*109*	*1303*	*51*	*1052*	*251*	*742*	*310*	*714*	*28*
Divergence	0.0238	0.0161	0.0238	0.0259	0.0233	0.0260	0.0228	0.0247	0.0225	0.0285
P-value		< 0.001		0.087		< 0.001		< 0.001		0.001

CHIMPANZEE SD	X-Chromosome vs. Autosomes		Y-Chromosome vs. Autosomes		Telomeres vs. rest of genome		Centromeres vs. rest of genome		Chromosome 19	
	SDs in autosomes	SDs in the HSAX.	SDs in autosomes	SDs in the HSAY.	SDs outside Telomeres	SDs within Telomeres	SDs outside Centromeres	SDs within Centromeres	SDs outside HSA19	SDs within HSA19
N	*1415*	*110*	*1415*	*87*	*1224*	*191*	*789*	*435*	*779*	*10*
Divergence	0.0222	0.0156	0.0222	0.0223	0.0217	0.0252	0.0210	0.0231	0.0207	0.0380
P-value		< 0.001		0.891		< 0.001		< 0.001		< 0.001

HUMAN SD	Rearranged vs. Colinear chromosomes			SDs within vs. outside rearranged regions (excluding HSA2)		
	SDs in colinear chr.	SDs in Rearranged chr.	P-value	SDs Outside inversions	SDs inside inversions	P-value
N	*267*	*447*		*280*	*112*	
Divergence	0.0236	0.0219	0.010	0.0218	0.0219	0.934

CHIMPANZEE SD	Rearranged vs. Colinear chromosomes			SDs within vs. outside rearranged regions (excluding HSA2)		
	SDs in colinear chr.	SDs in Rearranged chr.	P-value	SDs Outside inversions	SDs inside inversions	P-value
N	*256*	*523*		*312*	*160*	
Divergence	0.0216	0.0203	0.025	0.0202	0.0199	0.693

HUMAN SD	Breakpoints vs. inverted chromosomes (excludingHSA2)		
	SDs in Rearranged chr.	SDs in BKPs	P-value
N	*370*	*22*	
Divergence	0.0217	0.0242	0.136

CHIMPANZEE SD	Breakpoints vs. inverted chromosomes (excludingHSA2)		
	SDs in Rearranged chr.	SDs in BKPs	P-value
N	*441*	*31*	
Divergence	0.0201	0.0194	0.552

HUMAN SD	SDs within vs. outside rearranged regions (excluding breakpoints and HSA2)		
	SDs Outside inversions	**SDs inside inversions**	***P-value***
N	*264*	*106*	
Divergence	0.0216	0.0220	*0.619*

CHIMPANZEE SD	SDs within vs. outside rearranged regions (excluding breakpoints and HSA2)		
	SDs Outside inversions	**SDs inside inversions**	**P-value**
N	*291*	*150*	
Divergence	0.0202	0.0199	0.583

HUMAN SD	inversions (Newman et al. 2005) versus rest chromosomes		
	SDs Outside inversions	**SDs inside inversions**	**P-value**
N	*670*	*44*	
Divergence	0.0227	0.0196	0.015

CHIMPANZEE SD	inversions (Newman et al. 2005) chromosomes		
	SDs Outside inversions	**SDs inside inversions**	**P-value**
N	*713*	*66*	
Divergence	0.0210	0.0183	0.004

Table 5. Average of inter-divergences in human SDs and chimpanzee SDs in individual chromosomes as to major rearrangements between human and chimpanzee.

Hs Chr	**Human SDs Outside rearranged regions**	**Inside rearranged regions**	P-value	N_{out}	N_{in}	**Chimpanzee SDs Outside rearranged regions**	**Inside rearranged regions**	P-value	N_{out}	N_{in}	Lineage of the rearrangement
HSA1	0.0209	0.0070	0.043	*106*	*1*	0.0203			*105*	*0*	HUMAN
HSA4	0.0246	0.0263	0.608	*17*	*13*	0.0247			*13*	*0*	CHIMPANZEE
HSA5	0.0226	0.0170	0.116	*10*	*16*	0.0197	0.0161	0.065	*10*	*57*	CHIMPANZEE
HSA9	0.0230	0.0246	0.440	*35*	*26*	0.0184	0.0232	< 0.001	*49*	*38*	CHIMPANZEE
HSA12	0.0201	0.0243	0.286	*9*	*7*	0.0216	0.0190	0.875	*1*	*8*	CHIMPANZEE
HSA15	0.0251	0.0239	0.661	*41*	*7*	0.0224	0.0239	0.418	*48*	*10*	CHIMPANZEE
HSA16	0.0188	0.0319	0.008	*41*	*2*	0.0181	0.0413	0.008	*68*	*1*	CHIMPANZEE
HSA17	0.0215	0.0198	0.427	*16*	*40*	0.0225	0.0206	0.352	*17*	*46*	CHIMPANZEE
HSA18	0.0245			*5*	*0*	0.0252			*1*	*0*	HUMAN
TOTAL (without HSA2)	**0.0218**	**0.0219**	**0.934**	***280***	***112***	**0.0202**	**0.0199**	**0.693**	***312***	***160***	

Figure 1. Distribution of SDs identities relative to major rearrangements between humans and chimpanzees.

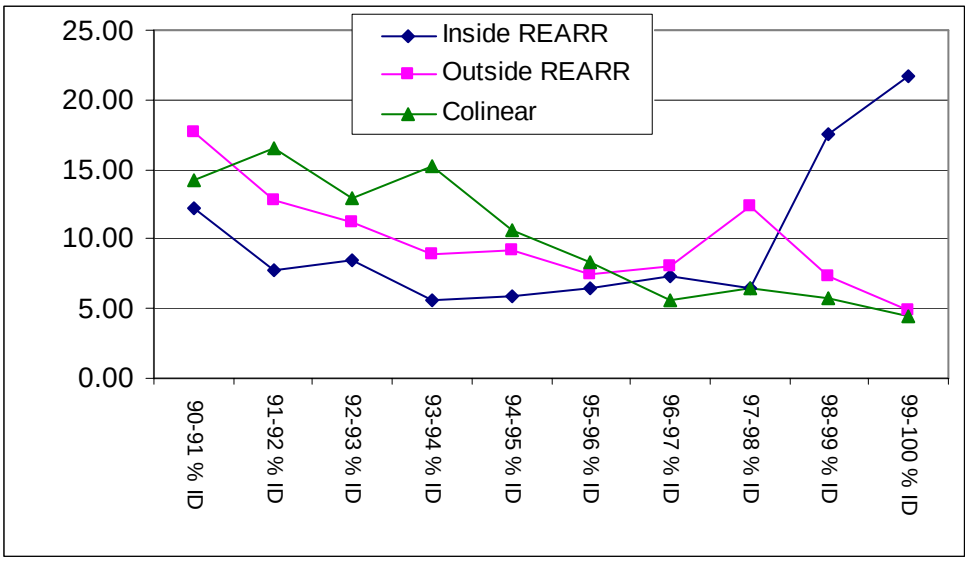
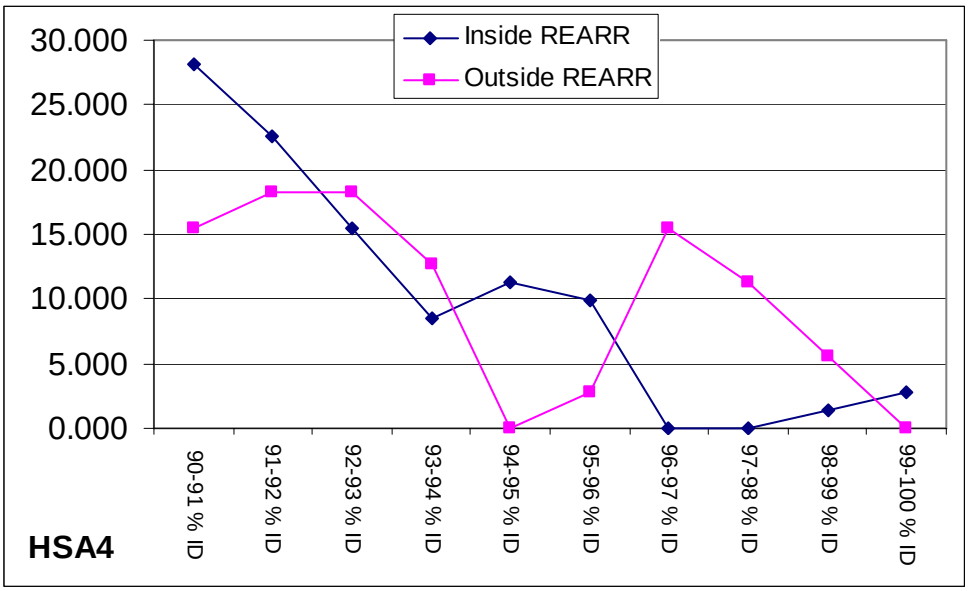
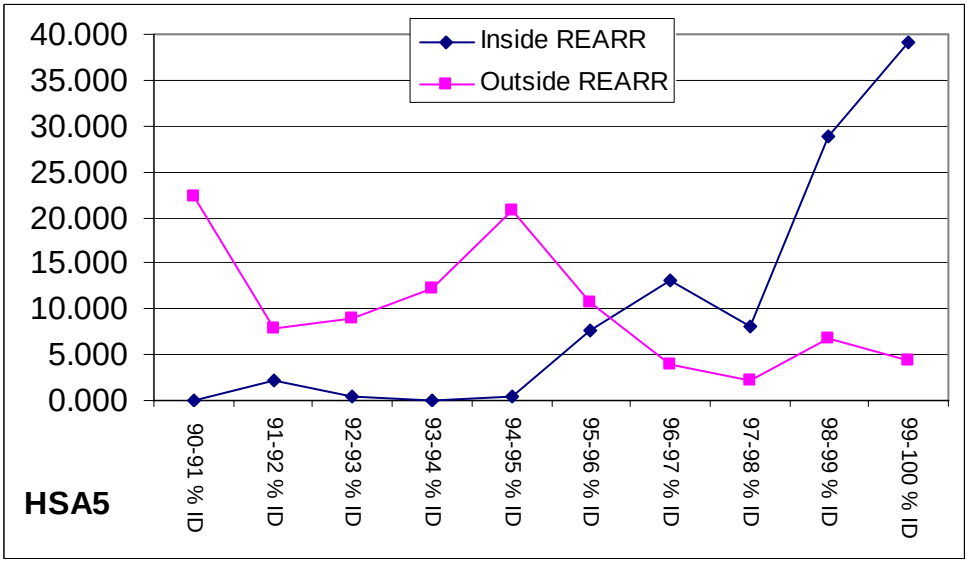
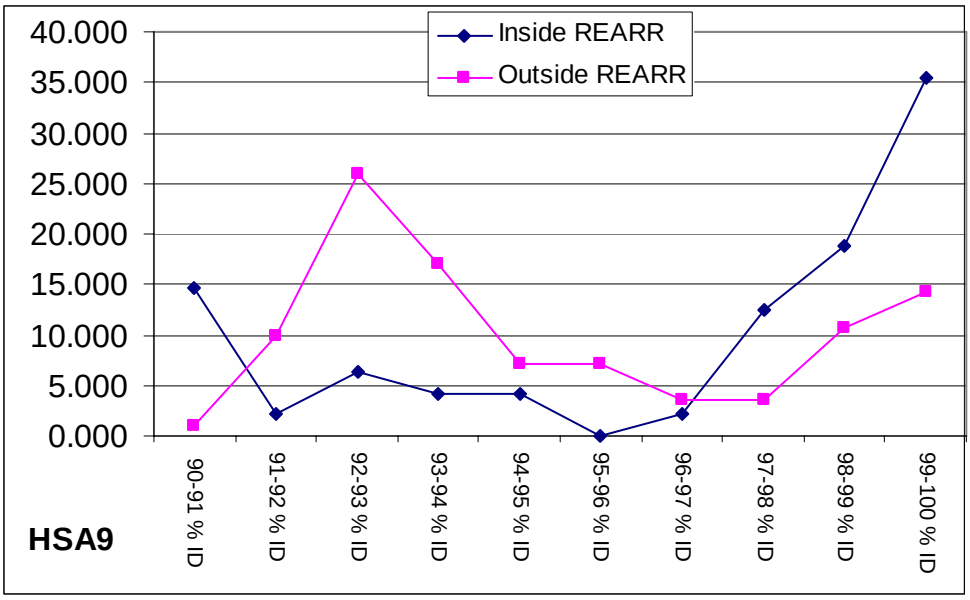
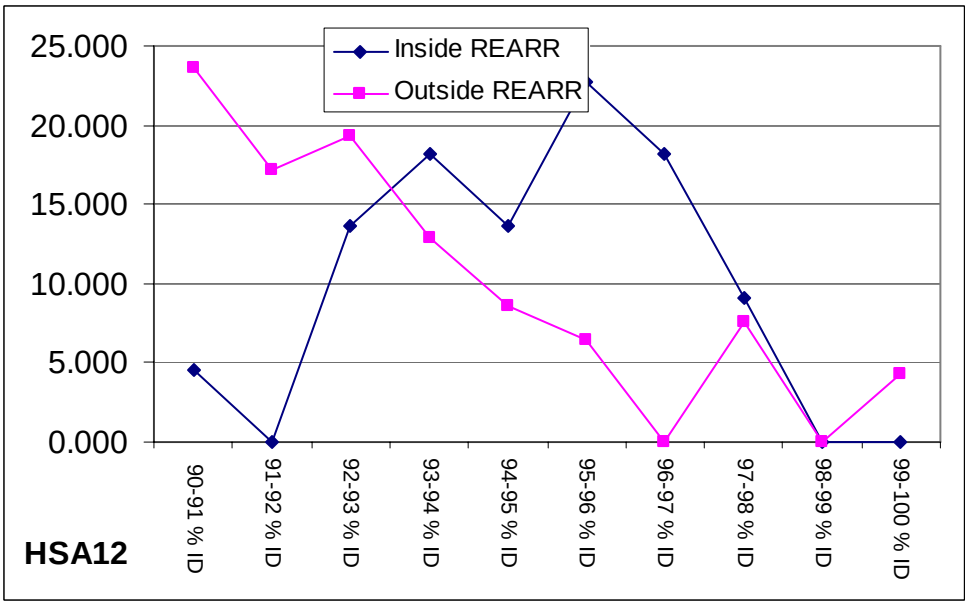


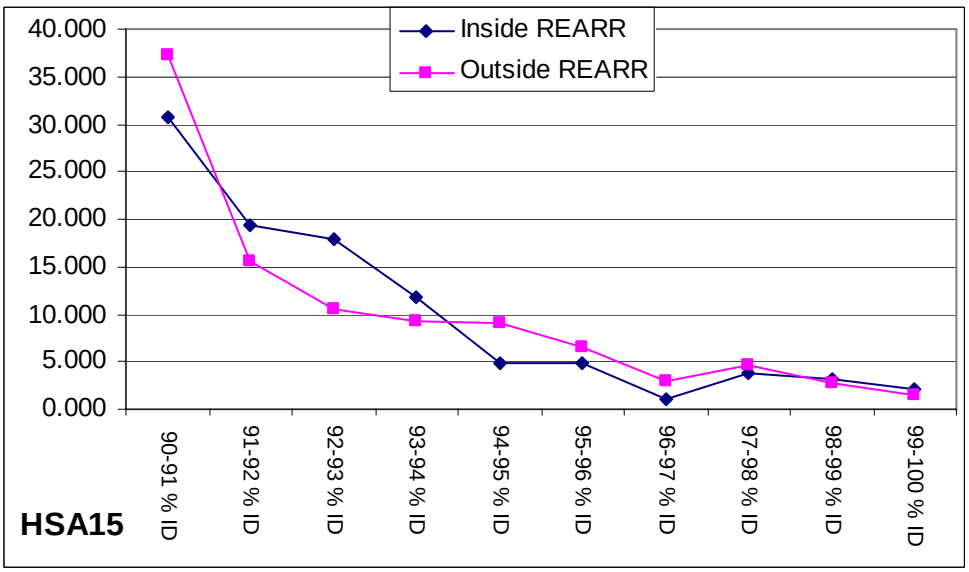
Figure 2. Distribution of SDs identities relative to major rearrangements between humans and chimpanzees in every individual chromosome. Chromosomes that did not have any SDs (both copies) within rearrangements are not shown (see Material & Methods).

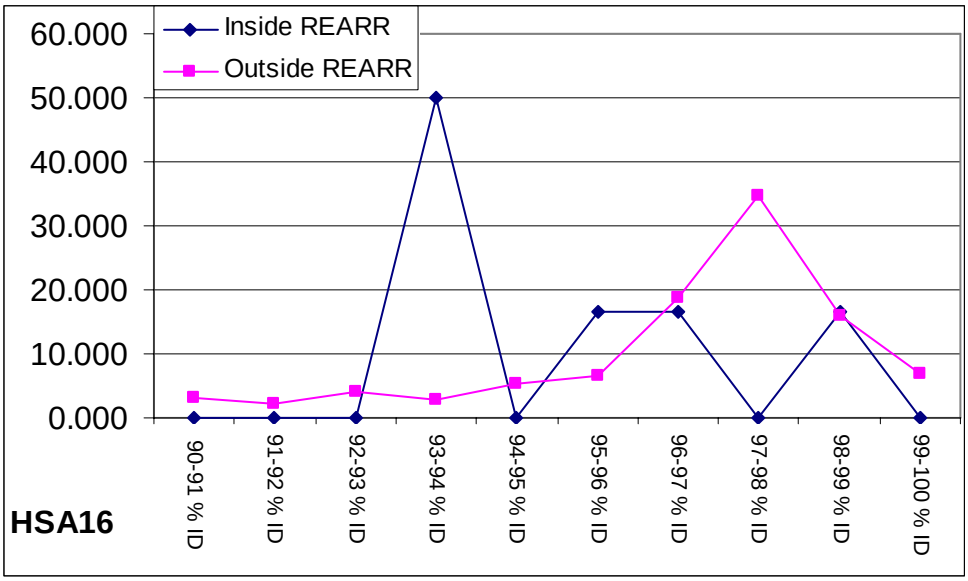





40.000
35.000
30.000
25.000
20.000
15.000
10.000
5.000
0.000
Inside REARR
Outside REARR
90-91 % ID
91-92 % ID
92-93 % ID
93-94 % ID
94-95 % ID
95-96 % ID
96-97 % ID
97-98 % ID
98-99 % ID
99-100 % ID
HSA9


25.000
20.000
15.000
10.000
5.000
0.000
Inside REARR
Outside REARR
90-91 % ID
91-92 % ID
92-93 % ID
93-94 % ID
94-95 % ID
95-96 % ID
96-97 % ID
97-98 % ID
98-99 % ID
99-100 % ID
HSA12


40.000
35.000
30.000
25.000
20.000
15.000
10.000
5.000
0.000
Inside REARR
Outside REARR
90-91 % ID
91-92 % ID
92-93 % ID
93-94 % ID
94-95 % ID
95-96 % ID
96-97 % ID
97-98 % ID
98-99 % ID
99-100 % ID
HSA15


Inside REARR
Outside REARR
60.000
50.000
40.000
30.000
20.000
10.000
0.000
90-91 % ID
91-92 % ID
92-93 % ID
93-94 % ID
94-95 % ID
95-96 % ID
96-97 % ID
97-98 % ID
98-99 % ID
99-100 % ID
HSA16

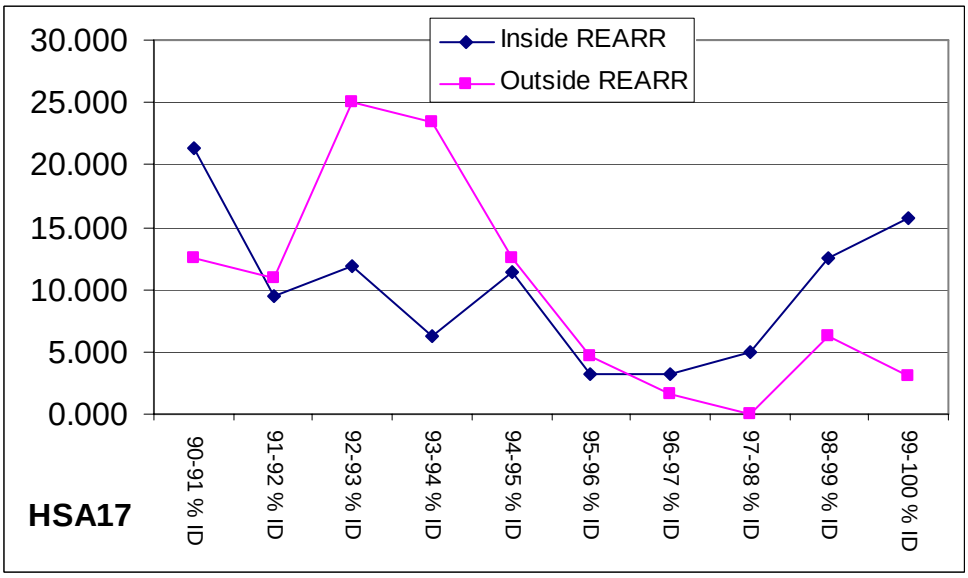

Inside REARR
Outside REARR
30.000
25.000
20.000
15.000
10.000
5.000
0.000
90-91 % ID
91-92 % ID
92-93 % ID
93-94 % ID
94-95 % ID
95-96 % ID
96-97 % ID
97-98 % ID
98-99 % ID
99-100 % ID
HSA17

Figure 3. Distribution of SDs identities relative to simulated pericentromeric rearrangements in collinear chromosomes between humans and chimpanzees.

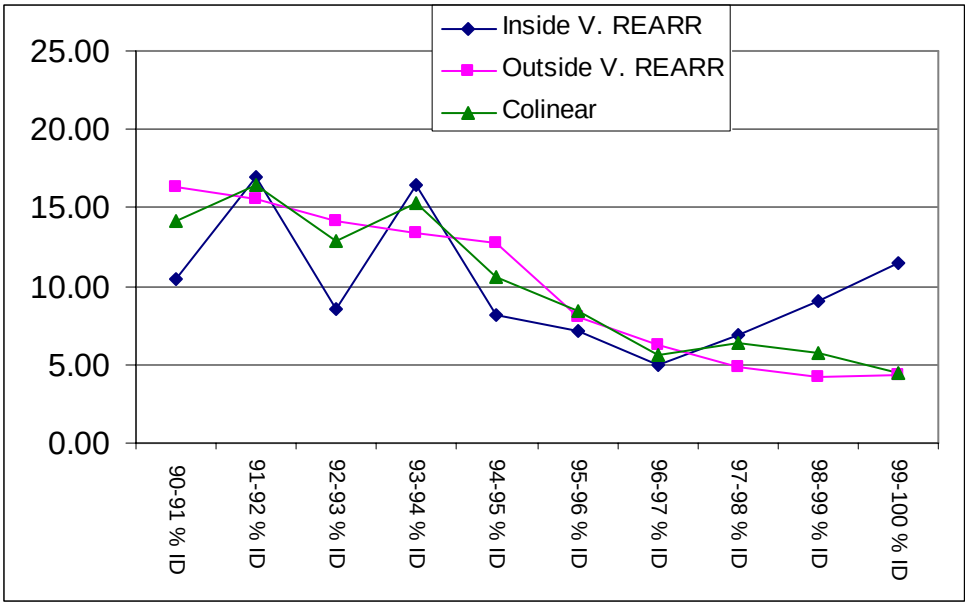
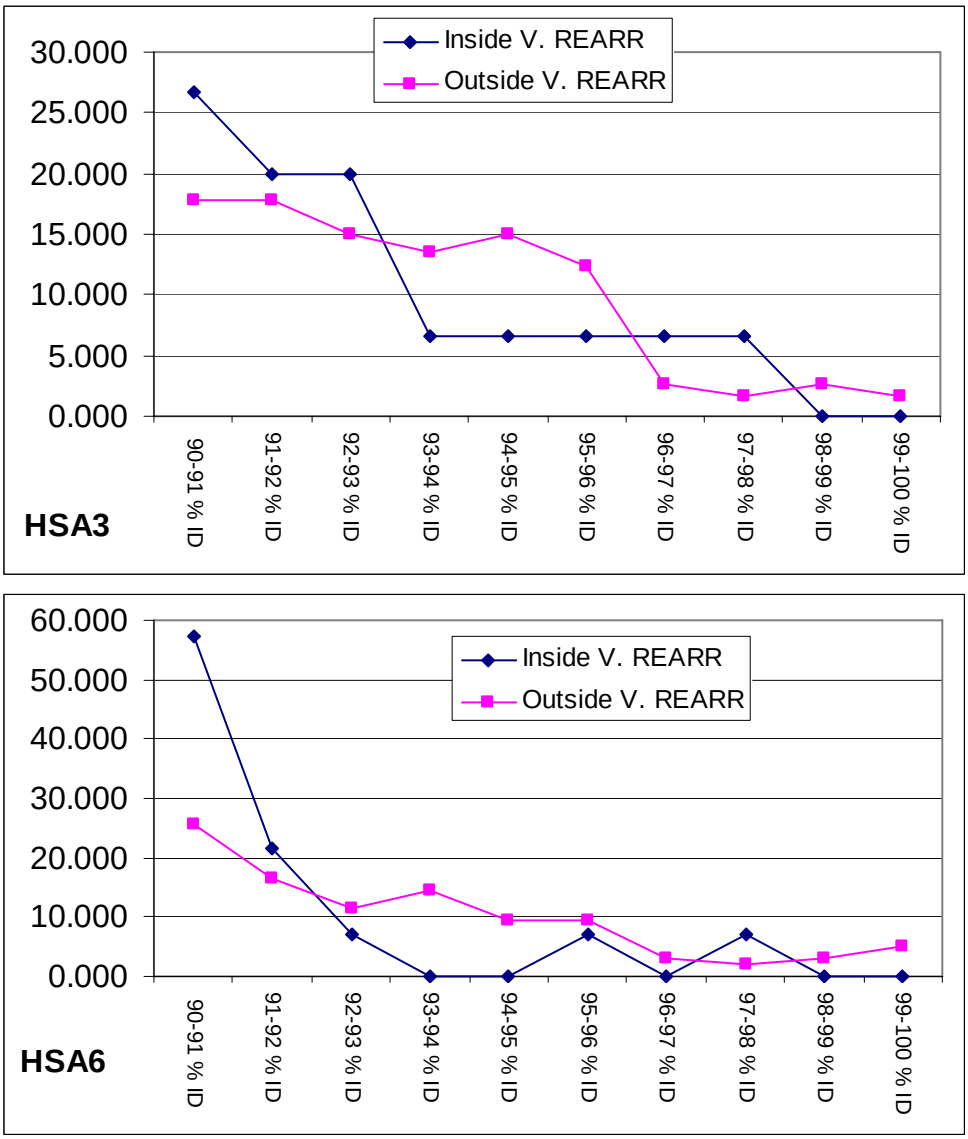
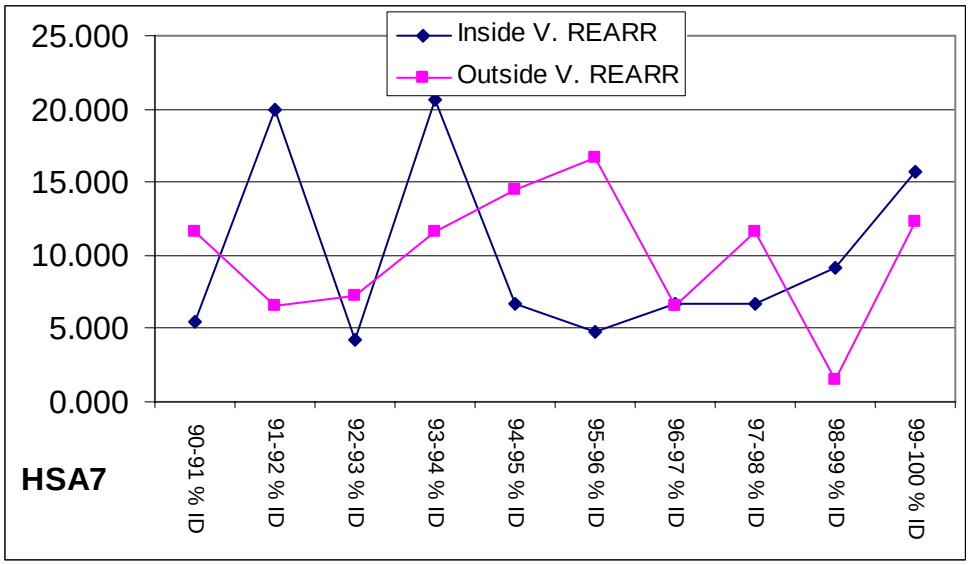
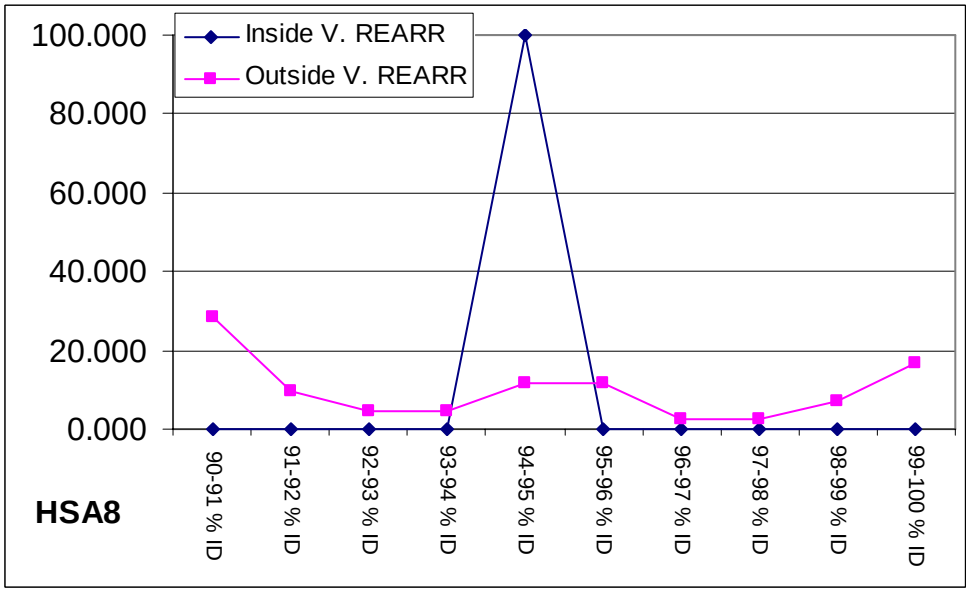
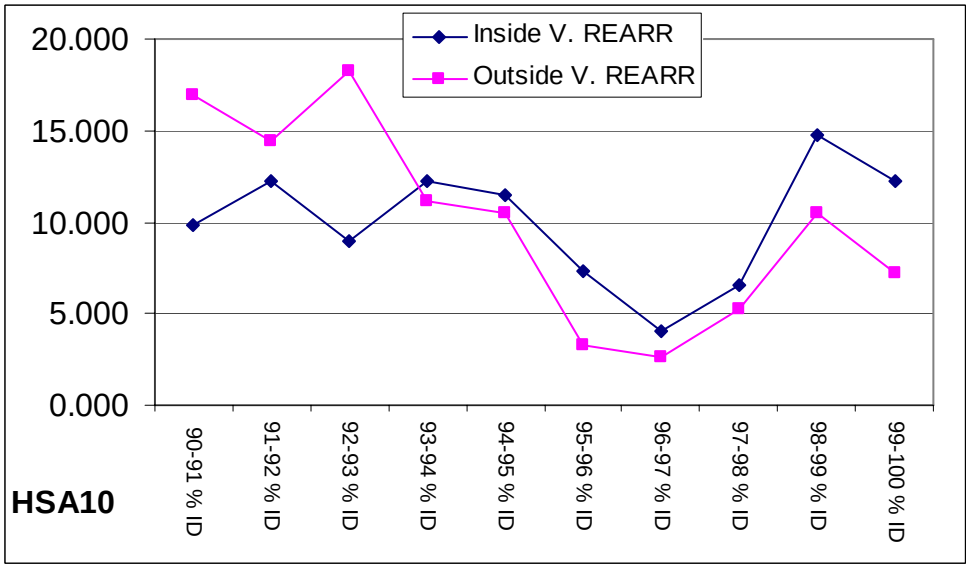


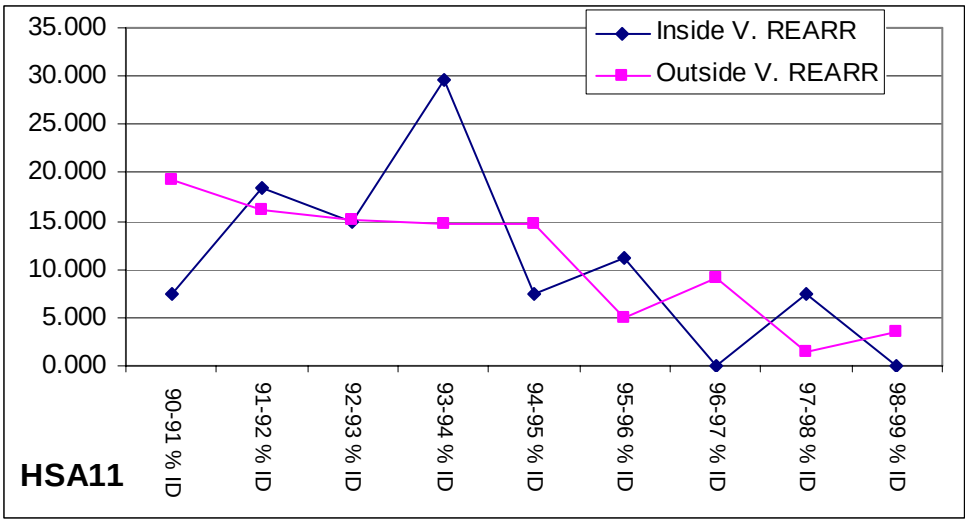
Figure 4. Distribution of SDs identities relative to simulated pericentric rearrangements in colinear chromosomes between humans and chimpanzees in every individual chromosome. Chromosomes that did not have any SDs (both copies) within rearrangements are not shown (see Material & Methods).

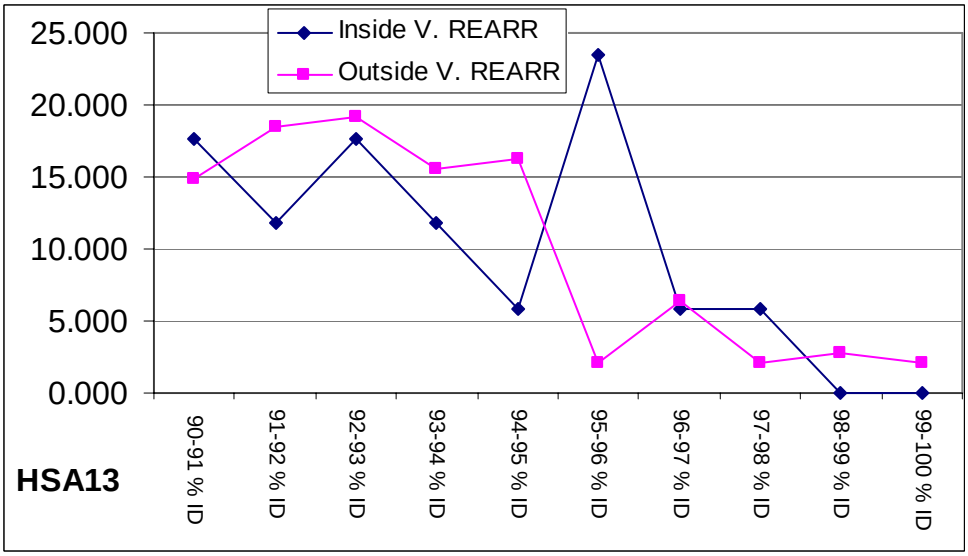


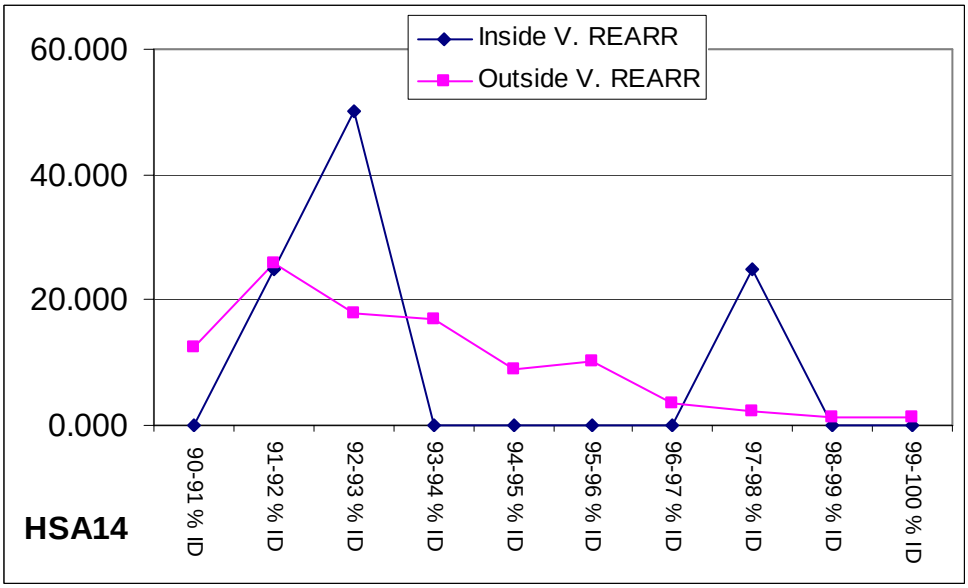

Inside V. REARR
Outside V. REARR
25.000
20.000
15.000
10.000
5.000
0.000
90-91 % ID
91-92 % ID
92-93 % ID
93-94 % ID
94-95 % ID
95-96 % ID
96-97 % ID
97-98 % ID
98-99 % ID
99-100 % ID
HSA7

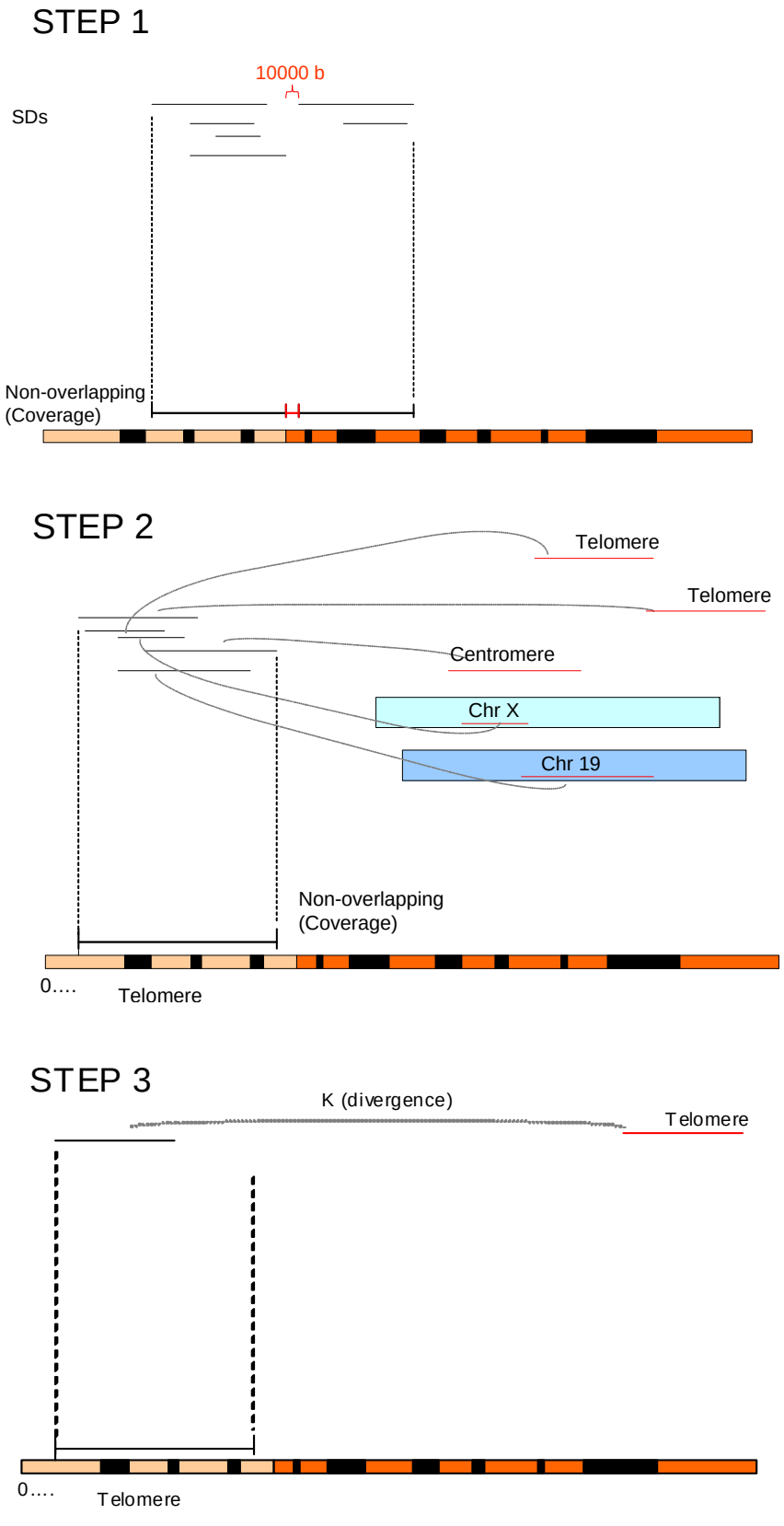

Inside V. REARR
Outside V. REARR
100.000
80.000
60.000
40.000
20.000
0.000
90-91 % ID
91-92 % ID
92-93 % ID
93-94 % ID
94-95 % ID
95-96 % ID
96-97 % ID
97-98 % ID
98-99 % ID
99-100 % ID
HSA8


Inside V. REARR
Outside V. REARR
20.000
15.000
10.000
5.000
0.000
90-91 % ID
91-92 % ID
92-93 % ID
93-94 % ID
94-95 % ID
95-96 % ID
96-97 % ID
97-98 % ID
98-99 % ID
99-100 % ID
HSA10

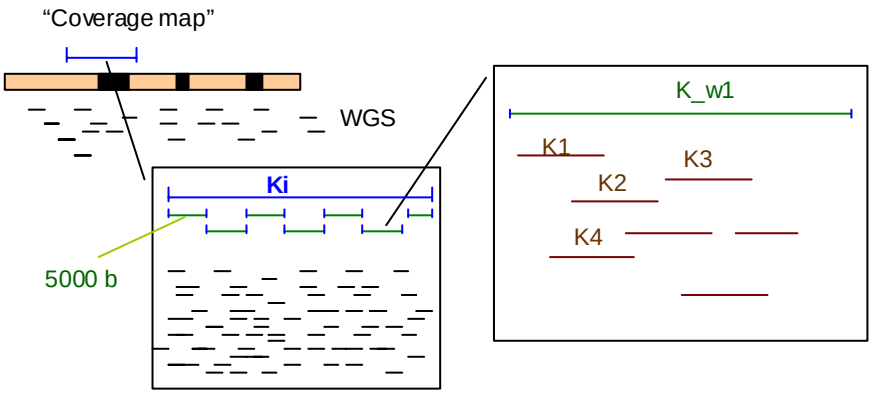

35.000
30.000
25.000
20.000
15.000
10.000
5.000
0.000
Inside V. REARR
Outside V. REARR
90-91 % ID
91-92 % ID
92-93 % ID
93-94 % ID
94-95 % ID
95-96 % ID
96-97 % ID
97-98 % ID
98-99 % ID
HSA11


25.000
20.000
15.000
10.000
5.000
0.000
Inside V. REARR
Outside V. REARR
90-91 % ID
91-92 % ID
92-93 % ID
93-94 % ID
94-95 % ID
95-96 % ID
96-97 % ID
97-98 % ID
98-99 % ID
99-100 % ID
HSA13


60.000
40.000
20.000
0.000
Inside V. REARR
Outside V. REARR
90-91 % ID
91-92 % ID
92-93 % ID
93-94 % ID
94-95 % ID
95-96 % ID
96-97 % ID
97-98 % ID
98-99 % ID
99-100 % ID
HSA14



Supplementary Figure 1. Dataset 2 construction (non-overlapping intraspecific dataset). There are the 3 main steps to construct Dataset 2. STEP1, we constructed the "coverage map", basically we recorded the bound coordinates of overlapping SDs. STEP 2, we labeled every SD as belonging to telomeres, centromeres, HSA19, sexual chromosomes, inverted and non-rearranged zones and breakpoints. STEP 3, we kept as a sample of the region in the "coverage map" those SDs which had the longer paralogous copy in the same labeled region.



Supplementary Figure 2. Dataset 3 construction (non-overlapping interdivergence dataset). We split every zone in the coverage map in non overlapping windows of 5000 b. For every one of those inner windows, divergence (K_w1) was calculated as the average for all chimpanzee WGS (v1) against human sequence (B35). Finally the averages of all windows were joined in a single average divergence of the coverage zone (Ki).

Overview and Discussion

1. Testing chromosomal speciation in human and chimpanzee genomes.

1.1 Initial steps

The speciation of humans from the great apes is one of the most debated issues in human evolution. In particular, human speciation from our closest living relatives, the chimpanzees, besides being of clear anthropological interest constitutes an interesting case study for the theory of speciation, since both species present several characteristics which hint at them having been separated by processes of chromosomal speciation. First, humans and chimpanzees differ not only in their gene sequence but also in their chromosomal structure. In addition, fossil records show that early hominids and chimpanzees had lived in the same part of Africa and thus, at least geographically, parapatric speciation is possible. Finally, it is known that rearrangements occurred early in the divergence of the ancestral populations (Szamalek et al. 2006b), and thus, they are likely to have been present at the time of the divergence of both species, and at least theoretically, they could have participated in chromosomal speciation.

In the first study to test predictions of suppressed-recombination chromosomal speciation models, Navarro and Barton (2003b) reported an association between chromosomal rearrangements and higher evolutionary rates based on 115 autosomal orthologous genes from humans and chimpanzees. The two main results of this analysis were first, that of the 26 genes with Ka/Ks ratios > 1, 20 were located in rearranged chromosomes and only 6 were found in colinear chromosomes and second, that the Ka/Ks averages of genes in rearranged chromosomes (0.84) was more than two-fold the average for colinear chromosomes (0.37). Several interpretations for those results were given, the most controversial being the suggestion that this 2-fold difference in Ka/Ks could be explained, under the hypothesis tested, only if the chromosomal rearrangements have been barriers in parapatry for no less than half of the time of divergence between humans and chimpanzees (Rieseberg and Livingstone 2003). This conclusion was striking because this was against what anthropological data and the molecular dating of rearrangements suggest (Szamalek et al. 2006b).

Several other analyses have questioned results and/or inferences from the initial paper. For example, Lu et al. (2003) reanalyzed the initial dataset from Navarro and Barton (2003b) and besides confirming the previous results, they also made use of outgroups for a subset of those genes (85 genes). They reported that the huge differences in the ratios of divergences (Ka/Ks) between genes within rearranged chromosomes and within colinear chromosomes was also found when comparing human genes with orthologous genes from the outgroups. Their interpretation was that rapidly evolving genes may not have a homogeneous distribution among chromosomes and that rearranged chromosomes may have been linked to rapidly evolving genes due to factors unrelated to speciation. As an example, they commented that faster evolution is known to happen for certain genes which are located in rearranged chromosomes (e.g glycophorins and protamins).

In our answer to the work by Lu et al. (2003), we confirmed the existence of their two findings, clusters of positively selected genes which, in turn, show a strong association with recently rearranged chromosomes (Navarro et al. 2003). In addition, we presented several further analyses. Our first additional observation was a possible bias associated with the way of calculating Ka/Ks ratios. When computing Ka/Ks averages, the initial study by Navarro and Barton (2003b) ignored genes with Ks = 0. In contrast, Lu et al. (2003) computed the Ka/Ks values of these genes by setting Ks = 0.002. This introduced a bias toward much higher Ka/Ks ratios in smaller genes, and moreover, this arbitrary lower limit to Ks had a greater impact on the Ka/Ks ratios of the human chimpanzee comparison than on the human-outgroup comparison (because the latter had fewer Ks = 0 values). When we ignored genes with Ks = 0 in the initial data set, the Ka/Ks ratio for rearranged over colinear chromosomes (R/C) for the human-chimpanzee comparison was found larger than the R/C ratio for the human-outgroup comparison (1.65 *versus* 1.42). Although these were not large differences, they hinted to an acceleration of protein evolution in association with recent rearrangements.

We also applied an improved methodology to detect selection in the same dataset. We estimated the Ka/Ks ratios of the individual branches leading to humans, chimpanzees, and outgroups. Interestingly, the branch leading to humans always showed higher R/C ratios than that leading to outgroups (1.76 versus 1.58 using unrooted trees, and 2.05 versus 1.15 using rooted trees). Again, this suggested a recent increment in the association between rearrangements and high Ka/Ks ratios.

As to why that reported association among chromosomal rearrangements and higher evolutionary rates existed, we suggested three potential causes: First, an association between fast divergence genes and chromosomal conformations without any relation with speciation. The establishment of rearrangements could have changed the expression of associated genes (for instance by disrupting the promoter sequence), which might have triggered amino-acid changes; or maybe rearrangements may have been established more easily if they hitchhiked with positively selected variants, which may occur more frequently near clusters of rapidly evolving genes. A second explanation was that rearranged regions with clusters of positively selected genes would have taken part in recurrent chromosomal speciation processes. In primates, some chromosomes have undergone intense rearrangements, whereas some syntenic groups have been conserved (Haig 1993; Muller and Wienberg 2001) and breakpoints for rearrangements are highly conserved within primates (Mouse Consortium 2002; Murphy et al. 2005; Ruiz-Herrera et al. 2002a; Ruiz-Herrera et al. 2002b). Al these facts suggested that the genomic distribution of rearrangements had not been random, and that the same regions might have been rearranged over and over in different branches within the primate lineage and, perhaps, had taken part in different speciation processes.

The last potential explanation that we proposed was that the GenBank sequences used by the initial study and by Lu et al. (2003) were not a random sample of the whole genome. By the time of Lu's work and our reply, Hellmann et al. (2003b) had already published an analysis of human and chimpanzee divergence by comparing more than 1200 random, highly-expressed chimpanzee expressed sequence tags (ESTs) to their human orthologs. The Ka/Ks ratio calculated in that study was 0.22, almost three-fold smaller than the one based on the limited dataset from GenBank used in the paper by Navarro and Barton (2003b). This suggested that the small dataset used in the initial study from Navarro and Barton (2003b) was biased and was not representative of the whole human genome and that the main results in that paper might have been an artifact.

1.2 Follow-up studies and integration of different kinds of data

Other studies on sequence data.

After these initial approaches to the problem, other studies tackled the question from different points of view and using different kinds of data. The first study revisiting the topic was by Zhang et al. (2004). In that work, BAC sequences covering ~4 Mb of aligned sequences of the human and chimpanzee genomes (< 0.1 % of the human genome) were used. The average nucleotide divergence (mainly from non-coding DNA) for sequences in the two kinds of chromosomes (rearranged versus colinear) were 1.20% and 1.34% respectively. Although this difference was not statistically different, it was in the opposite direction from the predictions and observations by Navarro and Barton (2003b). They also performed a study of the genomic distribution of genes with Ka/Ks > 1 and found an accumulation of those genes in colinear chromosomes, again contrary to the predictions of the model (P-value < 0.01). Moreover, they carried out a comparison of non-synonymous and synonymous substitutions in 69 genes, failing to detect any acceleration of protein evolution rates in rearranged chromosomes. Finally, the paper by Zhang et al. (2004) was the first one using gene expresion data to test a model of chromosomal speciation. Basically, they reanalyzed a dataset from Caceres et al. (2003) and calculated differences in gene expression intensity between humans and chimpanzees for rearranged and colinear chromosomes. They did not find any statistical differences between chromosomal classes. However, Zhang et al. (2004) did not use the whole dataset, but considered only those genes with statistical different expression pattern in humans and chimpanzees, and, thus, drastically reduced their statistical power (they only used 152 genes). Among other conclusions, Zhang et al. (2004) suggested that, since linked genes tend to have similar rates of protein evolution (Williams and Hurst 2000), it is possible that the small sample used by Navarro and Barton (2003b) was biased towards some clustering of rapidly evolving genes in rearranged chromosomes.

A few months later, a new analysis by Vallender and Lahn (2004) was published. This analysis re-examined the hypothesis of chromosomal speciation between humans and chimpanzees making use of the largest database available at that moment. They analyzed more than 7000 ESTs from the study by Clark et al. (2003). The analysis performed was basically a repetition of the previous ones, only that they used a

larger dataset. They did not detect any difference in the average of Ka/Ks ratio between chromosomal classes. In addition, they unexpectedly detected a non significant excess of genes with Ka/Ks > 1 in colinear chromosomes . Making use of rodent outgroups they also detected accelerated genes in primate rearranged chromosomes when comparing with rat or mouse orthologous sequences, which let them to state that these results favoured the hypothesis of clustering of rapidly evolving genes rather than being related any speciation process. The second interesting result of the paper was to show that, when comparing two close species (such as humans and chimpanzees) using small datasets (as the one used in the study from Navarro and Barton (2003b)), stochastic noise in substitution rates are prone to appear. They showed that Ks seems to have huge stochastic variances leading to huge fluctuations on the estimations of the ratio Ka/Ks and this biasing the results from small datasets.

The next paper that addressed the question using DNA sequences, was the chimpanzee genome paper (Mikkelsen et al. 2005) in which the Chimpanzee Genome Consortium analyzed evolutionary rates for more than 13000 unambiguously orthologous genes between humans and chimpanzees. This is by far the most complete database available nowadays and thus results from that paper are, in principle, reliable. Overall, and according to that paper, rearranged chromosomes do not present any acceleration in evolutionary rates and, thus, current data do not appear to support the predictions of chromosomal speciation models. However, they also compared evolutionary rates within and outside the inversion for each chromosome (as previously done by Marques-Bonet et al. (2004)). From the results, they claimed that only HSA4 (*Homo sapiens*, chromosome 4) and HSA5 have a significant increase in Ka/Ks for genes located within the inversion. Still, those differences are small compared to the initially reported figure of a two-fold of increase. Mikkelsen et al. (2005) attributed those increases of Ka/Ks to low and local Ks values rather than any speciation related explanation. This work, thus, confirmed the previous worries ((Navarro et al. 2003; Vallender and Lahn 2004)) that the scarce amount of data and a bias in GenBank sequences were the causes of the original results by Navarro and Barton (2003b). Indeed, the average Ka/Ks ratio between between human and chimpanzee genes was 0.23, quite close to the one suggested by Hellmann et al. (2003b) and certainly far away from the figure 0.61 oftained the initial paper from GenBank sequences.

Studies based on gene expression data

The next study to address the topic made use of the divergence in gene expression intensity between humans and chimpanzee was authored by us (Marques-Bonet et al. 2004). From several array-based comparative gene expression studies, we gathered published human and chimpanzee gene expression data from several tissues: cerebral cortex, liver, heart and fibroblasts (Caceres et al. 2003; Enard et al. 2002; Karaman et al. 2003). After excluding known factors that could have biased our analyses, we observed that, in brain cortex tissue, the average of differences in expressions patterns between humans and chimpanzees was statistically higher in rearranged chromosomes than in colinear chromosomes (1.543 *vs.* 1.463, P-value < 0.001). This observation was replicated by (Khaitovich et al. 2004) who found that rearranged chromosomes harbor an excess of genes differentially expressed between humans and chimpanzees.

After this initial results, we took into consideration previous criticisms and tested the possibility that rearrangements would have happened in regions of faster gene expression divergence. To do so, we used macaque as an outgroup. Our results suggested that genomic regions that have been rearranged between humans and chimpanzees do not show larger expression divergence from macaques, thus rejecting this explanation. We also looked for clustering of highly divergent genes, but the analysis showed that this could not be the main cause for our results.

The large amount of data available in gene-expression arrays allowed us to perform a detailed chromosome-per-chromosome analysis, together with an exhaustive study of gene expression divergence along chromosomes. One of our focuses was on inversion breakpoints and we found that regions surrounding breakpoints present the largest expression divergence between humans and chimpanzees. This was not surprisingly, as chromosomal rearrangements have direct effects over gene expression. Chromosomal rearrangements can break enhancers or promoters that control the expression of certain genes (Puig et al. 2004; Spitz et al. 2003; Tanimoto et al. 1999) or change the genomic context in which genes are immersed, thus altering mutation rates and, possibly, the rates of gene-expression divergence. However, not all the breakpoints did show an increase in gene expression divergence. In fact, only breakpoints in four chromosomes (HSA1, HSA5, HSA9 and HSA16) presented clearly higher differences than genes in colinear chromosomes. In addition, the breakpoint effect did not explain our main observation: after removing genes in regions

sorrounding breakpoints, rearranged chromosomes still presented statistically higher differences in gene expression divergence, thus supporting an association between chromosomal rearrangements and larger differences in gene expression.

Once an association between rearrangements higher gene expression divergence had been established and the association determined to be exclusive of humans and chimpanzees we made use of an outgroup to ascertain if it had been caused by direc or indirect effects of rearrangements. We devised a phylogenetic test based on the predictions of different scenarios in which rearrangements could have been involved. For instance, if there were direct effects of the rearrangements over expression divergence, we should detect acceleration of gene expression divergence only in the lineage in which the rearrangements took place. Thus, we used macaque data to build gene-expression phylogenetic trees for every gene, and were able to identify genes presenting faster rates of gene expression change in the same lineages in which the rearrangement took place. We removed these genes and repeated our analysis. Results still showed a strong association between rearrangements and higher gene expression divergence (FC=1.59 in rearranged chromosomes versus 1.46 in colinear chromosomes, P-value<0.001) thus suggesting that indirect effects of the rearrangements, perhaps related to speciation, could help explaining our observations.

As to the discrepancies between our results and those from Zhang et al. (2004), there are two main reasons as to why they did not detect any statistically significant association between chromosomal rearrangements and gene-expression divergence in the human cortex. First, when they analyzed the same dataset that we used in our paper (Caceres et al. 2003), they limited their study to genes for which significant expression differences between human and chimpanzee had been detected. Moreover, when testing for a possible accumulation of these differentially expressed genes in rearranged chromosomes, they estimated the proportion of differentially expressed genes per Mb (or per predicted number of genes) in rearranged versus colinear chromosomes. Instead, we used what we consider a better estimate: the proportion of differentially expressed genes in one or the other class of chromosome per number of genes in microarrays that map to these chromosomes. At any rate, all the results in that particular part of their analysis followed the same trend than ours.

1.3 The importance of chromosomal structure

One of the main issues in any study of the effects of chromosomal rearrangements in human-chimpanzee divergence is to determine the boundaries of inversions. The first studies testing chromosomal speciation could only map inversions at the level of cytological bands (Navarro and Barton 2003b), while subsequent studies could make use of more refined maps that combined data from experimental and computational studies (Marques-Bonet et al. 2004; Mikkelsen et al. 2005). One of the most important providers of refined cytological data has been the laboratory of Prof. Hildegard Kehrer-Sawatzi. This group has published several papers precisely mapping the breakpoints of most of the pericentromeric inversions at the Kb level (Goidts et al. 2005; Kehrer-Sawatzki et al. 2005a; Kehrer-Sawatzki et al. 2005b; Kehrer-Sawatzki et al. 2002; Kehrer-Sawatzki et al. 2005c; Szamalek et al. 2005). Recently they have also published a comparative study of inversion breakpoints in the genomes of common chimpanzees (*Pan troglodytes*) and bonobos (*Pan paniscus*) (Szamalek et al. 2006b). In that paper, they show that *P. troglodytes* and *P. paniscus* share exactly the same pericentric inversions, and thus these inversions predate the separation of the two species of chimpanzee (0.86-2 myr). They also claim that those inversions are likely to have been present at the time of the separation of the human/chimpanzee clade. This conclusion is obviously important in the context of the old debate about the anagenic or cladogenic origins of the inversions separating humans and chimpanzees.

Other studies have tried to map structural variation between humans and chimpanzees at the sub-cytological level (Feuk et al. 2005; Newman et al. 2005; Szamalek et al. 2006a). Newman et al. (2005) were the first to study structural variation sites. They mapped chimpanzee fosmid paired-end sequences against the human genome to systematically identify sites of structural variation (larger than 12 kb) between the two species. Among other structural variants, they detected 174 potential new inversions spanning from 12 kb up to 1000 kb. Experimental validation confirmed ~80% of those findings. Alternatively, Feuk et al. (2005) used net alignments for the human and chimpanzee genome assemblies to identify a total of 1576 putative regions of inverted orientation, covering more than 154 Mb of DNA (ranging inversions from 23 base pairs up to 62 mega-bases). Experimental validation confirmed 23 (85%) of 27 semi-randomly chosen regions. One of the new aspects of the paper was the use of gorilla as an outgroup to assign ancestral status to the

variants. They also found that some of those variants were polymorphic in the human genome, thus suggesting recent and lineage-specific rarrangement events. As a result, they claimed that this type of structural variation may be a more common feature of our genomes than previously though and it is an important source of variation in primate genome evolution. Finally, Szamalek et al. (2006a), compared the order in humans and chimpanzees of more than 10.000 orthologous genes, and found 71 putative micro-rearrangements with a validation percentage of ~60%. This validation percentage is lower than in the studies summarized above, most likely due to the low quality of the first assembly of the chimpanzee genome. In the same paper the authors showed that four of the validated inversions were polymorphic in the two species (3 in chimpanzees and 1 in humans).

In summary, the mini-inversions detected in the more recent publications illustrate that structural variants are more common than previously thought reinforces that idea that they may have played an important role on the evolution of primates. The fact that some of them are still segregating in either human or chimpanzee populations opens interesting new aspects in the study of structural genomic variability.

1.4 Latest results

The most recent two papers addressing the relationship between DNA sequence divergence and chromosomal rearrangements in humans and chimpanzees come from our lab (Marques-Bonet et al. 2006a; Marques-Bonet et al. 2006b). The first work is a detailed reanalysis of the most complete dataset of human-chimpanzee orthologous genes (Mikkelsen et al. 2005). The second paper presents a different approach to the same problem and focuses on SDs, which, to date, had been ignored in this context to the point of being removed in all previous analyses. Studying SDs is important, since they have been shown to be involved in rearrangements, genic novelties, differential expression in the brain and higher evolutionary rates, and, therefore, they are good candidates to be playing a role in chromosomal speciation. Those papers are recent enough to have an updated discussion (they were written in parallel to this document), so I will not discuss them here. What follows is a short common conclusion from both papers

Our most recent study of single-copy genes present results that contradict all previous literature. Overall, genes located in chromosomes rearranged between humans and chimpanzees, specially genes within the rearrangements themselves, present lower non-coding divergence than genes elsewhere in the genome. These results suggest that chromosomal speciation processes have not been common along the human and chimpanzee lineage. As to SDs, interespecific-divergence measures present a similar general relationship with rearrangements than as single copy DNA. Again, therefore, there is no evidence of a recurrent and direct role of SDs in chromosomal speciation processes between humans and chimpanzees.

However, if rearrangements did really play a role in some speciation event, we do not expect that all the chromosomes would have made the same contribution and, thus, speciation would not have left the same signals in all chromosomes. This leaded us to perform a chromosome-per-chromosome analysis in both works. Interesting patterns were unveiled in several chromosomes. HSA9, for example, stands out as a good candidate for a complex relationship among SDs, rearrangements and higher divergence among copies within inversions. Also, HSA4 was a clear outlier, presenting higher divergence within its inversion for both coding and non-coding sequences (Marques-Bonet et al. 2006b). Finally other chromosomes (such as HSA1, HSA4, HSA5,

HSA9, HSA15 and HSA16) were pinpointed as the most different in terms of gene expression intensity in the comparative gene-expression study (Marques-Bonet et al. 2004).

All these analyses were performed with the working draft version of the chimpanzee genome (commonly known as version 1). As noted by Taudien et al. (2006), a significant proportion of differences detected comparing the human reference genome versus this chimpanzee draft are sequencing errors rather than real differences. Only subsequent analyses making use of the new version of the chimpanzee genome (2.0, available from July, 2006), together with other primate genomes (gorilla, macaque or orangutan) as outgroups will help settling this interesting subject.

2. Testing chromosomal speciation in the rodent lineage.

The order Rodentia presents high karyotipical diversity. Huge chromosomal differentiation between species and great variety of chromosomal races within species provide evidence of a recurrent relationship between chromosomal rearrangements and speciation in rodents (Patton and Sherwood 1983). For instance, in the case of the house mice (*Mus musculus*) there are at least 40 local ''karyotypic races'' distributed all over West Europe and North Africa that are characterized by a reduced chromosome number, down to 2n = 22 in the most extreme cases (Searle 1998). Estimated rates of chromosomal change for rodents are among the highest observed in mammals (Mouse Consortium 2002; Murphy et al. 2005), but little is know about the effects of those rearrangements on the speciation process that separated mouse and rat.

The mouse and rat genomes have diverged during 12-24 Myrs (Mouse Consortium 2002; Gibbs et al. 2004) and are very divergent from the common ancestor of eutherian mammals, both in terms of their highly rearranged genomes, and in the relatively large number of nucleotide substitutions in selectively neutral sites that they have accumulated (Bourque et al. 2004; Mouse Consortium 2002; Gibbs et al. 2004; Webber and Ponting 2005). Considering all these facts, rodents are an excellent model to test for an association between chromosomal evolution and evolutionary rates.

Our first study (Marques-Bonet and Navarro 2005), was centered on testing this association in coding sequences of more than 12,000 human-mouse orthologous genes. We explored the distribution of evolutionary rates relative to genome rearrangements that we detected "de novo" from gene order comparisons. As the human–mouse ancestor is thought to be as distant as 87 Myr (Springer et al. 2003) and rodent genomes have been rearranged over and over during that period, we did not expect a clear relationship between evolutionary rates and rearrangements. Besides, we had to consider other factors that have been shown to affect evolutionary rates, such as GC content (Castresana 2002; Ebersberger et al. 2002; Matassi et al. 1999), local positioning of genes (Lercher et al. 2001; Williams and Hurst 2000; Williams and Hurst 2002), CpG rich islands (Bernardi 2000; Hardison et al. 2003), recombination rates (Hellmann et al. 2003a; Nekrutenko and Li 2000), or

duplications (Jordan et al. 2004; Lynch and Conery 2000a). Controlling for these factors implied, as usual, a strict filtering process that resulted in a considerable reduction of our dataset.

In spite of all those potential pitfalls, our results showed that genes in highly rearranged regions presented larger divergence that genes elsewhere in the genome, thus supporting an association among rearrangements and faster evolution. Furthermore, we also found that genes close to individual breakpoints (of both, translocations and inversions) had higher neutral divergence (Ks) than genes located in colinear regions. Neither GC content nor recombination rates, when used as covariates, could explain these differences in divergence. Two main explanations were put forward in the paper. First, we underlined the idea that if rearrangements took place recursively along a lineage (and they participated in speciation processes) the reported association can be expected. Breakpoints tend to be reused during the evolution (Bourque et al. 2004; Mouse Consortium 2002; Murphy et al. 2005; Ruiz-Herrera et al. 2002a; Ruiz-Herrera et al. 2002b; Zhao et al. 2004) and, thus, some of them could have participated not only in the initial speciation between human and mice, but in other speciation events all along their lineages. The second potential explanation we suggested was that regions in which the rearrangements took place are in fact special regions of faster genic evolution.

Our second study (Armengol et al. 2005) was the result of a collaboration with Dr. X. Estivill's group, who had previous experience on the study of the relationship between chromosomal rearrangements and segmental duplications in other species (Armengol et al. 2003). They extended their initial studies analyzing the association between SDs and rearrangements in the rodent lineage by comparing rat and mouse. The main goal of the paper was to provide further evidence of the role of segmental duplications and chromosomal rearrangements in the evolution of the architecture of mammalian chromosomes. Within this project, we were responsible of analyzing the genomic distribution of genic evolutionary rates relative to major rearrangements separating the genomes of rats and mice. The main result of our work was the finding that genes located within inversions evolved faster than genes located outside inversions. However, this result was significant for Ka and Ks, but not for the ratio (Ka/Ks). In addition, we showed that genes close to breakpoints also had higher neutral divergence (Ks).

In summary, human-mouse and mouse-rat results are congruent with the predictions of the model. Still, this is not enough to accept chromosomal speciation as the only causal explanation for these results. Indeed, it is clear that many other factors could be contributing to our observation and only making use of outgroups would help to clarify this issue.

Perhaps the clearest alternative to chromosomal speciation would be that certain genomic regions with special features are prone to be rearranged. These special features could be higher mutation rates or lower constraints and, thus, weaker purifying of selection. Given the results of recent papers, this kind of phenomena may well be a major contributor to our observations. The relationship between rearrangement breakpoints and higher neutral evolution found in our papers has been recently confirmed by Webber and Ponting (2005), who analyzed the dog genome, thus suggesting a consistent effect in the whole mammalian evolution. These authors found significant negative correlations between either G+C content or Ks and distance to a synteny breakpoint. This suggested that chromosomal rearrangements (especially chromosomal fissions) tend to happen in regions of ancestral high GC content. In mammalian genomes, CpG dinucleotides tend to be mutated to TpG through a methilation and deamination processes, leading to higher divergence measures (Castresana 2002; Ebersberger et al. 2002; Matassi et al. 1999; Webber and Ponting 2005; Yi et al. 2002). Thus, higher GC content could act at the same time as source for chromosomal rearrangements (by an as yet undetermined mechanism) and as a source for higher mutation rates.

Other potential explanations for our observations have been suggested by authors who have linked recombination, divergence and GC content by means of biased gene conversion (BGC) (Eyrewalker 1993; Galtier 2003; Marais 2003). Under the BGC model, chromosomal locations where recombination is highest should, over time, increase their GC content while decreasing nucleotide substitution rates (Lindblad-Toh et al. 2005; Marais 2003). Thus, among other predictions these models suggest that GC-rich isochores would correspond to highly recombining regions of the mammalian genome (Galtier 2003). We find higher GC-content associated with rearranged zones, but we did find neither an increase on recombination rates nor a decrease of divergence in regions inside inversions, or close to evolutionary breakpoints. Therefore, our results render no support to the BGC model. Finally, Jensen-Seaman et al. (2004) found that there was a positive correlation among

recombination and mutation in rat but a negative correlation in mouse, and, thus, the potential interaction among recombination and divergence should be revised carefully.

Finally, SDs could be contributing to our observations. There is a well-known relationship between segmental duplications (SDs), evolutionary breakpoints and genic evolution. SDs are one of the main forces of evolution and are usually found having higher evolutionary rates (Gu 2003; Marques-Bonet et al. 2006a; Marques-Bonet and Navarro 2005; Marques-Bonet et al. 2006b; Sharp et al. 2006). Also, and as mentioned above, many studies have noted a significant association between the location of segmental duplications and regions of chromosomal instability or evolutionary rearrangement (Armengol et al. 2003; Bailey et al. 2004; Locke et al. 2003). Given all these facts, SDs would be good candidates to be associated with both rearrangements and higher evolution rates (especially in primates, where it is found to be an expansion of SDs). However, in our results we carefully removed any SDs before the strudy of evolutionary rates relative to rearrangements, so they could not be the main responsible for them. The only exception might be ancient SDs, that could have diverged enough to remain undetected as duplications and, thus, could be adding noise to our results. Again, outgroups will help to clarify this issue.

In summary, other potential factors besides chromosomal speciation may be explaining the association between divergence and chromosomal rearrangements in mammals. Whatever the exact nature of these factors, our observations would in fact be reflecting local variation in underlying mutation rates. The causes of such variation may be diverse, including differences in nucleotide composition, DNA metabolism, recombination-associated mutagenesis, transcription-associated mutagenesis or even replication timing. Nuclear location may also be involved, including proximity to heterochromatin, nuclear membrane, and origins of replication (Mouse Consortium 2002). Thus, it is clear that the status of our observations in the rodent lineage will only be clarified by further studies using other complete genomes as outgroups and performing deeper analysis of non-coding DNA sequences.

3. Differences among lineages

The relationship between chromosomal rearrangements and genic evolution is different in the lineages of rodents and primates. Two main reasons can account for such discrepancy. First, and as mentioned above, chromosomal rearrangements do not *have to* promote speciation. Speciation is a complex issue, and one can not oversimplify the topic trying to explain all extant pairs of species as a result of simple speciation processes or trying to prove that every rearrangement was involved in speciation. Thus, it is possible that certain lineages undergo recurrent chromosomal speciation processes, whereas speciation in other lineages tends to be achieved by other mechanisms. The lineages of primates and rodents present particularities that are consistent with this hypothesis of "diversity of modes of speciation". For instance, both lineages differ, among other features, in their rates of chromosomal rearrangements or the presence of chromosomal races (all of them being higher in rodents), thus, apparently chromosomal speciation would be easier in rodents. Besides, any potential third factor linking at the same time rearrangements and genic evolution, could have had different impact in different lineages, maybe because of different biological features of the organisms, such as metabolic rates or generation time.

Also, it is important to consider that the amount of available information is not equivalent in both lineages. While the human genome has been deeply explored, rodents, although being central in medical research, have been relatively overlooked. Indeed, this bias can also be seen in the present work. Primate genomic information is, by far, more complete and detailed than that of the rodents and, thus, any lineage-specific conclusions should be taken with caution.

4. Discrepancy between divergence in gene expression studies and DNA sequence in primates

While our study of the distribution of DNA sequence divergence between humans and chimpanzees showed that rearrangements are associated to lower evolutionary rates, our analysis of gene expression data detected increased differences in gene expression intensity within certain inversions. This is a remarkable discrepancy, since gene sequences and gene expression have been found to evolve similarly (Khaitovich et al. 2005). Several causes could underly this discrepancy. First, differences in gene expression and divergence in coding sequences do not necessarily have to be linked when considering individual genes or regions, or specific tissues. Mutations in promoter sequences are more likely to be the main cause of changes in gene expression rather than changes in coding sequences. However, it is also true that substitution rates in coding and noncoding regions are correlated (Makalowski and Boguski 1998). Second, the comparison of gene expression in different species is a complex issue. Gene expression levels change over developmental stages, across different tissues, and in different environments. Moreover, tissues available for such studies do come from corpses and mRNA degradation after death is a difficult variable to control for. Finally, although we strived to control for this factor in our study, nucleotide differences can also account for biases in gene expression microarrays, since probes are usually constructed based on DNA from a single species. That is, probes are based the human reference genome, thus ignoring both human and chimpanzee polymorphim. Because of these and other problems, mRNA expression measurements include a large noise component, so any conclusions based on them should be carefully revisited in the light of new data.

Final Conclusions

The final conclusions of the present work are:

a) Current analyses do not support the hypothesis of recurrent chromosomal speciation processes along the human and chimpanzee lineages.

Current human-chimpanzee DNA divergence data do not fit the predictions of the supressed-recombination chromosomal speciation model. However, several questions remain open. What was the role (if any) of certain individual pericentromeric inversions separating humans and chimpanzees in the speciation processes along the two lineages? This is an interesting issue, particularly as it is seems likely that at least some inversions were segregating at the time of the original speciation between humans and chimpanzees and some rearrangements appear to fulfil the prediction of higher divergence. Another remaining issue is that of the relationship between DNA sequence and gene expression differences between humans and chimpanzees. Why are our results apparently contradictory? And finally, have SDs played any role in speciation processes?

b) Available data is congruent with recurrent chromosomal speciation in rodents, but that is not the only potential explanation for the observation.

Rodent data match some of the predictions of suppressed-recombination chromosomal speciation models. However, these results are not conclusive and other explanations cannot be ruled-out. Although we carefully controlled for many confounding factors, it is still possible that we did not control for everything. Only deeper analysis of protein coding genes, non-coding DNA, SDs and gene expression data (especially making use of outgroups) would help to resolve this issue.

c) Chromosomal and molecular evolution are linked by several factors, but their relationship is different in different lineages.

Based on the results of this work and many others, chromosomal rearrangements are linked to molecular evolution by means of several mechanisms. In this work we have shown, for example, that GC content is higher near breakpoints with subsequent effects on divergence of nearby genomic sequence; or that genes close to breakpoints show higher differences in gene expression intensity between close

species. All these results, and others, suggest that structural and molecular influence each other, but that their relationship differs in different lineages. Only a better understanding of the variables that model chromosome dynamics, basically by means of whole-genome analyses, will allow us to generalize about processes linking different levels of genome evolution.

References

Allendorf, F.W. 1978. Protein polymorphism and the rate of loss of duplicate gene expression. *Nature* **272:** 76-78.
Allendorf, F.W. 1979. Rapid loss of duplicate gene expression by natural selection. *Heredity* **43:** 247-258.
Andolfatto, P., F. Depaulis, and A. Navarro. 2001. Inversion polymorphisms and nucleotide variability in Drosophila. *Genetical Research* **77:** 1-8.
Armengol, L., T. Marques-Bonet, J. Cheung, R. Khaja, J.R. Gonzalez, S.W. Scherer, A. Navarro, and X. Estivill. 2005. Murine segmental duplications are hot spots for chromosome and gene evolution. *Genomics* **86:** 692-700.
Armengol, L., M.A. Pujana, J. Cheung, S.W. Scherer, and X. Estivill. 2003. Enrichment of segmental duplications in regions of breaks of synteny between the human and mouse genomes suggest their involvement in evolutionary rearrangements. *Human Molecular Genetics* **12:** 2201-2208.
Ayala, F.J. and M. Coluzzi. 2005. Chromosome speciation: Humans, Drosophila, and mosquitoes. *Proceedings of the National Academy of Sciences of the United States of America* **102:** 6535-6542.
Bailey, J.A., R. Baertsch, W.J. Kent, D. Haussler, and E.E. Eichler. 2004. Hotspots of mammalian chromosomal evolution. *Genome Biology* **5:** -.
Bailey, J.A., A.M. Yavor, H.F. Massa, B.J. Trask, and E.E. Eichler. 2001. Segmental duplications: Organization and impact within the current Human Genome Project assembly. *Genome Research* **11:** 1005-1017.
Balding, D., M. Bishop, and Cannings. 2003. *{Handbook of Statistical Genetics}*. Wiley.
Barton, N.H. 2006. Evolutionary biology: how did the human species form? *Curr Biol* **16:** R647-650.
Basset, P., Yannic, G, Brünner, H, and Hausser J. 2006. Restricted gene flow at specific parts of the shrew genome in chromosomal hybrid zones. *Evolution (In press)*.
Bedell, M.A., N.A. Jenkins, and N.G. Copeland. 1996. Good genes in bad neighbourhoods. *Nat Genet* **12:** 229-232.
Bernardi, G. 2000. Isochores and the evolutionary genomics of vertebrates. *Gene* **241:** 3-17.
Besansky, N.J., J. Krzywinski, T. Lehmann, F. Simard, M. Kern, O. Mukabayire, D. Fontenille, Y. Toure, and N. Sagnon. 2003. Semipermeable species boundaries between Anopheles gambiae and Anopheles arabiensis: evidence from multilocus DNA sequence variation. *Proc Natl Acad Sci U S A* **100:** 10818-10823.
Bonhoeffer, S., E.C. Holmes, and M.A. Nowak. 1995. Causes of HIV diversity. *Nature* **376:** 125.
Bourque, G., P.A. Pevzner, and G. Tesler. 2004. Reconstructing the genomic architecture of ancestral mammals: lessons from human, mouse, and rat genomes. *Genome Res* **14:** 507-516.
Britten, R.J. 2002. Divergence between samples of chimpanzee and human DNA sequences is 5%, counting indels. *Proc Natl Acad Sci U S A* **99:** 13633-13635.
Bromham, L., A. Rambaut, and P.H. Harvey. 1996. Determinants of rate variation in mammalian DNA sequence evolution. *J Mol Evol* **43:** 610-621.
Caceres, M., J. Lachuer, M.A. Zapala, J.C. Redmond, L. Kudo, D.H. Geschwind, D.J. Lockhart, T.M. Preuss, and C. Barlow. 2003. Elevated gene expression levels distinguish human from non-human primate brains. *Proc. Nat. Acad. Sci. (USA)* **100:** 13030-13035.
Castresana, J. 2002. Genes on human chromosome 19 show extreme divergence from the mouse orthologs and a high GC content. *Nucleic Acids Research* **30:** 1751-1756.

Chen, F.C. and W.H. Li. 2001. Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. *American Journal of Human Genetics* **68:** 444-456.
Cheng, Z., M. Ventura, X.W. She, P. Khaitovich, T. Graves, K. Osoegawa, D. Church, P. DeJong, R.K. Wilson, S. Paabo, M. Rocchi, and E.E. Eichler. 2005. A genome-wide comparison of recent chimpanzee and human segmental duplications. *Nature* **437:** 88-93.
Clark, A.G., S. Glanowski, R. Nielsen, P.D. Thomas, A. Kejariwal, M.A. Todd, D.M. Tanenbaum, D. Civello, F. Lu, B. Murphy, S. Ferriera, G. Wang, X. Zheng, T.J. White, J.J. Sninsky, M.D. Adams, and M. Cargill. 2003. Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios. *Science* **302:** 1960-1963.
Coluzzi, M., A. Sabatini, A. della Torre, M.A. Di Deco, and V. Petrarca. 2002. A polytene chromosome analysis of the Anopheles gambiae species complex. *Science* **298:** 1415-1418.
Conant, G.C. and A. Wagner. 2003. Asymmetric sequence divergence of duplicate genes. *Genome Research* **13:** 2052-2058.
Consortium, I.H.G.S. 2001. Initial sequencing and analysis of the human genome. *Nature***:** 860-921.
Consortium, M.G.S. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420:** 520-562.
Cooper, D.N. and H. Youssoufian. 1988. The CpG dinucleotide and human genetic disease. *Human Genetics* **78:** 151-155.
Crow, J.F. 2000. A new study challenges the current belief of a high human male : female mutation ratio. *Trends in Genetics* **16:** 525-526.
Deeb, S.S. 2005. The molecular basis of variation in human color vision. *Clin Genet* **67:** 369-377.
Delneri, D., I. Colson, S. Grammenoudi, I.N. Roberts, E.J. Louis, and S.G. Oliver. 2003. Engineering evolution to study speciation in yeasts. *Nature* **422:** 68-72.
Dennehey, B.K., D.G. Gutches, E.H. McConkey, and K.S. Krauter. 2004. Inversion, duplication, and changes in gene context are associated with human chromosome 18 evolution. *Genomics* **83:** 493-501.
Ebersberger, I., D. Metzler, C. Schwarz, and S. Paabo. 2002. Genomewide Comparison of DNA Sequences between Humans and Chimpanzees. *American Journal of Human Genetics* **70:** 1490-1497.
Enard, W., P. Khaitovich, J. Klose, S. Zollner, F. Heissig, P. Giavalisco, K. Nieselt-Struwe, E. Muchmore, A. Varki, and R. Ravid. 2002. Intra- and Interspecific Variation in Primate Gene Expression Patterns. *Science* **296:** 340-342.
Eyrewalker, A. 1993. Recombination and Mammalian Genome Evolution. *Proceedings of the Royal Society of London Series B-Biological Sciences* **252:** 237-243.
Feuk, L., J.R. MacDonald, T. Tang, A.R. Carson, M. Li, G. Rao, R. Khaja, and S.W. Scherer. 2005. Discovery of human inversion polymorphisms by comparative analysis of human and chimpanzee DNA sequence assemblies. *Plos Genetics* **1:** 489-498.
Galtier, N. 2003. Gene conversion drives GC content evolution in mammalian histones. *Trends Genet* **19:** 65-68.
Gibbs, R.A. G.M. Weinstock M.L. Metzker D.M. Muzny E.J. Sodergren S. Scherer G. Scott D. Steffen K.C. Worley P.E. Burch G. Okwuonu S. Hines L. Lewis C. DeRamo O. Delgado S. Dugan-Rocha G. Miner M. Morgan A. Hawes R. Gill Celera R.A. Holt M.D. Adams P.G. Amanatides H. Baden-Tillson M. Barnstead S. Chin C.A. Evans S. Ferriera C. Fosler A. Glodek Z. Gu D. Jennings C.L. Kraft T. Nguyen C.M. Pfannkoch C. Sitter G.G. Sutton J.C. Venter T. Woodage D. Smith H.M. Lee E. Gustafson P. Cahill A. Kana L. Doucette-Stamm K. Weinstock K. Fechtel R.B. Weiss D.M. Dunn E.D. Green R.W. Blakesley G.G.

Bouffard P.J. De Jong K. Osoegawa B. Zhu M. Marra J. Schein I. Bosdet C. Fjell S. Jones M. Krzywinski C. Mathewson A. Siddiqui N. Wye J. McPherson S. Zhao C.M. Fraser J. Shetty S. Shatsman K. Geer Y. Chen S. Abramzon W.C. Nierman P.H. Havlak R. Chen K.J. Durbin A. Egan Y. Ren X.Z. Song B. Li Y. Liu X. Qin S. Cawley K.C. Worley A.J. Cooney L.M. D'Souza K. Martin J.Q. Wu M.L. Gonzalez-Garay A.R. Jackson K.J. Kalafus M.P. McLeod A. Milosavljevic D. Virk A. Volkov D.A. Wheeler Z. Zhang J.A. Bailey E.E. Eichler E. Tuzun E. Birney E. Mongin A. Ureta-Vidal C. Woodwark E. Zdobnov P. Bork M. Suyama D. Torrents M. Alexandersson B.J. Trask J.M. Young H. Huang H. Wang H. Xing S. Daniels D. Gietzen J. Schmidt K. Stevens U. Vitt J. Wingrove F. Camara M. Mar Alba J.F. Abril R. Guigo A. Smit I. Dubchak E.M. Rubin O. Couronne A. Poliakov N. Hubner D. Ganten C. Goesele O. Hummel T. Kreitler Y.A. Lee J. Monti H. Schulz H. Zimdahl H. Himmelbauer H. Lehrach H.J. Jacob S. Bromberg J. Gullings-Handley M.I. Jensen-Seaman A.E. Kwitek J. Lazar D. Pasko P.J. Tonellato S. Twigger C.P. Ponting J.M. Duarte S. Rice L. Goodstadt S.A. Beatson R.D. Emes E.E. Winter C. Webber P. Brandt G. Nyakatura M. Adetobi F. Chiaromonte L. Elnitski P. Eswara R.C. Hardison M. Hou D. Kolbe K. Makova W. Miller A. Nekrutenko C. Riemer S. Schwartz J. Taylor S. Yang Y. Zhang K. Lindpaintner T.D. Andrews M. Caccamo M. Clamp L. Clarke V. Curwen R. Durbin E. Eyras S.M. Searle G.M. Cooper S. Batzoglou M. Brudno A. Sidow E.A. Stone J.C. Venter B.A. Payseur G. Bourque C. Lopez-Otin X.S. Puente K. Chakrabarti S. Chatterji C. Dewey L. Pachter N. Bray V.B. Yap A. Caspi G. Tesler P.A. Pevzner D. Haussler K.M. Roskin R. Baertsch H. Clawson T.S. Furey A.S. Hinrichs D. Karolchik W.J. Kent K.R. Rosenbloom H. Trumbower M. Weirauch D.N. Cooper P.D. Stenson B. Ma M. Brent M. Arumugam D. Shteynberg R.R. Copley M.S. Taylor H. Riethman U. Mudunuri J. Peterson M. Guyer A. Felsenfeld S. Old S. Mockrin and F. Collins. 2004. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428:** 493-521.

Gilad, Y., A. Oshlack, G.K. Smyth, T.P. Speed, and K.P. White. 2006. Expression profiling in primates reveals a rapid evolution of human transcription factors. *Nature* **440:** 242-245.

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

Goidts, V., J.M. Szamalek, H. Hameister, and H. Kehrer-Sawatzki. 2004. 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. *Human Genetics* **115:** 116-122.

Gojobori, T., E.N. Moriyama, and M. Kimura. 1990. Molecular clock of viral evolution, and the neutral theory. *Proc Natl Acad Sci U S A* **87:** 10015-10018.

Gu, X. 2003. Evolution of duplicate genes versus genetic robustness against null mutations. *Trends in Genetics* **19:** 354-356.

Gu, Z., D. Nicolae, H.H. Lu, and W.H. Li. 2002. Rapid divergence in expression between duplicate genes inferred from microarray data. *Trends in Genetics* **18:** 609-613.

Haig, D. 1993. The evolution of unusual chromosomal systems in coccoids: extraordinary sex ratios revisited. *J. Evol. Biol.* **6:** 69-78.

Hardison, R.C., K.M. Roskin, S. Yang, M. Diekhans, W.J. Kent, R. Weber, L. Elnitski, J. Li, M. O'Connor, D. Kolbe, S. Schwartz, T.S. Furey, S. Whelan, N. Goldman, A. Smit, W. Miller, F. Chiaromonte, and D. Haussler. 2003. Covariation in

frequencies of substitution, deletion, transposition, and recombination during eutherian evolution. *Genome Res* **13:** 13-26.
Hartl, D. and A. Clark. 1998. *Principles of Population Genetics*. Sinauer Associates.
Hellmann, I., I. Ebersberger, S.E. Ptak, S. Paabo, and M. Przeworski. 2003a. A neutral explanation for the correlation of diversity with recombination rates in humans. *American Journal of Human Genetics* **72:** 1527-1535.
Hellmann, I., S. Zollner, W. Enard, I. Ebersberger, B. Nickel, and S. Paabo. 2003b. Selection on human genes as revealed by comparisons to chimpanzee cDNA. *Genome Res* **13:** 831-837.
Hey, J. 2003. Speciation and inversions: chimps and humans. *Bioessays* **25:** 825-828.
Hinds, D.A., L.L. Stuve, G.B. Nilsen, E. Halperin, E. Eskin, D.G. Ballinger, K.A. Frazer, and D.R. Cox. 2005. Whole-genome patterns of common DNA variation in three human populations. *Science* **307:** 1072-1079.
Hughes, A.L. and M. Nei. 1988. Pattern Of Nucleotide Substitution At Major Histocompatibility Complex Class-I Loci Reveals Overdominant Selection. *Nature* **335:** 167-170.
Hurst, L.D. and H. Ellegren. 1998. Sex biases in the mutation rate. *Trends in Genetics* **14:** 446-452.
Innan, H. and H. Watanabe. 2006. The effect of gene flow on the coalescent time in the human-chimpanzee ancestral population. *Mol Biol Evol* **23:** 1040-1047.
Jensen-Seaman, M.I., T.S. Furey, B.A. Payseur, Y. Lu, K.M. Roskin, C.F. Chen, M.A. Thomas, D. Haussler, and H.J. Jacob. 2004. Comparative recombination rates in the rat, mouse, and human genomes. *Genome Res* **14:** 528-538.
Jordan, I.K., Y.I. Wolf, and E.V. Koonin. 2004. Duplicated genes evolve slower than singletons despite the initial rate increase. *BMC Evol Biol* **4:** 22.
Karaman, M.W., M.L. Houck, L.G. Chemnick, S. Nagpal, D. Chawannakul, D. Sudano, B.L. Pike, V.V. Ho, O.A. Ryder, and J.G. Hacia. 2003. Comparative Analysis of Gene-Expression Patterns in Human and African Great Ape Cultured Fibroblasts. *Genome Research* **13:** 1619-1630.
Kehrer-Sawatzki, H., C. Sandig, N. Chuzhanova, V. Goidts, J.M. Szamalek, S. Tanzer, S. Muller, M. Platzer, D.N. Cooper, and H. Hameister. 2005a. Breakpoint analysis of the pericentric inversion distinguishing human chromosome 4 from the homologous chromosome in the chimpanzee (Pan troglodytes). *Human Mutation* **25:** 45-55.
Kehrer-Sawatzki, H., C.A. Sandig, V. Goidts, and H. Hameister. 2005b. Breakpoint analysis of the pericentric inversion between chimpanzee chromosome 10 and the homologous chromosome 12 in humans. *Cytogenetic and Genome Research* **108:** 91-97.
Kehrer-Sawatzki, H., B. Schreiner, S. Tanzer, M. Platzer, S. Muller, and H. Hameister. 2002. Molecular Characterization of the Pericentric Inversion That Causes Differences Between Chimpanzee Chromosome 19 and Human Chromosome 17. *American Journal of Human Genetics* **71:** 375-388.
Kehrer-Sawatzki, H., J.M. Szamalek, S. Tanzer, M. Platzer, and H. Hameister. 2005c. Molecular characterization of the pericentric inversion of chimpanzee chromosome 11 homologous to human chromosome 9. *Genomics* **85:** 542-550.
Khaitovich, P., W. Enard, M. Lachmann, and S. Paabo. 2006. Evolution of primate gene expression. *Nat Rev Genet* **7:** 693-702.
Khaitovich, P., I. Hellmann, W. Enard, K. Nowick, M. Leinweber, H. Franz, G. Weiss, M. Lachmann, and S. Paabo. 2005. Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. *Science* **309:** 1850-1854.
Khaitovich, P., B. Muetzel, X. She, M. Lachmann, I. Hellmann, J. Dietzsch, S. Steigele, H.H. Do, G. Weiss, W. Enard, F. Heissig, T. Arendt, K. Nieselt-

Struwe, E.E. Eichler, and S. Paabo. 2004. Regional patterns of gene expression in human and chimpanzee brains. *Genome Res* **14:** 1462-1473.
Kimura, M. 1983. The neutral theory of molecular evolution. *Cambridge University Press, Cambridge*.
King, M. 1993. Species evolution: the role of chromosome change. *Cambridge University Press, Cambridge, UK*.
Kleinjan, D.J. and V. van Heyningen. 1998. Position effect in human genetic disease. *Hum Mol Genet* **7:** 1611-1618.
Kondrashov, F.A., I.B. Rogozin, Y.I. Wolf, and E.V. Koonin. 2002. Selection in the evolution of gene duplications. *Genome Biology* **3:** RESEARCH0008.
Kong, A., D.F. Gudbjartsson, J. Sainz, G.M. Jonsdottir, S.A. Gudjonsson, B. Richardsson, S. Sigurdardottir, J. Barnard, B. Hallbeck, G. Masson, A. Shlien, S.T. Palsson, M.L. Frigge, T.E. Thorgeirsson, J.R. Gulcher, and K. Stefansson. 2002b. A high-resolution recombination map of the human genome. *Nature Genetics* **31:** 241-247.
Kumar, S. and S. Subramanian. 2002. Mutation rates in mammalian genomes. *Proc Natl Acad Sci U S A* **99:** 803-808.
Laird, et, and al. 1969. Rate of fixation of nucleotide substitutions in evolution. *Nature* **224:** 149-?
Lander, E.S. L.M. Linton B. Birren C. Nusbaum M.C. Zody J. Baldwin K. Devon K. Dewar M. Doyle W. FitzHugh R. Funke D. Gage K. Harris A. Heaford J. Howland L. Kann J. Lehoczky R. LeVine P. McEwan K. McKernan J. Meldrim J.P. Mesirov C. Miranda W. Morris J. Naylor C. Raymond M. Rosetti R. Santos A. Sheridan C. Sougnez N. Stange-Thomann N. Stojanovic A. Subramanian D. Wyman J. Rogers J. Sulston R. Ainscough S. Beck D. Bentley J. Burton C. Clee N. Carter A. Coulson R. Deadman P. Deloukas A. Dunham I. Dunham R. Durbin L. French D. Grafham S. Gregory T. Hubbard S. Humphray A. Hunt M. Jones C. Lloyd A. McMurray L. Matthews S. Mercer S. Milne J.C. Mullikin A. Mungall R. Plumb M. Ross R. Shownkeen S. Sims R.H. Waterston R.K. Wilson L.W. Hillier J.D. McPherson M.A. Marra E.R. Mardis L.A. Fulton A.T. Chinwalla K.H. Pepin W.R. Gish S.L. Chissoe M.C. Wendl K.D. Delehaunty T.L. Miner A. Delehaunty J.B. Kramer L.L. Cook R.S. Fulton D.L. Johnson P.J. Minx S.W. Clifton T. Hawkins E. Branscomb P. Predki P. Richardson S. Wenning T. Slezak N. Doggett J.F. Cheng A. Olsen S. Lucas C. Elkin E. Uberbacher M. Frazier R.A. Gibbs D.M. Muzny S.E. Scherer J.B. Bouck E.J. Sodergren K.C. Worley C.M. Rives J.H. Gorrell M.L. Metzker S.L. Naylor R.S. Kucherlapati D.L. Nelson G.M. Weinstock Y. Sakaki A. Fujiyama M. Hattori T. Yada A. Toyoda T. Itoh C. Kawagoe H. Watanabe Y. Totoki T. Taylor J. Weissenbach R. Heilig W. Saurin F. Artiguenave P. Brottier T. Bruls E. Pelletier C. Robert P. Wincker D.R. Smith L. Doucette-Stamm M. Rubenfield K. Weinstock H.M. Lee J. Dubois A. Rosenthal M. Platzer G. Nyakatura S. Taudien A. Rump H. Yang J. Yu J. Wang G. Huang J. Gu L. Hood L. Rowen A. Madan S. Qin R.W. Davis N.A. Federspiel A.P. Abola M.J. Proctor R.M. Myers J. Schmutz M. Dickson J. Grimwood D.R. Cox M.V. Olson R. Kaul N. Shimizu K. Kawasaki S. Minoshima G.A. Evans M. Athanasiou R. Schultz B.A. Roe F. Chen H. Pan J. Ramser H. Lehrach R. Reinhardt W.R. McCombie M. de la Bastide N. Dedhia H. Blocker K. Hornischer G. Nordsiek R. Agarwala L. Aravind J.A. Bailey A. Bateman S. Batzoglou E. Birney P. Bork D.G. Brown C.B. Burge L. Cerutti H.C. Chen D. Church M. Clamp R.R. Copley T. Doerks S.R. Eddy E.E. Eichler T.S. Furey J. Galagan J.G. Gilbert C. Harmon Y. Hayashizaki D. Haussler H. Hermjakob K. Hokamp W. Jang L.S. Johnson T.A. Jones S. Kasif A. Kaspryzk S. Kennedy W.J. Kent P. Kitts E.V. Koonin I. Korf D. Kulp D. Lancet T.M. Lowe A. McLysaght T. Mikkelsen J.V. Moran N. Mulder V.J. Pollara C.P. Ponting G. Schuler J. Schultz G. Slater A.F. Smit E. Stupka J. Szustakowki D. Thierry-Mieg L. Wagner J.

Wallis R. Wheeler A. Williams Y.I. Wolf K.H. Wolfe S.P. Yang R.F. Yeh F. Collins M.S. Guyer J. Peterson A. Felsenfeld K.A. Wetterstrand A. Patrinos and M.J. Morgan. 2001. Initial sequencing and analysis of the human genome. *Nature* **409:** 860-921.

Lercher, M.J., E.J.B. Williams, and L.D. Hurst. 2001. Local Similarity in Evolutionary Rates Extends over Whole Chromosomes in Human-Rodent and Mouse-Rat Comparisons: Implications for Understanding the Mechanistic Basis of the Male Mutation Bias. *Molecular Biology and Evolution* **18:** 2032-2039.

Li. 1997. Molecular Evolution. In *Molecular Evolution*.

Li, W.H., D.L. Ellsworth, J. Krushkal, B.H. Chang, and D. Hewett-Emmett. 1996. Rates of nucleotide substitution in primates and rodents and the generation-time effect hypothesis. *Mol Phylogenet Evol* **5:** 182-187.

Li, W.H., M. Tanimura, and P.M. Sharp. 1988. Rates and dates of divergence between AIDS virus nucleotide sequences. *Mol Biol Evol* **5:** 313-330.

Li, W.H. and C.I. Wu. 1987. Rates of nucleotide substitution are evidently higher in rodents than in man. *Mol.Biol.Evol.* **4:** 74-77.

Li, W.H., S.J. Yi, and K. Makova. 2002. Male-driven evolution. *Current Opinion in Genetics & Development* **12:** 650-656.

Lindblad-Toh, K. C.M. Wade T.S. Mikkelsen E.K. Karlsson D.B. Jaffe M. Kamal M. Clamp J.L. Chang E.J. Kulbokas, 3rd M.C. Zody E. Mauceli X. Xie M. Breen R.K. Wayne E.A. Ostrander C.P. Ponting F. Galibert D.R. Smith P.J. DeJong E. Kirkness P. Alvarez T. Biagi W. Brockman J. Butler C.W. Chin A. Cook J. Cuff M.J. Daly D. DeCaprio S. Gnerre M. Grabherr M. Kellis M. Kleber C. Bardeleben L. Goodstadt A. Heger C. Hitte L. Kim K.P. Koepfli H.G. Parker J.P. Pollinger S.M. Searle N.B. Sutter R. Thomas C. Webber J. Baldwin A. Abebe A. Abouelleil L. Aftuck M. Ait-Zahra T. Aldredge N. Allen P. An S. Anderson C. Antoine H. Arachchi A. Aslam L. Ayotte P. Bachantsang A. Barry T. Bayul M. Benamara A. Berlin D. Bessette B. Blitshteyn T. Bloom J. Blye L. Boguslavskiy C. Bonnet B. Boukhgalter A. Brown P. Cahill N. Calixte J. Camarata Y. Cheshatsang J. Chu M. Citroen A. Collymore P. Cooke T. Dawoe R. Daza K. Decktor S. DeGray N. Dhargay K. Dooley K. Dooley P. Dorje K. Dorjee L. Dorris N. Duffey A. Dupes O. Egbiremolen R. Elong J. Falk A. Farina S. Faro D. Ferguson P. Ferreira S. Fisher M. FitzGerald K. Foley C. Foley A. Franke D. Friedrich D. Gage M. Garber G. Gearin G. Giannoukos T. Goode A. Goyette J. Graham E. Grandbois K. Gyaltsen N. Hafez D. Hagopian B. Hagos J. Hall C. Healy R. Hegarty T. Honan A. Horn N. Houde L. Hughes L. Hunnicutt M. Husby B. Jester C. Jones A. Kamat B. Kanga C. Kells D. Khazanovich A.C. Kieu P. Kisner M. Kumar K. Lance T. Landers M. Lara W. Lee J.P. Leger N. Lennon L. Leuper S. LeVine J. Liu X. Liu Y. Lokyitsang T. Lokyitsang A. Lui J. Macdonald J. Major R. Marabella K. Maru C. Matthews S. McDonough T. Mehta J. Meldrim A. Melnikov L. Meneus A. Mihalev T. Mihova K. Miller R. Mittelman V. Mlenga L. Mulrain G. Munson A. Navidi J. Naylor T. Nguyen N. Nguyen C. Nguyen T. Nguyen R. Nicol N. Norbu C. Norbu N. Novod T. Nyima P. Olandt B. O'Neill K. O'Neill S. Osman L. Oyono C. Patti D. Perrin P. Phunkhang F. Pierre M. Priest A. Rachupka S. Raghuraman R. Rameau V. Ray C. Raymond F. Rege C. Rise J. Rogers P. Rogov J. Sahalie S. Settipalli T. Sharpe T. Shea M. Sheehan N. Sherpa J. Shi D. Shih J. Sloan C. Smith T. Sparrow J. Stalker N. Stange-Thomann S. Stavropoulos C. Stone S. Stone S. Sykes P. Tchuinga P. Tenzing S. Tesfaye D. Thoulutsang Y. Thoulutsang K. Topham I. Topping T. Tsamla H. Vassiliev V. Venkataraman A. Vo T. Wangchuk T. Wangdi M. Weiand J. Wilkinson A. Wilson S. Yadav S. Yang X. Yang G. Young Q. Yu J. Zainoun L. Zembek A. Zimmer and E.S. Lander. 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438:** 803-819.

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

Lu, J., W.H. Li, and C.I. Wu. 2003. Comment on " Chromosomal speciation and molecular divergence - Accelerated evolution in rearranged chromosomes". *Science* **302:** 988.

Lynch, M. and J.S. Conery. 2000b. The evolutionary fate and consequences of duplicate genes. *Science* **290:** 1151-1155.

Machado, C.A., R.M. Kliman, J.A. Markert, and J. Hey. 2002. Inferring the history of speciation from multilocus DNA sequence data: The case of Drosophila pseudoobscura and close relatives. *Molecular Biology and Evolution* **19:** 472-488.

Makalowski, W. and M.S. Boguski. 1998. Evolutionary parameters of the transcribed mammalian genome: an analysis of 2,820 orthologous rodent and human sequences. *Proc Natl Acad Sci U S A* **95:** 9407-9412.

Makova, K.D. and W.H. Li. 2002. Strong male-driven evolution of DNA sequences in humans and apes. *Nature* **416:** 624-626.

Marais, G. 2003. Biased gene conversion: implications for genome and sex evolution. *Trends Genet* **19:** 330-338.

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

Marques-Bonet, T., Z. Cheng, X. She, E.E. Eichler, and A. Navarro. 2006a. The genomic distribution of intra and inter divergence of human SD. *Genomics (Submitted)*.

Marques-Bonet, T. and A. Navarro. 2005. Chromosomal rearrangements are associated with higher rates of molecular evolution in mammals. *Gene* **353:** 147-154.

Marques-Bonet, T., J. Sànchez-Ruiz, L. Armengol, R. Khaja, J. Bertranpetit, M. Rocchi, E. Gazave, and A. Navarro. 2006b. On the association between chromosomal rearrangements and genic evolution in humans and chimpanzees. *Genome Biology (Submitted)*.

Martin, A.P. and S.R. Palumbi. 1993. Body Size, Metabolic-Rate, Generation Time, and the Molecular Clock. *Proceedings Of the National Academy Of Sciences Of the United States Of America* **90:** 4087-4091.

Matassi, G., P.M. Sharp, and C. Gautier. 1999. Chromosomal location effects on gene sequence evolution in mammals. *Current Biology* **9:** 786-791.

Messier, W. and C.B. Stewart. 1997. Episodic adaptive evolution of primate lysozymes. *Nature* **385:** 151-154.

Mikkelsen, T.S., L.W. Hillier, E.E. Eichler, M.C. Zody, D.B. Jaffe, S.P. Yang, W. Enard, I. Hellmann, K. Lindblad-Toh, T.K. Altheide, N. Archidiacono, P. Bork, J. Butler, J.L. Chang, Z. Cheng, A.T. Chinwalla, P. deJong, K.D. Delehaunty, C.C. Fronick, L.L. Fulton, Y. Gilad, G. Glusman, S. Gnerre, T.A. Graves, T. Hayakawa, K.E. Hayden, X.Q. Huang, H.K. Ji, W.J. Kent, M.C. King, E.J. Kulbokas, M.K. Lee, G. Liu, C. Lopez-Otin, K.D. Makova, O. Man, E.R. Mardis, E. Mauceli, T.L. Miner, W.E. Nash, J.O. Nelson, S. Paabo, N.J. Patterson, C.S. Pohl, K.S. Pollard, K. Prufer, X.S. Puente, D. Reich, M. Rocchi, K. Rosenbloom, M. Ruvolo, D.J. Richter, S.F. Schaffner, A.F.A. Smit, S.M. Smith, M. Suyama, J. Taylor, D. Torrents, E. Tuzun, A. Varki, G. Velasco, M. Ventura, J.W. Wallis, M.C. Wendl, R.K. Wilson, E.S. Lander, R.H. Waterston, and C.S.A. Consortium. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437:** 69-87.

Muller, S. and J. Wienberg. 2001. "Bar-coding" primate chromosomes: molecular cytogenetic screening for the ancestral hominoid karyotype. *Human Genetics* **109:** 85-94.

Murphy, W.J., D.M. Larkin, A. Everts-van der Wind, G. Bourque, G. Tesler, L. Auvil, J.E. Beever, B.P. Chowdhary, F. Galibert, L. Gatzke, C. Hitte, S.N. Meyers, D. Milan, E.A. Ostrander, G. Pape, H.G. Parker, T. Raudsepp, M.B. Rogatcheva, L.B. Schook, L.C. Skow, M. Welge, J.E. Womack, J. O'Brien S, P.A. Pevzner, and H.A. Lewin. 2005. Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. *Science* **309:** 613-617.

Myers, S., L. Bottolo, C. Freeman, G. McVean, and P. Donnelly. 2005. A fine-scale map of recombination rates and hotspots across the human genome. *Science* **310:** 321-324.

Navarro, A. and N.H. Barton. 2003a. Accumulating postzygotic isolation genes in parapatry: A new twist on chromosomal speciation. *Evolution* **57:** 447-459.

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

Navarro, A., T. Marques-Bonet, and N.H. Barton. 2003. Response to comment on " chromosomal speciation and molecular divergence - Accelerated evolution in rearranged chromosomes". *Science* **302:** 988.

Nekrutenko, A. and W.H. Li. 2000. Assessment of Compositional Heterogeneity Within and Between Eukaryotic Genomes. *Genome Research* **10:** 1986-1995.

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

Noor, M.A.F., K.L. Grams, L.A. Bertucci, Y. Almendarez, J. Reiland, and K.R. Smith. 2001a. The genetics of reproductive isolation and the potential for gene exchange between *Drosophila pseudoobscura* and *D. persimilis* via backcross hybrid males. *Evolution* **55:** 512-521.

Noor, M.A.F., K.L. Grams, L.A. Bertucci, and J. Reiland. 2001b. Chromosomal inversions and the reproductive isolation of species. *Proceedings of the National Academy of Sciences of the United States of America* **98:** 12084-12088.

O'Brien, S.J., M. Menotti-Raymond, W.J. Murphy, W.G. Nash, J. Wienberg, R. Stanyon, N.G. Copeland, N.A. Jenkins, J.E. Womack, and J.A.M. Graves. 1999. The Promise of Comparative Genomics in Mammals. *Science***:** 458-462.

Ohno, S. 1970. *Evolution by gene duplication*. Springer-Verlag.

Ohta, T. 1984. Some models of gene conversion for treating the evolution of multigene families. *Genetics* **106:** 517-528.

Ohta, T. 1987. Very slightly deleterious mutations and the molecular clock. *J Mol Evol* **26:** 1-6.

Ohta, T. 1993. Amino-Acid Substitution at the Adh Locus of Drosophila Is Facilitated by Small Population-Size. *Proceedings of the National Academy of Sciences of the United States of America* **90:** 4548-4551.

Orr, H.A. 1997. Dobzhansky, Bateson and the genetics of speciation. *Genetics* **144:** 1331-1335.

Osada, N. and C.I. Wu. 2005. Inferring the mode of speciation from genomic data: a study of the great apes. *Genetics* **169:** 259-264.

Ota, T. 1971. Evolutionary rate of cistrons and DNA divergence. *J Mol Evol* **1:** 150-157.

Paabo, S. 2003. The mosaic that is our genome. *Nature* **421:** 409-412.

Patterson, N., D.J. Richter, S. Gnerre, E.S. Lander, and D. Reich. 2006. Genetic evidence for complex speciation of humans and chimpanzees. *Nature* **441:** 1103-1108.

Patton, J.L. and S.W. Sherwood. 1983. Chromosome evolution and speciation in rodents. *Ann.Rev.Syst. & Ecol.* **14:** 139-158.
Ptak, S.E., D.A. Hinds, K. Koehler, B. Nickel, N. Patil, D.G. Ballinger, M. Przeworski, K.A. Frazer, and S. Paabo. 2005. Fine-scale recombination patterns differ between chimpanzees and humans (vol 37, pg 429, 2005). *Nature Genetics* **37:** 445-445.
Puig, M., M. Caceres, and A. Ruiz. 2004. Silencing of a gene adjacent to the breakpoint of a widespread Drosophila inversion by a transposon-induced antisense RNA. *Proc Natl Acad Sci U S A* **101:** 9013-9018.
Rieseberg, L.H. 2001. Chromosomal rearrangements and speciation. *Trends in Ecology & Evolution* **16:** 351-358.
Rieseberg, L.H. and K. Livingstone. 2003. Evolution. Chromosomal speciation in primates. *Science* **300:** 267-268.
Rieseberg, L.H., C. Vanfossen, and A.M. Desrochers. 1995. Hybrid Speciation Accompanied by Genomic Reorganization in Wild Sunflowers. *Nature* **375:** 313-316.
Rieseberg, L.H., J. Whitton, and K. Gardner. 1999. Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics* **152:** 713-727.
Rifkin, S.A., J. Kim, and K.P. White. 2003. Evolution of gene expression in the Drosophila melanogaster subgroup. *Nat Genet* **33:** 138-144.
Rudd, M.K. and H.F. Willard. 2004. Analysis of the centromeric regions of the human genome assembly. *Trends in Genetics* **20:** 529-533.
Ruiz-Herrera, A., F. Garcia, C. Azzalin, E. Giulotto, J. Egozcue, M. Ponsa, and M. Garcia. 2002a. Distribution of intrachromosomal telomeric sequences (ITS) on Macaca fascicularis (Primates) chromosomes and their implication for chromosome evolution. *Hum Genet* **110:** 578-586.
Ruiz-Herrera, A., M. Ponsa, F. Garcia, J. Egozcue, and M. Garcia. 2002b. Fragile sites in human and Macaca fascicularis chromosomes are breakpoints in chromosome evolution. *Chromosome Res* **10:** 33-44.
Schmitz, J. and R.F. Moritz. 1998. Sociality and the rate of rDNA sequence evolution in wasps (Vespidae) and honeybees (Apis). *J Mol Evol* **47:** 606-612.
Searle, J.B. 1998. Speciation, chromosomes, and genomes. *Genome Res* **8:** 1-3.
Seuanez, H.N., H.J. Evans, D.E. Martin, and J. Fletcher. 1979. Inversion of Chromosome-2 That Distinguishes between Bornean and Sumatran Orangutans. *Cytogenetics and Cell Genetics* **23:** 137-140.
Sharp, A.J., Z. Cheng, and E.E. Eichler. 2006. Structural Variation of the Human Genome. *Annu Rev Genomics Hum Genet*.
Sharp, P.M. 1997. In search of molecular darwinism. *Nature* **385:** 111-112.
She, X.W., J.E. Horvath, Z.S. Jiang, G. Liu, T.S. Furey, L. Christ, R. Clark, T. Graves, C.L. Gulden, C. Alkan, J.A. Bailey, C. Sahinalp, M. Rocchi, D. Haussler, R.K. Wilson, W. Miller, S. Schwartz, and E.E. Eichler. 2004a. The structure and evolution of centromeric transition regions within the human genome. *Nature* **430:** 857-864.
She, X.W., Z.X. Jiang, R.L. Clark, G. Liu, Z. Cheng, E. Tuzun, D.M. Church, G. Sutton, A.L. Halpern, and E.E. Eichler. 2004b. Shotgun sequence assembly and recent segmental duplications within the human genome. *Nature* **431:** 927-930.
She, X.W., G. Liu, M. Ventura, S. Zhao, D. Misceo, R. Roberto, M.F. Cardone, M. Rocchi, E.D. Green, N. Archidiacano, and E.E. Eichler. 2006. A preliminary comparative analysis of primate segmental duplications shows elevated substitution rates and a great-ape expansion of intrachromosomal duplications. *Genome Research* **16:** 576-583.

Spirito, F. 2000. The role of chromosomal rearrangements in speciation. In *Endless forms* (eds. D.J. Howard and S.H. Berlocher), pp. 320-329. Oxford, Oxford University Press.
Spitz, F., F. Gonzalez, and D. Duboule. 2003. A global control region defines a chromosomal regulatory landscape containing the HoxD cluster. *Cell* **113:** 405-417.
Springer, M.S., W.J. Murphy, E. Eizirik, and S.J. O'Brien. 2003. Placental mammal diversification and the Cretaceous-Tertiary boundary. *Proc Natl Acad Sci U S A* **100:** 1056-1061.
Subramanian, S. and S. Kumar. 2003. Neutral substitutions occur at a faster rate in exons than in noncoding DNA in primate genomes. *Genome Res* **13:** 838-844.
Szamalek, J.M., D.N. Cooper, W. Schempp, P. Minich, M. Kohn, J. Hoegel, V. Goidts, H. Hameister, and H. Kehrer-Sawatzki. 2006a. Polymorphic micro-inversions contribute to the genomic variability of humans and chimpanzees. *Hum Genet* **119:** 103-112.
Szamalek, J.M., V. Goidts, N. Chuzhanova, H. Hameister, D.N. Cooper, and H. Kehrer-Sawatzki. 2005. Molecular characterisation of the pericentric inversion that distinguishes human chromosome 5 from the homologous chimpanzee chromosome. *Human Genetics* **117:** 168-176.
Szamalek, J.M., V. Goidts, J.B. Searle, D.N. Cooper, H. Hameister, and H. Kehrer-Sawatzki. 2006b. The chimpanzee-specific pericentric inversions that distinguish humans and chimpanzees have identical breakpoints in Pan troglodytes and Pan paniscus. *Genomics* **87:** 39-45.
Tanimoto, K., Q. Liu, J. Bungert, and J.D. Engel. 1999. Effects of altered gene order or orientation of the locus control region on human-globin gene expression in mice. *Nature***:** 344-347.
Taudien, S., I. Ebersberger, G. Glockner, and M. Platzer. 2006. Should the draft chimpanzee sequence be finished? *Trends Genet* **22:** 122-125.
Taylor, J.S. and J. Raes. 2004. Duplication and divergence: the evolution of new genes and old ideas. *Annu Rev Genet* **38:** 615-643.
Vallender, E.J. and B.T. Lahn. 2004. Effects of chromosomal rearrangements on human-chimpanzee molecular evolution. *Genomics* **84:** 757-761.
Wang, R.L., J. Wakeley, and J. Hey. 1997. Gene flow and natural selection in the origin of *Drosophila pseudoobscura* and close relatives. *Genetics* **147:** 1091-1106.
Webber, C. and C.P. Ponting. 2005. Hotspots of mutation and breakage in dog and human chromosomes. *Genome Res* **15:** 1787-1797.
Weinreich, D.M. 2001. The rates of molecular evolution in rodent and primate mitochondrial DNA. *J Mol Evol* **52:** 40-50.
White, M.J.D. 1968. Models of speciation. *Science* **158:** 1065-1070.
White, M.J.D. 1978. Chain processes in chromosomal speciation. *Syst.Zool.* **27:** 285-298.
Whitehead, A. and D.L. Crawford. 2006. Variation within and among species in gene expression: raw material for evolution. *Mol Ecol* **15:** 1197-1211.
Williams, E.J. and L.D. Hurst. 2000. The proteins of linked genes evolve at similar rates. *Nature* **407:** 900-903.
Williams, E.J.B. and L.D. Hurst. 2002. Is the Synonymous Substitution Rate in Mammals Gene-Specific? *Molecular Biology and Evolution* **19:** 1395-1398.
Wolfe, K.H. and W.H. Li. 2003. Molecular evolution meets the genomics revolution. *Nat Genet* **33 Suppl:** 255-265.
Wolfe, K.H. and P.M. Sharp. 1993. Mammalian Gene Evolution - Nucleotide-Sequence Divergence between Mouse and Rat. *Journal of Molecular Evolution* **37:** 441-456.

Wolfe, K.H., P.M. Sharp, and W.H. Li. 1989. Mutation-Rates Differ among Regions of the Mammalian Genome. *Nature* **337:** 283-285.
Woolfit, M. and L. Bromham. 2003. Increased rates of sequence evolution in endosymbiotic bacteria and fungi with small effective population sizes. *Mol Biol Evol* **20:** 1545-1555.
Wu, C. 2001. The genic view of the process of speciation. *J. Evol. Biol.* **14:** 851-865.
Wu, C.I. and W.H. Li. 1985. Evidence for higher rates of nucleotide substitution in rodents than in man. *Proc.Nat.Acad.Sci.(USA)* **82:** 1741-1745.
Wu, C.I. and C.T. Ting. 2004. Genes and speciation. *Nature Reviews Genetics* **5:** 114-122.
Yang, Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol Biol Evol* **15:** 568-573.
Yang, Z. and R. Nielsen. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol Biol Evol* **19:** 908-917.
Yang, Z.H. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer Applications in the Biosciences* **13:** 555-556.
Yang, Z.H., R. Nielsen, N. Goldman, and A.M.K. Pedersen. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155:** 431-449.
Yi, S., D.L. Ellsworth, and W.H. Li. 2002. Slow Molecular Clocks in Old World Monkeys, Apes, and Humans. *Molecular Biology and Evolution* **19:** 2191-2198.
Yunis, J.J. and O. Prakash. 1982. The Origin of Man - a Chromosomal Pictorial Legacy. *Science* **215:** 1525-1530.
Zhang, J., X. Wang, and O. Podlaha. 2004. Testing the chromosomal speciation hypothesis for humans and chimpanzees. *Genome Research* **14:** 845-851.
Zhang, P., Z.L. Gu, and W.H. Li. 2003. Different evolutionary patterns between young duplicate genes in the human genome. *Genome Biology* **4:** R56.
Zhao, S., J. Shetty, L. Hou, A. Delcher, B. Zhu, K. Osoegawa, P. de Jong, W.C. Nierman, R.L. Strausberg, and C.M. Fraser. 2004. Human, mouse, and rat genome large-scale rearrangements: stability versus speciation. *Genome Res* **14:** 1851-1860.
Zuckerlandl, E.a.P., L. . 1962. Molecular disease, evolution, and genetic heterogeneity. *Horizons in Biochemistry***:** pp. 189–225.