Available online at www.sciencedirect.com
SCIENCE DIRECT®
Gene 317 (2003) 67–77
GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES
www.elsevier.com/locate/gene

High mutation rates in human and ape pseudoautosomal genes

Dmitry A. Filatov*, Dave T. Gerrard

School of Biosciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK

Received 30 July 2002; received in revised form 19 November 2002; accepted 12 May 2003

Abstract

It has been suggested that recombination may be mutagenic, which, if true, would inflate intraspecies diversity and interspecies silent divergence in regions of high recombination. Here, we test this hypothesis comparing human/orangutan genome-wide non-coding divergence (K) to that in the pseudoautosomal genes which were reported to recombine much more frequently than the rest of the genome. We demonstrate that, compared to the average human/orangutan non-coding divergence (K=3%), the substitution rate is significantly elevated in the introns of *SHOX* (K=5.7%), *PPP2R3L* (K=8.7%) and *ASMT* (K=6.5%) genes located in the human and orangutan Xp/Yp pseudoautosomal region (p-PAR), where recombination is over 20-fold higher than the genomic average. On the other hand, human/orangutan non-coding divergence at the Xp/Yp pseudoautosomal boundary (K=3.5%) and in the *SYBL1* gene (K=2.7%), located in the human Xq/Yq pseudoautosomal region (q-PAR), where recombination is known to be less frequent than in p-PAR, was not significantly higher than the genome average. The data are consistent with the hypothesis that recombination may be mutagenic.

Keywords: Pseudoautosomal region; Human; Orangutan; Substitution rate; Mutation rate

1. Introduction

There is experimental evidence that, at least in yeast, recombination may cause a substantial increase in the mutation rate (Strathern et al., 1995; Rattray et al., 2001). If recombination is mutagenic by itself, regions of higher recombination might have a higher mutation rate, resulting in elevated intraspecies diversity and between species divergence in such regions. Drosophila and human single nucleotide polymorphism (SNP) variability was shown to positively correlate with recombination (Aquadro et al., 2001; Nachman, 2001), but this correlation is thought to be due to background selection (Charlesworth et al., 1993) and selective sweeps (Rice, 1987), which reduce diversity in regions of low recombination. Recently, Lercher and Hurst (2002) found a weak (r^2<0.1), but significant correlation between human SNP diversity and recombination rate throughout the genome. As less than 5% of these SNPs are in the proximity of the coding regions, it seems unlikely that background selection and selective sweeps may cause such a correlation, suggesting a mutagenic role for recombination (Lercher and Hurst, 2002).

The silent substitution rate is proportional to the mutation rate (Kimura, 1983) and may be used as a proxy in the comparison of mutation rates in different regions. Silent divergence between human and mouse orthologous genes was shown to correlate with recombination rate, suggesting that recombination may elevate mutation rates (Lercher and Hurst, 2002). In humans, the greatest density of recombination events yet observed for the human genome was in the pseudoautosomal region (PAR) on the p-arms of the human sex chromosomes (Lien et al., 2000), far above even the 'jungles' of high recombination elsewhere in the genome (Yu et al., 2001a). If recombination is mutagenic, then there is a good chance that this will be detected in the PAR. Perry and Ashworth (1999) demonstrated that silent divergence between the pseudoautosomal part of the mouse gene *Fxy*, and its rat homologue is much higher than for the non-pseudoautosomal part of this gene, suggesting a higher mutation rate in the highly recombining pseudoautosomal portion. It would be interesting to know whether a high mutation rate is a property of PARs in other mammals, especially in humans. Elevated genetic diversity was indeed reported for the human *PPP2R3L* gene located in the PAR

Abbreviations: PAR, pseudo autosomal region; SNP, single-nucleotide polymorphisms; HKA; Hudson; Kreitman; Aguade test.

* Corresponding author. Tel.: +44-121-4142500; fax: +44-121-4145925.
E-mail address: d.filatov@bham.ac.uk (D.A. Filatov).

0378-1119/$ - see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/S0378-1119(03)00697-8

(Schiebel et al., 2000). On the other hand, diversity in another human pseudoautosomal gene, *SHOX*, is not higher than elsewhere in the genome (May et al., 2002). A non-coding pseudoautosomal region close to the Xp/Yp telomere was reported to have a high substitution rate (Cooke et al., 1985; Baird and Royle, 1997), however, subtelomeric regions are known to evolve very fast, perhaps, due to ectopic recombination with homologous sequences in the subtelomeric regions of the other chromosomes (Mefford and Trask, 2002).

Human X and Y chromosomes pair and recombine in two small pseudoautosomal regions (PARs) at both ends of the sex chromosomes (Cooke et al., 1985; Freije et al., 1992). The short arms (Xp/Yp) of the sex chromosomes contain the larger p-PAR, which is about 2.6 Mb in size (Brown, 1988; Petit et al., 1988). Chiasmata between X and Y chromosomes in the p-PAR are essential for the correct segregation of the sex chromosomes in male meiosis (Burgoyne et al., 1992), resulting in a high recombination frequency in this region. High resolution sperm typing demonstrated that the recombination rate in the p-PAR is greater than 20 times the genomic average of ~1 cM/Mb (Lien et al., 2000). The long (Xq/Yq) arms contain a much smaller q-PAR, which is about 0.4 Mb long (Ciccodicola et al., 2000). Though the p-PAR was demonstrated to be the same in humans, great apes and old world monkeys (Ellis et al., 1990), the q-PAR exists only in humans (this region is X-linked in apes) and it started to recombine with the Y chromosome very recently, after the human/chimpanzee split, probably due to a translocation of the distal part of the Xq arm to the Yq (D'Esposito et al., 1997; Ciccodicola et al., 2000). The recombination rate in the q-PAR is lower than in the p-PAR, but it is still about six times more frequent than on the rest of the X chromosome (Ciccodicola et al., 2000).

Here, we report the estimates of human/ape silent divergence for three p-PAR genes, *SHOX* (Rao et al., 1997; Belin et al., 1998), *PPP2R3L* (Schiebel et al., 2000) and *ASMT* (Rodriguez et al., 1994). Also, we measure the human/ape non-coding divergence for the Xp/Yp pseudoautosomal boundary (PAB) region (Ellis et al., 1990), and one q-PAR gene, *SYBL1* (Matarazzo et al., 1999). We compare these estimates to the genome wide human/ape divergence values (e.g. Chen and Li, 2001) and demonstrate that all three p-PAR genes, located in the region with high frequency of recombination, accumulate substitutions significantly faster than the autosomal and X-linked regions. We also report an unusually high mouse/rat divergence in the mouse pseudoautosomal *Sts* gene, which is consistent with an earlier observation of the elevated rate of silent substitutions in the pseudoautosomal portion of the mouse *Fxy* gene (Perry and Ashworth, 1999). These observations suggest that mammalian pseudoautosomal regions mutate with unusually high frequency, perhaps due to the elevated recombination rate in these regions. The involvement of recombination is also suggested by the fact that the *SYBL1* gene, located in the human q-PAR and undergoing less recombination than the p-PAR genes, does not diverge significantly faster than the autosomal genes.

2. Materials and methods

2.1. Orangutan sequences

A sample of orangutan blood was kindly provided by the Zoological Institute (London). Orangutan genomic DNA was extracted from 0.25 ml of blood using the Qiagen Blood DNA extraction kit. Using human genomic sequence, we designed the primers for PCR amplification and sequencing of orangutan homologues (Table 1). Three fragments (940, 946 and 1028 bp long) of the orangutan *SHOX* gene were amplified using primer pairs SHOX+1 and SHOX−12, SHOX+15 and SHOX−17 and SHOX+7 and SHOX−9 (hereafter "SHOX+1−12", "SHOX+15−17" and "SHOX+7−9", respectively). A 1.3-kb fragment of the orangutan *PPP2R3L* gene was amplified with a primer pair PPP+1 and PPP-5. A 0.95-kb fragment of the orangutan *ASMT* gene was amplified with ASMT+6 and ASMT-7 primers. A 2-kb region around the pseudoautosomal bound-

Table 1
PCR and sequencing primers

Gene	Name	Sequence
ASMT	ASMT+6	TGGGCGTGTTTGACCTTCTC
ASMT	ASMT−7	TGACCGTGGTCAGGTAGTCG
ASMT	ASMT+30	GGTTGCAGTGAGCCGAGATCG
PAB	PAB+1	ACCGTGTCCAGCCTCTGGTATAC
PAB	PAB−2	GCCTTTGTATAACAGCACTGGC
PAB	PAB−4X	TCAGCGTGACTATCGACCTTGC
PAB	PAB−5X	GGGAGAAGGCCTGGAAATTAT
PPP2R3L	PPP+1	GAAGGGGCCGCTCTTCTATG
PPP2R3L	PPP−2	CAAGAAGGGGACAAAGTCCTCC
PPP2R3L	PPP−3	TGTAGCGCGAGTGGAACTCG
PPP2R3L	PPP+4	AATCCTCCAGAACTGCCACG
PPP2R3L	PPP−5	AGAGACCCCCAGAGAGCCTC
PPP2R3L	PPP+6	CTACACCACGCGCCTCGTGTC
SHOX	SHOX+1	AGCTCACGGCTTTTGTATCC
SHOX	SHOX−2	ATTGTCTACGTGGTCCTTGAAC
SHOX	SHOX−5	CCCAGGCGCTGGCTGAGCTC
SHOX	SHOX+6	GAGTGGACCCGACCGGAGAC
SHOX	SHOX+7	AGAACCGGAGAGCCAAGTGC
SHOX	SHOX−9	CCTTCTTAACCAGGCAGCAAG
SHOX	SHOX−12	ACACATCCTAAGCCGTCAGG
SHOX	SHOX+13	GCGCGAGGACGTGAAGTCGG
SHOX	SHOX+15	TGGGGATAGCGTCTCTCCGTAG
SHOX	SHOX−17	CAAACGCAATGAACCCATCC
SHOX	SHOX+18	ACCTGCACAAAGAACCTGCTC
SHOX	SHOX−19	TCATGCCATTATACTCCAGCC
XG	XG+1	CCCACCAAGAAGCCAAACTCAG
XG	XG−2	CAACATACCGTGTTAGCCAGGA
XG	XG−3	TTCACAGTGTTGTGCAGCCATC
XG	XG+4	AGGTGAATCCCTAGAACCAGAAG
SYBL1	SYBL1+5	GATTTTGAACGTTCCCGAGC
SYBL1	SYBL1−6	GGACGCAGAGACACTGGAAC
SYBL1	SYBL1−7	TGTTTCTGACCATGATTCCTTTC
SYBL1	SYBL1+8	TTTGTGTCTTGCTTGTTTCAGTC

ary was amplified with the pseudoautosomal PAB+1 and X-specific PAB-4X primers. A 1.3-kb long intron of the *XG* gene, located 6.5 kb distal to the Xp/Yp pseudoautosomal boundary was amplified with primers XG+1 and XG2. A 2.3-kb long intron of the orangutan *SYBL1* gene was amplified using primers SYBL1+5 and SYBL1−7. All the regions were PCR amplified using Roche High Fidelity PCR kit with the following conditions: 95 °C, 2.5 min, 58 °C, 1 min, 68 °C 3 min, followed by 34 cycles of 94 °C 0.5 min, *X* °C, 0.5 min and 68 °C 2.5 min (*X* is primer-specific annealing temperature). The PCR products were gel-purified, extracted from the agarose gel using the Qiagen Gel Extraction kit, cloned into pCR plasmid using the TA cloning kit (Invitrogen) and sequenced using the BigDye v3 sequencing system (ABI) on the ABI3700 automated sequencer. For sequencing, the primers listed in Table 1 were used. GenBank accession numbers for the orangutan PAR sequences: AY181053–AY181060.

All the newly sequenced orangutan sequences were aligned with human homologues using ProSeq software (Filatov, 2002), and the regions annotated as coding in the human homologues were removed. Pairwise human/orangutan divergence for non-coding regions was calculated as a proportion of differences per nucleotide, or as a Kimura's distance measure (Kimura, 1980), using ProSeq software (Filatov, 2002). The probability to observe the given or larger number of substitutions between human and orangutan homologues by chance was calculated from Poisson distribution formula (Sokal and Rohlf, 1995).

For the PAR sequences, the GC content, the proportion of CG-pairs, the number of (A or T)→(C or G) and (C or G)→(T or A) mutations (GC% stationarity test, Eyre-Walker, 1994), mutation spectra, transition/transversion ratios and the number of mutations in CpG pairs in the PARs were calculated using ProSeq software (Filatov, 2002). The role of methylation, which causes frequent C→T transitions in CG dinucleotides (Robertson and Wolffe, 2000), was studied as described in Filatov and Charlesworth (2002).

2.2. Building a distribution of human-orangutan divergence

A list of all GenBank (Benson et al., 2002) entries from the genus Pongo (both orangutan sub-species) was obtained using Entrez at the NCBI web-site. All against all comparisons were then made using FASTA (Wisconsin Package, GCG) and all duplicate sequences (100% identity) were removed to leave 450 sequences. Each of these sequences was then BLAST-searched against GenBank. Where the same piece of human sequence was matched by two or more orangutan sequences, the longest orangutan sequence was kept. This resulted in a set of 204 orangutan sequences with putative human homologues. As the previous BLAST search could have left us with very similar orangutan sequences matching the same human sequence but under a different accession, we then BLAST-searched these sequences against the non-redundant set of human genome sequences (RefSeq). This process automatically removed sequences containing human repeats (45 sequences). After again filtering for orangutan sequences sharing the same best match, we were left with 91 good homologous pairs. Of these, 51 sequences had non-coding regions over 100 nucleotides in length. Human and orangutan homologous sequences were aligned using ProSeq (Filatov, 2002) and checked by eye. All the coding sequences were excluded from further analysis, resulting in 74130 non-coding sites from 51 different orangutan/human alignments. The pairwise human/orangutan divergence was measured as a proportion of differences per site between the two species.

2.3. Mouse-rat PAR divergence

Similarly, we also looked at the PAR divergence between mouse and rat using the mouse PAR gene, *Sts* (GenBank accession numbers U37545, U37138 and M16505 for the mouse and rat *Sts*, and human *STS* genes, respectively), and the portion of *Fxy* gene pseudoautosomal in *Mus musculus* (accession numbers: AF026565, AF186460 and AF186461 for *M. musculus*, *M. spretus* and *Homo sapiens* homologues). As non-coding regions were unalignable, only coding regions were analysed. The translated coding regions were aligned using CLUSTAL W (Thompson et al., 1994) and the protein alignments were used to align the nucleotide sequences by eye in ProSeq. The synonymous divergence (K_s) was calculated using MEGA (Kumar et al., 1994) using the method of Li et al. (1985). The data of Wolfe and Sharp (1993) were used to build the distribution of mouse/rat K_s values across the genome.

To count the number of mutations of different types in the mouse lineage, we reconstructed the sequence ancestral to mouse and rat *Sts* genes using the human sequence as an outgroup. Similarly, we reconstructed the sequence of the *Fxy* gene in the ancestor of the *M. musculus* and *M. spretus*, using the rat sequence as an outgroup. The reconstruction of ancestral sequences was conducted by the baseml program from the PAML package (Yang, 2001). The actual and the reconstructed ancestral sequences were imported into ProSeq and the number of mutations of different types were analysed using the "Nucleotide mutation patterns" tool in ProSeq.

3. Results

3.1. Human/orangutan divergence

To compare mutation rates in the frequently recombining human p-arm pseudoautosomal region to those throughout the genome, we sequenced three fragments of the orangutan *SHOX* gene and one fragment from each of the *PPP2R3L* and *ASMT* orangutan genes from the p-PAR, totalling 5.8 kb of intron sequences (Table 2). We also sequenced two intron regions (3.3 kb in total) from the pseudoautosomal portion

Table 2
Mutation rates and patterns in p-PAR and q-PAR intron regions

Gene (fragment)	Location	Distance from p-telomere	Recombination ~cM/Mb[a]	Length analysed	Mutations observed[b]	Divergence[c] K±SD	Mutation patterns				GC%
							Ts	Tv	Ts/Tv	CpG	
PPP2R3LB	p-PAR	150 kb[d]	15[e]	854	72***	0.090±0.011	55	17	3.4	30	70.4
SHOX (+1–12)	p-PAR	500 kb[f]	27[e]	681	37***	0.057±0.009	18	19	0.95	8	63.3
SHOX (+15–17)	p-PAR	500 kb[f]	200–350[g]	600	25*	0.043±0.009	13	12	1.1	3	59
SHOX (+7–9)	p-PAR	500 kb[f]	27[e]	877	58**	0.070±0.009	30	28	1.1	11	55.1
SHOX (all 3)	p-PAR	500 kb[f]	27[e]	2157	120***	0.058±0.005	61	59	1	22	58.8
ASMT	p-PAR	1000 kb[h]	36[e]	722	51***	0.065±0.009	35	16	2.19	19	55.5
XG (+1–2)	p-PAR	2.6 Mb[i]	26[e]	956	30	0.033±0.006	19	11	1.7	4	39
XG (PABxy)	p-PAR	2.6 Mb[i]	26[e]	1534	51	0.035±0.005	29	22	1.32	11	43.2
XG (PABx)	X-linked	2.6 Mb[i]	0	507	14	0.028±0.008	10	4	2.5	1	39.6
SYBL1	q-PAR	151.7 Mb[j]	6[k]	2200	60	0.027±0.004	44	16	2.75	8	35.2

[a] Recombination in human male meiosis.
[b] Significantly more than expected from Poisson distribution assuming 3% average divergence (***) $P<0.0001$, (**) $P<0.001$, (*) $P<0.05$.
[c] Kimura's distance (Kimura, 1980).
[d] Schiebel et al., 2000
[e] Lien et al., 2000.
[f] Rao et al., 1997.
[g] May et al. (2002) described this region as a recombination hot spot.
[h] Rodriguez et al. (1994) and also from NCBI map of the X chromosome, build 29.
[i] Rappold, 1993.
[j] NCBI map of the X chromosome, build 29.
[k] Ciccodicola et al., 2000.

of the *XG* gene. One of these *XG* introns contains the Xp/Yp pseudoautosomal boundary region (PAB) and another is located about 6 kb distally to the PAB. Finally, a 2.2-kb long intron of the orangutan X-linked *SYBL1* gene was sequenced, which is located in the human q-PAR (Ciccodicola et al., 2000).

Human/orangutan intron divergence in the *SHOX*, *PPP2R3L* and *ASMT* genes (Table 2), was significantly higher than the average 2.5–3% human/orangutan silent divergence reported before (Li, 1997; Chen and Li, 2001). Divergence in the pseudoautosomal *XG* introns was slightly, but not significantly higher than the divergence in the X-linked portion of the *XG* intron. Divergence in the q-PAR gene, *SYBL1*, was 2.8%, close to the human/orangutan divergence value reported before (Li, 1997; Chen and Li, 2001).

To obtain a better estimate of the genome wide human/orangutan silent divergence and to study the correlation of divergence with recombination rate (see Fig. 2 below), we downloaded all the orangutan sequences available in GenBank, removed all the redundant and repetitive sequences and aligned with human homologues, as described in Materials and methods section. A total of 74130 non-coding sites from 51 different orangutan X-linked and autosomal sequences were aligned with their human homologues. There were 2116 differences between the human and orangutan sequences, giving an average divergence of 2.85±0.061%. The distribution of human/orangutan divergence values is shown in Fig. 1A. *SHOX*, *PPP2R3L* and *ASMT* genes, are located in the right tail of the distribution. The probability that these genes show such high divergence by chance is negligible (Table 2).

3.2. Mouse/rat divergence

The high substitution rate in the human and ape p-PAR is in agreement with the elevated substitution rate in the pseudoautosomal part of *Fxy* (Perry and Ashworth, 1999), the gene spanning the PAR boundary of *Mus musculus* (Palmer et al., 1997). To investigate whether this effect is just a peculiarity of the *Fxy* gene, or it is the property of the entire rodent PAR, we estimated divergence between mouse and rat *Sts* genes, the only other known mouse pseudoautosomal gene (Salido et al., 1996). As mouse and rat *Sts* genes are unalignable in introns, we used synonymous divergence, K_s, for this comparison. The mouse/rat divergence at silent sites in the *Sts* gene is 81.4%, which is much higher than the average mouse/rat silent divergence in X-linked (14.63±0.85%) and autosomal (23±0.44%) genes (McVean and Hurst, 1997).

3.3. Substitution patterns in the pseudoautosomal genes

In total, 324 substitutions (199 transitions and 125 transversions) in non-coding regions between human and orangutan p-PAR genes were observed (Table 2). The average transition/transversion ratio (ts/tv) for p-PAR genes is 1.59, close to the ts/tv ratio of 1.4 reported for human pseudogenes (Li et al., 1984). However, the ts/tv ratio varied between the genes studied (Table 2). All the three regions sequenced for the *SHOX* gene showed fairly low ts/tv ratio. The ts/tv ratio for the pseudoautosomal part of the p-PAR boundary was also fairly low (Table 2). This contrasts with the mutation pattern in the *PPP2R3L*, *ASMT* and in the *SYBL1* genes. The high ts/tv ratio in the *PPP2R3L* and

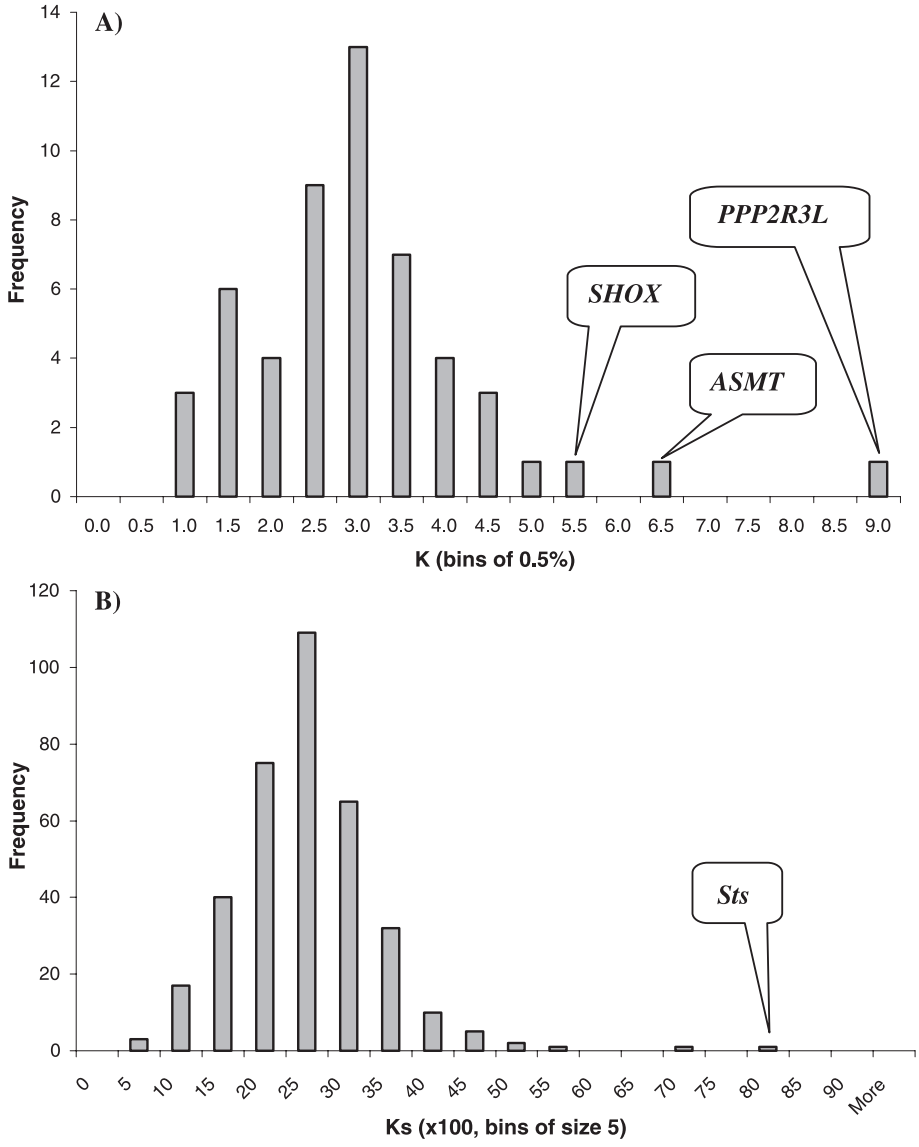

Fig. 1. Distribution of (A) non-coding human/orangutan divergence and (B) synonymous mouse/rat divergence. The divergence of pseudoautosomal genes is shown by the callouts.

ASMT genes may partly be explained by the high proportion of C→T and G→A transitions in CpG dinucleotides. Although the proportion of CpG mutations in the other genes is lower than in *PPP2R3L* and *ASMT*, all the regions showed significant lack of CG pairs, compared to the expected from local GC content. Neither of the p-PAR genes, nor the *SYBL1* gene deviated significantly from GC content stationarity: the number of (A or T)→(G or C) substitutions was not significantly different from the number of substitutions in the opposite direction (GC% stationarity test, Eyre-Walker, 1994).

We also studied substitution patterns in the mouse pseudoautosomal *Sts* gene and in the 3′ part of *Fxy*, which is pseudoautosomal in *M. musculus* (Palmer et al., 1997). Using the rat *Fxy* sequence as an outgroup, we reconstructed ancestral sequence and counted the number of mutations of each type in the *M. musculus* lineage after the split from *M. spretus*. Out of the 126 substitutions at all three codon positions, 42 were A→G and 53 were T→C transitions. Among the remaining 31 substitutions, 11 were T→G and 14 A→C transversions, and only six were of all the other types (ts/tv=3.35±0.02). Given such a biased mutation pattern, it is not surprising that GC content in the pseudoautosomal portion of *M. musculus Fxy* gene increased to 55.5%, compared to 49.1% in *M. spretus*.

A similar, but less pronounced pattern was detected in the *Sts* gene in the mouse lineage after the mouse/rat split, using human *STS* sequence as an outgroup. This gene is pseudoautosomal in *M. musculus*, but X-linked in rat and humans (Salido et al., 1996; Li et al., 1996). Of the 381 substitutions detected among the 1602 sites analysed, the most frequent were A→G (79) and T→C (83) transitions.

The second most frequent substitution type was G→C (60) and C→G (58) transversions, followed by A→C (43) and T→G (29) transversions. All these substitutions either increase, or do not affect the GC content. The number of substitutions decreasing GC content was much smaller, only 29 (G or C)→(T or A) substitutions were observed. The total GC content in the mouse *Sts* gene rose to 75.4%, compared to 62.6% and 52.2% in rat and human homologs, respectively. Four-fold degenerate sites in mouse *Sts* are almost exclusively represented by G or C (GC%=98%), while in rat and human homologs GC%=73.3 and 56.9%, respectively.

4. Discussion

4.1. Substitution rates in the PARs

We have demonstrated that the human p-PAR genes *SHOX*, *ASMT* and *PPP2R3L* exhibit significantly higher human/orangutan silent divergence compared to the autosomal and X-linked regions (Fig. 1A and Table 2). Similarly, the mouse/rat synonymous divergence in the pseudoautosomal *Sts* gene is substantially higher than the average mouse/rat autosomal divergence (Fig. 1B). Our observations are consistent with an elevated synonymous substitution rate in the pseudoautosomal portion of the *Fxy* gene in *Mus musculus* (Perry and Ashworth, 1999). As the silent substitution rate is proportional to the mutation rate (Kimura, 1983), such unusually high divergence might reflect an elevated rate of mutations in the human and orangutan p-PAR and also in the mouse PAR. On the other hand, the divergence in the *SYBL1* gene is fairly close to the average autosomal divergence between the two species, suggesting that either q-PAR genes do not experience an elevated substitution rate, or that the q-PAR is too recent to detect any elevation of substitution rate. The age of the q-PAR is unknown, but it must be younger than the split between humans and chimpanzees, as this region is X-linked in apes (Ciccodicola et al., 2000).

It is interesting that the pseudoautosomal portion of the *XG* gene near the PAR boundary only shows a moderate (and non-significant) increase in substitution rate compared to the X-linked portion of the same gene, and to X-linked and autosomal genes. This may suggest that the frequency of recombination close to the pseudoautosomal boundary is not very high (if recombination is the source of elevated mutation rate in the p-PAR genes). Unfortunately, the best available estimates of recombination in the p-PAR (Lien et al., 2000) are not detailed enough to measure local recombination rate within several kilobases of the pseudoautosomal boundary.

Although divergence in the *SHOX*, *ASMT* and *PPP2R3L* genes is consistently higher than in the autosomal and X-linked genes, there is considerable variation in divergence between these genes. The *PPP2R3L* gene diverges faster than the two other p-PAR genes. Partly, it may be explained by high GC content in the *PPP2R3L* (70%) and frequent C→T transitions in the CG dinucleotides, which are thought to be due to deamination of 5-methylcytosine in methylated CG (Robertson and Wolffe, 2000). *PPP2R3L* is also fairly close to the telomere which may result in much higher divergence in that region (Mefford and Trask, 2002).

Three separate regions sequenced from the orangutan *SHOX* gene also show somewhat different divergence (Table 2), with a maximum of 7% at the "SHOX+7−9", and a minimum of 4.3% at the "SHOX+15−17". It is interesting that the slowly diverging "SHOX+15−17", is located in the region which was reported to be a hot spot of recombination in humans, with recombination being as high as 190−370 cM/Mb (May et al., 2002). Another recombination hot spot in the human *TAP2* gene showed high human diversity, but no elevation in the human/ape divergence (Jeffreys et al., 2000). If recombination is mutagenic, recombination hot spots are expected to show higher human/ape silent divergence than the other regions, unless such hot spots are not of very recent origin. We do not know anything about the timescale of the recombination hot spots. On theoretical grounds, it was argued that such hot spots should be fairly short-lived (Boulton et al., 1997). Jeffreys and Neumann (2002) reported an interesting case where a recombinational hot spot in the human major histocompatibility complex depended on heterozygosity in a single SNP in the middle of the hot spot, suggesting that the high rate of recombination in this region is not a long-lasting phenomenon. Thus, the hot spot in the *SHOX* gene may not have existed for long enough to elevate the number of substitutions in this region. It would be interesting to test whether this recombinational hot spot exists in chimps, which would address the issue of its age.

4.2. Causes of elevated substitution rate in pseudoautosomal regions

Y chromosomes are known to have elevated substitution rate compared to the X chromosomes and autosomes (Shimmin et al., 1993; Agulnik et al., 1997; Filatov and Charlesworth, 2002; Makova and Li, 2002, but see Bohossian et al., 2000 for the opposite conclusion). The reasons for this are thought to be the higher number of cell divisions in the male germ line (Miyata et al., 1987) and a higher per replication mutation rate on the Y chromosome (Smith and Hurst, 1999; Filatov and Charlesworth, 2002). If the *SHOX*, *ASMT* and *PPP2R3L* genes are Y-linked in orangutan, this may result in the elevated human/orangutan divergence. However, the p-PAR boundary was demonstrated to be the same in humans, apes and old world monkeys (Ellis et al., 1990). It was suggested that the modern boundary of the p-PAR was formed in the progenitor of simian primates due to a translocation of *SRY* into the larger ancestral PAR (Glaser et al., 1999). Several human p-PAR genes (*SHOX*, *IL3RA*, *CSF2RA* and *ANT3*) are known to be pseudoautosomal in

lemurs, sheep and dogs (Blaschke and Rappold 1997; Glaser et al., 1999). Thus, the p-PAR is fairly old and conserved, and the homologs of the *SHOX*, *PPP2R3L* and *ASMT* should be pseudoautosomal in orangutan.

Pseudoautosomal genes spend an equal amount of time in males and females, so the differences in the number of cell divisions between males and females cannot affect substitution rates in the PARs. However, different chromosomes may have different mutation rates (Lercher et al., 2001), and if Y chromosomes accumulate substitutions faster than the autosomes (e.g. Filatov and Charlesworth, 2002), this may elevate the mutation rate in the PAR, which spends one fourth of the time as a part of the Y chromosome. If this was the case, the substitution rate in the PAR genes should only be a fraction of that in the Y-linked genes. The human/orangutan silent divergence available for five Y-linked regions ranges from 3.9% in a *ZFY* intron (Shimmin et al., 1993) to 8.2% in a *TSPY* intron (Kim and Takenaka, 1996), and is not significantly higher than in the p-PAR (data not shown), suggesting that a higher mutation rate on the Y chromosome cannot explain the elevated substitution rate in the p-PAR.

Pseudoautosomal genes are known to experience a storm of recombination (Lien et al., 2000), so a high rate of silent substitutions is highly suggestive that recombination is a source of mutations. This hypothesis explains why pseudoautosomal genes diverge much faster than autosomal and X-linked ones. It is also consistent with the positive correlation of human recombination rate with genome-wide human SNP diversity and with human/mouse divergence reported by Lercher and Hurst (2002). Furthermore, we would expect to see a positive correlation of human/orangutan divergence with recombination rate in humans. The correlation detected by Lercher and Hurst (2002) was fairly weak (r^2<0.1). However, for a human/orangutan comparison, one might expect to see a stronger effect, as the recombinational map might be conserved between these species. However, our attempt to detect such a correlation was not successful (r^2=0.061, Fig. 2), suggesting that the effect (if real) is too weak to be detected with the amount of data available for orangutan. Moreover, the slowly diverging SHOX+15−17 region, located in the recombination hot spot, and the low divergence in the *XG* introns seems to contradict the hypothesis that recombination is mutagenic, however, a more detailed recombination map for the *XG* gene is needed to draw the conclusion.

4.3. Substitution patterns

Recombination rate is known to correlate with GC content, for example GC-rich isochores usually have an elevated recombination frequency (Eyre-Walker, 1993; Fullerton et al., 2001). The causal relationship between GC content and the recombination rate is not clear (reviewed by Eyre-Walker and Hurst, 2001). On one hand, the sequence itself can affect recombination frequency (Wahls, 1998; Majewski and Ott, 2000). On the other hand, recombination may change GC content due to biased gene conversion (Lamb, 1986; Galtier et al., 2001). Selection may also play a substantial role in establishing GC content of certain genomic regions (Eyre-Walker, 1999; Bernardi, 2001). In the pseudoautosomal regions, however, it may be possible to disentangle these processes. As frequent recombination in the p-PAR is probably due to obligate chiasmata in male meiosis, it might be likely that high GC content in the p-PAR is the consequence rather than the cause of the elevated recombination rate.

The p-PAR is quite old, as the Xp/Yp pseudoautosomal boundary was demonstrated to be the same in humans, great apes and old world monkeys (Ellis et al., 1990), thus it is not

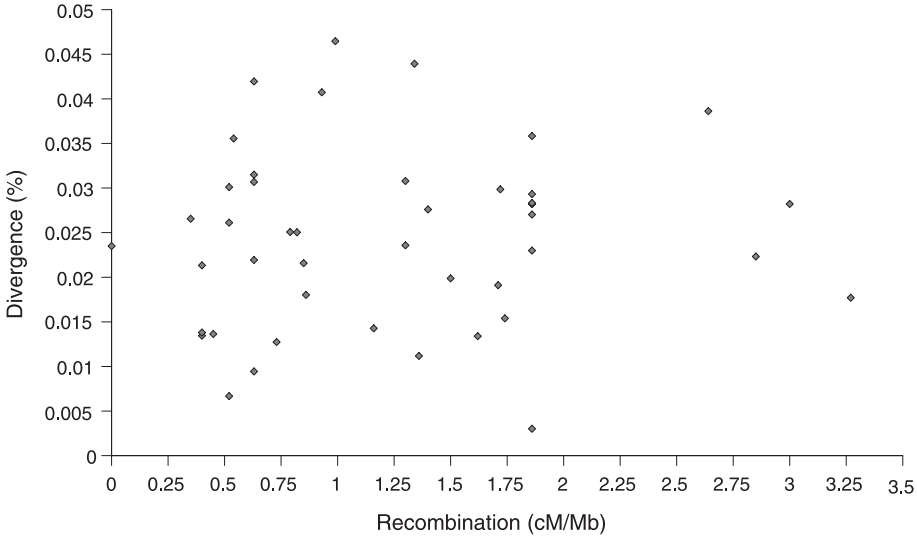

Fig. 2. Plot of human/orangutan non-coding divergence versus human recombination rate (Kong et al., 2002) for the 51 pair of homologous X-linked and autosomal regions in the two species.

surprising that we have not detected significant deviation from GC content stationarity. The human q-PAR, however, is very young, perhaps younger than the split of humans and chimpanzees, since in the apes, this region is X-specific (Ciccodicola et al., 2000). Although recombination in the q-PAR is about six-fold higher than elsewhere on the X chromosome (Ciccodicola et al., 2000), the proximal part of the q-PAR is GC-poor (GC content is 32%). We hypothesise that this is due to a very recent origin of the q-PAR and the GC content of this region will increase with time, i.e. there should be more (A or T)→(G or C) mutations than the mutations in the opposite direction. Unfortunately, we failed to detect a significant non-stationarity of GC content in the *SYBL1* gene, which may be either due to too few substitutions observed, or due to a very recent origin of the human q-PAR. We did detect, however, a significant deviation from GC content stationarity in the *Mus musculus Fxy* and *Sts* genes. The entire *Sts* gene and the 3′ part of the *Fxy* gene are pseudoautosomal in the house mouse (Salido et al., 1996; Palmer et al., 1997), but both *Sts* and *Fxy* genes are located in the differentiated part of the rat X chromosome (Li et al., 1996; Perry et al., 1998). The number of (A or T)→(G or C) mutations in the mouse lineage massively exceeds the number of mutations in the opposite direction, resulting in the elevated GC content in both *M. musculus* genes, compared to the rat homologues, consistent with the hypothesis that recombination biases GC content upwards (e.g. Galtier et al., 2001).

It is interesting that in the mouse *Fxy* and *Sts* genes and in the human *PPP2R3L* and *ASMT* genes the ts/tv ratio is fairly high. In mouse genes, where we can establish the direction of mutations, it is clear that the ts/tv ratio is inflated due to A→G and T→C transitions. The excess of such transitions may be the result of biased gene conversion events (Lamb, 1986; Galtier et al., 2001). In the human/orangutan comparison, we also see a substantial excess of A↔G and T↔C transitions, however, we cannot establish the direction of mutations since no outgroup sequence is available. It may be either due to biased gene conversion which would inflate the number of A→G and T→C mutations, or due to spontaneous deamination of 5-methylcytosine in methylated CG dinucleotides which would cause C→T and G→A transitions with high frequency (Robertson and Wolffe, 2000). The fact that most of the transitions in the *PPP2R3L* and *ASMT* genes occurred in CG dinucleotides (Table 2) suggests that they might be C→T and G→A transitions.

4.4. Mutation rates and human DNA diversity

A high mutation rate in the p-PAR genes may result in the elevated human genetic diversity in this region. Indeed, Schiebel et al. (2000) reported that the human DNA diversity in the *PPP2R3L* gene is as high as E(θ)=0.5%. On the other hand, May et al. (2002) reported much lower diversity (θ=0.07%) in the *SHOX* gene. May et al. (2002) reported 61 SNPs detected in a screen of six regions in and around the *SHOX* gene (26 kb in total) in seven individuals, which gives us E(θ)=S/($\Sigma 1/i$)=61/2.45=24.9 per sequence (where S is the number of segregating sites, and i increases from one to the number of individuals sampled minus one, Watterson, 1975), or about 0.1% per nucleotide, slightly higher than θ=0.07% reported by May et al. (2002). The diversity in the *SHOX* gene was measured by allele-specific oligonucleotide hybridization (May et al., 2002), which may have resulted in rare polymorphic sites being missed and the θ reported by May et al. (2002) being biased downwards. However, the difference between the two θ estimates is not large and both values are close to the estimates of DNA diversity in the autosomal and X-linked regions (Kaessmann et al., 1999; Jaruzelska et al., 1999; Nachman and Crowell, 2000; Zhao et al., 2000; Yu et al., 2001b), and substantially lower than the E(θ)=0.5%, reported for the *PPP2R3L* gene (Schiebel et al., 2000).

Variation in mutation rates between the regions may be one of the causes of genetic diversity differences between the genes. Using orangutan sequences as the outgroups, we

Table 3
HKA tests for PAR, X-linked and autosomal genes

Locus	*SHOX*[a]	*PPP2R3L*[b]	Zfx[c]	Xq13.3[d]	DMD-i7[e]	DMD-i44[e]	Chr1[f]	Chr22[g]
SNP	61	120	10	33	9	19	48	75
Length (bp)	26,000	20,299	1089	10,000	2389	3000	10,000	10,000
Sample size	7	2	335	69	41	41	122	128
Substitutions	120	72	32	300	48	84	245	306
Length (bp)	2157	854	1213	10,000	1383	2982	9017	9777
HKA *P*-values[h]								
SHOX	–	<0.00001***	0.0004***	0.143	0.171	0.0001***	0.071	0.004**
PPP2R3L	<0.00001***	–	0.923	0.019*	0.127	0.855	0.019*	0.105

[a] Diversity data from May et al., 2002, divergence from this study.
[b] Diversity data from Schiebel et al., 2000, divergence from this study.
[c] Jaruzelska et al., 1999.
[d] Kaessmann et al., 1999.
[e] Nachman and Crowell, 2000.
[f] Yu et al., 2001b.
[g] Zhao et al., 2000.
[h] The significant values (***P<0.001, **P<0.01, *P<0.05) indicate a lack of diversity in the *SHOX* and an excess of diversity in the *PPP2R3L* gene.

can test whether the differences in mutation rates between the regions are sufficient to explain the observed differences in human DNA diversity (HKA test, Hudson et al., 1987). To conduct the HKA tests for the *SHOX* gene, we used 61 segregating sites detected by May et al. (2002) in 26 kb sequenced from seven individuals. We cannot use the larger number of individuals genotyped in this study because the number of polymorphic sites (SNPs) comes from the original sequence of seven individuals, and further genotyping did not add additional polymorphic sites. For the *PPP2R3L* gene, we used the number of polymorphic sites (S=120) estimated according to Watterson (1975) formula from the diversity estimate π=0.5% reported for two 20-kb sequences by Schiebel et al. (2000). The divergence estimates for *PPP2R3L* and *SHOX* used for the HKA test are shown in Table 2 (three *SHOX* regions pooled). The comparison of diversity in the *SHOX* and *PPP2R3L* genes by the HKA test shows a significant (P<0.00001) difference in diversity between the two regions, suggesting that the difference in mutation rate between the two regions is not sufficient to explain the difference in diversity (Table 3).

The results of the pairwise HKA tests for the comparison of the two p-PAR genes with the diversity in several human X-linked and autosomal genes for which an orangutan outgroup sequence is available (Kaessmann et al., 1999; Jaruzelska et al., 1999; Nachman and Crowell, 2000; Zhao et al., 2000; Yu et al., 2001b) is shown in Table 3. The diversity in the *PPP2R3L* gene, corrected for the higher mutation rate, is marginally higher than in the Xq13 and chromosome 1 non-coding sequences, and not significantly higher than the other X-linked and autosomal genes, suggesting that the excess of diversity in the *PPP2R3L* gene may, at least partly, be explained by the higher mutation rate in this gene we reported here. On the other hand, with the correction for mutation rate differences, *SHOX* shows a highly significant lack of diversity compared to the last *ZFX* intron, intron 44 of the DMD gene and the non-coding region on chromosome 22. Thus, it is quite likely that the diversity in the *SHOX* gene is lower than expected from the mutation rate in this region and, perhaps, it is reduced by some additional factor. Mutations in the *SHOX* gene might be highly deleterious, as they have been associated with Leri-Weill dyschondrosteosis (Shears et al., 1998), however, it is not clear how this can reduce diversity in a region with such a high frequency of recombination. Overall, the results of the HKA tests suggest that the high diversity in the *PPP2R3L* gene is mostly due to a high underlying mutation rate, while the diversity in the *SHOX* is reduced, compared to that expected from the mutation rate in this gene.

5. Conclusions

We demonstrated that the human and mouse pseudoautosomal genes, undergoing frequent recombination, accumulate silent substitutions significantly faster than the autosomal and the X-linked genes, which seems to suggest that recombination elevates mutation rate in the pseudoautosomal regions. However, this effect might be fairly weak, as the human q-PAR and the *XG* gene located at the p-PAR boundary do not show an elevated substitution rate. Overall, our data do not contradict the hypothesis that recombination is mutagenic, but further studies are needed to draw the final conclusion.

Acknowledgements

We thank Alec Jeffreys, Gil McVean, and Joe Ironside for suggestions and critical reading of the manuscript. This work was supported by a grant to D.A.F. from the Wellcome Trust (grant No 068193). D.T.G. was supported by a PhD studentship from the BBSRC.

References

Agulnik, A.I., Bishop, C.E., Lerner, J.L., Agulnik, S.I., Solovyev, V.V., 1997. Analysis of mutation rates in the *SMCY/SMCX* genes shows that mammalian evolution is male driven. Mamm. Genome 8, 134–138.

Aquadro, C.F., DuMont, V.B., Reed, F.A., 2001. Genome-wide variation in the human and fruit fly: a comparison. Curr. Opin. Genet. Dev. 11, 627–634.

Baird, D.M., Royle, N.J., 1997. Sequences from higher primates orthologous to the human Xp/Yp telomere junction region reveal gross rearrangements and high levels of divergence. Hum. Mol. Genet. 6, 2291–2299.

Belin, V., Cusin, V., Viot, G., Girlich, D., Toutain, A., Moncla, A., Vekemans, M., Le Merrer, M., Munnich, A., Cormier-Daire, V., 1998. *SHOX* mutations in dyschondrosteosis (Leri-Weill syndrome). Nat. Genet. 19, 67–69.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Rapp, B.A., Wheeler, D.L., 2002. GenBank. Nucleic Acids Res. 30, 17–20.

Bernardi, G., 2001. Misunderstandings about isochores: Part 1. Gene 276, 3–13.

Blaschke, R.J., Rappold, G.A., 1997. Man to mouse—lessons learned from the distal end of the human X chromosome. Genome Res. 7, 1114–1117.

Bohossian, H.B., Skaletsky, H., Page, D.C., 2000. Unexpectedly similar rates of nucleotide substitution found in male and female hominids. Nature 146, 622–625.

Boulton, A., Meyers, R.S., Redfield, R.J., 1997. The hotspot conversion paradox and the evolution of meiotic recombination. Proc. Natl. Acad. Sci. U. S. A. 94, 8058–8063.

Brown, W.R.A., 1988. A physical map of the human pseudoautosomal region. EMBO J. 7, 2377–2385.

Burgoyne, P.S., Mahadevaiah, S.K., Sutcliffe, M.J., Palmer, S.J., 1992. Fertility in mice requires X-Y pairing and a Y-chromosomal spermiogenesis gene mapping to the long arm. Cell 71, 391–398.

Charlesworth, B., Morgan, M.T., Charlesworth, D., 1993. The effect of deleterious mutations on neutral molecular variation. Genetics 134, 1289–1303.

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

Ciccodicola, A., D'Esposito, M., Esposito, T., et al., 2000. Differentially regulated and evolved genes in the fully sequenced Xq/Yq pseudoautosomal region. Hum. Mol. Genet. 9, 395–401.

Cooke, H.J., Brown, W.R., Rappold, G.A., 1985. Hypervariable telomeric sequences from the human sex chromosomes are pseudoautosomal. Nature 317, 687–692.

Ellis, N., Yen, P., Neiswanger, K., Shapiro, L.J., Goodfellow, P.N., 1990. Evolution of the pseudoautosomal boundary in the old world monkeys and great apes. Cell 63, 977–986.

D'Esposito, M., Matarazzo, M.R., Ciccodicola, A., Strazzullo, M., Mazzarella, R., Quaderi, N.A., Fujiwara, H., Ko, M.S., Rowe, L.B., Ricco, A., Archidiacono, N., Rocchi, M., Schlessinger, D., D'Urso, M., 1997. Differential expression pattern of XqPAR-linked genes SYBL1 and IL9R correlates with the structure and evolution of the region. Hum. Mol. Genet. 6, 1917–1923.

Eyre-Walker, A., 1993. Recombination and mammalian genome evolution. Proc. Roy. Soc. Lond., B Bio. 252, 237–243.

Eyre-Walker, A., 1994. DNA mismatch repair and synonymous codon evolution in mammals. Mol. Biol. Evol. 11, 88–98.

Eyre-Walker, A., 1999. Evidence of selection on silent site base composition in mammals: potential implications for the evolution of isochores and junk DNA. Genetics 152, 675–683.

Eyre-Walker, A., Hurst, L., 2001. The evolution of isochores. Nat. Rev., Genet. 2, 549–555.

Filatov, D.A., 2002. ProSeq: a software for preparation and evolutionary analysis of DNA sequence data sets. Mol. Ecol. Notes 2, 621–624.

Filatov, D.A., Charlesworth, D., 2002. Substitution rates in the X- and Y-linked genes of the plants Silene latifolia and *S. dioica*. Mol. Biol. Evol. 19, 898–907.

Freije, D., Helms, C., Watson, M.S., Donis-Keller, H., 1992. Identification of a second pseudoautosomal region near the Xq and Yq telomeres. Science 258, 1784–1787.

Fullerton, S.M., Carvalho, A.B., Clark, A.G., 2001. Local rates of recombination are positively correlated with GC content in the human genome. Mol. Biol. Evol. 18, 1142–1339.

Galtier, N., Piganeau, G., Mouchiroud, D., Duret, L., 2001. GC-content evolution in mammalian genomes: the biased gene conversion hypothesis. Genetics 159, 907–911.

Glaser, B., Myrtek, D., Rumpler, Y., Schiebel, K., Hauwy, M., Rappold, G.A., Schempp, W., 1999. Transposition of SRY into ancestral pseudoautosomal region creates a new pseudoautosomal boundary in a progenitor of simian primates. Hum. Mol. Genet. 8, 2071–2078.

Hudson, R.R., Kreitman, M., Aguade, M., 1987. A test of neutral molecular evolution based on nucleotide data. Genetics 116, 153–159.

Jaruzelska, J., Zietkiewicz, E., Batzer, M., Cole, D.E.C., Moisan, J.-P., Scozzari, R., Tavare, S., Labuda, D., 1999. Spatial and temporal distribution of the neutral polymorphisms in the last ZFX intron: analysis of the haplotype structure and genealogy. Genetics 152, 1091–1101.

Jeffreys, A.J., Neumann, R., 2002. Reciprocal crossover asymmetry and meiotic drive in a human recombination hot spot. Nat. Genet. 31, 267–271.

Jeffreys, A.J., Ritchie, A., Neumann, R., 2000. High-resolution analysis of haplotype diversity and meiotic crossover in the human TAP2 recombination hotspot. Hum. Mol. Genet. 9, 725–733.

Kaessmann, H., Heissig, F., von Haeseler, A., Paabo, S., 1999. DNA sequence variation in a non-coding region of low recombination on the human X chromosome. Nat. Genet. 22, 78–81.

Kim, H.S., Takenaka, O., 1996. A comparison of TSPY genes from Y-chromosomal DNA of the great apes and humans: sequence, evolution, and phylogeny. Am. J. Phys. Anthropol. 100, 301–309.

Kimura, M., 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16, 111–120.

Kimura, M., 1983. The Neutral Theory of Molecular Evolution. Cambridge Univ. Press, Cambridge.

Kong, A., Gudbjartsson, D.F., et al., 2002. A high-resolution recombination map of the human genome. Nat. Genet. 31, 241–247.

Kumar, S., Tamura, K., Nei, M., 1994. MEGA—molecular evolutionary genetics analysis software for microcomputers. Comput. Appl. Biosci. 10, 189–191.

Lamb, B.C., 1986. Gene conversion disparity—factors influencing its direction and extent, with tests of assumptions and predictions in its evolutionary effects. Genetics 114, 611–632.

Lercher, M.J., Hurst, L.D., 2002. Human SNP variability and mutation rate are higher in regions of high recombination. Trends Genet. 18, 337–340.

Lercher, M.J., Williams, E.J.B., Hurst, L.D., 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. Mol. Biol. Evol. 18, 2032–2039.

Li, W.-H., 1997. Molecular Evolution. Sinauer Ass., Sunderland, MA.

Li, W.-H., Wu, C.-I., Luo, C.-C., 1984. Nonrandomness of point mutations as reflected in nucleotide substitutions in pseudogenes and its evolutionary implications. J. Mol. Evol. 21, 58–71.

Li, W.-H., Wu, C.-I., Luo, C.-C., 1985. A new method for estimating synonymous and non-synonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol. Biol. Evol. 2, 150–174.

Li, X.M., Salido, E.C., Gong, Y., Kitada, K., Serikawa, T., Yen, P.H., Shapiro, L.J., 1996. Cloning of the rat steroid sulfatase gene (*Sts*), a non-pseudoautosomal X-linked gene that undergoes X inactivation. Mamm. Genome 7, 420–424.

Lien, S., Szyda, J., Schechinger, B., Rappold, G., Arnheim, N., 2000. Evidence for heterogeneity in recombination in the human pseudoautosomal region: high resolution analysis by sperm typing and radiation-hybrid mapping. Am. J. Hum. Genet. 66, 557–566.

Majewski, J., Ott, J., 2000. GT repeats are associated with recombination on human chromosome 22. Genome Res. 10, 1108–1114.

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

Matarazzo, M.R., Cuccurese, M., Strazzullo, M., Vacca, M., Curci, A., Giuseppina, M.M., Cocchia, M., Mercadante, G., Torino, A., D'Urso, M., Ciccodicola, A., D'Esposito, M., 1999. Human and mouse SYBL1 gene structure and expression. Gene 240, 233–238.

May, C.A., Shone, A.C., Kalaydjieva, L., Sajantila, A., Jeffreys, A.J., 2002. Crossover clustering and rapid decay of linkage disequilibrium in the Xp/Yp pseudoautosomal gene *SHOX*. Nat. Genet. 31, 272–275.

McVean, G.T., Hurst, L.D., 1997. Evidence for a selectively favourable reduction in the mutation rate of the X chromosome. Nature 386, 388–392.

Mefford, H.C., Trask, B.J., 2002. The complex structure and dynamic evolution of human subtelomeres. Nat. Rev., Genet. 3, 91–102.

Miyata, T., Hayashida, H., Kuma, K., Mitsuyasa, K., Yasunaga, T., 1987. Male-driven molecular evolution: a model and nucleotide sequence analysis. Cold Spring Harbor Sym. 52, 863–867.

Nachman, M.W., 2001. Single nucleotide polymorphisms and recombination rate in humans. Trends Genet. 17, 481–485.

Nachman, M.W., Crowell, S., 2000. Contrasting evolutionary histories of two intrins of the Duchenne muscular dystrophy gene, *Dmd*, in humans. Genetics 155, 1855–1864.

Palmer, S., Perry, J., Kipling, D., Ashworth, A., 1997. A gene spans the pseudoautosomal boundary in mice. Proc. Natl. Acad. Sci. U. S. A. 94, 12030–12035.

Perry, J., Ashworth, A., 1999. Evolutionary rate of a gene affected by chromosomal position. Curr. Biol. 9, 987–989.

Perry, J., Feather, S., Smith, A., Palmer, S., Ashworth, A., 1998. The human Fxy gene is located within Xp22.3: implications for evolution of the mammalian X chromosome. Hum. Mol. Genet. 7, 299–305.

Petit, C., Levilliers, J., Weissenbach, J., 1988. Physical mapping of the human pseudoautosomal region; comparison with the genetic linkage map. EMBO J. 7, 2369–2376.

Rao, E., Weiss, B., Fukami, M., Rump, A., Niesler, B., Mertz, A., Muroya, K., Binder, G., Kirsch, S., Winkelmann, M., Nordsiek, G., Heinrich, U., Breuning, M.H., Ranke, M.B., Rosenthal, A., Ogata, T., Rappold, G.A., 1997. Pseudoautosomal deletions encompassing a novel homeobox gene cause growth failure in idiopathic short stature and Turner syndrome. Nat. Genet. 16, 54–63.

Rappold, G.A., 1993. The pseudoautosomal regions of the human sex chromosomes. Hum. Genet. 92, 315–324.

Rattray, A.J., McGill, C.B., Shafer, B.K., Strathern, J.N., 2001. Fidelity of mitotic double-strand-break repair in *Saccharomyces cerevisiae*: a role for SAE2/COM1. Genetics 158, 109–122.

Rice, W.R., 1987. Genetic hitchhiking and the evolution of reduced genetic activity of the Y sex chromosome. Genetics 116, 161–167.

Robertson, K.D., Wolffe, A.P., 2000. DNA methylation in health and disease. Nat. Rev., Genet. 1, 11–19.

Rodriguez, I.R., Mazuruk, K., Schoen, T.J., Chader, G.J., 1994. Structural analysis of the human hydroxyindole-*O*-methyltransferase gene. Presence of two distinct promoters. J. Biol. Chem. 269, 31969–31977.

Salido, E.C., Li, X.M., Yen, P.H., Martin, N., Mohandas, T.K., Shapiro, L.J., 1996. Cloning and expression of the mouse pseudoautosomal steroid sulphatase gene (*Sts*). Nat. Genet. 13, 83–86.

Schiebel, K., Meder, J., Rump, A., Rosenthal, A., Winkelmann, M., Fischer, C., Bonk, T., Humeny, A., Rappold, G., 2000. Elevated DNA sequence diversity in the genomic region of the phosphatase *PPP2R3L* gene in the human pseudoautosomal region. Cytogenet. Cell Genet. 91, 224–230.

Shears, D.J., Vassal, H.J., Goodman, F.R., Palmer, R.W., Reardon, W., Superti-Furga, A., Scambler, P.J., Winter, R.M., 1998. Mutation and deletion of the pseudoautosomal gene *SHOX* cause LeriWeill dyschondrosteosis. Nat. Genet. 19, 70–73.

Shimmin, L.C., Chang, B.H.J., Li, W.-H., 1993. Maledriven evolution of DNA sequences. Nature 362, 745–747.

Smith, N.G.C., Hurst, L.D., 1999. The causes of synonymous rate variation in the rodent genome: can substitution rates be used to estimate the sex bias in mutation rate? Genetics 152, 661–673.

Sokal, R.R., Rohlf, F.J., 1995. Biometry, 3rd ed. Freeman, San Francisco.

Strathern, J.N., Shafer, B.K., McGill, C.B., 1995. DNA synthesis errors associated with double-strand-break repair. Genetics 140, 965–972.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.

Wahls, W.P., 1998. Meiotic recombination hotspots: shaping the genome and insights into hypervariable minisatellite DNA change. Curr. Top. Dev. Biol. 37, 37–75.

Watterson, G.A., 1975. On the number of polymorphic sites in genetical models without recombination. Theor. Popul. Biol. 7, 256–276.

Wolfe, K.H., Sharp, P.M., 1993. Mammalian gene evolution nucleotide sequence divergence between mouse and rat. J. Mol. Evol. 37, 441–456.

Yang, Z., 2001. Phylogenetic Analysis by Maximum Likelihood (PAML). Version 3e. University College London.

Yu, A., Zhao, C.F., Fan, Y., Jang, W.H., Mungall, A.J., Deloukas, P., Olsen, A., Doggett, N.A., Ghebranious, N., Broman, K.W., Weber, J.L., 2001a. Comparison of human genetic and sequence-based physical maps. Nature 409, 951–953.

Yu, N., Zhao, Z., Fu, Y.-X., Sambuughin, N., Ramsay, M., Jenkins, T., Leskinen, E., Patthy, L., Jorde, L.B., Kuromori, T., Li, W.-H., 2001b. Global patterns of human DNA sequence variation in a 10 kb region on chromosome 1. Mol. Biol. Evol. 18, 214–222.

Zhao, Z., Jin, L., Fu, Y.-X., Ramsay, M., Jenkins, T., Leskinen, E., Pamilo, P., Trexler, M., Patthy, L., Jorde, L.B., Yu, N., Li, W.-H., 2000. Worldwide DNA sequence variation in a 10 kilobase noncoding region on human chromosome 22. Proc. Natl. Acad. Sci. U. S. A. 97, 11354–11358.