Significant positive correlation between the recombination rate and GC content in the human pseudoautosomal region

Jin-Feng Chen[1,2]， Fei Lu[1,2]， Su-Shing Chen[2,3] and Shi-Heng Tao[1,2]*
1 School of Life Science, Northwest A & F University， Yangling, Shaanxi 712100, China
2 Institute of Bioinformatics, Northwest A & F University, Yangling, Shaanxi 712100, China
3 Institute of Genetics, University of Florida, Gainesville, FL 32611, USA

*Corresponding author
Current address:
School of Life Science, Northwest A & F University
Yangling, Shaanxi 712100, China
Phone: +86-029-87091526
Fax: +86-029-87092262
shihengt@nwsuaf.edu.cn

Keywords: Pseudoautosomal region, GC-content, base composition, evolution, recombination

Abstract

This paper establishes that recombination drives the evolution of GC-content in a significant way. As the human P-arm pseudoautosomal region (PAR1) has demonstrated to have a high recombination rate, at least 20-fold more frequent than the genomic average of ~1 cM/Mb, this region provides an ideal system to study the role of recombination in the evolution of base composition. Nine non-coding regions of PAR1 are analyzed in this study. We have observed a highly significant positive correlation between the recombination rate and GC-content (ρ=0.837, $P \leq 0.005$). Five regions that lie in the distal part of PAR1 are shown to be significantly higher than genomic average divergence. By comparing the intra- and inter-specific AT→GC: GC→AT ratios, we have detected no fixation bias toward GC alleles except for *L254915*, which has excessive AT→GC changes in the human lineage. Thus we conclude that the high GC-content of the PAR1 genes fits better the BGC (Biased Gene Conversion) model.

Introduction

The mammalian genomes consist of large regions (>300kb) of relatively homogeneous base composition, which are known as isochores (Bernardi 2000). This characteristic of genome has been a fundamental problem to understand the organization of genomes. The evolution of base composition has drawn great interests recently (Meunier and Duret 2004; Huang et al. 2004; Yi and Li 2005; Ebersberger et al. 2005). Studies have focused on the role of recombination in GC-content evolution. A positive correlation between recombination rate and GC-content has been found in human (Kong et al. 2002; Meunier and Duret 2004), bird (Hurst, Brunton, and Smith 1999), rodent (Williams and Hurst 2000), worm (Marais et al. 2001), insect (Takano-Shimizu 2001), plant (Birdsell, J. A. 2002) and yeast (Gerton et al. 2000). Alternatively, recombination has been shown to be mutagenic in mammals (Brown and jiricny 1987, 1988, 1989) and in yeast (Strathern et al. 1995). It is believed that recombination should be the primary determinant of the evolution of base composition (Montoya-Burgos et al. 2003).

Although it is believed that recombination drives the evolution of base composition, the exact mechanism is not clearly known (Meunier and Duret 2004). Two models are most likely to explain the mechanism: the Biased Gene Conversion (BGC) model proposes a biased fixation toward GC-alleles in GC/AT heterozygous site due to biased DNA mismatch repair process during meiotic recombination (Galtier et al. 2001; Marais, G 2003). This process results in an increase of GC-content in the frequent recombination region. The Regional Mutation Biased (RMB) model suggests a variation in the ratios of GC→AT and AT→GC mutation rates among genomic region to account for their specific base composition (Wolfe, Sharp, and Li 1989). The mutagenicity of recombination would raise the possibility that new AT→GC mutations are introduced during recombination. Then, under neutrality, the regional GC-content will be increased if the mutation process of this region is biased toward AT→GC and decreased if reverse process occur. Supporting evidences have accumulated for both the Regional Mutation Bias model (Casane et al. 1997; Francino and Ochman 1999; Ebersberger, I et al. 2005) and Biased Gene Conversion model (Eyre-Walker 1993; Fullerton et al. 2001; Galtier et al. 2001; Smith and Eyre-Walker 2001; Meunier and Duret 2004). Since the BGC model predicts a fixation bias in favor of GC alleles while the RMB model does not - actually neutrality - we can distinguish them by studying the process of alleles fixation. Therefore whether or not there is a fixation bias toward GC alleles will be a significant evidence to

choose a mechanism over another.

The human X and Y-chromosomes recombine and pair in two small pseudoautosomal regions (PARs) at the end of the sex chromosomes (Cooke, Brown, and Rappold 1985; Freije et al. 1992). The ~2.6-Mb long PAR1, which comprises the tip of X/Y short arm, is homologous to the long PAR of other great apes and old world monkeys (Ellis et al. 1990; Graves et al. 1998); The PAR1 recombination rate in male meiosis was reported ~20-fold more frequent than the genome-average rate (Lien et al. 2000). The long (Xq/Yq) arms contain a smaller PAR2, which is 330kb in size and were reported to have a lower recombination rate about ~5-fold more frequent than the genome-average rate (Ciccodicola et al. 2000). The long PAR1 provides an ideal system to study the influence of recombination on GC-content. The GC-content of these genes that lie in PAR1 forms a gradient, rising with the distance from the PAR1 boundary (PAB) to telomere (Filatov 2004). As it is proposed that recombination should be much more frequent near the telomere than in the more proximal region of the PAR1 (Filatov 2004), we expect a positive correlation between the recombination rate and GC-content. But no significant correlation has been found in previous studies (Filatov and Gerrard 2003; Filatov 2004; Yi and Li 2005), which is not consistent with the hypothesis that recombination drives the evolution of GC-content. In this study, we analyze nine non-coding regions of the PAR1. With our newly calculated recombination rate, we have found a strong and significant positive correlation between the recombination rate and the GC-content (ρ=0.837, $P \leq 0.005$). By comparing the intra- and inter-specific AT→GC: GC→AT ratios, we have detected no fixation bias toward GC alleles except for *L254915*, which has excessive AT→GC changes in the human lineage. Thus we conclude that the high GC-content of PAR1 genes fits better the BGC model.

Materials and Methods

All human sequences used in this study are taken from the UCSC bioinformatics site using Human BLAT Search tool (http://genome.ucsc.edu). The chimpanzee and orangutan sequences are taken from the previous studies (Filatov and Gerrard 2003, Filatov 2004). Repetitive elements are masked using RepeatMasker web server (A.F.A. Smit and P. Green, http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker). For each region we construct a human-chimpanzee-orangutan alignment using ClustalW with the default parameters (Thompson et al.1994). Further analyses of mutation patterns are performed using ProSeq (Filatov 2002). The neighbor-joining (NJ) trees are built using the Kimura 2-parameter model in MEGA3 (Kumar, Tamura, Nei 2004).

To test for the stationary-ness of the base composition, we compare the number of sites

fixed for GC in human but exhibiting AT in chimpanzee with the number of sites fixed for AT in human but exhibiting GC in chimpanzee. Equality of these values is consistent with equal rates of AT→GC and GC→AT substitutions between species and thus implies stationary-ness. We have compared the observed ratios of hum_{GC}, chi_{AT}: hum_{AT}, chi_{GC} to the neutral expectation of 1:1 as described for intra-species changes. Significance is tested using two-tailed binomial test.

Recombination rate estimates are stemmed from a comparison of the physical map of PAR1 region with a genetic map that is used directly to estimate the recombination rate in the human PAR1 (Lien et al. 2000). Markers are assigned to the sequence of the May 2004 freeze of the Human Genome at UCSC (http://genome.ucsc.edu). Then recombination per unit of physical distance between markers of PAR1 is recalculated according to the new marker positions. In this study, recombination rate, referred to recombination per unit of physical distance, is compared to the genome-average rate of 1 cM/Mb.

Result

Nine regions analyzed in this study are listed in Table1. The alignment of the DNA sequences for each region from human, chimpanzee, orangutan, after exclusion of exon sequences and repetitive elements, results in a total of 13kb compared nucleotide positions. Recombination rate for each studied region is assigned according to the physical location of these regions in the PAR1 Marker intervals. Recombination rate for each marker interval is shown in Table2 compared to the data from Lien et al. 2000

GC-content

The GC-content of these PAR regions has a tendency to rise with distance from the PAR1 boundary (Figure1). The GC-content of the proximal part adjacent to the PAR1 boundary, PAB, XG and MIC2, is from 37.9% to 43.2%. The GC-content of the region including *L254916*, *L254915*, *DHRSXY* rises to 41.9% to 47.3%. The GC-content of the region farther away from PAR1 boundary including *ASMT* and *ASMTL* reaches 54.5% and 58.2% respectively. For more distal *SHOX* gene the GC-content is as high as 62.9% (Table1). The gradient exists still after exclusion of CpGs (Figure1). The positive correlation of GC-content with the distance from the PAR1 boundary is highly significant ($\rho=0.945$, $p \leq 0.001$). With the newly calculated recombination rate, we have found a significant positive correlation between the GC-content and the recombination rate ($\rho=0.837$, $p \leq 0.005$). These correlations remain significant after excluding CpGs ($\rho=0.926$, $p \leq 0.001$ and $\rho=0.870$, $p \leq 0.002$ respectively).

Additionally, the average GC-content of these PAR1 genes is 47.7%, which is nearly equal to the GC-content of PAR1 (46.64%) but much higher than the GC-content of the entire X

chromosome (37.68%) or of the human genome (41%) (Lander et al. 2001; Yi et al 2004).

Mutation rate

The substitution rate of these PAR1 regions, expressed as the total tree length, was shown to rise with the distance from the PAR boundary (Filatov 2004). We present this work using human-orangutan divergence as well as total tree length of three species NJ tree. Both of them are positively correlated with the recombination rate (ρ=0.679, $P \leq 0.044$ for the human-orangutan divergence, ρ=0.762, $P \leq 0.017$ for the total tree length). We compare the divergence estimated between human and orangutan sequences of these regions with the average genomic divergence between human and orangutan, which has been estimated to be 3% (Filatov and Gerrard 2003). Five of them are significantly greater than the average genomic divergence between human and orangutan (Table3). All the five regions that have a higher mutation rate lie in the distal part of the PAR1. As the recombination rate in the distal PAR1 is much higher than the proximal part and the positive correlation between the human-orangutan divergence and the recombination rate is significant (ρ=0.679, $P \leq 0.044$), we infer that it is recombination that increase the substitution rate, at least in PAR1.

The patterns of substitution along the human lineage are inferred using the human-chimpanzee-orangutan alignment. Among the studied regions, we infer 97 GC$\leftrightarrow$AT changes: 55 AT$\rightarrow$GC and 42 GC$\rightarrow$AT (Table4). The frequency of AT$\rightarrow$GC changes does not differ significantly from 0.5 ($p \leq 0.187$). Thus, the results are consistent with a 1:1 ratio of AT$\rightarrow$GC: GC$\rightarrow$AT changes within species. The ratios for these individual regions do not differ significantly from 1:1 except for *L254915*, which has excessive AT$\rightarrow$GC changes in the human lineage (Table4). In addition, we have tested for the stationary-ness of base composition between species (Wernegreen and Funk 2004; Eyre-Walker 1994,1997). A number of 101 sites are fixed for AT in human, but in chimpanzee and 113 sites for GC the converse is true (Table4). The ratio did not differ significantly from a frequency of 0.5 ($p \leq 0.412$), as well as the ratios for those individual regions (Table4). Thus, the results are consistent with a 1:1 ratio of AT$\rightarrow$GC: GC$\rightarrow$AT changes since the divergence of these species from their common ancestor and, thus, are stationary. Since the theory predicts that the ratio of the two mutational categories within and between species should be very sensitive to even slight fitness differences (Kimura 1968,1983), this similarity of intra- and inter-specific AT$\rightarrow$GC: GC$\rightarrow$AT ratios is highly consistent with neutral expectation and suggests that neither type of mutational change is selectively preferred. These results suggest that mutation may play a much larger role in the evolution of GC-content in the PAR1.

Discussion

In this study we have found a significant positive correlation between the GC-content and the recombination rate in these PAR1 regions. The GC-content of these PAR1 regions forms a gradient, rising with the distance from the PAR boundary (Filatov 2004). As it is proposed that recombination is much more frequent near the telomere than in the more proximal region in the PAR1 (Filatov 2004), we expect a positive correlation between them. But no significant correlation was found in previous studies (Filatov and Gerrard 2003; Filatov 2004; Yi and Li 2005). The main cause of this non-correlation may be due to a lack of more detailed recombination map of PAR1. To date, the most extensive study that directly estimates the recombination rate in the human PAR1 has been done by Lien et al. 2000. The recombination rate taken from the literature - known as recombination per physical unit compared to the genome-average rate of 1cM/Mb - was calculated by comparison of the genetic map and the physical map. The physical map of human genome has modified for three times since the Lien et al work was done. So the recombination rate may have changed if the new physical positions of these markers are used. This is true in the present study (Table2). Another possible cause of non-correlation in Yi and Li 2005 may be due to the loci used in analysis. The recombination hotspots are mostly short-lived and may not influence the current GC-content (Boulton, Myers, and Redfield 1997). *XG11* and *XG12* are X-linked and proposed to have moved from PAR1 to an X-linked location (Galtier, N. 2004). The GC-contents of these two loci remain as high as the average GC-content in PAR1 and may not reflect the local recombination rate (52.6%, 45.9% and 46.5%, respectively). The PAR1 is highly conserved in human, great apes and old world monkeys (Eills et al. 1990; Graves et al. 1998). Because obligatory meiotic crossover between the X- and Y- chromosomes is confined to the short PAR1, the recombination rate of this region would be higher and the recombination map might be conserved in the great apes (Lien et al 2000; Filatov and Gerrard 2003). Thus, the correlation that is obtained using only PAR1 regions should be reasonable.

Comparative data sets are collected from previous studies (Table 5). Only regions that lie in the PAR1 are selected. With the new recombination rate, we have found a significant positive correlation between the recombination rate and the GC-content using the data set of Yi and Li 2005 (ρ=0.843, $P \leq 0.017$). But, as the physical map of human genome has changed, we have found that *ASMT* is assigned to a different markers interval (now between *DXYS234* and *DXYS85* and the new recombination rate is 17.95 cM/Mb). This change does not influence the result, but leads to a stronger correlation (ρ=0.91, $P \leq 0.004$). For the data set of Filatov 2004, only when *PPP2R3L* is not included in the analysis can we find a significant positive correlation between the GC-content and the new recombination rate (ρ=0.785, $P \leq 0.012$). *PPP2R3L* is assigned to the markers interval between *TEL* and *DXYS201*, which

lie in the more distal part of the PAR1. The recombination rate of this region was supposed to be higher than the more proximal region in the PAR1 (Filatov 2004). But the recombination rate of the region is 13.3 cM/Mb (15.38 cM/Mb from Lien et al 2000, Table2), and is not in accordance with the theoretical prediction. So it is not surprised to get a result of non-correlation between the recombination rate and the GC-content when the analysis includes *PPP2R3L*.

As discussed in Filatov 2004, the most likely cause of the substitution rate gradient in PAR1 is the mutagenic effect of recombination. The positive correlation between the substitution rate and the recombination rate found in this study confirms the hypothesis that it is recombination that increases the substitution rate in the PAR1. Five of these PAR1 regions show significantly higher human-orangutan silent divergence compared to the average genomic divergence between human and orangutan. However, the PAR1 boundary was demonstrated to be the same in humans and apes (Ellis et al 1990). Due to the obligatory crossover between the X and Y chromosome during male meiosis, PAR1 genes are known to experience a storm of recombination (Lien et al 2000). Since all the five regions lie in the distal part of the PAR1 and the recombination rate in the distal PAR1 is much higher than proximal part, this is also consistent with the hypothesis that recombination is a source of mutations.

The mammalian genomes are undergoing a large excess of GC→AT substitutions over AT→GC substitutions (Eyre-Walker 1999; Duret, Semon, and Piganeau 2002). In a set of sequences that were generally stationary, it was found that the frequencies of GC→AT substitution were significantly higher than those of AT→GC in three mammalian species: human, mouse and bovid (Eyre-Walker 1999). This discrepancy between inter- and intra-specific patterns suggests that GC alleles might have a higher probability of fixation than AT alleles. But, by comparing directional base changes within and between species we have found a similar ratio of AT→GC: GC→AT changes between inter- and intra-specific except for *L254915*, which has excessive AT→GC changes in human lineage but is stationary between species. This similarity of inter- and intra-specific AT→GC: GC→AT ratio is highly consistent with the neutral expectation. Since the human PAR1 are homologous to the PAR of other great apes (Ellis et al. 1990; Graves et al. 1998) and the GC-content for each studied region in human is nearly equal to the other two studied organisms we suggest that mutation plays a much lager role in the evolution of base composition in the human PAR1, possibly not to increase the GC-content but to maintain the GC-rich region.

However, how does the GC-content of these PAR1 genes increase to a much higher level? The BGC model is most likely to explain this process. Indeed, different observations support

the BGC model: experiments in mammalian cells have shown that the repair of mismatches is biased in favor of GC; and genes that frequently undergo gene conversion are GC rich. And there is an overall positive correlation between the GC-content and the recombination rate (reviewed in Galtier et al. 2001). But the only evidence in this study is *L254915*, which has excessive AT→GC than GC→AT substitutions in the human lineage in accordance with the prediction of BGC model. This region was proposed as a newly evolving recombination/gene conversion hotpot in the hominoid genomes (Yi and Li 2005). As it has accumulated excessive AT to GC substitutions, Yi and Li (2005) suggested that this is possibly a result of BGC. This implies that the BGC model may account for the increase of GC-content in the PAR1 but has no effect when stationary. Since the GC-content of these PAR1 genes has a significantly positive correlation with their regional recombination rate, we conclude that the high GC-content of PAR1 genes is a result of the BGC model and then maintained by this mutation model.

Acknowledgements

We thank Hong-Xia Pang, Xin Li and Jing-Jing Li for a helpful discussion and critical reading of the manuscript.

References

Arndt, P. F., C. B. Burge, and T. Hwa. 2003. DNA sequence evolution with neighbor-dependent mutation. J. Comput. Biol. **10**:313–322.

Brown, T. C. and J. Jiricny. 1987. A specific mismatch repair event protects mammalian cells from loss of 5-methylcytosine. Cell **50**:945-950.

Brown, T. C. and J. Jiricny. 1988. Different base/base mispairs are corrected with different efficiencies and specificities in monkey kidney cells. Cell **54**:705-711.

Brown, T. C. and J. Jiricny. 1989. Repair of base-base mismatches in simian and human cells. Genome **31**:578-583

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

Bernardi, G. 2000. Isochores and the evolutionary genomics of vertebrates. Gene **241**:3–17.

Birdsell, J. A. 2002. Integrating genomics, bioinformatics, and classical genetics to study the effects of recombination on genome evolution. Mol. Biol. Evol. **19**:1181–1197.

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

Casane, D., S. Boissinot, B. H. Chang, L. C. Shimmin and W. Li. 1997. Mutation pattern

variation among regions of the primate genome. J. Mol. Evol. **45**:216-226.

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.

Duret, L., Semon, M., Piganeau, G., Mouchiroud, D., and Galtier, N. 2002. Vanishing GC-rich isochores in mammalian genomes. *Genetics* **162:** 1837-1847.

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

Eyer-Walker. 1993. Recombination and mammalian genome evolution. Proc. Biol. Sci. **252**:237–243.

Eyer-Walker. 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.

Ebersberger, I and Meyer, M. 2005. A genomic region evolving toward different GC contents in humans and chimpanzees indicates a recent and regionally limited shift in the mutation pattern. Mol. Biol. Evol. **22**:1240-1245.

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.

Francino, M. P., and H. Ochman. 1999. Isochores result from mutation not selection. Nature **400**:30-31.

Fullerton, S.M., Bernardo Carvalho, A., and Clark, A.G. 2001. Local rates of recombination are positively correlated with GCcontent in the human genome. Mol. Biol. Evol. **18**: 1139–1142.

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. and Gerrard, D.T. 2003. High mutation rates in humans and ape pseudoautosomal genes. Gene **317**: 67–77.

Filatov, D. A. 2004. A gradient of silent substitution rate in the human pseudoautosomal region. Mol. Biol. Evol. **21**: 410-417.

Graves, J.A.M., Wakefield, M.J., and Toder, R. 1998. The origin and evolution of the pseudoautosomal regions of human sex chromosomes. Hum. Mol. Genet. **7**: 1991–1996.

Gerton, J. L., J. Derisi, R. Shroff, M. Lichten, P. O. Brown, and T. D. petes. 2000. Global mapping of meiotic recombination hotspots and coldspots in the yeast *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA **97**:11383–11390.

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

Galtier, N. 2004. Recombination, GC-content and the human pseudoautosomal boundary paradox. Trends Genet. **20**:347–349.

Hurst, L. D., C. F. Brunton, and N. G. Smith. 1999. Small introns tend to occur in GC-rich regions in some but not all vertebrates. Trends Genet. **15**:437–439.

Huang, S.-W., R. Friedman, A. Yu and W.-H. Li. 2004. How strong is the mutagenecity of recombination in mammals ? Mol. Biol. Evol. **22**: 426-431.

Kong, A., D. F. Gudbjartsson, G. M. Jonsdottir, S. A. Gudjonsson, B. Richardsson, S. Sigurdardottir, J. Barnard, B. Hallbeck, G. Masson, A. Shlien, et al. 2002. A highresolution recombination map of the human genome. Nat. Genet. **31**:241–247.

Lien, S., J. Szyda, B. Schechinger, G. Rappold and N. Arnheim. 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.

Lander, E. S., Linton, L. M., Birren, B. , Nusbaum, C., Zody, M. C., Baldwin, J. et al. (2001). Initial sequencing and analysis of the human genome. Nature. **409**:860-921.

Marais, G., Mouchiroud, D., and Duret, L. 2001. Does recombination improve selection on codon usage? Lessons from nematode and fly complete genomes. Proc. Natl. Acad. Sci. USA **98**: 5688–5692.

Montoya-Burgos, J.I., Boursot, P., and Galtier, N. 2003. Recombination explains isochores in mammalian genomes. Trends Genet. **19**: 128–130.

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

Meunier, J., and L. Duret. 2004. Recombination drives the Evolution of GC-content in the Human Genome. Mol. Biol. Evol. **21**:984-990.

Strathern, J. N., B. K. Shafer, and C. B. McGill. 1995. DNA synthesis errors associated with double-strand-break repair. Genetics **140**:965-72.

Smith, N. G., and A. Eyre-Walker. 2001. Synonymous codon bias is not caused by mutation bias in g+c-rich genes in humans. Mol. Biol. Evol. **18**:982-986.

S Kumar, K Tamura, and M Nei. 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Brief Bioinform. **5**:150-163.

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

Takano-Shimizu, T. 2001. Local changes in GC/AT substitution biases and in crossover frequencies on *Drosophila* chromosomes. Mol. Biol. Evol. **18**:606–619.

Wolfe, K. H., P. M. Sharp and W. H. Li. 1989. Mutation rates differ among regions of the

mammalian genome. Nature **337**:283-285.

Williams, E. J., and L. D. Hurst. 2000. The proteins of linked genes evolve at similar rates. Nature **407**:900–903.

Yi, S., T. J. Summers, N. M. Pearson and W.-H. Li. 2004. Recombination Has Little Effect on the Rate of Sequence Divergence in Pseudoautosomal Boundary 1 Among Humans and Great Apes. Genome Res. **14**:37-43.

Yi, S and W. –H. Li. 2005. Molecular evolution of recombination hotspots and highly recombining pseudoautosomal regions in hominoids. Mol. Biol. Evol. **22**:1223 - 1230.

Table1

p-PAR Regions Analyzed in this study

Regions	GC[a] in human	Length	Kb to PAB[b]	Recombination rate[c]
PAB	43.2 (41.7)	1520	0	14.35
XG	38.8 (37.9)	930	7	14.35
MIC2	37.9 (37.0)	2192	60	14.35
L254916	45.2 (44.3)	1990	156	14.9
L254915	47.3 (44.3)	1750	200	14.9
DHRSXY	41.9 (40.1)	1845	350	14.9
ASMT	54.5 (49.5)	855	910	17.95
ASMTL	58.2 (54.9)	1245	1095	35.71
SHOX	62.9 (56.8)	666	1730	29.19

[a] Data in the parentheses are the GC-content after the exclusion of CpGs.

[b] Region positions were determined according to the UCSC Genome Brower.

[c] Recombination rate was recalculated from Lien et al. 2000.

Table2

Recombination rate of the human PAR1

Markers	Physical Map (kb)		Genetic Map (cM)	Recombination rate[a] (cM/Mb)	
	Lien	New		Lien	New
TEL	0	0	0		
DXYS201	468	540.646	7.2	15.38	13.3
DXYS15	699	756.466	13.5	27.27	29.19
DXYS218	831	910.388	17.7	31.82	27.29
GGAT3F08	1162	1437.086	30.2	37.76	23.73
DXYS234	1597	1711.530	40.0	22.53	35.71
DXYS85	1850	1951.099	44.3	17	17.95
MIC2	2831	2602.310	54.0	12.86	14.9
PAB	2881	2692.881	55.3	26	14.35

[a] Recombination rate refers to recombination per unit of physical distance, compared to the genome-average rate of 1 cM/Mb. Lien and New are stand for data from Lien et al. 2000 and this study, respectively.

Table3

Substitution rate for each region

Regions	Total tree length[a]	Divergence[b] K ± SD	Length[c]	Mutations observed[d]
PAB	0.0485	0.033 ± 0.005	1535	50
XG	0.0329	0.030 ± 0.006	930	27
MIC2	0.0351	0.026 ± 0.004	2195	57
L254916	0.0327	0.027 ± 0.004	1992	52
L254915	0.0461	0.040 ± 0.005	1712	66*
DHRSXY	0.0794	0.066 ± 0.006	1847	115***
ASMT	0.0849	0.074 ± 0.010	857	61***
ASMTL	0.0978	0.077 ± 0.008	1245	89***
SHOX	0.0885	0.063 ± 0.010	667	40***

[a] Tree length were calculated using three species NJ tree.

[b] Divergence between Human and Orangutan.

[c] Length of sequence used to analyze Human-orangutan pairwise distance.

[d] Significantly more than expected from Poisson distribution assuming 3% average divergence (***) P<0.0001, (**) P<0.001, (*) P<0.05.

Table4

AT→GC and GC→AT changes within and between species

Regions	$N_{AT\rightarrow GC}$:$N_{GC\rightarrow AT}$[a]	p-value	hum_{GC}, chi_{AT}[b]	hum_{AT}, chi_{GC}	p-value
PAB	5: 4	0.739	12	13	0.841
XG	3: 2	0.655	3	4	0.705
MIC2	7: 5	0.564	14	16	0.715
L254916	6: 1	0.059	10	6	0.317
L254915	14: 2**	0.003	16	8	0.102
DHRSXY	8: 6	0.593	21	14	0.237
ASMT	2: 4	0.414	6	5	0.763
ASMTL	8:11	0.491	19	24	0.446
SHOX	2: 7	0.096	12	11	0.835
total	55: 42	0.187	113	101	0.412

[a] $N_{AT\rightarrow GC}$ is the number of AT→GC changes in human lineage and $N_{GC\rightarrow AT}$ is number of reverse process. p-value from two-tailed binomial test.

[b] hum_{GC}, chi_{AT} is the number of positions fixed for GC in human that are AT in chimpanzee. hum_{AT}, chi_{GC} is the converse.

Table 5
Comparative data sets from previous studies

Regions	Yi and Li 2005			Filatov 2004		
	GC in human	Lien[a]	New[a]	GC in human	Lien	New
PAB	-	-	-	43	26	14.35
XG	40.7	26	14.35	43	26	14.35
MIC2	34.3	26	14.35	38	26	14.35
L254916	-	-	-	45	12.86	14.9
L254915	44.9	12.8	14.9	47	12.86	14.9
DHRSXY	42.1	12.8	14.9	42	12.86	14.9
ASMT	51.2	12.8	14.9	56	17	17.95
ASMTL	-	-	-	56	22.53	35.71
SHOX	-	-	-	64	27.27	29.19
AY181056	54.9	37.76	23.73	-	-	-
AY181054	64.3	37.76	23.73	-	-	-
PPP2R3L	-	-	-	67	15.38	13.3

[a] Lien and New are stand for recombination rate from Lien et al. 2000 and this study, respectively

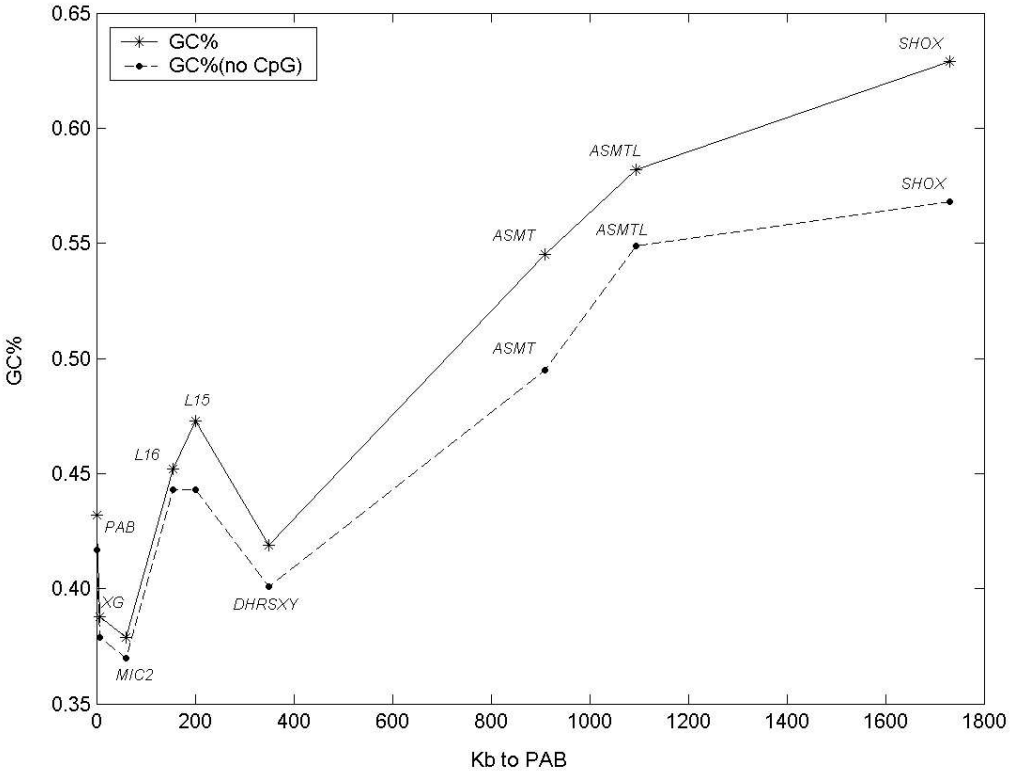

Fig.1. The gradient of GC-content in the human PAR1. The solid line joins the data points for all the intron sites included; the dashed line is the GC-content after exclusion of CpG.