BIOINFORMATICS DISCOVERY NOTE Vol. 23 no. 12 2007, pages 1441–1443
doi:10.1093/bioinformatics/btm137

Genetics and population analysis

Human recombination rates are increased around accelerated conserved regions—evidence for continued selection?

Jan Freudenberg[1,*], Ying-Hui Fu[1] and Louis J. Ptáček[1,2]

[1]University of California San Francisco, Department of Neurology, Institute of Human Genetics, 1550 4th Street, San Francisco, CA 94158-2922 and [2]Howard Hughes Medical Institute, University of California San Francisco, USA

Received on November 30, 2006; revised on March 30, 2007; accepted on April 2, 2007

Advance Access publication April 26, 2007

Associate Editor: Keith Crandall

ABSTRACT

Motivation: We hypothesized that recombination rates might be increased at genetic loci that are subject to more intense selection. Here, we test this hypothesis by using a recently published set of accelerated conserved regions and fine-scale recombination rate estimates provided by the HapMap project.

Results: We observed that fine-scale recombination rates are increased around conserved noncoding regions that show accelerated evolution in human or chimp, as compared to noncoding regions showing accelerated evolution in mouse and those being conserved between human and fugu. Recombination rates around hominid accelerated conserved regions (ACRs) are furthermore increased as compared to exonic regions. On the other hand, GC-content is reduced around ACRs, excluding a major confounding influence of GC-content on the observed variation in recombination rate.

Conclusion: Our observations indicate that selection intensity could be an important determinant of local recombination rate variation and that continued positive selection might act at many ACR loci. Alternatively, a confounding factor needs to be found that causes a congruent signal in recombination rate estimates based on human polymorphism data and in the comparative genomic data. Researchers who consider the explanation involving selection as more likely may expect more common functional sequence variants at ACRs in genetic association studies.

Contact: jan.freudenberg@ucsf.edu

Supplementary information: Supplementary data are available at *Bioinformatics* online.

In a recent study, we noticed that recombination hotspots are enriched around channel activity or developmental and particularly neurodevelopmental genes (Freudenberg *et al.*, 2007). The earlier observation that these functional categories had shown accelerated protein evolution in hominids (Dorus *et al.*, 2004; International _Chimp_Genome_Consortium, 2005) had led us to the hypothesis that selection intensity could be a determinant of recombination hotspot location. Intriguingly, the neurodevelopmental gene category 'cell adhesion' not only belonged to the 30 categories most strongly associated with hotspot predictions, but additionally was recently reported to be enriched near conserved noncoding regions that show accelerated evolution in hominids (Prabhakar *et al.*, 2006). Therefore, we hypothesized that recombination rates might be increased around hominid accelerated conserved regions (ACRs).

In order to test this hypothesis, we retrieved the recently published set of conserved noncoding regions that show evidence for accelerated evolution in human, chimp or mouse (Prabhakar *et al.*, 2006). Furthermore, we retrieved another recently published genome-wide list of noncoding regions that are conserved between human and fugu (Pennacchio *et al.*, 2006). The former set of ACRs was found associated with neuronal cell adhesion function and might have a regulatory role (Prabhakar *et al.*, 2006). The latter set of conserved regions (CRs) can be viewed as strong candidates for acting as gene enhancer (Pennacchio *et al.*, 2006). To compare those noncoding regions with protein coding regions, we also retrieved a list of central exons that belong to 1000 randomly picked human autosomal genes. We next calculated for each region the 30 Kilobase (Kb) interval centered on the respective ACR, CR or exon. We discarded all overlapping intervals in order to obtain signals that are specific for the respective type of region. This resulted in 6342 disjoint autosomal intervals that are located around 672 human ACRs, 721 chimp ACRs, 3293 mouse ACRs, 744 human-fugu CRs and 912 exonic regions. Then, we retrieved fine-scale recombination rate estimates between all HapMap Phase I SNPs (International _HapMap_Consortium, 2005; McVean *et al.*, 2004) from the UCSC database (http://genome.ucsc.edu) (Karolchik *et al.*, 2004). For each interval, we now calculated its average fine-scale recombination rate estimate by weighting the recombination rate of sub-intervals by their relative size.

When comparing recombination rate estimates of these intervals, we found a significant difference between genomic intervals around ACRs, CRs and exons ($P = 5.03 \times 10^{-9}$ by two-sided Kruskal–Wallis test) (Fig. 1). Consistent with the reported parallel acceleration of conserved noncoding sequence evolution near the same type of genes in both species (Prabhakar *et al.*, 2006), no significant difference between human ACRs and chimp ACRs ($P = 0.53$ by two-sided Mann–Whitney test) exists. Furthermore, mouse ACRs may be considered similar to human-fugu CRs in their relationship to hominid ACRs, because they were predicted on a

*To whom correspondence should be addressed.

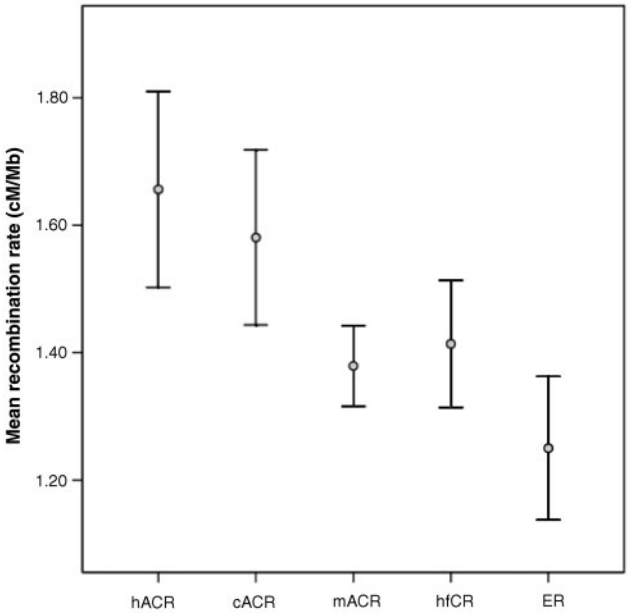

Fig. 1. Mean Fine-scale recombination rate estimates of 30 Kb sequence intervals that are located around conserved noncoding regions with accelerated evolution in human (hACR), chimp (cACR), mouse (mACR), human-fugu conserved regions (hfCRs) and exonic regions (ER). Error bars denote 2 SEM.

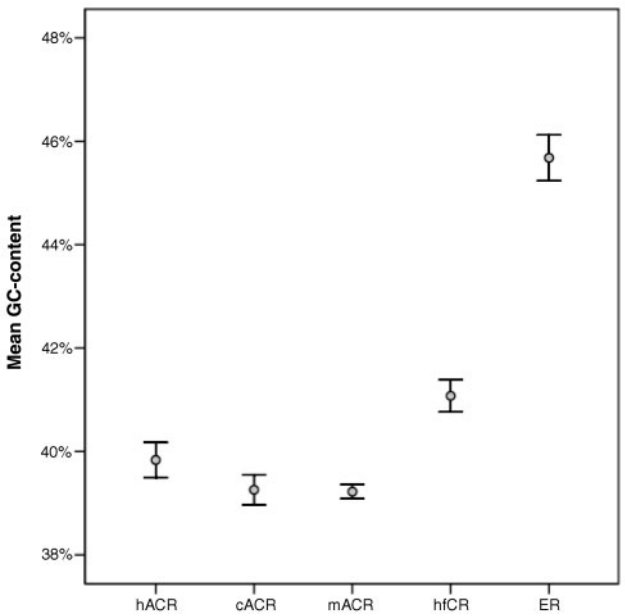

Fig. 2. Mean GC-content of 30 Kb sequence intervals around conserved noncoding regions with accelerated evolution in human (hACR), chimp (cACR), mouse (mACR), human-fugu conserved regions (hfCRs) and around central exons (ER) of human genes. Error bars denote 2 SEM.

background of multi-species conservation and do not show any evidence for accelerated evolution in hominids. Accordingly, we do not observe any difference in recombination rate estimates between mouse ACRs and human-fugu CRs ($P = 0.74$ by two-sided Mann–Whitney test). On the other hand, higher recombination rate estimates are displayed by the joined group of hominid ACRs as compared to the joined group of mouse ACRs and human-fugu CRs ($P = 9.55 \times 10^{-7}$ by two-sided Mann–Whitney test). The group of hominid ACRs furthermore displays higher recombination rate estimates than the exonic regions ($P = 2.34 \times 10^{-10}$ by two-sided Mann–Whitney test). Taken together, these observations strongly support that human recombination rates are increased around hominid ACRs, consistent with the hypothesis that selection intensity is an important determinant of local recombination rate variation. If the causal direction of this relationship was exclusively into the opposite direction, strong constraints would be imposed on the evolution of a genomic region by its chromosomal location. To further compare noncoding ACRs with those exon regions that show evidence for accelerated evolution, more work is required.

We next asked if the increase of recombination rates around hominid ACRs is correlated with an increase of GC-content in those regions. To this end, we calculated for each included genomic interval its average GC-content. Contrary to the expectation of increased GC-content in regions of high recombination (Meunier and Duret, 2004), we found low GC-content around ACRs and high GC-content around exons ($P = 1.5 \times 10^{-169}$) (Fig. 2). Hominid ACRs display a significantly lower GC-content than exonic regions ($P = 6.6 \times 10^{-117}$), but no significant difference to the joined group of mouse ACRs and CRs ($P = 0.263$). Thus, the above mentioned increase of recombination rates around hominid ACRs cannot be explained by a confounding influence of local GC-content variation.

Earlier studies had proposed biased gene conversion as an explanation for the correlation between GC-content and recombination (Fearnhead and Smith, 2005; Meunier and Duret, 2004; Spencer *et al.*, 2006). However, there may exist additional factors that contribute to this relationship, for instance the relation of expression to both selection (Subramanian and Kumar, 2004) and GC-content (Vinogradov, 2005). GC-content at noncoding intervals might be lower due to different transcriptional properties. Furthermore, it is known that most recombination events take place in recombination hotspots (Fearnhead and Smith, 2005; McVean *et al.*, 2004; Templeton *et al.*, 2000). The genomic location of hotspots is hardly conserved between species (Ptak *et al.*, 2005), appears to have a population specific component (Fearnhead and Smith, 2005) and could be heterogeneous across a population (Calabrese, 2007; Coop *et al.*, 2007). This gives new room to a potential relationship between local recombination rate and selection.

We focused the analysis on 30 Kb intervals, because this scale falls below the average spacing of recombination hotspots but may easily accommodate single hotspots (International_HapMap_Consortium, 2005; McVean *et al.*, 2004). Nevertheless, it might be worth mentioning that a congruent pattern can be found, when repeating the analysis using 20 Kb or 50 Kb intervals instead of 30 Kb intervals (Supplementary Figs).

In conclusion, we argue that human recombination rates might be influenced by local selection intensity. This explanation of our data assumes that due to the Hill–Robertson effect (Hill and Robertson, 1966), selection can favor increased recombination rates at genetic loci that are subject to more intense directional selection. Since recombination rate estimates are based on the HapMap SNP data and reflect the past 200 000 years of human evolution, this explanation suggests continued positive selection around ACR loci in humans. Continued positive selection is a plausible conclusion because there is no reason to assume that positive selection generates a signature at these loci over millions of years and then stops with the emergence of present humans. Based on this latter argument, one could also reason in the opposite direction: the fact that recombination rate estimates are increased around ACRs supports the notion that local recombination rates are traits that evolve under the influence of natural selection. If alternatively searching for a neutral explanation, a confounding factor is required that causes a congruent signal in the comparative genomic data and the human polymorphism data, the former acquired over millions of years of evolution and the latter acquired over the short period of present human evolution. Researchers who consider the explanation involving selection as more likely may expect a relative excess of common functional variation around ACRs. This may be relevant in the design of genetic association studies and provides an experimentally testable prediction of our hypothesis.

ACKNOWLEDGEMENTS

This study was supported by the Howard Hughes Medical Institute and a Sandler grant for neurogenetics. We thank Yun Freudenberg–Hua and the anonymous review for their comments.

Conflict of Interest: none declared.

REFERENCES

Calabrese,P. (2007) A population genetics model with recombination hotspots that are heterogeneous across the population. *Proc. Natl Acad. Sci. USA*, **104**, 4748–4752.

Coop,G. *et al.* (2007) Transmission distortion in recombination hotspots. *PLoS Genet.*, **3**, e35.

Dorus,S. *et al.* (2004) Accelerated evolution of nervous system genes in the origin of Homo sapiens. *Cell*, **119**, 1027–1040.

Fearnhead,P. and Smith,N.G. (2005) A novel method with improved power to detect recombination hotspots from polymorphism data reveals multiple hotspots in human genes. *Am. J. Hum. Genet.*, **77**, 781–794.

Freudenberg,J. *et al.* (2007) Enrichment of HapMap recombination hotspot predictions around human nervous system genes. submitted.

Hill,W.G. and Robertson,A. (1966) The effect of linkage on limits to artificial selection. *Genet. Res.*, **8**, 269–294.

International_Chimp_Genome_Consortium (2005) Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature*, **437**, 69–87.

International_HapMap_Consortium (2005) A haplotype map of the human genome. *Nature*, **437**, 1299–1320.

Karolchik,D. *et al.* (2004) The UCSC table browser data retrieval tool. *Nucleic Acids Res.*, **32**, D493–D496.

McVean,G.A. *et al.* (2004) The fine-scale structure of recombination rate variation in the human genome. *Science*, **304**, 581–584.

Meunier,J. and Duret,L. (2004) Recombination drives the evolution of GC-content in the human genome. *Mol. Biol. Evol.*, **21**, 984–990.

Pennacchio,L.A. *et al.* (2006) In vivo enhancer analysis of human conserved non-coding sequences. *Nature*, **444**, 499–502.

Prabhakar,S. *et al.* (2006) Accelerated evolution of conserved noncoding sequences in humans. *Science*, **314**, 786.

Ptak,S.E. *et al.* (2005) Fine-scale recombination patterns differ between chimpanzees and humans. *Nat. Genet.*, **37**, 429–434.

Spencer,C.C. *et al.* (2006) The influence of recombination on human genetic diversity. *PLoS Genet.*, **2**.

Subramanian,S. and Kumar,S. (2004) Gene expression intensity shapes evolutionary rates of the proteins encoded by the vertebrate genome. *Genetics*, **168**, 373–381.

Templeton,A.R. *et al.* (2000) Cladistic structure within the human lipoprotein lipase gene and its implications for phenotypic association studies. *Genetics*, **156**, 1259–1275.

Vinogradov,A.E. (2005) Non-coding DNA, isochores and gene expression: nucleosome formation potential. *Nucleic Acids Res.*, **33**, 559–563.