Note

Positive Selection Can Create False Hotspots of Recombination

Floyd A. Reed[1] and Sarah A. Tishkoff

Department of Biology, University of Maryland, College Park, Maryland 20742

Manuscript received October 8, 2005
Accepted for publication December 23, 2005

ABSTRACT

Simulations of positive directional selection, under parameter values appropriate for approximating human genetic diversity and rates of recombination, reveal that the effects of strong selective sweeps on patterns of linkage disequilibrium (LD) mimic the pattern expected with recombinant hotspots.

IN several cases, the local distribution of meiotic recombination in humans is nonuniform and concentrated into small regions, on the order of 1 kbp in size, termed recombinant hotspots (for reviews see PETES 2001; ARNHEIM *et al.* 2003; DE MASSY 2003; WALL and PRITCHARD 2003; KAUPPI *et al.* 2004). These recombinant hotspots appear to be a common feature in the human genome (CRAWFORD *et al.* 2004; MCVEAN *et al.* 2004) and are rarely shared between humans and closely related species (WALL *et al.* 2003; PTAK *et al.* 2004, 2005; WINCKLER *et al.* 2005). This suggests recombinant hotspots may be rapidly evolving and species specific. These hotspots contribute to the "haplotype block" pattern of genetic regions with high linkage disequilibrium (LD; nonindependent associations between alleles at different positions) separated by boundaries of low LD, which is actively being characterized to optimize marker choice for association mapping studies (INTERNATIONAL HAPMAP CONSORTIUM 2003, 2005; TISHKOFF and VERRELLI 2003). Knowledge of the distribution of LD is critical to mapping the genetic basis of complex phenotypes (WEISS and CLARK 2002). Methods have been developed to detect recombinant hotspots from DNA sequence data, which utilize patterns of LD to infer the existence of these hotspots (*e.g.*, CHAKRAVARTI *et al.* 1984; LI and STEPHENS 2003; ZHANG *et al.* 2004; for review see STUMPF and MCVEAN 2003). In particular, LI and STEPHENS (2003) developed a coalescent-based "product of approximate conditionals" model, which uses the distribution of haplotypes to estimate the likelihood of the underlying recombination rate.

Positive directional selection, in which a new mutant rises in frequency and quickly fixes in a population (*i.e.*, a selective sweep), can be rapid on an evolutionary timescale and/or population specific (for reviews see ANDOLFATTO 2001; AQUADRO *et al.* 2001; SCHLÖTTERER 2002a; BAMSHAD and WOODING 2003). One predicted effect of this type of selection on a sample of DNA sequences is an increase in LD in regions flanking the site undergoing selection (KIM and STEPHAN 2002; PRZEWORSKI 2002), but a reduction of LD across the site of selection (KIM and NIELSEN 2004). This dual pattern may not be intuitive at first, so consider that reducing the genealogical history of a sequence reduces the number of recombination events, thus generally increasing LD in the region. However, for linked genetic variation to be present immediately after a selective sweep, it must either be a new mutation, and therefore rare and contribute little to overall LD, or have experienced recombination during the sweep with the target of selection. This suggests that haplotypes with LD between polymorphic alleles that span the target of selection will not persist beyond the fixation of the selected allele. Here we do not address gene conversion, which could preserve LD over the site of selection. This pattern of two regions of high LD, separated by low LD, is similar to the pattern of LD expected with a recombinant hotspot. Furthermore, the speed and species specificity of selective sweeps may also mimic the species-specific distribution of recombinant hotspots.

To explore the possible effect of selective sweeps on the inference of recombinant hotspots, we simulated positive directional selection of varying intensities (using SelSim 2.1, SPENCER and COOP 2004) and applied hotspot detection software (Hotspotter 1.0, LI and STEPHENS 2003) to the resulting simulated sequences. The parameter values for the simulations were picked to

[1]*Corresponding author:* 4231 Biology, Department of Biology, University of Maryland, College Park, MD 20742.   E-mail: freed@umd.edu
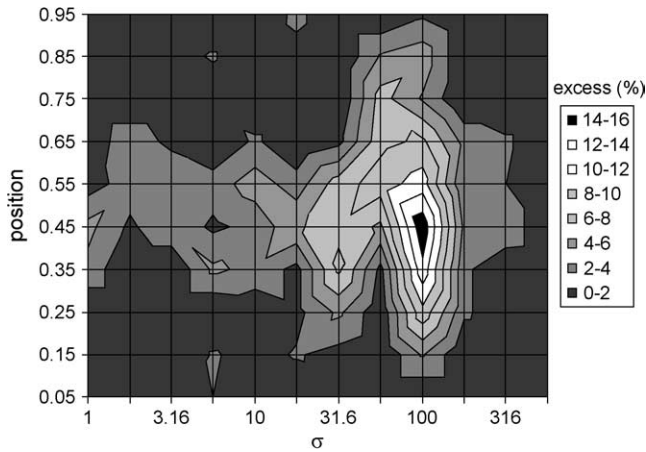
FIGURE 1.—A plot of the relative FPR of inferring significantly elevated local recombination rates along a simulated sequence of recombining DNA. The increase (in percentage) under a selective sweep scenario is plotted relative to the FPR from neutral simulations. The position of the site under positive selection was fixed at 0.45. Each population sample consists of 100 sequences. The recombination parameter $\rho \equiv 4Nr$ (for a diploid, where r is the per-generation recombination rate) was uniform and set to a value of 10 over the total region. The mutation parameter $\theta \equiv 4N\mu$ (where μ is the per-generation mutation rate) was also set to 10 for the region. A stochastic model of positive selection was used and the fixation of the selected allele was modeled as just completing in the sampled generation. One hundred replicates at each value of selection, σ, were generated and analyzed assuming a single hotspot 0.1 units wide of fixed location (0–0.1, 0.1–0.2, ..., 0.9–1). A significantly elevated recombination rate was called when the lower 95% confidence interval of the local estimated recombination rate was higher than the background recombination rate estimate (this includes both hotspots and "warm" spots). In general, FPR excesses higher than 4 (counts out of 100) are significantly elevated ($P < 0.05$), assuming a Poisson distribution of false positives with a mean equal to the neutral rate, which is an average of six false positives.

approximate a 10-kbp sequence of DNA from a human population sample. We found that for strong positive selection ($\sigma \approx 100$, where $\sigma \equiv 2Ns$, N is the diploid population size, and s is the relative strength of selection), a locally elevated recombination rate can falsely be inferred in the region of selection with statistical significance 22% of the time, corresponding to a 16% excess over the false positive rate (FPR) from neutral simulations (Figure 1). This selective elevation over the neutral FPR is highly significant ($P = 1 \times 10^{-7}$; see Figure 1 legend). However, as the strength of selection becomes even stronger ($\sigma \geq 300$), there is a rapid drop in the FPR—probably due to a loss in power associated with a paucity of genetic variation remaining immediately after a strong selective sweep. The patterns of LD resulting from positive selection can produce locally elevated estimated rates of recombination that are similar to the relative rates reported in the literature (*e.g.*, Figure 2; *cf.* JEFFREYS and NEUMANN 2002; WALL *et al.* 2003; McVEAN *et al.* 2004; PTAK *et al.* 2004; VERRELLI
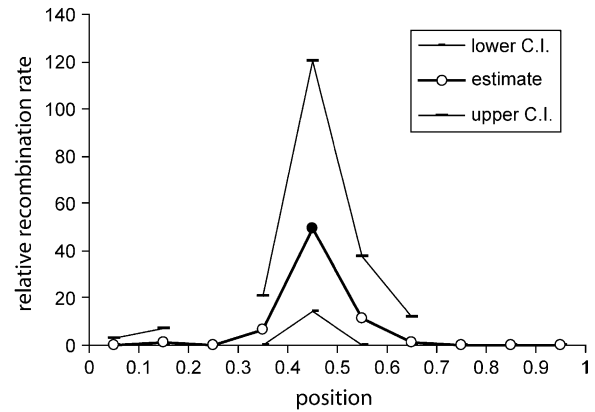

FIGURE 2.—An example of the inferred relative recombination rate from a sample simulated under the conditions described in Figure 1 with $\sigma = 100$. The relative recombination rate between each window, with a width 1/10th of the total sequence, and the remaining region were estimated. This example illustrates how a false hotspot of recombination may be inferred. In this case, the "hotspot" is at position 0.45 and has a local recombination rate estimate 49 times higher than that of the surrounding sequence. The upper and lower 95% confidence intervals are also plotted and a significantly elevated point estimate is represented by a solid circle.

and TISHKOFF 2004; WINCKLER *et al.* 2005). The geometric mean of estimated recombination rates at the site of selection, from 100 replicates at $\sigma = 100$, is 13.25 times higher than the background rate. Furthermore, this elevated FPR can persist for up to N generations ($0.5 \times 2N$ generations) after the selective sweep has ended (Figure 3). Assuming an effective human population size of 10,000 and an average generation time of 25 years, this corresponds to a maximum persistence time of ~250,000 years.

We do not mean to imply that true recombinant hotspots do not exist in humans; they have certainly been verified by experimental means (*e.g.*, HUBERT *et al.* 1994; CULLEN *et al.* 1995; SMITH *et al.* 1998; YIP *et al.* 1999;
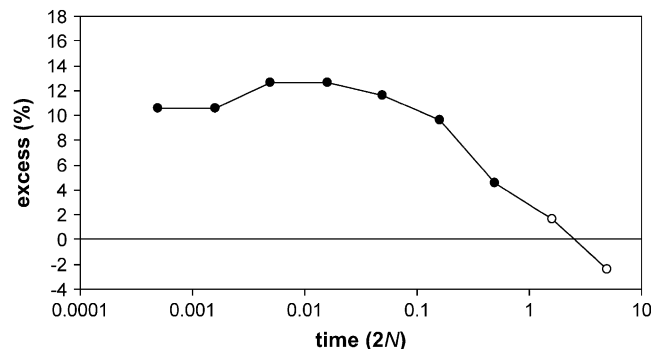

FIGURE 3.—A plot of the excess, over neutrality, of the FPR of inferring a significantly elevated recombination rate at the position of selection *vs.* time in units of $2N$ generations. Solid circles denote a significantly elevated FPR relative to neutral simulations (see Figure 1 legend).

JEFFREYS *et al.* 2001). But we do suggest caution when inferring the existence of hotspots solely on the basis of patterns of LD. The transient nature of positive selection, both over time and between populations, may easily mimic the rapidly evolving nature of recombination in primates. When a hotspot is inferred, it may be useful to also address the relative levels of genetic variation compared to levels of divergence (PTAK *et al.* 2004) to help rule out past positive selection—particularly since recombination may be associated with a mutagenic process (RATTRAY *et al.* 2002; HELLMANN *et al.* 2003) and selective sweeps can quickly remove genetic variation. However, recombinant hotspots and selective sweeps may be linked at a more basic level. There is evidence that hotspot crossover asymmetry can result in a form of meiotic drive (JEFFREYS and NEUMANN 2002), which itself is a "selfish" form of positive selection (for review see REED *et al.* 2005 and references therein). This crossover asymmetry predicts that a derived recombination-suppressing allele will eventually fix in the population (JEFFREYS and NEUMANN 2002), resulting in the co-occurrence of both a recombinant hotspot and a progressing selective sweep.

The possibility exists that inferred recombinant hotspots in gene regions that also appear to have undergone positive selection (*e.g.*, VERRELLI and TISHKOFF 2004) are not due to nonuniform densities of meiotic recombination, but may simply be a by-product of positive selection. In the same vein, estimates of the rate of hotspot sharing between species, based on LD analysis (PTAK *et al.* 2005), may be underestimated, and short-scale LD may be lower than expected (*e.g.*, PRITCHARD and PRZEWORSKI 2001) if recent positive selection plays a significant role. If selective sweeps do make a significant contribution to the patterns of LD in the human genome, then a better understanding of the effects of positive selection may have important implications for projects that characterize LD for association studies, particularly to the extent that selective pressures may have varied among human populations (*e.g.*, HAMBLIN and DI RIENZO 2000; SCHLÖTTERER 2002b; AKEY *et al.* 2004; STORZ *et al.* 2004).

LITERATURE CITED

AKEY, J. M., M. A. EBERLE, M. J. RIEDER, C. S. CARLSON, M. D. SHRIVER et al., 2004 Population history and natural selection shape patterns of genetic variation in 132 genes. PLoS Biol. **2:** 1591–1599.

ANDOLFATTO, P., 2001 Adaptive hitchhiking effects on genome variability. Curr. Opin. Genet. Dev. **11:** 635–641.

AQUADRO, C. F., V. BAUER DUMONT, and F. A. REED, 2001 Genome-wide variation in the human and fruitfly: a comparison. Curr. Opin. Genet. Dev. **11:** 627–634.

ARNHEIM, N., P. CALABRESE and M. NORDBORG, 2003 Hot and cold spots of recombination in the human genome: the reason we should find them and how this can be achieved. Am. J. Hum. Genet. **73:** 5–16.

BAMSHAD, M., and S. P. WOODING, 2003 Signatures of natural selection in the human genome. Nat. Rev. Genet. **4:** 99–111.

CHAKRAVARTI, A., K. H. BUETOW, S. E. ANTONARAKIS, P. G. WABER, C. D. BOEHM et al., 1984 Nonuniform recombination within the human β-globin gene cluster. Am. J. Hum. Genet. **63:** 861–869.

CRAWFORD, D. C., T. BHANGALE, N. LI, G. HELLENTHAL, M. J. RIEDER et al., 2004 Evidence for substantial fine-scale variation in recombination rates across the human genome. Nat. Genet. **36:** 700–706.

CULLEN, M., H. ERLICH, W. KLITZ and M. CARRINGTON, 1995 Molecular mapping of a recombination hotspot located in the second intron of the human *TAP2* locus. Am. J. Hum. Genet. **56:** 1350–1358.

DE MASSY, B., 2003 Distribution of meiotic recombination sites. Trends Genet. **19:** 514–522.

HAMBLIN, M. T., and A. DI RIENZO, 2000 Detection of the signature of natural selection in humans: evidence from the Duffy blood group locus. Am. J. Hum. Genet. **66:** 1669–1679.

HELLMANN, I., I. EBERSBERGER, S. PTAK, S. PÄÄBO and M. PRZEWORSKI, 2003 A neutral explanation for the correlation of diversity with recombination in humans. Am. J. Hum. Genet. **72:** 1527–1535.

HUBERT, R., M. MACDONALD, J. GUSELLA and N. ARNHEIM, 1994 High resolution localization of recombination hot spots using sperm typing. Nat. Genet. **7:** 420–424.

INTERNATIONAL HAPMAP CONSORTIUM, 2003 The international hapmap project. Nature **426:** 789–796.

INTERNATIONAL HAPMAP CONSORTIUM, 2005 A haplotype map of the human genome. Nature **437:** 1299–1320.

JEFFREYS, A., and R. NEUMANN, 2002 Reciprocal crossover asymmetry and meiotic drive in a human recombination hot spot. Nat. Genet. **31:** 267–271.

JEFFREYS, A. J., L. KAUPPI and R. NEUMANN, 2001 Intensely punctuate meiotic recombination in the class II region of the major histocompatibility complex. Nature Genet. **29:** 217–222.

KAUPPI, L., A. J. JEFFREYS and S. KEENEY, 2004 Where the crossovers are: recombination distributions in mammals. Nat. Rev. Genet. **5:** 413–424.

KIM, Y., and R. NIELSEN, 2004 Linkage disequilibrium as a signature of selective sweeps. Genetics **157:** 1513–1524.

KIM, Y., and W. STEPHAN, 2002 Detecting a local signature of genetic hitchhiking along a recombining chromosome. Genetics **160:** 765–777.

LI, N., and M. STEPHENS, 2003 Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. Genetics **165:** 2213–2233.

MCVEAN, G. A. T., S. R. MYERS, S. HUNT, P. DELOUKAS, D. R. BENTLEY et al., 2004 The fine-scale structure of recombination variation in the human genome. Science **304:** 581–584.

PETES, T. D., 2001 Meiotic recombination hot spots and cold spots. Nat. Rev. Genet. **2:** 360–369.

PRITCHARD, J. K., and M. PRZEWORSKI, 2001 Linkage disequilibrium in humans: models and data. Am. J. Hum. Genet. **69:** 1–14.

PRZEWORSKI, M., 2002 The signature of positive selection at randomly chosen loci. Genetics **160:** 1179–1189.

PTAK, S. E., A. D. ROEDER, M. STEPHENS, Y. GILAD, S. PÄÄBO et al., 2004 Absence of the TAP2 human recombination hotspot in chimpanzees. PLoS Biol. **2:** 849–855.

PTAK, S. E., D. A. HINDS, K. KOEHLER, B. NICKEL, N. PATIL et al., 2005 Fine-scale recombination patterns differ between chimpanzees and humans. Nat. Genet. **37:** 429–434.

RATTRAY, A. J., B. K. SHAFER, C. B. MCGILL and J. N. STRATHERN, 2002 The roles of REV3 and RAD57 in double-strand-break-repair-induced mutagenesis in *Saccharomyces cerevisiae.* Genetics **162:** 1063–1077.

REED, F. A., R. G. REEVES and C. F. AQUADRO, 2005 Evidence of susceptibility and resistance to cryptic X-linked meiotic drive in natural populations of *Drosophila melanogaster.* Evolution **59:** 1280–1291.

SCHLÖTTERER, C., 2002a Towards a molecular characterization of adaptation in local populations. Curr. Opin. Genet. Dev. **12:** 683–687.

Schlötterer, C., 2002b   A microsatellite-based multilocus screen for the identification of local selective sweeps. Genetics **160:** 753–763.

Smith, R. A., P. J. Ho, J. B. Clegg, J. R. Kidd and S. L. Thein, 1998   Recombination breakpoints in the human β-globin gene cluster. Blood **92:** 4415–4421.

Spencer, C. C. A, and G. Coop, 2004   SelSim: a program to simulate population genetic data with natural selection and recombination. Bioinformatics **20:** 3673–3675.

Storz, J. F., B. A. Payseur and M. W. Nachman, 2004   Genome scans of DNA variability in humans reveal evidence for selective sweeps outside of Africa. Mol. Biol. Evol. **21:** 1800–1811.

Stumpf, M. P., and G. A. McVean, 2003   Estimating recombination rates from population-genetic data. Nat. Rev. Genet. **4:** 959–968.

Tishkoff, S. A., and B. C. Verrelli, 2003   Role of evolutionary history on haplotype block structure in the human genome: implications for disease mapping. Curr. Opin. Genet. Dev. **13:** 569–575.

Verrelli, B. C., and S. A. Tishkoff, 2004   Signatures of selection and gene conversion associated with human color vision variation. Am. J. Hum. Genet. **75:** 363–375.

Wall, J. D., and J. K. Pritchard, 2003   Haplotype blocks and the structure of linkage disequilibrium in the human genome. Nat. Rev. Genet. **4:** 587–597.

Wall, J. D., L. A. Frisse, R. R. Hudson and A. Di Rienzo, 2003   Comparative linkage disequilibrium analysis of the β-globin hotspot in primates. Am. J. Hum. Genet. **73:** 1330–1340.

Weiss, K. M., and A. G. Clark, 2002   Linkage disequilibrium and the mapping of complex human traits. Trends Genet. **18:** 19–24.

Winckler, W., S. R. Myers, D. J. Richter, R. C. Onofrio, G. J. McDonald *et al.*, 2005   Comparison of fine-scale recombination rates in humans and chimpanzees. Science **308:** 107–111.

Yip, S. P., J. U. Lovegrove, N. A. Rana, D. A. Hopkinson and D. B. Whitehouse, 1999   Mapping recombination hot spots in human phosphoglucomutase (*PGM1*). Hum. Mol. Genet. **9:** 1699–1706.

Zhang, J., F. Li, J. Li, M. Q. Zhang and X. Zhang, 2004   Evidence and characteristics of putative human α recombination hotspots. Hum. Mol. Genet. **13:** 2823–2828.