nature genetics

Fine-scale recombination patterns differ between chimpanzees and humans

Susan E Ptak[1], David A Hinds[2], Kathrin Koehler[1], Birgit Nickel[1], Nila Patil[2], Dennis G Ballinger[2], Molly Przeworski[3,4], Kelly A Frazer[2,4] & Svante Pääbo[1,4]

Recombination rates seem to vary extensively along the human genome. Pedigree analysis suggests that rates vary by an order of magnitude when measured at the megabase scale[1], and at a finer scale, sperm typing studies point to the existence of recombination hotspots[2]. These are short regions (1–2 kb) in which recombination rates are 10–1,000 times higher than the background rate. Less is known about how recombination rates change over time. Here we determined to what degree recombination rates are conserved among closely related species by estimating recombination rates from 14 Mb of linkage disequilibrium data in central chimpanzee and human populations. The results suggest that recombination hotspots are not conserved between the two species and that recombination rates in larger (50 kb) genomic regions are only weakly conserved. Therefore, the recombination landscape has changed markedly between the two species.

Recombination is a fundamental parameter in both evolutionary and human genetics. In humans, empirical methods to study recombination rates are limited by the difficulty of obtaining large pedigrees and the labor-intensiveness of sperm typing. Consequently, we are unlikely to soon obtain direct fine-scale estimates of recombination for the whole genome. As an alternative, researchers have developed statistical methods to infer recombination rates based on patterns of linkage disequilibrium (LD)[3]. These statistical methods estimate a population recombination rate, ρ, from LD data, assuming a constant rate throughout the genomic region of interest. Estimates of ρ equals to $4N_er$ (where N_e is the effective population size and r is the recombination rate per generation) represent the amount of recombination needed in the population to produce the observed levels of LD under a simple model. Statistical methods have also been developed to infer variation in recombination rates in a region[4–7]. Application of these methods to human data confirms that fine-scale recombination rates vary markedly along the genome and suggests that hotspots are present at least every 60–200 kb (refs. 4,7).

We are thus gaining a better understanding of variation in recombination rate along the genome, but little is known about how these rates change temporally. In humans, recombination rates differ among individuals[1,8], and this variation, at least in females, seems to be heritable[9]. These observations suggest that recombination rates could evolve over time. At the fine scale, two recent studies suggest that human hotspots

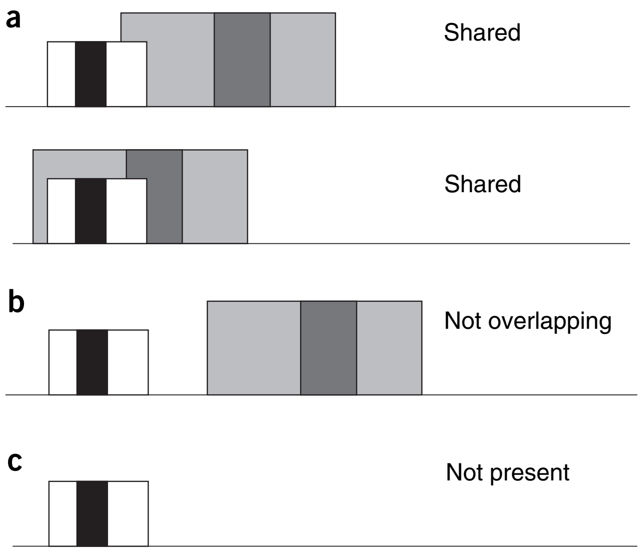

Figure 1 Comparison of hotspots in the two species. Depicted here is a 70-kb window, with the inferred hotspot in chimpanzee sample indicated by the black box and its 95% credible intervals indicated by the white boxes. Similarly, the inferred human hotspot is indicated by larger gray boxes. There are three possibilities: (**a**) the hotspot could be shared in the two species (we treat hotspots as the same if their credible intervals overlap); (**b**) the hotspots could be in different locations (*i.e.*, nonoverlapping credible intervals); or (**c**) the hotspot could be present in only one species. There will be errors in our classification into these three categories. We could infer the location of a hotspot incorrectly and erroneously conclude that there is a shared hotspot or two nonoverlapping hotspots. Alternatively, we could fail to detect a hotspot or incorrectly infer that one is present.

[1]Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6 04103, Leipzig, Germany. [2]Perlegen Sciences, Inc., 2021 Stierlin Court, Mountain View, California 94043, USA. [3]Department of Ecology and Evolutionary Biology, Brown University, 80 Waterman Street, Box G-W, Providence, Rhode Island 02912, USA. [4]These authors contributed equally to this work. Correspondence should be addressed to S.E.P. (ptak@email.eva.mpg.de).

Published online 18 February 2005; doi:10.1038/ng1529

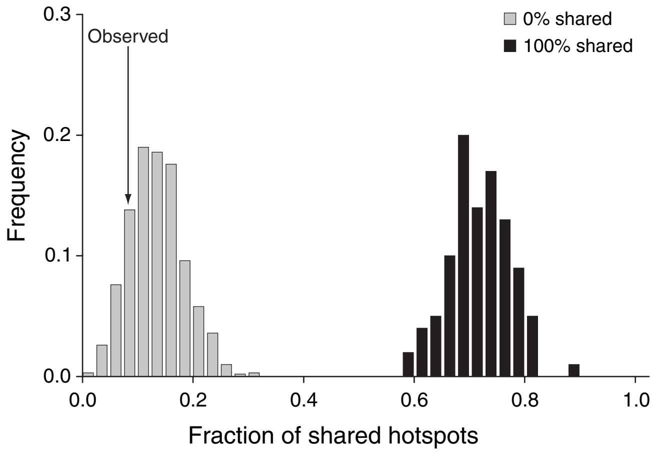

Figure 2 Percentage of hotspots detected as shared under two models. The first model assumes that all hotspots are shared; the second assumes that hotspots are independently distributed along the sequence in humans and chimpanzees. The arrow indicates what was observed as shared (8%). We can reject the model that all hotspots are shared in the two species but cannot reject the model that none are shared.

are not always conserved in other primates: the β-globin hotspot seems to be absent in rhesus macaques (*Macaca mulatta*)[10], and the TAP2 hotspot seems to be absent in chimpanzees (*Pan troglodytes*)[11]. Although these results suggest that fine-scale recombination can evolve rapidly, it is difficult to generalize because they are based on only two hotspots and the two regions were known *a priori* to contain a human hotspot.

In this paper, we investigated how often fine-scale recombination rates differ between humans and chimpanzees over large genomic regions. We used a previously described statistical approach[4] to infer recombination rates in two regions totaling 14 Mb of genomic sequence in chimpanzees and the orthologous regions in humans. This method uses a Bayesian approach to reconstruct haplotypes from a mosaic of previously considered haplotypes. The smaller the pieces needed to reconstruct a given haplotype, the more recombination has occurred. The frequency of breaks along the sequence provides information about the existence of possible hotspots.

Our first question was whether the frequency of hotspots is the same in humans and chimpanzees. In humans, 58% of windows contain an inferred hotspot, whereas only 15% do in chimpanzees. But simulations suggest that the lower fraction in chimpanzee may be explained by differences in power (**Supplementary Tables 1–3** online). Therefore, there is no evidence that the frequency of hotspots differs between the two species.

Next, we asked how often hotspots are shared between humans and chimpanzees (**Fig. 1**). We selected putative chimpanzee hotspots, because our power simulations suggested that both the power and the false-positive rate were lower in the chimpanzee sample, and then asked how often the 95% credible interval for the chimpanzee hotspot overlapped with a hotspot detected in the human sample. By this criterion, only 3 of the 39 inferred hotspots (8%) overlapped in the two species.

Even if all hotspots present in chimpanzees were also present in humans, we would not expect to detect them all as shared. To estimate how many we would expect to find overlapping under this scenario, we considered every window for which a hotspot was inferred in chimpanzees and simulated a corresponding hotspot in humans on the basis of the hotspot location and intensity estimated in chimpanzees. We then determined how many of the hotspots would have been detected. The average in 100 simulations was 72% and the minimum was 59%, well above the observed 8% (**Fig. 2**). Hence, we can reject the hypothesis that all hotspots estimated in the chimpanzee sample are also present in the human sample ($P < 0.01$ from 100 simulations). Although this finding relies on a simplistic model, our conclusion is probably robust to model mis-specification, given the large discrepancy between expected and observed values (**Supplementary Note** online).

We then asked the opposite question: whether we could reject the hypothesis that the hotspots are distributed independently in the two species. We noted where the inferred hotspots are in the 14 Mb of chimpanzee sequence. We then distributed the inferred human hotspots randomly along the sequence, preserving the estimated length of the hotspot, and determined how many of the inferred chimpanzee hotspots overlapped, by chance, with inferred human hotspots. On average, 13% of the hotspots overlapped, more than we observed (**Fig. 2**). Hence, we cannot reject the hypothesis that hotspots are independently distributed along the genomes of the two species ($P = 0.90$ from 1,000 simulations).

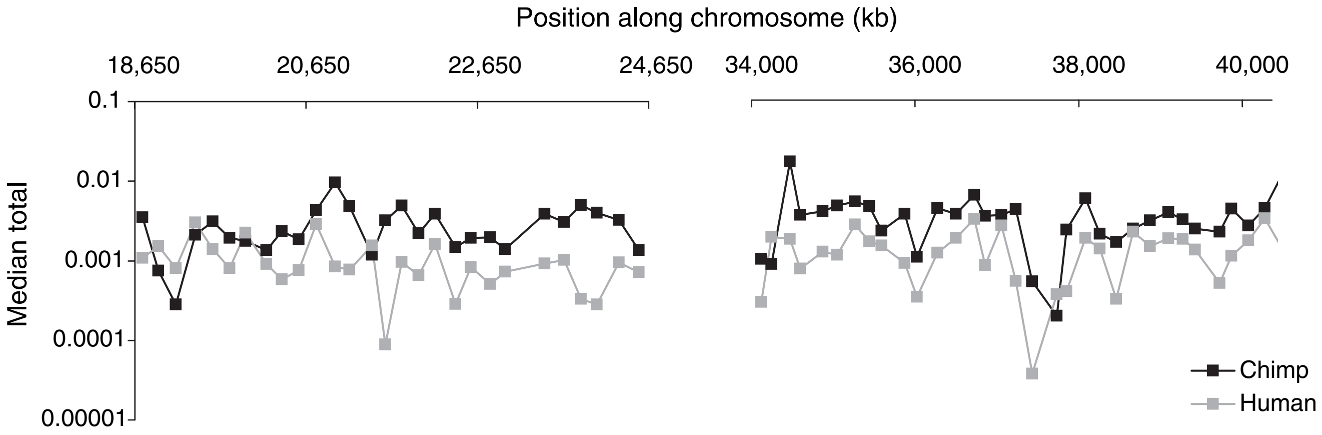

Figure 3 Concordance of total ρ for chimpanzee and human samples. Depicted here are the population recombination estimates (per base pair) averaged over every four windows for the observed data. The position is measured in kb and is the midpoint of the appropriate four windows. Total recombination rates, as estimated at the 50-kb scale, are significantly but weakly correlated in humans and chimpanzees.

Although hotspot location seems not to be conserved between species, it is possible that total recombination rates are under stronger constraint. For example, hotspots may move freely within circumscribed regions of the genome, while the total recombination rate for a region (*i.e.*, the background rate plus recombination in hotspots) remains the same. To evaluate this possibility, we examined the correlation in the total recombination rate between the two species for the 50-kb windows. The total recombination estimates were significantly but weakly correlated between chimpanzees and humans ($r_S = 0.216$, $P = 0.002$, degrees of freedom (d.f.) = 178, one-tailed test; **Fig. 3**). Given that the rates are estimated with error in both species, it is not certain how strong a correlation would be expected if rates were perfectly correlated. To examine this correlation, we generated 100 pairs of data, assuming that total recombination rates were identical in the two species. The average correlation coefficient was 0.69 and the minimum was 0.61, well above the observed value of 0.22 (**Fig. 4**). Hence, we can reject the hypothesis that total recombination rates at the 50-kb scale are perfectly correlated in humans and chimpanzees ($P < 0.01$ from 100 simulations).

The weak conservation in total recombination rates could reflect conservation in the background rates of recombination in the two species. When we examined the correlation in the background recombination rate for the 50-kb windows, we found that the estimated background rates of recombination were also significantly but weakly correlated between the chimpanzee and human samples ($r_S = 0.276$, $P < 0.001$, d.f. = 147, one-tailed test). To examine how strong a correlation would be expected if background rates were perfectly correlated, we generated 100 pairs of simulated data, assuming identical background recombination rates in the two species. The average correlation coefficient was 0.77 and the minimum was 0.70, well above the observed 0.28. Hence, we can reject the hypothesis that background recombination rates at the 50-kb scale are perfectly correlated in humans and chimpanzees ($P < 0.01$ from 100 simulations). This finding and the analogous one for total recombination depend on our model of recombination. Given the large discrepancy between expected and observed values in both cases, however, our conclusion is probably robust to mis-specification of the recombination model (**Supplementary Note** online). Therefore, because background rates are weakly conserved but hotspots are not, we hypothesize that the conservation in total recombination rate reflects the conservation in background recombination rates.

As shown here, the recent development of statistical techniques to identify hotspots and measure local rates of recombination, and the increasing availability of polymorphism data, make it possible to create very detailed recombination maps in humans. In particular, a new data set[12] will make possible a fine-scale genetic map for the entire human genome. We can then compare these local maps with the large-scale maps to begin addressing questions about variation in recombination rate. Homologous data from closely related species, such as chimpanzee, will also help to increase our understanding of the determinants of recombination rates. Already, our data indicate that estimated fine-scale recombination rates are markedly different between primate species that share 95–99% of their primary sequence[13,14]. Therefore, if fine-scale recombination rates in chimpanzees or other primate species need to be known, they will have to be measured anew.

Our results also raise a number of questions about the relationship between small- and large-scale recombination rates and the role of primary sequence in determining recombination rates. We found a weak correlation in background recombination rates at the 50-kb scale. Background rates may be influenced by sequence properties that are largely the same between the two species. Previous studies have found some sequence properties to be correlated with estimated fine-scale recombination rates[4,7]. Given these results, it will be interesting to examine the degree to which recombination rates are conserved at large scales, because the degree of conservation will probably vary with the scale at which recombination is measured. If sequence properties do predict recombination rates, the strength of the correlation between these two (and hence the conservation in rates between species) may depend on the scale at which these properties interact with recombination. In addition, although the exact location of hotspots does not seem to be under constraint, the density of hotspots in a larger area might be. To address these questions of scale and the role of primary sequence, data are needed from many unlinked genomic regions for which we can obtain both fine- and large-scale estimates in the two species. To this end, a large-scale recombination map in chimpanzees would be valuable.

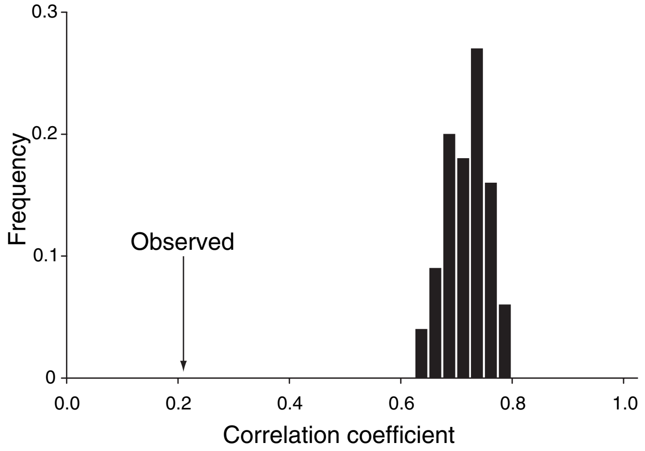

Figure 4 Correlation in total ρ if recombination rates are the same in the two species. The arrow indicates the observed value (0.22), which is significantly below the expected value under this model and significantly above zero.

The lack of concordance between hotspots also raises the question of how hotspots change so quickly between species with such high sequence similarity. Although no sequence motifs are known to predict recombination hotspots in humans, putative sequence motifs have been identified in other species[15,16]. Furthermore, the examination of hotspots in yeast and mice shows that DNA sequence located at or near a hotspot can modify its activity[17,18]. Alleles at a single site can alter the hotspot activity for two human hotspots[19,20]. This suggests that substitutions in *cis* may be responsible for the differences in the hotspot landscape of humans and chimpanzees. Given the large number of changes that seem to have occurred and the low sequence divergence, however, this scenario would require a large number of sites to be potential modifiers of hotspot activity, if hotspots are evolving neutrally. In light of the evidence that hotspot location and intensity are regulated at both local and global scales[2,16], it seems likely that in addition to substitutions in *cis*, changes in *trans* (*e.g.*, the binding affinity to sequence motifs) or in epigenetic factors (*e.g.*, chromatin structure) contribute to the evolution of recombination rates. Much more work is needed to unravel the mechanisms underlying the rapid changes in fine-scale recombination rates.

METHODS

Data collection. We used ~14 Mb of human genotype data[12], consisting of two disjoint regions of approximately equal length on chromosome 21. These regions were chosen such that one segment (build 34 coordinates 18635000–24650000) is 'gene rich' (~90 genes) and the other segment (build 34 coordinates 34050000–41800000) is 'gene-poor' (~2 genes). Together, these two regions represent 41% of the euchromatic chromosome 21. The 11,642 genotyped SNPs were distributed with an average spacing of 1.2 kb per SNP in two segments of roughly equal length in a sample of 23 African Americans, 24 Han Chinese Americans and 24 European Americans from Coriell. We identified ~80% of the SNPs by resequencing 20 haploid chromosomes from the National Institutes of Health Polymorphism Discovery Resource[21]; the remaining 20% consisted of validated SNPs from dbSNP selected to achieve more uniform spacing. We genotyped the human SNPs using high density oligonucleotide arrays designed for SNP genotyping[12]. The genotyping call rate was ~98–99%.

We also resequenced the homologous region (chromosome 22) in eight central chimpanzees (*Pan troglodytes troglodytes*), whose samples were donated by J. Wickings (Unité de Génétique des Écosystèmes Tropicaux, Centre International de Recherches Médicales de Franceville (CIRMF)) and come from the CIRMF (Gabon). Five of the chimpanzees were wild-born. The other three were born at CIRMF (first generation) to parents other than the chimpanzees used in this study (as best as could be established for the fathers). Therefore, the eight chimpanzees are putatively unrelated. To resequence the homologous ~14-Mb region in chimpanzee, we used 1,248 overlapping amplicons on high-density oligonucleotide arrays designed to match the sequence of chimpanzee chromosome 22, based on a previously described technique[21], and discovered 30,611 SNPs with an average spacing of 440 bp per SNP. By resequencing DNA from haploid chimp-hamster fusions, we determined haplotypes for the sequences corresponding to each of the 16 chimpanzee chromosomes empirically. This approach greatly increased our ability to estimate recombination rates, because we avoided the uncertainty of estimating haplotypes from diploid genotypes. For the chimpanzee data from resequencing arrays, 13.6% of bases could not be called.

The human and chimpanzee data vary in a number of aspects, particularly in whether the data were phased or not. Previous researchers obtained very different ρ estimates when they assumed known phase versus when they did not[10]. For the algorithm used here, however, we obtained similar results from simulations (results not shown), as well as for empirical data from the TAP2 hotspot (ref. 11 and S.E.P. and M.P, unpublished results).

Map between humans and chimpanzees. To compare inferred hotspot locations (and windows) in humans and chimpanzees, we aligned each chimpanzee SNP and the surrounding 24 bases against the human genome using BLAST[22], retaining only those SNPs with a unique perfect match. We took the two chimpanzee SNPs that flanked the edge of the hotspot (or window) and calculated the average distance between the location of the SNPs on the chimpanzee and human genomes. We then added this average distance to the chimpanzee genome position to determine the location of the chimpanzee hotspot (window) in humans. This procedure does not exactly preserve the length of the hotspot (or window), although the discrepancy is usually small (90% of the windows are between 49 kb and 51 kb) and the differences in length were accounted for in simulations.

Data analyses: division into windows. To detect hotspots, we divided the 14 Mb into 70-kb windows, with a 20-kb overlap to avoid missing hotspots that lie on the edge of windows. To estimate the recombination rate in a window, we divided the 14 Mb into 50-kb nonoverlapping windows. In the human data, we excluded variable sites with a minor allele frequency less than or equal to 5%, and in the chimpanzee data, we excluded all singletons, because these sites contribute little or no information about LD. Likewise, we removed all sites with more than 75% missing data (this never occurred in the human data). Finally, we considered only windows with at least 20 segregating sites to ensure sufficient power to estimate recombination rates[4]. This procedure left us with 226–260 70-kb windows in the three human populations and the chimpanzee samples to compare putative hotspots and 210 pairs of 50-kb windows to compare recombination rate estimates in the two species.

Estimation method. To estimate recombination rates, we used a method[4] that allows missing data and unknown phase and with which credible intervals are easily computed. This method makes a number of simplistic assumptions about recombination and population history. Results from simulations suggest that this method is relatively robust to deviations from these assumptions[4,6]. To examine the performance of the method in the context of our data collection scheme, we applied the same approach to LD data from the MHC region, for which there exist estimates of the human recombination rate from two studies using sperm-typing (**Supplementary Fig. 1** online).

We ran the simple hotspot model of PHASE[23,24] (version 2.1.1) to obtain estimates of three recombination parameters for each window (option –MR1 1)[4,6]: the population rate of recombination outside the hotspot, referred to as the background ρ; the relative intensity of recombination within the hotspot, λ; and the location of the hotspot. We obtained 1,000 draws from the posterior distribution (option –X10) using only haplotypes with frequency above 0.1 to speed up the run time (option –F0.1; simulations suggested this had little effect on the recombination parameter estimates). For chimpanzees we had haploid sequence data, and so we ran the program assuming known phase but with missing data (option –k999 with –l8).

Estimates of recombination parameters. We computed all recombination parameter estimates based on the hotspot output file, which provides a sample from the posterior distribution of the background ρ per base pair, of λ, and, for $\lambda > 1$, of the location of the hotspot demarcated by a left and right boundary. For every input file, we ran PHASE three times with different seeds[4]. We thus obtained three output files. For each output file, we computed the median and the 5th and 95th percentiles of the 1,000 draws from the posterior distributions of the background ρ, total ρ and λ. We computed the total ρ by a weighted average of the median background ρ and median $\rho\lambda$ (*i.e.*, the recombination rate outside and inside the hotspot, respectively). We estimated the posterior probability that λ was greater than 10 and computed the associated Bayes Factor to assess the statistical support for a hotspot, where a Bayes Factor ≥ 50 was taken to indicate the presence of a hotspot. This is a similar but more stringent criterion than the one previously used[4]. To obtain point estimates for the right and left boundaries of the hotspot, we considered the median of all draws where $\lambda > 1$. To be conservative, we considered the credible intervals of the hotspot to extend from the 5th percentile of the posterior probability for the left boundary to the 95th percentile for the right end. We then combined the three output files, by taking the median value for each statistic across the three but retaining the credible intervals for the median value as our measure of uncertainty.

Because the windows overlap, the same hotspot could be identified in adjacent windows. We therefore treated inferred hotspots with overlapping credible intervals as the same hotspot (**Fig. 1**). We then arbitrarily chose the estimates from the second window to characterize this hotspot (this choice should not introduce a bias). We chose not to average the two estimates, because this would lead to increased precision for a subset of windows. We then excluded the first window from estimates of the frequency of hotspots. In comparing human and chimpanzee hotspots, we treated hotspots with overlapping credible intervals for location as shared between the two species.

Choice of populations. We examined the frequency of hotspots in three human populations (African American, Han Chinese American and European American), but focused only on the African American population in subsequent analyses because the power simulations suggested both that the power to detect hotspots was highest in this population (**Supplementary Tables 1–3** online) and that the estimates of ρ were more accurate and precise (results not shown). The power simulations further suggested that this difference results from the higher effective population size of African Americans.

Correlation in recombination rates between species. To address the conservation of total and background levels of recombination in the two species, we examined the correlation in the estimates of total ρ and background ρ of the chimpanzee and African American data sets. The 210 windows show evidence of autocorrelation as assessed by a runs test (results not shown). To correct for this, we used the method of Dutilleul[25] implemented in the program modttest to adjust the degrees of freedom. Another problem is that the chimpanzee and human windows are not perfectly aligned. To compare estimates of ρ for roughly homologous regions, we mapped the chimpanzee windows onto the human data (as described above) and then reran these windows for both samples. For this data, we altered the prior used by PHASE such that the maximum λ is 100. Simulations suggest that this leads to more accurate estimates of total levels of recombination (data not shown).

Simulations. For the coalescent simulations examining different demographic scenarios, we used the program ms[26]. For all other coalescent simulations, we use a modification of ms that models recombination hotspots, provided by J. Wall (University of Southern California, Dept. of Biological Sciences). In all simulations, we matched the number of individuals and the number of base pairs. We randomly converted a fraction of the simulated sequence data to be coded as missing, where the appropriate fraction was chosen to match the amount of missing data in the two species. The simulations also matched the process of site removal (as described above). If the estimate of ρ was 0, then it was replaced by 10^{-6} per base pair in simulations (the limit in accuracy of PHASE). Finally, as in the actual data, we ran PHASE on the human simulated data, assuming that we did not know phase, and on the chimpanzee simulated data, assuming that we did.

In these simulations, we assumed the standard neutral model of a randomly mating population of constant size, unless otherwise stated. Recombination is modeled as a single crossing-over, with no gene conversion, and is assumed to be piece-wise constant. We chose a θ value that approximately matched the observed allele frequency, after excluding singletons (in chimpanzees) or rare alleles (in humans). For the chimpanzee data, this involved choosing a value of θ somewhat lower than the observed value, because our sample of chimpanzees has an excess of rare alleles relative to the standard neutral model. For the human data, we instead chose a value of θ somewhat higher than the observed value, because, owing to ascertainment bias, the human samples had an excess of intermediate alleles. This approach yields a rough match to the observed allele frequencies (**Supplementary Fig. 2** online). Although our standard neutral model is unrealistic, simulations suggest that the approach is relatively robust to departures from model assumptions[4,6].

Power to detect a hotspot. To estimate the type I and II error rates of PHASE, we ran simulations with and without a hotspot. All simulations used the same value of θ across all windows, chosen to reproduce the median observed number of segregating sites per window in each population. The simulations were run with one of three values of total ρ: the 25th percentile, the median or the 75th percentile of the total ρ estimated from the data, computed for each population separately. For simulations with a hotspot, the hotspot was uniformly distributed from 10 kb to 60 kb in a 70-kb window. The hotspot was 1,500 bases long and its intensity λ = 50, 100 or 200. We ran 100 sets of simulations for each set of parameters and for each population, for a total of 3,200 simulations.

We ran an additional set of simulations to explore the false positive rate under more realistic demographic models for each population. For the chimpanzee data, we simulated a 10-fold bottleneck from 900 to 1,200 generations ago. The ancestral and present population sizes were 30,000. For the African American sample, we simulated growth with an ancestral population size of 10,000, which expanded to 90,000 over the past 400 generations. For the non-African human data sets, we simulated a 10-fold bottleneck from an ancestral population size of 10,000, beginning 350 generations ago and with a 60% recovery 340 generations ago. These parameters were chosen so that the mean Tajima's D would roughly match the observed value for each population (for human populations, the observed Tajima's D value came from the Seattle SNPS database; for the chimpanzee population, the observed Tajima's D came from ref. 27). We also scaled θ and ρ to match the average number of segregating sites (keeping ρ/θ fixed). Because θ and ρ estimates for European Americans and Chinese Americans were so similar, and the demography parameters identical, we did not run separate simulations for these two populations.

Fraction of shared hotspots. To examine how much concordance would be expected if no hotspots were shared between humans and chimpanzees, we retained the location of all the inferred chimpanzee hotspots and placed the hotspots inferred in the African American sample randomly along the sequence. We preserved the observed length of the hotspot credible intervals and required that the hotspot be completely within the entire region surveyed, that only one hotspot mapped to each window and that it did not overlap with any other hotspots (by choosing a new location if any requirement was violated). We then tabulated the percentage of these randomly placed hotspots that overlapped with the ones inferred in chimpanzees. This entire process was repeated 1,000 times.

We also wanted to quantify what would be expected if all hotspots were conserved between humans and chimpanzees. To do this, we first determined which human window each of the inferred chimpanzee hotspots mapped into. We then simulated 100 sets of 39 human windows that contained an inferred chimpanzee hotspot by matching θ and the background ρ values estimated in the African American sample, using the λ and the hotspot location values estimated in chimpanzees. For the intensity, we divided the observed λ by 2 and simulated the hotspot with this lowered value, because the estimates of λ tended to be two times higher than the true value in the power simulations of chimpanzee data. For the location, we used the midpoint of the median estimate for the right and left boundaries of the hotspot. We then used a fixed length of 1,500 bases surrounding this midpoint, because the power simulations suggested an upward bias to the length and intensity. We also ran the simulations without correcting for these two biases (**Supplementary Note** online). In no case did an inferred chimpanzee hotspot span multiple human windows or did multiple hotspots map to the same window. Windows with fewer than 20 SNPs were not considered, just as they were not analyzed in the observed data. (As a result, in both simulated and actual data, we could not have detected all hotspots even if we had 100% power.) Finally, we tabulated how many of the chimpanzee hotspots were detected in humans and identified as the same, given our criterion of overlapping credible intervals. This entire analysis was repeated using the hotspot length and λ as directly estimated from the data, rather than with a fixed length and lowered intensity.

Correlation in recombination rates between species. To examine the strength of the correlation that would be expected if the recombination rate (r) did not vary between the species, we simulated each of the 210 windows used in the correlation 10 times, matching the observed value of θ in that window for each species. We computed an average total recombination rate by dividing the observed total ρ in each species by an estimate of N_e (15,000 for African Americans[28] and 25,000 for central chimpanzees[27]; both these estimates were obtained from diversity and divergence data). We then ran the simulations with this average rate of recombination multiplied by the respective N_e for each species. Thus, the background rate for these windows is also identical in the two species. Because our results suggest that most windows in both species have a hotspot, we simulated a hotspot in each window with an intensity λ of 50 and length of 1,500 bases. These were distributed independently in the two species (because this is what the data suggest) and placed randomly along the sequence. We then computed a one-tailed nonparametric correlation coefficient (Spearman's ρ) for all pairwise combinations of the ten data sets, giving us a total of 100 correlation coefficients. We also ran ten data sets where we assumed that the number of hotspots per window followed an approximate Poisson distribution based on the average frequency of putative hotspots in the two species (**Supplementary Note** online). The number of hotspots for a given window was chosen independently for the two species. For these ten data sets, we again examined all pairwise combinations, giving us a total of 100 correlation coefficients.

URLs. The program modttest is available at http://www.bio.umontreal.ca/legendre/index.html. The program ms is available from http://home.uchicago.edu/~rhudson1/source.html. The Seattle SNPS database is available at http://pga.gs.washington.edu/summary_stats.html.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank M. Lachmann, M. Stephens, K. Voepel and J. Wall for computer programs; M. Lachmann, M. Stephens and Rechenzentrum Garching (especially H. Lederer and W. Nagel) for computer support; D. Cox, M. Lachmann, S. Myers, J. Pritchard and M. Stephens for discussions; A. Di Rienzo for comments on the manuscript; N. Shen for her efforts in the hybrid analyses and resequencing array hybridizations; and M. Jen and K. Pant for their contributions to SNP discovery and base calling efforts. P. Morin helped oversee the collection and management of the chimpanzee samples. Support for this work was provided by the Max Planck Society and a Deutsche Forshungsgesellschaft grant to S.P.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 28 October 2004; accepted 11 February 2005
Published online at http://www.nature.com/naturegenetics/

1. Kong, A. *et al.* A high-resolution recombination map of the human genome. *Nat. Genet.* **31**, 241–247 (2002).

2. Kauppi, L., Jeffreys, A.J. & Keeney, S. Where the crossovers are: recombination distributions in mammals. *Nat. Rev. Genet.* **5**, 413–424 (2004).
3. Stumpf, M.P. & McVean, G.A. Estimating recombination rates from population-genetic data. *Nat. Rev. Genet.* **4**, 959–968 (2003).
4. Crawford, D.C. *et al.* Evidence for substantial fine-scale variation in recombination rates across the human genome. *Nat. Genet.* **36**, 700–706 (2004).
5. Fearnhead, P. Consistency of estimators of the population–scaled recombination rate. *Theor. Popul. Biol.* **64**, 67–79 (2003).
6. Li, N. & Stephens, M. Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics* **165**, 2213–2233 (2003).
7. McVean, G.A. *et al.* The fine-scale structure of recombination rate variation in the human genome. *Science* **304**, 581–584 (2004).
8. Cullen, M., Perfetto, S.P., Klitz, W., Nelson, G. & Carrington, M. High-resolution patterns of meiotic recombination across the human major histocompatibility complex. *Am. J. Hum. Genet.* **71**, 759–776 (2002).
9. Kong, A. *et al.* Recombination rate and reproductive success in humans. *Nat. Genet.* **36**, 1203–1206 (2004).
10. Wall, J.D., Frisse, L.A., Hudson, R.R. & Di Rienzo, A. Comparative linkage-disequilibrium analysis of the beta-globin hotspot in primates. *Am. J. Hum. Genet.* **73**, 1330–1340 (2003).
11. Ptak, S.E. *et al.* Absence of the TAP2 human recombination hotspot in chimpanzees. *PLoS Biol.* **2**, 849–855 (2004).
12. Hinds, D.A. *et al.* Whole genome patterns of common DNA variation in diverse human populations. *Science* (in the press).
13. Britten, R.J. Divergence between samples of chimpanzee and human DNA sequences is 5%, counting indels. *Proc. Natl. Acad. Sci. USA* **99**, 13633–13635 (2002).
14. Ebersberger, I., Metzler, D., Schwarz, C. & Paabo, S. Genomewide comparison of DNA sequences between humans and chimpanzees. *Am. J. Hum. Genet.* **70**, 1490–1497 (2002).
15. de Massy, B. Distribution of meiotic recombination sites. *Trends Genet.* **19**, 514–522 (2003).
16. Petes, T.D. Meiotic recombination hot spots and cold spots. *Nat. Rev. Genet.* **2**, 360–369 (2001).
17. Haring, S.J., Halley, G.R., Jones, A.J. & Malone, R.E. Properties of natural double-strand-break sites at a recombination hotspot in Saccharomyces cerevisiae. *Genetics* **165**, 101–114 (2003).
18. Shiroishi, T., Sagai, T., Hanzawa, N., Gotoh, H. & Moriwaki, K. Genetic control of sex–dependent meiotic recombination in the major histocompatibility complex of the mouse. *EMBO J.* **10**, 681–686 (1991).
19. Jeffreys, A.J. & Neumann, R. Reciprocal crossover asymmetry and meiotic drive in a human recombination hot spot. *Nat. Genet.* **31**, 267–271 (2002).
20. Jeffreys, A.J., Murray, J. & Neumann, R. High-resolution mapping of crossovers in human sperm defines a minisatellite-associated recombination hotspot. *Mol. Cell* **2**, 267–273 (1998).
21. Patil, N. *et al.* Blocks of limited haplotype diversity revealed by high–resolution scanning of human chromosome 21. *Science* **294**, 1719–1723 (2001).
22. Altshul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
23. Stephens, M. & Donnelly, P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am. J. Hum. Genet.* **73**, 1162–1169 (2003).
24. Stephens, M., Smith, N.J. & Donnelly, P. A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* **68**, 978–989 (2001).
25. Dutilleul, P. Modifying the t-test for assessing the correlation between two spatial processes. *Biometrics* **49**, 305–314 (1993).
26. Hudson, R.R. Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* **18**, 337–338 (2002).
27. Fischer, A., Wiebe, V., Paabo, S. & Przeworski, M. Evidence for a complex demographic history of chimpanzees. *Mol. Biol. Evol.* **21**, 799–808 (2004).
28. Ptak, S.E., Voelpel, K. & Przeworski, M. Insights into recombination from patterns of linkage disequilibrium in humans. *Genetics* **167**, 387–397 (2004).