Bayesian Inference of Shared Recombination Hotspots Between Humans and Chimpanzees

Ying Wang* and Bruce Rannala†,1
*Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China and †Department of Evolution and Ecology, University of California, Davis, California 95616
ORCID ID: 0000-0002-8355-9955 (B.R.)

ABSTRACT Recombination generates variation and facilitates evolution. Recombination (or lack thereof) also contributes to human genetic disease. Methods for mapping genes influencing complex genetic diseases via association rely on linkage disequilibrium (LD) in human populations, which is influenced by rates of recombination across the genome. Comparative population genomic analyses of recombination using related primate species can identify factors influencing rates of recombination in humans. Such studies can indicate how variable hotspots for recombination may be both among individuals (or populations) and over evolutionary timescales. Previous studies have suggested that locations of recombination hotspots are not conserved between humans and chimpanzees. We made use of the data sets from recent resequencing projects and applied a Bayesian method for identifying hotspots and estimating recombination rates. We also reanalyzed SNP data sets for regions with known hotspots in humans using samples from the human and chimpanzee. The Bayes factors (BF) of shared recombination hotspots between human and chimpanzee across regions were obtained. Based on the analysis of the aligned regions of human chromosome 21, locations where the two species show evidence of shared recombination hotspots (with high BFs) were identified. Interestingly, previous comparative studies of human and chimpanzee that focused on the known human recombination hotspots within the β-globin and HLA regions did not find overlapping of hotspots. Our results show high BFs of shared hotspots at locations within both regions, and the estimated locations of shared hotspots overlap with the locations of human recombination hotspots obtained from sperm-typing studies.

RECOMBINATION plays an essential role during meiosis, ensuring proper chromosomal segregation; improper recombination and/or segregation may cause serious diseases (Lynn *et al.* 2004; Arnheim *et al.* 2007). By exchanging genetic material between chromosomes, recombination also influences natural selection (Coop and Przeworski 2007) and is an important factor shaping the patterns of linkage disequilibrium, which have a fundamental role in disease association analysis (Pritchard and Przeworski 2001; Slatkin 2008). Moreover, studies have suggested that nonallelic homologous recombination (NAHR) may be related to allelic homologous recombination (De Raedt *et al.* 2006; Lindsay *et al.* 2006). NAHR can result in chromosomal rearrangements, which are responsible for many diseases in humans (Lupski 2004). Despite the importance of recombination, many aspects of recombination are still unknown, although the recent discovery of PRDM9 has furthered understanding of the regulation of recombination (Paigen and Petkov 2010; Baudat *et al.* 2013). Studying the evolution of recombination hotspots among related species could potentially aid in our understanding about the mechanisms of recombination.

In recent years, our understanding of the rates of homologous recombination across the human genome has been advanced by high-resolution pedigree analyses (Kong *et al.* 2002; Coop *et al.* 2008), sperm-typing studies (reviewed in Kauppi *et al.* 2004; Arnheim *et al.* 2007), and statistical inference based on population genomic data (Crawford *et al.* 2004; Myers *et al.* 2005). Such studies have been facilitated by recent advances in genotyping technologies and statistical inference methods. Pedigree analyses have provided insights concerning recombination rates over broad scales across the human genome, but the resolution of such analyses is limited by the number of meioses in pedigrees. Sperm-typing studies

doi: 10.1534/genetics.114.168377
Manuscript received July 14, 2014; accepted for publication September 24, 2014; published Early Online September 26, 2014.
Supporting information is available online at http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.114.168377/-/DC1.
[1]Corresponding author: 1 Shields Ave., University of California, Davis, CA 95616. E-mail: brannala@ucdavis.edu

have provided insights regarding fine-scale recombination rates for several genomic regions, but such methods are laborious and expensive and are currently difficult to scale up to the whole genome. In addition, only male recombination rates can be inferred. Genome-scale analyses of fine-scale recombination rates and comparative studies of recombination hotspots among primates more generally are currently being pursued by applying statistical methods to population genomic data.

Comparative studies of recombination hotspot locations between closely related species can provide insights concerning shared sequence features that may be involved in regulating recombination. Several comparative analyses have been carried out on the basis of population genomic data, mostly between human and chimpanzee (Ptak *et al.* 2004, 2005; Winckler *et al.* 2005; Myers *et al.* 2010; Auton *et al.* 2012), with one study considering human, chimpanzee, and macaque (Wall *et al.* 2003). In general, these studies have concluded that few hotspots are shared between the two species, including both regions for which prior information about human recombination hotspots is available, such as the β-globin and the human leukocyte antigen (HLA) regions, and regions for which no such prior knowledge is available (*e.g.*, comparative analyses that are not focused to regions with known human recombination hotspots). These results raise many interesting questions regarding the evolution of recombination hotspots and the biological factors influencing recombination rates, given that human and chimpanzee possess highly similar DNA sequences (Coop and Przeworski 2007). One possible explanation is that shared hotspots exist but the statistical methodologies perform differently due to various approximations applied. The computational complexity of population genomic inference of fine-scale recombination rates has led to the use of statistical methods that rely on various approximations and it is therefore possible that methodology may play a role in generating the observed dissimilarities between species. Another possibility is that the nature of the mechanism of recombination differs even between closely related species. Previous studies (Myers *et al.* 2008, 2010) show that a sequence motif is associated with a certain percentage of human recombination hotspots and a zinc finger protein PRDM9 binds to the motif, but the motif is not active in chimpanzees. This might be due to *trans*-acting factors regulating recombination in chimpanzees that differ from humans.

To explore these questions, we applied our Bayesian Markov chain Monte Carlo method for simultaneously estimating recombination rates and detecting recombination hotspots (Wang and Rannala 2008, 2009) to the data from two recent resequencing projects (Abecasis *et al.* 2010; Prado-Martinez *et al.* 2013) and to the data sets from a previous study concerning known human recombination hotspots (Winckler *et al.* 2005). A full-likelihood method is used if the sampled genomic regions are not large; otherwise a composite-likelihood method is used by splitting the larger regions into subregions. The full-likelihood method is feasible for use in analyzing moderate-size intervals through the use of a SNP genealogy in which the ancestral markers are efficiently modeled by marker ancestry vectors (Wang and Rannala 2008). A composite-likelihood method applied to larger regions takes advantage of additional information in the larger subintervals, providing improved performance over methods that use small number of SNPs. Moreover, the population mutation rate parameter θ and the expected background recombination rates are integrated (or estimated) in the Markov chain, so that the estimated recombination rates are not influenced by prespecified values of these parameters, which are usually unknown, especially for species other than human. Currently, a Jukes–Cantor (JC69) DNA substitution model is assumed, although it is straightforward to use other models. Empirical data analyses using the JC69 and the F81 substitution models (Felsenstein 1981) suggest that estimated hotspots and recombination rates are not influenced by the substitution model. This is often the case for closely related sequences. By incorporating a realistic model of recombination hotspots and background rates, estimates of these parameters are expected to be improved. Both simulation studies and analyses of population human genomic data for regions that have been studied previously by sperm typing suggest that our method performs well and that the results are consistent over different scenarios (Wang and Rannala 2009). In the present study, we extended our analysis framework so that it can be used for estimating, at any location, the Bayes factors (BFs) that the location is within a hotspot for both species, in this case human and chimpanzee. In other words, we are interested in obtaining the BFs of shared hotspots between the two species across aligned regions.

Material and Methods

Syntenic region

We made use of the existing human–chimpanzee two-way alignment (hg19–panTro3), available from the Uuniversity of California—Santa Cruz (UCSC) genome browser, to obtain fine-scale syntenic regions. First, all syntenic regions that were $\geq$100 bp were noted. The overlapping regions were identified and only the largest regions were retained. For those regions that aligned to minus strands of the human genome, the coordinates for the regions were converted to coordinates on plus strands. The reference sequences for these regions were then obtained, and the human–chimpanzee sequence divergence d (percentage of bases that differ between reference sequences of human and chimpanzee) for each region was calculated. Only regions with $d \leq 10\%$ were retained so that 1319 regions on chromosome 21 were removed and 46092 regions remained. The distance (percentage of different bases) and size distributions of the syntenic regions located on chromosome 21 are illustrated in Supporting Information, Figure S1 and Figure S2. The average distance was 1.98%.

Recombination hotspots are defined as genomic regions with elevated recombination rates compared with surrounding regions. It is impossible to identify hotspots if the regions are too small or if there is not enough data (polymorphic ites) to distinguish hotspot rates and background rates. To identify hotspots based on the sytenic regions, we merged regions into "syntenic blocks" if the gap between the two adjacent regions was <20 bp. Finally only those blocks that were larger than (or equal to) 10 kb and contained at least 20 SNPs from both species were retained. There were 653 such blocks identified on human chromosome 21, spanning 9,828,471 bp, and ranging from 10,000–38809 bp in size. The size distribution of the syntenic blocks is illustrated in Figure S3. The blocks covered 20.4% of the genome on chr21.

Polymorphism data for humans and chimpanzees

We used the human polymorphism data from the 1000 genomes project (Abecasis *et al.* 2010). Variant calls from release v. 3.20101123 were downloaded. Only SNPs with two alleles were retained. We chose 10 YRI individuals in our analysis. Their sample IDs are NA18486, NA18487, NA18489, NA18498, NA18499, NA18501, NA18502, NA18504, NA18505, and NA18508. For chimpanzees, we used the SNP data from Prado-Martinez *et al.* (2013). The chimpanzee samples in their study included individuals from four subspecies, including 10 Nigeria–Cameroon, 6 Eastern, 4 Central, and 4 Western chimpanzees, and only SNPs from the 10 Nigeria–Cameroon individuals were used in our analysis, since it may violate the assumption of the model if combining individuals from different subspecies. Note that these SNPs were identified by mapping reads to the human genome build 36 (hg18). The SNP positions were converted to hg19 using the liftover tool from the UCSC genome browser. For this study, we focused on human chromosome 21. Given the coordinates for the syntenic blocks, SNPs for human and chimpanzee that fall into these regions were identified and used in the recombination hotspot analyses. There are in total 37,118 SNPs for human and 42,403 SNPs for chimpanzee.

For the studies concerning the known human recombination hotspots, including the β-globin hotspot and the several hotspots at the HLA region, we analyzed the data set from a previous study (Winckler *et al.* 2005). The data are summarized in Table S1.

Bayesian inference of recombination hotspots

We applied our method to the data sets described above to identify recombination hotspots and studied the degree of sharing in hotspot usage. Our analysis framework is described elsewhere (Wang and Rannala 2008, 2009), so here we only briefly outline the method. We used a Markov chain Monte Carlo (MCMC) method to estimate the posterior distributions of recombination hotspots and recombination rates across regions. The haplotype phase, missing data (if they exist), ancestral haplotypes, genealogy relating the sample, background recombination rates, and mutation rate parameters are integrated over in the MCMC. Two main models were assumed: one is for the genealogy and the other is for the distribution of recombination hotspots. The genealogy of the sample is described using the coalescent with recombination, and the hotspot distribution is described by a model consisting of two exponential distributions for the waiting distance and duration distance of the hotspot and a log-normal distribution for the intensity of the hotspot.

Table 1 The location and intensity of the hotspots that were considered for generating the 4 simulated data sets in the simulation study

Data set	Hotspot location (bp)	Hotspot intensity (ρ/kb)
S0	—	—
S1	15,000–16,500	40
S2	15,000–16,500	10
S3	15,700–17,200	10

The interval starts from 0 bp with a size of 30,000 bp. Each data set includes 100 replicates. For ρ = 40/kb, if assuming an effective population size (N_e) of 10^4 (for example, for human; Morton, 1982), the recombination rate is 100 cM/Mb and is 25 cM/Mb for ρ = 10/kb.

Let x denote data and G denote the genealogy underlying the data, $\rho = \{\rho_H, \rho_B\}$ be recombination rates consisting of the hotspot rate (ρ_H) and background rate (ρ_B), and θ be the mutation rate. Let λ_1 denote the expected waiting distance between hotspots and λ_2 denote the expected width of the hotspot. Note that based on the current knowledge about these two values, λ_1 and λ_2 were fixed to be 1/50,000 and 1/1000 (Jeffreys *et al.* 2001). The height (intensity) of the hotspot is log-normal distributed with parameters μ and σ. Parameters μ and σ were fixed to be 9 and 1, respectively, so the 95% interval is ~(1141, 57532). Each hotspot is described by three variables: X_1, X_2, Z, representing the start and end positions, and the height of the hotspot. So the prior on ρ_H is $p(\rho_H|\lambda_1, \lambda_2, \mu, \sigma)$. The prior on ρ_B is an exponential distribution with parameter λ_B. The prior on θ is a Gamma distribution with shape parameter equal to 0.25 and scale parameter equal to 2. The priors were chosen to be diffuse with the 95% interval to be $(5.3 \times 10^{-7}, 3.4)$. The posterior probability of hotspots at location i is

$$p(H_i|x) = \sum_G \iint\iint p(x|G,\theta)p(G|\rho_H,\rho_B)p(\rho_H)p(\rho_B) \times p(\theta)p(\lambda_B)\mathrm{d}\rho_H\mathrm{d}\rho_B\ \mathrm{d}\theta\ \mathrm{d}\lambda_B. \quad (1)$$

Let x_{S_1} and x_{S_2} denote the data from species S_1 and S_2 respectively. The BF of shared hotspot between S_1 and S_2 is

$$\frac{p(H_i|x_{S_1})p(H_i|x_{S_2})/(1-p(H_i|x_{S_1})p(H_i|x_{S_2}))}{p(H_i)^2/\left(1-p(H_i)^2\right)}, \quad (2)$$

where $p(H_i)$ is defined by the recombination hotspot model in the analysis framework and is based on the values of λ_1

Table 2 Six analysis scenarios that were considered for examining the performance of the method for detecting shared hotspots between species in the simulation studies

Set	Data sets	No. pairs	Note
1	S1	4950	Hotspots are the same between two species with ρ = 40/kb
2	S2	4950	Hotspots are the same between two species with ρ = 10/kb
3	S0/S1	10000	One species has a hotspot with ρ = 40/kb and the other species lacks a hotspot
4	S0/S2	10000	One species has a hotspot with ρ = 10/kb and the other species lacks a hotspot
5	S1/S3	10000	Hotspots are partially overlapped with intensities of ρ = 40/kb for one species and ρ = 10 per kb for the other species
6	S2/S3	10000	Hotspots are partially overlapped with intensities of ρ = 10/kb for both species

The distribution of the hotspots for simulated data set S0–3 are listed in Table 1. The number of pairs that were used in each set are given in the table. For sets 1 and 2, since only 1 simulated data set was used and all possible pairs were considered, the number of pairs is 4950. For set 3–6, two different simulated data sets were used, and the number of pairs is 10,000.

and λ_2. Since in our model, a hotspot that spans the beginning of the interval is not allowed, the prior probability of a hotspot across the region is thus not uniform, with the probability smaller toward the start of the interval. The prior probability of hotspots across the region were obtained by simulation and is plotted in Figure S4 for a region with size 1,000,000 bp.

Simulation studies

Since the current data set for chr21 contains only 10 YRI individuals and 10 chimpanzee individuals, to obtain a rough idea of the interpretations of different values of BFs of shared hotspots, we first conducted two simulation studies using the same parameters as those in our previous analysis (Wang and Rannala 2009). The data were simulated using msHOT (Hudson 2002; Hellenthal and Stephens 2007). The parameters used in the simulation are given in Wang and Rannala (2009). The data sets used here were labeled S1 and S2 and are the same as those used in Wang and Rannala (2009) except that 20 haplotypes were used (as 10 genotypes). Data sets S1 and S2 each includes 100 simulated replicates. The interval was 30 kb long with 0 as the starting point. The true hotspot was located between 15 and 16.5 kb. The hotspot intensity ρ_H was set to be 40/kb for S1 and 10/kb for S2. We added two additional simulated data sets illustrating two different cases: in the first case the hotspot is present only in one species, and in the second case, hotspots from two species are partially overlapped. The two data sets were labeled as S0 and S3. The location and hotspot intensities for the four simulated data sets are listed in Table 1. In summary, we considered six scenarios including a completely shared hotspot, no sharing of a hotspot, and partially overlapped hotspots between species. The six combinations are listed in Table 2.

First, for each simulated data set, we obtain the posterior probability of a hotspot across the region. The region was divided into 200-bp nonoverlapping windows. So we have a posterior probability of a hotspot for each window. We then calculated the BF of shared hotspots for each pair of simulated data sets for each set. Sets 1 and 2 contain 4950 pairs and sets 3–6 each contain 10,000 pairs. For each set, we calculated power and false-positive rate for each window across the regions. Power is defined as the probability that the window is correctly identified as a shared hotspot (when overlapped with a shared hotspot). False-positive rate is defined as the probability that the window is identified as a shared hotspot when it is not. We explored several BF thresholds for identifying shared hotspots. In the plots (Figure S8, Figure S9, and Figure S10), we used a BF threshold of 100 for calculating power and false-positive rate. The distributions of the BFs, power, and false-positive rate for locations across the regions were plotted in Figure S8, Figure S9, and Figure S10.

Since the hotspot locations are estimated on the basis of SNPs, the estimated locations of hotspots are usually not precise (it is affected by SNP density, location, etc.). The false-positive rate is relatively high around the shared hotspots and decreases rapidly while moving away from the shared hotspots (Figure S8 and Figure S10, right). This is true for both completely shared hotspots (Figure S8) and partially overlapped hotspots (Figure S10). When there is no overlapped hotspots, the false-positive rate is very low and the maximum false-positive rate across the region is 0.0037 from 15,900 to 16,100 bp, which is within the hotspot of one species (Figure S9). Based on the simulation studies, a BF threshold of 100 was used for inferring shared hotspot in the following data analysis.

MCMC analyses

For the chr21 data sets, we analyzed all regions using two independent runs. For each run, four chains were used. The number of burn-in and sample iterations were set to be 100,000 and 100,000, respectively. The temperature parameter for the parallel chains was set to be 1.2. The tuning parameters for changing λ_B and ρ_B are 170 and 300. For other tuning parameters, default values of the program were used. Usually the tuning parameters are chosen on the basis of the percentage of accepted moves (swapping chains or parameter changes). We usually start some testing runs first to examine the percentage of accepted moves and then adjust the tuning parameters before running the entire MCMC. To assess the consistency between the two independent runs, we first calculated the BF of shared hotspots between YRI and chimpanzees across regions for all 653 syntenic

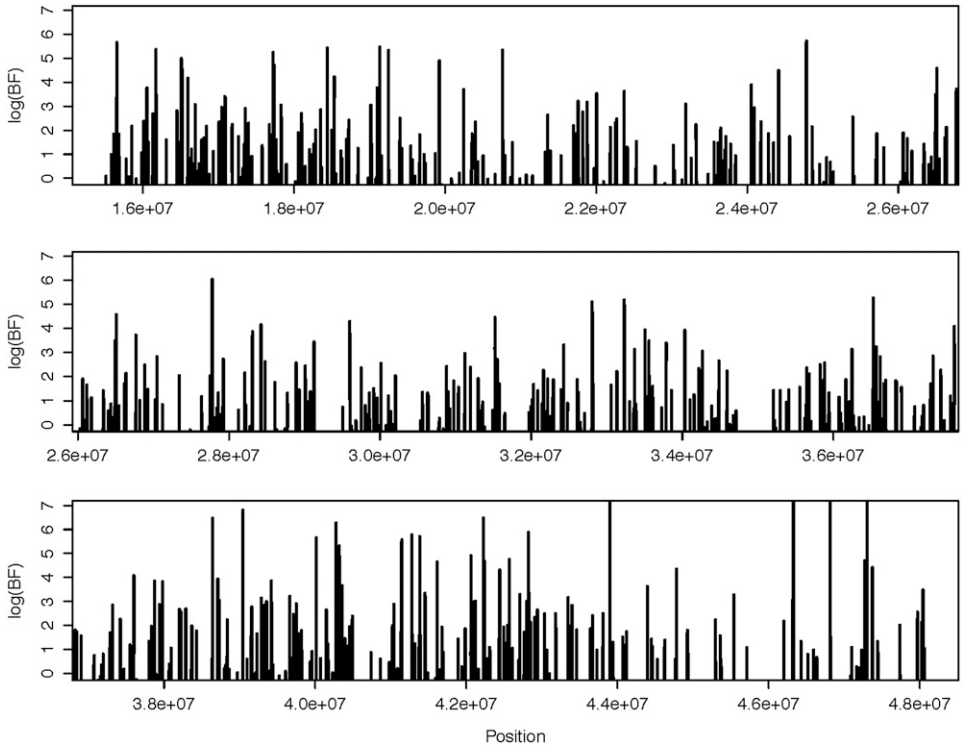

Figure 1 Log BF of shared recombination hotspots between humans and chimpanzees as a function of location for the 653 aligned syntenic regions across chromosome 21 estimated using 10 YRI samples and 10 chimpanzee samples. The coordinates are for the hg19 reference genome.

regions from the two runs. We then calculate the correlation of BFs from the two independent runs; the correlation is 0.979.

For β-globin data set, we analyzed the data set from three populations, including BEN, CEU, and Chimp, using four independent runs. Again, four parallel chains were used for each run. The number of burn-in and sample iterations were set to be 80,000 and 120,000, respectively. The temperature parameter was set to be 1.1 for Chimp and CEPH data sets and 1.05 for BEN data set. The tuning parameters for changing λ_B and ρ_B were set to be 250 and 250 for all three data sets. The default values were used for other tuning parameters. The results from the four independent runs are largely consistent (Figure S11).

Convergence of the MCMC is sometimes problematic for the HLA region in chimpanzee. One possible explanation is that more heterogeneity of hotspots exists in chimpanzee. For the analyses across regions of chr21 and the β-globin region, the interval sizes are small enough that they can be analyzed using the full-likelihood method implemented in InferRho (IR). The HLA data sets were first split into 20-SNP intervals and were analyzed using the composite-likelihood method implemented in the program. The parameters were specified as follows: the number of burn-in iterations was 80,000, the number of sampling iterations was 120,000, the tuning parameter for λ_B was 150, the tuning parameter for ρ_B was 300, the temperature was 1.03, and the number of parallel chains was eight. We analyzed the data sets from three populations using five different runs. Due to the existence of slight inconsistencies between runs, we calculated the BFs of shared hotspots using the runs with the highest sum of log priors and log likelihood (which indicates a better mixing) for all thee data sets.

Results

Chromosome 21

We first identified syntenic regions of the genomes in humans and chimpanzees. In this analysis, we focus on regions located on human chromosome 21. We used the 1000 genomes project data for 10 Yoruban (YRI) samples (Abecasis *et al.* 2010) and the data from a recent ape resequencing project for 10 chimpanzee individuals (Prado-Martinez *et al.* 2013) for inferring recombination hotspots and estimating recombination rates. The YRI sample was chosen because it has a very ancient population history, which can be expected to increase the power for identifying hotspots. The coordinates of syntenic regions were obtained on the basis of the two-way alignment of human and chimpanzee. The regions were verified by calculating the distances (percentage of different bases) for each region to ensure synteny. There are 653 such syntenic blocks identified for human chromosome 21, spanning 9,828,471 bp and ranging in size from 10,000 to 38,809 bp (see Figure S3). The SNPs in these regions from the 1000 genomes project and the ape resequencing project were used for the recombination rate analyses. The posterior probabilities of recombination hotspots across regions were obtained using our program IR (Wang and Rannala 2009). The BF of shared hotspots between the two species was then calculated on the basis of the prior and posterior probabilities of hotspot in each species across

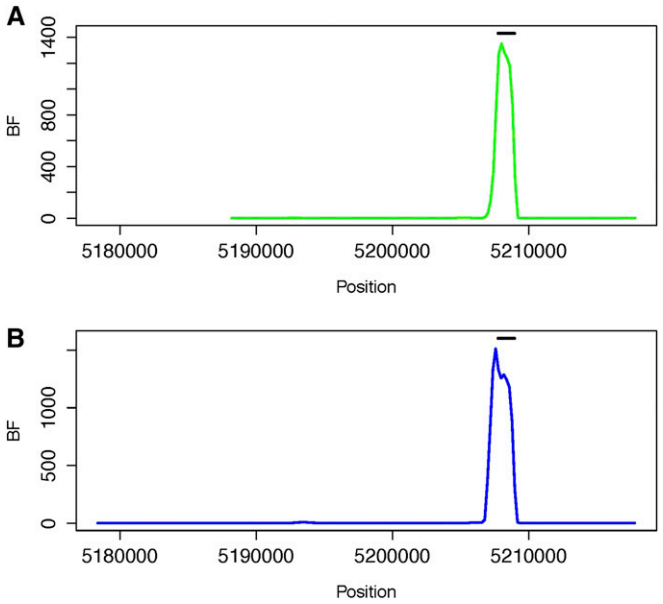

Figure 2 BF of shared recombination hotspots between (A) BEN and Chimp and between (B) CEU and Chimp as a function of location for the β-globin region. The coordinates are given on the basis of coordinates in the original data set (Winckler *et al.* 2005), which used the hg15 reference genome. The human β-globin recombination hotspot estimated from sperm-typing studies (Holloway *et al.* 2006) is illustrated with the horizontal bar.

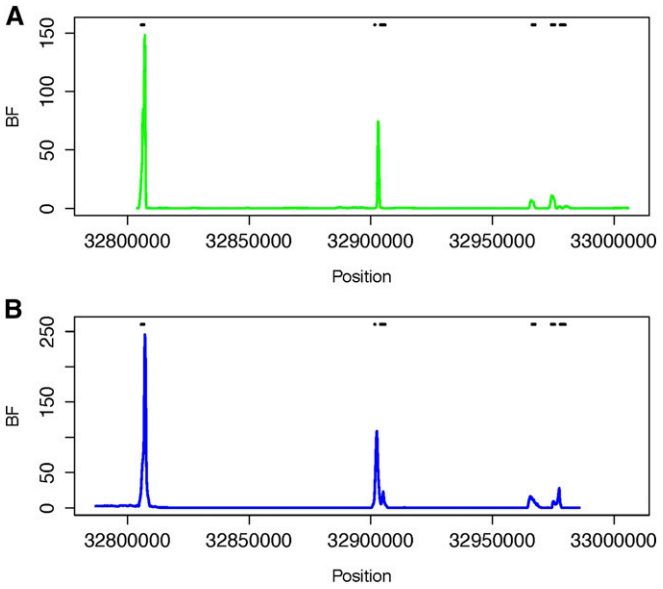

Figure 3 BF of shared recombination hotspots between (A) BEN and Chimp and between (B) CEU and Chimp as a function of location for the HLA region. The coordinates are based on the coordinates in the original data set (Winckler *et al.* 2005), which used the hg15 reference genome. Six human recombination hotspots within the region, including DNA1, DNA2, DNA3, DMB1, DMB2, and TAP2, estimated from sperm-typing studies (Jeffreys *et al.* 2001), are illustrated with the horizontal bars.

locations, with the assumption that the posterior and prior probabilities of a recombination hotspot in the two species are independent.

To evaluate the validity of our method for detecting shared hotspots in two species we conducted six simulation analyses that explore the distribution of the BFs of shared hotspots for regions with (or without) a hotspot and that assessed how the locations of hotspots affect the distribution of the BFs. The simulation studies provide a reference for interpreting the estimated BFs obtained using the real data. Although the 95% credible interval of the BFs for locations within a hotspot is large, the BF decreases rapidly as one moves away from the hotspot along a chromosome. For regions that do not contain a common hotspot, the BF of a shared hotspot is uniformly low. For example, if we sample one location of 3900 bp over the 30-kb region and the location is relatively far away from the shared hotspot, which is located between 15 and 16.5 kb, the 95% interval of the BF is (0.000, 0.263) for data set S1 and is (0.000, 0.321) for data set S2. The simulation study results suggest that even a moderate BF (>10) can indicate a shared hotspot with high confidence.

Across chr21 over all aligned blocks, the BFs of shared recombination hotspots between YRI and chimpanzee samples were plotted across the chromosome and are illustrated in Figure 1. The plot suggests that numerous hotspots that are shared between the two species with high confidence exist (BF $\geq$ 100). The estimated recombination rates for YRI and chimpanzee are presented in Figures S5, A–C. The locations with BF of shared hotspot $\geq$100 for chromosome 21 are given in Table S3, A–D.

β-Globin and HLA regions

We also examined regions that contain well-established human recombination hotspots and examined whether these hotspots also exist in chimpanzee. The data sets that were analyzed are from Winckler *et al.* (2005). The hotspots include the β-globin hotspot and the several hotspots in the HLA region. In total, 48 individuals from the CEPH resource (CEU), 47 individuals from the Beni population sampled from Nigeria (BEN) and 37 individuals from the western African chimpanzee (Chimp) were sampled. The samples were genotyped at 26, 30, and 39 SNP loci, respectively. BFs of shared recombination hotspots between BEN and Chimp, and between CEU and Chimp, across the region were estimated. The results are shown in Figure 2. The human β-globin recombination hotspot obtained from a sperm-typing study overlaps with the interval showing high BFs of a shared hotspot. Our findings differ from those of previous analyses in suggesting that the well-established human β-globin hotspot is also present in chimpanzee. The estimated recombination rates across the region using samples of Beni, CEPH, and Chimp are plotted in Figure S6. The locations with BF of shared hotspot $\geq$100 for the β-globin region are given in Table S4 and Table S5.

Sperm-typing studies have previously revealed six recombination hotspots in the human HLA region, including three hotspots in the DNA1-3 cluster, two hotspots in the

DMB1-2 cluster, and a TAP2 hotspot Jeffreys *et al.* (2001). Winckler *et al.* (2005) examined the HLA regions spanning all six human recombination hotspots using the same sample of human and chimpanzee individuals as were used for the β-globin region study described above, genotyping 114, 111, and 98 SNP loci for Chimp, CEU, and BEN, respectively. The results based on a reanalysis of their data using IR are presented in Figure 3. The locations with higher BFs of sharing overlap with the locations of the human recombination hotspots identified from sperm-typing studies, although the BFs are smaller than those within the β-globin region. Thus, there is evidence that the TAP2 hotspot is shared between human and chimpanzee and weaker evidence of sharing for the two DMB hotspots. The estimated recombination rates across the region using samples of Beni, CEPH, and Chimp are plotted in Figure S7. The locations with BF of shared hotspot $\geq$100 for the HLA region are given in Table S6 and Table S7.

Discussion

Our results partially support the findings of several previous studies that sharing of hotspots between humans and chimpanzees does not appear to be universal, but differ from these studies in that we identified a fraction of human recombination hotspots that are clearly present in both humans and chimpanzees. This difference may be due to methodology, for example the statistical power of different methods or parameter handling of different models. Conversely, the difference may be due to the existence of a higher degree of population heterogeneity in chimpanzee than in human. Especially based on our results, for example, the hotspot at the β-globin region is shared between the two species, but the recombination rates within the hotspot are much higher than the surrounding regions for human, and it is not that obvious for chimpanzees (Figure S6).

It may also be suspected that the degree of sharing is due to random effects (*i.e.*, that independently arising hotspots in the two species overlap purely by chance). We conducted a simulation study to roughly examine how often the hotspots are shared if locations of hotspots are determined independently in each of the two species. Assuming that the length of the shared chromosome region is 9,828,471 bp (the size of the syntenic regions from human and chimpanzee based on above analysis of human chromosome 21), we simulated 1000 regions using our model of hotspot distribution with parameters as estimated from our empirical analysis (described in the above section). There are 499,500 pairs of regions. The simulated hotspots were examined for each pair to calculate the percentage of hotspots overlapping with the other species. On average the percentage of sharing is 0.039 if an overlap of at least 1 bp was considered as sharing and is 0.012 if a 1000-bp overlap was the criterion for sharing. The histograms in Figure S12 show the distribution of the percentages of sharing from the simulation. Based on our rough analysis, the expected degree of sharing is quite small if the distributions of hotspots from the two species arise by independent processes rather than being due to shared (homologous) genomic features. Given the evidence for shared hotspots, it may be useful to examine whether shared sequences are associated with the conserved hotspots and, if so, whether there are factors that are common in humans and chimpanzees that interact with certain sequences and regulate recombination activity. The program inferRho is available for download from http://rannala.org.

Acknowledgments

B.R. was supported by a grant from the National Institutes of Health/National Human Genome Research Institute (HG01988).

Literature Cited

Abecasis, G. R., D. Altshuler, A. Auton, L. D. Brooks, R. M. Durbin *et al.*, 2010 A map of human genome variation from population-scale sequencing. Nature 467: 1061–1073.

Arnheim, N., P. Calabrese, and I. Tiemann-Boege, 2007 Mammalian meiotic recombination hot spots. Annu. Rev. Genet. 41: 369–399.

Auton, A., A. Fledel-Alon, S. Pfeifer, O. Venn, L. Sgurel *et al.*, 2012 A fine-scale chimpanzee genetic map from population sequencing. Science 336: 193–198.

Baudat, F., Y. Imai, and B. de Massy, 2013 Meiotic recombination in mammals: localization and regulation. Nat. Rev. Genet. 14: 794–806.

Coop, G., and M. Przeworski, 2007 An evolutionary view of human recombination. Nat. Rev. Genet. 8: 23–34.

Coop, G., X. Q. Wen, C. Ober, J. K. Pritchard, and M. Przeworski, 2008 High-resolution mapping of crossovers reveals extensive variation in fine-scale recombination patterns among humans. Science 319: 1395–1398.

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.

De Raedt, T., M. Stephens, I. Heyns, H. Brems, D. Thijs *et al.*, 2006 Conservation of hotspots for recombination in low-copy repeats associated with the nf1 microdeletion. Nat. Genet. 38: 1419–1423.

Felsenstein, J., 1981 Evolutionary trees from dna-sequences: a maximum-likelihood approach. J. Mol. Evol. 17: 368–376.

Hellenthal, G., and M. Stephens, 2007 mshot: modifying hudson's ms simulator to incorporate crossover and gene conversion hotspots. Bioinformatics 23: 520–521.

Holloway, K., V. E. Lawson, and A. J. Jeffreys, 2006 Allelic recombination and de novo deletions in sperm in the human beta-globin gene region. Hum. Mol. Genet. 15: 1099–1111.

Hudson, R. R., 2002 Generating samples under a Wright–Fisher neutral model of genetic variation. Bioinformatics 18: 337–338.

Jeffreys, A. J., L. Kauppi, and R. Neumann, 2001 Intensely punctate meiotic recombination in the class ii region of the major histocompatibility complex. Nat. 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.

Kong, A., D. F. Gudbjartsson, J. Sainz, G. M. Jonsdottir, S. A. Gudjonsson *et al.*, 2002 A high-resolution recombination map of the human genome. Nat. Genet. 31: 241–247.

Lindsay, S. J., M. Khajavi, J. R. Lupski, and M. E. Hurles, 2006 A chromosomal rearrangement hotspot can be identified from population genetic variation and is coincident with a hotspot for allelic recombination. Am. J. Hum. Genet. 79: 890–902.

Lupski, J. R., 2004 Hotspots of homologous recombination in the human genome: not all homologous sequences are equal. Genome Biol. 5: 242.

Lynn, A., T. Ashley, and T. Hassold, 2004 Variation inhuman meiotic recombination. Annu. Rev. Genomics Hum. Genet. 5: 317–349.

Morton, N. E., 1982 *Outline of Genetic Epidemiology*. Karger, New York.

Myers, S., L. Bottolo, C. Freeman, G. McVean, and P. Donnelly, 2005 A fine-scale map of recombination rates and hotspots across the human genome. Science 310: 321–324.

Myers, S., C. Freeman, A. Auton, P. Donnelly, and G. McVean, 2008 A common sequence motif associated with recombination hot spots and genome instability in humans. Nat. Genet. 40: 1124–1129.

Myers, S., R. Bowden, A. Tumian, R. E. Bontrop, C. Freeman *et al.*, 2010 Drive against hotspot motifs in primates implicates the prdm9 gene in meiotic recombination. Science 327: 876–879.

Paigen, K., and P. Petkov, 2010 Mammalian recombination hot spots: properties, control and evolution. Nat. Rev. Genet. 11: 221–233.

Prado-Martinez, J., P. H. Sudmant, J. M. Kidd, H. Li, J. L. Kelley *et al.*, 2013 Great ape genetic diversity and population history. Nature 499: 471–475.

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

Ptak, S. E., A. D. Roeder, M. Stephens, Y. Gilad, S. Paabo *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.

Slatkin, M., 2008 Linkage disequilibrium: understanding the evolutionary past and mapping the medical future. Nat. Rev. Genet. 9: 477–485.

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

Wang, Y., and B. Rannala, 2008 Bayesian inference of fine-scale recombination rates using population genomic data. Philos. Trans. R. Soc. Lond. B Biol. Sci. 363: 3921–3930.

Wang, Y., and B. Rannala, 2009 Population genomic inference of recombination rates and hotspots. Proc. Natl. Acad. Sci. USA 106: 6215–6219.

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.

Communicating editor: J. D. Wall

GENETICS

Supporting Information
http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.114.168377/-/DC1

Bayesian Inference of Shared Recombination Hotspots Between Humans and Chimpanzees

Ying Wang and Bruce Rannala

DOI: 10.1534/genetics.114.168377

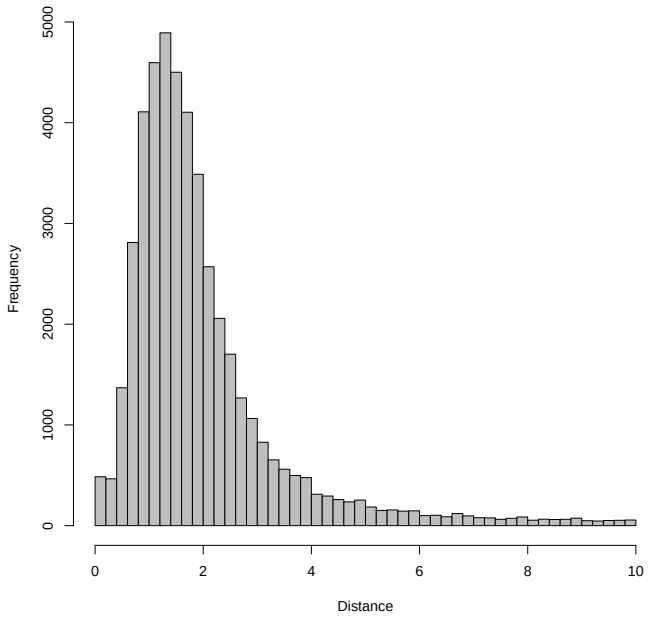

Figure S1. The distribution of human-chimpanzee sequence divergence for syntenic regions of humans and chimpanzees on human chromosome 21. Distance refers to the percentage of different bases across region between the two species.

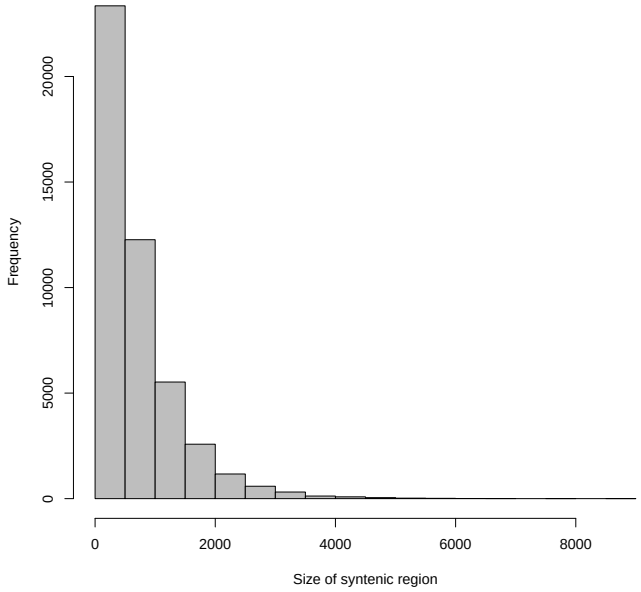

Figure S2. The distribution of the sizes (in bp) of syntenic regions of humans and chimpanzees on human chromosome 21.

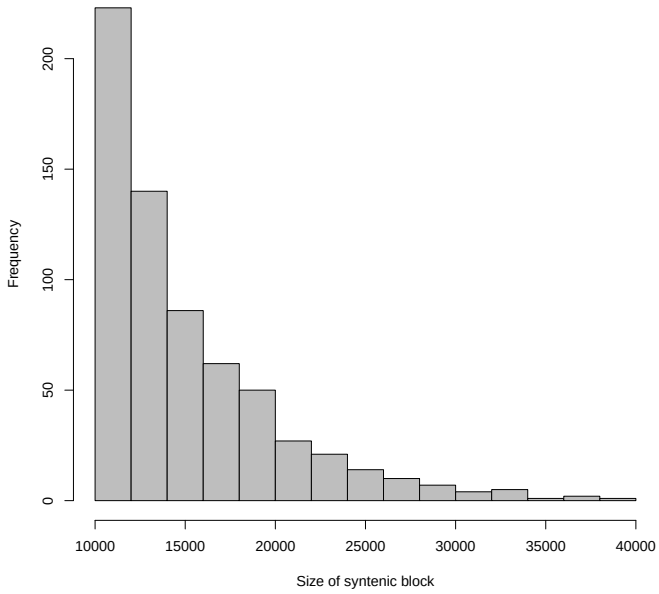

Figure S3. The distribution of the sizes (in bp) of syntenic blocks of humans and chimpanzees on human chromosome 21.

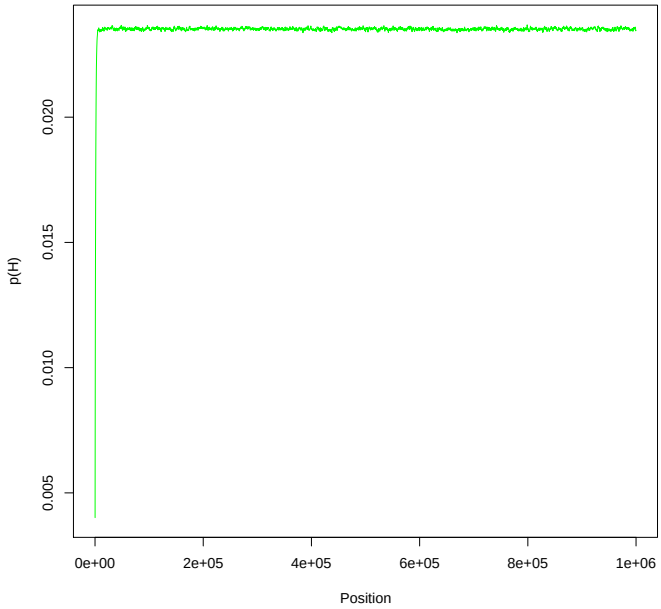

Figure S4. The prior probability of hotspot estimated from simulation. The parameters λ_1 and λ_2 were fixed to be $1/50000$ and $1/1000$, respectively.

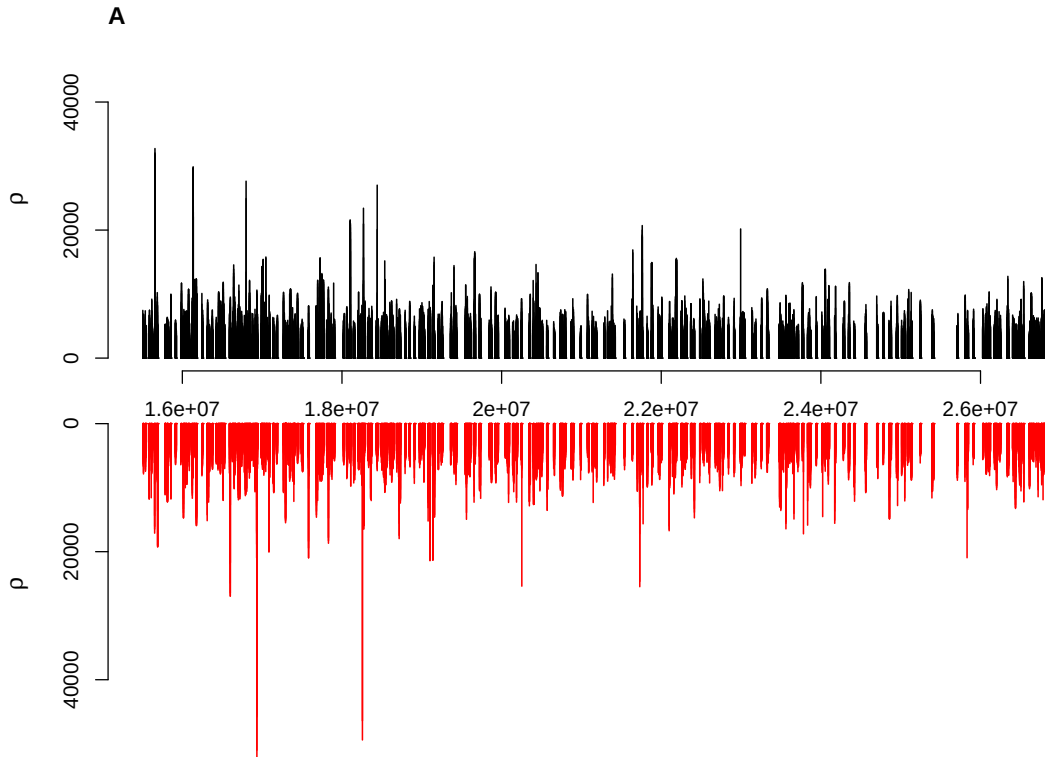

Figure S5A. The estimated recombination rates (ρ) from 15504000 to 26362000 bp on chromosome 21 for YRI (black lines) and chimpanzee (red lines).

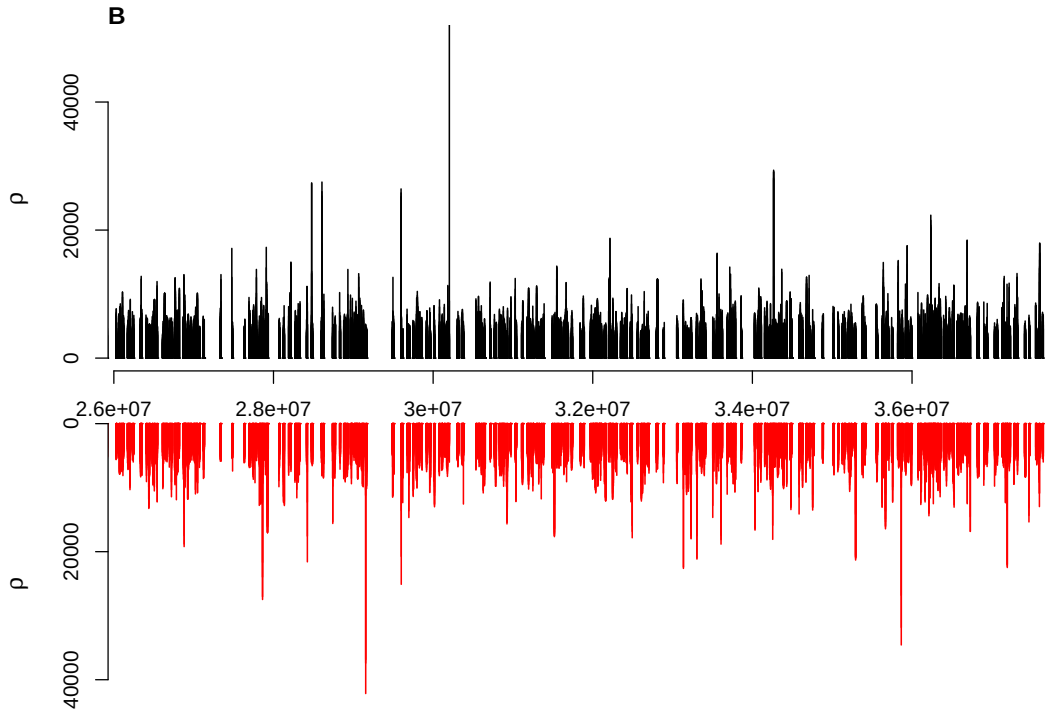

Figure S5B. The estimated recombination rates (ρ) from 26362000 to 37220000 bp on chromosome 21 for YRI (black lines) and chimpanzee (red lines).

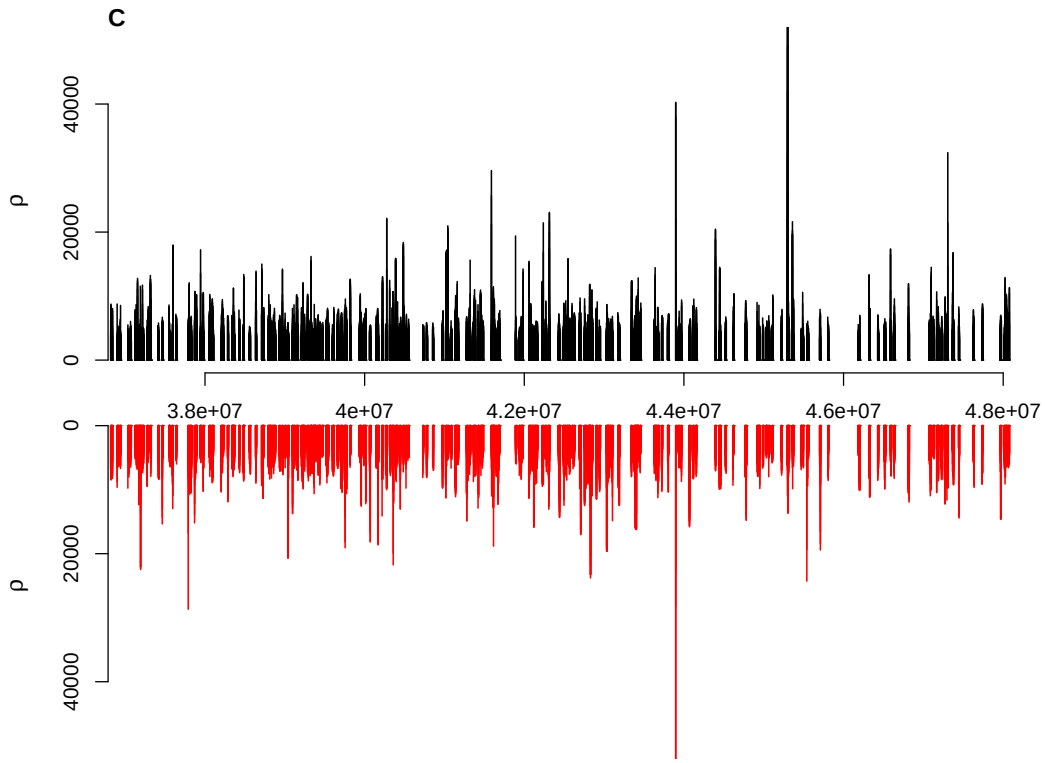

Figure S5C. The estimated recombination rates (ρ) from 37220000 to 48078000 bp on chromosome 21 for YRI (black lines) and chimpanzee (red lines).

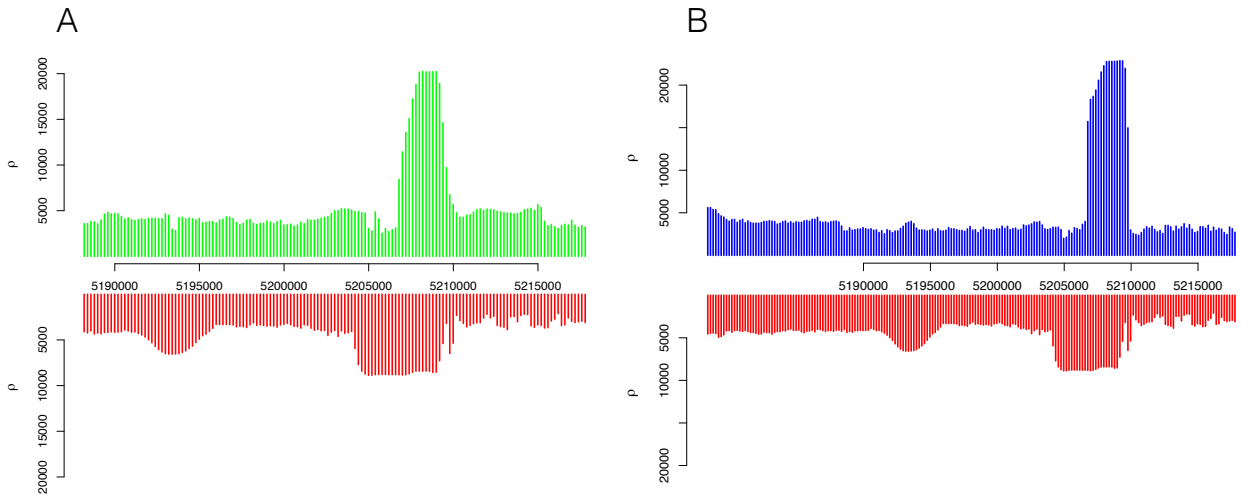

Figure S6. The estimated recombination rates (ρ) across β-globin region using samples of Beni (green) and Chimpanzee (red) **(A)** and samples of CEPH (blue) and Chimpanzee (red) **(B)**.

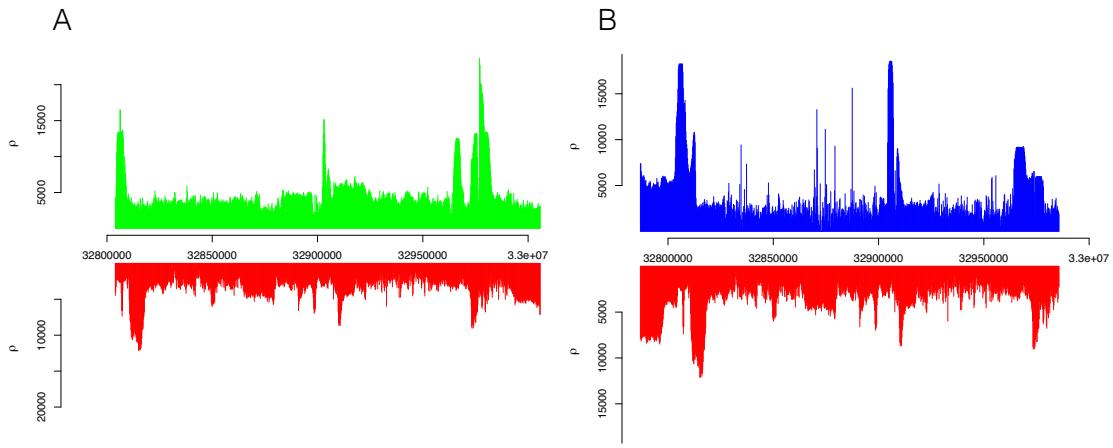

Figure S7. The estimated recombination rates (ρ) across HLA region using samples of Beni (green) and Chimpanzee (red) **(A)** and samples of CEPH (blue) and Chimpanzee (red) **B**.

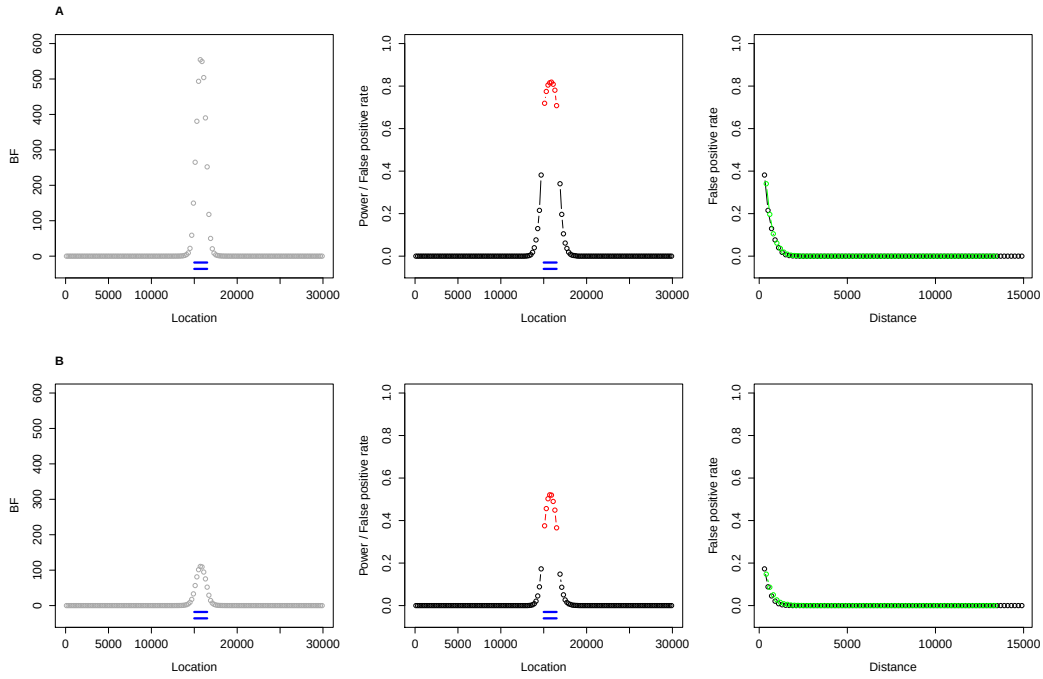

Figure S8. Simulation study results from set 1 **(A)** and results from set 2 **(B)**. The true hotspot was located between 15 kb and 16.5 kb and indicated by blue horizontal lines. Note that the locations of the hotspot for the two species are the same, so the shared hotspot is between 15 kb and 16.5 kb. **Left panel:** The 50% quantile of the BFs of shared hotspots over locations. **Middle panel**: Power (red) and false positive rate (black) of identifying shared hotspot estimated using the 4950 pairs across locations. **Right panel**: False positive rate of identifying shared hotspot as a function of distance to the left bound of the shared hotspot (15 kb) (black dots/line) and to the right bound of the shared hotspot (16.5 kb) (green dots/line)

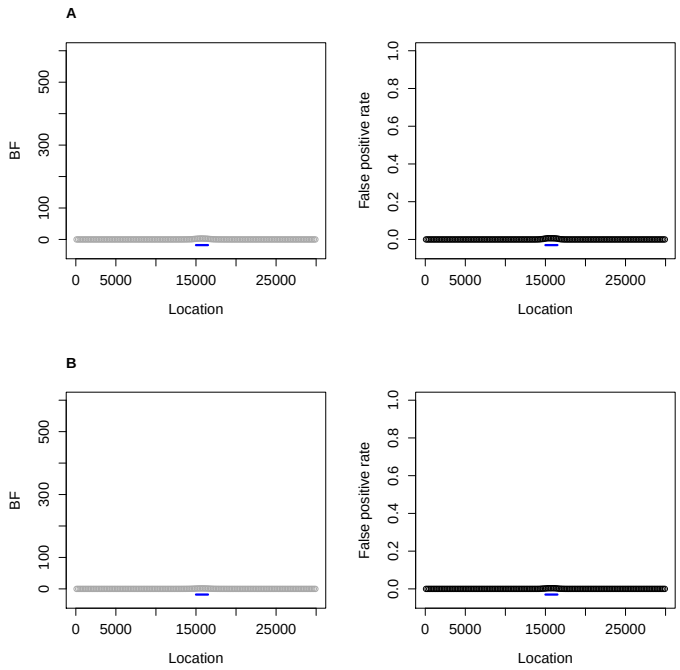

Figure S9. Simulation study results from set 3 **(A)** and results from set 4 **(B)**. The true hotspot from one species was located between 15 kb and 16.5 kb and indicated by a blue horizontal line. Note that only one species contains a hotspot, so the shared hotspot is absent. **Left panel:** The 50% quantile of the BFs of shared hotspots over locations. **Right panel**: False positive rate of identifying shared hotspot estimated using the 10000 pairs across locations.

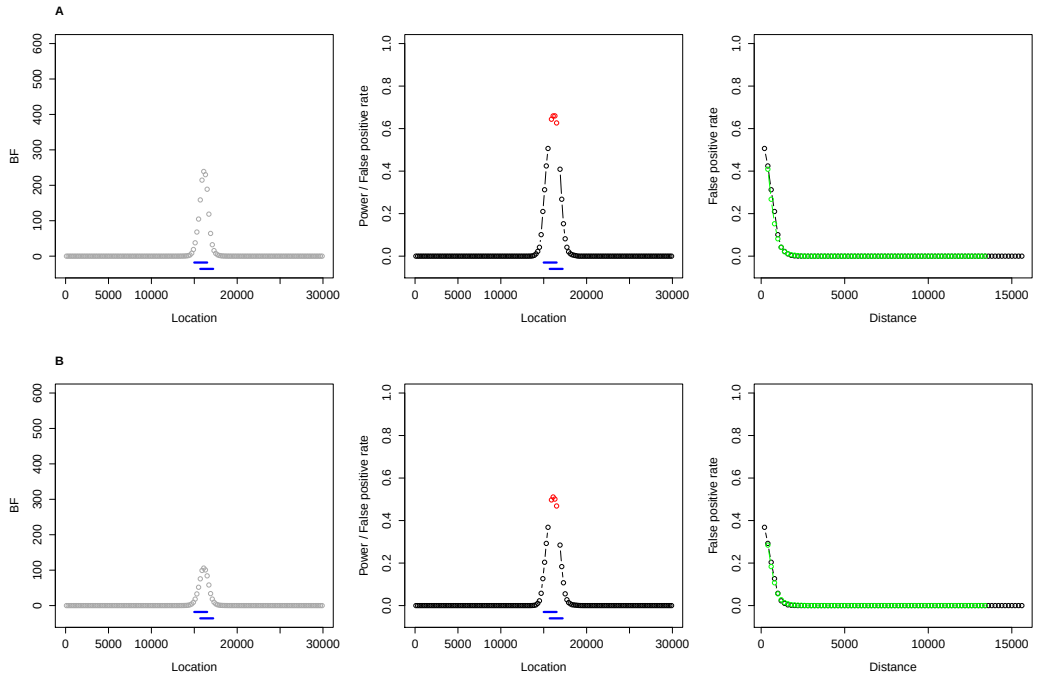

Figure S10. Simulation study results from set 5 **(A)** and results from set 6 **(B)**. The true hotspot was located between 15 kb and 16.5 kb for one species and was located between 15.7 kb and 17.2 kb for the other speices. The hotspots were indicated by blue horizontal lines. Note that the shared hotspot is between 15.7 kb and 16.5 kb. **Left panel:** The 50% quantile of the BFs of shared hotspots over locations. **Middle panel**: Power (red) and false positive rate (black) of identifying shared hotspot estimated using the 10000 pairs across locations. **Right panel**: False positive rate of identifying shared hotspot as a function of distance to the left bound of the shared hotspot (15.7 kb) (black dots/line) and to the right bound of the shared hotspot (16.5 kb) (green dots/line)

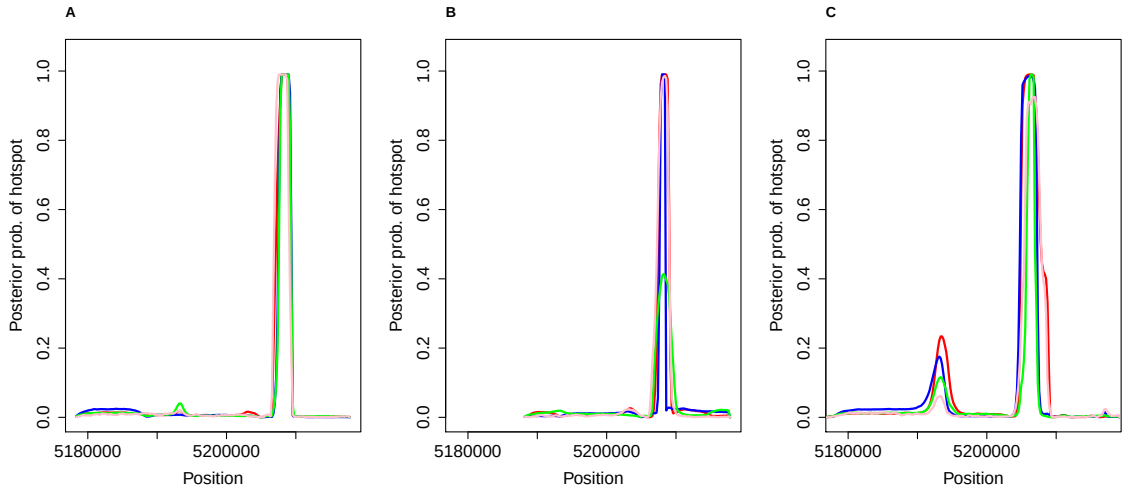

Figure S11. The posterior probability of hotspot across the region for CEU (A), BEN (B), Chimp (C) from four independent runs (red, blue, green, and pink lines).

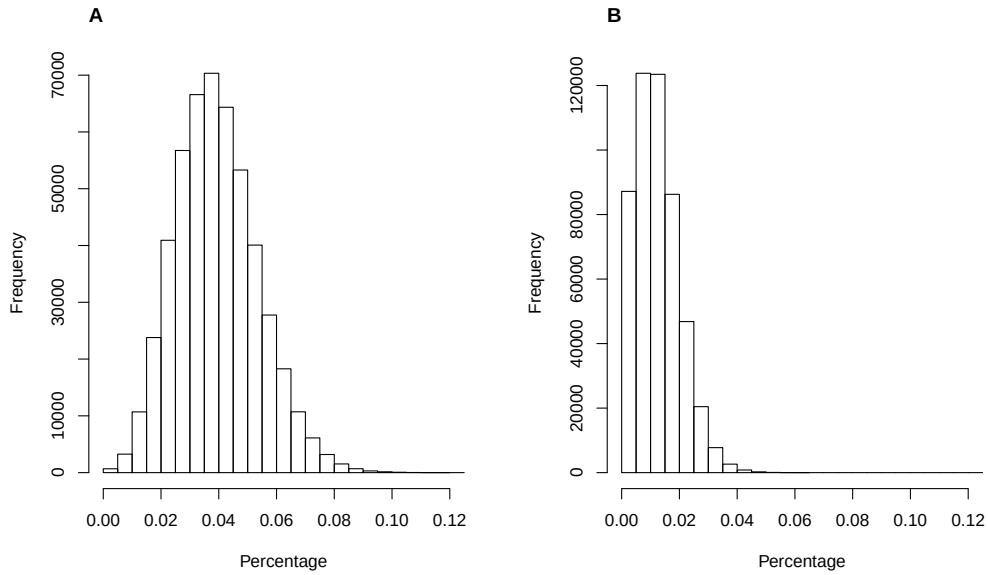

Figure S12. The percentage of hotspot overlapping between species calculated using simulated recombination hotspots based on the prior model used in the analysis. Either overlapping at least 1 bp (A) or 1000 bp (B) was assumed.

Table S1. Summary of the polymorphic data from humans and chimpanzees for the β-globin and HLA regions.

Region	Number of individuals			Number of SNPs		
	CEU	BEN	Chimp	CEU	BEN	Chimp
β-globin	48	47	38	26	30	39
HLA	48	47	37	111	98	114

Table S2. Positions of human recombination hotspots identified from sperm typing studies for the β-globin (HOLLOWAY *et al.*, 2006) and the HLA regions (JEFFREYS *et al.*, 2001) indicated in Figures 2 and 3 in the main text. The coordinates are given in both hg15 and hg19.

		hg15		hg19	
Hotspot	chr	start	end	start	end
β-globin	chr11	5207738	5208938	5248701	5249901
DNA1	chr6	32977952	32979852	32976505	32978405
DNA2	chr6	32974244	32975544	32972797	32974097
DNA3	chr6	32966323	32967522	32964870	32966070
DMB1	chr6	32904007	32905807	32902527	32904327
DMB2	chr6	32901655	32901657	32900176	32900178
TAP2	chr6	32805677	32806677	32804157	32805157

Table S3A. The location, BF of shared hotspot, posterior and prior probabilities of hotspot, and recombination rate for the locations within chromosome 21 with BF of shared hotspot for YRI and chimpanzee that is ≥ 100. Note that the bin size was set to be 200 bp in the analysis.

	BF of	Posterior prob.			ρ per kb	
Location	shared hotspot	YRI	Chimp	Prior prob.	YRI	Chimp
15656560	285.494	0.870	0.156	0.023	31539.316	6929.126
15656760	164.886	0.908	0.092	0.023	31667.729	5494.949
15656960	111.596	0.971	0.059	0.023	31866.629	4600.050
16176300	183.542	0.231	0.398	0.023	10007.385	14574.489
16176500	213.912	0.191	0.552	0.023	10033.589	15936.120
16176700	177.297	0.159	0.562	0.023	10029.131	15885.957
16176900	140.780	0.131	0.553	0.024	9983.577	15879.444
16177100	100.458	0.108	0.485	0.023	9964.658	15440.234
16512113	147.039	0.232	0.324	0.023	10857.111	10928.991
17725644	184.334	0.486	0.190	0.024	15400.158	8741.137
17725844	188.580	0.481	0.196	0.023	15453.427	8987.934
17726044	176.494	0.476	0.186	0.023	15535.003	8718.583
17726244	159.890	0.472	0.171	0.023	15649.587	8516.929
17726444	109.519	0.446	0.127	0.023	15632.720	8274.600
17742085	101.937	0.018	0.618	0.011	6775.410	8969.137
17742285	110.475	0.024	0.751	0.013	6825.274	8778.151
17742485	108.812	0.029	0.813	0.015	6790.138	8620.067
17742685	100.120	0.031	0.857	0.016	6876.244	8548.510
18441284	227.144	0.841	0.133	0.024	27012.989	7350.433
19138301	121.491	0.061	0.990	0.023	10145.817	21274.234
19138501	228.562	0.110	0.990	0.023	11359.481	21274.234
19138701	236.943	0.114	0.990	0.023	11330.807	21290.425
19254018	138.807	0.266	0.268	0.024	5237.969	6057.802
19254218	208.283	0.411	0.251	0.023	5847.392	5937.443
19254418	176.956	0.656	0.135	0.023	6985.160	5191.916
19254618	110.026	0.899	0.064	0.023	8421.662	4404.301
19926150	134.071	0.856	0.081	0.023	9786.664	8080.429
19926350	130.642	0.611	0.110	0.023	8873.195	8718.400
19926550	109.032	0.445	0.128	0.023	8420.547	8935.445
20763118	117.891	0.062	0.980	0.023	4821.258	11329.398
20763318	167.235	0.087	0.975	0.024	5221.899	11306.070
20763518	208.727	0.117	0.883	0.024	5596.319	11051.005
20763718	196.094	0.144	0.680	0.023	5869.359	10940.450
20763918	110.488	0.179	0.320	0.023	6464.270	9175.157
24783846	188.423	0.931	0.102	0.024	7063.042	5412.213
24784046	288.036	0.939	0.147	0.024	7095.379	6348.382
24784246	301.998	0.955	0.150	0.024	7127.590	6090.196
24784446	220.955	0.955	0.114	0.023	7127.887	4819.261
24784646	154.536	0.945	0.083	0.023	7102.932	3928.361
24784846	105.676	0.809	0.068	0.023	6834.772	3726.665

Table S3B. The location, BF of shared hotspot, posterior and prior probabilities of hotspot, and recombination rate for the locations within chromosome 21 with BF of shared hotspot for YRI and chimpanzee that is ≥ 100. Note that the bin size was set to be 200 bp in the analysis.

	BF of	Posterior prob.			ρ per kb	
Location	shared hotspot	YRI	Chimp	Prior prob.	YRI	Chimp
27777974	188.311	0.736	0.128	0.023	6599.743	5083.321
27778174	253.744	0.738	0.166	0.023	6599.454	5188.921
27778374	284.893	0.735	0.184	0.023	6608.043	5106.888
27778574	303.014	0.729	0.196	0.023	6616.955	4807.713
27778774	337.623	0.709	0.220	0.023	6638.771	4821.803
27778974	403.944	0.677	0.268	0.023	6689.960	5069.792
27779174	420.534	0.611	0.308	0.023	6642.483	5422.041
27779374	329.147	0.496	0.309	0.023	6305.686	5689.206
27779574	232.713	0.399	0.285	0.024	6051.800	5714.347
27779774	135.506	0.303	0.230	0.023	5892.818	5261.390
32802712	106.588	0.706	0.079	0.023	10935.955	3665.847
32802912	163.649	0.585	0.142	0.023	10672.446	4579.081
32803112	101.419	0.307	0.172	0.023	10012.219	4888.483
32810512	107.456	0.990	0.057	0.024	12253.264	7098.364
32810712	100.499	0.990	0.053	0.024	12241.436	7030.004
33231354	126.159	0.763	0.085	0.023	6002.980	17892.281
33231554	160.735	0.773	0.105	0.023	5999.274	17992.985
33231754	177.498	0.777	0.115	0.024	6002.654	17483.822
33231954	175.287	0.782	0.113	0.023	6030.923	15835.609
33232154	149.576	0.790	0.097	0.023	6093.150	13010.088
33232354	105.112	0.798	0.069	0.023	6181.327	8416.882
36526950	161.245	0.094	0.872	0.024	5689.540	12675.394
36527150	192.922	0.115	0.840	0.024	7023.180	13048.529
38639126	211.700	0.111	0.943	0.023	6955.501	7719.173
38639326	640.579	0.272	0.958	0.023	13896.870	7722.063
38639526	644.525	0.273	0.963	0.023	13650.934	7721.398
38639726	468.696	0.218	0.945	0.024	11312.170	7696.826
38639926	229.270	0.165	0.683	0.023	9802.986	7655.443
39039063	211.825	0.241	0.420	0.023	5590.545	17544.685
39039263	532.540	0.331	0.670	0.023	5842.511	19470.258
39039463	774.496	0.355	0.830	0.023	5846.362	20047.953
39039663	900.318	0.331	0.990	0.023	5515.199	20643.148
39039863	832.826	0.315	0.990	0.023	5368.431	20658.812
39040063	239.249	0.304	0.381	0.023	5355.595	20690.915

Table S3C. The location, BF of shared hotspot, posterior and prior probabilities of hotspot, and recombination rate for the locations within chromosome 21 with BF of shared hotspot for YRI and chimpanzee that is ≥ 100. Note that the bin size was set to be 200 bp in the analysis.

	BF of	Posterior prob.			ρ per kb	
Location	shared hotspot	YRI	Chimp	Prior prob.	YRI	Chimp
40014559	129.356	0.078	0.857	0.024	5695.848	11674.009
40014759	183.894	0.108	0.855	0.024	6243.054	11712.068
40014959	239.070	0.137	0.855	0.024	6727.649	11733.253
40015159	270.714	0.154	0.848	0.024	6943.860	11781.198
40015359	287.940	0.168	0.822	0.024	7660.602	11845.215
40015559	166.870	0.160	0.527	0.024	7913.520	10115.673
40276752	168.105	0.172	0.493	0.023	14171.093	8830.327
40276952	528.521	0.737	0.307	0.024	22147.608	8162.475
40277152	297.215	0.775	0.183	0.024	22035.319	7650.460
40316392	133.867	0.117	0.573	0.023	3102.026	4172.360
40316592	159.642	0.136	0.581	0.023	3226.749	4225.450
40316792	184.309	0.152	0.597	0.023	3354.056	4403.064
40316992	198.542	0.163	0.597	0.023	3442.941	4415.224
40317192	204.654	0.171	0.591	0.023	3471.018	4379.222
40317392	199.521	0.181	0.548	0.023	3512.845	4147.033
40317592	188.951	0.195	0.483	0.023	3608.689	4063.566
40317792	181.629	0.216	0.421	0.023	4183.590	4049.927
40317992	168.738	0.238	0.358	0.023	4987.293	4058.472
40318192	137.340	0.248	0.285	0.024	5112.354	4012.951
41143825	181.261	0.462	0.198	0.024	6220.698	3541.143
41144025	210.568	0.526	0.199	0.024	6247.228	3489.888
41144225	244.985	0.591	0.202	0.024	6446.650	3453.253
41144425	262.282	0.617	0.206	0.024	6570.222	3451.275
41144625	251.811	0.584	0.210	0.024	6439.340	3465.952
41144825	220.723	0.513	0.212	0.024	6291.991	3475.548
41145025	205.815	0.477	0.214	0.024	6201.248	3479.447
41145225	205.056	0.476	0.214	0.023	6370.582	3472.832
41145425	199.117	0.461	0.214	0.023	6433.713	3477.599
41145625	167.788	0.394	0.215	0.023	6270.621	3481.081
41277919	209.166	0.177	0.477	0.021	6037.182	10536.394
41278119	321.375	0.252	0.514	0.021	6771.420	10607.349
41278319	321.469	0.256	0.523	0.022	6729.899	10593.630
41278519	295.549	0.245	0.519	0.022	6383.708	10598.074
41278719	250.905	0.227	0.494	0.022	5954.922	10656.032
41278919	170.159	0.192	0.416	0.023	5525.400	10736.870
41383598	121.176	0.190	0.330	0.023	7128.685	6481.764
41383798	217.252	0.234	0.457	0.023	8032.590	6753.986
41383998	301.386	0.214	0.666	0.023	7475.461	7357.192
41384198	268.828	0.183	0.710	0.024	6950.994	7398.506
41384398	222.401	0.147	0.746	0.024	6662.363	7435.505
41384598	153.612	0.103	0.761	0.024	6467.212	7450.173
41612186	104.358	0.074	0.736	0.023	11466.139	12251.212

Table S3D. The location, BF of shared hotspot, posterior and prior probabilities of hotspot, and recombination rate for the locations within chromosome 21 with BF of shared hotspot for YRI and chimpanzee that is ≥ 100. Note that the bin size was set to be 200 bp in the analysis.

	BF of	Posterior prob.			ρ per kb	
Location	shared hotspot	YRI	Chimp	Prior prob.	YRI	Chimp
42059206	135.427	0.209	0.139	0.015	12562.039	11208.861
42059406	131.325	0.211	0.163	0.016	13694.442	11318.255
42226149	626.224	0.920	0.011	0.004	12325.392	5020.142
42226349	410.638	0.972	0.024	0.008	12264.831	5103.089
42226549	481.869	0.986	0.052	0.011	12222.542	5750.803
42226749	596.665	0.983	0.092	0.013	12158.181	8700.793
42226949	655.856	0.979	0.129	0.015	12129.305	9504.365
42227149	501.579	0.775	0.155	0.016	12079.859	9233.073
42227349	168.123	0.289	0.174	0.018	10833.864	9022.659
42238349	107.716	0.735	0.076	0.023	21446.246	8928.534
42570781	115.065	0.310	0.192	0.023	6233.344	7506.878
42570981	108.709	0.324	0.175	0.024	6480.774	6740.126
42571181	107.410	0.307	0.183	0.024	6578.037	6684.103
42571381	104.378	0.249	0.220	0.024	6368.259	7095.729
42819800	105.549	0.970	0.056	0.023	11806.235	5158.828
42820000	292.836	0.958	0.145	0.023	11831.413	7030.569
42820200	356.870	0.521	0.315	0.023	9029.626	9422.601
42820400	225.951	0.312	0.355	0.023	6618.535	9853.427
42820600	180.871	0.259	0.351	0.023	6419.125	9837.760
43897282	3060835.838	0.990	0.990	0.004	40191.718	90686.510
43897482	150708.119	0.957	0.938	0.008	40266.913	89782.656
43897682	561.031	0.430	0.137	0.011	37978.474	28271.525
46327614	191.394	0.232	0.412	0.023	6406.955	10933.605
46327814	1020.907	0.443	0.812	0.023	7173.002	11046.233
46328014	2208.526	0.610	0.899	0.023	7994.535	11116.588
46328214	2598.796	0.644	0.912	0.023	8039.030	11142.919
46328414	2488.649	0.650	0.889	0.023	8014.472	11157.102
46328614	1481.975	0.642	0.700	0.023	8010.726	10980.290
46328814	362.140	0.623	0.267	0.023	7975.626	9450.552
46329014	132.833	0.614	0.111	0.024	7968.658	8636.293
46813401	282.342	0.247	0.427	0.020	8419.912	9646.114
46813601	2236.206	0.842	0.590	0.021	11898.599	10162.635
46813801	2676.036	0.885	0.625	0.021	11932.661	10520.865
46814001	1655.596	0.736	0.602	0.022	11264.752	10452.088
46814201	284.072	0.326	0.376	0.022	8426.337	9852.642
47272127	111.177	0.383	0.151	0.023	9433.282	11739.240
47305828	247.356	0.161	0.745	0.023	21436.346	9604.348
47306028	1960.835	0.952	0.545	0.023	32402.412	9181.073

Table S4. The location, BF of shared hotspot, posterior and prior probabilities of hotspot, and recombination rate for the locations within β-globin region with BF of shared hotspot for Beni and chimpanzee that is ≥ 100. Note that the bin size was set to be 200 bp in the analysis.

	BF of	Posterior prob.		Prior prob.		ρ per kb	
Location	shared hotspot	Beni	Chimp	Beni	Chimp	Beni	Chimp
5207191	135.56	0.08	0.89	0.02	0.02	13586.45	8816.91
5207391	355.93	0.21	0.79	0.02	0.02	15075.34	8714.64
5207591	842.27	0.50	0.64	0.02	0.02	17242.87	8537.15
5207791	1274.07	0.81	0.51	0.02	0.02	18809.18	8420.40
5207991	1350.57	0.96	0.45	0.02	0.02	20149.54	8419.20
5208191	1286.51	0.99	0.42	0.02	0.02	20258.55	8418.88
5208391	1241.97	0.99	0.41	0.02	0.02	20194.92	8400.70
5208591	1180.78	0.99	0.40	0.02	0.02	20197.72	8427.60
5208791	884.80	0.99	0.33	0.02	0.02	20227.12	8566.47
5208991	290.92	0.98	0.14	0.02	0.02	20220.12	8529.29

Table S5. The location, BF of shared hotspot, posterior and prior probabilities of hotspot, and recombination rate for the locations within β-globin region with BF of shared hotspot for CEPH and chimpanzee that is ≥ 100. Note that the bin size was set to be 200 bp in the analysis.

	BF of	Posterior prob.		Prior prob.		ρ per kb	
Location	shared hotspot	CEPH	Chimp	CEPH	Chimp	CEPH	Chimp
5206961	371.22	0.18	0.93	0.02	0.02	18326.23	8852.13
5207161	839.78	0.36	0.89	0.02	0.02	18673.54	8816.91
5207361	1331.00	0.54	0.79	0.02	0.02	19411.68	8714.64
5207561	1511.18	0.71	0.64	0.02	0.02	20566.67	8537.15
5207761	1328.14	0.83	0.51	0.02	0.02	21541.59	8420.40
5207961	1256.48	0.92	0.45	0.02	0.02	22266.50	8419.20
5208161	1286.98	0.99	0.42	0.02	0.02	22760.68	8418.88
5208361	1243.76	0.99	0.41	0.02	0.02	22787.70	8400.70
5208561	1180.72	0.99	0.40	0.02	0.02	22788.52	8427.60
5208761	883.43	0.99	0.33	0.02	0.02	22787.79	8566.47
5208961	295.27	0.99	0.14	0.02	0.02	22814.71	8529.29

Table S6. The location, BF of shared hotspot, posterior and prior probabilities of hotspot, and recombination rate for the locations within HLA region with BF of shared hotspot for Beni and chimpanzee that is ≥ 100. Note that the bin size was set to be 200 bp in the analysis.

	BF of	Posterior prob.		Prior prob.		ρ per kb	
Location	shared hotspot	Beni	Chimp	Beni	Chimp	Beni	Chimp
32806970	143.693	0.485	0.146	0.022	0.024	13405.120	6546.155
32807170	148.140	0.449	0.163	0.023	0.024	13494.462	7360.673

Table S7. The location, BF of shared hotspot, posterior and prior probabilities of hotspot, and recombination rate for the locations within HLA region with BF of shared hotspot for CEPH and chimpanzee that is ≥ 100. Note that the bin size was set to be 200 bp in the analysis.

	BF of	Posterior prob.		Prior prob.		ρ per kb	
Location	shared hotspot	CEPH	Chimp	CEPH	Chimp	CEPH	Chimp
32806899	168.752	0.990	0.086	0.024	0.024	18246.979	4179.673
32807099	245.422	0.823	0.146	0.024	0.024	17367.281	6546.155
32807299	239.016	0.718	0.163	0.024	0.024	15555.192	7360.673
32807499	167.397	0.634	0.134	0.023	0.024	13960.294	6622.008
32902299	100.630	0.629	0.084	0.024	0.024	3603.851	3096.905
32902499	108.641	0.604	0.094	0.024	0.024	3584.889	3095.336

References

HOLLOWAY, K., V. E. LAWSON, and A. J. JEFFREYS, 2006 Allelic recombination and de novo deletions in sperm in the human beta-globin gene region. Human Molecular Genetics **15**: 1099–1111.

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