

Report

The β -Globin Recombinational Hotspot Reduces the Effects of Strong Selection around HbC, a Recently Arisen Mutation Providing Resistance to Malaria

Elizabeth T. Wood,^{1,2} Daryn A. Stover,¹ Montgomery Slatkin,³ Michael W. Nachman,² and Michael F. Hammer^{1,2}

¹Division of Biotechnology and ²Department of Ecology and Evolutionary Biology, University of Arizona, Tucson; and ³Department of Integrative Biology, University of California, Berkeley

Recombination is expected to reduce the effect of selection on the extent of linkage disequilibrium (LD), but the impact that recombinational hotspots have on sites linked to selected mutations has not been investigated. We empirically determine chromosomal linkage phase for 5.2 kb spanning the β -globin gene and hotspot. We estimate that the HbC mutation, which is positively selected because of malaria, originated <5,000 years ago and that selection coefficients are 0.04–0.09. Despite strong selection and the recent origin of the HbC allele, recombination (crossing-over or gene conversion) is observed within 1 kb 5' of the selected site on more than one-third of the HbC chromosomes sampled. The rapid decay in LD upstream of the HbC allele demonstrates the large effect the β -globin hotspot has in mitigating the effects of positive selection on linked variation.

Recombinational hotspots are a ubiquitous feature of the human genome, occurring every 60–200 kb, and likely contribute to the observed pattern of large haplotypic blocks punctuated by low linkage disequilibrium (LD) over very short (1–2-kb) distances (Gabriel et al. 2002; Jeffreys and May 2004; McVean et al. 2004). Recombination breaks up ancestral LD and produces new combinations of alleles on which natural selection can act. Positive selection increases the frequency of beneficial mutations, creating LD via genetic hitchhiking (Smith and Haigh 1974). LD has been observed over great physical distances at several genes experiencing recent selection, including loci associated with malarial resistance (Tishkoff et al. 2001; Sabeti et al. 2002; Saunders et al. 2002; Ohashi et al. 2004). The β -globin hotspot spans ~1 kb and is located ~500 bp from the selected site at the β -globin gene (Harding et al. 1997; Wall et al. 2003). The close proximity of these β -globin regions allows us, for the first time, to empirically ex-

amine the signature of selection across a region that recombines at a rate 50 (Wall et al. 2003; Winckler et al. 2005) to 90 (Schneider et al. 2002) times higher than the genomic average of 1.1 cM/Mb (Kong et al. 2004).

More than 50 years ago, Haldane (1949) and Allison (1954) proposed that elevated frequencies of hemoglobinopathies such as thalassemia and sickle cell disease, which are caused by mutations at the β -globin locus, are maintained via balancing selection (“the malarial hypothesis”). Since then, malarial-resistant alleles have been identified at several other loci, including G6PD, TNF, and HLA (Kwiatkowski 2005). Early studies of the β -globin HbC polymorphism (β 6Glu→Lys) suggested that this allele was also subject to balancing selection (Allison 1956). More recently, it has been shown that HbC provides protection against *Plasmodium falciparum* malaria without significantly reducing fitness, indicating that this allele is increasing in frequency as a result of positive directional selection (Agarwal et al. 2000; Modiano et al. 2001; Hedrick 2004; Rihet et al. 2004). Because the African HbC allele rarely exceeds frequencies of 20% and is geographically concentrated in central West Africa, it is thought that this mutation is very young (Livingstone 1976; Trabuchet et al. 1991; Modiano et al. 2001). Here, we examine the extent of LD surrounding the African HbC allele, to estimate its

Received May 17, 2005; accepted for publication July 25, 2005; electronically published August 29, 2005.

Address for correspondence and reprints: Dr. Elizabeth Wood, Department EEB, Biosciences West, University of Arizona, Tucson, AZ 85721. E-mail: ewood@u.arizona.edu

© 2005 by The American Society of Human Genetics. All rights reserved. 0002-9297/2005/7704-0012\$15.00

age and the strength of selection acting on this mutation and to test the hypothesis that the β -globin recombinational hotspot decouples the selected HbC allele from nearby upstream regions (fig. 1a).

To generate 5.2 kb of contiguous phased sequence data, we cloned two \sim 3-kb fragments that were tiled together using polymorphic sites in the overlapping region (fig. 1b). We examined 17 heterozygous and 2 homozygous individuals carrying the HbC mutation and supplemented the data set by including 18 additional African HbA chromosomes (fig. 2). Within the 5.2-kb fragment spanning the β -globin hotspot and gene region, we observe 46 segregating sites. To delineate the boundaries of the hotspot in our data, we examine the extent of LD among 35 HbA chromosomes (fig. 1c). The boundaries for the three recombinational regions correspond well with those observed in a sample of 16 African Hausa individuals (Wall et al. 2003) and in a larger study of 349 globally distributed individuals (Harding et al. 1997).

To examine the effects of selection, we compared patterns of LD on selected chromosomes with those on nonselected chromosomes. Substantially more LD is observed among HbC than HbA chromosomes (fig. 2). We employed a coalescent-based method that incorporates recombination (c), effective population size (N_e), and population growth (r) to jointly estimate the selection coefficient (s) and the time since the origin of the HbC allele (t_1). The selection coefficient is estimated using Slatkin and Bertorelle's (2001) method for estimating the likelihood of s from allele frequency and the extent of LD with a neutral marker locus. The allele age is estimated from the posterior probability distribution of t_1 as a function of s . Analyses were performed on the basis of haplotypes generated from sites -2906 and 16 (HbC), where we assume that the A allele at site 16 arose on a chromosome carrying a C at -2906 and that the frequency of HbC is 15% in the present generation. Figure 3a shows the likelihood of s given the data, under the assumption that $N_e = 10,000$ and $r = 0$, for a range of recombination rates (c). For all values of c , we reject neutrality ($P < 10^{-5}$) and conclude that s is most likely in the range 0.04 – 0.09 . Neutrality is also rejected when we incorporate a population growth rate as high as $r = 0.009$ for $c = 0.004$ and $r = 0.02$ for $c = 0.008$ (data not shown). Figure 3b shows the posterior distributions of the age of HbC for the three maximum-likelihood estimates of s obtained using three different values of c . The age of the allele using this method depends primarily on s and is affected little by c (data not shown). The estimated age is 75–150 generations ago (or 1,875–3,750 years, under the assumption of a 25-year generation time), with an upper bound <275 generations. Past growth would make the allele slightly younger, so estimates of age based on the $r = 0$ results are conservative.

Our estimates accord well with those obtained using theoretical predictions from epidemiological data, where $t_1 < 150$ generations and $s \sim 0.08$ (Hedrick 2004) (see the “Material and Methods” section [online only]).

At least three HbC chromosomes in our survey (15.8%), given at the bottom of figure 2, show evidence for crossing-over and/or gene conversion within the hotspot. Dgn66_{HbC} contains a 5' haplotype motif identical to seven HbA chromosomes (AGCGTCTGCGA from sites -3092 to -1944), which is likely the result of a single crossover event occurring between sites -1944 and 16 . Dgn99_{HbC} possesses polymorphisms ATCTC, from sites -835 to -598 , that are also found in four HbA chromosomes and may be due to gene conversion or double crossing-over. Dgn83_{HbC} has a CACA motif, at sites -1927 , -835 , -598 , and -543 , that is also found in three HbA chromosomes. However, Dgn83_{HbC} does not contain intervening polymorphisms that are identical to HbA chromosomes surveyed here and may be the result of a single cross-over event where the recombining HbA chromosome is not represented in our data set, or it may be the result of multiple cross-over and/or gene conversion events. Because of the higher rate of gene conversion relative to crossing-over (4:1 to 15:1) (Jeffreys and May 2004), it is likely that gene conversion is responsible for some of the patterns we observe. Evidence for recombination is observed in an additional five individuals (for a total of 38.1%), under the assumption that the HbC mutation occurred on the most common HbC haplotype observed (e.g., Dgn06_{HbC}) (fig. 2). An alternative explanation is that recurrent mutation occurred at site 16 -HbC or at sites described in the “recombinant” motifs in Dgn66, Dgn99, or Dgn83. However, this hypothesis requires either (1) independent mutations at site 16 -HbC on four different HbC haplotypic backgrounds or (2) that three (Dgn99), four (Dgn83), or five (Dgn66) sites have mutated 5' of the gene on both the HbA and HbC backgrounds. Thus, it seems unlikely that elevated rates of mutation would cause decay in LD observed on the HbC chromosomes.

The extent of LD surrounding a selected allele is expected to depend on the age of the allele, the strength of selection, and the rate of recombination. The recent origin ($<5,000$ years) and high selection coefficient ($0.04 < s < 0.09$) of the HbC allele estimated here are roughly comparable to those of other malarial-resistance loci, including G6PD/A- ($<11,800$ years; $0.02 < s < 0.2$) (Tishkoff et al. 2001; Saunders et al. 2002, 2005), and G6PD_{med} ($<6,500$ years; $0.01 < s < 0.2$) (Tishkoff et al. 2001; M. Saunders, M. Slatkin, M. F. Hammer, and M. W. Nachman, unpublished data). Long-range LD extends 400–1,600 kb at loci recombining near the genome average, including G6PD/A-, G6PD_{med}, and TNFSF5/726C (Tishkoff et al. 2001; Sabeti et al. 2002; Saunders et al.

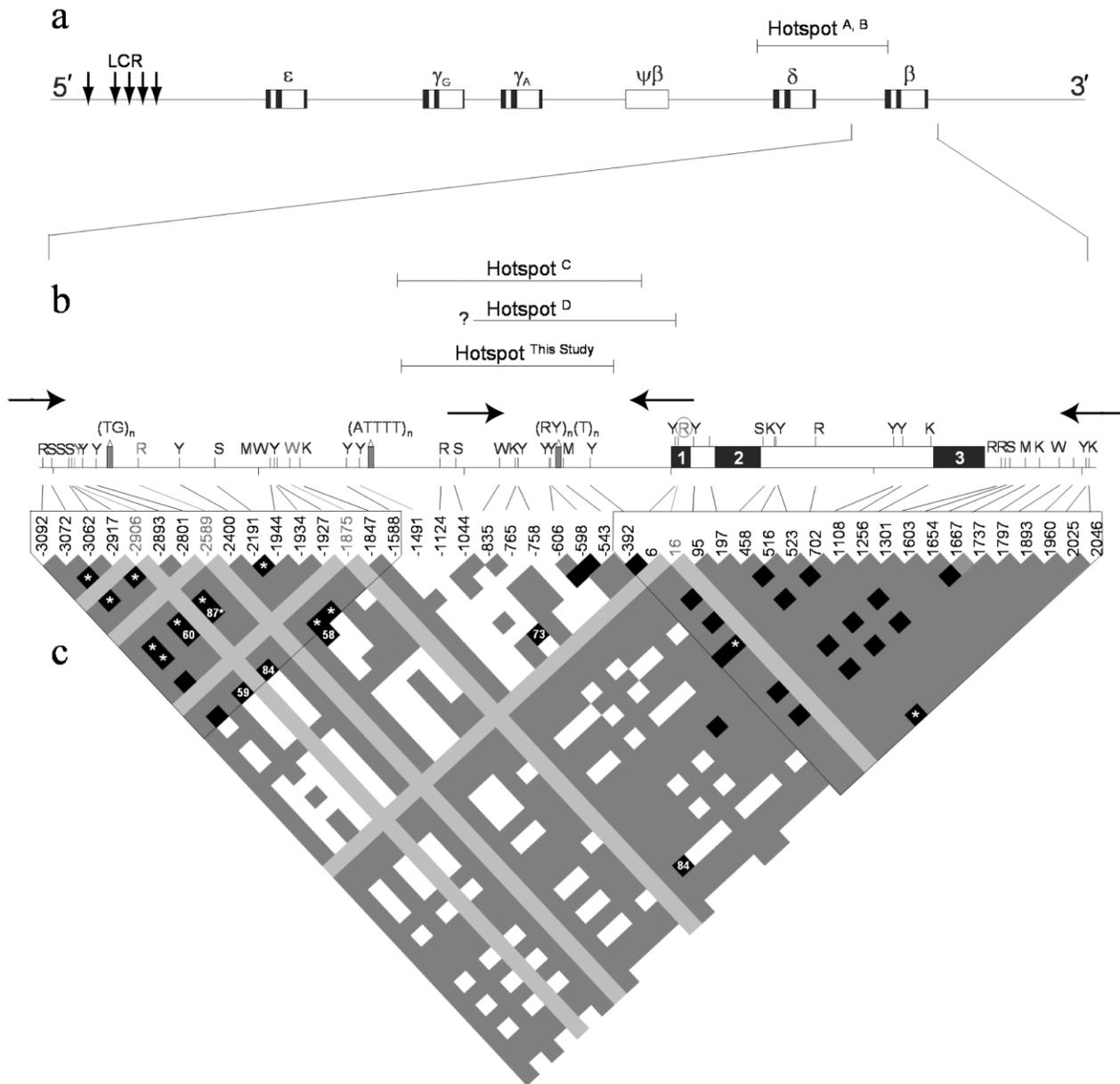


Figure 1 a, Map of the β -globin gene cluster found on 11p15.5. Four functional globin genes, one pseudogene, and the locus control region (LCR) that controls transcription and replication of the globin gene cluster (Aladjem et al. 1995) are located 5' of the β -globin gene. The position of the recombinational hotspot identified using RFLP analysis (A) (Chakravarti et al. 1984) and single-sperm typing (B) (Schneider et al. 2002) are shown above. b, Magnified view of β -globin gene and hotspot, as identified using population sequence data in 16 Africans (C) (Wall et al. 2003) and 349 globally distributed individuals (D) (Harding et al. 1997). The locations of primers that amplify the two 3-kb fragments are shown with horizontal arrows. The polymorphisms observed in a 5.2-kb region are indicated with International Union of Pure and Applied Chemistry (IUPAC) codes. Indels were excluded. Exons are shown in horizontal black boxes; introns are depicted by white boxes. Grey boxes indicate repeat motifs. The HbC mutation is circled. c, LD triangle plot of 35 HbA chromosomes in all pairwise comparisons between 42 sites. Black boxes indicate a significant D' value ($P < .01$). Values of $|D'|$ are 1 unless indicated in white numbers where $|D'|$ is multiplied by 100. Dark gray boxes indicate $|D'| = 1$ ($P > .01$), and white boxes indicate $|D'| < 1$ ($P > .01$). Light gray boxes correspond to the four segregating sites that are observed only on HbC chromosomes. Asterisks (*) indicate that pairwise comparisons are significant with a Bonferroni correction. The two regions of low recombination are outlined.

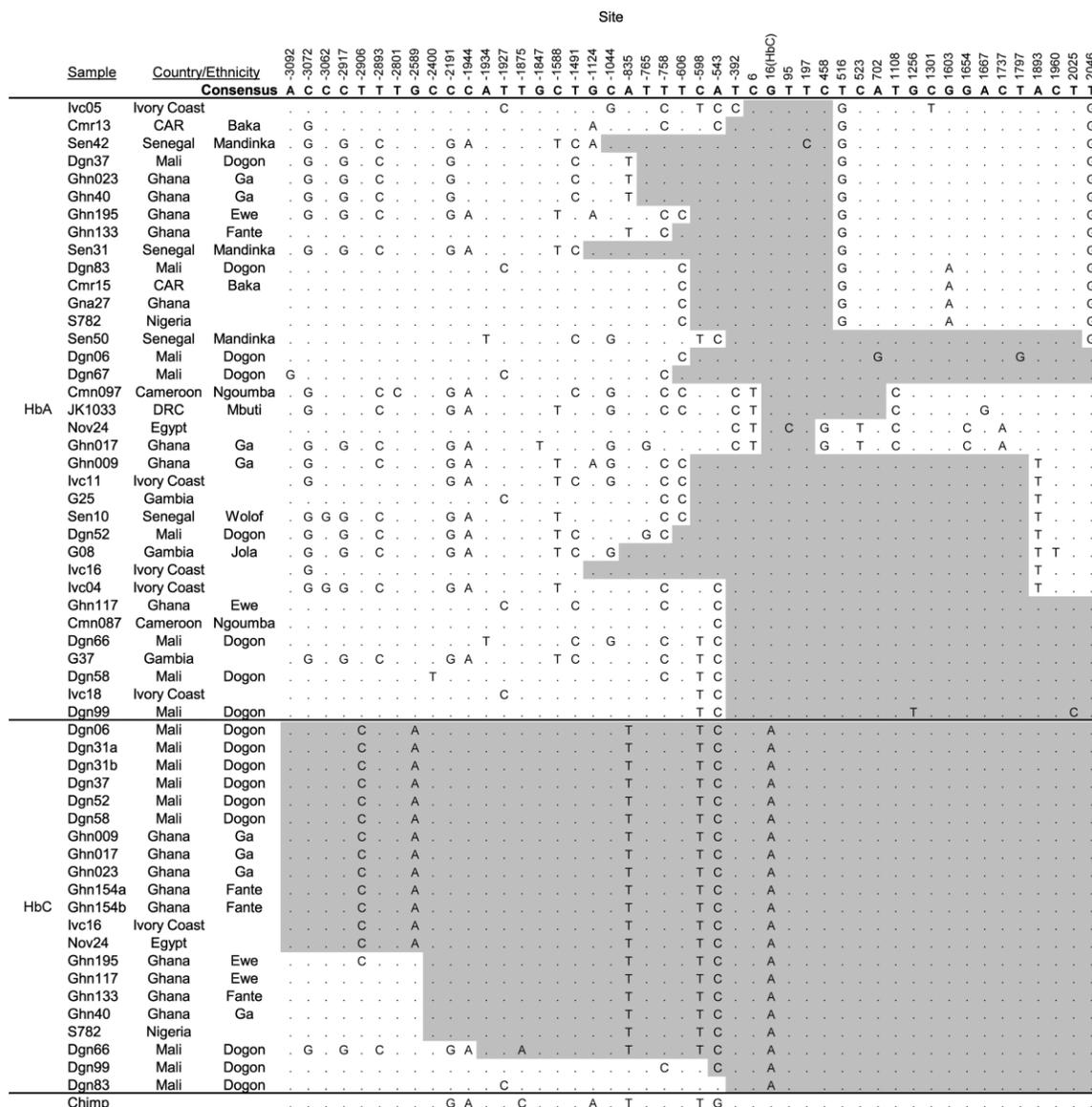


Figure 2 Polymorphisms observed in 35 HbA (top) and 21 HbC (bottom) chromosomes. To visually represent the decay in haplotype sharing, the conserved most-common long-range haplotype within the HbA and HbC chromosomes is shaded. Each site was examined sequentially, moving away from site 16. Singletons were excluded. Shaded areas include sites that are most frequent within the conserved haplotype. Recombination, as well as mutation, can cause a transition from gray to white in this depiction. CAR = Central African Republic. DRC = Democratic Republic of Congo.

2002, 2005; M. Saunders, M. Slatkin, M. F. Hammer, and M. W. Nachman, unpublished data). The β -globin locus, however, is unique in that elevated rates of recombination are sufficiently strong to reduce the effect of genetic hitchhiking on the HbC alleles to <1 kb upstream of the gene. Thus, recombinational hotspots appear to quickly erase the signal of strong and recent selection.

It has been hypothesized that recombination is evolutionarily advantageous because selection can operate

more efficiently when genes are decoupled from one another; this hypothesis predicts that recombination will be higher in gene-dense regions (Barton and Charlesworth 1998). A positive correlation between recombination and gene density (or associated features such as CpG motifs) has been observed in humans (Kong et al. 2004; McVean et al. 2004). Recombinational hotspots have been characterized at the intensely selected β -globin gene cluster and the gene-dense HLA region (Jeffreys et al. 2000, 2001; Schneider et al. 2002; Winckler et al.

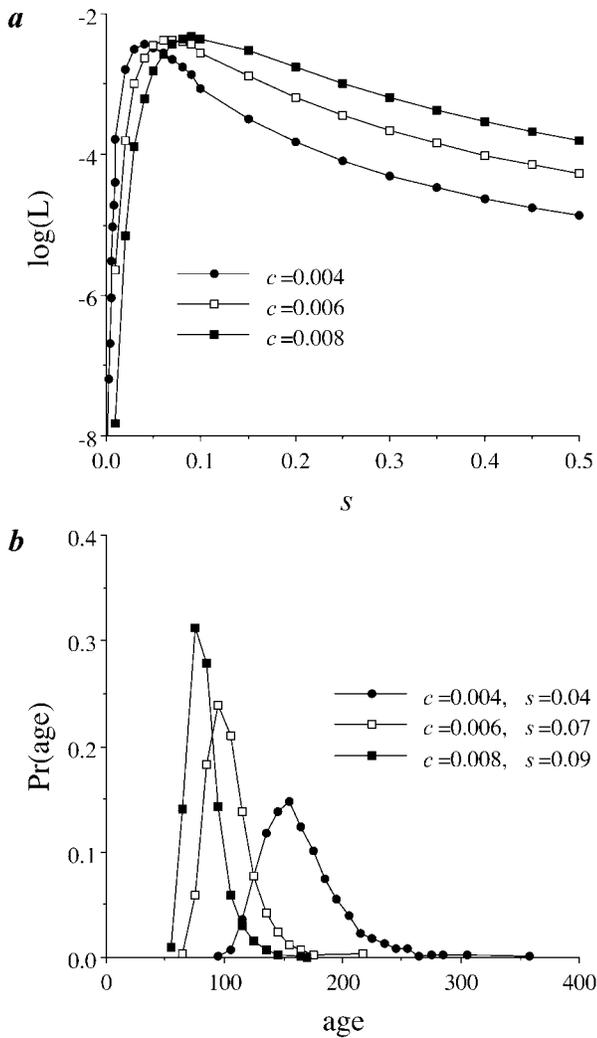


Figure 3 a, Log-likelihood surface of selection coefficient (s) under the assumption of no growth ($r = 0$), an effective population size (N_e) of 10,000, and a recombination fraction (c) ranging from 0.4%–0.8%. b, The posterior distribution of the age (in generations) of the HbC allele over the three values of s (0.04, 0.07, and 0.09) that correspond with the maximum-likelihood estimate obtained under the different values of c shown in panel a.

2005). Although it remains unclear whether this correlation is the result of selection favoring recombination near genes, our data are consistent with these observations. The location of the β -globin recombinational hotspot may allow selection to act more efficiently on malaria-resistance alleles at β -globin without significantly affecting the upstream globin gene copies and regulating locus control region sequence. Future studies will help determine the extent to which recombinational hotspots are preferentially located near regions that are under strong selective pressure.

Acknowledgments

We thank Matt Saunders and Jason Wilder, for fun discussions and helpful suggestions on the manuscript; Phil Hedrick and two anonymous reviewers, for thoughtful comments; Beverly Strassmann, for Dogon DNA samples and suggestions; and Matt Kaplan and Leigh Hunnicutt, for technical advice. This manuscript was made possible by National Science Foundation (NSF) doctoral dissertation improvement grant BCS-0220737 (to E.T.W.) and by National Institute of General Medical Sciences grant GM-53566 (to M.F.H.). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NSF or the National Institutes of Health.

References

Agarwal A, Guindo A, Cissoko Y, Taylor JG, Coulibaly D, Kone A, Kayentao K, Djimde A, Plowe CV, Doumbo O, Wellems TE, Diallo D (2000) Hemoglobin C associated with protection from severe malaria in the Dogon of Mali, a West African population with a low prevalence of hemoglobin S. *Blood* 96:2358–2363

Aladjem MI, Groudine M, Brody LL, Dieken ES, Fournier RE, Wahl GM, Epner EM (1995) Participation of the human beta-globin locus control region in initiation of DNA replication. *Science* 270:815–819

Allison AC (1954) Notes on sickle-cell polymorphism. *Ann Hum Genet* 19:39–51

——— (1956) The sickle-cell and haemoglobin C genes in some African populations. *Ann Hum Genet* 21:67–89

Barton NH, Charlesworth B (1998) Why sex and recombination? *Science* 281:1986–1990

Chakravarti A, Buetow KH, Antonarakis SE, Waber PG, Boehm CD, Kazazian HH (1984) Nonuniform recombination within the human beta-globin gene cluster. *Am J Hum Genet* 36:1239–1258

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D (2002) The structure of haplotype blocks in the human genome. *Science* 296:2225–2229

Haldane J (1949) The rate of mutation of human genes. *Hereditas Suppl* 35:267–273

Harding RM, Fullerton SM, Griffiths RC, Bond J, Cox MJ, Schneider JA, Moulin DS, Clegg JB (1997) Archaic African and Asian lineages in the genetic ancestry of modern humans. *Am J Hum Genet* 60:772–789

Hedrick P (2004) Estimation of relative fitnesses from relative risk data and the predicted future of haemoglobin alleles S and C. *J Evol Biol* 17:221–224

Jeffreys AJ, Kauppi L, Neumann R (2001) Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nat Genet* 29:217–222

Jeffreys AJ, May CA (2004) Intense and highly localized gene conversion activity in human meiotic crossover hot spots. *Nat Genet* 36:151–156

Jeffreys AJ, Ritchie A, Neumann R (2000) High resolution

- analysis of haplotype diversity and meiotic crossover in the human TAP2 recombination hotspot. *Hum Mol Genet* 9: 725–733
- Kong A, Barnard J, Gudbjartsson DF, Thorleifsson G, Jonsdottir G, Sigurdardottir S, Richardsson B, Jonsdottir J, Thorgeirsson T, Frigge ML, Lamb NE, Sherman S, Gulcher JR, Stefansson K (2004) Recombination rate and reproductive success in humans. *Nat Genet* 36:1203–1206
- Kwiatkowski DP (2005) How malaria has affected the human genome and what human genetics can teach us about malaria. *Am J Hum Genet* 77:171–190
- Livingstone FB (1976) Hemoglobin history in West Africa. *Hum Biol* 48:487–500
- McVean GA, Myers SR, Hunt S, Deloukas P, Bentley DR, Donnelly P (2004) The fine-scale structure of recombination rate variation in the human genome. *Science* 304:581–584
- Modiano D, Luoni G, Sirima BS, Simporé J, Verra F, Konate A, Rastrelli E, Olivieri A, Calissano C, Paganotti GM, D'Urbano L, Sanou I, Sawadogo A, Modiano G, Coluzzi M (2001) Haemoglobin C protects against clinical *Plasmodium falciparum* malaria. *Nature* 414:305–308
- Ohashi J, Naka I, Patarapotikul J, Hananantachai H, Brittenham G, Looareesuwan S, Clark AG, Tokunaga K (2004) Extended linkage disequilibrium surrounding the hemoglobin E variant due to malarial selection. *Am J Hum Genet* 74:1198–1208
- Rihet P, Flori L, Tall F, Traore AS, Fumoux F (2004) Hemoglobin C is associated with reduced *Plasmodium falciparum* parasitemia and low risk of mild malaria attack. *Hum Mol Genet* 13:1–6
- Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, Ackerman HC, Campbell SJ, Altshuler D, Cooper R, Kwiatkowski D, Ward R, Lander ES (2002) Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419:832–837
- Saunders MA, Hammer MF, Nachman MW (2002) Nucleotide variability at G6PD and the signature of malarial selection in humans. *Genetics* 162:1849–1861
- Saunders MA, Slatkin M, Garner C, Hammer MF, Nachman MW (2005) The span of linkage disequilibrium caused by selection on *G6PD* in humans. *Genetics* (<http://www.genetics.org/cgi/content/abstract/genetics.105.048140v1>) (electronically published July 14, 2005; accessed August 23, 2005)
- Schneider JA, Peto TE, Boone RA, Boyce AJ, Clegg JB (2002) Direct measurement of the male recombination fraction in the human beta-globin hot spot. *Hum Mol Genet* 11:207–215
- Slatkin M, Bertorelle G (2001) The use of intraallelic variability for testing neutrality and estimating population growth rate. *Genetics* 158:865–874
- Smith JM, Haigh J (1974) The hitch-hiking effect of a favourable gene. *Genet Res* 23:23–35
- Tishkoff SA, Varkonyi R, Cahinhinan N, Abbes S, Argyropoulos G, Destro-Bisol G, Drousiotou A, Dangerfield B, Lefranc G, Loiselet J, Piro A, Stoneking M, Tagarelli A, Tagarelli G, Touma EH, Williams SM, Clark AG (2001) Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. *Science* 293:455–462
- Trabuchet G, Elion J, Dunda O, Lapoumeroulie C, Ducrocq R, Nadifi S, Zohoun I, et al (1991) Nucleotide sequence evidence of the unicentric origin of the beta C mutation in Africa. *Hum Genet* 87:597–601
- Wall JD, Frisse LA, Hudson RR, Di Rienzo A (2003) Comparative linkage-disequilibrium analysis of the β -globin hot spot in primates. *Am J Hum Genet* 73:1330–1340
- Winckler W, Myers SR, Richter DJ, Onofrio RC, McDonald GJ, Bontrop RE, McVean GA, Gabriel SB, Reich D, Donnelly P, Altshuler D (2005) Comparison of fine-scale recombination rates in humans and chimpanzees. *Science* 308: 107–111