

Published in final edited form as:

Nat Genet. 2009 May ; 41(5): 625–629. doi:10.1038/ng.346.

The rise and fall of a human recombination hot spot

Alec J. Jeffreys* and Rita Neumann

Department of Genetics, University of Leicester, Leicester LE1 7RH, UK

Abstract

Human meiotic crossovers mainly cluster into narrow hot spots¹ that profoundly influence patterns of haplotype diversity² and which may also impact on genome instability³ and sequence evolution⁴⁻⁶. Hot spots also appear to be ephemeral⁷⁻⁹ but processes of hot-spot activation and their subsequent evolutionary dynamics remain unknown. We now analyse the life cycle of a recombination hot spot. Sperm typing revealed a polymorphic hot spot that was activated in *cis* by a single base change, providing evidence for a primary sequence determinant necessary, though not sufficient, to activate recombination. This activating mutation occurred roughly 70,000 years ago and has persisted to the present, most likely fortuitously through genetic drift despite its systematic elimination by biased gene conversion. Nonetheless, this self-destructive conversion will eventually lead to hot-spot extinction. These findings define a subclass of highly transient hot spots and highlight the importance of understanding hot-spot turnover and how it influences haplotype diversity.

Human crossover hot spots can be efficiently inferred from patterns of haplotype diversity¹, but direct high-resolution analysis is only possible by screening sperm for recombinant DNA molecules². Most sperm studies have targeted regions showing linkage disequilibrium (LD) breakdown, revealing narrow (1-2 kb) hot spots^{2,10}. An alternative approach is now possible through mapping crossovers in large pedigrees by high density SNP typing¹¹, although resolution is currently limited by marker density and small numbers of offspring. We therefore combined the pedigree and LD approaches by screening the linkage maps of Coop *et al.* (ref. 11) for clusters of male crossovers that mapped to a single LD hot spot inferred by coalescent analysis of Phase II HapMap data¹². The strongest cluster was in the terminal region of chromosome 3p (Fig. 1a), with 7 crossovers seen in 364 male meioses implying an intense hot spot with a recombination frequency (RF) of 2%. The lack of female crossovers does not imply male specificity – the crossover data are compatible with a female RF as high as 0.8%.

SNP genotypes of a panel of 94 semen donors of north European origin confirmed the presence of an LD hot spot, as shown by metric LD mapping¹³ (Fig. 1b), which was further analysed by recovering and mapping sperm crossover molecules. The first man tested (man 1) showed a single 1.5-kb wide crossover hot spot termed S1 that coincided with the region of LD breakdown (Fig. 1c). The RF was 0.15%, lower than the 2% predicted from the linkage map. This discrepancy almost certainly results from our selecting the most extreme cluster of familial crossovers; also, not all of the familial exchanges necessarily map to this hot spot. A second man (man 2) showed hot-spot S1 plus an additional nearby hot spot termed S2 with an RF of 0.1% and width of 1.0 kb. This very active hot spot however lies in a region of intense LD and is not evident in the metric LD map of the semen donors, nor of any of the populations typed by HapMap (not shown).

* to whom correspondence should be addressed. tel. +44 116 2523435 fax. +44 116 2523378 email ajj@le.ac.uk.

AUTHOR CONTRIBUTIONS

A.J.J. and R.N. designed the study and performed the analyses, and A.J.J. wrote the paper.

Hot-spots S1 and S2 both showed biased gene conversion accompanying crossover¹⁴, with significantly distorted transmission ratios into crossover progeny ($P < 0.001$) especially for markers closest to the hot-spot centre (S7.1C/T in S1 and S9.1G/A in S2, located ~3 bp and ~60 bp from centre respectively) (Fig. 1d). Man 2 showed opposed conversion bias, with alleles S7.1C and S9.1G on the same haplotype being over- and under-transmitted respectively. These distortions arise either through biased mismatch repair of heteroduplex DNA or through disparities in the frequency of crossover initiation on different chromosomes leading to under-transmission of markers from the more active chromosome¹⁴. If the latter is correct, then man 2 shows hot-spot S1 preferentially active on one chromosome and hot-spot S2 active on the other, with allele S9.1G marking the S2-active chromosome.

To explore this polymorphism, we analysed 15 additional informative men (see Supplementary Fig. 1). Hot-spot S1 was active in everyone, with little variation in crossover activity (mean RF 0.28%, range 0.11–0.56%) (Fig. 2a). In contrast, hot-spot S2 was only active in the seven S9.1G carriers (mean RF 0.17%, range 0.10–0.34%), while all ten S9.1A/A homozygotes were suppressed at least 100-fold (mean RF 0.0017% across S2, including possible background crossovers²). This perfect association between hot-spot S2 activity and the presence of S9.1G is highly significant ($P = 0.00005$, or 0.0012 after Bonferroni correction for additional markers tested across this region).

Most S7.1C/T heterozygotes showed biased gene conversion in hot-spot S1 but with modest transmission distortion (typically 62:38)¹⁰ and inconsistency in direction in favour of C or T alleles (Fig. 2b). Biased conversion is not therefore controlled by S7.1C/T itself, nor does it correlate with allelic status at any nearby SNP, and presumably arises through minor variation in crossover initiation efficiency on different chromosomes^{10,14}. In contrast, all men active at hot-spot S2 showed strong under-transmission of the S9.1G allele (mean 74:26 A:G), pointing to a significant association between this allele and an active haplotype ($P = 0.008$). S9.1 is the only SNP in the genotyped region to show perfect association with hot-spot activity and is also the SNP nearest the centre of the hot spot, a significant positional correlation ($P = 0.032$). Together, this strongly suggests that the S9.1G variant is directly responsible for promoting recombination initiation at hot-spot S2.

The evolutionary origin of this activating variant was explored by establishing haplotypes across hot-spot S2 for all 94 semen donors (Fig. 3). Eleven of the 16 different haplotypes identified, including all common haplotypes, could be assembled into a simple phylogeny evolving by sequential base substitution without recombination, with only one SNP (at a CpG doublet) showing homoplasmy through recombination or recurrent mutation (Fig. 3b). This is consistent with the strong LD seen across this hot spot (Fig. 1b). While this minimum recombination phylogeny might not be accurate, alternative less parsimonious phylogenies (Supplementary Fig. 2) were less plausible. Most phylogenies showed that the activating substitution S9.1A->G, restricted to haplotypes J and K, arose by mutation on the inactive haplotype I, the most common haplotype (23% frequency) that is otherwise identical in sequence to haplotype J. Subsequently, only one other variant has arisen on the S9.1G lineage. This low sequence divergence on lineages I–K indicates that the S9.1A->G mutation occurred recently (very roughly 70,000 years ago, see Methods).

The minimum recombination phylogeny identified only six recombinant haplotypes (five different, L–P), each of which could be explained by simple exchange between two non-recombinant haplotypes (Fig. 3c). The association between recombinants and rare haplotypes is significant ($P = 0.005$) and is consistent with their recent origin. Five recombinants showed exchange in hot-spot S2, with at least two (haplotypes L, M) and up to four being derived from the active lineages J and K. This enrichment of recombinant

haplotypes derived from the low-frequency J/K clade is significant ($P = 0.003$) and suggests historical crossovers occurring preferentially on these active lineages. Such historical crossovers are expected, given the contemporary sperm RF (see Methods). Furthermore, analysis of semen donor genotypes using LDhat1 predicted a historical crossover frequency of 0.006% across the hot-spot S2 region, in good agreement with the current RF of 0.006% estimated from sperm RFs averaged over all men and assuming no female crossover. All evidence therefore points to an early switch-on of hot-spot S2 activity following the S9.1A->G mutation.

The activating mutation S9.1G is systematically under-transmitted to only 26% of crossover progeny from S9.1G/A heterozygotes (Fig. 2b). This predicts that 0.04% of S9.1G alleles will be lost per generation through this recombination-based meiotic drive, sufficient to block the spread of the activating mutation through a population^{14,15}. The question thus arises as to how the S9.1G mutation could have persisted so long. We therefore carried out forward simulations of populations each seeded with a single founding mutation and asked whether any could evolve to the current S9.1G population frequency of 3.4% after 70,000 years and if so, how likely this outcome was compared to an unbiased founder mutation showing strictly 50:50 segregation. Surprisingly, the S9.1G mutation was perfectly capable of achieving this target (Fig. 4), and did so about 60% as often as an unbiased mutation. However, further evolution of current populations resulted in elimination of the recombination-promoting S9.1G allele in >99.9% of simulations. Using historical effective population sizes, this extinction occurred on average after 50,000 years though with wide bounds (8,000–150,000 years). Hot-spot S2 was therefore doomed to extinction from the outset by the very mutation that created it, although genetic drift along the road to extinction can result in high, if transient, hot-spot frequencies in a population (Fig. 4).

Combining these birth and death data for historical population sizes allowed the lifespan of the hot spot to be predicted at just 120,000 years, though spans as short as 30,000 years or as long as 270,000 years are possible. However, large contemporary population sizes and changes in the future will alter the time to extinction to an unknowable extent, making prediction of future lifespan impossible. If hot-spot S2 is also active in gene conversion without crossover¹⁶, this could increase the strength of meiotic drive and shorten the lifespan; however, preliminary sperm analyses indicated that such conversions at S9.1A/G are uncommon, at <40% of the crossover frequency, and will not significantly perturb these age estimates.

How did the S9.1A->G mutation activate the hot spot? It does not affect a sequence motif associated with human crossover hot spots¹⁷ but instead maps within an *AluSq* repeat. We identified 68 other *Alus* in the human genome perfectly matched to the 21 bp sequence centred on S9.1G, but only 7 were located within an LD hot spot identified by Phase II HapMap¹², similar to the 4.5 expected for no association. Instead, 78% mapped to regions of complete LD with no evidence whatsoever for historical recombination in CEU individuals. The S9.1A->G mutation therefore appears necessary to activate crossover initiation at S2, but by itself is not sufficient to create a hot spot, suggesting that activation requires additional factors. One possible clue comes from the close proximity of hot-spots S1 and S2. We have described five other hot-spot pairs separated by 1.4–2.3 kb, with two cases where one member of the pair exhibits presence/absence polymorphism as seen at hot-spot S29,10,18. Maybe the presence of a hot spot somehow favours the formation of a new hot spot nearby, though it is unclear how this operates mechanistically to create the fairly uniform 2 kb spacing.

Previous work has shown substantial divergence in hot-spot locations between humans and chimpanzees as inferred from LD studies^{7,8}, implying major hot-spot turnover on a

timescale of 5 Myr. The present study has revealed a very fast-evolving and ephemeral hot spot, appearing and disappearing on a much shorter timescale. It also provides the first direct evidence that a hot spot can be activated in *cis* by a primary DNA sequence change. The persistence of this hot spot can be explained by fortuitous genetic drift; there is no need to invoke hypotheses such as activation in *trans*¹⁹ or early immunity to meiotic drive²⁰ to explain survival despite systematic elimination of the activating mutation by biased gene conversion^{14,19-21}. This “hot-spot conversion paradox”²² thus disappears for hot-spot S2. This is however not the case for active hot spots such as S1 which are largely if not completely fixed in humans¹⁰; drift will not allow such hot spots to achieve high population frequencies unless they are activated without a primary sequence change¹⁸ and succeed in avoiding subsequent silencing mutations in *cis* or somehow become immune to the effects of these mutations^{19,20}. One challenge now is to find other examples of *cis*-activated ephemeral hot spots, perhaps by searching the genome for regions that harbour recombinant haplotypes preferentially derived from a restricted haplotypic lineage, as seen in this study. Another challenge is to understand how apparently fixed hot spots can arise and persist in the human genome.

METHODS

Genotyping, haplotyping and sequencing

Semen samples were collected with approval from the Leicestershire Health Authority Research Ethics Committee and with informed consent. DNAs were extracted and manipulated under conditions designed to minimise the risk of contamination²³. Routine genotyping of dbSNPs was performed as described elsewhere²⁴ on genomic DNAs that had been whole genome amplified using a GenomiPhi HY DNA amplification kit (GE Healthcare Bio-Sciences, Little Chalfont, Bucks, UK). Fifteen semen donors, including 14 suitable for sperm crossover analysis, were further analysed by haplotype separation using allele-specific PCR followed by DNA sequencing. The 30 fully sequenced haplotypes were derived from lineages A, B, C, F, G, I, J, K, L and M (8, 2, 1, 5, 1, 5, 4, 2, 1 and 1 copies respectively) and yielded 14 additional SNPs and indels (see Supplementary Methods), all of which were typed across the semen donor panel. The experimentally determined haplotypes were used to infer the phase of haplotypes from diplotype data on the remaining 79 semen donors. All rare haplotypes were verified by haplotype separation and SNP typing.

Sperm crossover assays

Seventeen men were selected for crossover analysis. Details of the crossover assays are provided in Supplementary Methods. In brief, the linkage phase of markers across hot-spot S1 in each man was established using allele-specific PCR primers (ASPs) designed for crossover recovery. For eleven men, these ASPs were then used in repulsion phase to selectively amplify crossover molecules from batches of sperm DNA containing typically 0.5-2.0 crossover molecules per PCR reaction, followed by a second round of nested repulsion-phase allele-specific PCR. Sperm DNA concentrations were quantified on a NanoDrop 1000 spectrophotometer, and crossover frequencies calculated assuming a single DNA molecule amplification efficiency of 50% (one amplifiable molecule of each haplotype per 12 pg sperm DNA, as established from extensive previous data on single-molecule long PCR^{2,24}). In five cases lacking suitable heterozygosities, one ASP in the initial PCR was replaced by a universal (non-allele specific) primer; selective amplification of crossovers was still sufficient in the secondary PCR for crossover detection. In one case lacking any suitable heterozygosities downstream of hot-spot S2, a “half-crossover assay” was employed on very small pools of sperm DNA each containing 50 amplifiable molecules per haplotype, using nested ASPs upstream of S1 plus universal primers downstream to selectively amplify one haplotype. Allele-specific oligonucleotide (ASO) hybridisation was then used to detect

recombinants by the presence of markers from the non-amplified haplotype16. These assays typically yielded 120 crossover molecules per man per orientation of the assay, generating in total 4200 crossover molecules recovered from 2.8×10^6 sperm. All were mapped by ASO hybridisation to locate crossover exchange points, as described elsewhere9.

Bioinformatic and population genetic analyses

Familial crossover intervals were taken from Coop *et al.* (ref. 11). LD hot-spot locations identified by Phase II HapMap (release 21 Phase I + II data) were downloaded from the International HapMap Project (<http://www.hapmap.org>). Human, chimp and orangutan DNA sequences were downloaded from Ensembl. *Alu* motifs were identified by BLAT searches of the human genome sequence (build 36) using the UCSC genome browser. Metric LD maps were constructed using LDMAP (<http://cedar.genetics.soton.ac.uk/pub/PROGRAMS/LDMAP>) and historical recombination rates were estimated by coalescent analysis using LDhat1 (<http://www.stats.ox.ac.uk/~mcvean/LDhat/>), as described previously10. The population frequency of the activating S9.1G allele (3.4%) was estimated from the semen donor panel of 94 individuals and from 92 UK and 170 French individuals plus 114 CEPH parents/grandparents, all of north European origin.

Phylogenetic analysis

Haplotype phylogenies were constructed by reference to the ancestral sequence determined by comparing human, chimpanzee and orangutan sequences. All possible combinations of different haplotypes, including the ancestral haplotype, were analysed by the four-gamete test25 over all segregating sites to identify the largest cluster showing no evidence of historical recombination. Remaining haplotypes had to be recombinant with respect to this cluster, and their likely parental haplotypes could be easily determined by comparison with non-recombinant haplotypes. Alternative phylogenies were explored by forcing each of these “recombinant” haplotypes to be an obligate non-recombinant in the cluster analysis, and are discussed in Supplementary Fig. 2.

Estimating the age of the S9.1G mutation

A base substitution rate of 1.96×10^{-9} /year was estimated from the number of substitutions accumulated in the 4.3-kb interval spanning hot-spot S2 (Fig. 3) since the human-chimp divergence, assumed to be 5.5 Myr ago. This rate is 1.7-fold higher than predicted from overall human/chimp divergence and may reflect atypical mutation processes in this subtelomeric region26. We used this local rate plus the method of Thomson *et al.* (ref. 27) to date the age of the haplotype JK clade (6 fully sequenced representatives) and IJK clades (11 sequenced members) at 40,000 and 90,000 years respectively, yielding a rough estimate of the age of the S9.1G mutation of 70,000 years (Fig. 3). A similar estimate of 80,000 years was obtained by maximum-likelihood analysis of mutations on the IJK phylogeny with varying branch lengths, assuming that mutations arise at random and that the two mutations seen in the IJK clade show the same mutation rate as at other sites in this region of DNA. Confidence intervals, estimated for clade ages by the method of Hudson (ref. 28) and for the age of the S9.1G mutation from likelihood ratios, were all very broad; for example, 4,000–280,000 years for the age of the IJK clade and 7,000–270,000 years for the age of the mutation.

Evolutionary trajectories of the S9.1G mutation

Forward simulations to investigate the persistence and extinction of S9.1G used a randomly-mating population of 10,000 diploid individuals29 and a generation time of 20 years. Simulations commencing with a single founder mutation were repeated until 100 successful simulations over 3500 generations (70,000 years), with or without crossover and biased gene

conversion in S9.1G/A heterozygotes, had been accumulated, each of which had arrived at the contemporary mutant frequency of 3.4%. Subsequent extinctions were monitored using an initial population frequency of 3.4%, then continuing mating until S9.1G extinction.

Estimating the incidence of historical crossovers

Monitoring genealogies in forward simulations showed that 7,900 generations (range 4,500–14,000 in 100 simulations) would have accumulated since the founder mutation on the lineages leading to the seven S9.1G haplotypes sampled in Fig. 3. Given a sex-averaged RF of 0.083% plus the observed bias against S9.1G in crossover progeny, then 1.2 detectable crossovers should have occurred along these lineages, compared with one observed (haplotype L). Note that 24% of crossovers (e.g. exchanges between haplotype J and I) will not yield a detectable recombinant. Historical crossovers were investigated in more detail by forward simulations of populations seeded with haplotypes A-I at the frequency shown in Fig. 3, plus a single founder J haplotype, then monitoring the generation of detectable recombinants derived from J or from recombinant descendants still carrying the S9.1G mutation. In those simulations achieving the contemporary S9.1G frequency, seven S9.1G haplotypes were selected at random (to match the sampling in Fig. 3), together with any inactive S9.1A recombinant descendants in the same sample. The majority of simulations (99%) yielded recombinant haplotypes in the sample but their incidence was highly variable, reflecting the age of exchanges and the underlying genealogy of haplotypes; the most common outcome, seen in 40% of simulations, was 1-10 recombinants compared with 2-4 recombinants in Fig. 3.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank J. Blower and volunteers for providing semen samples, T.E. King and P. Balaesque for DNA samples, A. Webb for bioinformatics support, colleagues for helpful discussions, and the Medical Research Council, the Wellcome Trust (ref. 081227/Z/06/Z), the Royal Society and the Louis-Jeantet Foundation for funding support.

REFERENCES

1. McVean GA, et al. The fine-scale structure of recombination rate variation in the human genome. *Science*. 2004; 304:581–584. [PubMed: 15105499]
2. Jeffreys AJ, Kauppi L, Neumann R. Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nat. Genet.* 2001; 29:217–222. [PubMed: 11586303]
3. Lindsay SJ, Khajavi M, Lupski JR, Hurler ME. A chromosomal rearrangement hotspot can be identified from population genetic variation and is coincident with a hotspot for allelic recombination. *Am. J. Hum. Genet.* 2006; 79:890–902. [PubMed: 17033965]
4. Spencer CCA, et al. The influence of recombination on human genetic diversity. *PLoS Genet.* 2006; 2:e148. [PubMed: 17044736]
5. Dreszer TR, Wall GD, Haussler D, Pollard KS. Biased clustered substitutions in the human genome: the footprints of male-driven biased gene conversion. *Genome Res.* 2007; 17:1420–1430. [PubMed: 17785536]
6. Duret L, Arndt PF. The impact of recombination on nucleotide substitutions in the human genome. *PLoS Genet.* 2008; 4:e1000071. [PubMed: 18464896]
7. Ptak SE, et al. Fine-scale recombination patterns differ between chimpanzees and humans. *Nat. Genet.* 2005; 37:429–434. [PubMed: 15723063]
8. Winckler W, et al. Comparison of fine-scale recombination rates in humans and chimpanzees. *Science*. 2005; 308:107–111. [PubMed: 15705809]

9. Jeffreys AJ, Neumann R, Panayi M, Myers S, Donnelly P. Human recombination hot spots hidden in regions of strong marker association. *Nat. Genet.* 2005; 37:601–606. [PubMed: 15880103]
10. Webb AJ, Berg IL, Jeffreys A. Sperm cross-over activity in regions of the human genome showing extreme breakdown of marker association. *Proc. Natl. Acad. Sci. USA.* 2008; 105:10471–10476. [PubMed: 18650392]
11. Coop G, Wen X, Ober C, Pritchard JK, Przeworski M. High-resolution mapping of crossovers reveals extensive variation in fine-scale recombination patterns among humans. *Science.* 2008; 319:1395–1398. [PubMed: 18239090]
12. The International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature.* 2007; 449:851–861. [PubMed: 17943122]
13. Maniatis N, et al. The first linkage disequilibrium (LD) maps: delineation of hot and cold blocks by diplotype analysis. *Proc. Natl. Acad. Sci. USA.* 2002; 99:2228–2233. [PubMed: 11842208]
14. Jeffreys AJ, Neumann R. Reciprocal crossover asymmetry and meiotic drive in a human recombination hot spot. *Nat. Genet.* 2002; 31:267–271. [PubMed: 12089523]
15. Jeffreys AJ, Neumann R. Factors influencing recombination frequency and distribution in a human meiotic crossover hotspot. *Hum. Mol. Genet.* 2005; 14:2277–2287. [PubMed: 15987698]
16. Jeffreys AJ, May CA. Intense and highly localized gene conversion activity in human meiotic crossover hot spots. *Nat. Genet.* 2004; 36:151–156. [PubMed: 14704667]
17. Myers S, Freeman C, Auton A, Donnelly P, McVean G. A common sequence motif associated with recombination hot spots and genome instability in humans. *Nat. Genet.* 2008; 40:1124–1129. [PubMed: 19165926]
18. Neumann R, Jeffreys AJ. Polymorphism in the activity of human crossover hotspots independent of local DNA sequence variation. *Hum. Mol. Genet.* 2006; 15:1401–1411. [PubMed: 16543360]
19. Peters AD. A combination of *cis* and *trans* control can solve the hotspot conversion paradox. *Genetics.* 2008; 178:1579–1593. [PubMed: 18245829]
20. Coop G, Myers SR. Live hot, die young: transmission distortion in recombination hotspots. *PLoS Genet.* 2007; 3:e35. [PubMed: 17352536]
21. Pineda-Krch M, Redfield RJ. Persistence and loss of meiotic recombination hotspots. *Genetics.* 2005; 169:2319–2333. [PubMed: 15687277]
22. Boulton A, Myers RS, Redfield RJ. The hotspot conversion paradox and the evolution of meiotic recombination. *Proc. Natl. Acad. Sci. USA.* 1997; 94:8058–8063. [PubMed: 9223314]
23. Jeffreys AJ, Murray J, Neumann R. High-resolution mapping of crossovers in human sperm defines a minisatellite-associated recombination hotspot. *Mol. Cell.* 1998; 2:267–273. [PubMed: 9734365]
24. Jeffreys AJ, Ritchie A, Neumann R. High resolution analysis of haplotype diversity and meiotic crossover in the human *TAP2* recombination hotspot. *Hum. Mol. Genet.* 2000; 9:725–733. [PubMed: 10749979]
25. Hudson RR, Kaplan NL. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics.* 1985; 111:147–164. [PubMed: 4029609]
26. Arndt PF, Hwa T, Petrov DA. Substantial regional variation in substitution rates in the human genome: importance of GC content, gene density, and telomere-specific effects. *J. Mol. Evol.* 2005; 60:748–763. [PubMed: 15959677]
27. Thomson R, Pritchard JK, Shen P, Oefner PJ, Feldman MW. Recent common ancestry of human Y chromosomes: evidence from DNA sequence data. *Proc. Natl. Acad. Sci. USA.* 2000; 97:7360–7365. [PubMed: 10861004]
28. Hudson RR. The variance of coalescent time estimates from DNA sequences. *J. Mol. Evol.* 2007; 64:702–705. [PubMed: 17487522]
29. Morton, NE. *Outline of genetic epidemiology.* Karger; Basel: 1982.

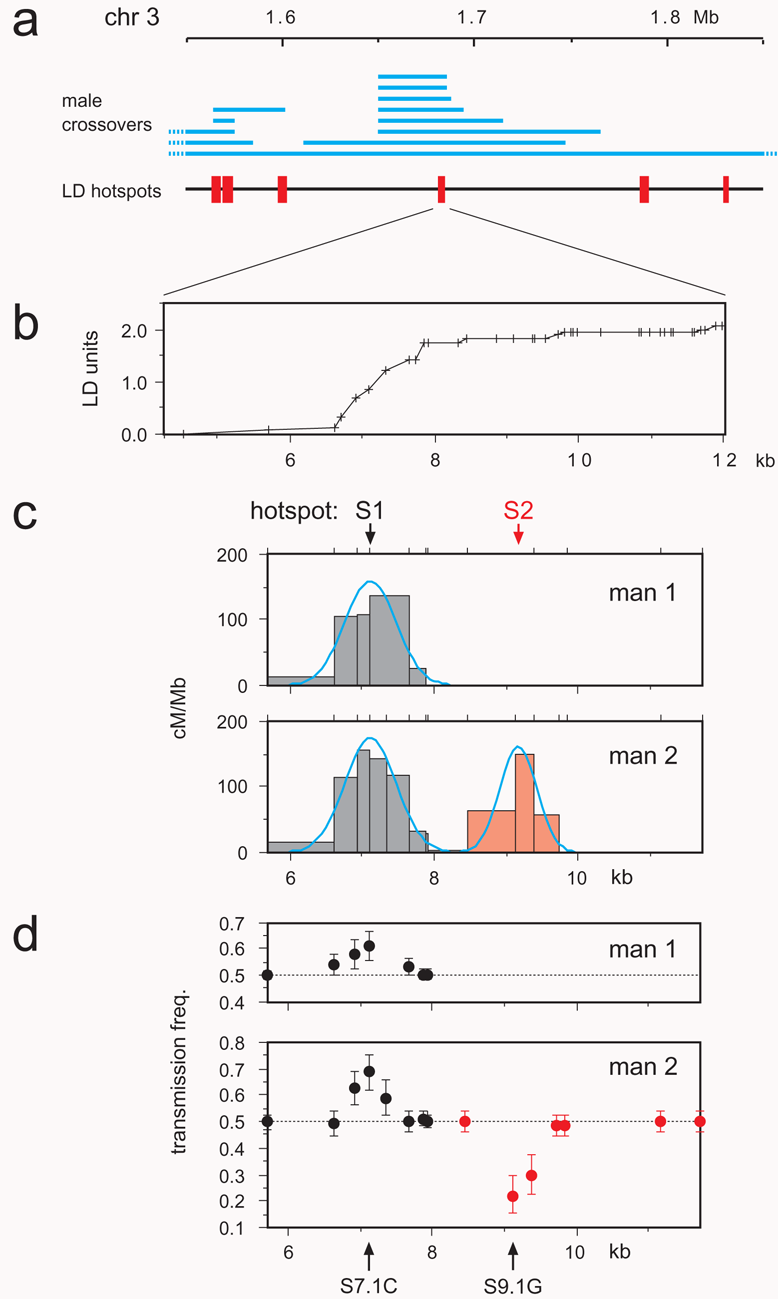


Figure 1. Identification of a sperm crossover hot spot from family linkage maps. **(a)** the strongest cluster of male crossovers in the linkage maps of Coop *et al.* (ref. 11) assignable to a single LD hot spot predicted from Phase II HapMap data¹². Crossover intervals are shown in blue. **(b)** Metric LD map across this hot spot determined from SNP genotypes of a panel of 94 semen donors, with marker locations in local units. **(c)** Sperm crossover activity across this region in two men. The number of reciprocal crossover molecules with exchange points mapping to a given interval between adjacent heterozygous SNP markers (ticks above each plot) was used to estimate crossover activity in cM/Mb. Least-square best-fit normal distributions² are shown in blue. **(d)** Biased gene conversion in crossover progeny. The

frequency of transmission of each allele on one of the two haplotypes into crossover progeny, together with 95% confidence intervals, was normalised to equal numbers of reciprocal crossover molecules¹⁴. Estimates were performed separately for the two hot spots in man 2; the alleles showing maximum transmission distortion (S7.1C, S9.1G) are on the same haplotype and therefore show opposed conversion bias.

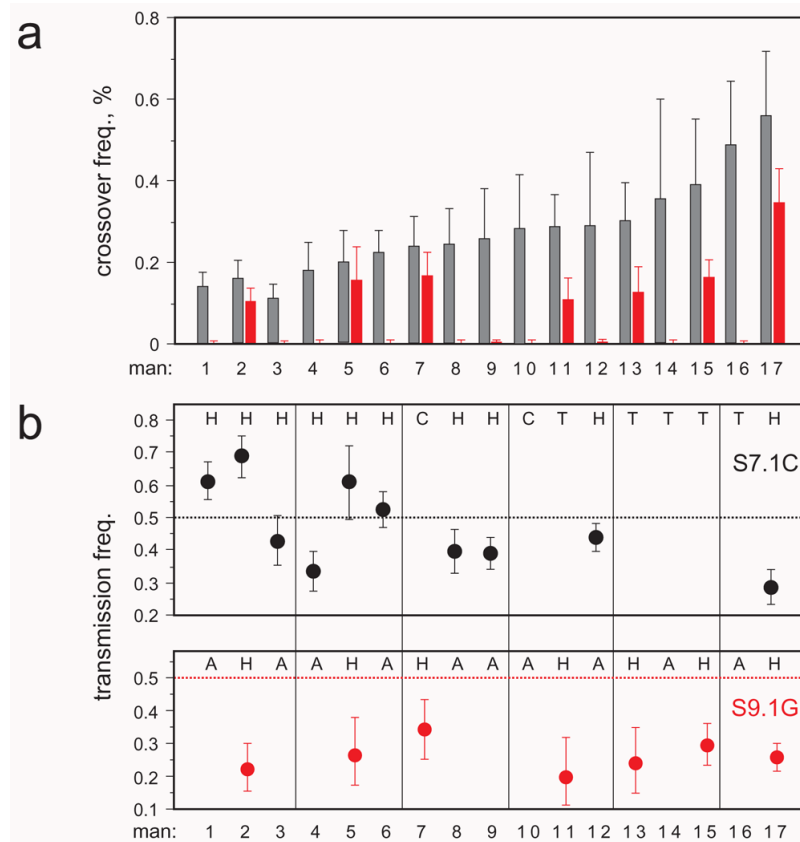


Figure 2. Variation between men in sperm crossover activity and biased gene conversion. **(a)** Crossover frequencies, with upper 95% CIs, at hot-spot S1 (grey) and S2 (red) in 17 men. Men 3-17 were ranked by activity at hot-spot S1. **(b)** Transmission frequencies at markers S7.1C/T and S9.1G/A located closest to the centre of hot-spot S1 and S2 respectively, estimated as in Fig. 1d. Genotypes for each donor are shown above, with heterozygotes indicated by H and homozygotes by the appropriate single letter.

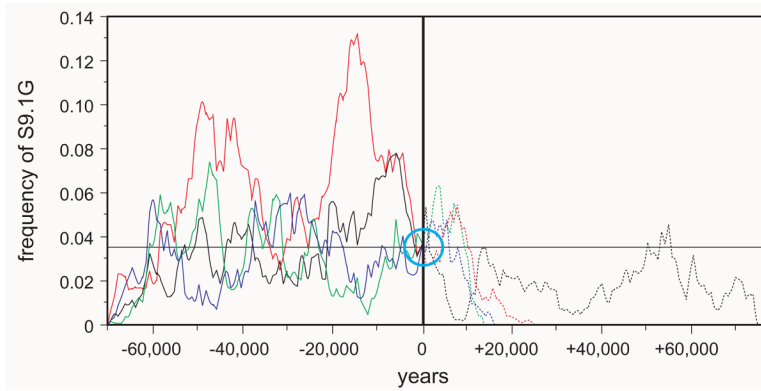


Figure 4.

Evolutionary dynamics of hot-spot S2. Simulations were performed on populations each seeded 70,000 years ago with a single S9.1A->G mutation, then allowed to evolve by random mating with crossovers occurring in S9.1A/G heterozygotes at the mean frequency seen in sperm and assuming no female crossover activity (sex-averaged crossover rate of 0.083%). Transmissions into crossover progeny were biased against S9.1G at the mean level seen in sperm crossovers (A:G 0.743:0.257). Four examples of simulations that achieved the contemporary S9.1G frequency of 3.4% (circle) are shown (solid lines). Simulations into the future (dashed lines) used the same population and crossover parameters.