Published in final edited form as:

Hum Genet. 2012 July ; 131(7): 1205–1216. doi:10.1007/s00439-012-1147-5.

Evolutionary genetics of the human Rh blood group system

George H. Perry^{1,2,*}, Yali Xue^{3,*}, Richard S. Smith^{2,4,*}, Wynn K. Meyer⁵, Minal Çalışkan⁵, Omar Yanez-Cuna^{2,6}, Arthur S. Lee², María Gutiérrez-Arcelus^{2,6}, Carole Ober⁵, Edward J. Hollox^{7,^}, Chris Tyler-Smith^{3,^}, and Charles Lee^{2,4,^}

¹ Department of Anthropology, Pennsylvania State University, University Park, PA 16801 USA

² Department of Pathology, Brigham & Women's Hospital, Boston, MA 02115, USA

³ Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK

⁴ Harvard Medical School, Boston, MA 02115 USA

⁵ Department of Human Genetics, University of Chicago, Chicago, IL 60637, USA

⁶ Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, 62210, México

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

Abstract

The evolutionary history of variation in the human Rh blood group system, determined by variants in the *RHD* and *RHCE* genes, has long been an unresolved puzzle in human genetics. Prior to medical treatments and interventions developed in the last century, the D-positive children of D-negative women were at risk for hemolytic disease of the newborn, if the mother produced anti-D antibodies following sensitization to the blood of a previous D-positive child. Given the deleterious fitness consequences of this disease, the appreciable frequencies in European populations of the responsible *RHD* gene deletion variant (for example, 0.43 in our study) seem surprising. In this study, we used new molecular and genomic data generated from four HapMap population samples to test the idea that positive selection for an as-of-yet unknown fitness benefit of the *RHD* deletion may have offset the otherwise negative fitness effects of hemolytic disease of the newborn. We found no evidence that positive natural selection affected the frequency of the *RHD* deletion. Thus, the initial rise to intermediate frequency of the *RHD* deletion in European populations may simply be explained by genetic drift/ founder effect, or by an older or more complex sweep that we are insufficiently powered to detect. However, our simulations recapitulate previous findings that selection on the *RHD* deletion is frequency dependent, and weak or absent near 0.5. Therefore, once such a frequency was achieved, it could have been maintained by a relatively small amount of genetic drift. We unexpectedly observed evidence for positive selection on the C allele of *RHCE* in non-African populations (on chromosomes with intact copies of the *RHD* gene) in the form of an unusually high F_{ST} value and the high frequency of a single haplotype carrying the C allele. RhCE function is not well understood, but the C/c antigenic variant is clinically relevant and can result in hemolytic disease of the newborn, albeit much less commonly and severely than that related to the D-negative blood type. Therefore, the potential fitness benefits of the *RHCE* allele are currently unknown but merit further exploration.

Corresponding author: Charles Lee, Ph.D. Department of Pathology Brigham & Women's Hospital Harvard Medical School 221 Longwood Ave., EBRC 422A Boston, MA 02115, USA Phone: 1-617-278-0031 Fax: 1-617-264-6861 clee@rics.bwh.harvard.edu.

*Equal contribution

^Equal contribution

Keywords

Blood group polymorphism; copy number variation; human evolution; balancing selection

INTRODUCTION

The human Rh blood group system is a collection of antigens expressed on erythrocyte cell membranes (Avent and Reid 2000) that may play a role in the transport of ammonia (Marini et al. 2000) or carbon dioxide (Endeward et al. 2008; Kustu and Inwood 2006), or not in transport but in erythrocyte membrane structure (Westhoff 2004; Westhoff and Wylie 2006). Functional (antigen) variation in the Rh blood group system is determined by insertions/deletions, single nucleotide polymorphisms (SNPs), and gene conversion events in the *RHD* and *RHCE* genes (Colin et al. 1991; Flegel 2011; Mouro et al. 1993). Homozygous deletion of the entire *RHD* gene results in the D-negative blood phenotype (Wagner and Flegel 2000), whereas the D-positive phenotype is conferred by the presence of either one or two intact copies of the *RHD* gene.

D-negative mothers may produce anti-D antibodies following exposure to red blood cells from a D-positive fetus during pregnancy or childbirth. Subsequent D-positive offspring of a D-negative mother may develop hemolytic disease of the newborn resulting in fetal death or severe disability (Levine et al. 1941; Urbaniak and Greiss 2000). Prior to treatments introduced beginning in the 1940s (culminating with Rho(D) Immune Globulin (RhoGAM) in 1968) that largely obviated these health issues (Urbaniak and Greiss 2000), D-negative mothers may have suffered reduced reproductive fitness; pre-treatment mortality from hemolytic disease of the newborn was reportedly 1 in every 56 births to D-negative women in a European-American population (1 in 392 births among all women, regardless of D status) (Potter 1947). Certain non-synonymous SNPs in the *RHCE* gene result in antigenic variation in the encoded RHCE protein (rs676785, Ser103Pro: C/c; rs609320, Pro226Ala: E/e) (Avent and Reid 2000), and differential expression of these antigens can also lead to hemolytic disease of the newborn, but with much lower frequency than that caused by anti-D antibodies, at least in European populations (Monchamont et al. 1991).

Based on the potential consequence of hemolytic disease of the newborn, one might expect strong purifying selection to have acted against the D-negative phenotype and minor alleles of the C/c and E/e antigens. Yet in European populations the D-negative phenotype is observed at substantial frequencies, typically 0.15-0.17 and up to 0.29 in the Basque (Touinssi et al. 2004; Urbaniak and Greiss 2000), and the frequency of the *RHCE*C allele is ~0.44 (Urbaniak and Greiss 2000). Similar to the effects of sickle cell and thalassemia hemoglobin heterozygosity on malarial resistance (Allen et al. 1997; Allison 1954; Flint et al. 1986; Kwiatkowski 2005), we asked whether these alleles confer an unknown fitness benefit whereby positive or balancing selection explains their otherwise surprisingly high frequencies (Feldman et al. 1969; Westhoff 2004). Because such a history may have left detectable genomic signatures, in this study we used a population genetics framework to test evolutionary hypotheses concerning these functional genetic variants of the Rh blood group system.

MATERIALS AND METHODS

Genotyping

DNA samples and cell lines from the HapMap individuals were obtained from the Coriell Institute for Medical Research. To genotype the *RHD* deletion we used a previously-developed TaqMan quantitative PCR (qPCR) assay (Lo et al. 1998), in which the forward

primer for the *RHD* amplicon (all 5'-3'; CCTCTCACTGTTGCCTGCATT) maps to the very 3' end of the gene and the reverse primer (AGTGCCTGCGCGAACATT) maps to the 5' end of the segmental duplication flanking and unique to *RHD* (Figure 1a). Therefore, the amplification is specific to *RHD* (i.e., there is no complicating signal from *RHCE*). Primers and the internal probe

([FAM]TACGTGAGAAACGCTCATGACAGCAAAGTCT[TAMRA]) were purchased from Integrated DNA Technologies (Coralville, IA). To calibrate for minor input DNA quantity differences across samples we multiplexed the *RHD* assay with a second TaqMan assay from the single-copy control gene *RNaseP* with a VIC-labeled probe (Applied Biosystems, Foster City, CA). Quadruplicate reactions of 8 ng DNA in 10 μ L volumes with TaqMan Genotyping Master Mix (Applied Biosystems) were run on Applied Biosystem's 7900HT Real-Time PCR System in a 384-well plate. One individual with 2 *RHD* copies, NA10857, was run on each plate and used to calculate the estimated diploid copy numbers for all unknowns based on $\Delta\Delta C_T$ (Supplemental Table 1). To genotype the *RHCE* SNPs rs676785 (C/c) and rs609320 (E/e) we developed an assay based on co-amplification of these regions (for both *RHCE* and the corresponding regions of *RHD*) using dye-labelled primers followed by digestion of the internal products with the restriction enzyme *MnII* (New England Biolabs). Products were digested variably depending on rs676785 and rs609320, and the different-sized products were compared quantitatively to estimate Cc and Ee genotypes, taking *RHD* gene copy number into account. Details of the assay are provided in Supplemental Figure 1.

We validated the TaqMan-estimated *RHD* genotypes by checking consistency with Mendelian inheritance in the CEPH and Yoruba (these samples are comprised of parent-offspring trios), comparison to genotype estimates from analysis of data from the Affymetrix 6.0 SNP genotyping array (McCarroll et al. 2008), and high-resolution fluorescence *in situ* hybridization on stretched DNA fibers (fiber FISH). For fiber FISH analyses we used a fosmid probe G248P84657G6 (labeled with digoxigenin-11-dUTP; green in images) that maps to the 5' end of *RHD* but also hybridizes strongly to *RHCE* due to the high homology between these genes. We also created a PCR probe (labeled with biotin-16-dUTP; red in images) for the pair of segmental duplications that flank *RHD* only and contain the deletion breakpoint (Figure 1a). We first used long-range PCR to amplify nearly the entire segmental duplications (primers general to both copies; TAAATGCTCTTCTGAAGGCTGATACG and TTTACAAAGGGGAGAACGGTAAGAAG) from NA10857 in a 25 μ L reaction with 100 ng DNA and TripleMaster *Taq* Polymerase (Eppendorf) with an initial step of 93° C for 3 min followed by 40 cycles of 93° C for 30 sec, 64° C for 30 sec, and 68° C for 10 min. Nested PCR was then used to amplify two internal, overlapping fragments ~3.5 kb each in size (fragment 1: TCTTCTGAAGGCTGATACGACA and ATGATAGGGTTTGGTTGTGTCC; fragment 2: GGACACAACCAAACCCTATCAT and AATCACCGTCAAGGAGTCAGAT) using 0.1 μ L of the long-range PCR product in 25 μ L reactions with HotMaster *Taq* Polymerase (Eppendorf) for 3 min at 93° C followed by 40 cycles of 93° C for 30 sec, 60° C for 30 sec, and 70° C for 5 min. Probe labeling, preparation of the fibers and slides, hybridization, washes, detection, and imaging were performed as described previously (Perry et al. 2007) except that the fosmid probe was labeled for 12 rather than 5 hours.

Nucleotide sequence analyses

We used long-range PCR to amplify two unique (i.e., not duplicated) regions of the *RHD*/*RHCE* gene locus, each approximately 6 kb in size. One region is directly upstream of the segmental duplication at the 5' end of the *RHD* gene, while the second region is directly downstream of the segmental duplication at the 3' end of *RHD*, and between the *RHD* and *RHCE* genes (Figure 1a). The ~6kb amplified products were then used in nested PCR

reactions to amplify products ~600 bp in size with overlap ~280 bp, which were then sequenced in two directions as described previously (Xue et al. 2008). PCR primers are provided in Supplemental Table 2. Putative SNPs were flagged using Mutation Surveyor (SoftGenetics, State College, PA, USA) and then manually checked; each SNP call was supported by at least four reads. Summary statistic tests of neutrality were carried out and their *P* values estimated based on the percentile of the test value in the null distribution from 1000 simulations conditioned on the number of segregating sites using the program ms (Hudson 2002) with the best-fit demographic model (Schaffner et al. 2005) via a custom Perl script as described previously (Xue et al. 2006).

Simulations

We simulated selection at the *RHD* locus in a population of infinite size, considering the following: *i*) probability of sensitization if the mother is D-negative and the child is D-positive (*ps*), *ii*) probability that an individual child of a sensitized mother would die from hemolytic disease of the newborn (*pa*), *iii*) mean family size in families not potentially affected by hemolytic disease of the newborn (*fs*), and *iv*) maximum number of attempts to conceive (*na*). In turn, we fixed three of the parameters at reasonable values (including the assumption of some negative effect on individual fitness due to hemolytic disease of the newborn: *ps* = 0.2 and *pa* = 0.9) and varied the fourth, calculating the change in *RHD* deletion frequency over one generation given different initial *RHD* deletion population frequencies

Family size in families not potentially affected by hemolytic disease of the newborn was modeled as a binomial, with $n = na$ and $p = fs/na$. For families potentially affected by hemolytic disease of the newborn, we took a random draw from this binomial as the initial number of offspring and then applied selection due to hemolytic disease of the newborn. Heterozygous offspring of D-negative mothers paired with D-positive fathers were split into “pre-” and “post-sensitization” sets, with sensitization occurring at the *x*-th pregnancy, where *x* is a random draw from a geometric distribution with $p = ps$. Selection occurred on the heterozygous post-sensitization offspring (if any), with the number of surviving heterozygous post-sensitization offspring determined by a random draw from a binomial with n = the number of possible heterozygous offspring remaining, and $p = 1 - pa$. In families in which the father was heterozygous for the *RHD* deletion, each offspring was assigned to be homozygous or heterozygous with equal probability. We performed 10,000 simulations as described above and calculated the mean number of heterozygous offspring from families with D-negative mothers paired with m_1) D-positive fathers homozygous for the non-deletion allele and m_2) D-positive fathers heterozygous for the deletion. We defined selection against heterozygous offspring of m_1 as $s_1 = 1 - (\text{mean number of offspring of } m_1 \text{ couples} / fs)$, and selection against heterozygous offspring of m_2 as $s_2 = 1 - (\text{mean number of heterozygous offspring of } m_2 \text{ couples} / (fs/2))$. We then assumed an initial frequency of the *RHD* deletion and calculated the change in frequency over one generation, assuming a population of infinite size, as follows. We started with a population at Hardy-Weinberg equilibrium and calculated the proportion of families of each type potentially affected by hemolytic disease of the newborn (m_1 and m_2). We then reduced the frequency of *RHD* deletion heterozygotes in the subsequent generation by the expected proportion of heterozygous offspring from each potentially affected family type, multiplied by 1 - selection against heterozygotes of that family type (s_1 or s_2). We calculated the new frequency of the *RHD* deletion in the population after selection had occurred and subtracted the initial frequency to obtain the change in frequency over one generation.

RESULTS

RHD/ RHCE genotype frequencies

We genotyped the *RHD* gene deletion and the *RHCE* gene SNPs marking the C/c and E/e antigen variants in individuals from four HapMap human populations: European-Americans from Utah (CEU; $n = 90$), Yoruba from Ibadan, Nigeria (YRI; $n = 90$), Chinese Han from Beijing (CHB; $n = 45$), and Japanese from Tokyo (JPT; $n = 45$). In a subset of these individuals ($n = 68$; CEU, $n=23$; YRI, $n=23$; and CHB, $n=23$), we also resequenced two genomic regions, each ~6 kb in size: one region flanked the 5' end of the *RHD* gene, and the second region was located between the 3' ends of *RHD* and *RHCE* (Figure 1a). We used the HapMap population samples to integrate our *RHD* and *RHCE* genotype data with the HapMap Phase II SNP data (International HapMap Project Consortium 2007) for genome-wide comparisons of population differentiation and analyses of extended SNP haplotypes.

We estimated *RHD* deletion frequencies of 0.43, 0.19, and 0.07 in the CEU, YRI, and CHB +JPT population samples, respectively (Table 1; Figure 1b,c), consistent with previous studies (Fisher and Race 1946; Urbaniak and Greiss 2000; Wagner et al. 2003). Unexpectedly, we estimated three *RHD* copies per diploid cell for one YRI (NA19204) and one JPT (NA18952) individual. Fluorescent *in situ* hybridization analysis on stretched DNA fibers (fiber FISH) confirmed the presence of an *RHD* duplication allele (Figure 1d), the formation of which was likely mediated by non-allelic homologous recombination of the *RHD*-flanking segmental duplications (i.e., the reciprocal product of the *RHD* deletion). The genotypes of an additional YRI parent-offspring trio initially seemed inconsistent with a Mendelian pattern of inheritance, but this result was explained by the presence of both the *RHD* duplication and *RHD* deletion alleles in the father (Figure 1e). A similar duplication has been reported previously in a different Japanese population sample (Suto et al. 2000). Such a duplication may also explain qPCR results of greater *RHD* than *RHCE* copy number in eight individuals from Germany with relatively large densities of the D antigen in red blood cells despite Rhesus box (the *RHD*-flanking segmental duplications; see Figure 1a)-based genotyping assay suggestions of heterozygosity for the *RHD* deletion in those individuals (Yu et al. 2006). If so, then this result would indirectly suggest that the *RHD* duplication (which may cause Rhesus box-based *RHD* genotyping methods to fail) has a functional effect on antigen density.

For *RHCE*, the estimated frequency of the derived C allele was considerably higher in the CHB+JPT (0.78) and CEU (0.45) population samples than in YRI (0.08). Frequencies of the derived E allele were less variable among populations: 0.25, 0.09, and 0.08 in CHB+JPT, CEU, and YRI, respectively (Table 1).

Evolutionary analyses

We first examined the relative degree of population differentiation (F_{ST}) for the *RHD* and *RHCE* functional variants compared to genome-wide SNPs that have been genotyped in the same samples (International HapMap Project Consortium 2007). Extreme F_{ST} values may reflect past population-specific positive selection (Barreiro et al. 2008; Sabeti et al. 2006; Xue et al. 2009) (but see ref. (Coop et al. 2009)). While the *RHD* deletion and the E/e variant of *RHCE* are not unusual in this respect, the (CHB+JPT)-YRI F_{ST} value for the C/c variant of *RHCE* is exceptional ($F_{ST} = 0.64$; genome-wide percentile = 98.8; Figure 2a).

It is possible that F_{ST} values of variants within a duplicated locus, which may be subject to recurrent gene conversion (Chen et al. 2007), might not be comparable to those of unique variants. Indeed, the *RHCE* C allele itself is likely the result of a gene conversion event that transferred sequence from *RHD* to *RHCE* (Carritt et al. 1997). To examine this issue, we resequenced ~12 kb of this genomic region from 68 HapMap individuals. When we

integrated the identified SNP genotype data from this resequencing effort with the *RHD* deletion and *RHCE* C/c and E/e genotype data and examined the estimated haplotypes, it was clear that the high frequency of the C allele in non-African populations is explained almost exclusively by a single mutation event (Figure 2b; see discussion below), alleviating this concern.

We next performed tests of neutrality on the patterns of nucleotide diversity and the haplotypes that were estimated from the *RHD/RHCE* genotypes combined with the ~12 kb resequencing data. These tests evaluate whether or not the observed allele and haplotype frequency distributions are consistent with neutrality, given the demographic history of the populations (Schaffner et al. 2005). We observed significant deviations from neutral expectations in non-African populations (Table 2), including an excess of variants in external branches of the phylogeny in the CEU (Fu and Li's *D* test (Fu and Li 1993); $P=0.038$) and a single haplotype that has risen to unexpectedly high frequency in both CEU and CHB (Figure 2b; 47 out of 90 chromosomes; $P=0.045$ by the common haplotype frequency test (Xue et al. 2006)). This result is not explained by haplotypes containing the *RHD* deletion. Rather, the high frequency haplotype is comprised of *RHD*-positive chromosomes that carry the C allele of *RHCE* (Figure 2b). In total, we sampled 90 CHB and CEU chromosomes, 56 of which contain the *RHCE* C allele. Of these 56 chromosomes, 47 are identical across the ~12 kb region, and an additional four chromosomes differ from this common haplotype by only one nucleotide each.

Finally, we integrated the *RHD* deletion and *RHCE* C/c and E/e genotypes with HapMap Phase II SNP haplotype data to examine the patterns of linkage disequilibrium extending from these functional genetic variants. Positive selection may drive beneficial genetic variants to high frequencies at faster rates than neutral variants (which increase in frequency only by genetic drift). If the selection was recent and strong, then such variants might be associated with relatively long, low-diversity haplotypes, because recombination has not had sufficient opportunity to break down the extended haplotype on which the mutation occurred (Sabeti et al. 2002). To quantify the unusualness of extended haplotypes, we computed iHS scores (Voight et al. 2006) for the *RHD* and *RHCE* functional variants. None of the variants in any population had $|iHS| > 2.5$, a cutoff corresponding to the highest 1% of genome-wide iHS scores for SNPs in that population. The iHS score for the *RHD* deletion variant in CEU was 2.12, a moderate outlier, with the positive iHS score indicating that a relatively long, low-diversity haplotype is associated with the ancestral (non-deletion) allele. No other *RHD* or *RHCE* $|iHS|$ score exceeded 1.05. The $|iHS|$ scores associated with the C allele in CEU and CHB+JPT were 0.069 and 0.670, respectively.

DISCUSSION

Potential evidence for positive selection on the C allele of RHCE

Several observations from our evolutionary analyses – an unusually high (CHB+JPT)-YRI F_{ST} value for the *RHCE* C/c variant, the unusual Fu and Li's *D* value in the CEU and the high frequency of a single haplotype carrying the derived C allele in non-Africans – suggest a departure from neutral evolution at the *RHD/RHCE* locus. These findings all point towards the same conclusion: that the high frequency of the C allele in European and especially East Asian populations may reflect a history of positive natural selection. Enthusiasm for this hypothesis is somewhat tempered by the lack of a positive selection signal from the analysis of associated extended haplotypes. This discrepancy, however, might be explained by the moderate frequency of the putatively-selected haplotype or the different temporal resolutions of the population genetic tests we used. The F_{ST} statistic and the allele and haplotype frequency analyses from resequencing data are generally more sensitive to older selection events that have reached intermediate to high frequency, while

the iHS test is most powerful for detecting evidence of recent, strong, positive selection when there is good SNP density across the region of haplotype breakdown (Voight et al. 2006).

The derived C allele resulted from an *RHD* to *RHCE* gene conversion event (Carritt et al. 1997) that effectively reduced the difference between the two encoded proteins, raising the possibility that the C allele reached high frequency in non-African populations by conferring protection against *RHD*-related hemolytic disease of the newborn to D-negative mothers (i.e., if an antigenic response to exposure to blood from an D-positive fetus is less likely or less severe if the mother's RhCE antigen is more similar to the child's RhD). However, this idea can be dismissed based on two lines of evidence. One, the *RHCE* C allele is found almost exclusively on chromosomes with an intact *RHD* gene, consistent with previous observations (Carritt et al. 1997). Among the 90 CHB+CEU chromosomes for which we collected the ~12 kb resequencing data, 56 were found to carry the *RHCE* C allele. Of these 56 chromosomes, 55 (98%) were associated with intact *RHD* genes while only one (2%) was associated with the *RHD* deletion (Figure 2b). Two, the estimated *RHD* deletion frequency is < 0.10 in CHB+JPT (Table 1). The majority of D-negative blood type cases in an East Asian population resulted from *RHD* deletion (Luettringhaus et al. 2006). Together, these results suggest a very low occurrence of D-negative blood type and thus a naturally low rate of *RHD*-related hemolytic disease of the newborn in East Asian populations, consistent with previous observations (Urbaniak and Greiss 2000).

We hope that our report of possible positive selection on the *RHCE* C allele will motivate more detailed studies of the functional effects of this variant, both for those associated with the antigenicity phenotype and for any not associated with antigenicity. While Ser103Pro (which was the focus of our genotyping via the underlying SNP rs676785) explains C/c antigenicity, there are three other amino acid variants of RHCE in linkage disequilibrium with Ser103Pro (Cys16Trp, Ile60Leu, and Ser68Asn) (Avent and Reid 2000) that should also be considered candidates for fitness-relevant functional effects. In the absence of a current functional and testable hypothesis, we consider the suggestion of positive selection on the *RHCE* C allele to be preliminary but intriguing, especially given the previously-demonstrated importance of functional variation of other red blood cell membrane proteins in human evolution, for example of the Duffy antigen (*FY* gene) and susceptibility to *Plasmodium vivax* malaria (Kwiatkowski 2005).

No evidence for positive selection on the RHD deletion allele

Given that D-negative mothers – prior to now-common medical interventions – would have faced the potential fitness-reducing consequences of hemolytic disease of the newborn, we were somewhat surprised that in Europeans, with *RHD* deletion allele frequency = 0.43, our evolutionary analyses failed to uncover any strong evidence of positive selection on this allele that may have resulted from some unknown, offsetting fitness benefit. Our tests may have had insufficient power to reject neutrality for this variant if its high frequency initially resulted from a relatively ancient selective event. A second selective sweep in a genomic region (e.g., on the *RHCE* C allele) might also make it more difficult to detect an earlier selective event (e.g., on the *RHD* deletion). Extending this notion further, we note that there is considerable functional diversity at the *RHD*/*RHCE* gene locus even beyond the common, well-known variants studied here. For example, a pseudogene-causing (and D-negative phenotype-producing) 37 bp frameshift duplication in *RHD* exon 4 has a reported frequency of 7% in at least one African population (Singleton et al. 2000), and there are numerous, rare antigen-affecting SNPs, gene conversion events, and partial deletions of *RHD* and *RHCE* (Avent and Reid 2000; Westhoff 2004). Such a pattern of diversity could reflect low functional constraint at this locus, but if these mutations also resulted in or contributed to fitness-beneficial phenotypes, then their collective frequency increases may

have further clouded any genomic signatures of positive selection associated with the *RHD* deletion itself (Pennings and Hermisson 2006). Therefore, based only on the results of our analyses, it is not possible to determine that selection for the *RHD* deletion allele did not occur.

However, as an alternative to the above scenarios, the negative fitness effect of the D-negative blood type in the past might have been less strong than previously thought; the combined probabilities of (i) sensitization to a D-positive fetus and (ii) penetrance/morbidity of hemolytic disease of the newborn may have been too low to prevent the *RHD* deletion from reaching high frequency in Europe by genetic drift/ founder effect alone, given the effective size of this population. There are few appropriate, pre-medical intervention data with which to address this issue: Knox and Walker (Knox and Walker 1957) used antenatal antibody test data collected from 1947 to 1956 in the United Kingdom to estimate that the risk of sensitization for D-negative women was 1 in 17 pregnancies with a D-positive fetus. Between 1940 and 1946 at the Chicago Lying-In Hospital in the United States, incidence of hemolytic disease among the newborns of D-negative women was 1 in 37 births (1:252 when considering all women regardless of D status), with approximately two-thirds of these cases resulting in newborn death or stillbirth (Potter 1947).

While the above data suggest a higher mortality rate for the children of D-negative mothers, it is less clear whether this affected differences in the total number of surviving children between D-positive and D-negative mothers. Reed (1971) analyzed and reviewed multiple birth statistic datasets from the 1940s to 1960s in the United Kingdom, Canada, and United States and observed a significant association between D status and fertility in only one population sample, in which there were fewer live births to couples with a D-negative mother and a D-positive father than to couples with a D-positive mother or a D-negative father or both.

To address this issue further, we genotyped the *RHD* deletion in 942 Hutterites from South Dakota, U.S.A. The Hutterites are a founder population of European descent with high reproductive rates and large family sizes (Hostetler 1974; Ober et al. 1999). These data were combined with Rh blood group types for 504 Hutterites, which were previously determined by serology and birth and pedigree information for 1623 Hutterites (including all individuals with either Rh genotypes or serotypes) (Abney et al. 2000). We considered all families with at least one child born prior to 1968 (the first year of RhoGAM availability) and compared potentially affected families (defined as families with a D-negative mother and D-positive father) to unaffected families (defined as families with a D-positive mother and a D-positive father, a D-positive mother and a D-negative father, or a D-negative mother and a D-negative father, an unknown D status mother and a D-negative father, or a D-positive mother and an unknown D status father). There was no significant difference between the two groups in either total number of children (t-test; $P = 0.73$; Figure 3a) or mean interbirth interval (t-test; $P = 0.75$; Figure 3b).

The lack of a significant difference in family size or mean interbirth interval between these two groups of families may reflect the small sample size of potentially affected families (there were only 24 families with a D-negative mother and a D-positive father). We also caution that some (non-first) children included in this analysis were born after 1968 (Figure 3c), and therefore RhoGAM treatment may have helped to prevent sensitization in some women, if sensitization had not occurred in earlier pregnancies. In addition, recent general medical advances may have lessened the fitness disadvantage for D-negative mothers, even without RhoGAM treatment. Still, our results are consistent with most previous analyses that also failed to observe significant differences in fertility rates between D-negative and D-positive mothers (Reed 1971).

Apart from any uncertainty about the underlying fitness effect of the D-negative blood type, an additional evolutionary aspect that should be considered is the effect of frequency-dependent selection on the *RHD* deletion. Specifically, once the deletion reaches intermediate frequency, D-negative children of D-negative mothers and heterozygous D-positive fathers would have an advantage over D-positive children (Haldane 1942). Previous models of under-dominance at the *RHD* locus (Feldman et al. 1969; Levin 1967; Li 1953; Yokoyama 1981) were developed primarily to derive equilibrium frequencies of the allele, and few besides Haldane (1942) have provided illustrations of the effects of variation in relevant parameters at all frequencies of the deletion. Therefore, to evaluate how such variables might influence the strength of frequency-dependent selection, we simulated selection at the *RHD* locus in a population of infinite size, considering the following (see *Methods*): *i*) probability of sensitization, *ii*) probability of death from hemolytic disease of the newborn, *iii*) mean family size, and *iv*) maximum number of attempts to conceive. We fixed three of the parameters at reasonable values and varied the fourth, calculating the change in *RHD* deletion frequency over one generation given different initial *RHD* deletion population frequencies. We found that the strength of selection was affected by the probability of sensitization (Figure 4a), the probability of child death due to hemolytic disease of the newborn (Figure 4b), and family size (Figure 4c); due to modeling assumptions and the region of parameter space explored, the strength of selection was unaffected by the maximum number of attempts to conceive (Figure 4d).

These simulations confirm what Haldane (1942) and others (Feldman et al. 1969; Levin 1967; Li 1953) demonstrated through mathematical evaluation: that selection for the *RHD* deletion, under a model in which there is a potential fitness effect on D-negative mothers and in populations of infinite size, is frequency-dependent. If deletion frequencies are initially < 0.5 , then the population frequency becomes lower still in subsequent generations. If deletion frequencies are initially > 0.5 , then the population frequency becomes higher in subsequent generations. However, at intermediate frequencies – near *RHD* deletion frequency of ~ 0.5 , or similar to that observed in populations of European ancestry – the effect of selection is relatively small (at frequency = 0.5, there is no change; Figure 4). In populations of finite size, weak selective pressures may have been overwhelmed by the stochastic effects of genetic drift. In addition, the simulations demonstrate that family size has a dramatic effect on the strength of selection (also see Yokoyama 1981); if family sizes were small during the initial rise in frequency of the *RHD* deletion allele, then this also may have reduced the magnitude of selection.

Overall, evolutionary interpretations of the *RHD* gene deletion and the D-negative phenotype are mixed. Pre-treatment reported hemolytic disease mortality for births to D-negative mothers was 1 in 56 (Potter 1947). Once sensitization occurred, all subsequent pregnancies with a D-positive fetus would have been at risk, yet the observed fertility rates in the Hutterites generally did not differ between D-positive and D-negative mothers. None of our population genetic analyses revealed convincing evidence of past positive selection on the *RHD* deletion. While this negative result could be explained by a number of evolutionary scenarios, our current working explanation for the initial rise to relatively high frequency of the *RHD* deletion and the D-negative phenotype in European populations must be the null hypothesis: genetic drift/ founder effect. The level of purifying selection acting against this phenotype may have been less strong than thought, or simply not strong enough to overcome the effects of genetic drift in founding European populations. However, once an intermediate frequency had been achieved, selective pressures would have been relatively weak and potentially overcome by small amounts of genetic drift, even with considerable negative fitness consequences for individual D-negative mothers.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Rachael Cartledge for assistance with initial development of the *RHCE* genotyping assays, David Hopkinson for DNA samples with known Rh serotype from the former MRC Blood Group Unit, Luis Barreiro for the HapMap Phase II F_{ST} value database, Joe Pickrell for assistance with the iHS test, Richard Hudson for assistance with simulations, the Sanger Faculty Small Sequencing Projects Group for generating the sequence data, Steve McCarroll, Pardis Sabeti, and Molly Przeworski for helpful discussions, and two reviewers for insightful comments and suggestions on the manuscript. We acknowledge the participants who contributed samples for this study. This work was funded by National Institutes of Health Grant P41-HG004221 (to C.L.), The Wellcome Trust (WT098051, C.T.-S. and Y.X.), Medical Research Council New Investigator Award GO801123 (to E.J.H.), and National Institutes of Health Grant R01-HD21244 (to C.O.).

LITERATURE CITED

- Abney M, McPeck MS, Ober C. Estimation of variance components of quantitative traits in inbred populations. *American journal of human genetics*. 2000; 66:629–50. [PubMed: 10677322]
- Allen SJ, O'Donnell A, Alexander ND, Alpers MP, Peto TE, Clegg JB, Weatherall DJ. alpha+-Thalassemia protects children against disease caused by other infections as well as malaria. *Proc Natl Acad Sci U S A*. 1997; 94:14736–41. [PubMed: 9405682]
- Allison AC. The distribution of the sickle-cell trait in East Africa and elsewhere, and its apparent relationship to the incidence of subtertian malaria. *Trans R Soc Trop Med Hyg*. 1954; 48:312–8. [PubMed: 13187561]
- Avent ND, Reid ME. The Rh blood group system: a review. *Blood*. 2000; 95:375–87. [PubMed: 10627438]
- Barreiro LB, Laval G, Quach H, Patin E, Quintana-Murci L. Natural selection has driven population differentiation in modern humans. *Nat Genet*. 2008; 40:340–5. [PubMed: 18246066]
- Carritt B, Kemp TJ, Poulter M. Evolution of the human RH (rhesus) blood group genes: a 50 year old prediction (partially) fulfilled. *Hum Mol Genet*. 1997; 6:843–50. [PubMed: 9175729]
- Chen JM, Cooper DN, Chuzhanova N, Ferec C, Patrinos GP. Gene conversion: mechanisms, evolution and human disease. *Nat Rev Genet*. 2007; 8:762–75. [PubMed: 17846636]
- Colin Y, Cherif-Zahar B, Le Van Kim C, Raynal V, Van Huffel V, Cartron JP. Genetic basis of the RhD-positive and RhD-negative blood group polymorphism as determined by Southern analysis. *Blood*. 1991; 78:2747–52. [PubMed: 1824267]
- Coop G, Pickrell JK, Novembre J, Kudaravalli S, Li J, Absher D, Myers RM, Cavalli-Sforza LL, Feldman MW, Pritchard JK. The role of geography in human adaptation. *PLoS Genet*. 2009; 5:e1000500. [PubMed: 19503611]
- Endeward V, Cartron JP, Ripoché P, Gros G. RhAG protein of the Rhesus complex is a CO₂ channel in the human red cell membrane. *FASEB J*. 2008; 22:64–73. [PubMed: 17712059]
- Feldman MW, Nabholz M, Bodmer WF. Evolution of the Rh polymorphism: a model for the interaction of incompatibility, reproductive compensation, and heterozygote advantage. *Am J Hum Genet*. 1969; 21:171–93. [PubMed: 4976959]
- Fisher RA, Race RR. Rh gene frequencies in Britain. *Nature*. 1946; 157:48–9. [PubMed: 21013978]
- Flegel WA. Molecular genetics and clinical applications for RH. *Transfus Apher Sci*. 2011; 44:81–91. [PubMed: 21277262]
- Flint J, Hill AV, Bowden DK, Oppenheimer SJ, Sill PR, Serjeantson SW, Bana-Koiri J, Bhatia K, Alpers MP, Boyce AJ, et al. High frequencies of alpha-thalassaemia are the result of natural selection by malaria. *Nature*. 1986; 321:744–50. [PubMed: 3713863]
- Fu YX, Li WH. Statistical tests of neutrality of mutations. *Genetics*. 1993; 133:693–709. [PubMed: 8454210]
- Haldane JBS. Selection against heterozygosis in man. *Annals Eugenics*. 1942; 11:333–40.
- Hostetler, J. Hutterite Society Johns Hopkins University Press; Baltimore, MD: 1974.

- Hudson RR. Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics*. 2002; 18:337–8. [PubMed: 11847089]
- International HapMap Project Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature*. 2007; 449:851–61. [PubMed: 17943122]
- Knox G, Walker W. Nature of the determinants of rhesus isoimmunization. *Br J Prev Soc Med*. 1957; 11:126–30. [PubMed: 13471904]
- Kustu S, Inwood W. Biological gas channels for NH₃ and CO₂: evidence that Rh (Rhesus) proteins are CO₂ channels. *Transfus Clin Biol*. 2006; 13:103–10. [PubMed: 16563833]
- Kwiatkowski DP. How malaria has affected the human genome and what human genetics can teach us about malaria. *Am J Hum Genet*. 2005; 77:171–92. [PubMed: 16001361]
- Levin BR. The effect of reproductive compensation on the long term maintenance of the Rh polymorphism: the Rh crossroad revisited. *Am J Hum Genet*. 1967; 19:288–302. [PubMed: 4961094]
- Levine P, Vogel P, Katzin EM, Burnham L. Pathogenesis of Erythroblastosis Fetalis: Statistical Evidence. *Science*. 1941; 94:371–372. [PubMed: 17820878]
- Li CC. Is Rh facing a crossroad? A critique of the compensation effect. *Am Nat*. 1953; 87:257–61.
- Lo YM, Hjelm NM, Fidler C, Sargent IL, Murphy MF, Chamberlain PF, Poon PM, Redman CW, Wainscoat JS. Prenatal diagnosis of fetal RhD status by molecular analysis of maternal plasma. *N Engl J Med*. 1998; 339:1734–8. [PubMed: 9845707]
- Luettringhaus TA, Cho D, Ryang DW, Flegel WA. An easy RHD genotyping strategy for D-East Asian persons applied to Korean blood donors. *Transfusion*. 2006; 46:2128–37. [PubMed: 17176325]
- Marini AM, Matassi G, Raynal V, Andre B, Cartron JP, Cherif-Zahar B. The human Rhesus-associated RhAG protein and a kidney homologue promote ammonium transport in yeast. *Nat Genet*. 2000; 26:341–4. [PubMed: 11062476]
- McCarroll SA, Kuruvilla FG, Korn JM, Cawley S, Nemesh J, Wysoker A, Shapero MH, de Bakker PI, Maller JB, Kirby A, Elliott AL, Parkin M, Hubbell E, Webster T, Mei R, Veitch J, Collins PJ, Handsaker R, Lincoln S, Nizzari M, Blume J, Jones KW, Rava R, Daly MJ, Gabriel SB, Altshuler D. Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nat Genet*. 2008; 40:1166–74. [PubMed: 18776908]
- Moncharmont P, Dupraz F, Juron, Vignal M, Rigal D, Meyer F, Debeaux P. Haemolytic disease of the newborn infant. Long term efficiency of the screening and the prevention of alloimmunization in the mother: thirty years of experience. *Arch Gynecol Obstet*. 1991; 248:175–80. [PubMed: 1910321]
- Mouro I, Colin Y, Cherif-Zahar B, Cartron JP, Le Van Kim C. Molecular genetic basis of the human Rhesus blood group system. *Nat Genet*. 1993; 5:62–5. [PubMed: 8220426]
- Ober C, Hyslop T, Hauck WW. Inbreeding effects on fertility in humans: evidence for reproductive compensation. *Am J Hum Genet*. 1999; 64:225–31. [PubMed: 9915962]
- Pennings PS, Hermisson J. Soft sweeps III: the signature of positive selection from recurrent mutation. *PLoS Genet*. 2006; 2:e186. [PubMed: 17173482]
- Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, Redon R, Werner J, Villanea FA, Mountain JL, Misra R, Carter NP, Lee C, Stone AC. Diet and the evolution of human amylase gene copy number variation. *Nat Genet*. 2007; 39:1256–60. [PubMed: 17828263]
- Potter, EL. Rh... Its relation to congenital hemolytic disease and to intragroup transfusion reactions. Year Book Publishers; Chicago: 1947.
- Reed TE. Does reproductive compensation exist? An analysis of Rh data. *Am J Hum Genet*. 1971; 23:215–24. [PubMed: 4997205]
- 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. Detecting recent positive selection in the human genome from haplotype structure. *Nature*. 2002; 419:832–7. [PubMed: 12397357]
- Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilyl P, Shamovsky O, Palma A, Mikkelsen TS, Altshuler D, Lander ES. Positive natural selection in the human lineage. *Science*. 2006; 312:1614–20. [PubMed: 16778047]

- Schaffner SF, Foo C, Gabriel S, Reich D, Daly MJ, Altshuler D. Calibrating a coalescent simulation of human genome sequence variation. *Genome Res.* 2005; 15:1576–83. [PubMed: 16251467]
- Singleton BK, Green CA, Avent ND, Martin PG, Smart E, Daka A, Narter-Olaga EG, Hawthorne LM, Daniels G. The presence of an RHD pseudogene containing a 37 base pair duplication and a nonsense mutation in africans with the Rh D-negative blood group phenotype. *Blood.* 2000; 95:12–8. [PubMed: 10607679]
- Suto Y, Ishikawa Y, Hyodo H, Uchikawa M, Juji T. Gene organization and rearrangements at the human Rhesus blood group locus revealed by fiber-FISH analysis. *Hum Genet.* 2000; 106:164–71. [PubMed: 10746557]
- Touinssi M, Chiaroni J, Degioanni A, De Micco P, Dutour O, Bauduer F. Distribution of rhesus blood group system in the French basques: a reappraisal using the allele-specific primers PCR method. *Hum Hered.* 2004; 58:69–72. [PubMed: 15711086]
- Urbaniak SJ, Greiss MA. RhD haemolytic disease of the fetus and the newborn. *Blood Rev.* 2000; 14:44–61. [PubMed: 10805260]
- Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. *PLoS Biol.* 2006; 4:e72. [PubMed: 16494531]
- Wagner FF, Flegel WA. RHD gene deletion occurred in the Rhesus box. *Blood.* 2000; 95:3662–8. [PubMed: 10845894]
- Wagner FF, Moulds JM, Touunkara A, Kouriba B, Flegel WA. RHD allele distribution in Africans of Mali. *BMC Genet.* 2003; 4:14. [PubMed: 14505497]
- Westhoff CM. The Rh blood group system in review: a new face for the next decade. *Transfusion.* 2004; 44:1663–73. [PubMed: 15504174]
- Westhoff CM, Wylie DE. Transport characteristics of mammalian Rh and Rh glycoproteins expressed in heterologous systems. *Transfus Clin Biol.* 2006; 13:132–8. [PubMed: 16563829]
- Xue Y, Daly A, Yngvadottir B, Liu M, Coop G, Kim Y, Sabeti P, Chen Y, Stalker J, Huckle E, Burton J, Leonard S, Rogers J, Tyler-Smith C. Spread of an inactive form of caspase-12 in humans is due to recent positive selection. *Am J Hum Genet.* 2006; 78:659–70. [PubMed: 16532395]
- Xue Y, Sun D, Daly A, Yang F, Zhou X, Zhao M, Huang N, Zerjal T, Lee C, Carter NP, Hurles ME, Tyler-Smith C. Adaptive evolution of UGT2B17 copy-number variation. *Am J Hum Genet.* 2008; 83:337–46. [PubMed: 18760392]
- Xue Y, Zhang X, Huang N, Daly A, Gillson CJ, Macarthur DG, Yngvadottir B, Nica AC, Woodwark C, Chen Y, Conrad DF, Ayub Q, Mehdi SQ, Li P, Tyler-Smith C. Population differentiation as an indicator of recent positive selection in humans: an empirical evaluation. *Genetics.* 2009; 183:1065–77. [PubMed: 19737746]
- Yokoyama S. Family size and evolution of Rh polymorphism. *J Theor Biol.* 1981; 92:119–25. [PubMed: 6799698]
- Yu X, Wagner FF, Witter B, Flegel WA. Outliers in RhD membrane integration are explained by variant RH haplotypes. *Transfusion.* 2006; 46:1343–51. [PubMed: 16934070]

Figure 1. *RHD* deletion genotyping and fiber FISH validation

(a) The paralogous *RHD* and *RHCE* genes (98% nucleotide sequence similarity) are located on human chromosome 1p36.11 in inverted orientation. The genes are ~60 kb each in size, separated by ~30 kb. Smaller “Rhesus box” segmental duplications (~9 kb each) with 99% sequence similarity flank *RHD* and provided the substrate for gene deletion via non-allelic homologous recombination (Wagner and Flegel 2000). We used these features to isolate and distinguish *RHD* in TaqMan genotyping and fiber FISH validation assays of the deletion. We also resequenced ~12 kb from two unique regions of this locus, as indicated, in a subset of the individuals in our study. (b) TaqMan results for the HapMap population samples and allele frequencies. Note that one YRI and one JPT individual were estimated to have 3 *RHD* copies. (c-e) Fiber FISH validation for three HapMap trios, with representative images from the two alleles for each individual. The CEU trio depicted in (c) represents the scenario in which the largest proportion of offspring are at for hemolytic disease of the newborn, in which the mother is D-negative and the father has two copies of *RHD*, such that all of their offspring would be D-positive. The YRI trio depicted in (d) includes one of the individuals estimated to have 3 *RHD* copies (GM19204). We confirmed the duplication and infer that its origin was also mediated by non-allelic homologous recombination of the Rhesus box segmental duplications; this is the expected reciprocal of the *RHD* gene deletion. The duplication allele was not transmitted to the child (GM19205). The Yoruba trio depicted in (e) initially seemed a case of non-Mendelian inheritance based on total diploid copy numbers from our TaqMan copy number estimates; however, direct determinations of the copy numbers of each chromosome by fiber FISH show that the father (GM19153) has one copy each of the *RHD* duplication and deletion alleles, and the child (GM19154) inherited the deletion allele.

Figure 2. Evidence for positive selection on the C allele of *RHCE*

(a) Cumulative F_{ST} distributions of HapMap Phase II SNPs for each population comparison (CHB and JPT are considered together), with F_{ST} values for the functional *RHCE* variants indicated by the circle, triangle, and square symbols. The (CHB+JPT)-YRI F_{ST} value for the Cc variant is exceptional compared to the distribution of HapMap SNPs, consistent with positive selection. (b) Network analysis of inferred haplotypes from *RHCE* functional variants and SNPs within and flanking the locus (from ~12 kb resequencing data), for CEU ($n = 44$ chromosomes) and CHB ($n = 46$ chromosomes) together. Circles represent haplotypes with the area of each proportional to frequency; the total number of chromosomes is shown next to each haplotype circle where $n > 1$. The shortest lines between haplotype circles indicate one mutational step. *RHCE* haplotypes are indicated by circle colors (historical *RHCE* haplotype nomenclature is provided in parentheses). The frequency of the principal DCE haplotype ($n = 47$ chromosomes) is unexpected given the overall pattern of haplotype diversity at this locus, a signature consistent with positive selection.

Figure 3. Fertility rate and inter-birth interval analysis by D status in the Hutterites

(a) Frequency distributions of the number of children per family, for families with at least one child born prior to 1968 (the first year of RhoGAM availability for this population), separately for families potentially affected by hemolytic disease of the newborn (D-negative mother and D-positive father) and unaffected families (D-positive mother or D-negative father or both). (b) Cumulative distributions of inter-birth intervals (days) for potentially affected (black) and unaffected (grey) families. (c) Frequency distributions of birth years for children in potentially affected and unaffected families.

Figure 4. Influence of parameter variation on frequency-dependent selection at the *RHD* locus
 Simulated change in *RHD* deletion frequency over one generation (y-axis), given initial *RHD* deletion frequency (x-axis) and evaluated under varying (a) probability of sensitization if the mother is D-negative and the child is D-positive (ps), (b) probability that an individual child of a sensitized mother would die from hemolytic disease of the newborn (pa), (c) mean family sizes, in families not potentially affected by hemolytic disease of the newborn (fs), and (d) maximum number of attempts to conceive (na). For each round of simulations (a-d), the remaining three of the four variables were fixed at reasonable values ($ps = 0.2$, $pa = 0.9$, $fs = 3$, $na = 20$).

Table 1*RHD* and *RHCE* derived allele frequencies in HapMap populations

Population ¹	<i>RHD/CE</i> derived allele frequencies ²		
	<i>RHD</i> : +/– (–)	<i>RHCE</i> : C/c (C)	<i>RHCE</i> : E/e (E)
YRI (Yoruba from Ibadan, Nigeria)	0.19	0.08	0.08
CEU (CEPH European-Americans from Utah)	0.43	0.45	0.09
CHB (Chinese Han from Beijing)	0.09	0.70	0.31
JPT (Japanese from Tokyo)	0.05	0.84	0.20

¹Notes: Considering unrelated individuals only.²Derived alleles in parentheses.

Table 2

Summary statistics for *RHD/RHCE* region resequencing data

Population	Chr. ¹	S ²	π ³	Tajima's <i>D</i>		Fu and Li's <i>D</i>		Fay and Wu's <i>H</i>	
				<i>D</i>	<i>P</i> value ⁴	<i>D</i>	<i>P</i> value ⁴	<i>H</i>	<i>P</i> value ⁴
YRI	46	61	0.083%	-0.926	0.660	-1.000	0.678	2.334	0.754
CHB	46	18	0.035%	0.038	0.466	-0.389	0.346	-1.754	0.205
CEU	44	21	0.037%	-0.124	0.816	-1.968	0.038	1.197	0.930

¹Notes: Number of chromosomes sequenced for each population.²Number of segregating sites.³Mean pairwise nucleotide diversity.⁴*P* values based on simulations using a best-fit demographic model (Schaffner et al. 2005).