June 2004

Genome Scans of DNA Variability in Humans Reveal Evidence for Selective Sweeps Outside of Africa

Jay F. Storz
University of Nebraska - Lincoln, jstorz2@unl.edu

Bret A. Payseur
Department of Molecular Biology and Genetics, Cornell University

Michael W. Nachman
University of Arizona

Genome Scans of DNA Variability in Humans Reveal Evidence for Selective Sweeps Outside of Africa

Jay F. Storz,[1] Bret A. Payseur,[2] and Michael W. Nachman

Department of Ecology and Evolutionary Biology, Biosciences West, University of Arizona
[1] Present address: Department of Biology, San Francisco State University
[2] Present address: Department of Molecular Biology and Genetics, Cornell University

Abstract

The last 50,000–150,000 years of human history have been characterized by rapid demographic expansions and the colonization of novel environments outside of sub-Saharan Africa. Mass migrations outside the ancestral species range likely entailed many new selection pressures, suggesting that genetic adaptation to local environmental conditions may have been more prevalent in colonizing populations outside of sub-Saharan Africa. Here we report a test of this hypothesis using genome-wide patterns of DNA polymorphism. We conducted a multilocus scan of microsatellite variability to identify regions of the human genome that may have been subject to continent-specific hitchhiking events. Using published polymorphism data for a total of 624 autosomal loci in multiple populations of humans, we used coalescent simulations to identify candidate loci for geographically restricted selective sweeps. We identified a total of 13 loci that appeared as outliers in replicated population comparisons involving different reference samples for Africa. A disproportionate number of these loci exhibited reduced levels of relative variability in non-African populations alone, suggesting that recent episodes of positive selection have been more prevalent outside of sub-Saharan Africa.

Keywords: adaptation, positive selection, genetic hitchhiking, human genome, microsatellite DNA, natural selection

Introduction

Genetic and archaeological evidence suggest that the demographic history of anatomically modern humans involved a range expansion out of Africa approximately 50,000–150,000 years before present (BP) (Cavalli-Sforza, Menozzi, and Piazza 1994; Harpending and Rogers 2000; Relethford 2001; Excoffier 2002). The colonization of novel environments outside the ancestral species range likely entailed many new selection pressures caused by emergent infectious diseases, changes in diet, and exposure to new climatic conditions. Adaptive challenges posed by new biological, cultural, and physical environments were likely to have been particularly acute after the start of the Neolithic, 10,000 years BP, which marked the advent of agriculture and associated increases in population density. These considerations suggest that, within the last 10,000 years of human history, genetic adaptation to local environmental conditions may have been more prevalent in colonizing populations outside of sub-Saharan Africa.

Adaptive explanations have been offered for many physiological and morphological differences among human populations (Diamond 1992), although compelling empirical support is generally lacking. In recent years, some progress has been made by analyzing DNA sequence variation underlying phenotypes that distinguish human populations in different continental regions. For example, the incidence of infectious disease is known to vary among different geographic regions, and selection has probably played a significant role in shaping patterns of variation at disease-resistance genes in different human populations (Cavalli-Sforza, Menozzi, and Piazza 1994). Consistent with this idea, geographically localized selection for disease resistance appears to have driven the frequency of the CCR5-Δ32 allele to unusually high frequencies only in northern Europe (Stephens *et al.* 1998). Another physiological trait that is thought to have been subject to positive selection in populations outside of sub-Saharan Africa is lactase persistence (the persistence of intestinal lactase activity into adulthood; Bodmer and Cavalli-Sforza 1976). Consistent with this idea, positive selection appears to have driven a recently derived "lactase-persistence" haplotype to high frequency only in northern Europe (Hollox *et al.* 2001). One of the most obvious phenotypic differences among human populations from different continental regions is skin pigmentation. Some form of natural and/or sexual selection is probably responsible for shaping global patterns of variation in human skin color, as differentiation in the level of skin pigmentation among continental populations vastly exceeds the level of differentiation observed at neutral genetic markers (Relethford 2002).

Although locus-specific surveys of DNA sequence variation can potentially elucidate the role of past selection in driving the differentiation of specific traits among human populations, individual case studies cannot provide an answer to the more general question of whether recent positive selection has been more prevalent in colonizing populations outside of sub-Saharan Africa. In principle, this hypothesis could be tested by conducting genomic scans of DNA variability in samples from different continental populations of humans. If recent positive selection has been more prevalent outside of Africa, multilocus neutrality tests based on DNA polymorphism or haplotype structure should identify a disproportionate number of candidate genes for selective sweeps in non-African populations. The underlying premise of multilocus neutrality tests is that demographic processes have relatively uniform effects across the entire genome, whereas the effects of selection are generally locus-spe-

Table 1. Data Sets Used for the Analysis of Microsatellite Variability in Samples from Each of Three Continents (Africa, Europe, and Asia).

Data Set	D1	D2	D3	D4
Number of loci	95	92	60	377
Repeat motif	Dinucleotide	Dinucleotide	Tetranucleotide	Di-, tri-, and tetranucleotide
African population samples	Biaka ($n = 132.27$)	Biaka ($n = 34.35$)	Biaka ($n = 9.03$)	Biaka ($n = 70.05$)
	Mbuti ($n = 75.47$)	Mbuti ($n = 30.41$)	Mbuti ($n = 9.60$)	Mbuti ($n = 28.93$)
			San ($n = 27.38$)	San ($n = 13.58$)
European population samples	Danish ($n = 111.44$)	N. Italian ($n = 40.37$)	French ($n = 37.07$)	French ($n = 55.89$)
Asian population samples	Han Chinese ($n = 93.48$)	Han Chinese ($n = 29.87$)	Han Chinese ($n = 33.08$)	Han Chinese ($n = 87.20$)

Note: For each continent, n = mean number of chromosomes sampled per locus.

cific and can be inferred from patterns of variation at linked sites (Cavalli-Sforza 1966; Lewontin and Krakauer 1973). For example, the spread and fixation of adaptive mutations results in the joint fixation of linked neutral variants ("genetic hitchhiking;" Maynard Smith and Haigh 1974). The strength of this hitchhiking effect is a function of the selection coefficient and recombinational distance from the selected site (Maynard Smith and Haigh 1974; Kaplan, Hudson, and Langley 1989; Stephan, Wiehe, and Lenz 1992; Wiehe and Stephan 1993). In general, genetic hitchhiking results in a reduced level of variability and a skewed distribution of allele frequencies at linked neutral loci (Tajima 1989; Fu and Li 1993; Braverman *et al.* 1995; Simonsen, Churchill, and Aquadro 1995).

In the case of geographically subdivided populations, hitchhiking with a locally adaptive mutation is expected to greatly increase the level of differentiation at linked neutral loci (Stephan and Mitchell 1992; Begun and Aquadro 1993; Stephan 1994; Stephan *et al.* 1998). Thus, loci that have undergone geographically restricted selective sweeps should be characterized by levels of differentiation that greatly exceed the genome-wide average.

Here we test the hypothesis that local adaptation in humans has been more prevalent in colonizing populations outside of sub-Saharan Africa. Using publicly available data, we conducted multilocus scans of microsatellite variability to identify regions of the human genome that may have been subject to continent-specific hitchhiking events. A previous genome scan of microsatellite variability in humans identified a number of candidate regions for divergent selection between African and European populations (Kayser, Brauer, and Stoneking 2003). Intriguingly, the authors reported an excess of loci that exhibited reduced variability in Europe relative to sub-Saharan Africa. Here we use a model-based approach to investigate this pattern in more detail using independent samples from European, Asian, and African populations and a much larger sample of marker loci. To the extent that our sample of marker loci is representative of the genome as a whole, results of the analysis provide further evidence that recent selective sweeps have been more prevalent in Europe and Asia than in sub-Saharan Africa.

Methods

Population Samples

We analyzed four microsatellite data sets that included polymorphism data for a total of 624 autosomal loci in multiple human populations (Table 1). If we assume that hitchhiking effects typically extend 0.25–0.5 cM on either side of an adaptive mutation (e.g., Sabeti *et al.* 2002; Saunders, Hammer, and Nachman 2002), and that the sex-averaged length of the human genome is ~3,615 cM (Kong *et al.* 2002), then our survey of 624 unlinked loci has effectively screened 4.3%–8.6% of the genome.

The first data set, D1 ($n = 95$ dinucleotide repeat loci), is available online at http://info.med.yale.edu/genetics/kkidd/abiinfo.html. Data were available from a total of 102 loci. However, we excluded loci that were missing data from one or more of the African samples. The second data set, D2 ($n = 92$ dinucleotide repeat loci), has been used in studies of demographic population structure (Bowcock *et al.* 1994; Jin *et al.* 2000), and these data are available online at http://www-evo.stanford.edu. The third data set, D3 ($n = 60$ tetranucleotide repeat loci), has been used in studies of population structure (Jorde *et al.* 1995, 1997; Eller 1999) and these data were kindly provided by L. B. Jorde (University of Utah Health Sciences Center). The fourth data set, D4 ($n = 377$ di-, tri-, and tetranucleotide repeat loci), has also been used in studies of population structure (Rosenberg *et al.* 2002; Zhivotovsky, Rosenberg, and Feldman 2003) and is available online at http://research.marshfieldclinic.org/genetics/Freq/FreqInfo.htm. For each of the four data sets, we used a single reference population to represent each of three continents: Africa, Asia, and Europe (table 1). For each comparison, we used a single, geographically restricted sample from each continent because using pooled allele counts from multiple, geographically dispersed localities could produce an artifactual skew in the distribution of allele frequencies within continents (Wakeley *et al.* 2001; Przeworski 2002; Ptak and Przeworski 2002; Hammer *et al.* 2003). This sampling problem would be especially severe within Africa because of the high degree of population subdivision (Yu *et al.* 2002). Because of the high degree of subdivision within Africa, and as a means of evaluating the consistency of locus-specific patterns of variation, we repeated all population comparisons using the Biaka and Mbuti as separate reference samples for Africa. The Biaka and Mbuti were the only African samples represented in the D1 and D2 data sets and they were the only samples that were common to all four data sets. For the D3 and D4 data sets, we also included the San as a third reference sample for Africa. We chose to use the San sample (rather than any of the other samples available in the D3 and D4 data sets) because previous studies have indicated that the San represent one of the most basal groups of modern humans (e.g., Chen *et al.* 2000; Ingman *et al.* 2000). Thus, inclusion of the San sample maximizes the coverage of African diversity in our analysis.

Three-way Population Comparisons

For each three-way population comparison, we obtained single-locus estimates of F_{ST} (a standardized measure of genetic differentiation) using the β statistic of Cockerham and Weir (1993):

$$\beta = (F_0 - F_1) / (1 - F_1),   (1)$$

where $1 - F_1$ is the average probability of identity-by-descent for two alleles sampled randomly from separate subpopulations, and $1 - F_0$ is the average probability of identity-by-descent for two alleles sampled randomly from the same subpopulation. We estimated expected heterozygosity, H, as $1 - F_1$. For each data set, we tested for evidence of selection by comparing observed F_{ST} values (conditional on heterozygosity) to a null distribution generated by a coalescent-based simulation model.

Simulation results of Beaumont and Nichols (1996) indicate that conditional distributions of F_{ST} generated under an infinite island model are robust to a range of different nonequilibrium conditions. However, it is not clear whether this apparent robustness holds up in cases involving small numbers of demes. We therefore conducted coalescent simulations under a nonequilibrium model that incorporated population bottlenecks to account for founder effects associated with migrations out of Africa. Specifically, we extended the "multi-epoch" model of Marth *et al.* (2004) to consider the case of three populations that diverged at a specified time in the past. Following the initial divergence (going forward in time), one of the populations remained stationary at the ancestral effective population size ($N_3 = 10,000$), whereas the other two populations each underwent an instant reduction in effective size ($N_2 = 1,000$). The duration of the population bottleneck was $T_2 = 550$ generations. The bottleneck was then followed by a stepwise increase of effective size to $N_1 = 10,000$, which occurred $T_1 = 3100$ generations before the present. For T_1 and T_2, we used the average of maximum-likelihood parameter estimates obtained for Asian and European populations in the study of Marth *et al.* (2004). Similarly, we used a size expansion ratio ($N_1/N_2 = 10$) close to the average of maximum-likelihood parameter estimates obtained for these same samples under the "three-epoch" model of Marth *et al.* (2004). To evaluate the effect of gene flow, we included a migration matrix that defined a three-deme island model. We conducted coalescent simulations under this nonequilibrium model of population structure to generate null distributions of F_{ST} over a range of migration rates.

Coalescent simulations were conducted under the stepwise mutation model (SMM). The mutation rate varied randomly over the range 1×10^{-5} to 5×10^{-4} (to generate a roughly uniform distribution of H values), and allele sizes were constrained to a range of 5 to 20 repeat units. In each set of iterations, sample sizes were set equal to the median of actual sample sizes in the specific data set under consideration. Coalescent simulations were used to generate a total of 50,000 paired values of F_{ST} and H, which was then used to compute the 0.95 and 0.50 quantiles of the conditional distribution (Beaumont and Nichols 1996; Storz and Nachman 2003). We used F_{ST} as an estimator of differentiation rather than microsatellite-specific statistics such as R_{ST} because the former statistic is generally character-

ized by a lower sampling variance (Slatkin 1995; Gaggiotti *et al.* 1999; Balloux and Goudet 2002).

We used an iterative fitting procedure to generate the expected neutral distribution for each three-way population comparison. In each set of iterations, values for the migration matrix were varied over the range 1×10^{-5} to 5×10^{-4} to produce a null distribution in which equal numbers of loci fell above and below the median quantile. Loci with F_{ST} values that exceeded the 0.95 quantile of the resultant distribution were considered as preliminary candidates for continent-specific selective sweeps.

The set of candidate loci identified in each set of simulations may contain loci that are tracking selection at linked sites as well as loci that are false positives. All genome-wide scans for detecting selection face the problem of identifying and excluding false positives when multiple tests are conducted (e.g., Huttley *et al.* 1999; Akey *et al.* 2002; Payseur, Cutter, and Nachman 2002; Vigouroux *et al.* 2002; Luikart *et al.* 2003; Schlötterer 2003). To minimize this problem, we restricted further consideration to the subset of loci that appeared as outliers in separate comparisons using different reference populations for Africa (Biaka and Mbuti). For the D3 and D4 data sets, we also considered the smaller subset of loci that appeared as outliers in comparisons using all three different reference populations (Biaka, Mbuti, and San).

African Versus Non-African Comparisons

Although higher-than-expected F_{ST} values should be indicative of continent-specific selective sweeps, F_{ST} values alone cannot identify the populations in which the allele frequency changes took place. Thus, after identifying candidate loci that may have been subject to continent-specific sweeps, we then used pairwise comparisons to assess whether equal numbers of these loci exhibited reduced levels of relative variability in African and non-African populations. If recent selective sweeps have been more prevalent outside of Africa, then a disproportionate number of candidate loci should exhibit reduced levels of relative variability in non-African populations alone.

Under the SMM at mutation-drift equilibrium, expected heterozygosity (or "gene diversity," H; Nei 1978) can be used to estimate the neutral parameter $\theta = 4N_e\mu$, where N_e = effective population size and μ = mutation rate (Ohta and Kimura 1973; Moran 1975). For two populations, the expected ratio of heterozygosity-based estimators of θ is:

$$E[RH] = \frac{\left(\left(\frac{1}{1-H_{pop1}}\right)^2 - 1\right)\frac{1}{8\mu}}{\left(\left(\frac{1}{1-H_{pop2}}\right)^2 - 1\right)\frac{1}{8\mu}} \approx \frac{\left(\frac{1}{1-H_{pop1}}\right)^2 - 1}{\left(\frac{1}{1-H_{pop2}}\right)^2 - 1}.   (2)$$

For each locus we used estimates of African gene diversity in the numerator of the ratio and the arithmetic average of European and Asian gene diversities in the denominator, following Schlötterer (2002). We considered Europe and Asia jointly because we are specifically interested in testing the hypothesis that selective sweeps have been more prevalent outside of sub-

Table 2. Genetic Differentiation Among African, European, and Asian Population Samples.

Data Set	Populations	Weighted Mean FST	Median No. Chromosomes/Sample
D1 (n = 95 loci)	Biaka-Danish-Han Chinese	0.112	111
	Mbuti-Danish-Han Chinese	0.124	92
D2 (n = 92 loci)	Biaka-N. Italian-Han Chinese	0.122	23
	Mbuti-N. Italian-Han Chinese	0.134	23
D3 (n = 60 loci)	Biaka-French-Han Chinese	0.151	26
	Mbuti-French-Han Chinese	0.164	26
	San-French-Han Chinese	0.122	33
D4 (n = 377 loci)	Biaka-French-Han Chinese	0.137	70
	Mbuti-French-Han Chinese	0.157	57
	San-French-Han Chinese	0.166	52

Saharan Africa. If the majority of higher-than-expected F_{ST} values are attributable to selective sweeps outside of Africa, then a disproportionate number of candidate loci should have RH values that fall above the median of the genome-wide distribution in each pairwise comparison between African and non-African populations. Since the estimate of African gene diversity is in the numerator of the ratio, RH values that fall above the median of the distribution indicate a reduced level of relative variability in non-African populations. Loci that show only a modest difference in gene diversity between populations (and which therefore fall just slightly above or below the median value) are more likely to be false-positives than loci that show a more pronounced discrepancy in relative levels of gene diversity (and which therefore show a higher deviation from the genome-wide average). To maximize the probability of excluding false positives from the African versus non-African comparisons, we only considered loci with ln-transformed RH values that fell above or below 0.5 standard deviations (SDs) of the genome-wide average. In cases where loci were monomorphic in one population, we substituted an H value of 0.0001 for zero.

By considering ratios of θ values, we are able to control for interlocus differences in levels of variability that are attributable to differences in mutation rate and/or recombinational environment. For the ratios of θ values, we used estimators based on gene diversity (lnRH) rather than variance in allele size (lnRV; Schlötterer 2002), because lnRH is characterized by a much lower variance and is less sensitive to departures from the SMM (Kauer, Dieringer, and Schlötterer 2003).
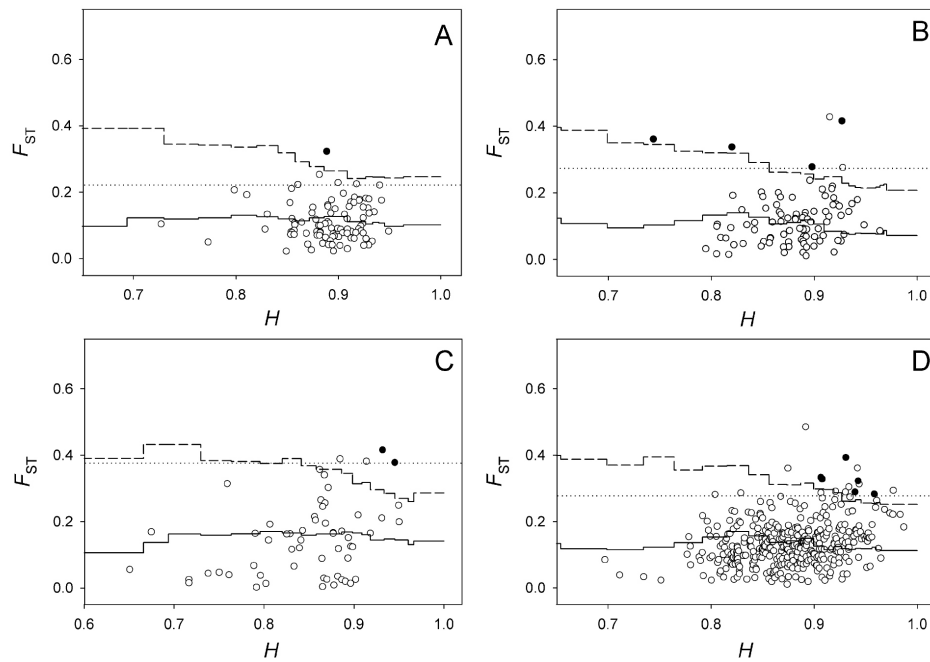
Results

Three-way Population Comparisons

Mean F_{ST} estimates for the three-way population comparisons ranged from 0.1117 to 0.1661 (Table 2). In the four comparisons using Biaka as the reference sample for Africa, a total of 26 loci were identified as candidates for selection (one locus from the D1 data set, six from D2, four from D3, and 15 from D4; Figure 1). In the comparisons using Mbuti as the reference sample for Africa, a total of 25 loci were identified as candi-

dates for selection (four loci from the D1 data set, four from D2, five from D3, and 12 from D4; Figure 2). In the separate comparisons using Biaka and Mbuti samples, 18 of the same loci were identified as candidates for selection. In the two comparisons using San as the reference sample for Africa, a total of 16 loci were identified as candidates for selection (three loci from the D3 data set and 13 from D4; Figure 3). In the three separate comparisons using Biaka, Mbuti, and San samples (D3 and D4 data sets only), seven of the same loci were identified as candidates for selection.

African Versus Non-African Comparisons

Consistent with the results of previous surveys of DNA variability in humans (reviewed by Harpending and Rogers 2000), average levels of gene diversity tended to be slightly higher in African than in non-African populations (table 3). The exceptions to this general pattern can probably be attributed to the fact that the Biaka, Mbuti, and San samples represent restricted subsets of African diversity.

Of the 18 loci identified as candidates for selection in the separate three-way comparisons using Biaka and Mbuti samples, 13 loci exhibited differences in relative levels of variability between African and non-African populations (i.e., lnRH values fell more than 0.5 SDs from the means of both distributions). If recent selective sweeps have occurred with equal frequency inside and outside of Africa, then roughly equal numbers of candidate loci should exhibit reduced levels of relative variability in African and non-African populations. We can reject this null hypothesis, as only one locus exhibited a reduced level of relative variability in Africa and 12 loci exhibited reduced levels of relative variability outside of Africa (χ^2_1 = 9.31, P = 0.0055; table 4). Of the seven loci identified as candidates for selection in the separate comparisons using Biaka, Mbuti, and San samples (D3 and D4 data sets only), five loci exhibited discrepancies in relative levels of variability between African and non-African populations (using the same 0.5 SD cut-off for all three distributions). All five loci exhibited reduced levels of relative variability outside of Africa. The result for this smaller subset of loci further bolsters the conclusion that a dis-
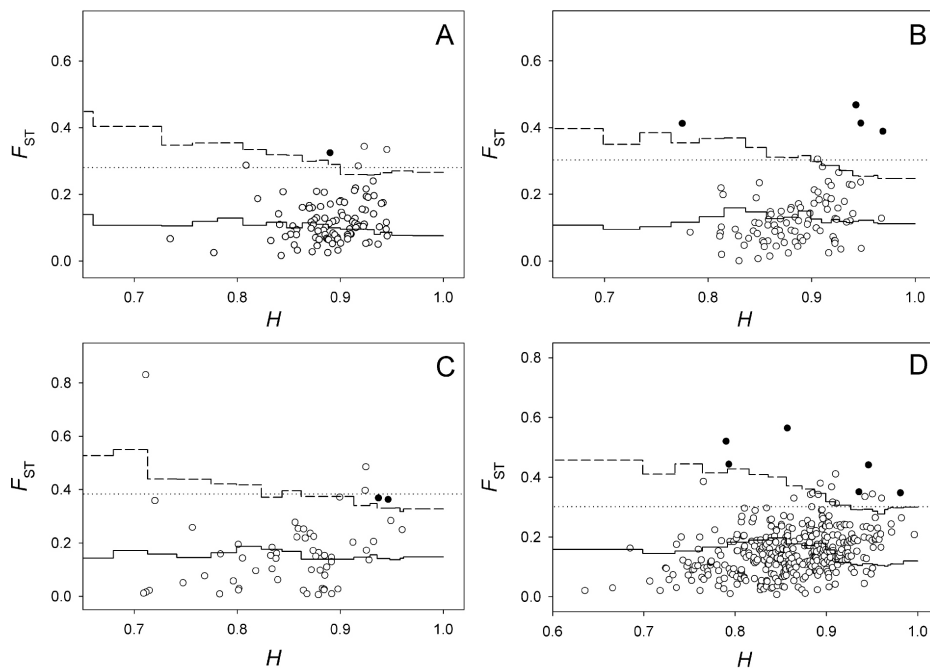
Figure 1. Results of three-way population comparisons using the Biaka as a reference sample for Africa. Estimated F_{ST} values for microsatellite loci are plotted as a function of heterozygosity. The 0.95 and 0.50 quantiles of the simulation-based distribution are denoted by dashed and solid lines, respectively. The horizontal dotted line denotes the 0.95 quantile of the empirical distribution of F_{ST}. Filled symbols denote candidate loci for geographically restricted selective sweeps (based on criteria described in the text). **(A)** Results for the D1 data set ($n = 95$ loci); **(B)** results for the D2 data set ($n = 92$ loci); **(C)** results for the D3 data set ($n = 60$ loci); **(D)** results for the D4 data set ($n = 377$ loci)
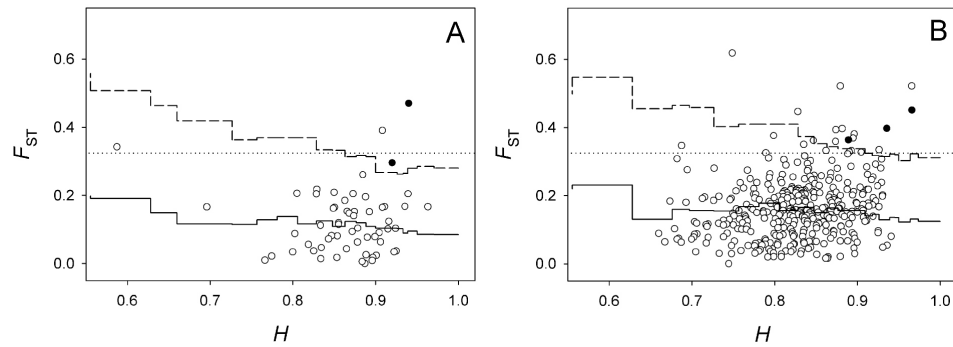
proportionate number of candidate loci exhibit reduced variability outside of Africa.

Loci that were identified as outliers in the model-based analysis were also characterized by extreme values in the empirical distribution of F_{ST} for each three-way population comparison

(table 5). Of the 13 loci identified as candidates for selection in the comparisons using the Biaka and Mbuti samples (table 4), 11 were characterized by F_{ST} values that fell in the upper 0.05 tails of both empirical distributions, and all 13 fell in the upper 0.10 tails (table 5). Of the five loci identified as candidates for

Figure 2. Results of three-way population comparisons using the Mbuti as a reference sample for Africa. Estimated F_{ST} values for microsatellite loci are plotted as a function of heterozygosity. The 0.95 and 0.50 quantiles of the simulation-based distribution are denoted by dashed and solid lines, respectively. The horizontal dotted line denotes the 0.95 quantile of the empirical distribution of F_{ST}. Filled symbols denote candidate loci for geographically restricted selective sweeps (based on criteria described in the text). **(A)** Results for the D1 data set ($n = 95$ loci); **(B)** results for the D2 data set ($n = 92$ loci); **(C)** results for the D3 data set ($n = 60$ loci); **(D)** results for the D4 data set ($n = 377$ loci)

Figure 3. Results of three-way population comparisons using the San as a reference sample for Africa. Estimated F_{ST} values for microsatellite loci are plotted as a function of heterozygosity. The 0.95 and 0.50 quantiles of the simulation-based distribution are denoted by dashed and solid lines, respectively. The horizontal dotted line denotes the 0.95 quantile of the empirical distribution of F_{ST}. Filled symbols denote candidate loci for geographically restricted selective sweeps (based on criteria described in the text). **(A)** Results for the D3 data set ($n = 60$ loci); **(B)** results for the D4 data set ($n = 377$ loci)

selection in the separate comparisons using the Biaka, Mbuti, and San samples (D3 and D4 data sets only), three were characterized by F_{ST} values that fell in the upper 0.05 tails of each of the three empirical distributions, and all five fell in the upper 0.10 tails (table 5).

Discussion

We analyzed genome-wide patterns of DNA polymorphism to test the hypothesis that local adaptation in humans has been more prevalent in colonizing populations outside of sub-Saharan Africa. We conducted a genome scan of microsatellite variability within and among the three major continental populations of humans and identified a number of candidate loci for geographically restricted hitchhiking events. A disproportionate number of these loci exhibited reduced levels of relative variability in non-African populations alone. This result is consis-

tent with the hypothesis that recent positive selection has been more prevalent outside of sub-Saharan Africa. This result is also consistent with the results of another recent study (Kayser, Brauer, and Stoneking 2003) that identified a larger number of candidate loci for positive selection in non-African populations than in African populations. None of our 13 candidate loci identified the same genomic regions as those identified by Kayser, Brauer, and Stoneking (2003), as locations of candidate loci on the same chromosome were never closer than 20 cM. This lack of correspondence is not too surprising given the relatively low combined marker density of the two studies.

Of the 11 outlier loci identified as candidates for selection by Kayser, Brauer, and Stoneking (2003), five were also included in the D4 data set that we analyzed. Although these five loci were characterized by unusually large R_{ST} values between African and European population samples in the study of Kayser, Brauer, and Stoneking (2003), none of them were identified as outliers in our F_{ST}-based simulation analyses. Since Kayser,

Table 3. Summary Statistics for Microsatellite Variability Within Population Samples from Three Continents, from Each of Four Independent Data Sets.

Data Set	Population	NA	H	V	Skewness	Kurtosis
D1 ($n = 95$ loci)	Biaka	10.37	0.807	9.661	0.087	3.330
	Mbuti	8.91	0.783	7.265	0.106	3.350
	Danish	9.50	0.777	7.512	20.104	3.172
	Han Chinese	8.43	0.711	6.805	20.073	3.730
D2 ($n = 92$ loci)	Biaka	6.53	0.763	7.441	0.077	2.554
	Mbuti	5.24	0.682	7.969	20.059	2.659
	N. Italian	6.53	0.736	7.955	20.047	2.885
	Han Chinese	6.11	0.746	8.743	0.077	2.606
D3 ($n = 60$ loci)	Biaka	4.42	0.672	5.093	20.074	1.975
	Mbuti	4.57	0.663	5.321	0.056	2.176
	San	6.57	0.728	3.963	20.192	2.942
	French	6.40	0.713	4.305	20.079	3.017
	Han Chinese	6.02	0.671	5.124	0.079	3.300
D4 ($n = 377$ loci)	Biaka	7.72	0.762	13.032	20.099	2.950
	Mbuti	6.55	0.743	12.816	20.015	2.817
	San	5.08	0.704	12.470	20.016	2.400
	French	6.89	0.739	10.996	20.080	2.893
	Han Chinese	7.14	0.717	10.165	0.064	3.259

Note: NA, number of alleles; H, expected heterozygosity; V, variance in allele size.

Table 4. Loci Identified as Outliers in the Three-way Population Comparisons that Also Exhibited a Discrepancy in Relative Levels of Variability Between African and Non-African Populations.

Locus	Chromosome	Map Position (Marshfield map; Broman et al. 1998)	Local Recombination Rate (cM/Mb)	Continental Population Showing Reduced Variability	No. of Genes Within a 0.5 cM Window Centered on the Marker Locus	Identity of Known Genes Within the 0.5 cM Window
D3S4523	3	138.00 cM	0.8	non-African	56	IGSF11, FLJ32859, UPK1B, B4GALT4, CDGAP, FLJ10902, MDS010, C3orf1, CD80, CSRP2P, ADPRH, PLA1A, POPDC2, COX17, AAT-1, NR1I2, GSK3B, GPR156, FSTL1, NDUFB4, HGD, RABL3, GTF2E1, KIAA1006, POLQ, HCLS1, GOLGB1, KIAA0036, EAF2, SLC15A2, MGC50831, CD86, CASR, CSTA, E2IG5, WDR5B, KPNA1, BAL, BBAP, FLJ40597, KIAA1268, HSPBAP1
D3S2418	3	215.84 cM	2.0	non-African	2	FGF12
D4S3360	4	0.0 cM	1.3	African	2	FLJ90036
D6S400	6	63.28 cM	1.7	non-African	14	TFEB, PGC, FRS3, C6orf49, MGC20741, USP49, BYSL, CCND3, TBN, GUCA1A, GUCA1B, MRPS10, TRERF
D10S597	10	128.73 cM	0.5	non-African	8	XPNPEP1, RNU4P5, ADD3, MXI1, SMNDC1
D10S212	10	170.94 cM	2.2	non-African	31	BNIP3, C10orf39, DPYSL4, PKE, KIAA1674, INPP5A, NKX6-2, DKFZp434A1721, FLJ25954, GPR123, KIAA1768, RASGEF2, UTF1, VENTX2, ADAM8, TUBGCP2, CALCYON, UPA, FLJ26016, ECHS1, PAOX, Sprn, OR6L1P
D15S165	15	20.24 cM	1.0	non-African	23	DKFZP434L187, CHRFAM7A, KIAA1018, FLJ20313, TRPM1
D15S148	15	52.33 cM	3.2	non-African	18	LIPC, ADAM10, KIAA1164, FLJ13213, RNF111, CCNB2, MYO1E, LDHL, MGC26690, GCNT3, GTF2A2, BNIP2, RPS3A-like, FOXB1
D16S508	16	57.79	0.6	non-African	50	XP06, LAT1-3TM, MVP, PSK-1, PSK, ITGAL, ITGAM, ZNF267, TP53TG3, SHCBP1, GPT2, NETO2, CDA08, PHKB, ABCC12, ABCC11, LONP, N4BP1, OAZ, FLJ20718, BRD7, NKD1, CYLD, TNRC9, FLJ12178, RBL2, KIAA1005, FTO, FLJ20481, SLC6A2, GNAO1, AMFR, KIAA0095, SLC6A3, NOD27, CPNE2, KIAA1972, TM4SF11, GPR56, DKFZp434I099, KIFC3, CNGB1, CSNK2A2, NDRG4, KIAA1007, CDH8
D17S2180	17	66.85 cM	0.8	non-African	60	NSF, WNT3, WNT9B, GOSR2, CDC27, MYL4, ITGB3, FLJ40342, NPEPPS, KPNB1, ProSAPiP2, TBX21, OSBPL7, MRPL10, MGC16309, SCRN2, SP6, SP2, PNPO, MGC11242, CDK5RAP3, COPZ2, NFE2L1, CBX1, SNX11, SCAP1, HOXB1, HOXB2, HOXB3, HOXB4, HOXB5, HOXB6, HOXB7, HOXB8, HOXB9, PRAC, HOXB13, FLJ35808, NDP52, ATP5G1, FLJ13855, EAP30, GIP, IMP-1, GALGT2, GNGT2, NESH, PHOSPHO1, KIAA0924, PHB, NGFR

Table 4, continued

Locus	Chromosome	Map Position (Marshfield map; Broman et al. 1998)	Local Recombination Rate (cM/Mb)	Population Showing Reduced Variability	Continental No. of Genes Within a 0.5 cM Window Centered on the Marker Locus	Identity of Known Genes Within the 0.5 cM Window
D19S403	19	32.94 cM	1.8	non-African	27	ZNF177, ZNF266, ZNF560, ZNF426, ZNF561, ZNF562, MGC45408, FLJ20079, UBE2L4, FBXL12, PIN1, OLFM2, COL5A3, RDH8, FLJ11286, ANGPTL-6, PPAN, P2RY11, EIF3S4, DNMT1
D20S103	20	2.13 cM	5.2	non-African	2	CSNK2A1, TCF15
UTSW1523	unknown	unknown		non-African		

Brauer, and Stoneking (2003) used different population samples for Europe and sub-Saharan Africa, a lack of correspondence between the results of our studies would be expected if selection was geographically restricted to specific subpopulations within continents. However, a replicated test using different population samples should identify loci that were subject to continent-wide selective sweeps. To assess whether these five loci also exhibit unusually high levels of differentiation in the D4 data set, we estimated locus-specific F_{ST} and R_{ST} values for each of the six possible pairwise comparisons between our African and non-African population samples. Two of the five loci (D9S2169 and D2S1400) fell in the upper 0.05 tail of the empirical distribution of R_{ST} values in each of the six pairwise comparisons (table 6). The D2S1400 locus also exhibited the highest F_{ST} value in the pairwise comparison between the San and Han Chinese samples, but neither D9S2169 nor any of the other three loci fell in the upper tails of the empirical distributions of F_{ST}. Rank correlations between single-locus F_{ST} and R_{ST} values

Table 5. Levels of Differentiation at Candidate Loci for Geographically Restricted Sweeps.

Data set	Locus	Three-way Comparison	F_{ST}	Ranking (%)
D1	D10S597	Biaka-Danish-Han Chinese	0.323	1 of 95 (0.01)
		Mbuti-Danish-Han Chinese	0.325	3 of 95 (0.03)
D2	D15S148	Biaka-N. Italian-Han Chinese	0.416	2 of 92 (0.02)
		Mbuti-N. Italian-Han Chinese	0.414	2 of 92 (0.02)
	D15S165	Biaka-N. Italian-Han Chinese	0.278	5 of 92 (0.05)
		Mbuti-N. Italian-Han Chinese	0.389	4 of 92 (0.04)
	D16S508	Biaka-N. Italian-Han Chinese	0.361	3 of 92 (0.03)
		Mbuti-N. Italian-Han Chinese	0.413	3 of 92 (0.03)
	UTSW1523	Biaka-N. Italian-Han Chinese	0.337	4 of 92 (0.04)
		Mbuti-N. Italian-Han Chinese	0.468	1 of 92 (0.01)
D3	D19S403	Biaka-French-Han Chinese	0.378	4 of 60 (0.07)
		Mbuti-French-Han Chinese	0.363	6 of 60 (0.10)
		San-French-Han Chinese	0.470	1 of 60 (0.02)
	D6S400	Biaka-French-Han Chinese	0.416	1 of 60 (0.02)
		Mbuti-French-Han Chinese	0.369	5 of 60 (0.08)
		San-French-Han Chinese	0.296	4 of 60 (0.07)
D4	D4S3360	Biaka-French-Han Chinese	0.322	8 of 377 (0.02)
		Mbuti-French-Han Chinese	0.344	12 of 377 (0.03)
		San-French-Han Chinese	0.224	76 of 377 (0.20)
	D10S212	Biaka-French-Han Chinese	0.393	2 of 377 (0.01)
		Mbuti-French-Han Chinese	0.368	8 of 377 (0.02)
		San-French-Han Chinese	0.362	13 of 377 (0.03)
	D3S2418	Biaka-French-Han Chinese	0.283	17 of 377 (0.05)
		Mbuti-French-Han Chinese	0.330	17 of 377 (0.05)
		San-French-Han Chinese	0.451	4 of 377 (0.01)
	D20S103	Biaka-French-Han Chinese	0.289	14 of 377 (0.04)
		Mbuti-French-Han Chinese	0.335	15 of 377 (0.04)
		San-French-Han Chinese	0.397	6 of 377 (0.02)
	D3S4523	Biaka-French-Han Chinese	0.333	6 of 377 (0.02)
		Mbuti-French-Han Chinese	0.411	5 of 377 (0.01)
		San-French-Han Chinese	0.118	259 of 377 (0.69)
	D17S2180	Biaka-French-Han Chinese	0.328	7 of 377 (0.02)
		Mbuti-French-Han Chinese	0.380	7 of 377 (0.02)
		San-French-Han Chinese	0.364	12 of 377 (0.03)

Note: The ranking of locus-specific F_{ST} values (in descending order) and their associated percentiles are specific to each of the four data sets.

Table 6 Levels of Differentiation at Five Loci that Were Identified as Outliers by Kayser, Brauer, and Stoneking (2003).

Locus	Biaka-French		Biaka-Han Chinese		Mbuti-French		Mbuti-Han Chinese		San-French		San-Han Chinese	
	F_{ST}	R_{ST}	F_{ST}	R_{ST}	F_{ST}	R_{ST}	F_{ST}	R_{ST}	F_{ST}	R_{ST}	F_{ST}	R_{ST}
D16S539	0.116	0.070	0.081	−0.005	0.038	0.177	0.032	0.023	0.300	0.036	0.225	−0.010
	149	101	244	302	304	49	333	183	32	145	106	354
	(0.40)	(0.27)	(0.65)	(0.80)	(0.81)	(0.13)	(0.88)	(0.49)	(0.08)	(0.39)	(0.28)	(0.94)
D21S1437	0.117	0.129	0.264	0.006	0.124	0.068	0.350	−0.001	0.119	0.224	0.385	0.073
	142	51	31	252	149	116	28	363	161	26	32	104
	(0.38)	(0.14)	(0.08)	(0.67)	(0.40)	(0.31)	(0.07)	(0.96)	(0.43)	(0.07)	(0.08)	(0.28)
D9S2169	0.103	**0.256**	0.064	**0.287**	0.084	**0.331**	0.059	**0.342**	0.078	**0.351**	0.134	**0.351**
	166	10	277	15	221	11	277	13	235	6	179	6
	(0.44)	(0.03)	(0.73)	(0.04)	(0.59)	(0.03)	(0.73)	(0.03)	(0.62)	(0.02)	(0.47)	(0.02)
D2S1400	0.085	**0.293**	0.043	**0.410**	0.066	**0.477**	0.014	**0.560**	0.294	**0.345**	**0.704**	**0.431**
	204	7	320	5	259	3	357	3	40	7	1	3
	(0.54)	(0.02)	(0.85)	(0.01)	(0.69)	(0.01)	(0.95)	(0.01)	(0.11)	(0.02)	(0.00)	(0.01)
D6S1031	0.170	**0.253**	0.228	0.087	0.140	0.090	0.184	−0.001	0.098	**0.328**	0.138	0.065
	68	11	46	91	134	99	128	316	197	8	176	111
	(0.18)	(0.03)	(0.12)	(0.24)	(0.04)	(0.26)	(0.34)	(0.84)	(0.52)	(0.02)	(0.47)	(0.29)
Average ($n = 5$)	0.118	0.200	0.136	0.157	0.090	0.229	0.128	0.185	0.178	0.257	0.317	0.182
Genome-wide average ($n = 377$)	0.111	0.066	0.129	0.068	0.125	0.075	0.163	0.069	0.138	0.054	0.175	0.048

Note: Locus-specific F_{ST} and R_{ST} values and their rank order in the empirical distributions are given for each of six pairwise comparisons between African and non-African populations. The ranking of locus-specific F_{ST} and R_{ST} values (in descending order) and their associated percentiles are specific to the D4 data set ($n = 377$ loci). Values that fell in the upper 0.05 tail of the empirical distributions are in bold.

were not statistically significant in any of the six pairwise comparisons (Spearman's rank correlation coefficient ranged from 0.077 to 0.024). In summary, two of the outliers identified by Kayser, Brauer, and Stoneking (2003) also appear as outliers in our analysis using completely independent population samples. However, patterns of variation at these loci are only unusual with respect to variance in allele size, not heterozygosity.

Aside from the generally lower sampling variance of F_{ST} relative to R_{ST} (Slatkin 1995; Gaggiotti *et al.* 1999; Balloux and Goudet 2002), statistics based on different measures of polymorphism may be sensitive to the effects of selection over different time scales. Since heterozygosity is expected to return to its equilibrium value more rapidly than variance in allele size following a population bottleneck or selective sweep (Kimmel *et al.* 1998), test statistics based on allele identity may have more power to detect sweeps that have occurred during a relatively recent time interval, whereas test statistics based on allele size may have more power to detect sweeps that occurred in the more distant past. If this is the case, then candidate loci identified by Kayser, Brauer, and Stoneking (2003) may be implicated in more ancient selective sweeps than the candidate loci that we have identified. In any case, the two loci that appeared as outliers in the empirical distributions of R_{ST} both show severely reduced levels of variability in non-African populations relative to Africa, consistent with the pattern exhibited by the candidate loci identified by our F_{ST}-based tests.

The nonequilibrium demographic model that we used was highly conservative with regard to the identification of outliers. Simulations under an equilibrium island model of population

structure typically predicted a sharper decline in F_{ST} at high values of H, consistent with the results of Flint *et al.* (1999). Consequently, greater numbers of loci exceeded the upper quantile of the distribution null at H values above ~0.90. However, the same 13 loci listed in table 4 were consistent outliers in simulations that covered a wide range of different equilibrium and nonequilibrium models of population structure (data not shown).

Because African and non-African populations have experienced different demographic histories, it is reasonable to expect them to be characterized by different standing levels of DNA variability and different distributions of allele frequencies. However, the use of multilocus data allows us to distinguish between the locus-specific effects of selection and the genome-wide effects of demographic processes. Moreover, by considering the distribution of gene-diversity ratios in the pairwise population comparisons, we are able to control for differences in absolute levels of variability between continents.

The microsatellite loci listed in table 4 are flanked by 2–60 genes within a 0.25 cM window on either side. One or more loci within these chromosomal regions may be expected to harbor adaptive mutations that have been driven to unusually high frequencies within a single continental region. This list of candidate genes can be considered as a starting point for efforts to characterize functional polymorphisms that underlie adaptive differentiation in the human gene pool. Using estimates of local recombination rates taken from Kong *et al.* (2002), we assessed whether candidate loci were disproportionately represented in genomic regions of low recombination. We found this not to be the case, as only five candidate loci were characterized by local

recombination rates that fell below the genome-wide average of 1.1 cM/Mb (Kong *et al.* 2002). Thus, if the candidate loci are in fact tracking selection at linked sites, they may be tightly linked to the targets of selection.

Using Genome Scans to Infer Selective Sweeps

The results presented here will be a valuable complement to data from genome-wide surveys of nucleotide polymorphism once such data become more widely available. Comparing multilocus patterns of nucleotide and microsatellite polymorphism can be expected to provide many novel insights into the evolutionary processes that shape patterns of genomic diversity. For example, following a selective sweep, single nucleotide polymorphism and microsatellite length polymorphism will exhibit very different rates of return to mutation-drift equilibrium because mutation rates differ by several orders of magnitude. Deterministic and stochastic models of hitchhiking at microsatellite loci have demonstrated that the strength of hitchhiking effects depends strongly on mutation rates in addition to recombination rates and the selection coefficient (Wiehe 1998). Thus, comparing patterns of nucleotide and microsatellite polymorphism can be expected to provide valuable information about the timing of hitchhiking events. Given the high mutation rates at human microsatellite loci (3×10^{-3} to 6×10^{-4}; Ellegren 2000), theoretical results of Wiehe (1998) suggest that surveys of microsatellite variation are most appropriate for detecting selective sweeps that were both strong and recent (e.g., during the Neolithic). This stands in contrast to nucleotide variation, which may be more appropriate for detecting relatively ancient sweeps, as mutation rates for single nucleotide changes are generally at least four orders of magnitude lower ($\sim 2 \times 10^{-8}$; Nachman and Crowell 2000*a*). In addition, microsatellites may generally provide more power than single nucleotide changes for detecting sweeps, because reductions in microsatellite variability will occur relative to a much higher baseline level of standing variation.

Recent studies of both humans and Drosophila have revealed substantial differences in patterns of DNA variability between African and non-African populations (Begun and Aquadro 1993; Andolfatto 2001; Harr, Kauer, and Schlötterer 2002; Kauer *et al.* 2002; Glinka *et al.* 2003; Kauer, Dieringer, and Schlötterer 2003). These differences may largely reflect founder effects associated with migrations out of Africa. However, patterns of DNA variability may also reflect the effects of positive selection associated with the colonization of novel environments outside the ancestral species range. Consistent with this hypothesis, patterns of nucleotide and microsatellite variability in African and non-African populations of *D. melanogaster* have been interpreted as evidence for geographically restricted selective sweeps associated with adaptation to temperate zone environments (Andolfatto 2001; Harr, Kauer, and Schlötterer 2002; Kauer *et al.* 2002; Glinka *et al.* 2003; Kauer, Dieringer, and Schlötterer 2003). Although several studies have documented directional selection on loci involved in disease resistance within African populations of humans (Hamblin and Di Rienzo 2000; Tishkoff *et al.* 2001; Hamblin, Thompson, and Di Rienzo 2002; Sabeti *et al.* 2002; Saunders,

Hammer, and Nachman 2002), in each case there were a priori reasons to expect that the genes under consideration were involved in local adaptation. By contrast, geographic surveys of X-linked loci in humans that were not previously implicated as candidates for local adaptation have revealed evidence of genetic hitchhiking events that appear to have been largely restricted to non-African populations (e.g., Harris and Hey 1999, 2001; Nachman and Crowell 2000*b*). Our finding that selective sweeps may have been more prevalent outside of Africa appears to be consistent with the picture emerging from global surveys of DNA sequence variation in humans. It remains to be seen whether surveys of X-linked loci will reveal additional evidence for selective sweeps outside of Africa, as in the case of Drosophila (Andolfatto 2001; Harr, Kauer, and Schlötterer 2002; Kauer *et al.* 2002; Glinka *et al.* 2003; Kauer, Dieringer, and Schlötterer 2003).

Literature Cited

Akey, J. M., G. Zheng, K. Zhang, L. Jin, and M. D. Shriver. 2002. Interrogating a high-density SNP map for signatures of natural selection. Genome Res. **12**:1805-1814.

Andolfatto, P. 2001. Contrasting patterns of X-linked and autosomal nucleotide variation in *Drosophila melanogaster* and *Drosophila simulans*. Mol. Biol. Evol. **18**:279-290.

Balloux, F., and J. Goudet. 2002. Statistical properties of population differentiation estimators under stepwise mutation in a finite island model. Mol. Ecol. **11**:771-783.

Beaumont, M. A., and R. A. Nichols. 1996. Evaluating loci for use in the genetic analysis of population structure. Proc. R. Soc. Lond. B **263**:1619-1626.

Begun, D. J., and C. F. Aquadro. 1993. African and North American populations of *Drosophila melanogaster* are very different at the DNA level. Nature **365**:548-550.

Bodmer, W. F., and L. L. Cavalli-Sforza. 1976. Genetics, evolution, and man. W. H. Freeeman, San Francisco, Calif.

Bowcock, A. M., A. Ruiz-Linares, J. Tomfohrde, E. Minch, J. R. Kidd, and L. L. Cavalli-Sforza. 1994. High resolution of human evolutionary trees with polymorphic microsatellites. Nature **368**:455-457.

Braverman, J. M., R. R. Hudson, N. L. Kaplan, C. H. Langley, and W. Stephan. 1995. The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. Genetics **140**:783-796.

Broman, K. W., J. C. Murray, V. C. Sheffield, R. L. White, and J. L. Weber. 1998. Comprehensive human genetic maps: individual and sex-specific variation in recombination. Am. J. Hum. Genet. **63**:861-869.

Cavalli-Sforza, L. L. 1966. Population structure and human evolution. Proc. R. Soc. Lond. B **164**:362-379.

Cavalli-Sforza, L. L., P. Menozzi, and A. Piazza. 1994. The history and geography of human genes. Princeton University Press, Princeton, N.J.

Chen, Y.-S., A. Olckers, T. G. Schurr, A. M. Kogelnik, K. Huoponen, and D. C. Wallace. 2000. MtDNA variation in the South African Kung and Khwe—and their genetic relationships to other African populations. Am. J. Hum. Genet. **66**:1362-1383.

Cockerham, C. C., and B. S. Weir. 1993. Estimation of gene flow from *F*-statistics. Evolution **47**:855-863.

Diamond, J. 1992. The third chimpanzee. Harper Perennial, New York.

Ellegren, H. 2000. Microsatellite mutations in the germline: implications for evolutionary inference. Trends Genet. **16**:551-558.

Eller, E. 1999. Population substructure and isolation by distance in three continental regions. Am. J. Phys. Anthropol. **108**:147-159.

Excoffier, L. 2002. Human demographic history: refining the recent African origin model. Curr. Opin. Genet. Dev. **12**:675-682.

Flint, J., J. Bond, D. C. Rees, A. J. Boyce, J. M. Roberts-Thomson, L. Excoffier, J. B. Clegg, M. A. Beaumont, R. A. Nichols, and R. M. Harding. 1999. Minisatellite mutational processes reduce F_{ST} estimates. Hum. Genet. **105**:567-576.

Fu, Y.-X., and W.-H. Li. 1993. Statistical tests of neutrality of mutations. Genetics **133**:693-709.

Gaggiotti, O. E., O. Lange, K. Rassmann, and C. Gliddon. 1999. A comparison of two indirect methods for estimating average levels of gene flow using microsatellite data. Mol. Ecol. **8**:1513-1520.

Glinka, S., L. Ometto, S. Mousset, W. Stephan, and D. De Lorenzo. 2003. Demography and natural selection have shaped genetic variation in *Drosophila melanogaster*: a multi-locus approach. Genetics **165**:1269-1278.

Hamblin, M. T., and A. Di Rienzo. 2000. Detection of the signature of natural selection in humans: evidence from the Duffy blood group locus. Am. J. Hum. Genet. **66**:1669-1679.

Hamblin, M. T., E. E. Thompson, and A. Di Rienzo. 2002. Complex signatures of natural selection at the Duffy blood group locus. Am. J. Hum. Genet. **70**:369-383.

Hammer, M. F., F. Blackmer, D. Garrigan, M. W. Nachman, and J. A. Wilder. 2003. Human population structure and its effects on sampling Y-chromosome variation. Genetics **164**:1495-1509.

Harpending, H., and A. Rogers. 2000. Genetic perspectives on human origins and differentiation. Annu. Rev. Genomics Hum. Genet. **1**:361-385.

Harr, B., M. Kauer, and C. Schlötterer. 2002. Hitchhiking mapping—a population-based fine mapping strategy for adaptive mutations in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **99**:12949-12954.

Harris, E. E., and J. Hey. 1999. X chromosomal evidence for ancient human histories. Proc. Natl. Acad. Sci. USA **96**:3320-3324.

Harris, E. E., and J. Hey. 2001. Human populations show reduced DNA sequence variation at the Factor IX locus. Curr. Biol. **11**:774-778.

Hollox, E. J., M. Poulter, M. Zvarik, V. Ferak, A. Krause, T. Jenkins, N. Saha, A. I. Kozlov, and D. M. Swallow. 2001. Lactase haplotype diversity in the Old World. Am. J. Hum. Genet. **68**:160-172.

Huttley, G. A., M. W. Smith, M. Carrington, and S. J. O'Brien. 1999. A scan for linkage disequilibrium across the human genome. Genetics **152**:1711-1722.

Ingman, M., H. Kaessmann, S. Pääbo, and U. Gyllensten. 2000. Mitochondrial genome variation and the origin of modern humans. Nature **408**:708-713.

Jin, L., M. L. Baskett, L. L. Cavalli-Sforza, L. A. Zhivotovsky, M. W. Feldman, and N. A. Rosenberg. 2000. Microsatellite evolution in modern humans: a comparison of two data sets from the same populations. Ann. Hum. Genet. **64**:117-134.

Jorde, L. B., M. J. Bamshad, W. S. Watkins, R. Zenger, A. E. Fraley, P. Krakowiak, H. Soodyall, T. Jenkins, and A. R. Rogers. 1995. Origins and affinities of modern humans: a comparison of mitochondrial and nuclear genetic data. Am. J. Hum. Genet. **57**:523-538.

Jorde, L. B., A. R. Rogers, M. Bamshad, W. S. Watkins, P. Krakowiak, S. Sung, J. Kere, and H. C. Harpending. 1997. Microsatellite diversity and the demographic history of modern humans. Proc. Natl. Acad. Sci. USA **94**:3100-3103.

Kaplan, N. L., R. R. Hudson, and C. H. Langley. 1989. The "hitchhiking effect" revisited. Genetics **123**:887-899.

Kauer, M., D. Dieringer, and C. Schlötterer. 2003. A microsatellite variability screen for positive selection associated with the "out of Africa" habitat expansion of *Drosophila melanogaster*. Genetics **165**:1137-1148.

Kauer, M., B. Zangerl, D. Dieringer, and C. Schlötterer. 2002. Chromosomal patterns of microsatellite variability contrast sharply in African and non-African populations of *Drosophila melanogaster*. Genetics **160**:247-256.

Kayser, M., S. Brauer, and M. Stoneking. 2003. A genome scan to detect candidate regions influenced by local natural selection in human populations. Mol. Biol. Evol. **20**:893-900.

Kimmel, M., R. Chakraborty, J. P. King, M. Bamshad, W. S. Watkins, and L. B. Jorde. 1998. Signatures of population expansion in microsatellite repeat data. Genetics **148**:1921-1930.

Kong, A., D. F. Gudbjartsson, J. Sainz, G. M. Jonsdottir, and S. A. Gudjonsson, *et al.* (13 co-authors). 2002. A high-resolution recombination map of the human genome. Nat. Genet. **31**:241-247.

Lewontin, R. C., and J. Krakauer. 1973. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. Genetics **74**:175-195.

Luikart, G., P. R. England, D. Tallmon, S. Jordon, and P. Taberlet. 2003. The power and promise of population genomics: from genotyping to genome typing. Nat. Rev. Genet. **4**:981-994.

Marth, G. T., E. Czabarka, J. Murvai, and S. T. Sherry. 2004. The allele frequency spectrum in genome-wide human variation data reveals signals of differential demographic history in three large world populations. Genetics **166**:351-372.

Maynard Smith, J., and J. Haigh. 1974. The hitch-hiking effect of a favourable gene. Genet. Res. **23**:23-35.

Moran, P. A. P. 1975. Wandering distributions and the electrophoretic profile. Theor. Popul. Biol. **8**:318-330.

Nachman, M. W., and S. L. Crowell. 2000*a*. Estimate of the mutation rate per nucleotide in humans. Genetics **156**:297-304.

Nachman, M. W., and S. L. Crowell. 2000*b*. Contrasting evolutionary histories of two introns of the Duchenne muscular dystrophy gene, *Dmd*, in humans. Genetics **155**:1855-1864.

Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics **89**:583-590.

Ohta, T., and M. Kimura. 1973. A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. Genet. Res. **22**:201-204.

Payseur, B. A., A. D. Cutter, and M. W. Nachman. 2002. Searching for evidence of positive selection in the human genome using patterns of microsatellite variability. Mol. Biol. Evol. **19**:1143-1153.

Przeworski, M. 2002. The signature of positive selection at randomly chosen loci. Genetics **160**:1179-1189.

Ptak, S. E., and M. Przeworski. 2002. Evidence for population growth in humans is confounded by fine-scale population structure. Trends Genet. **18**:559-563.

Relethford, J. H. 2001. Genetic history of the human species. Pp. 813–846 *in* D. J. Balding, M. Bishop, and C. Cannings, eds. Handbook of statistical genetics. John Wiley and Sons, New York.

Relethford, J. H. 2002. Apportionment of global human genetic diversity based on craniometrics and skin color. Am. J. Phys. Anthropol. **118**:393-398.

Rosenberg, N. A., J. K. Pritchard, J. L. Weber, H. M. Cann, K. K. Kidd, L. A. Zhivotovsky, and M. W. Feldman. 2002. Genetic structure of human populations. Science **298**:2381-2385.

Sabeti, P. C., D. E. Reich, J. M. Higgins, H. Z. P. Levine, and D. J. Richter, *et al.* (14 co-authors). 2002. Detecting recent positive selection in the human genome from haplotype structure. Nature **419**:832-837.

Saunders, M. A., M. F. Hammer, and M. W. Nachman. 2002. Nucleotide variability at *G6pd* and the signature of malarial selection in humans. Genetics **162**:1849-1861.

Schlötterer, C. 2002. A microsatellite-based multilocus screen for the identification of local selective sweeps. Genetics **160**:753-763.

Schlötterer, C. 2003. Hitchhiking mapping—functional genomics from the population genetics perspective. Trends Genet. **19**:32-38.

Simonsen, K. L., G. A. Churchill, and C. F. Aquadro. 1995. Properties of statistical tests of neutrality for DNA polymorphism data. Genetics **141**:413-439.

Slatkin, M. 1995. A measure of population subdivision based on microsatellite allele frequencies. Genetics **139**:457-462.

Slatkin, M., and T. Wiehe. 1998. Genetic hitchhiking in a subdivided population. Genet. Res. **71**:155-160.

Stephan, W. 1994. Effects of recombination and population subdivision on nucleotide sequence variation in *Drosophila ananassae*. Pp. 57–66 *in* B. Golding, ed. Non-neutral evolution: theories and molecular data. Chapman Hall, New York.

Stephan, W., and S. J. Mitchell. 1992. Reduced levels of DNA polymorphism and fixed between-population differences in the centromeric region of *Drosophila ananassae*. Genetics **132**:1039-1045.

Stephan, W., T. Wiehe, and M. W. Lenz. 1992. The effect of strongly selected substitutions on neutral polymorphism: analytical results based on diffusion theory. Theor. Popul. Biol. **41**:237-254.

Stephan, W., L. Xing, D. A. Kirby, and J. M. Braverman. 1998. A test of the background selection hypothesis based on nucleotide data from *Drosophila ananassae*. Proc. Natl. Acad. Sci. USA **95**:5649-5654.

Stephens, J. C., D. E. Reich, D. B. Goldstein, H. D. Shin, and M. W. Smith, *et al.* (36 co-authors). 1998. Dating the origin of the CCR5-Δ32 AIDS-resistance allele by the coalescence of haplotypes. Am. J. Hum. Genet. **62**:1507-1515.

Storz, J. F., and M. W. Nachman. 2003. Natural selection on protein polymorphism in the rodent genus *Peromyscus*: evidence from interlocus contrasts. Evolution **57**:2628-2635.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123**:585-595.

Tishkoff, S. A., R. Varkonyi, N. Cahinhinan, S. Abbes, and G. Argyropoulos, *et al.* (14 co-authors). 2001. Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. Science **293**:455-462.

Vigouroux, Y., M. McMullen, C. T. Hittinger, K. Houchins, L. Schulz, S. Kresovich, Y. Matsuoka, and J. Doebley. 2002. Identifying genes of agronomic importance in maize by screening microsatellites for evidence of selection during domestication. Proc. Natl. Acad. Sci. USA **99**:9650-9655.

Wakeley, J., R. Nielsen, S. N. Liu-Cordero, and K. Ardlie. 2001. The discovery of single-nucleotide polymorphisms—and inferences about human demographic history. Am. J. Hum. Genet. **69**:1332-1347.

Wiehe, T. 1998. The effect of selective sweeps on the variance of the allele distribution of a linked multiallele locus: hitchhiking of microsatellites. Theor. Popul. Biol. **53**:272-283.

Wiehe, T., and W. Stephan. 1993. Analysis of a genetic hitchhiking model and its application to DNA polymorphism data from *Drosophila melanogaster*. Mol. Biol. Evol. **10**:842-854.

Yu, N., F. C. Chen, S. Ota, L. B. Jorde, P. Pamilo, L. Patthy, M. Ramsay, T. Jenkins, S. K. Shyue, and W. H. Li. 2002. Larger genetic differences within Africans than between Africans Genetics **161**:269-274.

Zhivotovsky, L. A., N. A. Rosenberg, and M. W. Feldman. 2003. Features of evolution and expansion of modern humans, inferred from genome-wide microsatellite markers. Am. J. Phys. Anthropol. **72**:1171-1186.