Am. J. Hum. Genet. 66:969–978, 2000

Significant Admixture Linkage Disequilibrium across 30 cM around the FY Locus in African Americans

James A. Lautenberger,[1] J. Claiborne Stephens,[1,*] Stephen J. O'Brien,[1] and Michael W. Smith[2]

[1]Laboratory of Genomic Diversity, National Cancer Institute, and [2]Intramural Research Support Program, Science Applications International Corporation–Frederick, National Cancer Institute, Frederick Cancer Research and Development Center, Frederick, MD

Scientists, to understand the importance of allelic polymorphisms on phenotypes that are quantitative and environmentally interacting, are now turning to population-association screens, especially in instances in which pedigree analysis is difficult. Because association screens require linkage disequilibrium between markers and disease loci, maximizing the degree of linkage disequilibrium increases the chances of discovering functional gene-marker associations. One theoretically valid approach—mapping by admixture linkage disequilibrium (MALD), using recently admixed African Americans—is empirically evaluated here by measurement of marker associations with 15 short tandem repeats (STRs) and an insertion/deletion polymorphism of the AT3 locus in a 70-cM segment at 1q22-23, around the FY (Duffy) locus. The FY polymorphism (−46T→C) disrupts the GATA promoter motif, specifically blocking FY erythroid expression and has a nearly fixed allele-frequency difference between European Americans and native Africans that is likely a consequence of a selective advantage of FY−/− in malaria infections. Analysis of linkage disequilibrium around the FY gene has indicated that there is strong and consistent linkage disequilibrium between FY and three flanking loci (D1S303, SPTA1, and D1S484) spanning 8 cM. We observed significant linkage-disequilibrium signals over a 30-cM region from −4.4 to 16.3 cM (from D1S2777 to D1S196) for STRs and at 26.4 cM (AT3), which provided quantitative estimates of centimorgan limits, by MALD assessment in African American population-association analyses, of 5–10 cM.

Introduction

The detection of polymorphic genes that influence quantitative traits, disease states, and other characters is the goal of population-genetic–association studies, but such detection depends on the persistence of measurable linkage disequilibrium (i.e., haplotype-allele association) among markers and discovered loci. In European Americans, the extent and usefulness of linkage disequilibrium is limited by recent population history (Bodmer 1986; Laan and Pääbo 1997; Huttley et al. 1999). The power of this approach depends on how far linkage disequilibrium extends over a chromosomal interval, which in turn determines the spacing and number of markers required for a genome scan. An approach to maximization of the linkage-disequilibrium interval for gene-localization studies involves mapping by admixture linkage disequilibrium (MALD), whereby populations composed of recently mixed ethnic groups display transient linkage disequilibrium over longer centimorgan intervals, for ⩾20 generations, as a consequence of admixture (Briscoe et al. 1994; Stephens et al. 1994). Theoretical and simulation studies that predict the limits of population parameters influencing MALD assessment have been described (Chakraborty and Weiss 1988; Chakraborty et al. 1991; Briscoe et al. 1994; Stephens et al. 1994), but empirical demonstration of the effect in the human population has not been reported with short tandem repeats (STRs).

The African American descendants of native Africans, some of whom first arrived in the United States >450 years ago, are ideal subjects for MALD-based association ascertainment. Native African slaves were present in the Spanish colonies that are now part of the United States as early as 1526 and were present in British colonies as early as 1619. A total of ~380,000–570,000 native Africans were brought to the United States; the largest number arrived during 1790–1808 (Parra et al. 1998). Official population estimates indicate that the U.S. African American population has grown, by births and additional immigration, from 760,000 in 1790 to 34.9 million at present (U.S. Census Bureau). Studies have shown that African Americans represent an admixed population with significant genetic contributions from both African and European ancestors. Recent estimates of the proportion of Eu-

Received September 30, 1999; accepted for publication September 30, 1999; electronically published March 9, 2000.

Address for correspondence and reprints: Dr. Michael W. Smith, SAIC-Frederick, NCI/FCRDC, P.O. Box B, Frederick, MD 21702. E-mail: Michael_W_Smith@nih.gov

* Present affiliation: Genaissance Pharmaceuticals, Five Science Park, New Haven.

0002-9297/2000/6603-0020$02.00

ropean American genes in African American populations range from 6.8%, for Sapelo Island, GA, to 26%, for Chicago (Long 1991; Chakraborty et al. 1992; Parra et al. 1998; Destro-Bisol et al. 1999).

MALD assessment provides tremendous potential for discovery of novel genes involved in common diseases. As a test of the technique, we examined the region adjacent to the Duffy chemokine-receptor locus (FY on chromosome 1q22-23), which carries a biallelic polymorphism ("+" denotes an intact promoter GATA motif, and "−" denotes a promoter motif disrupted by a −46 T→C substitution, GACA), which abrogates Duffy receptor expression on erythroid cells. There are no known phenotypic costs for the lack of expression of the FY gene, since African Americans who do not express the Duffy antigen in their red blood cells are not known to suffer any adverse effects and have red blood cells with no apparent abnormalities. FY expression has been observed in other tissues, such as spleen and endothelial cells in FY−/− individuals (Chaudhuri et al. 1995; Hadley and Peiper 1997). Indeed, the FY− allele probably arose as a consequence of selection for resistance to malaria (Miller et al. 1976; Horuk et al. 1993; Hadley and Peiper 1997). The FY+ allele is effectively monomorphic in European Americans, whereas the FY− allele is nearly fixed in native Africans (Roychoudhury and Nei 1988). In this report, we quantify the length and extent of linkage disequilibrium due to population admixture, in an 80-cM interval around the FY locus, on the basis of the nonrandom association of FY group-specific alleles with 15 STR loci and with the AT3 insertion/deletion polymorphism.

Subjects and Methods

Subjects

Human DNA was extracted from immortalized lymphoblastoid B cell lines established from subjects in six cohorts developed for AIDS epidemiology studies (Smith et al. 1997). The cohorts were those of the AIDS Link to the Intravenous Experience (ALIVE [Vlahov et al. 1991]), D. C. Gay Cohort Study (DCG; Goedert et al. 1987), Human Growth and Development Study (HGDS [Loveland et al. 1994]), Multicenter AIDS Cohort Study (MACS [Ginzburg et al. 1988]), Multicenter Hemophiliac Cohort Study (MHCS [Goedert et al. 1989]), and San Francisco City Clinic Study (SFCC [Jaffe et al. 1985]).

Genotyping

Polymorphisms in the human AT3 gene that were due to the presence of 32- or 108-bp nonhomologous DNA (Bock and Levitan 1983; Wu et al. 1989), the STR loci (Hentati et al. 1992; Dib et al. 1996), and the FY polymorphism (Tournamille et al. 1995; Iwamoto et al. 1996) were assayed (table 1). The FY polymorphism was typed in 2,909 European American and African American individuals from the six HIV-1/AIDS cohorts listed above: ALIVE (34 European Americans and 665 African Americans), DCG (235 European Americans and 18 African Americans), HGDS (210 European Americans and 20 African Americans), MACS (790 European Americans and 105 African Americans), MHCS (553 European Americans and 70 African Americans), and SFCC (201 European Americans and 8 African Americans). A stratified sample of African Americans was selected on te basis of FY genotypes, which consisted of +/+ (n = 38), +/− (n = 121), and some −/− (n = 139) individuals from the ALIVE (n = 274), HGDS (n = 4), MACS (n = 10), MHCS (n = 9), and SFCC (n = 1) studies, along with European American FY +/+ (n = 48, from MHCS), for genotyping with the STRs and AT3. STR map locations and the GenBank accession numbers for flanking sequences were obtained from Généthon. The selection of STR primers and dyes (FAM, HEX, or TET) for multiplexing was made on the basis of product size and color. Primer synthesis was performed by Genosys.

PCR was performed in 96-well plates with PE Biosystems model 9600 thermal cyclers. Reactions were generally carried out in 15-μl volumes containing 50 ng of genomic DNA, 0.5 mM of each primer, 250 μM of each dNTP, 10 mM Tris-HCl, 50 mM KCl, 0.6 U of Ampli*Taq* Gold DNA polymerase (PE Biosystems), and specified $MgCl_2$ concentrations (table 1). Amplification consisted of 35 cycles of 30 s at 94°C, 30 s at the indicated annealing temperature, and 1 min at 72°C. Reactions were incubated for 9 min at 94°C before PCR and for 10 min at 72°C after PCR. PCR products of the individual STR reactions were combined into pooled panels, mixed with loading buffer and GENESCAN-350 TAMRA-labeled markers, and resolved on a PE Biosystems 377 sequencer with Long Ranger gel mix (FMC Bioproducts), under standard conditions specified by the manufacturer. Gels were analyzed with GENESCAN 2.0, and then alleles were binned and assigned on the basis of estimated sizes. Genotypes were determined by GENOTYPER 2.1 software (PE Biosystems).

Statistical Analysis

Commercial and custom software packages were used in the analysis. Fisher's exact test and a log-likelihood-ratio χ^2 test (G-test) were performed, with the SAS system, on the Advanced Biomedical Computing Center DEC Alpha AXP computer. Contingency analysis of linkage disequilibrium between FY genotypes and flanking markers was examined, in African Americans, with all of the alleles at each locus, by the log-likelihood-ratio χ^2 test in four genetic models and on the basis of the

Table 1

Summary of Loci Studied, PCR Conditions, and Primer Sequences

Locus[a]	Annealing Temperature (°C)	$MgCl_2$ (mM)	Allele Size (bp)	Primer Sequence[b] (5′→3′)
D1S206	55	1.5	195–223	…
D1S252	55	1.5	92–120	…
D1S514	60	2	310–336	TTTTCATCTACCTATCTCATCCAGC
				GTCAGACTTCCATCTGGACTAATAGG
D1S2346	60	2	122–158	TTTATCTTGCCCTGCACCTC
				GCTCCATCCTGTGTCCTCCC
D1S2777	60	2	314–344	GCACCACGGAACTCCAGTAT
				5′GTGACCTCCTGGCTCAAGC
D1S303	60	2	99–119	AGCGAAACTCCATCTCAATACA
				GTGTTTGTATGTATGCATGTATGTGT
SPTA1	60	2	233–251	TGTAAATATGCACACAAACACAAGC
				TGCAGCAATATATGGACCCA
D1S2635	60	2	138–162	TAGCAGATCCCCCGTC
				GTGAATCCTACCCCTAAGTAGAAT
FY[c]	59	1.5	150, 172	CCCTCATTAGTCCTTGGCTCgTA [d]
				AACAGCAGGGGAAATGAGG
D1S484	55	2	268–280	…
D1S2705	60	2	248–268	TGCCCATACCACATTGGTC
				GAAACAGGCCACACTCAATATT
D1S2768	60	2	193–213	GAGGCCAGGAGAAGTAAATGG
				GAAATCCCCGCACATATTTGA
D1S2844	60	2	169–205	CCCAGCCTTCCCTATTGTCT
				GTCCATCTTTGTGGAAATAAATGA
D1S196	55	2	309–323	…
D1S452	60	2	114–132	AAGCACTTTTCTCTCCCTCTCC
				GCATCCCTCCACTTGCAAATT
AT3[e]	57	2	496, 572	CCACAGGTGTAACATTGTGT
				GAGATAGTGTGATCTGAGGC
D1S238	55	1.5	291–317	…

[a] Polymorphisms at each locus are STRs, except for a single-nucleotide variant (FY) and an insertion/deletion (AT3).

[b] Primer sequences shown were designed with Primer 0.5 (Lincoln et al. 1991), except for those for D1S2635 (Dib et al. 1996), AT3 (Wu et al. 1989), and FY; an ellipsis denotes that primer sequences were from the PE-ABI alpha test set of microsatellite markers.

[c] Products were digested with *Rsa*I (5 U) for 2 h to overnight at 37°C and were resolved on agarose gels.

[d] Primer has a misincorporated g at the −3 position, which generates a *Rsa*I site when the C is present in the FY− allele.

[e] Products were directly resolved on agarose gels.

presence or absence of individual alleles by means of Fisher's exact test. The four genetic models considered were + dominant (+/+ vs. +/− or −/−), − dominant (−/− vs. +/− or +/+), codominant (+/+ vs. +/− vs. −/−), and homozygous only (+/+ vs. −/−). The resulting raw *P* values were corrected for multiple comparisons with the step-down Holm-Sidak procedure (Sidak 1967; Holm 1979; Sokal and Rohlf 1995; Weir 1996; Ludbrook 1998), for the number of tests performed per locus, for allelic tests, and for the 16 loci considered in the four genetic models tested (table 2 and fig. 1). The step-down Holm-Sidak adjustment for multiples tests uses the formula $P' = 1 - (1 - P)^n$, where n is the number of P values greater than or equal to that being corrected. This correction results in an unbiased probability with a uniform distribution from 0 to 1, unlike the overly conservative Bonferroni correction (mP, where m is the total number of tests), which has a distribution from 0 to >1 (Sokal and Rohlf 1995; Weir 1996; Ludbrook 1998). An evaluation of the *P* values serves as an indication of the strength of linkage disequilibrium for multiallelic loci such as STRs (Peterson et al. 1995).

Haplotypic Estimation

Approximate haplotype frequencies were estimated, for the purpose of calculation of linkage-disequilibrium coefficients, with FY+, by the EM algorithm (Long et al. 1995). Because of the stratified oversampling of +/+

Table 2

Significant Associations between FY and Flanking Polymorphic Loci and Their Alleles

Name	Distance[a] (cM)	Composite δ	Locus Tests			Allele Tests			
			Model[b]	P	P'[c]	Size	D'[d]	P[e]	P'[c]
D1S206	−32.5	.282	Codominant	.036		205	.129	.019	
						203	−.850	.034	
						213	.071	.047	
D1S252	−15.0	.214				110	.256	.011	
D1S514	−12.7	.353	Homozygous	.042		324	−.325	.011	
D1S2346	−6.6	.357	+ Dominant	.050		142	.421	.008	
			− Dominant	.0078		128	−.490	.014	
			Codominant	.017		134	1.000	.024	
			Homozygous	.0036	.028				
D1S2777	−4.4	.254	+ Dominant	.0078		322	−.646	.0015	.017
			− Dominant	.0064		326	.234	.011	
			Codominant	.0091					
			Homozygous	.0012	.010				
D1S303	−4.4	.312	+ Dominant	5.0×10^{-5}	.00065	115	.790	2.2×10^{-6}	2.0×10^{-5}
			− Dominant	2.6×10^{-8}	3.8×10^{-7}	111	.276	.00017	.0013
			Codominant	1.3×10^{-8}	1.9×10^{-7}	107	−.388	.004	.030
			Homozygous	1.1×10^{-8}	1.4×10^{-7}	109	−.659	.011	
SPTA1	−.9	.304	+ Dominant	1.9×10^{-6}	2.6×10^{-5}	237	.482	4.0×10^{-9}	4.0×10^{-8}
			− Dominant	2.5×10^{-6}	3.5×10^{-5}	239	−.398	.0035	.031
			Codominant	2.7×10^{-8}	3.5×10^{-7}	235	−.673	.0061	.048
			Homozygous	5.8×10^{-9}	8.2×10^{-8}				
D1S2635	.0	.332	+ Dominant	3.7×10^{-9}	5.9×10^{-8}	150	−.736	6.3×10^{-8}	8.2×10^{-7}
			− Dominant	3.3×10^{-11}	5.2×10^{-10}	146	.455	2.2×10^{-6}	2.7×10^{-5}
			Codominant	6.7×10^{-13}	1.1×10^{-11}	148	−.612	3.1×10^{-5}	.00035
			Homozygous	3.6×10^{-14}	5.7×10^{-13}	156	.663	.00022	.0022
						154	.203	.0046	.040
						158	.404	.027	
D1S484	3.8	.240	+ Dominant	1.5×10^{-7}	2.3×10^{-6}	268	−.894	3.8×10^{-7}	2.6×10^{-6}
			− Dominant	1.7×10^{-5}	.00022	276	.490	.00044	.0026
			Codominant	2.3×10^{-8}	3.2×10^{-7}				
			Homozygous	1.9×10^{-9}	2.8×10^{-8}				
D1S2705	5.0	.193	+ Dominant	.0016	.031	260	−.440	.033	
			Codominant	.0035		258	−.513	.037	
			Homozygous	.00056	.0056	268	.276	.044	
D1S2768	6.7	.311	+ Dominant	.0010	.012	193	−.998	.020	
			− Dominant	.0059		201	.217	.022	
			Codominant	.0016	.017	203	.183	.033	
			Homozygous	.00011	.0013				
D1S2844	9.1	.131				191	.185	.011	
						189	−.561	.019	

(*continued*)

Table 2 Continued

Name	Distance[a] (cM)	Composite δ	Locus Tests			Allele Tests			
			Model[b]	*P*	*P*′[c]	Size	*D*′[d]	*P*[e]	*P*′[c]
D1S196	16.3	.233	+ Dominant	.0010	.011	315	.459	3.0×10^{-6}	2.4×10^{-5}
			− Dominant	.011					
			Codominant	.0019	.018				
			Homozygous	.00020	.0022				
D1S452	23.7	.273	− Dominant	.0089		126	.158	.0079	
			Codominant	.045		122	−.761	.025	
			Homozygous	.010					
AT3	26.4	.358	− Dominant	2.1×10^{-5}	.00025	496	−.176	8.3×10^{-5}	.00017
			Codominant	8.4×10^{-5}	.0010	572	.176	8.3×10^{-5}	.00017
			Homozygous	.036					
D1S238	36.6	.247	− Dominant	.012		293	1.000	.0082	
			Codominant	.013					

NOTE.—Data are for significant allelic and locus associations based on a $P \leq .05$ model for the loci examined around the FY locus.

[a] Relative to FY (found at 170.1 cM on chromosome 1). Radiation-hybrid locations for FY and the linked coding genes SPTA1 and AT3 were obtained from GeneMap'98 (Deloukas et al. 1998) and were converted to equivalent centimorgan values by interpolation using the radiation-hybrid locations of flanking Généthon STR markers. The determination that two YACs (887G1 and 788B1) previously shown to contain SPTA1 and D1S2635 (Hudson et al. 1995) contain FY (a determination made by PCR analysis) localized these three markers to within a cloned 1.7-Mb segment, confirming their proximal locations.

[b] See Statistical Analysis subsection.

[c] Adjusted on the basis of either the number of alleles at a locus or the 16 loci considered in each genetic model. Only values ≤.05 are shown.

[d] For linkage with FY+ in African Americans.

[e] From two-tailed Fisher's exact tests performed on a 3 × 2 contingency table of African American FY genotypes, by frequencies of each allele (the number of alleles of that type vs. alternative alleles).

and +/− individuals, additional genotypes were estimated in the +/− and −/− classes. To accurately represent FY and STR genotype frequencies in the population, an additional 415 −/− samples were estimated, on the basis of allelic frequencies observed in the −/− sample and 158 +/− samples, by multiplying +/+ frequencies by −/− frequencies. The EM-algorithm disequilibrium coefficients were used to estimate the normalized linkage disequilibrium coefficient (D'), to judge relative magnitude (Lewontin 1964).

Results

Loci in Linkage Disequilibrium

Initially, the FY-promoter polymorphism was characterized in European Americans and African Americans (2,023 European Americans and 886 African Americans), in whom the frequencies were .991 and .203, respectively, for the FY+ allele (FY− allele frequencies are 1 − FY+). Allele frequencies for 48 European Americans and 298 African Americans were determined, and D' values are available from the Laboratory of Genomic Diversity. Statistical tests for admixture linkage disequilibrium (ALD) between the loci and alleles at each locus and the alleles of FY found substantial ALD (table 2 and fig. 1). Analyses with the four FY genetic models + dominance (+/+ vs. others), – dominance (−/− vs. others), codominance (+/+ vs. +/− vs. −/−), and homozygote-only comparisons (+/+ vs. −/−), in contingency tests with a second dimension of allele counts from the African American samples, found strong linkage disequilibrium across the region (table 2). For two models of genetic association—FY codominant and homozygote-only comparisons—the plots of the extent of pairwise ALD between FY and adjacent loci, illustrate the strong core of ALD over 8 cM and significant signals across 30 cM. Some linkage disequilibrium was seen between the AT3 locus and D1S452, located ~2.7 cM away (P = .05 to .002, for the equivalent four genetic models), in addition to the AT3 FY ALD described below. However, the more distant flanking markers were not significantly associated with AT3 from D1S2635 to D1S238 (P = .94 to .01 [analysis not shown]). Analyses with all four genetic models consistently found significant ALD, from −4.4 to 3.8 cM, between FY and the loci D1S303, SPTA1, D1S2365, and D1S484 (P' = .0002 to 6×10^{-13}). Around this core of ALD from −6.6 to 6.7 cM, at least one model resulted in significant

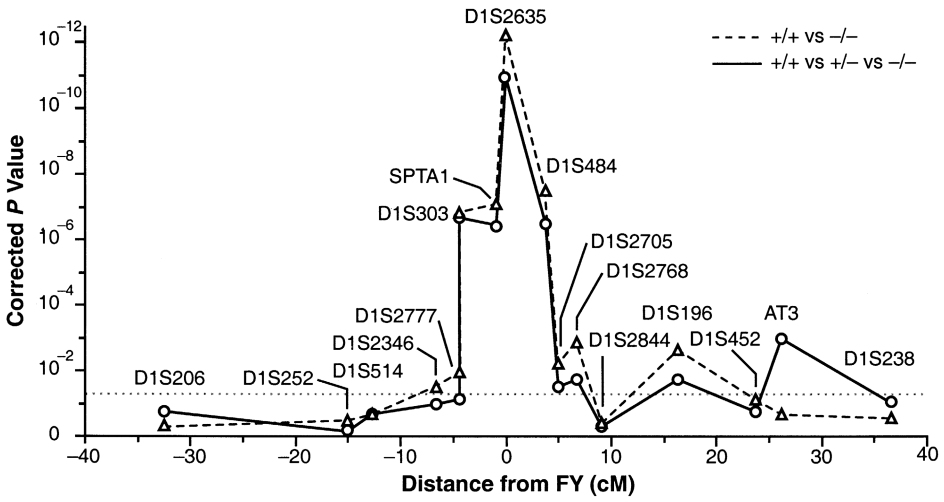

Figure 1 Degree of linkage disequilibrium in African Americans, at 16 loci spanning 70 cM around the FY locus. At each locus, multiple-test–corrected probabilities from two genetic models are shown, as both a codominant comparison of all three genotypes (+/+ vs. +/− vs. −/−) and a comparison of homozygotes (+/+ vs. −/−). Log-likelihood-ratio χ^2 contingency tests of FY genotypes on the basis of flanking marker alleles are shown as log-linear plots of P' in relation to distance (cM) from FY. The dotted line separates significant ($P' = .05$) results from the remainder.

ALD between FY and each of the additional flanking loci D1S2346, D1S2777, D1S2705, and D1S2768 ($P' = .03$ to $.001$). Additional significant associations, at 16.3 and 26.4 cM, for D1S196 and AT3, respectively, suggest that ALD can extend across 30 cM.

Alleles in Linkage Disequilibrium

A further analysis of linkage disequilibrium was made on the basis of an examination of individual alleles of loci linked to FY, through estimation of allele frequencies for each racial-genotypic category and statistical tests of these data on an allele-by-allele basis (table 2). Specific alleles at flanking loci were associated with the FY polymorphism even when raw P values were corrected (P'; see the Subjects and Methods section) for the number of alleles tested at a locus (table 2; 37 of 176 STR alleles, many of which are rare, for $P \leqslant .05$ and 15 alleles for $P' \leqslant .05$). Many of the significant alleles have frequencies in European Americans that are similar to those seen in +/+ African Americans, with +/− being intermediate and with −/− being at the extreme (fig. 2), as predicted for ALD. In particular, from −4.4 to 3.8 cM, alleles at the loci D1S2777, D1S303, SPTA1, D1S2635, and D1S484 showed strong and consistent ALD ($P' = .05$ to 4×10^{-8} and $D' = -.894$ to $.790$; table 2). Strong allelic associations were also seen at 16.3 and 26.4 cM (D1S196, allele 315, $P' = .00002$ and $D' = .46$; AT3, both alleles, $P = .0002$ and $D' = .18$). These allelic results also show both strong and consistent STR-based ALD at the same −4.4-to-3.8 cM core ($P' = .05$ to 4×10^{-8}) and additional strong signals at 16.3 and 26.4 cM ($P' = .0002$ to $.00002$). In the 23-cM genetic segment defined by D1S2346 (−6.6 cM) to D1S196 (16.3 cM), all 10 STR loci examined, except D1S2844 (9.1 cM), show at least one significant multiple-test–corrected association, and the significant associations with the AT3 insertion/deletion extends the ALD region to ~30 cM.

Discussion

Extent of ALD around FY

Linkage disequilibrium is fundamentally the nonrandom association of alleles among different loci. The FY-promoter polymorphism provides an excellent test of ALD because of the essentially fixed allele-frequency differences between native African and European American populations. Linkage disequilibrium between FY and the flanking STR loci is readily detectable over a wide range of genetic distances, from −4.4 to 16.3 cM. Seven of eight STR loci in this 23-cM region from D1S2346 to D1S196 yielded at least one significant corrected P' value ($P' = .05$ to 6×10^{-13} in one of the allelic or locus tests; table 2). The confirmation of ALD between FY and the AT3 locus previously reported in African Americans (Parra et al. 1998) extends the total region of ALD to 30 cM. The most striking and statistically consistent genetic-model–based locus tests were across an 8-cM core around FY (from D1S303 to D1S484, $P' = .0008$ to 6×10^{-13}).

Finding disease-causing lesions of either European

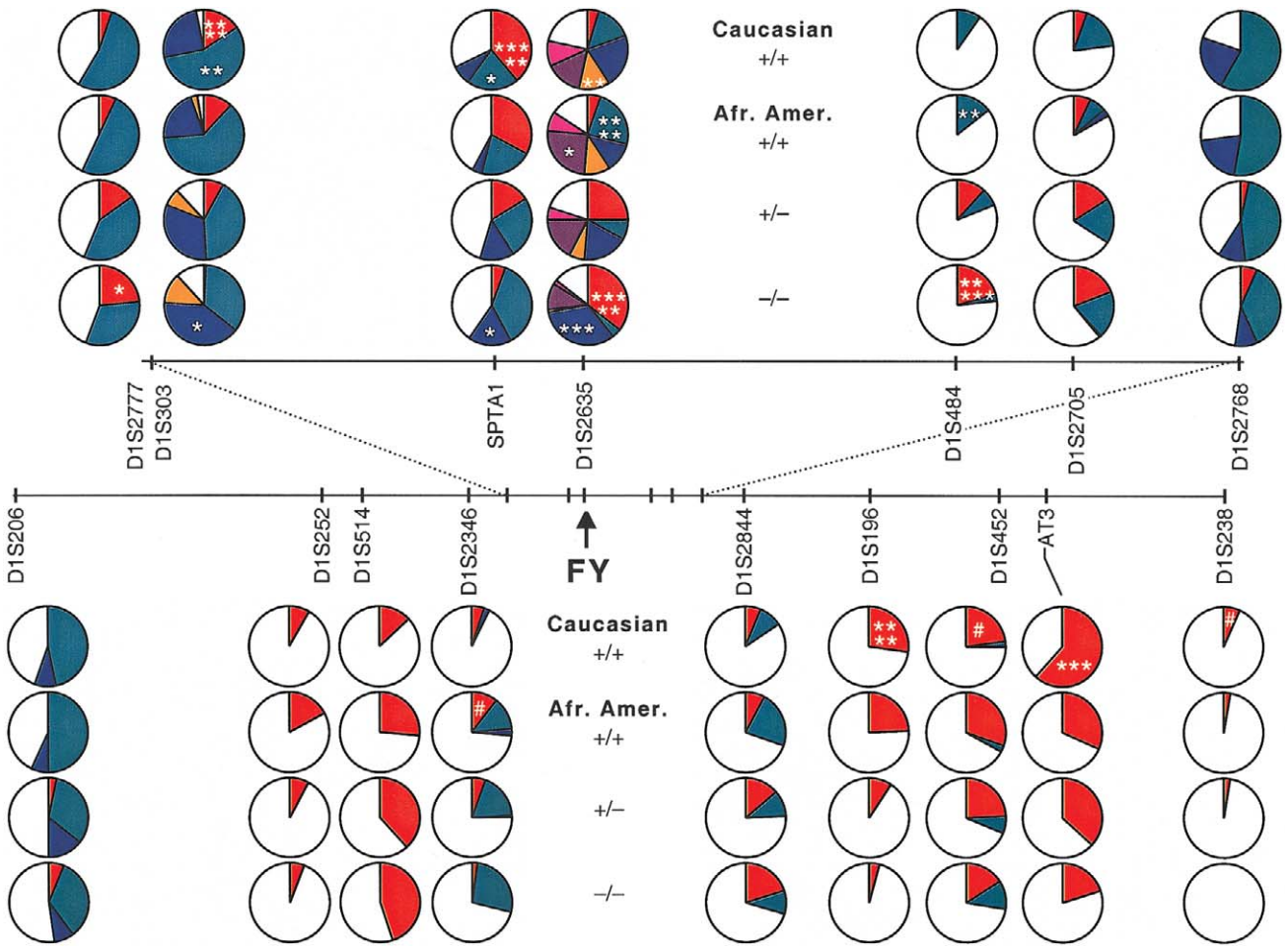

Figure 2 Allele frequencies from loci around the FY locus that have significant allelic differences between FY+/+ African Americans, FY+/− African Americans, and FY−/− African Americans, with color coding reflecting, in a clockwise pattern, the most-significant to the least-significant test results. Only alleles with $P \leq .05$ are shown individually, and those which were still significant after multiple-test correction (P') are marked as follows: *, $P' \leq .05$; **, $P' \leq .01$; ***, $P' \leq .001$; ****, $P' \leq .0001$; and *****, $P' \leq .00001$. Alleles whose initial P value was $\leq .01$ but was not significant after correction for multiple comparisons are denoted by a "pound" sign (#). Allele frequencies from European Americans (+/+) are shown for comparison.

American or native African origins in African Americans is practical in MALD, as judged by the ability of all four genetic models to efficiently reveal ALD around FY in this nearly ideal instance. The analysis of all African American individuals with European American-derived FY alleles versus those without (+ dominant model: +/+ and +/− vs. −/−) was significant for STRs from −4.4 to 16.3 cM (from D1S303 to D1S196, $P' = .01$ to $.0007$). The converse test—of all those African Americans with African FY alleles versus those without any (− dominant model: +/+ vs. +/− and −/−)—was significant over a smaller region, from −4.4 cM to 3.8 cM (from D1S303 to D1S484, $P' = .0002$ to 4×10^{-7}), with the AT3 locus at 26.4 cM also showing significance ($P' = .0003$). The codominance model of +/+ versus +/− versus −/− also detected ALD, from −4.4 to 16.3 cM (D1S303 to D1S196, $P' = .02$ to 2×10^{-7}) and at 26.4 cM with the AT3 insertion/deletion ($P' = .001$). The strongest results were obtained with the homozygotes-only comparisons of +/+ versus −/−, from −6.6 to 16.3 cM (from D1S2346 to D1S196, $P' = .03$ to 6×10^{-13}). The pattern of significance, which indicates the strength of linkage disequilibrium (Peterson et al. 1995), peaks at the locus closest to FY (D1S2635; $P' = 6 \times 10^{-8}$ to 6×10^{-13}) and declines in either direction. The strong 8-cM core of ALD is flanked by additional significant markers, extending it to ≥16 cM, and some signals indicate a 30-cM region. It is reasonable to suppose that a genome scan would have seen any of these signals. Further analysis saturating the region with markers similar to those

used in our study would then localize the strongest signal to within a few centimorgans of FY.

History of Admixture in African Americans

It is difficult to predict expected levels of disequilibrium, since actual levels depend on the interaction of multiple forces, including recombination, genetic drift, migration, and natural selection. In the case of disequilibrium caused by admixture, it also depends on the pattern of admixture and on the allele frequencies and their differences between the admixing populations. One can use the standard linkage-disequilibrium/recombination–decay calculation to calculate the "half-life" of initial levels of linkage disequilibrium in a simple scenario in which recombination is the major determinant. For instance, 13.5 generations is enough to reduce initial disequilibrium by 50% when markers are separated by 5 cM, whereas 69 generations are needed for the same effect at 1-cM separation. However, since only 6.6 generations are needed to reduce by half the disequilibrium at distances of 10 cM, it is reasonable to assume that appreciable ALD seen at $\geqslant$10 cM has been maintained, to some extent, by recurrent admixture in African Americans.

Implications for MALD

The analysis of the loci examined in the present report provides information on the suggested features of a MALD genome-scanning marker set. As predicted, those loci with smaller differences between African Americans and European Americans—as measured by one index of information content, composite δ (the sum of positive allele-frequency differences, at a locus, between African Americans and European Americans [Shriver et al. 1997; Stephens et al. 1999])—were only somewhat less able to allow detection of ALD than were loci at equivalent distances from FY (e.g., D1S2777 vs. D1S303; table 2). The locus-based P' values were generally more sensitive to linkage disequilibrium than were the allelic ones, by several orders of magnitude—with one important exception, at 16.3 cM, where the results of locus tests of D1S196 (P' = .02 to .002) were much less significant than those of its allele 315 (P' = .00002). Overall, these observations suggest that a 5–10-cM MALD map with marker-enriched differences between the founding populations used for MALD analysis will be sufficient for genome scans.

Recent theoretical analyses pointing out the power of the transmission/disequilibrium test (TDT [Spielman et al. 1993]) (McKeigue 1997; Rabinowitz 1997; Kaplan et al. 1998; McKeigue 1998; Zheng and Elston 1999) for MALD disease-gene discovery suggest that the TDT may yield larger regions of ALD or may require fewer samples. The present study is defined by the FY genotypes, similar to the situation in a case-control study, and it used (*a*) fewer patients who had a "disease" δ larger than that assumed in previous simulations (1.0 vs. .3) and (*b*) similar marker-allele-frequency differences (δ = .3, vs. an average composite δ of .27 [Briscoe et al. 1994; Stephens et al. 1994]). Although all of the assumptions of complex disease-gene mapping are not met, the analysis of FY suggests the utility of a case-control design in MALD. The case-control design offers advantages over the TDT, when age at onset is late and parents are frequently unavailable for study, but this design must be balanced against possible false-positive findings from population stratification (Lander and Schork 1994; Altshuler et al. 1998). Both the remarkably strong degree of ALD around FY and the success of multiple analyses to detect it empirically affirm MALD as an approach to the finding of genes associated with disease in African Americans and other recently admixed populations, while suggesting that a 5–10-cM marker density is appropriate for MALD genome scans.

Acknowledgments

This study made extensive use of the computational resources of the Advanced Biomedical Computing Center of the National Cancer Institute (Frederick, MD). We thank R. Boaze, K. Byrd, S. Cervario, D. Lomb, B. Henry, G. K. Pei, and S. Shrestha for technical assistance and Drs. G. Nelson and M. Dean for their helpful discussions. We are grateful to G. Washburn for assistance in designing the multiplex STR primer sets and to Z. Q. Chen for performing radiation-hybrid–mapping experiments. We thank J. Long for supplying source code and G. Smythers for implementation of the three-locus EM program. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. government. This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under contract NO1-CO-56000.

Electronic-Database Information

URLs for data in this article are as follows:

GeneMap'98, http://www.ncbi.nlm.nih.gov/genemap98 (for radiation-hybrid locations for FY and the linked coding genes SPTA1 and AT3)

Généthon, ftp://ftp.genethon.fr/pub/Gmap/Nature-1995/data (for map locations of the STR markers)

Laboratory of Genomic Diversity, The, http://lgd.nci.nih.gov (for additional allele-frequency data for each locus and for D')

US Census Bureau, http://www.census.gov (for population-size estimates)

References

Altshuler D, Kruglyak L, Lander E (1998) Genetic polymorphisms and disease. N Engl J Med 338:1626

Bock SC, Levitan DJ (1983) Characterization of an unusual DNA length polymorphism 5′ to the human antithrombin III gene. Nucleic Acids Res 11:8569–8582

Bodmer WF (1986) Human genetics: the molecular challenge. Cold Spring Harb Symp Quant Biol 51:1–13

Briscoe D, Stephens JC, O'Brien SJ (1994) Linkage disequilibrium in admixed populations: applications in gene mapping. J Hered 85:59–63

Chakraborty R, Kamboh MI, Ferrell RE (1991) "Unique" alleles in admixed populations: a strategy for determining "hereditary" population differences of disease frequencies. Ethn Dis 1:245–256

Chakraborty R, Kamboh MI, Nwankwo M, Ferrell RE (1992) Caucasian genes in American blacks: new data. Am J Hum Genet 50:145–155

Chakraborty R, Weiss KM (1988) Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. Proc Natl Acad Sci USA 85: 9119–9123

Chaudhuri A, Polyakova J, Zbrzezna V, Pogo AO (1995) The coding sequence of Duffy blood group gene in humans and simians: restriction fragment length polymorphism, antibody and malarial parasite specificities, and expression in nonerythroid tissues in Duffy-negative individuals. Blood 85:615–621

Deloukas P, Schuler GD, Gyapay G, Beasley EM, Soderlund C, Rodriguez-Tome P, Hui L, et al (1998) A physical map of 30,000 human genes. Science 282:744–746

Destro-Bisol G, Maviglia R, Caglia A, Boschi I, Spedini G, Pascali V, Clark A, et al (1999) Estimating European admixture in African Americans by using microsatellites and a microsatellite haplotype (CD4/Alu). Hum Genet 104: 149–157

Dib C, Faure S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, et al (1996) A comprehensive genetic map of the human genome based on 5,264 microsatellites. Nature 380:152–154

Ginzburg HM, Fleming PL, Miller KD (1988) Selected public health observations derived from Multicenter AIDS Cohort Study. J Acquir Immune Defic Syndr 1:2–7

Goedert JJ, Biggar RJ, Melbye M, Mann DL, Wilson S, Gail MH, Grossman RJ, et al (1987) Effect of T4 count and cofactors on the incidence of AIDS in homosexual men infected with human immunodeficiency virus. JAMA 257: 331–334

Goedert JJ, Kessler CM, Aledort LM, Biggar RJ, Andes WA, White GCD, Drummond JE, et al (1989) A prospective study of human immunodeficiency virus type 1 infection and the development of AIDS in subjects with hemophilia. N Engl J Med 321:1141–1148

Hadley TJ, Peiper SC (1997) From malaria to chemokine receptor: the emerging physiologic role of the Duffy blood group antigen. Blood 89:3077–3091

Hentati A, Hu P, Asgharzadeh S, Siddique T (1992) Dinucleotide repeat polymorphism at the human erythroid alpha spectrin (SPTA1) locus. Hum Mol Genet 1:218

Holm S (1979) A simple sequentially rejective multiple test procedure. Scand J Stat 6:65–70

Horuk R, Chitnis CE, Darbonne WC, Colby TJ, Rybicki A, Hadley TJ, Miller LH (1993) A receptor for the malarial parasite *Plasmodium vivax:* the erythrocyte chemokine receptor. Science 261:1182–1184

Hudson TJ, Stein LD, Gerety SS, Ma J, Castle AB, Silva J, Slonim DK, et al (1995) An STS-based map of the human genome. Science 270:1945–1954

Huttley GA, Smith MW, Carrington M, O'Brien SJ (1999) A scan for linkage disequilibrium across the human genome. Genetics 152:1711–1722

Iwamoto S, Li J, Sugimoto N, Okuda H, Kajii E (1996) Characterization of the Duffy gene promoter: evidence for tissue-specific abolishment of expression in Fy(a−b−)of black individuals. Biochem Biophys Res Commun 222:852–859

Jaffe HW, Darrow WW, Echenberg DF, O'Malley PM, Getchell JP, Kalyanaraman VS, Byers RH, et al (1985) The acquired immunodeficiency syndrome in a cohort of homosexual men: a six-year follow-up study. Ann Intern Med 103: 210–214

Kaplan NL, Martin ER, Morris RW, Weir BS (1998) Marker selection for the transmission/disequilibrium test, in recently admixed populations. Am J Hum Genet 62:703–712

Laan M, Pääbo S (1997) Demographic history and linkage disequilibrium in human populations. Nat Genet 17:435–438

Lander ES, Schork NJ (1994) Genetic dissection of complex traits. Science 265:2037–2048

Lewontin RC (1964) The interaction of selection and linkage. I. General considerations: heterotic models. Genetics 49: 49–67

Lincoln SE, Daly MJ, Lander ES (1991) PRIMER: a computer program for automatically selecting PCR primers. MIT Center for Genome Research, Cambridge, MA

Long JC (1991) The genetic structure of admixed populations. Genetics 127:417–428

Long JC, Williams RC, Urbanek M (1995) An E-M algorithm and testing strategy for multiple-locus haplotypes. Am J Hum Genet 56:799–810

Loveland KA, Stehbens J, Contant C, Bordeaux JD, Sirois P, Bell TS, Hill S, et al (1994) Hemophilia growth and development study: baseline neurodevelopmental findings. J Pediatr Psychol 19:223–239

Ludbrook J (1998) Multiple comparison procedures updated. Clin Exp Pharmacol Physiol 25:1032–1037

McKeigue PM (1997) Mapping genes underlying ethnic differences in disease risk by linkage disequilibrium in recently admixed populations. Am J Hum Genet 60:188–196

——— (1998) Mapping genes that underlie ethnic differences in disease risk: methods for detecting linkage in admixed populations, by conditioning on parental admixture. Am J Hum Genet 63:241–251

Miller LH, Mason SJ, Clyde DF, McGinniss MH (1976) The resistance factor to *Plasmodium vivax* in blacks: the Duffy-blood-group genotype, FyFy. N Engl J Med 295:302–304

Parra EJ, Marcini A, Akey J, Martinson J, Batzer MA, Cooper R, Forrester T, et al (1998) Estimating African American admixture proportions by use of population-specific alleles. Am J Hum Genet 63:1839–1851

Peterson AC, Di Rienzo A, Lehesjoki AE, de la Chapelle A,

Slatkin M, Freimer NB (1995) The distribution of linkage disequilibrium over anonymous genome regions. Hum Mol Genet 4:887–894

Rabinowitz D (1997) A transmission disequilibrium test for quantitative trait loci. Hum Hered 47:342–350

Roychoudhury AK, Nei M (1988) Human polymorphic genes: world distribution. Oxford University Press, New York

Shriver MD, Smith MW, Jin L, Marcini A, Akey JM, Deka R, Ferrell RE (1997) Ethnic-affiliation estimation by use of population-specific DNA markers. Am J Hum Genet 60: 957–964

Sidak Z (1967) Rectangular confidence regions for the means of multivariate normal distributions. J Am Stat Assoc 62: 626–633

Smith MW, Dean M, Carrington M, Winkler C, Huttley GA, Lomb DA, Goedert JJ, et al (1997) Contrasting genetic influence of CCR2 and CCR5 variants on HIV-1 infection and disease progression. Science 277:959–965

Sokal RR, Rohlf FJ (1995) Biometry: the principles and practice of statistics in biological research. WH Freeman, New York

Spielman RS, McGinnis RE, Ewens WJ (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes melltius (IDDM). Am J Hum Genet 52:506–516

Stephens JC, Briscoe D, O'Brien SJ (1994) Mapping by admixture linkage disequilibrium in human populations: limits and guidelines. Am J Hum Genet 55:809–824

Stephens JC, Smith MW, Shin HD, O'Brien SJ (1999) Tracking linkage disequilibrium in admixed populations with MALD using microsatellite loci. In: Goldstein DB, Schlötterer C (eds) Microsatellites evolution and applications. Oxford University Press, Oxford, pp 211–224

Tournamille C, Colin Y, Cartron JP, Le Van Kim C (1995) Disruption of a GATA motif in the Duffy gene promoter abolishes erythroid gene expression in Duffy-negative individuals. Nat Genet 10:224–228

Vlahov D, Anthony JC, Munoz A, Margolick J, Nelson KE, Celentano DD, Solomon L, et al (1991) The ALIVE study, a longitudinal study of HIV-1 infection in intravenous drug users: description of methods and characteristics of participants. NIDA Res Monogr 109:75–100

Weir BS (1996) Genetic data analysis. II. Methods for discrete population genetic data. Sinauer Associates, Sunderland, MA

Wu S, Seino S, Bell GI (1989) Human antithrombin II (AT3) gene length polymorphism revealed by the polymerase chain reaction. Nucleic Acids Res 17:6433

Zheng C, Elston RC (1999) Multipoint linkage disequilibrium mapping with particular reference to the African-American population. Genet Epidemiol 17:79–101