

M. Molokhia · C. Hoggart · A. L. Patrick · M. Shriver
E. Parra · J. Ye · A. J. Silman · P. M. McKeigue

Relation of risk of systemic lupus erythematosus to west African admixture in a Caribbean population

Received: 29 August 2002 / Accepted: 8 November 2002 / Published online: 24 January 2003

© Springer-Verlag 2003

Abstract Risk of systemic lupus erythematosus (SLE) is higher in people of west African descent than in Europeans. The objective of this study was to distinguish between genetic and environmental explanations for this ethnic difference by examining the relationship of disease risk to individual admixture (defined as the proportion of the genome that is of west African ancestry); 124 cases of SLE and 219 matched controls resident in Trinidad were studied. Analysis of admixture was restricted to 52 cases and 107 controls who reported no Indian or Chinese ancestry. These individuals were typed with a panel of 26 single-nucleotide polymorphisms and five insertion/deletion polymorphisms chosen to have large allele frequency differentials between west African, European and Native American populations. A Bayesian model for population admixture, individual admixture and locus ancestry was fitted by Markov chain simulation. Mean west African admixture (M) was 0.81 in cases and 0.74 in controls ($P=0.01$). The risk ratio for SLE associated with unit

change in M was estimated as 32.5 with a 95% confidence interval (CI) of 2.0–518. Adjustment for measures of socioeconomic status (household amenities in childhood and years of education) altered this risk ratio only slightly (adjusted risk ratio: 28.4, 95% CI 1.7–485). These results support an additive genetic model for the ethnic difference in risk of SLE between west Africans and Europeans, rather than an environmental explanation or an “overdominant” model in which risk is higher in heterozygous than in homozygous individuals. This conclusion lays a basis for localizing the genes underlying this ethnic difference in risk of SLE by admixture mapping.

Introduction

Systemic lupus erythematosus (SLE) is an auto-immune disease involving multiple organ systems and affects women approximately ten times as frequently as men. Prevalence in African-Americans and in people of west African descent who have migrated from the Caribbean to Europe has been reported to be six to eight times higher than in people of European descent living in the same countries (Fessel 1974; Hopkinson et al. 1994; Johnson et al. 1995; Nossent 1992). A recent study in England has shown that risk of SLE is high both in first-generation migrants from west Africa and in second-generation people of African Caribbean origin (Molokhia et al. 2001). The consistency with which increased risk of SLE has been reported in people of west African descent settled in different regions of the world, and the similarity of risk in recent migrants to risk in people settled overseas for many generations are consistent with a genetic explanation for this ethnic difference in risk. As SLE has been reported to be rare in west Africa itself (Greenwood 1968), environmental explanations cannot be excluded. There have been reports that risk of SLE is also high in people of Chinese and Indian descent (Catalano and Hoffmeier 1989; Frank 1979; Serdula and Rhoads 1979; Wang et al. 1997; Samanta et al. 1992).

Two main types of genetic explanation for the ethnic difference in risk can be hypothesized. For simplicity, we

Electronic database information: URLs for the data in this article are as follows:

dbSNP database, <http://www.ncbi.nlm.nih.gov/SNP>
WinBUGS version 1.3, <http://www.mrc-bsu.cam.ac.uk/bugs/winbugs>
ADMIXMAP, <http://www.lshtm.ac.uk/eph/eu/geneticepi/ADMIXMAP>

M. Molokhia (✉) · C. Hoggart · P. M. McKeigue
Epidemiology Unit,
London School of Hygiene and Tropical Medicine,
Keppel Street, London, WC1E 7HT, UK
Tel.: +44-207-9272633, Fax: +44-207-5806897,
e-mail: mariam.molokhia@lshtm.ac.uk

A. L. Patrick
Kavanagh St Clinic, Port of Spain, Trinidad

M. Shriver · E. Parra · J. Ye
Department of Anthropology, Penn State University,
Philadelphia, Pa., US

A. J. Silman
ARC Epidemiology Research Unit,
Oxford Street, Manchester, UK

outline these explanations based on a two-allele single locus model, although the same general principles would apply to models in which more than one locus or more than two alleles contribute to the ethnic difference in risk. The simplest type of genetic explanation is a model in which risk increases with the number of copies of the high-risk allele at the disease locus, and the frequency of the high-risk allele is higher in west Africans than in Europeans. We denote this as an “additive” model, in that additive effects would account for most of the genetic variance in risk. An alternative type of genetic explanation for the increased susceptibility to SLE in African-American and African Caribbean populations compared with European is a model in which risk is higher in individuals who are heterozygous at the disease locus than in individuals with either of the two homozygous genotypes, and different alleles are fixed in west African and European populations. Under this model, risk would be higher in individuals of mixed European and west African descent than in either of the two ancestral populations, and highest of all in individuals with one African and one European parent, because all such individuals would be heterozygous at the disease locus. We denote this as an “overdominant” model, in that dominance variance would account for most of the genetic variance in risk. Such a model has experimental support: the F1 offspring of a cross between two inbred strains of mice are more susceptible to lupus than either of the two parental strains (Theofilopoulos and Dixon 1985; Vyse and Kotzin 1998).

The most direct way to distinguish between environmental and genetic models for an ethnic difference in disease risk is to study the relationship of disease risk to individual admixture (Chakraborty and Weiss 1986). We define individual admixture (for the purposes of this study) as the proportion of the individual’s genome that is of African ancestry. From each of the three hypotheses outlined above to explain the observed ethnic difference in risk of SLE, environmental, additive genetic model, or overdominant genetic model, we can make a specific prediction. If the ethnic difference in risk has an environmental explanation, we predict no relationship between disease risk and individual admixture, after controlling for likely confounders such as socioeconomic status. Under an additive genetic model, we predict a linear relationship of disease risk to individual admixture, which would persist after controlling for any environmental confounders. Under an overdominant genetic model, we predict an inverse U-shaped relationship of disease risk to admixture and a positive relationship of disease risk to heterozygosity of ancestry (the proportion of an individual’s genome at which there is one gene copy of African ancestry and one of non-African ancestry). The objective of this study has been to distinguish between the three hypotheses (environmental risk factor, additive genetic model, or overdominant genetic model) for the increased susceptibility to SLE in African-American and African Caribbean people compared with people of European descent by studying the relationship of disease risk to admixture in the island of Trinidad.

Subjects and methods

Population history

The original inhabitants of Trinidad were Carib and Sarawak Native Americans. Their descendants are now almost entirely admixed with Europeans and west Africans. Although the first landing by Europeans was in 1498, the first Spanish settlement was not established until 1592 and large-scale settlement did not begin until 1783, when a Spanish decree encouraged Catholic (predominantly French) settlers to bring slaves and establish plantations. During the Napoleonic wars, the island was conquered by the British who, after ending African slavery in 1838, brought large numbers of labourers from India to work on the plantations. Over the last century, there has been further migration from China, from the eastern Mediterranean, and from neighbouring Caribbean islands. Until the last two generations, there has been very little admixture of Indian and Chinese people with the African population. In Trinidad, about 40% of the population classify their ethnic origin as African, 40% as Indian and 15% as mixed (Central Statistical Office Trinidad and Tobago 1992).

Study design

As SLE is a rare disease even in high-risk African Caribbean populations (prevalence about 2 per 1000 adult women), a case-control study is the most efficient design to estimate the relationship of risk to individual admixture. For this study, a target sample size of 120 cases and 240 controls was set, giving 90% power to detect at 5% significance a risk ratio of 2.5 between any two tertiles of admixture. As the objective was to study possible explanations for the west African/European difference in risk of SLE, rather than to study the effects of admixture with other high risk groups such as Indians and Chinese, the inclusion criteria for cases and controls were designed to restrict the sample to individuals without Indian or Chinese ancestry. Cases and controls were eligible to be included if they were resident in northern Trinidad (excluding the southern part of the island where Indians are in the majority) and they had Christian (rather than Hindu, Muslim or Chinese) first names. Genealogical data was also collected by questionnaire, obtaining parental and grandparental country of origin and reported ethnicity.

Case ascertainment and sampling of controls

We attempted to ascertain all surviving cases of SLE among adults resident in northern Trinidad. All physicians specializing in rheumatology, nephrology and dermatology at the two main public hospitals in northern Trinidad were approached and asked to provide a list of all patients with SLE seen at their out-patient clinics. A register of cases seen since 1992 at the main dermatology clinic was available. This was supplemented with a systematic search of out-patient records at the two hospitals, hospital laboratory test results positive for auto-antibodies (anti-nuclear or anti-double-stranded DNA antibody titre >1:256) and histological reports of skin biopsy examination consistent with SLE. Cases were also identified through the Lupus Society of Trinidad and Tobago, which circulated its members with information about the study; 90% of patients ascertained through the Society were also identified through one of the two main public hospitals.

Cases of SLE identified by these means were telephoned where possible and invited to participate in the study. Where telephones were not available, invitations to take part in the study were delivered to the patient’s home address and these addresses were visited by the field team. For each case, randomly chosen households in the same neighbourhood were sampled by the field team to obtain (where possible) two controls, matched with the case for sex and for 20-year age group. Strict matching for age is unnecessary where (as in this study) any residual confounding by age can be

adjusted for in the analysis. The exclusion criteria (based on first names) were the same for controls as for cases.

Collection of samples and clinical data

Cases and controls were interviewed at home or in the project office by using a questionnaire that included demographic items, socioeconomic indices in childhood and adult life, ancestry of each grandparent, birth order, medical history and reproductive history. As a proxy measure of socioeconomic status and possible exposure to infection in childhood, items were included regarding household amenities (running water, sanitation) when the participant was aged 12 years. The case definition of SLE was based on American Rheumatism Association (ARA) criteria (Tan et al. 1982), applied to medical records (which were available for more than 90% of cases), and to the medical history given by the patient in the questionnaire. Informed consent for blood sampling and the use of the sample for genetic studies including estimation of admixture was obtained from each participant.

Initial case ascertainment identified 264 possible cases of SLE. Of these, 72 (27%) were excluded either on the basis of their names or because their medical history did not meet ARA criteria for the diagnosis of SLE. Of the remaining 192 individuals, 54 had incomplete addresses or were not resident in northern Trinidad,

four were too ill to be interviewed, eight were aged less than 18 years and two refused to participate. For 80% (99/124) of cases, two matched controls were obtained: the response rate from those invited to participate as controls was 70%. The total sample consisted of 124 cases and 219 controls aged over 20 years who completed the questionnaire. Blood samples were obtained from 122 cases and 219 controls and DNA was successfully extracted from 93% (317/341) of these.

Markers used for estimation of individual admixture

For estimation of two-way admixture, the information content for ancestry of a marker locus can be measured by the f -value or Wahlund variance (McKeigue 1998) as defined in the Appendix; f -values lie between 0 (no information about ancestry) and 1 (fully informative for ancestry). Calculations based on a large-sample variance of the maximum-likelihood estimator show that to estimate two-way individual admixture with a standard error of no more than 0.1, approximately 40 markers with average f -values of 0.40 are required. The markers used in this study were identified from various sources, including published reports of polymorphisms that showed extreme variation of allele frequencies between human populations, and screening the 1494 single-nucleotide polymorphisms (SNPs) included on the HuSNP chip (Affy-

Table 1 Marker loci

Locus (by position)	Chromo- somal band	Distance from last marker (cM)	Allele frequencies			Marker information content f			
			Poly- morphism	Allele	West African	European	Native American	$f_{\text{Afr-Eur}}$	$f_{\text{Eur-NAm}}$
MID-187	1p32		I/D	I	0.759	0.388	0.318	0.14	0.01
FY-null	1q23.2		T/C	T	0.001	0.998	0.992	0.99	0.00
AT3-Indel	1q25.1	22	I/D	I	0.858	0.282	0.08	0.34	0.07
WI-11392	1q42.2				0.878	0.433	0.622	0.22	0.04
WI-16857	2p16.1		G/A	G	0.751	0.215	0.203	0.29	0.00
WI-11153	3p12.1		G/C	G	0.785	0.133	0.761	0.43	0.40
GC	4q13.3		T/G, C/A	1F (T,C) 1S (G,C)	0.853 0.069	0.156 0.607	0.323 0.547	0.49	0.05
SGC30610	5q11.2		T/A	T	0.401	0.255	0.682	0.02	0.18
SGC30055	5q22.1		A/G	A	0.054	0.511	0.729	0.26	0.05
WI-17163	5q33.1	40	G/A	G	0.054	0.175	0.695	0.04	0.28
WI-9231	7p22.3		C/G	C	0.129	0.147	0.517	0.00	0.15
WI-4019	7q21.3		A/G	A	0.43	0.306	0.603	0.02	0.09
CYP3A4	7q22.1	2			0.198	0.958	0.953	0.59	0.00
LPL	8p21.3		T/C	T	0.971	0.492	0.45	0.29	0.00
WI-11909	9q21.31		G/A	G	0.805	0.881	0.218	0.01	0.44
TYR-192	11q14.3		A/C	A	0.005	0.449	0.06	0.28	0.20
DRD2-Bcl I	11q23.2	20	C/T	C	0.063	0.144	0.628	0.02	0.25
DRD2-Taq D	11q23.2	0	T/C	T	0.118	0.67	0.089	0.32	0.36
APOA1-Alu	11q23.3	6	Alu+/-	Alu+	0.42	0.925	0.975	0.29	0.01
D11S429	11q13.3		T/C	T	0.087	0.516	0.14	0.22	0.16
GNB3_C825T	12p13.31		T/C	T	0.795	0.332	0.365	0.22	0.00
RB1	13q14.2		G/A	G	0.926	0.315	0.187	0.40	0.02
OCA2	15q12		A/G	A	0.115	0.746	0.484	0.41	0.07
WI-14319	15q14	15	C/T	C	0.386	0.201	0.696	0.04	0.25
CYP19-E2	15q21.2	20	T/G	T	0.34	0.296	0.719	0.00	0.18
PV92-Alu	16q23.3		Alu+/-	Alu+	0.225	0.152	0.776	0.01	0.39
MC1R-314	16q24.3		G/A	G	0.59	0.163	0.072	0.19	0.02
WI-7423	17p13.1		T/C	T	0	0.476	0.074	0.31	0.20
WI-14867	17p13.2	10	C/T	C	0.024	0.472	0.427	0.27	0.00
CKM	19q13.2		A/G	A	0.164	0.313	0.858	0.03	0.31
MID-154	20q11.23		I/D	I	0.806	0.362	0.439	0.20	0.01

metrix, Santa Clara, USA). Of the 31 markers available for this study, 21 were informative ($f > 0.05$) for west African/European ancestry (mean f -value: 0.34) and 18 were informative for Native American/European ancestry (mean f -value: 0.22). Details of these markers including allele frequencies in all parental populations, DNA sequences, exact positions of SNPs and the polymerase chain reaction (PCR) primers used are available from the dbSNP database (<http://www.ncbi.nlm.nih.gov/SNP>) under the handle PSU-ANTH. Allele frequencies, estimated from contemporary west African, European and Native American population samples are given in Table 1. All markers, except GC (a combination of two SNPs), are biallelic. Three of the markers are short insertion/deletion polymorphisms, two are Alu insertions and all others are SNPs. Nine of the 31 markers are linked to at least one other marker in the set and two are in the same gene (*DRD2*). For these two markers, ancestry-specific haplotype frequencies were estimated by the EM algorithm from genotypes of unrelated individuals.

SNP markers were scored by a melting-curve assay (McSNP), in which the target sequence containing the SNP was amplified by PCR by using a mismatched primer where necessary to create an artificial restriction site polymorphism. PCR products were digested with a restriction enzyme and the resulting restriction fragment length polymorphisms were scored by their melting curves in a Hybaid DASH machine (Akey et al. 2001). Scoring the triallelic GC marker required separate digests with *HaeIII* and *SlyI* and agarose gel electrophoresis in lieu of McSNP. PCR products of the insertion/deletion polymorphisms were scored without digestion by a restriction enzyme. Three of the insertion/deletion polymorphisms (*APOA1*, *AT3*, *PV92*) and one of the SNPs (*RB1*) were scored on gels and the others by McSNP.

Statistical analysis

As allele frequency data were not available for Indians or Chinese and markers informative for Indian or Chinese ancestry are also not available, analyses of admixture were restricted to the 159 individuals who reported no Indian or Chinese ancestry in their grandparents and for whom DNA was available. Admixture was modelled with an extension of the Bayesian statistical model described previously for studies of two-way admixture in African-American populations (McKeigue et al. 2000). In comparison with classical approaches for modelling admixture (Chakraborty 1975), this approach makes it possible to fit a hierarchical model for population and individual admixture, to deal adequately with nuisance variables and to allow for linkage between marker loci. Because the Bayesian estimates of individual admixture are based on a hierarchical model, they cannot be directly compared with estimates based on maximizing the likelihood independently for each individual. With a large sample size (more than 100 individuals typed at more than 20 unlinked loci) and non-informative priors (contributing less information than a single observation), the posterior mode (or mean, where the posterior distribution is symmetric) of any variable will closely approximate a maximum likelihood estimate.

The model specifies, as random variables, the distribution of parental admixture proportions in the population, the admixture of each individual's parents and the ancestry of the maternally derived and paternally derived gene copies at each locus. The probabilities of the observed genotypes, conditional on locus ancestry, are calculated from the ancestry-specific allele frequencies (or, at the two loci in the *DRD2* gene, from the ancestry-specific haplotype frequencies). Each individual's admixture proportions are calculated as the mean of the parental admixture proportions. Heterozygosity of ancestry is defined, for an individual, as the probability that a locus chosen at random will have one gene copy of African ancestry and one of non-African ancestry and is calculated as $M(1-M') + M'(1-M)$, where M and M' are the proportions of the paternal and maternal genomes that are of African ancestry. This probability can vary from 0 (in an unadmixed individual) to 1 (in an individual with one African and one non-African parent).

To allow for linkage between marker loci, it is necessary to model the stochastic variation of ancestry as a function of map distance on each set of chromosomes inherited from one parent. This

variation is modelled as a Markov process in which transitions to new states of ancestry (west African, European or Native American) occur as independent Poisson processes. The intensities of these three Poisson processes (α , β , γ) and the admixture proportions of the parent (M_1 , M_2 , M_3) are related to each other by the equations $M_1 = \alpha/s$, $M_2 = \beta/s$, $M_3 = \gamma/s$, where $s = \alpha + \beta + \gamma$. For the sum-of-intensities parameter s , expressed in Morgan⁻¹, a non-informative gamma prior distribution was assigned.

Two alternative models (discrete and continuous) for the population distribution of parental admixture proportions were fitted. In the discrete model, parental admixture was modelled as a categorical variable with 153 categories (corresponding to all possible three-way splits of fractions of 16) and assigned a non-informative Dirichlet prior distribution for the vector of category probabilities. In the continuous model, the population distribution of parental admixture was modelled as a Dirichlet distribution that was assigned a non-informative prior. The continuous model requires fewer parameters to specify it but the assumption that the distribution of parental admixture follows a Dirichlet distribution (which is smooth and unimodal) may be incorrect in a population where there has been recent gene flow from unadmixed populations. The correlation between estimates of individual admixture obtained under these two models was 0.985 and estimates of associations with SLE were similar with either model. Results presented in this paper are based on the discrete model for admixture.

The posterior distribution of all random variables, conditional on the observed marker genotypes, was generated by Markov chain simulation, by using either a general-purpose program for Bayesian modelling (WinBUGS version 1.3, <http://www.mrc-bsu.cam.ac.uk/bugs/winbugs>) or a specially-written program for modelling admixture (ADMIXMAP, <http://www.lshtm.ac.uk/eph/eu/geneticipi/ADMIXMAP>). Each run consisted of at least 20,000 iterations, and multiple runs with different starting values were used to check for adequate mixing of the sampler.

Score tests of various null hypotheses were obtained from the posterior distribution, as described previously (McKeigue et al. 2000). The method estimates the observed information (curvature of the log-likelihood function) as the complete information (expectation of the information over the posterior distribution) minus the missing information (variance of the score over the posterior distribution). The ratio of observed to complete information is the proportion of information extracted. Two types of score test were constructed for this analysis.

(i) a test for association between SLE and individual admixture. This tests the null hypothesis that $\beta = 0$, where β is the regression coefficient for the effect of west African admixture (M) on risk of SLE in a logistic regression model. The score for the i th individual is $M_{(d_i - p_i)}$ and the information is $M_i^2 p_{(1-p_i)}$, where M_i is the admixture of the i th individual, d_i is an indicator variable (0 in controls, 1 in cases) and p_i is the probability of being a case given $\beta = 0$. The score is summed over all cases and controls.

(ii) an affected-only test for linkage with the ethnic difference in risk of SLE, as described previously (McKeigue 1998). The score for each affected individual is the observed proportion of gene copies that have African ancestry minus the expected proportion given the admixture of that individual's parents. The score is summed over all affected individuals.

To estimate the slope of the relationship of SLE risk to individual admixture, two different approaches were used.

(i) extending the Bayesian model to include a logistic regression of SLE risk on individual admixture, specifying non-informative prior distributions for the regression parameters. This approach correctly takes into account the uncertainty in estimation of individual admixture but is laborious when several different models (with and without covariates) have to be examined.

(ii) using Bayesian estimates of individual admixture in a classical logistic regression model, fitted by using the Stata 7 package (StataCorp 2001).

Although the original design was based on matching cases and controls for age, sex and neighbourhood, after exclusion of individuals with Indian or Chinese ancestry, there were many strata (neighbourhoods) with no cases or no controls. Unconditional logistic regression models ignoring the matching in the design were

Table 2 Summary statistics for SLE cases and controls (NS not significant)

Factor	Cases (n=124)	Controls (n=219)	Significance
Mean age (range)	43.8 (21–70)	44.0 (20–77)	NS
Female	92% (114/124)	94% (205/219)	NS
Indian ancestry in grandparents	45% (56/124)	40% (88/219)	NS
Chinese ancestry in grandparents	17% (21/124)	12% (27/219)	NS
Mean [SD] birth order	4.0 [2.9]	3.5 [2.6]	NS
Household with water supply at age 12	62% (77/124)	74% (162/219)	<i>P</i> =0.02
Completed education after age 16 years	62% (136/219)	65% (81/124)	NS
Ever smoker	22 (33%)	44 (67%)	NS
Women only:			
Ever used oral contraceptives	55% (62/113)	51% (100/198)	NS
Ever pregnant	79% (89/113)	92% (188/204)	<i>P</i> =0.001
Excluding those with Indian or Chinese ancestry: n=52 n=107			
Mean African admixture	0.81 (0.09)	0.74 (0.18)	<i>P</i> =0.01
Mean European admixture	0.11 (0.07)	0.16 (0.14)	<i>P</i> =0.02
Mean Native American admixture	0.08 (0.03)	0.10 (0.05)	NS

therefore fitted. Ignoring matching in a case-control design may bias estimates of association towards the null value but will not yield false-positive results (Breslow and Day 1980). When neighbourhoods in the same district were combined into larger strata to eliminate zero cell entries and conditional logistic regression models (which take account of matching) were fitted, the estimates of association were similar (with larger standard errors) to those obtained when matching was ignored.

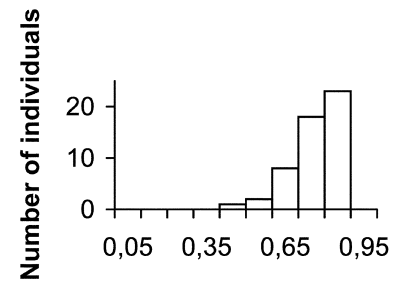
Results

Comparison of cases and controls

Table 2 compares all 124 cases and 219 controls. The proportion who had grown up in households with running water (a proxy measure of hygiene and socioeconomic status in childhood) was significantly lower in cases (62%) than in controls (73%). Among women, the proportion who reported that they had been pregnant was lower in cases than in controls (79% versus 92%). The proportion who reported their religious affiliation as Catholic was higher in cases than in controls (48% versus 36%). The ratio of Catholics to Protestants in the control group was similar to that estimated for the general population of Trinidad in the 1991 census. A follow-up enquiry of a subsample of cases and controls revealed that about 20% had converted from Catholic to other religious denominations during adult life. Religious affiliation was not therefore used in the analysis as a proxy marker of demographic background.

Analyses of admixture were restricted to the 159 individuals who reported no Indian or Chinese ancestry in their grandparents (Fig. 1). The correlation between marker-based estimates of African admixture and estimates based on reported ancestry of grandparents was 0.70 (*P*<0.001). Only the marker-based estimates of admixture were used in subsequent analyses. The mean crossover rate between African and non-African ancestry was estimated as 2.4 per 100 cM, with a 95% confidence interval (CI) of 0.7–7.4. The mean (\pm SD) proportions of west African ad-

(i) Cases



(ii) Controls

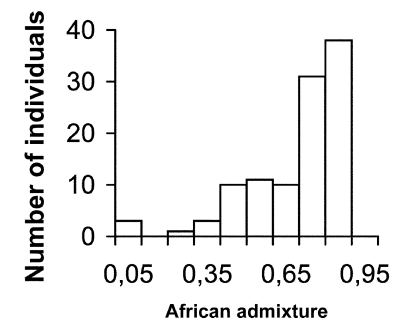


Fig. 1 Distribution of individual west African admixture estimates in cases (*top*) and controls (*bottom*)

mixture among cases and controls were 0.81 ± 0.09 and 0.74 ± 0.18 (*P*=0.01). Mean proportions of European admixture differed in the opposite direction (0.11 ± 0.07 in cases and 0.16 ± 0.14 in controls, *P*=0.02). The mean proportions of Native American admixture did not differ significantly between cases and controls (0.08 ± 0.03 and 0.10 ± 0.05 respectively, *P*=0.5). Among controls, several measures of socioeconomic status were inversely related to west African admixture. Mean west African admixture was lower in those who had grown up in households with running water than in those who had grown up in households lacking this amenity (0.72 vs 0.80, *P*=0.05), in those whose full-time education had continued after the age of

Table 3 Logistic regression analyses with case/control status as dependent variable, excluding those with Indian or Chinese ancestry

Factor	Odds ratio associated with unit change in African admixture	95% CI
Univariate analyses: risk factors examined one at a time, with age and sex as only other variables in the model		
African admixture	32.5	2.0–517
Lived at age 12 in household lacking water supply	2.2	1.1–4.4
Age completed full-time education ^a	1.5	0.9–2.4
Multivariate analysis		
African admixture	28.4	1.7–485
Lived at age 12 in household lacking water supply	2.7	1.2–5.8
Years of education	2.0	1.1–3.3

^aAge completed full-time education coded as: 1 less than 16 years, 2 16–18 years, 3 more than 18 years

16 years than in those who had completed their education by this age (0.72 vs 0.78, $P=0.025$) and in those currently living in households with access to a car compared with those living in households without a car (0.72 vs 0.78, $P=0.03$).

Association of SLE with west African admixture

A score test for association of SLE with west African admixture, with age and sex as the only covariates, was significant at $P=0.01$. In Bayesian logistic regression models adjusted for age and sex, the risk ratio associated with unit change (from 0 to 1) in the proportion of west African admixture, adjusted for age and sex, was estimated as 19.1 (95% posterior interval 1.02–256) and the risk ratio associated with unit change in heterozygosity of ancestry was estimated as 0.13 (95% posterior interval 0.01–1.25).

To examine the effect of adjusting for other covariates, classical logistic regression models were fitted by using the marker-based estimates of individual admixture from the Bayesian analysis (Table 3). In an unconditional logistic regression with SLE as a dependent variable, the age- and sex-adjusted risk ratio associated with unit change in African admixture was 32.5 (95% CI 2.0–518). Adjustment for measures of socioeconomic status before disease onset (access to running water in childhood and years of full-time education) altered this risk ratio only slightly (adjusted risk ratio: 28.4, 95% CI 1.7–485). In this multivariate analysis, lack of access to running water in childhood and a longer duration of full-time education were both significantly and independently associated with increased risk of SLE.

Testing marker loci for linkage with SLE

Each marker locus was tested for linkage to genes underlying the increased susceptibility of west Africans to SLE by using the affected-only score test described previously. This is equivalent to testing whether the observed proportion of gene copies that have African ancestry at the locus under study is higher than the expected proportion given the admixture of each individual's parents. In this analy-

sis, it is unnecessary to control for covariates, as each affected individual acts as their own control. None of the loci showed evidence of linkage to genes underlying the ethnic difference in SLE risk (Table 4). At most loci, the proportion of information extracted was low and, for two of the marker loci (WI-9231 and WI-11909), the test statistic could not be calculated because the observed information was zero or negative (implying that the likelihood function is not log-concave at the null value of the parameter). These two markers are informative for Native American versus European ancestry but have very low information content for African versus non-African ancestry.

Discussion

The theoretical basis for studying the relationship of disease risk to individual admixture was set out some years ago (Chakraborty and Weiss 1986). Under additive, recessive or dominant genetic models, we predict an approximately linear relation of disease risk to individual admixture and these three types of model cannot easily be distinguished. The risk ratio associated with unit change in M should equal the risk ratio observed between populations living in the same environment. This contrasts with the predictions from an environmental model (no relationship of disease risk to admixture, after controlling for confounders) or an "overdominant" model (inverse U-shaped relation of disease risk to admixture and positive relationship of disease risk to heterozygosity of ancestry).

Although this approach has been successfully applied to studying the genetics of hypertension in African-American and African-Caribbean populations by using markers such as blood groups and protein electrophoresis (Darlu et al. 1990; MacLean et al. 1974), it has not been widely used until recently to study other diseases, because it has not been technically possible to measure the admixture of individuals even approximately. Even with the 31 ancestry-informative markers used in this study, the posterior standard deviations (asymptotically equivalent to standard errors) in the estimates of individual admixture are large. The ability to estimate the slope of the relationship between disease risk and individual admixture also depends upon studying a population in which admixture propor-

Table 4 Affected-only tests for linkage with marker loci

Locus	Score	Observed information	Percent information extracted	Z	P-value (one-sided)
MID-187	-0.78	0.08	3%	-2.73	0.997
FY-null	1.53	2.19	71%	1.04	0.150
AT3-Indel	1.02	0.89	29%	1.08	0.141
WI-11392	-0.01	0.24	8%	-0.03	0.510
WI-16857	-0.75	0.20	7%	-1.67	0.953
WI-11153	0.65	0.29	9%	1.20	0.116
GC	-0.92	0.77	25%	-1.05	0.852
SGC30610	0.02	0.05	1%	0.10	0.459
SGC30055	0.82	0.83	27%	0.90	0.185
WI-17163	0.77	0.48	15%	1.12	0.131
WI-9231	-0.43	-0.01	0%	-	-
WI-4019	-0.41	0.66	21%	-0.51	0.695
CYP3A4	-0.47	0.89	29%	-0.50	0.692
LPL	0.96	0.74	24%	1.11	0.133
WI-11909	-0.62	-0.07	-2%	-	-
TYR-192	0.83	0.48	16%	1.20	0.116
DRD2	0.79	0.87	28%	0.85	0.198
APOA1-Alu	0.86	0.81	26%	0.96	0.169
D11S429	-0.11	0.22	7%	-0.23	0.590
GNB3_C825T	-0.54	0.17	6%	-1.30	0.903
RB1	-0.13	0.95	31%	-0.13	0.552
OCA2	0.14	0.57	18%	0.18	0.427
WI-14319	0.25	0.14	4%	0.68	0.247
CYP19-E2	0.60	0.16	5%	1.53	0.063
PV92-Alu	0.07	0.12	4%	0.21	0.416
MC1R-314	-0.78	0.16	5%	-1.94	0.974
WI-7423	-0.96	0.80	26%	-1.07	0.859
WI-14867	-0.45	0.68	22%	-0.54	0.707
CKM	0.48	0.39	13%	0.77	0.221
MID-154	0.24	0.33	11%	0.41	0.341

tions vary widely between individuals. The population of Trinidad meets this requirement, as even when individuals of Indian or Chinese ancestry are excluded, 19% of individuals have less than 60% west African admixture, even though the mean proportion of west African admixture is 0.73. The estimated African/non-African ancestry crossover rate of 2.4 per 100 cM suggests that the average time since admixture between Africans and non-Africans is at least five generations, consistent with historical evidence.

Studies of admixture in Trinidad are complicated by five-way admixture between west African, European, Native American, Indian and Chinese populations. Risk of SLE has been reported to be higher in people of Chinese and Indian descent settled overseas than in other ethnic groups settled in the same countries (Catalano and Hoffmeier 1989; Frank 1979; Serdula and Rhoads 1979; Wang et al. 1997; Samanta et al. 1992). This is consistent with the clinical impression that the risk of SLE is high in Indians and in Africans in Trinidad. At present, we are unable to model this five-way admixture adequately because we do not have markers that are informative for Indian and Chinese ancestry. In Trinidad, the admixture of Chinese and Indians with other groups has been uncommon until the last two generations. Excluding individuals who

reported Chinese or Indian ancestry in their grandparents should therefore be sufficient to restrict the study population to those with no more than three-way admixture between Europeans, west Africans and Native Americans. Application of the same inclusion/exclusion criteria to cases and controls ensures that both groups are representative of the same "population at risk".

In any study of the relationship of disease risk to individual admixture, socioeconomic and demographic factors may confound the association. We controlled for these possible confounders first by matching cases and controls for neighbourhood and second by adjusting, in the analysis, for variables such as household amenities in childhood and educational status.

The association of SLE with lack of access to piped water during childhood suggests a possible role for poor hygiene and exposure to infection in early life. This association is in the opposite direction to what would be predicted from Greenwood's 1968 hypothesis that exposure to parasitic infections in early life protects against autoimmune disease in developing countries. In any case, adjustment for this association had little effect on the estimated slope of the relationship between SLE risk and admixture. The analysis did not adjust for reproductive history, as dif-

ferences between cases and controls in reproductive history are likely to be a consequence rather than a cause of the disease. Women with SLE may avoid becoming pregnant because they are ill. Additionally, SLE is associated with increased risk of early fetal loss caused by antiphospholipid antibodies (Firkin et al 1980); these early fetal losses may occur before women realize that they are pregnant.

Although the score tests for association with admixture are based on the asymptotic properties of the likelihood in large samples, the Bayesian estimates, which do not rely on asymptotic arguments, yield similar results in that the 95% CI for the risk ratio associated with admixture does not overlap 1. The estimates of the risk ratio were similar whether estimated in the Bayesian model (which correctly allows for uncertainty in estimation of individual admixture), in a classical unconditional logistic regression model (which ignores the matching in the design) or in a conditional logistic regression model (which takes matching into account).

The relationship between risk of SLE and west African admixture demonstrated in this study is consistent with an additive genetic model for the ethnic difference in risk. We can reject not only a purely environmental explanation, from which we would predict no relation between risk of SLE and west African admixture, but also an "overdominant" model as an explanation for this ethnic difference. For an "overdominant" model to explain the more than five-fold higher risk of SLE in African-Caribbean and African-American populations compared with Europeans, the relationship of disease risk to heterozygosity of ancestry would have to be extremely strong (risk ratio >10), as the predicted heterozygosity of ancestry in African-American or African-Caribbean populations with 20% European admixture is only 32%. From our results, a risk ratio larger than 1.25 (the upper boundary of the 95% CI) associated with unit change (from 0 to 1) in heterozygosity of ancestry can be excluded.

Although the confidence limits for the slope of this relationship are wide, the estimates are consistent with the risk ratios of 6–8 that have been reported in comparisons between African-American or African Caribbean populations and people of European ancestry (Fessel 1974; Hopkinson et al. 1994; Johnson et al. 1995; Nossent 1992). This does not exclude a role for environmental factors. In this study, risk of SLE was strongly associated with lack of household access to running water in childhood and this association remained statistically significant after adjusting for admixture, suggesting a possible role for exposure to infection in early life.

Exploiting admixture to map genes

Several genetic associations with SLE have been identified, including associations with genes in the HLA region. However, even where there are ethnic differences in the frequencies of high-risk alleles or haplotypes at these loci, these differences do not account for the observed ethnic

difference in disease risk between people of west African and European descent (Molokhia and McKeigue 2000).

Demonstration of a relationship of disease risk to individual admixture lays the basis for exploiting admixture to localize the genes underlying this relationship. If the number of genes that underlie the effect of admixture is not too large, the theory of admixture mapping guarantees that these genes can be localized (McKeigue 1998). For a rare disease such as lupus, the most efficient approach is a case-only design, testing for regions of the genome where the proportion of gene copies that have African ancestry is higher in cases than expected from their parents' admixture proportions. In this paper, we have demonstrated the use of a score test based on this approach. A full admixture mapping study, however, would require a larger sample of cases, ideally from a population with a lower proportion of west African admixture. For example, to detect a locus that accounts for a risk ratio of 3 between west Africans and Europeans by admixture mapping, at least 130 cases with 50% west African admixture would be required (McKeigue 1998). About twice as many cases are required where the proportion of west African admixture in cases is 80% (as in this study) rather than 50%. Admixture mapping would also require a far denser map of markers informative for ancestry. With the current marker set, the test for linkage extracts only a small proportion of the information potentially available, except in regions such as 1q23 and 11q23 where the analysis uses information from two or more linked markers. We cannot therefore at this stage determine whether regions in which linkage to SLE have been previously detected in African-American populations also show linkage to genes underlying the ethnic difference in SLE in this Caribbean population.

Simulations indicate that an average marker spacing of 2–3 cM will be required to extract at least 70% of information about ancestry when testing for linkage (McKeigue 1998). The accumulation of data on SNP allele frequencies in various ethnic groups will soon make it possible to select hundreds of additional markers that are informative for west African, European and Native American ancestry (Collins-Schramm et al. 2002).

Acknowledgements Ethical approval was granted by the ethics committee of the London School of Hygiene and Tropical Medicine and by the Ministry of Health of the Republic of Trinidad and Tobago. This study was supported by the UK Arthritis Research Campaign. The development and testing of the ADMIXMAP program was supported by NIH grant no. IR01MH60343-01A1 to P.M.M. Marker development and typing was supported by grants from the NIH/NHGRI (HG002154) to M.D.S. IRB approval from the NIH was granted for this project. We thank those who participated in this research, Gloria Chan, Odette Mason, Martha Greenidge and Sylvia Liverpool for help with the data collection, Wayne LaBastide for help with programming, and Dr. Marilyn Suite, Dr. Kim Basdeo Maharaj, Dr. Leslie Roberts, Dr. Deepak Mahabir, Dr. Rajiv Serrano, Prof. Hylton McFarlane, Dr. Zinora Asgaralli, Ms Arlene Damanie, Dr. Peter Poon King, Dr. Richard Poon King, Dr. Dianne Sandy, Mr Curtis Wilson and others at the Lupus Society of Trinidad and Tobago for help with case ascertainment.

Appendix

Marker information content for ancestry (f)

We shall consider a haploid population formed by equal admixture between two populations X and Y. The ancestry at a marker locus in a randomly chosen haploid genome can be modelled as a Bernoulli random variable with parameter $1/2$. The prior variance of locus ancestry (before typing) is then $1/4$. If the marker is informative for ancestry, the posterior variance of locus ancestry (after typing the locus) will be less than the prior variance. The marker information content for ancestry f is defined as

$1 - (\text{the expected posterior variance of locus ancestry}/\text{the prior variance of locus ancestry})$.

For a locus with k alleles and allele frequencies p_{iX} and p_{iY}

$$f = 1 - 2 \sum_{i=1}^k \frac{p_{iX} p_{iY}}{p_{iX} + p_{iY}}$$

For a biallelic locus, f is equal to Wahlund's standardized variance of allele frequencies, defined as

$$\frac{H_T - H_S}{H_T}$$

where H_T is the heterozygosity of a total population formed by pooling equal numbers from the two subpopulations, and H_S is the mean heterozygosity of the two subpopulations.

References

- Akey JM, Sosnoski D, Parra E, Dios S, Hiester K, Su B, Bonilla C, Jin L, Shriver MD (2001) Melting curve analysis of SNPs (McSNP): a gel-free and inexpensive approach for SNP genotyping. *Biotechniques* 30:358–367
- Breslow NE, Day NE (1980) *Statistical methods in cancer research, vol I. The analysis of case-control studies*. International Agency for Research on Cancer, Lyon, pp 270–279
- Catalano MA, Hoffmeier M (1989) Frequency of SLE among the ethnic groups of Hawaii. *Arthritis Rheum* 32:s30
- Central Statistical Office Trinidad and Tobago (1992) *Annual Statistical Digest No. 39. Population and vital statistics*. Central Statistical Office, Trinidad and Tobago, p 9
- Chakraborty R (1975) Estimation of race admixture — a new method. *Am J Phys Anthropol* 42:507–511
- Chakraborty R, Weiss KM (1986) Frequencies of complex diseases in hybrid populations. *Am J Phys Anthropol* 70:489–503
- Collins-Schramm HE, Phillips CM, Operario DJ, Lee JS, Weber JL, Hanson RL, Knowler WC, Cooper R, Li H, Seldin MF (2002) Ethnic-difference markers for use in mapping by admixture linkage disequilibrium. *Am J Hum Genet* 70:737–750
- Darlu P, Sagnier PP, Bois E (1990) Genealogical and genetical African admixture estimations, blood pressure and hypertension in a Caribbean community. *Ann Hum Biol* 17:387–397
- Fessel J (1974) Systemic lupus erythematosus in the community: incidence, prevalence, outcome and first symptoms; the high prevalence in Black women. *Arch Intern Med* 134:1027–1035
- Firkin BG, Howard MA, Radford N (1980) Possible relationship between lupus inhibitor and recurrent abortion in young women. *Lancet* II:366
- Frank AO (1979) Apparent predisposition to systemic lupus erythematosus in Chinese patients in west Malaysia. *Ann Rheum Dis* 39:266–269
- Greenwood BM (1968) Autoimmune disease and parasitic infections in Nigerians. *Lancet* II:380–382
- Hopkinson ND, Doherty M, Powell RJ (1994) Clinical features and race specific incidence/prevalence rates of SLE in a geographically complete cohort of patients. *Ann Rheum Dis* 53:675–680
- Johnson AE, Gordon C, Palmer RG, Bacon PA (1995) The prevalence and incidence of systemic lupus erythematosus in Birmingham, England. Relationship to ethnicity and country of birth. *Arthritis Rheum* 38:551–558
- MacLean CH, Adams MS, Leyshon WC, Workman PL, Reed TE, Gershowitz H, Weitkamp LR (1974) Genetic studies on hybrid populations. III. Blood pressure in an American Black community. *Am J Hum Genet* 26:614–626
- McKeigue PM (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
- McKeigue PM, Carpenter JR, Parra EJ, Shriver MD (2000) Estimation of admixture and detection of linkage in admixed populations by a Bayesian approach: application to African-American populations. *Ann Hum Genet* 64:171–186
- Molokhia M, McKeigue PM (2000) Risk for rheumatic disease in relation to ethnicity and admixture. *Arthritis Rheum* 2:115–125
- Molokhia M, McKeigue PM, Cuadrado M, Hughes G (2001) Systemic lupus erythematosus in migrants from west Africa compared with African-Caribbean people in the UK. *Lancet* 357:1414–1415
- Nossent JC (1992) Systemic lupus erythematosus on the Caribbean island of Curacao: an epidemiological investigation. *Ann Rheum Dis* 51:1197–1201
- Samanta A, Roy S, Feehally J, Symmons DPM (1992) The prevalence of diagnosed systemic lupus erythematosus in Whites and Indian Asian immigrants in Leicester city, UK. *Br J Rheumatol* 31:679–682
- Serdula MK, Rhoads GG (1979) Frequency of systemic lupus erythematosus in different ethnic groups in Hawaii. *Arthritis Rheum* 22:328–333
- StataCorp (2001) *Stata statistical software: release 7.0*. Stata Corporation, College Station, Tex., USA
- Tan EM, Cohen AS, Fries JF, Masi AT, McShane DJ, Rothfield NF, Schaller JG, Talal N, Winchester RJ (1982) The 1982 revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum* 25:1271–1277
- Theofilopoulos AN, Dixon FJ (1985) Murine models of systemic lupus erythematosus. *Adv Immunol* 37:269–390
- Vyse TJ, Kotzin BL (1998) Genetic susceptibility to systemic lupus erythematosus. *Annu Rev Immunol* 16:261–292
- Wang F, Wang CL, Tan CT, Manivasagar M (1997) Systemic lupus erythematosus in Malaysia: a study of 539 patients and comparison of prevalence and disease expression in different racial and gender groups. *Lupus* 6:248–253