

The African Diaspora: Mitochondrial DNA and the Atlantic Slave Trade

Antonio Salas,^{1,2,3} Martin Richards,^{2,4} María-Victoria Lareu,¹ Rosaria Scozzari,⁵ Alfredo Coppa,⁶ Antonio Torroni,⁷ Vincent Macaulay,^{3,8} and Ángel Carracedo¹

¹Unidad de Genética Forense, Instituto de Medicina Legal, Universidad de Santiago de Compostela, Galicia, Spain; ²Department of Chemical and Biological Sciences, University of Huddersfield, Huddersfield, United Kingdom; ³Department of Statistics, University of Oxford, Oxford; ⁴Schools of Biology and Computing, University of Leeds, Leeds, United Kingdom; ⁵Dipartimento di Genetica e Biologia Molecolare and ⁶Dipartimento di Biologia Animale e dell'Uomo, Università "La Sapienza," Rome; ⁷Dipartimento di Genetica e Microbiologia, Università di Pavia, Pavia, Italy; and ⁸Department of Statistics, University of Glasgow, Glasgow

Between the 15th and 19th centuries AD, the Atlantic slave trade resulted in the forced movement of ~13 million people from Africa, mainly to the Americas. Only ~11 million survived the passage, and many more died in the early years of captivity. We have studied 481 mitochondrial DNAs (mtDNAs) of recent African ancestry in the Americas and in Eurasia, in an attempt to trace them back to particular regions of Africa. Our results show that mtDNAs in America and Eurasia can, in many cases, be traced to broad geographical regions within Africa, largely in accordance with historical evidence, and raise the possibility that a greater resolution may be possible in the future. However, they also indicate that, at least for the moment, considerable caution is warranted when assessing claims to be able to trace the ancestry of particular lineages to a particular locality within modern-day Africa.

Introduction

Slavery has been a feature of human societies since antiquity, but the scale of the Atlantic slave trade was unprecedented. This African diaspora was the result of the enslavement in the Americas of probably ~11 million people, and at least 2 million more died during the middle passage. The traders spread from northwestern Africa around the Atlantic coast, reaching Angola by the 17th century and Mozambique by the 19th. Historical records suggest that western Africa contributed ~8 million people, west-central Africa (from Cameroon down to Angola) ~4 million, and Mozambique/Madagascar in the southeast ~1 million more, although agreement on details is lacking (Curtin 1969; Fage 1969; Thomas 1998). The scale of the displacements and the European prejudice against African populations meant that, rather than the assimilation of the imported populations, as had often been the case in antiquity (Lewis 1990), the result was often the creation of new and distinctive populations in the Americas (Sans 2000).

African mtDNAs fall into several distinctive paralogous (paraphyletic clusters of mtDNA lineages) and monophyletic haplogroups. The nomenclature has been amended following the suggestion of Mishmar et al. (2003), so that L1a, L1d, L1f, and L1k are here renamed

L0a, L0d, L0f, and L0k, respectively. These haplogroups, along with L1b'c, L1e, L2, and L3A, evolved in sub-Saharan Africa (Chen et al. 1995, 2000; Watson et al. 1997), whereas U6 is thought to have evolved in northern Africa (Rando et al. 1998; Macaulay et al. 1999). Each of these major clusters can be divided on the basis of various diagnostic mutations into subhaplogroups (Torroni et al. 2001; Salas et al. 2002), which can be used as markers of recent migrations when found in other parts of the world.

Alves-Silva et al. (2000), for example, found that 28% of a sample of mainly "white" subjects from Brazil were of recent African maternal ancestry, with substantial variation from region to region. Some lineages could be readily attributed to arrivals from western Africa, which had already been extensively sampled, but almost 50% belonged to haplogroups L1c and L3e, which are rare in western Africa. Their presence at much higher frequencies in small west-central African data sets suggested that the Brazilian L1c and L3e mtDNAs might be of largely west-central African origin (see also Bandelt et al. 2001). Indeed, major sources for Brazilian slaves are thought, on the basis of historical records, to have been Congo and Angola (e.g., Curtin 1969; Thomas 1998). Alves-Silva et al. (2000) predicted that, when these regions were sampled for mtDNA variation, L1c and L3e would be found at high frequencies.

Brehm et al. (2002) identified mtDNA lineages transported proximally from the western Atlantic coast to the Cabo Verde islands, many en route to the Caribbean and Brazil, and Pereira et al. (2001) initiated work on the slave trade from southeastern Africa. They discovered numerous matches between Mozambican mtDNAs

Received October 1, 2003; accepted for publication December 15, 2003; electronically published February 10, 2004.

Address for correspondence and reprints: Dr. Antonio Salas, Unidad de Genética Forense, Universidad de Santiago de Compostela, Galicia, Spain. E-mail: apimlase@usc.es

© 2004 by The American Society of Human Genetics. All rights reserved. 0002-9297/2004/7403-00XX\$15.00

and those of America and, to a lesser extent, those of recent African ancestry in Europe (see also Salas et al. [2002]).

There have also been several attempts, using autosomal, mtDNA, and Y-chromosomal markers, to address the question of the extent of recent African versus European and Native American ancestry in American populations (Chakraborty et al. 1992; Parra et al. 1998, 2001; Mesa et al. 2000). However, attempts have not hitherto been made to trace American ancestry to particular regions within Africa. The accumulation of mtDNA data from both African populations and American populations with putative African ancestry has now reached the point where this can be pursued. In this study, we synthesize this body of evidence and attempt to quantify the contributions of the major subregions of Africa to mtDNAs in the Americas and Eurasia.

Material and Methods

Population Samples

We compiled a database from 481 individuals carrying African mtDNAs belonging to seven available American samples harboring a major African component. The sample from North America included 101 African Americans from the United States (HvrBase database; for details, see Handt et al. 1998). The sample from Central America/Caribbean included 8 from Mexico (Green et al. 2000), 25 Carib from Belize (Monsalve and Hagelberg 1997), 112 from the Dominican Republic (Torrioni et al. 2001; A.T., unpublished data), 41 Chocó from Columbia, and 37 Garífuna from Panama and Belize (A. Salas, M. Richards, M.-V. Lareu, S. Silva, M. Matorros, V. Macaulay, and A. Carracedo, unpublished data). The sample from South America included 157 Brazilians: 29 from the study by Bortolini et al. (1997), 68 from the study by Alves-Silva et al. (2000), and 60 from the study by Santos et al. (2002). The northern African sample (650 total) included 30 Mauritians, 25 Saharans, 60 Berbers from Morocco, and 32 Moroccans (Rando et al. 1998); 50 Moroccans from the Souss Valley (Brakez et al. 2001); 68 Egyptians (Klings et al. 1999); 85 Algerian Berbers (Côte-Real et al. 1996); and 300 Canarians (246 from the study by Rando et al. [1998] and 54 from the study by Pinto et al. [1996]). The western African sample (694 total) included 20 Hausa, 14 Kanuri, 60 Fulbe, 10 Songhai, 23 Tuareg, and 21 Yoruba (Watson et al. 1996); 14 further Yoruba (Vigilant et al. 1991); 23 Serer, 50 Senegalese, and 48 Wolof (Rando et al. 1998); 119 Mandenka (Graven et al. 1995); and 292 individuals from the Cabo Verde islands (Brehm et al. 2002). The west-central African sample (179 total) included 50 individuals from the islands of São Tomé and Príncipe and 45 Bubi from the

island of Bioko (Mateu et al. 1997); 11 Fang from Guinea (Pinto et al. 1996); 43 Angolans (Santos et al. 2002); 17 Biaka and 13 Mbuti (Vigilant et al. 1991). The eastern African sample (335 total) included 37 Turkana, 27 Somali, and 25 Kikuyu (Watson et al. 1996); 96 Nubians and 76 Sudanese (Klings et al. 1999); and 74 Ethiopians (Thomas et al. 2002). The southeastern African sample included 417 individuals from Mozambique (Pereira et al. 2001; Salas et al. 2002). The southern African sample (99 total) included 25 !Kung and 31 Khwe (Chen et al. 2000); and 43 further !Kung (Vigilant et al. 1991). We also surveyed the mtDNA first hypervariable segment (HVS-I) of ~15,000 individuals from Eurasia and found 113 mtDNAs of recent African ancestry. We excluded another 28 sequences from the database, possibly belonging to L3 but not characterized by any diagnostic sites in our data.

All sequences correspond to HVS-I of the control region; the nucleotide positions considered included at least positions 16090–16365, according to the numbering system of the Cambridge Reference Sequence (Andrews et al. 1999). For some purposes, we also used HVS-I information from outside the minimum segment, as well as RFLP and HVS-II data. Length variation was excluded. Transitions are indicated by the nucleotide position minus 16,000, and transversions are indicated by a suffix.

Phylogeographic Analysis

The haplogroup classification of mtDNA sequences was updated from Salas et al. (2002); see table 1 for HVS-I motifs (cf. Chen et al. 1995; Watson et al. 1997; Rando et al. 1998; Quintana-Murci et al. 1999; Alves-Silva et al. 2000; Bandelt et al. 2001; Pereira et al. 2001; Torrioni et al. 2001). Richards et al. (2003) discuss mtDNA haplogroups that may be regarded as of recent African ancestry.

We constructed phylogenetic networks (Bandelt et al. 1995, 1999) by hand, on the basis of the information from Salas et al. (2002), with the aid of the Network 3.0 package (Fluxus Engineering Web site). We performed principal component (PC) analysis based on haplogroup frequencies, broken down as follows: U6, M1, L0a, L0d/k, L1b, L1c, L2a, L2b, L2c, L3b, L3d, L3e, L1/L2 remainder, L3A remainder, and non-African haplogroups. The entire data sets were used in the case of Africans, whereas only mtDNAs of recent African ancestry were used in the case of Americans. We excluded some outliers from the PC analysis, as in Salas et al. (2002), including the Khoisan samples, with their high levels of L1d/L1k; some northern African populations, with moderate levels of U6; and the Biaka, with their very high frequency of L1c. However, it should be noted that three of the L1c HVS-I sequence types in the Biaka

Table 1
HVS-I Sequence Motifs Used for Haplogroup Classification

Haplogroup	HVS-I Sequence Motif ^a (np-16,000)
L1*	187-189-223-278-311
L0a*	129-148-172-187-188G-189-223-230-311-320
L0a1	129-148-168-172-187-188G-189-223-230-311-320
L0a2	148-172-187-188G-189-223-230-311-320
L0d	129-187-189-223-230-243-311
L1b	126-187-189-223-264-270-278-311
L1c*	129-187-189-223-278-294-311-360
L1c1	129-187-189-223-278-293-294-311-360
L1c2	129-187-189-223-265C-278-286G-294-311-360
L1c3	129-189-215-223-278-294-311-360
L1e	129-148-166-187-189-223-311
L2a	223-278-294-(390)
L2b	114A-129-213-223-278-(390)
L2c	223-278-(390)
L2d	223-278-(390)-(399)
L3*	223
L3b	124-223-278-362
L3d	124-223
L3e1*	223-327
L3e1a	185-223-327
L3e1b	223-325D-327
L3e2*	223-320
L3e2b	172-189-223-320
L3e3	223-265T
L3e4	223-264
L3f	209-223-311
L3g	223-293T-311-355-362

^a Those positions that fall outside the common HVS-I segment used in this article are in parentheses. D = deletion.

sample match American types (see fig. 3). We also excluded the Mexican sample because of its small size.

We estimated sequence diversity as $[n/(n-1)](1 - \sum_{i=1}^k p_i^2)$, where p_i is the frequency of each of the k different sequences in the sample and n is the sample size, and we evaluated the mean pairwise nucleotide difference, π , as well as the number of segregating sites, S , through use of Arlequin 2.0 (Arlequin's Home on the Web). We also computed haplotypic and segregation site mutation-drift statistics (θ_k and θ_s , respectively) (Helgason et al. 2003).

Admixture Analysis

To quantify the magnitude of the impact of each African region on cluster frequencies in the Americas, we fitted the following model. The number of mtDNAs in each cluster in the sample from a region of the Americas ($n_i; 1 \leq i \leq C$, the number of clusters) was assumed to be a draw from a multinomial distribution with parameters $N = \sum_{i=1}^C n_p$, the sample size in the American region, and $p_i = \sum_{j=1}^R \alpha_j f_{ji}$ ($1 \leq i \leq C$), where R is the number of source regions in Africa, f_{ji} is the frequency of the i th cluster in the j th source region (assumed to be known), and the α_j are the admixture coefficients. This

model describes samples from an urn with C different kinds of ball, where the urn has been created by mixing together R other urns in proportions given by the admixture coefficients. We chose to analyze this model in a Bayesian framework, which meant that we had to explore the distribution of the admixture coefficients, given the data. The prior distribution of the admixture coefficients was taken to be uninformative—namely, uniform on $\alpha_j \geq 0, \sum_j \alpha_j = 1$. The posterior distribution of α_j was explored with the Metropolis-Hastings algorithm, using a simple proposal, and was summarized by the posterior mean of each α_j and its root-mean-square deviation about the mean.

To assess model fit, we examined plots of standardized residuals. The analysis was first performed with all African regions present, and then, to test robustness, was repeated with those regions removed whose admixture coefficients were within 2 SDs of 0.

Results

Data Summaries and Demography

Genetic diversity indices are shown in table 2. The haplotype mutation-drift statistic (θ_k) and haplotype diversity (H) show a distinctive pattern that is less clearly reflected in the segregating sites mutation-drift statistic (θ_s) or the mean number of pairwise differences (π). θ_k and H are similar, with eastern Africa having the highest values and southern Africa the lowest, and with minor variations in ranking. They suggest that founder effects were not important during the formation of slave communities of the Americas but that reductions in population size may have subsequently affected populations in Central America.

The reduction in diversity for Central American lineages is also indicated by the estimator π , but it is not reflected in this statistic for southeastern Africa (where founder effects associated with Bantu dispersals have evidently elevated the frequencies of some types in several haplogroups [Salas et al. 2002]). This is likely to be because, as a measure of pairwise nucleotide differences, π exaggerates the effect of the genetic divergence *between* haplogroups that may survive even after a substantial reduction in haplotype diversity; thus, π is a poor indicator of population size reduction (Helgason et al. 2003). This is evident in southeastern Africa—for example, between L0a and L3e—and is even more pronounced in west-central Africa, where, for example, L0a, L1c, and L3e are all significant. Since the haplogroup compositions of North, Central, and South America are very similar, the reduction in diversity in Central America becomes evident even with this statistic. The mutation-drift parameter based on segregating sites (θ_s), by contrast, shows a reduction in southeastern Africa but is not sensitive to the

Table 2**HVS-I Diversity Indices in the African mtDNAs from North, Central, And South American Population Samples and in Africans from Different Regions in Africa**

Population	<i>n</i> (or <i>N</i>) ^a	<i>k</i> ^b	θ_k	<i>S</i>	θ_s	<i>H</i> (SE)	π
North America	(101)	76	138	70	13.5	.991 (.003)	7.8
Central America	(223)	109	83.6	80	13.4	.981 (.003)	7.2
South America	(157)	110	162	84	14.9	.993 (.002)	8.2
Northern Africa	651	259	159	115	16.3	.975 (.003)	5.1
Western Africa	687	285	182	108	15.2	.988 (.001)	6.6
West-central Africa	179	97	85.7	80	13.9	.987 (.002)	8.7
Eastern Africa	328	208	243	106	16.6	.994 (.001)	8.0
Southeastern Africa	396	133	69.9	78	11.9	.964 (.004)	8.4
Southern Africa	98	27	12.0	45	8.73	.925 (.013)	7.9

^a Values in parentheses (*N*) are numbers of L/U6/M1 haplotypes in the African-American samples; all other values (*n*) are total sample sizes.

^b Number of different sequences found among *N* in the American samples and *n* in the African samples.

effect in Central America. In general, it therefore appears that neither of these latter two statistics is consistently sensitive to demographic processes.

Haplogroup Frequency Profiles

Haplogroup frequency profiles for the regions studied are shown in figure 1, and a scatterplot of the first two principal components is shown in figure 2. The first principal component (PC1), accounting for 30% of the variance, separates the populations into a spectrum with northern Africans at one pole, followed by northeast Africans, eastern Africans, western Africans, and west-central Africans, with Bioko at the opposite pole. Southeastern Africans and Angolans sit tightly together within the west-central African cluster. The American samples also cluster fairly tightly together in PC1, with the Brazilian and several of the Central American samples localized very closely with the southeastern Africans, whereas the African American, São Tomé, and Caribbean samples are positioned closer to the western Africans.

PC2 accounts for 19% of the variance and shows an approximately east-west gradient, with northern Africans, west-central Africans, and southeastern Africans all positioned towards the eastern fringe of the western African variation. The Caribbean sample sits with the southeastern Africans, whereas the other American samples are positioned more towards the western pole. The African Americans, Brazilians, and mainland Central Americans cluster tightly together and with the São Toméan and Dominican samples, further towards the western pole.

Admixture Analysis

To quantify the proportion of ancestry from each of the major regions of Africa present in the African component of the modern American mtDNA pool, we fitted

an admixture model of the haplogroup profiles in the African regions to the profiles in the Americas (table 3). The estimated admixture coefficients for all regions other than western and west-central Africa were within 2 SDs of 0, and so we repeated the analysis with a more parsimonious model by suppressing those unsupported African regions (table 4).

The results for the Americas overall accord broadly with the picture built up by historians, in attributing major contributions from both western and west-central Africa. The proportion of western African ancestry is not significantly greater than that of west-central Africa, and a significant southeastern African component is not detected at this level of analysis.

Again in agreement with the historical evidence—and confirming the indications from the frequency profiles—the largest component in both North and Central America appears to derive from western Africa and to be somewhat greater in Central than North America. The southeast also seems to have left a minor signature, although the error margins are wide. South America is, however, problematic. A higher—and possibly a majority—contribution from west-central Africa is indicated, but the admixture model provides a poor description of the data, as judged from a plot of standardized residuals (not shown). For example, the frequency of L3e in South America is larger than in any of the potential source regions, an observation incompatible with the admixture model. (It is most closely approached by the frequency of L3e in Bioko, which, however, may be an unrepresentative sample, since it appears to have been subject to extensive drift [$\theta_k = 8.4$].) Several other population samples that contribute to our west-central African sample are unlikely slave sources and have also been subject to heavy drift, in particular the Mbuti ($\theta_k = 2.5$). This suggests that a major source region for South America remains unsampled; more data will be necessary to make

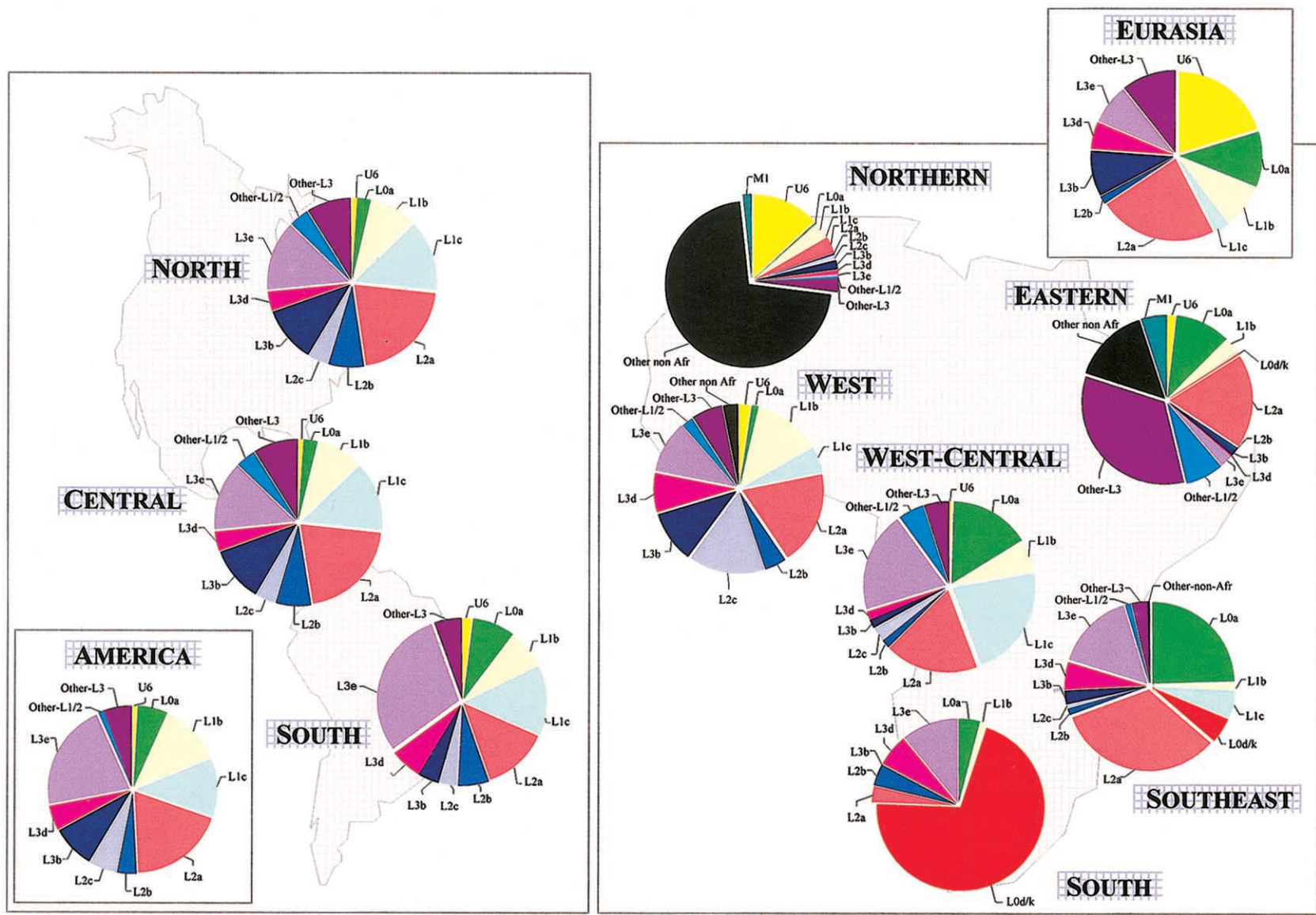


Figure 1 Frequency profiles of the major haplogroups in Africa and haplogroups of recent African ancestry in the Americas and Eurasia

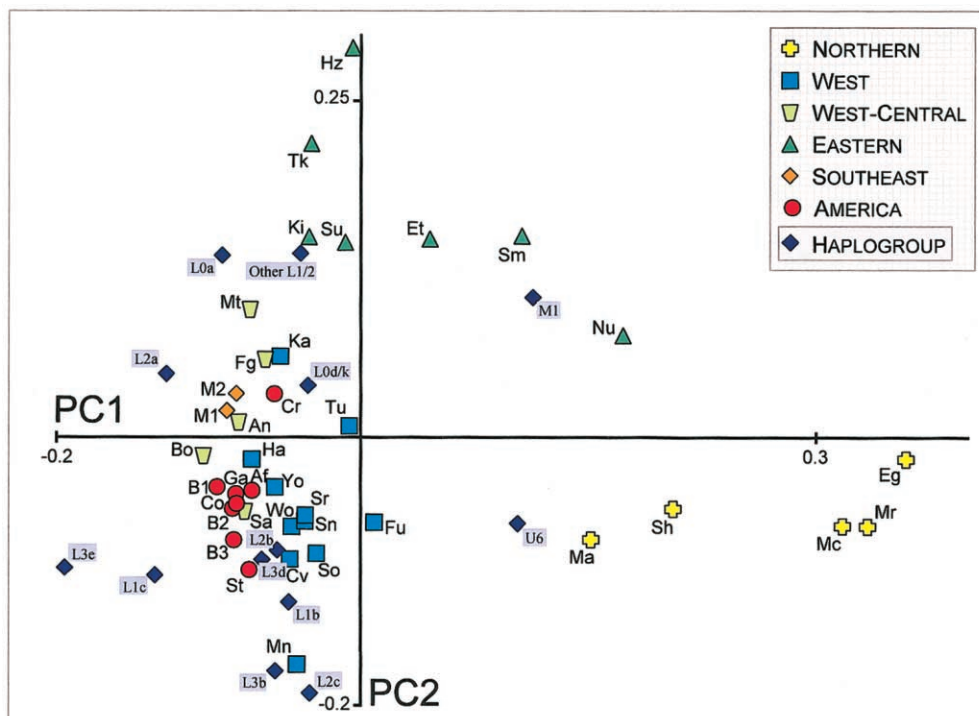


Figure 2 Scatterplot of the first two principal components based on haplogroup frequencies. For practical reasons, the “Other-L3” and “Other non Afr” haplogroup categories, with coordinates (0.0171, 0.4210) and (0.6680, 0.0768), respectively, are not displayed in the PC plot. Codes and references for populations are as follows: America: Af = African American (Vigilant et al. 1991; Handt et al. 1998); B1 = white Brazilian (Alves-Silva et al. 2000); B2 = African Brazilian (Bortolini et al. 1997); B3 = Bahía-Brazil (Santos et al. 2002); Co = Chocó, Ga = Garífuna (A. Salas, M. Richards, M.-V. Lareu, S. Silva, M. Matamoros, V. Macaulay, and A. Carracedo, unpublished data); Mx = Mexican (Green et al. 2000); Cr = Caribbean (Monsalve and Hagelberg 1997); St = Dominican Republic (Torrioni et al. 2001; A.T., unpublished data). Northern Africa: Sh = West Saharan, Ma = Mauritanian, Mo = Berber, Mr = Moroccan non-Berber (Rando et al. 1998); Mc = Moroccan (Brakez et al. 2001); Eg = Egyptian (Krings et al. 1999); Mb = Mozabite (Côrte-Real et al. 1996); Ca = Canarian (Pinto et al. 1996; Rando et al. 1998). Western Africa: Ha = Hausa, Ka = Kanuri, Fu = Fulbe, So = Songhai, Tu = Tuareg (Watson et al. 1997); Yo = Yoruba (Vigilant et al. 1991; Watson et al. 1997); Sn = Senegalese, Sr = Serer, Wo = Wolof (Rando et al. 1998); Mn = Mandenka (Graven et al. 1995); Cv = Cabo Verde (Brehm et al. 2002). West-central Africa: Bo = Bubi, Sa = São Tomé (Mateu et al. 1997); Fg = Fang (Pinto et al. 1996); Bi = Biaka, Mt = Mbuti (Vigilant et al. 1991); An = Angola (Santos et al. 2002). Eastern Africa: Tk = Turkana, Sm = Somalian, Ki = Kikuyu (Watson et al. 1997); Nu = Nubian, Su = Nuba, Shilio, Duba, Nuer (Krings et al. 1999); Et = Ethiopian (Thomas et al. 2002). Southeastern Africa: M1 = Mozambique (Pereira et al. 2001); M2 = Mozambique (Salas et al. 2002). South Africa: K2 = !Kung, Kw = Khwe (Chen et al. 2000).

further progress here. A further sampling issue arises because South America is represented only by Brazil (although this country bears the largest community of African descendants in America). The existing data suggest that Brazilian slaves came largely from west-central Africa rather than western Africa, but data from other parts of South America will be needed to generalize this observation.

Haplogroup Composition of American mtDNAs of African Ancestry

Frequencies of mtDNA lineages in Africa are shown in figure 1, alongside mtDNA lineages of African ancestry in North, Central, and South Americans (cf. Salas et al. 2002). A network of HVSI sequence types shared between Africans, Americans, and Eurasians is shown

in figure 3 (excluding U6 and one unclassified American L3* type, which matches a western African mtDNA). Eighty-eight mtDNA types from these clades are shared between the African samples and America. Although this is only 36% of the total different American sequence types of African origin in our database, it represents more than half (54%) of the total sample size of individuals, since a number of types are present in more than one individual in the database.

Remember that sequence types of Eurasian ancestry (within haplogroups M and N) are also excluded. However, in agreement with the historical evidence, the contribution of northern Africa (which includes many western Eurasian sequence types [Rando et al. 1998, 1999; Macaulay et al. 1999]) to the American gene pool appears to be very minor. This is indicated by the low

Table 3

Estimated Admixture Coefficients for American Samples, in Terms of the Source African Regions

AMERICAN REGION	ADMIXTURE COEFFICIENT ± SD IN AFRICAN (SOURCE) REGION ^a					
	N	W	WC	E	SE	S
North	.012 ± .012	.530 ± .101	.345 ± .102	.045 ± .037	.055 ± .049	.013 ± .013
Central	.005 ± .005	.691 ± .066	.212 ± .076	.012 ± .011	.073 ± .054	.007 ± .007
South	.010 ± .009	.321 ± .079	.585 ± .088	.016 ± .015	.057 ± .050	.011 ± .011
Overall	.003 ± .003	.559 ± .047	.399 ± .051	.007 ± .007	.029 ± .026	.003 ± .003

^a The posterior mean and root-mean-square deviation of the admixture coefficient for each African region is shown. Region abbreviations are as follows: N = northern; W = western; WC = west-central; C = central; E = eastern; SE = southeastern; S = southern.

frequency of the indigenous northern African haplogroup U6 in Americans: this haplogroup is found in <2% of American lineages of African origin, which is lower than the frequency of U6 in western Africa, the likely source for most American U6 mtDNAs.

Most, but not all, of the known African subhaplogroups within L0–L3A are present in America. Overall, the continental frequency profiles of the Americas, especially those of North and Central America, are strikingly similar to those of western and west-central Africa (fig. 1), confirming that these were the major source regions for American mtDNA lineages of African ancestry. South America has a markedly higher proportion of L3e and L0a lineages, which likely reflects a relatively higher contribution of both west-central and southeastern African mtDNAs.

The great majority of haplogroup L0a mtDNAs, concentrated in South America, belong to subclusters L0a1 and L0a2, both of which are typically southeastern African. All American L0a types are shared with individuals from Mozambique, where two are elevated to extremely high frequencies (fig. 3) (Pereira et al. 2001; Salas et al. 2002). Several of the southeastern African L0a founder types and a number of derived types are also found in the Angolan sample, indicating substantial commonality between the more southerly Eastern and Western Bantu-speaking communities and pointing also to a possible west-central African contribution, especially to South America.

In contrast, American L1b sequences are most often shared with samples from western Africa, where L1b is concentrated (fig. 3; see also fig. 4 of Salas et al. 2002). Shared L1b types are also found in northern Africa; however, in all cases except for one, these also match western Africans, and in the one exception there is also a match with southeastern Africa. Haplogroup L1c, which is likely to be of west-central African origin, is the African mtDNA clade with the most unmatched representatives in America, particularly in South America—only ~17% of American L1c types are shared with Africans. The Angolan sample, however, carries L1c at a frequency of ~21%, which confirms the prediction of Alves-Silva et

al. (2000) based on the pattern of L1c lineages in Brazil, where the frequency among mtDNAs of recent African ancestry is ~19%. One L1c3 type matches eight individuals from Bioko (mis-labeled as “Bi” in figure 5 of Salas et al. 2002). However, it also matches a number of individuals in Mozambique, suggesting a possible origin in west-central Africa for a lineage that was probably dispersed by the Bantu expansions (Salas et al. 2002). This is reinforced by the discovery of a derivative of this type in Angola. L1c is rare in Mozambique (~5%), suggesting that it may be a “west-central Bantu” marker that has been brought to southeastern Africa by interaction with western Bantu-speaking communities.

Haplogroup L2a is the most common and widely distributed sub-Saharan African haplogroup and is also frequent in the Americas (~19%). The wide distribution of L2a in Africa makes identifying geographical origins of lineages difficult. Nevertheless, almost all of the American L2a types match with or are one-step derivatives of some western African L2a type. Several common and widespread American types are shared with eastern Africa as well, but none are shared uniquely with eastern Africans alone. A number are also shared with southeastern Africans; again, however, all of these types are also present in western Africa. A largely western African provenance, with a possible minor southeastern African

Table 4

Estimated Admixture Coefficients, when Unsupported Source Regions Are Excluded

AMERICAN REGION	ESTIMATED ADMIXTURE COEFFICIENT ± SD IN AFRICAN (SOURCE) REGION ^a	
	W	WC
North	.592 ± .103	.408 ± .103
Central	.719 ± .066	.281 ± .066
South	.351 ± .080	.649 ± .080
Overall	.572 ± .047	.428 ± .047

^a The posterior mean and root-mean-square deviation of the admixture coefficient is shown. Region abbreviations are as follows: W = western; WC = west-central.

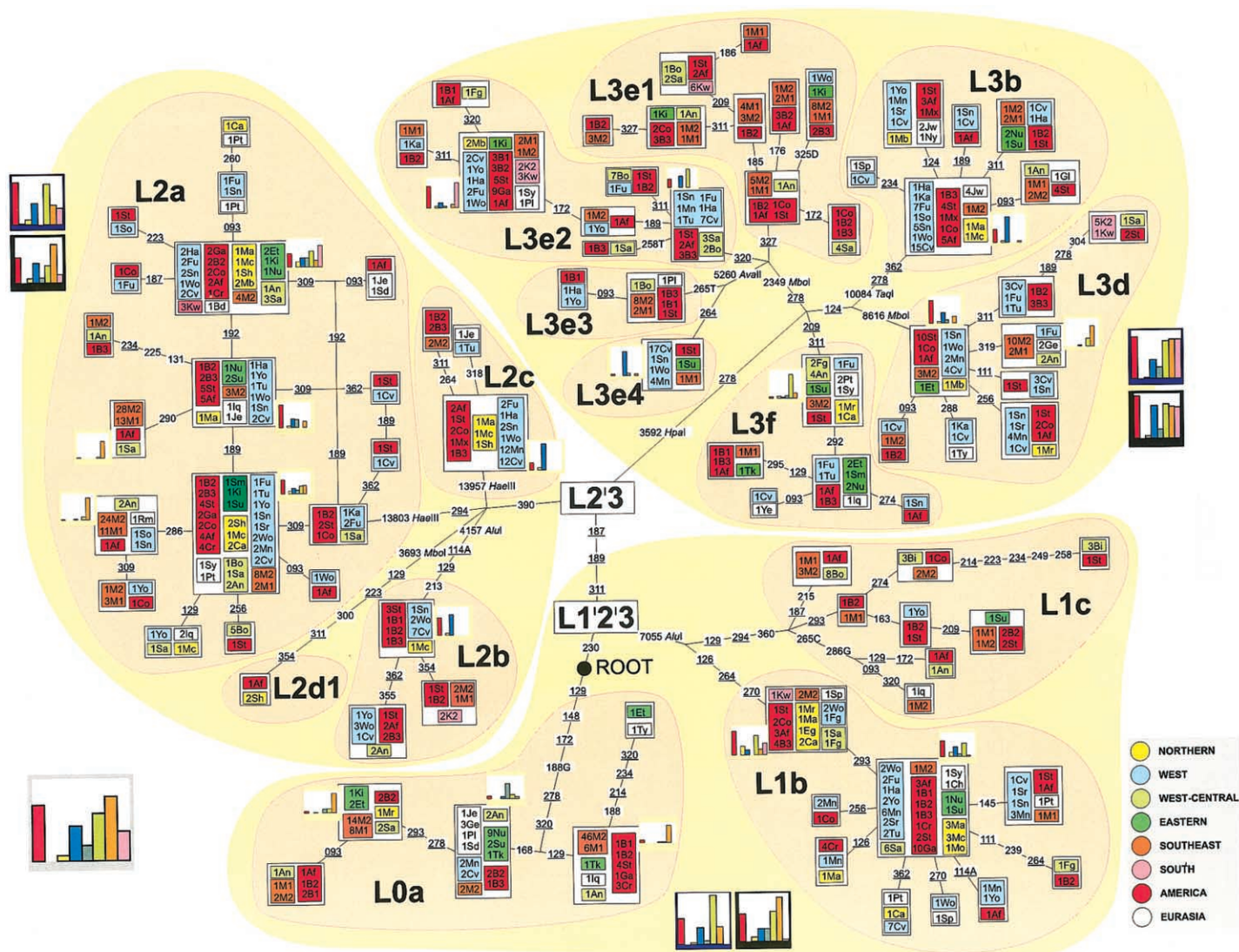


Figure 3 Phylogenetic network of shared mtDNA HVS-I sequence types of recent African ancestry in America, Eurasia, and Africa. The charts show: (a) the proportion in each main region of those specific haplotypes shared by >15 individuals (*yellow frame*), (b) the frequency distribution when the whole haplotypes are considered in each haplogroup with <16 individuals per haplotype (*black frame*), and (c) the frequency distribution of all shared haplotypes for each of the main lineages (L1, L2, and L3) (*blue frame*). In the bottom left corner, the chart shows the frequency distribution of the whole set of shared haplotypes (*gray frame*). Population codes correspond to those listed in the legend of figure 2.

contribution, seems to be the simplest explanation for this pattern. The much less frequent subhaplogroups—L2b, L2c, and L2d—are not found in eastern Africa and are again most likely to be of largely western African provenance in the Americas, although two American types are shared only with southeastern Africans and may, therefore, have their immediate origin there.

L3A types in the Americas are, again, probably of largely western African provenance, with lesser west-central and southeastern African components. L3b and L3d mtDNAs are mainly western African, with a few types shared with eastern, southeastern (and occasionally southern, through interaction between Bantu-speaking and Khoisan-speaking groups), and (for L3b) west-central Africans. The same is true of L3e2, which includes one of the most common and widespread American mtDNA types with a number of derivatives observed only in America, as well as the much less frequent L3e4. One American L3e3 type is shared with the Bubi of Bioko but is more common in Mozambique, with a one-step derived type in the Angolan sample; thus, it most likely derives from west-central or southeastern Africa. Other L3e types are also found in Bioko and neighboring São Tomé, as well as in Angola and Mozambique. American L3e1 lineages, however, largely match southeastern Africans, implying a likely origin either there or somewhere on the west-central coast of Africa, between Cameroon and Angola.

The proportion of L3e in the Angolan sample is 14%, of which 5/6 are members of L3e1, the subclade that is also particularly frequent in Mozambique, where the frequency is similar (Salas et al. 2002). On the basis of the phylogeographic distribution (see fig. 9 of Salas et al. 2002), L3e2 appears to be more northern and to have spread into western Africa, where the derived subclade L3e2b predominates, possibly accounting for its greater prevalence in the Caribbean. The minor subclade L3e4 seems to have a similar distribution. The wide distribution of L3e1 and L3e3 is likely to be due mainly to Bantu dispersals (Salas et al. 2002). However, the 18th-century transfer of slaves from Mozambique and Angola to work in the sugar plantations of São Tomé and probably also Bioko (Iliffe 1995) and their subsequent transfer to the Americas may also help to account for the sharing of mtDNA types across this wide area.

A few haplogroup L3f types in America match those of eastern Africans—again, however, never uniquely; they are also shared by western Africans, southeastern Africans, or both. At the same time, several L3f types are shared uniquely by western Africans only. L3f is likely of eastern African origin (Salas et al. 2002), but the derived subhaplogroup L3f1 is also present in western Africa, and it is this component that is most commonly found in Americans.

Some mtDNA clades are notable for their absence (or virtual absence) in the Americas. Neither of the Khoisan-

specific haplogroups, L1d and L1k, are found in America, nor are the eastern African haplogroups, L1e and L1f. This supports the historical view that neither Khoisan-speaking populations nor eastern Africa contributed significantly to the Atlantic slave trade. However, the eastern African haplogroup L3g is found in several American individuals, implying either a small eastern African influence (Thomas 1998, p. 706), more recent immigration from Africa into America, or hitherto undetected gene flow into western or southeastern Africa and thence to the Americas. There is indeed evidence for low levels of transcontinental gene flow in the modern mtDNA patterns of Africa (for example, “accidental” or “erratic” lineages of western African origin in eastern and southeastern Africa [Bandelt et al. 2001; Salas et al. 2002]). Eastern African L3f types, for example, may also have had an immediate origin in the eastern African component of west-central African Bantu-speaking populations.

Some of the American U6 types (those belonging to U6b1) are most likely of Canarian ancestry and not a direct consequence of the Atlantic slave trade. They probably reflect the genetic admixture of Spanish men with Guanches women after the Spanish colonization of the Canary Islands and the subsequent movements of these admixed populations to the Caribbean islands during the process of Spanish colonization of the region.

African mtDNAs in Eurasia

Eurasian matches to African mtDNAs are also indicated in figure 3, representing <1% of Eurasian mtDNAs. More than half of Eurasian U6 lineages occur in western Iberia (Salas et al. 1998; Pereira et al. 2000), and most of the remainder occur in southwest Asia. Both Portugal and the Near East are regions with known historical gene flow from northern Africa, but both were also centers for the importation of slaves. Eurasian L0–L3A are similarly concentrated in the Near East and Iberia. Eurasian L0a types match or are one-step derivatives of eastern African L0a types and are mainly Near Eastern or southern European. The L1b types can be readily derived from western or northern Africa and are similarly distributed. As in America, it is difficult to assign a provenance to L2a types in Eurasia; two types have a likely southeastern African origin, whereas the remainder could be from either eastern or western Africa. The precise ancestry of Eurasian L3b types is similarly difficult to trace. The eastern African-specific L3f is found mainly in the Near East and southern Europe.

African types in Eurasia, unlike those in America, can therefore be attributed to gene flow from both eastern Africa—perhaps partly via the Arab slave trade (Richards et al. 2003)—and western and southeastern Africa, more likely as a result of the Atlantic slave trade. This latter scenario is made more probable by the surprisingly

high proportion of shared types across all three continents, an observation made more remarkable by the fact that the shared types are not necessarily particularly common ones in Africa. A contribution from the medieval Arab/Berber conquest of Iberia and Sicily is also possible (Semino et al. 1989; Côrte-Real et al. 1996). All of the Iberian L0–L3A types are shared with either western or southeastern Africans, so that it is possible that all of them might be accounted for by the Atlantic slave trade (Pereira et al. 2000). More than a quarter of these are, however, also shared with northern Africans, suggesting that a proportion may alternatively be attributed to the Arab/Berber conquest, especially since the northern African sample size for these types is much smaller than that for western Africa. The ratio of the frequencies of haplogroup U6 to L0–L3A in Iberia (35%) seems too small to be attributed solely to northern African immigration but too large to be accounted for by western African immigration alone ($P = .0013$ and $P < 10^{-5}$, respectively, in a one-tailed Fisher's exact test). Therefore, it seems unlikely that the impact of mtDNAs of recent African ancestry on Iberia was the result of a single process; indeed, earlier gene flow may also have contributed some of the lineages.

Discussion

In this study, we have identified mtDNA lineages with a recent African ancestry in contemporary American populations. Of course, all modern human mtDNAs have what is often described in the literature as “recent African ancestry.” That is to say, Eurasian mtDNAs originated in a founder effect and dispersal of haplogroup L3 lineages (rapidly differentiating into Eurasian haplogroups M and N) from sub-Saharan Africa that probably took place ~60,000–80,000 years ago (Watson et al. 1997; Quintana-Murci et al. 1999). In this article, however, we define “recent African ancestry” as rather the result of gene flow within historical times; in the case of the Americas, this gene flow is largely (but not necessarily exclusively) due to the effects of the Atlantic slave trade.

A recent summary of the historical literature suggests that ~62% of slaves came from western Africa to America (~8 million slaves), ~30% from west-central Africa (4 million), and ~8% from southeastern Africa (1 million) (Thomas 1998, p. 806). The North American and Central American source regions are thought likely to have been mainly in western Africa, supplemented substantially in South America by sources in west-central Africa and, to some extent, southeastern Africa.

The mtDNA composition in America, although indicating only the female line of descent, broadly corroborates historical research. In particular, and unsurprisingly, our admixture estimates, PC plots, and phylogenetic networks all stress the overwhelming

impact of western and west-central Africa on the composition of American mtDNAs with recent African ancestry, with a likely small southeastern African component. Western Africa appears to have been the most important source for North and Central America, although, perhaps surprisingly, North America appears to harbor the larger (but probably still a minority) west-central African component. The results from South America are more problematic, but the likely picture appears to be of a large, possibly majority west-central African contribution, with a substantial western African component as well, at least for Brazil. However, the west-central African contribution most likely derives largely from an area that so far has not been sampled for mtDNA variation, such as the Congo basin. A contribution in Brazil from southeastern Africa also seems likely, as suggested by the high level of matching between Brazilian and Mozambican lineages within L3e1 (fig. 3).

The contributions of northern, eastern, and southern Africa to the American mtDNA gene pool appear to be very small, which, again, is in good agreement with the historical picture. Eastern Africa is, by contrast, an important donor of mtDNAs to the Near East. This may be the result, at least in part, of more ancient movements of slaves, such as the Arab trade through Red Sea and Indian Ocean ports (Richards et al. 2003). However, there are also discernible western and southeastern African components to the (relatively few) mtDNAs of recent African ancestry within Europe, which are likely to be mainly attributable to the more recent Atlantic trade. Portuguese western, southwestern, and southeastern Africa were the main sources for the Atlantic slave trade to Europe (Thomas 1998, p. 805). A striking finding of this study is the high number of three-way sequence matches between African, American, and European mtDNAs. Almost all of these are of likely western African (such as those in L1b) or southeastern African (such as those in L0a) origin. These matches are particularly prevalent in Portugal, which was indeed the principal destination for slaves within Europe. Nevertheless, the composition of Iberian mtDNAs of recent African ancestry suggests that other processes, such as the medieval Arab/Berber conquest, must also have been influential.

Overall, these results show that mtDNAs in America and Eurasia can, in many cases, be traced to broad geographical regions within Africa. This raises the possibility that a greater resolution may be possible in the near future. However, there are many difficulties with such an endeavor. A major problem at the moment is the poor sampling coverage in some parts of the continent, especially central Africa. Nevertheless, even when this is resolved, there are many features of African mtDNA variation that will persist to confound such an undertaking.

In the first place, some of the major African mtDNAs, such as haplogroup L2a, seem to have been widely distributed within the continent in prehistoric times, and individual mtDNA types are often difficult to localize geographically. Another important factor is the effect of the Bantu dispersals, which resulted in a fairly small number of mtDNA types from various haplogroups being widely dispersed quite recently throughout subtropical Africa. Further confounding influences for phylogeographic analyses result from the large-scale and widespread movements of people within the continent itself during the period at which the slave trade was in operation. It is likely that improved phylogenetic resolution, with the aid of more complete mtDNA sequences, will increase the phylogeographic resolution to some extent, but the problems of recent widespread movements will remain. The implication is that, although great progress can be made towards understanding the broad pattern of human dispersals using mtDNA, tracking the lineages of individuals is fraught with problems. Considerable caution is warranted when assessing claims of being able to trace the ancestry of certain American or European lineages to a particular region or population within modern-day Africa.

Acknowledgments

We thank James Walvin for a critical reading of the text. This work was supported by Ministerio de Ciencia y Tecnología grant DGCYT-P4. BIO2000-0145-P4-02 and by Ministerio de Sanidad y Consumo (Fondo de Investigación Sanitaria, Instituto de Salud Carlos III) grant PI030893; SCO/3425/2002. Financial support was also provided by the Italian Ministry of the University (Progetti Ricerca Interesse Nazionale 2002 and 2003) (to A.T., R.S., and A.C.), Progetto CNR-MIUR Genomica Funzionale-Legge 449/97 (to A.T.), Grandi Progetti di Ateneo (to R.S.), and the Instituto Pasteur Fondazione Cenci Bolognetti (to R.S.). A.S. is supported by the Isidro Parga Pondal program (Xunta de Galicia). V.M. was partly supported by a Research Career Development Fellowship from the Wellcome Trust.

Electronic-Database Information

The URLs for data presented herein are as follows:

Arlequin's Home on the Web, <http://lgb.unige.ch/arlequin/> (for Arlequin 2.0: a software for population genetic data analysis)

Fluxus Engineering, <http://www.fluxus-engineering.com/> (for Network 3.0: software for median network constructions)

References

- Alves-Silva J, Santos M, Guimaraes P, Ferreira AC, Bandelt H-J, Pena SD, Prado VF (2000) The ancestry of Brazilian mtDNA lineages. *Am J Hum Genet* 67:444–461 (erratum 67:775)
- Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23:147
- Bandelt H-J, Alves-Silva J, Guimaraes P, Santos M, Brehm M, Pereira L, Coppa A, Larruga JM, Rengo C, Scozzari R, Torroni A, Prata MJ, Amorim A, Prado VF, Pena SDJ (2001) Phylogeography of the human mitochondrial L3e: a snapshot of African prehistory and Atlantic slave trade. *Ann Hum Genet* 65:549–563
- Bandelt H-J, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16:37–48
- Bandelt H-J, Forster P, Sykes BC, Richards MB (1995) Mitochondrial portraits of human populations using median networks. *Genetics* 141:743–753
- Bortolini MC, Zago MA, Salzano FM, Silva WA, Bonatto SL, da Silva MC, Weimer TA (1997) Evolutionary and anthropological implications of mitochondrial DNA variation in African Brazilian populations. *Hum Biol* 69:141–159
- Brake Z, Bosch E, Izaabel H, Akhayat O, Comas D, Bertranpetit J, Calafell F (2001) Human mitochondrial DNA sequence variation in the Moroccan population of the Sous Area. *Ann Hum Biol* 28:295–307
- Brehm A, Pereira L, Bandelt H-J, Prata MJ, Amorim A (2002) Mitochondrial portrait of the Cabo Verde archipelago: the Senegambian outpost of the Atlantic slave trade. *Ann Hum Genet* 66:49–60
- Chakraborty R, Kamboh MI, Nwankwo M, Ferrell RE (1992) Caucasian genes in American blacks: new data. *Am J Hum Genet* 50:145–155
- Chen YS, Olckers A, Schurr TG, Kogelnik AM, Huoponen K, Wallace DC (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
- Chen YS, Torroni A, Excoffier L, Santachiara-Benerecetti AS, Wallace DC (1995) Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. *Am J Hum Genet* 57:133–149
- Côrte-Real HB, Macaulay VA, Richards MB, Hariti G, Issad MS, Cambon-Thomsen A, Papiha S, Bertranpetit J, Sykes BC (1996) Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis. *Ann Hum Genet* 60:331–350
- Curtin PD (1969) The dimensions of the Atlantic slave trade. University of Wisconsin, Madison
- Fage JD (1969) Slavery and the slave trade in the context of west African history. *J Afr Hist* 3:393–404
- Graven L, Passarino G, Semino O, Boursot P, Santachiara-Benerecetti S, Langaney A, Excoffier L (1995) Evolutionary correlation between control region sequence and restriction polymorphisms in the mitochondrial genome of a large Senegalese Mandenka sample. *Mol Biol Evol* 12:334–345
- Green LD, Derr JN, Knight A (2000) mtDNA affinities of the peoples of North-Central Mexico. *Am J Hum Genet* 66:989–998
- Handt O, Meyer S, von Haeseler A (1998) Compilation of human mtDNA control region sequences. *Nucleic Acids Res* 26:126–129
- Helgason A, Nicholson G, Stefánsson K, Donnelly P (2003)

- A reassessment of genetic diversity in Icelanders: strong evidence from multiple loci for relative homogeneity caused by genetic drift. *Ann Hum Genet* 67:281–297
- Illife J (1995) *Africans: the history of a continent*. Cambridge University Press, Cambridge
- Krings M, Salem AE, Bauer K, Geisert H, Malek AK, Chaix L, Simon C, Welsby D, Di Rienzo A, Utermann G, Sajantila A, Pääbo S, Stoneking M (1999) mtDNA analysis of Nile River Valley populations: a genetic corridor or a barrier to migration? *Am J Hum Genet* 64:1166–1176
- Lewis B (1990) *Race and slavery in the Middle East: an historical enquiry*. Oxford University Press, Oxford
- Macaulay V, Richards M, Hickey E, Vega E, Cruciani F, Guida V, Scozzari R, Bonnè-Tamir B, Sykes B, Torroni A (1999) The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *Am J Hum Genet* 64:232–249
- Mateu E, Comas D, Calafell F, Pérez-Lezaun A, Abade A, Bertranpetit J (1997) A tale of two islands: population history and mitochondrial DNA sequence variation of Bioko and São Tomé, Gulf of Guinea. *Ann Hum Genet* 61:507–518
- Mesa NR, Mondragon MC, Soto ID, Parra MV, Duque C, Ortiz-Barrientos D, Garcia LF, Velez ID, Bravo ML, Munera JG, Bedoya G, Bortolini MC, Ruiz-Linares A (2000) Autosomal, mtDNA, and Y-chromosome diversity in Amerinds: pre- and post-Columbian patterns of gene flow in South America. *Am J Hum Genet* 67:1277–1286
- Mishmar D, Ruiz-Pesini E, Golik P, Macaulay V, Clark AG, Hosseini S, Brandon M, Easley K, Chen E, Brown MD, Sukernik RI, Olckers A, Wallace DC (2003) Natural selection shaped regional mtDNA variation in humans. *Proc Natl Acad Sci USA* 100:171–176
- Monsalve MV, Hagelberg E (1997) Mitochondrial DNA polymorphisms in Carib people of Belize. *Proc R Soc Lond B Biol Sci* 264:1217–1224
- Parra EJ, Kittles RA, Argyropoulos G, Pfaff CL, Hiester K, Bonilla C, Sylvester N, Parrish-Gause D, Garvey WT, McKeigue PM, Kamboh MI, Ferrell RE, Pollitzer WS, Shriver MD (2001) Ancestral proportions and admixture dynamics in geographically defined African Americans living in South Carolina. *Am J Phys Anthropol* 114:18–29
- Parra EJ, Marcini A, Akey J, Martinson J, Batzer MA, Cooper R, Forrester T, Allison DB, Deka R, Ferrell RE, Shriver MD (1998) Estimating African American admixture proportions by use of population-specific alleles. *Am J Hum Genet* 63:1839–1851
- Pereira L, Macaulay V, Torroni A, Scozzari R, Prata MJ, Amorim A (2001) Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade. *Ann Hum Genet* 65:439–458
- Pereira L, Prata MJ, Amorim A (2000) Diversity of mtDNA lineages in Portugal: not a genetic edge of European variation. *Ann Hum Genet* 64:491–506
- Pinto F, González AM, Hernandez M, Larruga JM, Cabrera VM (1996) Genetic relationship between the Canary Islanders and their African and Spanish ancestors inferred from mitochondrial DNA sequences. *Ann Hum Genet* 60:321–330
- Quintana-Murci L, Semino O, Bandelt H-J, Passarino G, McElreavey K, Santachiara-Benerecetti AS (1999) Genetic evidence of an early exit of *Homo sapiens sapiens* from Africa through eastern Africa. *Nat Genet* 23:437–441
- Rando JC, Cabrera VM, Larruga JM, Hernández M, González AM, Pinto F, Bandelt H-J (1999) Phylogeographic patterns of mtDNA reflecting the colonization of the Canary Islands. *Ann Hum Genet* 63:413–428
- Rando JC, Pinto F, González AM, Hernández M, Larruga JM, Cabrera VM, Bandelt H-J (1998) Mitochondrial DNA analysis of northwest African populations reveals genetic exchanges with European, Near-Eastern, and sub-Saharan populations. *Ann Hum Genet* 62:531–550
- Richards M, Rengo C, Cruciani F, Gratrix F, Wilson JF, Scozzari R, Macaulay V, Torroni A (2003) Extensive female-mediated gene flow from sub-Saharan Africa into near eastern Arab populations. *Am J Hum Genet* 72:1058–1064
- Salas A, Comas D, Lareu MV, Bertranpetit J, Carracedo A (1998) mtDNA analysis of the Galician population: a genetic edge of European variation. *Eur J Hum Genet* 6:365–375
- Salas A, Richards M, De la Fé T, Lareu MV, Sobrino B, Sánchez-Diz P, Macaulay V, Carracedo A (2002) The making of the African mtDNA landscape. *Am J Hum Genet* 71:1082–1111
- Sans M (2000) Admixture studies in Latin America: from the 20th to the 21st century. *Hum Biol* 72:155–177
- Santos MV, Abreu A, Mendes C, Carvalho M, Anjos MJ, Andrade L, Lopes V, Côrte-Real F, Vieira DN, Vide MC (2002) Polimorfismos na sequência do DNA mitochondrial nas populações da Bahia-Brasil e de Angola. Paper presented at the Congress of the GEP-ISFG (Spanish and Portuguese Group of the International Society of Forensic Genetics), Barcelona, June 4–7
- Semino O, Torroni A, Scozzari R, Brega A, De Benedictis G, Santachiara-Benerecetti AS (1989) Mitochondrial DNA polymorphisms in Italy. III. Population data from Sicily: a possible quantitation of maternal African ancestry. *Ann Hum Genet* 53:193–202
- Thomas H (1998) *The slave trade—the history of the Atlantic slave trade: 1440–1840*. Macmillan, London
- Thomas MG, Weale ME, Jones AL, Richards M, Smith A, Redhead N, Torroni A, Scozzari R, Gratrix F, Tarekegn A, Wilson JF, Capelli C, Bradman N, Goldstein DB (2002) Founding mothers of Jewish communities: geographically separated Jewish groups were independently founded by very few female ancestors. *Am J Hum Genet* 70:1411–1420
- Torroni A, Rengo C, Guida V, Cruciani F, Sellitto D, Coppa A, Calderon FL, Simionati B, Valle G, Richards M, Macaulay V, Scozzari R (2001) Do the four clades of the mtDNA haplogroup L2 evolve at different rates? *Am J Hum Genet* 69:1348–1356
- Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC (1991) African populations and the evolution of mitochondrial DNA. *Science* 253:1503–1507
- Watson E, Bauer K, Aman R, Weiss G, von Haeseler A, Pääbo S (1996) mtDNA sequence diversity in Africa. *Am J Hum Genet* 59:437–444
- Watson E, Forster P, Richards M, Bandelt H-J (1997) Mitochondrial footprints of human expansions in Africa. *Am J Hum Genet* 61:691–704