Am. J. Hum. Genet. 77:676–680, 2005

Report

Charting the Ancestry of African Americans

Antonio Salas,[1] Ángel Carracedo,[1] Martin Richards,[2] and Vincent Macaulay[3]

[1]Unidade de Xenética, Instituto de Medicina Legal, Facultad de Medicina, Universidad de Santiago de Compostela, and Centro Nacional de Genotipado, Hospital Clínico Universitario, Galicia, Spain; [2]School of Biology, University of Leeds, Leeds, United Kingdom; and [3]Department of Statistics, University of Glasgow, Glasgow

The Atlantic slave trade promoted by West European empires (15th–19th centuries) forcibly moved at least 11 million people from Africa, including about one-third from west-central Africa, to European and American destinations. The mitochondrial DNA (mtDNA) genome has retained an imprint of this process, but previous analyses lacked west-central African data. Here, we make use of an African database of 4,860 mtDNAs, which include 948 mtDNA sequences from west-central Africa and a further 154 from the southwest, and compare these for the first time with a publicly available database of 1,148 African Americans from the United States that contains 1,053 mtDNAs of sub-Saharan ancestry. We show that >55% of the U.S. lineages have a West African ancestry, with <41% coming from west-central or southwestern Africa. These results are remarkably similar to the most up-to-date analyses of the historical record.

The number of people of recent African ancestry living in the Americas is comparable to that in Africa itself, largely because of the forced mass migrations of the Atlantic slave trade during the 15th–19th centuries. Historical documentation indicates that many Africans were taken from the West African coast but that a large proportion were also taken from the former Portuguese colonies of Angola and Mozambique. The perpetrators included the Portuguese, Spanish, and British, and it is estimated that they moved ~12 million people (with ~10.5 million surviving the voyages) in <4 centuries (Eltis et al. 1998).

There has been a growing interest in the historical roots of African Americans and other communities of recent African ancestry, and genetics can play an important role in the study of that ancestry. However, much of the work to date has focused either on the extent of the African contribution to the American gene pool (Chakraborty et al. 1992; Parra et al. 1998, 2001; Mesa et al. 2000; Sans 2000) or on particular mtDNA lineages carried to the Americas through the Atlantic slave trade (Alves-Silva et al. 2000; Bandelt et al. 2001; Pereira et al. 2001; Salas et al. 2002). Despite widespread public interest and the rapidly growing involvement of private companies in genetic ancestry testing, the only published attempt to date to assign African geographical ancestry to African American lineages by use of genetic data has been by Salas et al. (2004). This latter work suggested a majority contribution of West Africa to the mtDNA pool of North and Central Americans of recent African descent, with a slightly smaller component from west-central Africa, whereas African Americans from South America (Brazil) reversed the picture, with a larger west-central African component. No significant component was detected from other regions, including Mozambique, despite the historical evidence of the movement of ~1 million slaves from that region (Thomas 1998). Eurasian mtDNAs of recent African ancestry were found to be made up mainly of west-central and North African components, with a larger East African component in the Near East.

However, the analysis of Salas et al. (2004) had small sample sizes, particularly the small sample size from west-central Africa and the complete lack of samples from southwestern Africa, a known source of an estimated 3 million slaves (Thomas 1998). Here, we update the analysis, using 948 sequences from west-central Africa and a further 154 from Angola and Cabinda, as well as sub-

Received June 10, 2005; accepted for publication July 27, 2005; electronically published August 11, 2005.

Address for correspondence and reprints: Dr. Antonio Salas, Unidade de Xenética, Instituto de Medicina Legal, Facultad de Medicina, Universidad de Santiago de Compostela, Calle San Francisco, 15705 Santiago de Compostela, A Coruña, Galicia, Spain. E-mail: apimlase@usc.es

0002-9297/2005/7704-0017$15.00

stantially enhanced data sets from the other regions of Africa (table 1) and a database of African American lineages that is more than an order of magnitude larger than that used earlier (Monson et al. 2002; see also the work of Bandelt et al. [2004*a*, 2004*b*] and Budowle et al. [2004]). This allows us to perform a much more decisive test of African ancestry among African American maternal lineages than has previously been possible.

Africa is the most genetically diverse continent. A fine subdivision of African mtDNA lineages provides a powerful source of phylogeographic information: major regions of the continent display markedly different frequencies of the continent-specific mtDNA clades, or haplogroups (fig. 1*a*). However, the first point to make from this enhanced data set is the obvious similarity of the haplogroup frequency profiles of West Africa, west-central Africa, and southwestern Africa in comparison with the other major regions of the continent. Although there are certain differences (which become more obvious in a finer classification of the lineages than at the gross haplogroup level), these regions display a broad similarity to each other, with most of the major haplogroups showing a gradient through the three regions, with (unsurprisingly) a much closer similarity of west-central and southwestern Africa compared with West Africa.

The African American mtDNA sample from the United States is striking for its strong resemblance to the mtDNA composition of all three of these regions (fig. 1*b*). Principal-components analysis (PCA) graphically portrays the same pattern: in a scatter plot of the first two components (which account for 77% of the variance of haplogroup frequencies), African American mtDNA is almost indistinguishable from that of these three African regions (fig. 1*c*).

This picture can be quantified by fitting a linear model to the frequency profiles, as done by Salas et al. (2004). Our analysis (table 2) indicated, as before, that North Africa, East Africa, southern Africa, and even southeastern Africa (Mozambique) show no significant contribution to the African American mtDNA pool in the United States. Thus, the larger data set here offers no support for the weak signal from southeastern Africa detected elsewhere (Salas et al. 2004).

More significantly, we are now able to substantially revise and update the estimates for the main contributing regions. We performed the analysis in several ways. In the first, low-resolution analysis, for which we used the 15 haplogroups that are illustrated in figure 1*a* and were used for the PCA, West Africa contributed 56.0% ± 4.2%, southwestern Africa 27.8% ± 6.6%, and west-central Africa 13.5% ± 8.3% (table 2). Although the last is not significantly different from zero, the admixture values for the west-central and southwestern samples are highly negatively correlated (Pearson correlation coefficient of −0.9), indicating that the large SD of the posterior distribution is most plausibly explained by the similarity of haplogroup frequency profiles of our west-central and southwestern subdivisions. A reanalysis that combined these two similar regions (and also excluded the other uninformative regions of Africa) yielded values of 52.5% ± 4.1% for West Africa and 47.5% ± 4.1% for west-central Africa.

Table 1

African and African American Samples Used in the Present Study

The table is available in its entirety in the online edition of *The American Journal of Human Genetics*.

For a high-resolution analysis, to decrease the SD of the estimates, we divided the lineages into 56 haplogroups on phylogenetic grounds, such that no haplogroup included <20 samples. In this case, West Africa contributed 60% ± 3%, southwestern Africa 30% ± 2%, and west-central Africa 9% ± 3% (table 2). In this case, the correlation between west-central and southwestern African values was reduced to −0.6.

These values can be compared with classic estimates provided >30 years ago by Curtin (1969, p. 157) and with more up-to-date estimates. Curtin (1969) estimated that 50.4% of North American slaves came from West Africa and 47.8% from west-central Africa, with just 1.6% from southeastern Africa. A much more recent estimate for the United States (McMillin 2004, pp. 48, 54, and 70) suggests that ~64% were from West Africa, ~35% from west-central Africa, and ~1% from southeastern Africa. Interestingly, our values are closer to the more recent estimate. Given the uncertainties involved and that the admixture estimates would most likely be affected by demographic factors subsequent to the arrival in the Americas, these values derived from entirely different methodologies are strikingly similar, corroborating both lines of evidence.

Figure 1*a* illustrates that characteristic West African (L1b; L2b,c,d; and L3b,d) and west-central African (L1c and L3e) haplogroups have been carried into the Americas at high frequencies. Typically, eastern African lineages (e.g., within haplogroups L3*, L1f*, and L1g) are poorly represented in America, again mirroring the historically documented low contribution of East Africa to the slave trade. The small impact of North Africa is evident by the low frequency (<1%) of its autochthonous haplogroup U6. Haplogroups that characterize southern African Khoisan-speaking groups (L0d and L0k) are completely absent in America.

Of course, these results indicate only the maternal component of the American gene pool. Further research on Y-chromosome and autosomal markers is clearly needed to test whether the contribution of females and males to the African American gene pool has been symmetrical or

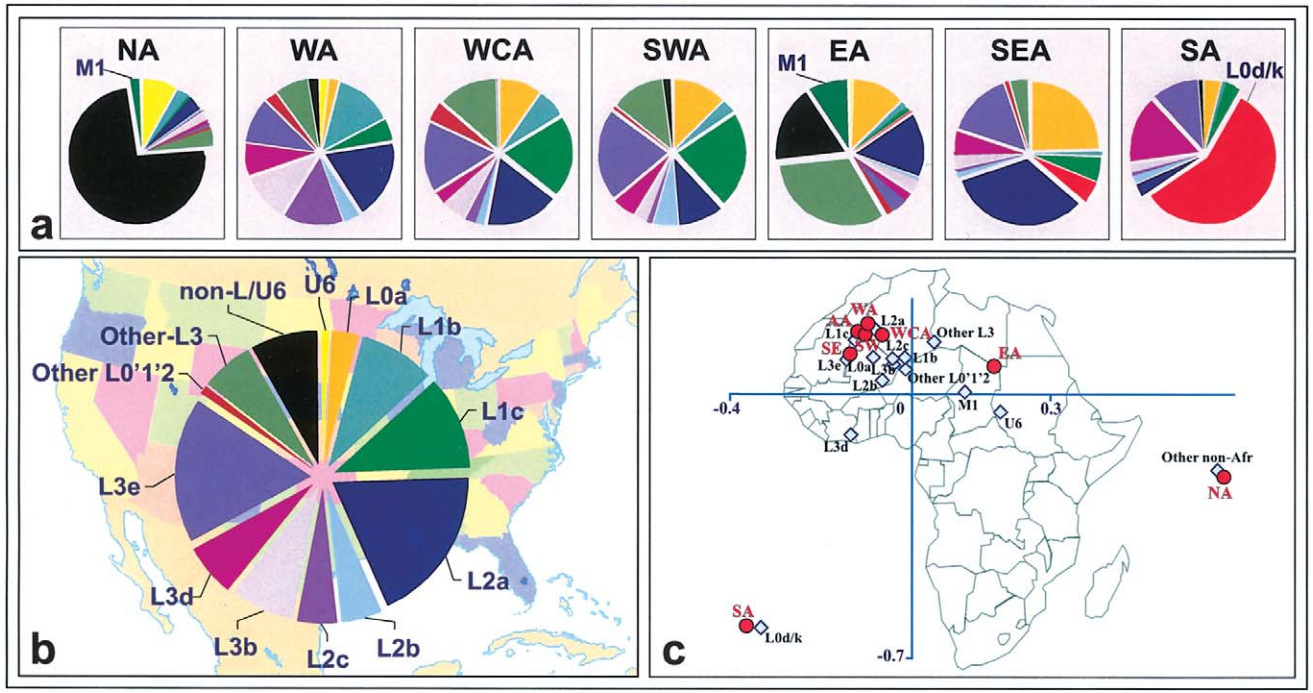

Figure 1 *a,* The mtDNA haplogroup distribution in Africa regions (NA = North Africa; WA = West Africa; WCA = west-central Africa; SWA = southwestern Africa; EA = East Africa; SEA = southeastern Africa; SA = southern Africa). Each sequence was assigned to 1 of 15 haplogroups (Salas et al. 2002). *b,* The mtDNA haplogroup distribution in African Americans. *c,* Scatter plot of the first two principal components of the African American and the African regional haplogroup frequency profiles.

Table 2

Estimated Admixture Coefficients (of Source African Regions) for African Americans

African Region	Sample Size	Posterior Mean and Root-Mean-Square Deviation[a]	
		Low Resolution[b]	High Resolution[c]
North Africa	978	.001 ± .001	.001 ± .001
West Africa	1,549	.560 ± .042	.599 ± .028
West-central Africa	855	.135 ± .083	.090 ± .027
Southwestern Africa	157	.278 ± .066	.296 ± .024
East Africa	717	.026 ± .003	.002 ± .002
Southeastern Africa	416	.021 ± .018	.009 ± .008
Southern Africa	266	.002 ± .002	.002 ± .002

[a] The posterior mean and root-mean-square deviation of the admixture coefficient for each African region.

[b] Values are based on a division of haplotypes into the 15 haplogroups shown in figure 1.

[c] Values are based on a division of haplotypes into 56 haplogroups.

asymmetrical—research for which, again, large African databases of such marker systems will be needed, such as are now becoming available for mtDNA. With the acceleration of data acquisition now under way, these analyses will no doubt be possible in the near future.

We conclude that mtDNA variation allows us to trace the maternal ancestry of African Americans to broad geographic regions of Africa, with results that are closely concordant with historical studies that now encompass documentation for between two-thirds and three-quarters of the estimated total voyages made during the course of the Atlantic slave trade (Eltis et al. 1998). We have previously raised the possibility of whether, with larger data sets and extensive phylogeographic analyses, more-specific reconstructions will be possible (Salas et al. 2004). However, even with this substantially augmented data set, we note that it is still not possible to go further at this stage. Even with greatly improved geographic coverage, it remains the case that many mtDNAs are very widely distributed throughout the African continent, most likely as a result largely of the Bantu dispersals (Salas et al. 2002), but no doubt also as a result of both earlier and more recent movements, including those that are due to the Atlantic slave trade itself (Salas et al. 2004). This problem will continue to hamper the allocation of African American mtDNAs to narrower geographic locations in Africa, even if the resolution of the molecular analyses is increased from the first hypervariable segment (HVS-I) to complete mtDNA genomes.

Considerable caution is therefore warranted when dealing with claims in the popular media (such as the acclaimed and prestigious BBC television documentary *Motherland: A Genetic Journey,* first shown in the United Kingdom in 2003) and those made by genetic ancestry–testing companies about their ability to trace the ancestry of certain American (or, for that matter, European) mtDNAs to a particular locale or population within modern-day Africa. Our analyses stand as a warning to unsuspecting members of the public who may be seduced by such promises.

Acknowledgments

We thank David Richardson and Mike Turner for advice on the historical evidence. This work was supported by grants from the Ministerio de Sanidad y Consumo (Fondo de Investigación Sanitaria [PI030893]) and Genoma España, Centro Nacional de Genotipado.

References

Alves-Silva J, da Silva Santos M, Guimarães PEM, Ferreira ACS, Bandelt H-J, Pena SD, Prado VF (2000) The ancestry of Brazilian mtDNA lineages. Am J Hum Genet 67:444–461 (erratum 67:775)

Bandelt H-J, Alves-Silva J, Guimarães 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, Salas A, Bravi C (2004*a*) Problems in FBI mtDNA database. Science 305:1402–1404

Bandelt H-J, Salas A, Lutz-Bonengel S (2004*b*) Artificial recombination in forensic mtDNA population databases. Int J Legal Med 118:267–273

Budowle B, Polanskey D, Allard MW, Chakraborty R (2004) Addressing the use of phylogenetics for identification of sequences in error in the SWGDAM mitochondrial DNA database. J Forensic Sci 49:1256–1261

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

Curtin PD (1969) The Atlantic slave trade: a census. University of Wisconsin Press, Madison

Eltis D, Behrendt SD, Richardson D, Klein HS (1998) The transatlantic slave trade 1527–1867: a database prepared at the W. E. B. Du Bois Institute, Harvard University. Cambridge University Press, New York

McMillin JA (2004) The final victims: foreign slave trade to North America 1783–1810. University of South Carolina Press, Columbia

Mesa NR, Mondragón MC, Soto ID, Parra MV, Duque C, Ortiz-Barrientos D, Garcia LF, Velez ID, Bravo ML, Múnera JG, Bedoya G, Bortolini M-C, 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

Monson KL, Miller KWP, Wilson MR, Dizinno JA, Budowle B (2002) The mtDNA population database: an integrated software and database resource for forensic comparison. Forensic Sci Comm 4, no 2 (http://www.fbi.gov/hq/lab/fsc/backissu/april2002/miller1.htm) (accessed August 11, 2005)

Parra EJ, Kittles RA, Argyropoulos G, Pfaff CL, Hiester K, Bonilla C, Sylvester N, Parrish-Gause D, Garvey WT, Mc-

Keigue 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, Amorin 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

Richards M, Macaulay V, Hill CH., Carracedo Á, Salas A (2004) The archaeogenetics of the dispersals of the Bantu-speaking peoples. In: Jones M (ed) Traces of ancestry: studies in honour of Colin Renfrew. McDonald Institute for Archaeological Research, Cambridge, United Kingdom, pp 75–87

Salas A, Richards M, De la Fe T, Lareu M-V, Sobrino B, Sánchez-Diz P, Macaulay V, Carracedo A (2002) The making of the African mtDNA landscape. Am J Hum Genet 71:1082–1111

Salas A, Richards M, Lareu M-V, Scozzari R, Coppa A, Torroni A, Macaulay V, Carracedo A (2004) The African diaspora: mitochondrial DNA and the Atlantic slave trade. Am J Hum Genet 74:454–465

Sans M (2000) Admixture studies in Latin America: from the 20th to the 21st century. Hum Biol 72:155–177

Thomas H (1998) The slave trade—the history of the Atlantic slave trade: 1440–1870. Macmillan, London