

The Phylogeography of African Brazilians

Vanessa F. Gonçalves^a Cláudia M.B. Carvalho^a Maria Cátira Bortolini^b
Sérgio P. Bydlowski^c Sérgio D.J. Pena^a

^aDepartamento de Bioquímica e Imunologia, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais,

^bDepartamento de Genética, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, and

^cDivisão de Pesquisa e Biologia Molecular, Fundação Pró-Sangue Hemocentro de São Paulo and Departamento de Hematologia, LIM-31, Faculdade de Medicina da Universidade de São Paulo, São Paulo, SP, Brazil

Key Words

MtDNA · Y chromosome · Africa · Slavery · Atlantic Slave Trade · Brazil · Admixture

Abstract

Background/Aims: Approximately four million Africans were taken as slaves to Brazil, where they interbred extensively with Amerindians and Europeans. We have previously shown that while most White Brazilians carry Y chromosomes of European origin, they display high proportions of African and Amerindian mtDNA lineages, because of sex-biased genetic admixture. **Methods:** We studied the Y chromosome and mtDNA haplogroup structure of 120 Black males from Sao Paulo, Brazil. **Results:** Only 48% of the Y chromosomes, but 85% of the mtDNA haplogroups were characteristic of sub-Saharan Africa, confirming our previous observation of sexually biased mating. We mined literature data for mtDNA and Y chromosome haplogroup frequencies for African native populations from regions involved in Atlantic Slave Trade. Principal Components Analysis and Bayesian analysis of population structure revealed no genetic differentiation of Y chromosome marker frequencies between the African regions. However, mtDNA examination unraveled considerable genetic structure, with three clusters at Central-West Africa, West Africa and Southeast Africa. A hypothesis is proposed to explain this structure. **Con-**

clusion: Using these mtDNA data we could obtain for the first time an estimate of the relative ancestral contribution of Central-West (0.445), West (0.431) and Southeast Africa (0.123) to African Brazilians from Sao Paulo. These estimates are consistent with historical information.

Copyright © 2008 S. Karger AG, Basel

Introduction

It is estimated that from 1550 to 1870, approximately four million Africans were taken to Brazil as slaves [1, 2]. Together with Amerindians and Europeans, Africans represent one of three major ancestral roots for the formation of Brazilian people. The impact of this extensive tripartite admixture can be assessed by the numbers of the 2000 Brazilian census: among 169,872,856 Brazilians, 53.7% were Whites, 38.5% were Brown and 6.2% were Blacks.

In reality the extent of genetic admixture was much larger than the number of Brown individuals might suggest. In a previous study based on mitochondrial DNA (mtDNA) sequencing [3] we showed that taking Brazil as a whole, 28% of individuals self-declared as White carried mtDNA African haplogroups and 33% had mtDNA Amerindian haplogroups. These data were further corroborated by estimates of admixture made with autosomal polymorphisms, which showed that 87% of Brazil-

ians (circa 146 million people) have more than 10% African genomic ancestry [4]. Indeed, our studies demonstrated that in Brazil, at an individual level, color, as determined by physical evaluation, was a poor predictor of genomic European or African ancestry as estimated by molecular markers [5, 6].

Considering the gigantic importance of Africans in the formation of the Brazilian people, it is of great historical, anthropological, linguistic and social importance to know their exact origins in Africa. Throughout the 19th and 20th centuries there was considerable debate and controversy about the relative importance of West African (Senegambia, Bight of Benin, Bight of Biafra, Windward Coast, Gold Coast and Sierra Leone) and West-Central African (Angola, Cabinda, Congo) origins of Afro-Brazilians. More recently, new historical data have revealed that Southeast Africa (Mozambique) was also a quite significant, albeit smaller, third geographical origin [7, 8].

There are several reasons why the question of the comparative significance of the three regions cannot be adequately settled solely using historical and demographic tools. First, the historical record is incomplete: to avoid claims of compensation from slave owners, the Brazilian government ordered, soon after the abolition of slavery in 1888, that all historical records of slave trade be burned [9]. The records of embarkation of ships from Africa to Brazilian ports in all time periods can be obtained from a comprehensive database [10], but only for 1.3 million slaves. These data show that West-Central Africa, West Africa and Southeast Africa were respectively responsible for 73.2, 9.5 and 17.3% of the total number of slaves embarked. As we change the data sources and the time period, the numbers change considerably. For instance, Klein [2] demonstrated that between 1701 and 1810, 68% of the slaves arriving in Brazil were from West-Central Africa and 32% from West Africa. However, in the period 1811–1830, the proportion of slaves from Southeast Africa rose to roughly 20% of the total, because the British anti-slave movement restricted trade from the Atlantic coast of Africa [7, 8]. There are almost no records for the huge number of slaves were brought illegally to Brazil after 1830, when Brazil signed a treaty with England to stop the Atlantic trade.

Second, the ports of embarkation do not reflect accurately the true geographical origin of the slaves, since they were captured from broad areas in Africa, from the coast to hundreds of miles inland [2]. Finally, we have to consider the place of disembarkation: the data clearly show that a larger proportion of slaves from West Africa landed in Bahia (Northeast Brazil) than in Rio de Janeiro [2]. These regional disparities were partially erased in the pe-

riod between the true ending of the slave trade (1850) and the abolition of slavery in Brazil (1888) because of considerable internal migration of slaves, especially from Bahia to the coffee plantations of the state of São Paulo [11].

One possible empirical approach to obtain estimates of the relative African regional importance is the use of molecular data. In 1992 Zago et al. [12] published the first such study, analyzing 74 β^S chromosomes from sickle cell anemia patients in São Paulo and discovered that only three of the five known β^S haplotypes could be observed: 73% of the chromosomes had the Bantu haplotype (also known as the Central Africa Republic type), while the remaining had haplotypes characteristic of West Africa (26% had the Benin haplotype and 1% had the Senegal haplotype). The authors attributed the Bantu haplotype predominance exclusively to a West-Central African origin, not taking into account the fact that many of the Bantu type β^S chromosomes could have originated from Southeast Africa, where this haplotype also predominates [13].

Thus, what we really need are molecular markers able to discriminate between the populations of West Africa, West-Central Africa and Southeast Africa. In the present article we describe the investigation of Black Brazilians from the city of São Paulo using two distinct lineage markers – the matrilineal human mitochondrial DNA (mtDNA) and the patrilineal non-recombinant portion of the Y chromosome (NRY) lineages. Our objective was to ascertain whether these markers could be useful in establishing the African geographical origin of African Brazilians from São Paulo. We here report that mtDNA data are particularly valuable by permitting estimates of the relative contribution from the three major regional geographical origins of Brazilian slaves in Africa.

Material and Methods

DNA Samples

We studied 120 DNA samples from Black Brazilian males from São Paulo, Brazil. For classification as Blacks, the individuals were evaluated by joint phenotypic and genealogical criteria as follows: subjects were asked about their color group and those of their parents and grand-parents, according to their own definition. Phenotypic analysis (facial characteristic and skin pigmentation in the axilla, a body region not exposed to the sun) was performed by the interviewer. Subjects were classified as Black when the characteristic phenotype was present and subjects described themselves, their parents and all their grand-parents as Blacks. We feel that with this methodology, color classification became as free as possible from subjective biases. All studies were anonymous, and the participants provided written consent. DNA was prepared from blood samples using a salting out technique.

Amplification and Sequencing of the Mitochondrial DNA (mtDNA) Control Region

PCR amplification of the hypervariable sequence I (HVS-I) of mtDNA was performed using primers L15989 and H16410 as described previously [3]. In some cases we also analyzed and sequenced the HVS-II region using primers L29 and H580. Amplified segments were precipitated with polyethylene glycol 8000 (20% w/v in 2.5 mM NaCl) and then sequenced with the DYEnamic™ ET dye terminator kit (GE Healthcare) using the same primers employed for the first amplification step (L15989/H16410 for HVS-I and L29/H580 for HVS-II). The sequence products were run on a MegaBace 1000 sequencer (GE Healthcare) and analyzed with the Sequence Analyzer v.3.0 program (GE Healthcare). Haplogroup-specific SNPs in the coding region of mtDNA were analyzed by PCR amplification and restriction enzyme digestion as described previously [3].

The information provided by HVS-I and RFLP analysis were used to sort the sequences into haplogroups. In the analysis we rigorously followed the quality control measures recommended by Bandelt et al. [14]. In order to compare the present results with those from Africans populations, we used data for HVS-I (positions 16065–16365) from a number of Sub-Saharan population groups found in references 15–37.

The mtDNA haplogroup nomenclature used was that of Salas et al. [34] with the additional renaming of haplogroup L1e as L5a1 [38] and haplogroup L3g as L7 [39].

Y Haplogroup Analysis

The Y chromosome haplotype analyses were done following the hierarchical order proposed by the Y Chromosome Consortium [40] updated by Jobling and Tyler-Smith [41]. The UEPs (Unique Event Polymorphisms) used here were specifically chosen to provide high resolution analysis of African haplogroups, especially haplogroup E: M145, PN2 (also known as P2), M174, sY81 (also known as M2), M33, M35, M75, M58, M154, M155, M149, M191, M116, M10, M78, M281, M123, M81, V6, M213, M60 [39]. Except for M154 and M155, all UEPs were analyzed using a minisequencing protocol. The principles of the methodology are described in details elsewhere [40] and the details are given below.

Before performing the minisequencing protocol it was necessary to perform the multiplex amplification of the specific regions where the SNPs were located. The primer sequences used and the PCR product size are listed in supplementary table 1 (www.karger.com/doi/10.1159/000106059). The Y chromosome sequence flanking regions for each marker were kindly provided by Dr. Peter A. Underhill from the Department of Genetics, Stanford University. Y chromosome SNPs were resolved using either multiplex or simplex PCR reactions (supplementary table 1). Primers to the flanking regions of the following SNPs composed the first multiplex (YD/Ea): M145, PN2, M174, sY81. PCR reactions were done in a volume of 12.5 µl with 50 ng genomic DNA, 0.4 µM each primer, 2.0 mM MgCl₂, 1 × AmpliTaq Gold buffer, 200 µM dNTPs, 1 unit of Taq DNA Polymerase (AmpliTaq Gold; Applied Biosystems, Foster City, Calif., USA). The second multiplex, (YD/Eb) was composed of the primers for M33, M35, M75 primers. PCR reactions were done using 0.32 µM (M33 and M35) and 0.4 µM (M75) of each primer. The third multiplex (YE3a) was composed of the following SNPs: M10, M58, M191, M116 and used 0.4 µM of each primer. M149 (simplex E3a3) was amplified alone using

0.4 µM of each primer. The fourth multiplex (YE3b) was composed of M78, M81, M281, M123 and V6. The PCR reactions used 0.4 µM (M78, M81, V6) and 0.6 µM (M123, M281) of each primer. Amplification parameters were 95°C for 5 min, followed by 29 cycles for 30 s at 95°C, 60 s at 50°C, 1 min at 72°C, and a final extension for 10 min at 72°C.

All samples were first typed with the multiplex YD/Ea. Samples with derived allele in M145 and PN2 markers were tested further with YD/Eb and YE3b, while samples with the derived allele at both M145 and sY81 were tested further with YE3a. Samples typed as having the ancestral allele at M145 were then tested sequentially with M60 and M213, both in simplex protocols.

Multiplex and simplex DNA minisequencing was performed according to Carvalho and Pena [40]. The following FAM-labeled dideoxynucleotides were used in the minisequencing: ddGTP for YD/Ea, YD/Eb, YE3b and M213 and ddATP for YE3a, M149 and M60. All minisequencing primers were used in 0.4 µM, except: sY81 (0.2 µM), M174 (0.04 µM) and M116 (0.08 µM). The minisequencing products were applied in a 6% denaturing polyacrylamide gel and resolved in a fluorescent automatic DNA sequencer ALF (GE Healthcare). The analyses were done in the software package Allelinks (GE Healthcare).

M154 and M155 were typed by means of the ARMS (Amplification Refractory Mutation System) technique. Common primers to M154 are the same used to amplify M81 (supplementary table 1) giving rise to a 248-bp amplicon and allele-specific primers were M154An (ACAAATTAGTGCGACACAAC) for the C ancestral allele and M154De (ACAAATTAGTGCGACACAAT) for the T-derived allele, both producing a 112-bp segment in the presence of common primers. Common primers to M155 were the same used to amplify M58 (table 1) giving rise to a 202-bp amplicon and allele-specific primers were M155An (GAGGAATCCTCACCTAACG) for the ancestral allele and M155De (GAGGAATCCTCACCTAACA) for the A-derived allele, both producing a 93-bp segment in the presence of common primers. The PCR products were resolved in 10% polyacrylamide gels and stained with silver. The *DYS199* T allele that defines the founder Amerindian haplogroup Q3* was typed according to Lell [43].

Data Analysis

The frequencies of mtDNA haplogroups were estimated by direct counting and genetic diversity was estimated using the Arlequin 2.0 software [44]. Analysis of Principal Component was performed using the software Statistica 6.0.

Bayesian analysis of genetic differentiation was performed as described by Corander et al. [45, 46] using their BAPS 2.2 software. A major advantage of this software is that the number of distinct populations is treated as an unknown parameter. That means that if the program perceives that two populations, because of high degree of gene flow or recent foundation, can be considered a single panmictic population it will then combine them [45]. All group combinations are considered a priori equally likely. From the data the software then establishes the most likely partition among the population groups in a certain number of empirically plausible clusters.

To calculate admixture of haplotype frequency data we used the maximum likelihood program LEADMIX [47] that estimates admixture proportions taking genetic drift data into account. The program also provides estimates of the method of moments of Roberts and Hiorns [48].

Table 1. Mitochondrial DNA haplogroups observed in Black Brazilian males from São Paulo, Brazil

Haplogroup	N	%
Sub-Saharan African	102	0.85
L0a1	1	0.01
L0a1a	8	0.07
L0a2	4	0.03
L1b	5	0.04
L1b1	4	0.03
L1c	5	0.04
L1c1	8	0.07
L1c2	4	0.03
L2a	8	0.07
L2a1	9	0.08
L2a1a	3	0.03
L2b	2	0.02
L2b1	4	0.03
L2c2	1	0.01
L2d2	1	0.01
L3b	6	0.05
L3b1	1	0.01
L3b2	3	0.03
L3d	2	0.02
L3d1	1	0.01
L3d2	1	0.01
L3e1	3	0.03
L3e1a	2	0.02
L3e2	7	0.06
L3e3	5	0.04
L3f1	4	0.03
Amerindian	14	0.12
A2	4	0.03
A5	1	0.01
B2	1	0.01
C1	6	0.05
D1	2	0.02
Eurasian	3	0.03
K	1	0.01
T2	1	0.01
V	1	0.01
North African	1	0.01
U6a	1	0.01
Total	120	1.00

Results

mtDNA Composition of the Black Brazilian Population

Among the 120 black Brazilian individuals analyzed, 85 different HVSI sequences were found, with a calculated haplotypic diversity of 0.991. The sequences obtained are shown in supplementary table 2, together with the RFLP data and haplogroup assignment and were de-

posited in GenBank with Accession numbers EF452532–EF452616. In three cases (SP22, SP29, SP71), sequencing of HVSI-II was needed for establishing the correct haplogroup.

Altogether, 102 sequences belonged to haplogroups characteristic of Sub-Saharan Africa (85.0%), whereas 14 were assigned to Native American (A–D; 11.7%), three to European haplogroups (2.5%) and one to the North African haplogroup U6a (table 1). The 102 Sub-Saharan African sequences were distributed into 68 haplotypes (supplementary table 2). For 45 of these haplotypes we could find identical matches in the literature, as detailed in supplementary table 3. Most of the haplotypes (40) gave complete matches with more than one, and occasionally a large number, of African populations.

Principal Component Analysis of mtDNA Data

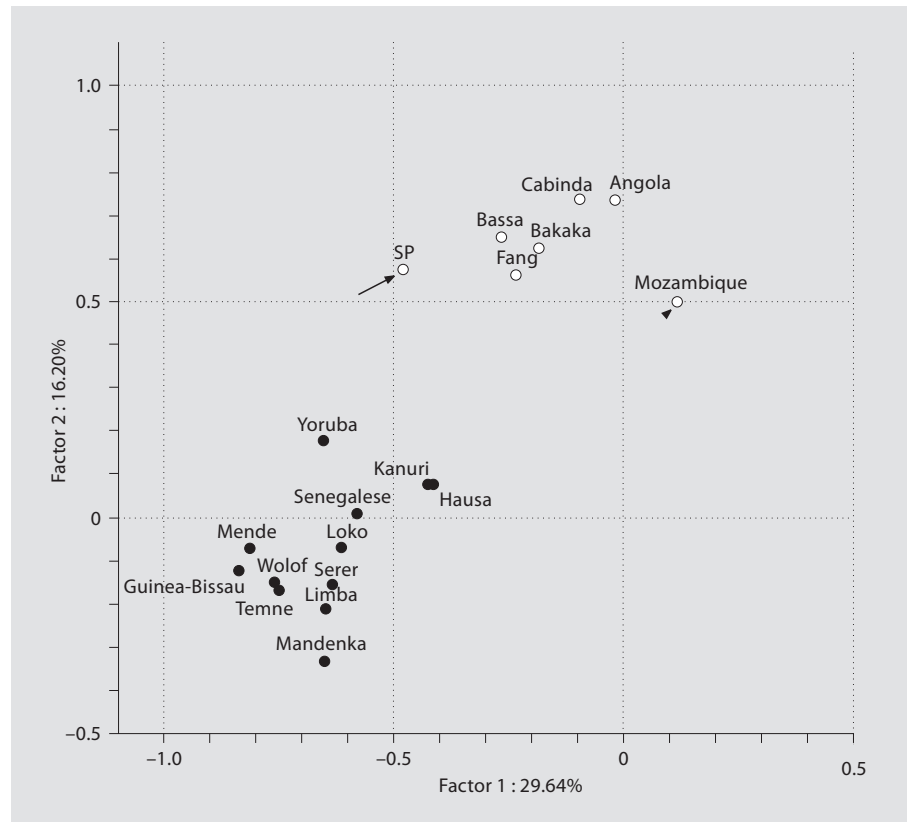
We combed the available literature for mtDNA and Y chromosome Sub-Saharan haplogroup frequencies for African native populations from regions involved in Atlantic Slave Trade, and extracted data from 19 different populations from West Africa, Central-West Africa or Southeast Africa. These are presented, with relevant references, in supplementary table 4. These data were then submitted to Analysis of Principal Components (APC), where a significant amount of geographical structure is evident on the projection of the variables on the 1 × 2 component plane (fig. 1). There are two clear clusters: one encompasses the populations of West Africa (black circles) and the second includes populations of West-Central Africa. Two populations lay outside the clusters: the Black Brazilians from São Paulo (SP), situated between the two main clusters (arrow) Mozambique, further to the right of the West-Central Africa cluster (arrowhead), suggesting a separate identity. We investigated these interesting suggestions by analyzing quantitatively the genetic structure of the data.

Bayesian Genetic Structure Analysis of mtDNA Data

For the Bayesian analysis of population structure we used the software BAPS 2.2 [45, 46]. The actual analysis was performed using a Markov chain Monte Carlo (MCMC) estimation with 100,000 steps (burn-in 5,000 steps). A structure with three geographically consistent clusters emerged, with the following posterior probabilities:

- (1) A Central-West Africa cluster, composed of Angola, Fang, Cabinda, Bakaka, Bassa and the Brazilian Blacks with a posterior probability of 0.910.

Fig. 1. Analysis of Principal Components of mtDNA haplogroup frequency data. The projection of the variables on the 1×2 component plane is shown. The black circles indicate populations from West Africa and the open circles designate populations of West-Central Africa. The arrow points toward the Black Brazilians from São Paulo and the arrowhead specifies the individuals from Mozambique.



- (2) A West Africa cluster composed of Kanuri, Yoruba, Senegalese, Serer, Wolof, Mandenka, Mende, Loko, Limba, Temne, Guinea-Bissau and Hausa with a posterior probability of 0.910.
- (3) A single-element Southeast Africa cluster (Mozambique) with a probability of 0.988.

The pairwise F_{st} values for the three clusters were 0.027 (SE 0.003) between clusters 1 and 2, 0.032 (SE 0.004) between clusters 1 and 3, and 0.054 (SE 0.004) between clusters 2 and 3 – all are significantly different from zero.

Admixture Analysis of mtDNA Data

Our next step was to try to estimate from the mtDNA data the relative ancestral contribution of West-Central, West and Southeast Africa to Brazilian Blacks from São Paulo using the LEADMIX maximum likelihood program [47]. The mtDNA data were entered with haplogroups as different alleles of a single locus, with the combined data of West-Central, West and Southeast Africa populations as parentals. 500 integration points were used in calculating the likelihood function. LEADMIX calculated for the Brazilian Blacks a West-Central Africa ancestry of 0.450 (95% confidence interval 0.106–0.702),

West Africa ancestry of 0.426 (95% CI 0.232–0.704) and Southeast Africa ancestry of 0.124 (0.000–0.324). The same program also outputs admixture proportion using the method of moments of Robert and Hiorns [48] which does not take drift into account – the proportions were 0.522, 0.326 and 0.153 for West-Central, West and Southeast Africa respectively (no confidence intervals are calculated because there was only one locus).

Y Chromosome Haplogroup Assignment

All samples were haplotyped using a hierarchical system according to the Y chromosome Consortium [38] as updated by Jobling and Tyler-Smith [39]. For classification we used 21 biallelic DNA markers that allowed high-resolution analysis of the E3a and E3b haplogroups. We also searched for the B (M60), for the *DYS199* (M3) T allele that defines the founder Amerindian haplogroup Q3* and for the F (M213) haplogroups. The F haplogroup is very broad, harboring patrilineages from Europe, Asia and America. However, our objective was to achieve high resolution in the haplogroups most characteristic of sub-Saharan Africa. Presumably individuals with the ancestral alleles in M60, M145 and M213 would belong to hap-

logroup A or C, but none was found. All individuals could be reliably assigned to the Y haplogroups.

Y Chromosome Population Analysis

The haplogroups found in Black Brazilians were E1, E2, E3a* (xE3a1, E3a2, E3a3, E3a4, E3a5, E3a6, E3a7), E3a4, E3a7, E3b1, E3b2, E3b3, B and F. The most frequent haplogroup overall was F (44%), which has a very wide geographical distribution worldwide and is quite heterogeneous. As shown in table 2, 48% of the individuals presented Y chromosome haplogroups characteristic of sub-Sahara Africans (E1, E2, E3a*, E3a4, E3a7 and B) and 52% presented non-Sub-Saharan African haplogroups (E3b1, E3b2, E3b3, Q3* and F (xQ3*)). Sub-lineages E3a* and E3a7 represented 88.3% of the African patrilineages.

Principal Component and Bayesian Genetic Structure Analysis of Y Chromosome Data

From the data presented by Cruciani et al. [49], Luis et al. [50], Beleza et al. [15] and Wood et al. [51] we compiled haplogroup frequency results for 20 population groups from West Africa [Wolof and Mandinka (Gambia/Senegal), Ewe, Ga, Fante (all Ghana), Fon (Benin), Mossi and Rimaibe (Burkina Faso)] and Central West Africa [Adamawa, Fali, Tali, Tupuri, Ewondo, Bakaka, Bamileke, Bantu, Bassa, Ngoumba (all Cameroon), Nande and Hema (Democratic Republic of Congo)]. The haplogroup frequencies calculated for these populations are presented in supplementary table 5. These data were then submitted to Analysis of Principal Components. As can be seen by inspection of figure 2, no geographically coherent genetic structure is evident – population from West Africa and West-Central Africa are intermixed without any pattern. We confirmed this observation using the BAPS 2.2 software as described above. The probability of all partitions was less than 0.01, thus corroborating the results of the APC that no genetic structure is observable from the NRY data.

Discussion

Directional Mating in Formation of the Brazilian Population

This study provides very relevant information about the structure of the Brazilian population. Our previous studies in self-declared White Brazilians from several geographical regions revealed a striking sexual asymmetry in the patterns of genetic admixture between the three

Table 2. Y chromosome haplogroup frequencies in 120 Black Brazilian males from São Paulo

Y Chromosome haplogroup	Black Brazilian males, %
A	0.0
B	3.2
C	0.0
E* (xE1, E2, E3)	0.0
E1	0.8
E2	0.8
E3* (xE3a, E3b)	0.0
E3a1	0.0
E3a2	0.0
E3a3	0.0
E3a4	0.8
E3a5	0.0
E3a6	0.0
E3a7	20.0
E3a*	22.4
E3b1	4.0
E3b2	3.0
E3b3	0.8
E3b4	0.0
E3b-V6	0.0
E3b*	0.0
F* (xQ3*)	42.5
Q3*	1.7

ancestral roots of the Brazilian population: Amerindians, Europeans and Africans. While the vast majority of Y-chromosomal lineages of Brazilian Whites were of European origin [52], almost two thirds of the mtDNA lineages were African or Amerindian [3]. This asymmetric mating, which led to directional introgression of Amerindian and African mitochondrial DNA, was perfectly consistent with the historical information that throughout Brazilian history, most importantly before the mid-19th Century, the vast majority of European immigrants was masculine. Similar patterns of sexually biased gene flow were described in Argentina [53], Chile [54], Colombia [55] and Costa Rica [56]. Indeed, it may prove to constitute a general characteristic of the Iberian colonization of the Americas.

In the present study we have found that 48% of the Y chromosome lineages of Brazilian Blacks and 85% of the mtDNA lineages were of Sub-Saharan African origin. Only 1.7% of the patrilineages were of Amerindian origin. On the other hand, there was significant introgression of European Y chromosome haplogroups into African Brazilians, certainly caused by the same demographic phenomena of directional mating between European males

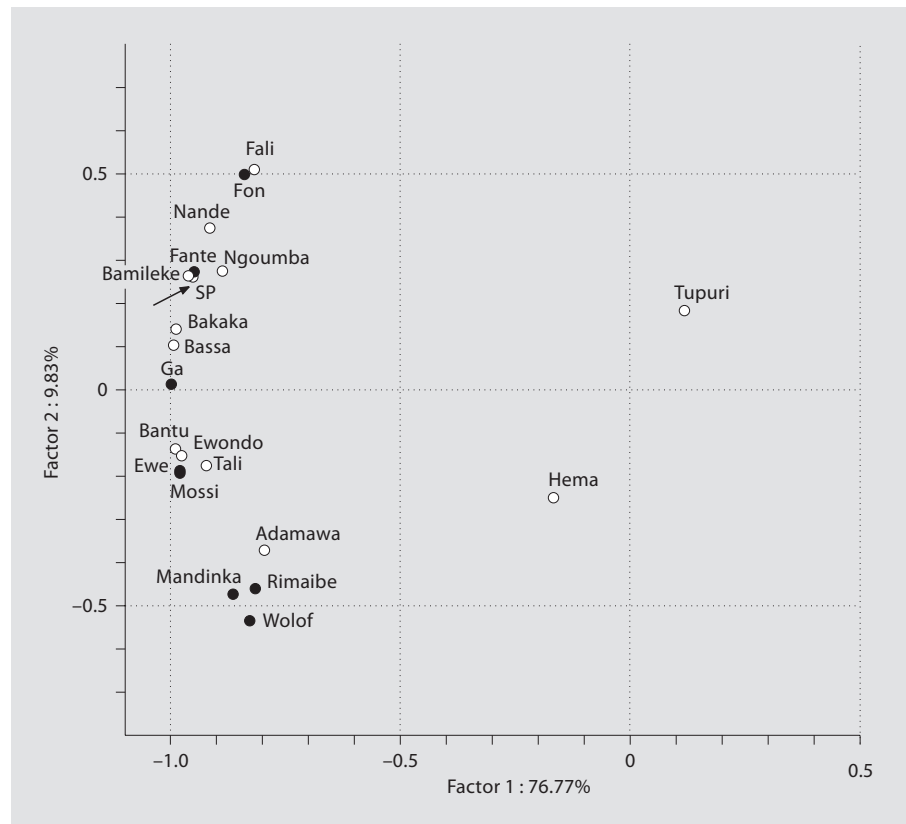


Fig. 2. Analysis of Principal Components of Y chromosome haplogroup frequency data. The projection of the variables on the 1×2 component plane is shown. The black circles indicate populations from West Africa and the open circles designate populations of West-Central Africa. The arrow points toward the Black Brazilians from São Paulo.

and African females in Brazil. Similar sexually asymmetric patterns were described in a Uruguayan population of African ancestry [57] and in African Americans [58].

Regional Assignment in Africa

The second question that we wished to examine was whether it was possible to use the African patrilineages and matrilineages found in Brazilian Blacks to ascertain their geographical origin in Africa. Of course this endeavor is predicated on the ability to differentiate genetically between the populations from the parts of Africa where slaves were captured to be taken to Brazil, especially West Africa, West-Central Africa and Southeast Africa, i.e. the Niger-Congo linguistic territory.

Interestingly, while most of the subfamilies of the Niger-Congo family are agglomerated in West Africa, more than half of the Niger-Congo speakers, those in subequatorial Africa, belong to the single Bantu subfamily [59]. This is due to the fact that circa 5,000 years before present a small group of farmers inhabiting part of North Cameroon and adjacent eastern Nigeria spread out of their homeland over most of subequatorial Africa in the so-called Bantu expansion [60, 61].

This territorial sweep by the Bantus originating from West Africa would be expected to produce significant genetic homogenization between West and subequatorial Africa and also within the latter. According to Diamond [61] this was 'how Africa became Black'. Thus, it is not surprising that we did not find significant stratification of Y chromosome haplogroup frequencies between the West African and West-Central African populations analyzed.

How then can we explain our observation of considerable genetic structure in mtDNA among West Africa, West-Central Africa and Southeast Africa, especially if we take into consideration the fact that West-Central Africa and Southeast Africa are both predominantly inhabited by Bantu speakers? A possible explanation would be that in the process of Bantu expansion there could have occurred directional introgression of mtDNA haplogroups from the inhabitants of the conquered territories into the invading Bantus. This would create genetic diversification of the settled Bantus in their new homes. Recently, Batini et al. [62] presented evidence in favor of this scenario. They showed that most of the Bantu mtDNA haplotypes belonging to L1c1a, L1c4 and L1c5

were in fact Pygmy haplotypes introgressed by directional mating. Indeed, the authors gathered ethnographic and genetic evidence to suggest that unions between Pygmy females and Bantu males were socially favored, whereas taboos make those between Pygmy males and Bantu females difficult. They also show that sharing of Y-chromosomal haplogroups between the two population groups was most probably due to introgression of Bantu lineages into Western Pygmies (haplogroup E3a), a process that should contribute to homogenization of regional Y chromosome haplogroup frequencies.

Thus, we could hypothesize the possible sequence of events to explain our observations: let us assume that originally the Bantus had a genetic makeup similar to their neighboring West African populations, as suggested by our Y chromosome results. As they moved into West-Central Africa, introgression of the mtDNA of Pigmies and other local populations led to their differentiation from West Africans. A similar proposal was put forward by Destro-Bisol et al. [63]. Afterwards, as Bantus moved into Southeast Africa, there was further mtDNA introgression from the invaded populations leading to differentiation from the Bantus of West-Central Africa. Consistent with this proposal, Salas et al. [34] reported that Khoisan lineages comprise circa 5% of the total lineages in the Bantu speakers of Mozambique. The observed pairwise *F*_{st} distances between the three geographical groups are quite compatible with our proposed scenario, being largest for the West Africa-Southeast Africa comparison (0.054 ± 0.004) and more or less equivalent the West-West-Central (0.032 ± 0.004) and West-Central-Southeast (0.027 ± 0.003) contrasts. Of course, genetic drift should also contribute to this genetic differentiation.

The Ancestral African Roots of Blacks from São Paulo

Since we found significant differences in the mtDNA haplogroup frequencies among West, West-Central and Southeast Africa, we next tried to estimate the relative ancestral contribution of these three regions to the studied Black individuals from São Paulo, Brazil. The LEADMIX software, using maximum likelihood estimates that take genetic drift into account calculated a West-Central Africa ancestry of 0.450 (95% CI 0.106–0.702), West Africa ancestry of 0.426 (95% CI 0.232–0.704) and Southeast Africa ancestry of 0.124 (95% CI 0.000–0.324). Two elements are very noteworthy in these data: an unexpectedly high ancestral involvement of West Africa and simultaneously a significant contribution of Southeast Africa to this group of Black individuals from São Paulo.

The relatively high proportion of West African ancestry in Blacks from São Paulo can be explained by two factors. First, Klein [2] presented data showing that between 1777 and 1829 65% of the slaves in 13 municipalities in the state of São Paulo had West African origin. Thus, it appears that São Paulo, for reasons that are not clear, already in the early years of the 19th century differed from the norm that in southeastern Brazil the bulk of slaves had been imported from West Central Africa. Second, a strong movement of internal migration occurred between the termination of the Atlantic slave trade in 1850 and the abolition of slavery in 1888 (Brazil was the last country in the Western Hemisphere to do so). The coffee industry was booming in São Paulo, and a significant number of slaves were relocated from Bahia (to where originally a high proportion of West African slaves had been taken) and other states of the then economically stagnant Brazilian Northeast [11, 64]. On the eve of abolition in 1887, almost three-quarters of the Brazilian slaves could be found in the Southeast region of Brazil, especially São Paulo [11].

The second element, a significant proportion of Southeast Africans, can be explained by the fact that between 1830 and 1850 the slave trade reached a historical all time high and that this was exactly when imports from Mozambique were reaching 20% of the total [7, 8]. Thus, although limited in time span, the contribution of Southeast Africa was numerically quite significant.

As we have seen, the special usefulness of mtDNA data derives from its maternal uniparental mode of transmission. Yet, the same characteristic also makes mtDNA data susceptible to hidden biases. For instance, it is known that roughly 70% of the African slaves taken to the Americas were males [65], but there were significant variations over time and space [66]. If any particular region of Africa contributed a larger proportion of females it might be overrepresented in ancestral estimates based on mtDNA. However, there is no evidence that this happened [65].

In conclusion, we have shown that the frequency of mitochondrial DNA haplogroups represents a useful new tool for estimating the regional ancestry of slaves in Africa, including differentiation within the Bantu linguistic group. Our next step will be to embark on mtDNA haplogroup determination in Black individuals from diverse regions of Brazil, including Rio de Janeiro, Minas Gerais and Bahia. From these studies we should be able to derive a more comprehensive picture of the ancestry of African Brazilian populations.

Acknowledgements

We specially thank the anonymous sample donors, who allowed this work to be carried out. We thank to Raquel Liboredo, Neuza Antunes, Kátia Barroso and Míriam Rodrigues for technical assistance. We are grateful to Dr. Peter A. Underhill from the

Department of Genetics, Stanford University for the Y chromosome sequence used to design the minisequencing primers and to Dr. Herbert Klein, of the Center for Latin American Studies of Stanford University for useful discussions and suggestions. This work was supported by grants from CNPq of Brazil.

References

- 1 Reis JJ: Presença negra: conflitos e encontros; in: IBGE – Instituto Brasileiro de Geografia Estatística (2000) Brasil: 500 Anos de Povoamento. IBGE, Rio de Janeiro, Brazil, 2000, pp 80–99.
- 2 Klein HS: As origens Africanas dos escravos brasileiros; in Pena SDJ (ed): *Homo brasilis*: aspectos genéticos, linguísticos, históricos e socioantropológicos da formação do povo brasileiro. Ribeirão Preto, FUNPEC, Brazil, 2002, pp 93–112.
- 3 Alves-Silva J, da Silva Santos M, Guimaraes PE, Ferreira AC, Bandelt HJ, Pena SDJ, Prado VF: The ancestry of Brazilian mtDNA lineages. *Am J Hum Genet* 2000;67:444–461.
- 4 Pena SDJ, Bortolini MC: Pode a genética definir quem deve se beneficiar das cotas universitárias e demais ações afirmativas? *Estudos Avançados* 2004;18:31–50.
- 5 Parra FC, Amado RC, Lambertucci JR, Rocha J, Antunes CM, Pena SDJ: Color and genomic ancestry in Brazilians. *Proc Natl Acad Sci USA* 2003;100:177–182.
- 6 Pimenta JR, Zuccherato LW, Debes AA, Masetti L, Soares RP, Moura-Neto RS, Rocha J, Bydlowski SP, Pena SDJ: Color and genomic ancestry in Brazilians: A study with forensic microsatellites. *Hum Hered* 2006;62:190–195.
- 7 Florentino M: *Em Costas Negras*. Companhia das Letras, São Paulo, Brazil, 1997.
- 8 Alpers EA: 'Moçambiques' in Brazil: another dimension of the African diaspora in the Atlantic world; in: Curto JC, Soulandre-LaFrance R (eds) *Africa and the Americas: Interconnections During the Slave Trade*. Lawrenceville, NJ, Africa World Press, pp 43–64.
- 9 Lacombe AJ, Silva E, Barbosa FA: *Rui Barbosa e a Queima dos Arquivos*. Rio de Janeiro, Brazil, Fundação Casa de Rui Barbosa, 1988.
- 10 Eltis D, Behrendt SD, Richardson D, Klein HS: *The Trans-Atlantic Slave Trade: A Database on CD-ROM*. Cambridge, Cambridge University Press, 2000.
- 11 Klein HS: *African Slavery in Latin America and the Caribbean*. New York, Oxford University Press, 1986.
- 12 Zago MA, Figueiredo MS, Ogo SH: Bantu beta S cluster haplotype predominates among Brazilian blacks *Am J Phys Anthropol* 1992;88:295–298.
- 13 Öner C, Dimovski AJ, Olivieri NF, Schiliro G, Codrington JF, Fattoum S, Adekile AD, Oner R, Yuregir GT, Altay C, Gurgey A, Gupta RB, Jogessar VB, Kitundu MN, Loukopoulos D, Tamagnini GP, Ribeiro MLS, Kutlar F, Gu L-H, Lanclos KD, Huisman THJ: Beta S haplotypes in various world populations. *Hum Genet* 1992;89:99–104.
- 14 Bandelt HJ, Quintana-Murci L, Salas A, Macaulay V: The fingerprint of phantom mutations in mitochondrial DNA data. *Am J Hum Genet* 2002;71:1150–1160.
- 15 Beleza S, Gusmão L, Amorim A, Carracedo A, Salas A: The genetic legacy of western Bantu migrations *Hum Genet* 2005;117:366–375.
- 16 Brakez Z, Bosch E, Izaabel H, Akhayat O, Comas D, Bertranpetit J, Calafell F: Human mitochondrial DNA sequence variation in the Moroccan population of the Souss area. *Ann Hum Biol* 2001;28:295–307.
- 17 Brandstatter A, Peterson CT, Irwin JA, Mpoke S, Koech DK, Parson W, Parsons TJ: Mitochondrial DNA control region sequences from Nairobi (Kenya): Inferring phylogenetic parameters for the establishment of a forensic database. *Int J Legal Med* 2004;118:294–306.
- 18 Brehm A, Pereira L, Bandelt HJ, Prata MJ, Amorim A: Mitochondrial portrait of the Cabo Verde archipelago: The Senegambian outpost of Atlantic slave trade. *Ann Hum Genet* 2002;66:49–60.
- 19 Cerny V, Hajek M, Cmejla R, Bruzek J, Brdiczka R: mtDNA sequences of Chadic-speaking populations from northern Cameroon suggest their affinities with eastern Africa. *Ann Hum Biol* 2004;31:554–569.
- 20 Chen YS, Olckers A, Schurr TG, Kogelnik AM, Huoponen K, Wallace DC: mtDNA variation in the South African Kung and Khwe-and their genetic relationships to other African populations. *Am J Hum Genet* 2000;66:1362–1383.
- 21 Coia V, Destro-Bisol G, Verginelli F, Battaglia C, Boschi I, Cruciani F, Spedini G, Comas D, Callafel F: mtDNA variation in North Cameroon: lack of Asian lineages and implications for back migration from Asia to sub-Saharan Africa. *Am J Phys Anthropol* 2005;128:678–681.
- 22 Destro-Bisol G, Donati F, Coia V, Boschi I, Verginelli F, Caglia A, Tofanelli S, Spedini G, Capelli C: Variation of female and male lineages in sub-Saharan populations: the importance of sociocultural factors. *Mol Biol Evol* 2004;21:1673–1682.
- 23 Graven L, Passarino G, Semino O, Boursot P, Santachiara-Benerecetti S, Langaney A, Excoffier L: Evolutionary correlation between control region and restriction polymorphisms in the mitochondrial genome of a large Senegalese Mandenka sample. *Mol Biol Evol* 1995;12:334–345.
- 24 Jackson BA, Wilson JL, Kirbah S, Sidney SS, Rosenberg J, Bessie N, Alie JA, McLean DC, Garvey WT, Ely B: Mitochondrial DNA genetic diversity among four ethnic groups in Sierra Leone. *Am J Phys Anthropol* 2005;128:156–163.
- 25 Kivisild T, Reidla M, Metspalu E, Rosa A, Brehm A, Pennarun E, Parik J, Geberhiwot T, Usanga E, Vilems R: Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears. *Am J Hum Genet* 2004;75:752–770.
- 26 Knight A, Underhill PA, Mortensen HM, Zhivotovsky LA, Lin AA, Henn BM, Louis D, Ruhlen M, Mountain JL: African Y chromosome and mtDNA divergence provides insight into the history of click languages. *Curr Biol* 2003;13:464–473.
- 27 Krings M, Salem AE, Bauer K, Geisert H, Malek AK, Chaix L, Simon C, Welsby D, Di Rienzo A, Utermann G, Sajantila A, Paabo S, Stoneking M: mtDNA analysis of Nile River Valley populations: A genetic corridor or a barrier to migration? *Am J Hum Genet* 1999;64:1166–1176.
- 28 McLean DC Jr, Spruill I, Gevaio S, Morrison EY, Bernard OS, Argyropoulos G, Garvey WT: Three novel mtDNA restriction site polymorphisms allow exploration of population affinities of African Americans. *Hum Biol* 2003;75:147–161.
- 29 Pereira L, Macaulay V, Torroni A, Scozzari R, Prata MJ, Amorim A: Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade. *Ann Hum Genet* 2001;65:439–458.
- 30 Pinto F, González AM, Hernández M, Laruga JM, Cabrera VN: Genetic relationship between the Canary islanders and their African and Spanish ancestors inferred from mitochondrial DNA sequences. *Ann Hum Genet* 1996;60:321–330.

- 31 Plaza S, Salas A, Calafell F, Corte-Real F, Bertranpetit J, Carracedo A, Comas D: Insights into the western Bantu dispersal: mtDNA lineage analysis in Angola. *Hum Genet* 2004; 115:439–447.
- 32 Rando JC, Pinto F, González AM, Hernández M, Larruga JM, Cabrera VM, Bandelt H-J: Mitochondrial DNA analysis of northwest African populations reveals genetic exchanges with European, Near-Eastern, and sub-Saharan populations. *Ann Hum Genet* 1998;65:531–550.
- 33 Rosa A, Brehm A, Kivisild T, Metspalu E, Villemers R: MtDNA profile of West Africa Guineans: towards a better understanding of the Senegambia region. *Ann Hum Genet* 2004;68:340–352.
- 34 Salas A, Richards M, De la Fe T, Lareu MV, Sobrino B, Sanchez-Diaz P, Macaulay V, Carracedo A: The making of the African mtDNA landscape. *Am J Hum Genet* 2002; 71:1082–1111.
- 35 Trovada MJ, Pereira L, Gusmao L, Abade A, Amorim A, Prata MJ: Pattern of mtDNA variation in three populations from Sao Tome e Principe. *Ann Hum Genet* 2004;68: 40–54.
- 36 Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC: African populations and the evolution of mitochondrial DNA. *Science* 1991;253:1503–1507.
- 37 Watson E, Forster P, Richards M, Bandelt H-J: Mitochondrial footprints of human expansions in Africa. *Am J Hum Genet* 1997; 61:691–704.
- 38 Shen P, Lavi T, Kivisild T, Chou V, Sengun D, Gefel D, Shpirer I, Woolf E, Hillel J, Feldman M, Oefner P: Reconstruction of patri- and matri-lineages of Samaritans and other Israeli populations from Y-chromosome and mitochondrial DNA sequence variation. *Hum Mut* 2004;24:248–260.
- 39 Torroni A, Achilli A, Macaulay V, Richards M, Bandelt H-J: Harvesting the fruit of the human mtDNA tree. *Trends Genet* 2006;22: 339–345.
- 40 Y Chromosome consortium: A nomenclature for the tree of human Y-chromosomal binary haplogroups. *Genome Res* 2002;12: 339–348.
- 41 Jobling M, Tyler-Smith C: The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet* 2003;4:598–612.
- 42 Carvalho CMB, Pena SDJ: Optimization of a multiplex minisequencing protocol for population studies and medical genetics. *Genet Mol Res* 2005;4:115–125.
- 43 Lell JT, Brown MD, Schurr TG, Sukernik RI, Starikovskaya YB, Torroni A, Moore LG, Troup GM, Wallace DC: Y chromosome polymorphisms in native American and Siberian populations: identification of native American Y chromosome haplotypes. *Hum Genet* 1997;100:536–543.
- 44 Schneider S, Roessli D, Excoffier L: Arlequin ver 2.000: A software for population genetics data. Genetics and Biometry Laboratory, University of Geneva, Switzerland, 2000.
- 45 Corander J, Waldmann P, Sillanpaa MJ: Bayesian analysis of genetic differentiation between populations. *Genetics* 2003;163: 367–374.
- 46 Corander J, Waldmann P, Marttinen P, Sillanpaa MJ: BAPS 2: enhanced possibilities for the analysis of genetic population structure. *Bioinformatics* 2004;20:2363–2369.
- 47 Wang J: Maximum likelihood estimation of admixture proportions from genetic data. *Genetics* 2003;164:747–765.
- 48 Roberts DF, Hiorns RW: Methods of analysis of the genetic composition of a hybrid population. *Hum Biol* 1965;37:38–43.
- 49 Cruciani F, La Fratta R, Santolamazza P, Sellitto D, Pascone R, Moral P, Watson E, Guida V, Colomb EB, Zaharova B, Lavinha J, Vona G, Aman R, Cali F, Akar N, Richards M, Torroni A, Novelletto A, Scozzari R: Phylogeographic analysis of haplogroup E3b (E-M215) Y chromosomes reveals multiple migratory events within and out of Africa. *Am J Hum Genet* 2004;74:1014–1022.
- 50 Luis JR, Rowold DJ, Regueiro M, Caeiro B, Cinnioglu C, Roseman C, Underhill PA, Cavalli-Sforza LL, Herrera RJ: The Levant versus the Horn of Africa: evidence for bidirectional corridors of human migrations. *Am J Hum Genet* 2004;74:532–544.
- 51 Wood ET, Stover DA, Ehret C, Destro-Bisol G, Spedini G, McLeod H, Louie L, Bamshad M, Strassmann BI, Soodyall H, Hammer MF: Contrasting patterns of Y chromosome and mtDNA variation in Africa: evidence for sex-biased demographic processes. *Eur J Hum Genet* 2005;13:867–876.
- 52 Carvalho-Silva DR, Santos FR, Rocha J, Pena SDJ: The phylogeography of Brazilian Y-chromosome lineages. *Am J Hum Genet* 2001;68:281–286.
- 53 Martinez-Marignac VL, Bertoni B, Parra EJ, Bianchi NO: Characterization of admixture in an urban sample from Buenos Aires, Argentina, using uniparentally and biparentally inherited genetic markers. *Hum Biol* 2004; 76:543–557.
- 54 Rocco PP, Morales CG, Moraga MV, Miquel JFP, Nervi FO, Llop ER, Carvallo PS, Rothhammer FE: Composición genética de la población chilena: Distribución de polimorfismos de DNA mitocondrial en grupos originarios y en la población mixta de Santiago. *Rev Med Chile* 2002;130:125–131.
- 55 Bedoya G, Montoya P, Garcia J, Soto I, Bourgeois S, Carvajal L, Labuda D, Alvarez V, Ospina J, Hedrick PW, Ruiz-Linares A: Admixture dynamics in Hispanics: a shift in the nuclear genetic ancestry of a South American population isolate. *Proc Natl Acad Sci USA* 2006;103:7234–7239.
- 56 Carvajal-Carmona LG, Ophoff R, Service S, Hartiala J, Molina J, Leon P, Ospina J, Bedoya G, Freimer N, Ruiz-Linares A: Genetic demography of Antioquia (Colombia) and the Central Valley of Costa Rica. *Hum Genet* 2003;112:534–541.
- 57 Bravi CM, Sans M, Bailliet G, Martinez-Marignac VL, Portas M, Barreto I, Bonilla C, Bianchi NO: Characterization of mitochondrial DNA and Y-chromosome haplotypes in a Uruguayan population of African ancestry. *Hum Biol* 1997;69:641–652.
- 58 Lind JM, Hutcheson-Dilks HB, Williams SM, Moore JH, Essex M, Ruiz-Pesini E, Wallace DC, Tishkoff SA, O'Brien SJ, Smith MW: Elevated male European and female African contributions to the genomes of African American individuals. *Hum Genet* 2007; 120:713–722.
- 59 Rexova K, Bastin Y, Frynta D: Cladistic analysis of Bantu languages: A new tree based on combined lexical and grammatical data. *Naturwissenschaften* 2006;93:189–194.
- 60 Hiernaux J: *The People of Africa*. New York, Charles Scribner's Sons, 1975.
- 61 Diamond J: *Guns, Germs and Steel: The Fates of Human Societies*. New York, W.W. Norton & Co., 1997.
- 62 Batini C, Coia V, Battaglia C, Rocha J, Pilkington MM, Spedini G, Comas D, Destro-Bisol G, Calafell F: Phylogeography of the human mitochondrial L1c haplogroup: Genetic signatures of the prehistory of Central Africa. *Mol Phylogenet Evol* 2006; [Epub ahead of print].
- 63 Destro-Bisol G, Donati F, Coia V, Boschi I, Verginelli F, Caglia A, Tofanelli S, Spedini G, Capelli C: Variation of female and male lineages in sub-Saharan populations: the importance of sociocultural factors. *Mol Biol Evol* 2004;21:1673–1682.
- 64 Andrews GR: *Blacks and Whites in São Paulo Brazil 1888–1988*. Madison, The University of Wisconsin Press, 1991.
- 65 Eltis D, Engerman SL: Was the slave trade dominated by men? *J Interdisc Hist* 1992;23: 237–257.
- 66 Eltis D, Engerman SL: Fluctuations in sex age ratios in the transatlantic slave trade, 1663–1864. *Econ Hist Rev* 1993;46:308–323.
- 67 Mateu E, Comas D, Calafell F, Perez-Lezaun A, Abade A, Bertranpetit J: A tale of two islands: Population history and mitochondrial DNA sequence variation of Bioko and Sao Tome, Gulf of Guinea. *Ann Hum Genet* 1997;61:507–518.
- 68 Rando JC, Cabrera VM, Larruga JM, Hernandez M, Gonzalez AM, Pinto F, Bandelt HJ: Phylogeographic patterns of mtDNA reflecting the colonization of the Canary Islands. *Ann Hum Genet* 1999;63:413–28.