Maternal traces of deep common ancestry and asymmetric gene flow between Pygmy hunter–gatherers and Bantu-speaking farmers

Lluís Quintana-Murci[a,b], Hélène Quach[a], Christine Harmant[a], Francesca Luca[a], Blandine Massonnet[a], Etienne Patin[a], Lucas Sica[c], Patrick Mouguiama-Daouda[d], David Comas[e], Shay Tzur[f], Oleg Balanovsky[g], Kenneth K. Kidd[h], Judith R. Kidd[h], Lolke van der Veen[d], Jean-Marie Hombert[d], Antoine Gessain[i], Paul Verdu[j], Alain Froment[j], Serge Bahuchet[j], Evelyne Heyer[j], Jean Dausset[b,k], Antonio Salas[l], and Doron M. Behar[f]

[a]Human Evolutionary Genetics Unit, Centre National de la Recherche Scientifique URA3012, Institut Pasteur, 75015 Paris, France; [c]Centre International de Recherches Médicales de Franceville, BP769 Franceville, Gabon; [d]Unité Mixte de Recherche 5596, Institut des Sciences de l'Homme, 69363 Lyon, France; [e]Unitat de Biologia Evolutiva, Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, 08003 Barcelona, Spain; [f]Molecular Medicine Laboratory, Rambam Health Care Campus, Haifa 31096, Israel; [g]Russian Academy of Medical Sciences, 109544 Moscow, Russia; [h]Department of Genetics, Yale University School of Medicine, SHM I-353, New Haven, CT 06510; [i]Unité d'Epidémiologie et Physiopathologie des Virus Oncogènes, Institut Pasteur, 75015 Paris, France; [j]Muséum National d'Histoire Naturelle–Centre National de la Recherche Scientifique-P7 Unité Mixte de Recherche 5145, Eco-Anthropologie, Musée de l'Homme, 75016 Paris, France; [k]Fondation Jean Dausset, Centre d'Etude du Polymorphisme Humain, 75010 Paris, France; and [l]Unidade de Xenética, Instituto de Medicina Legal, Universidad de Santiago de Compostela, 15782 Galicia, Spain

Contributed by Jean Dausset, December 6, 2007 (sent for review October 12, 2007)

Two groups of populations with completely different lifestyles—the Pygmy hunter–gatherers and the Bantu-speaking farmers—coexist in Central Africa. We investigated the origins of these two groups and the interactions between them, by analyzing mtDNA variation in 1,404 individuals from 20 farming populations and 9 Pygmy populations from Central Africa, with the aim of shedding light on one of the most fascinating cultural transitions in human evolution (the transition from hunting and gathering to agriculture). Our data indicate that this region was colonized gradually, with an initial L1c-rich ancestral population ultimately giving rise to current-day farmers, who display various L1c clades, and to Pygmies, in whom L1c1a is the only surviving clade. Detailed phylogenetic analysis of complete mtDNA sequences for L1c1a showed this clade to be autochthonous to Central Africa, with its most recent branches shared between farmers and Pygmies. Coalescence analyses revealed that these two groups arose through a complex evolutionary process characterized by (*i*) initial divergence of the ancestors of contemporary Pygmies from an ancestral Central African population no more than ≈70,000 years ago, (*ii*) a period of isolation between the two groups, accounting for their phenotypic differences, (*iii*) long-standing asymmetric maternal gene flow from Pygmies to the ancestors of the farming populations, beginning no more than ≈40,000 years ago and persisting until a few thousand years ago, and (*iv*) enrichment of the maternal gene pool of the ancestors of the farming populations by the arrival and/or subsequent demographic expansion of L0a, L2, and L3 carriers.

Africa | evolution | human | mtDNA | populations

Modern humans have undergone a major cultural and technological change: the transition from food collection (hunting–gathering) to food production (agriculture). This transition has occurred in many parts of the world and began ≈13–10,000 years before the present (YBP). It has made it possible for groups to increase in size and to shift from nomadism to sedentarism (1). In subSaharan Africa, agriculture spread much later, expanding from western Central Africa (i.e., eastern Nigeria and western Cameroon) to much of the East, Central, and southern Africa only 3–5,000 YBP. This spread of agriculture was related to the diffusion of Bantu languages ("Bantu expansions") and, possibly, the use of iron (2–4). A few populations, such as the Pygmy hunter–gatherers, did not adopt an agricultural lifestyle and have remained demographically and geographically restricted (5, 6). Modern-day western and eastern Pygmy populations in Central Africa (CA) share distinctive physical and cultural characteristics thought to result from long isolation and adaptation to the rainforest (7–9). Two features of CA make this a key region for understanding recent human evolution: (*i*) two groups of populations with completely different lifestyles coexist in this region—Pygmy hunter–gatherer (PHG) and Bantu-speaking agricultural (AGR) populations (5, 6), and (*ii*) this region is immediately adjacent to the putative site of origin of Bantu expansions (2–4).

Studies of contemporary genetic variation in human populations have proved an important tool for investigating human origins and migratory patterns (10–12). Variations in the maternally inherited mtDNA genome have provided evidence supporting both the African origin of modern humans and subsequent expansion throughout the world (13–20). However, peopling processes and migration dynamics remain poorly resolved in Africa, especially in CA. Patterns of mtDNA variation in AGR and PHG populations have been investigated in studies focusing on the control region [usually the hypervariable segment I (HVS-I)] and a few coding sites (14, 21–35). A number of mtDNA haplogroups (Hgs) have been identified as possible genetic footprints of Bantu expansions. These Hgs include L0a (21, 24, 29), L1c (23, 32), L2a (29, 30), L3b (26), and L3e (27, 28). Studies of small numbers of eastern Mbuti PHG, in which only HVS-I was investigated (14), have suggested that the mtDNA gene pool of eastern Pygmies differs substantially from that of western PHG, who display genetic similarities to neighboring farming populations (30, 33, 34). However, the actual origins of the PHG and AGR populations and their interactions in space and time remain unclear because of the small samples sizes and

Author contributions: L.Q.-M. designed research; H.Q., C.H., F.L., and B.M. performed research; E.P., L.S., P.M.-D., D.C., S.T., O.B., K.K.K., J.R.K., L.v.d.V., J.-M.H., A.G., P.V., A.F., S.B., E.H., and J.D. contributed new reagents/analytic tools; L.Q.-M. and D.M.B. analyzed data; and L.Q.-M., A.S., and D.M.B. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

Data deposition: The complete mtDNA sequences reported in this paper have been deposited in the GenBank database (accession nos. EU273476–EU273502).

[b]To whom correspondence may be addressed. E-mail: quintana@pasteur.fr or dausset@cephb.fr.

This article contains supporting information online at www.pnas.org/cgi/content/full/0711467105/DC1.

Table 1. General diversity indices and neutrality tests for the Bantu-speaking AGR and PHG populations studied

Population (code)	Location	*n**	Hg D (SE)†	Ht D (SE)‡	Pi (SE)§	Tajima's *D* (*P*)¶	Fu's F_S (*P*)¶
Agricultural							
Akele (KEL)	Gabon, west	48	0.925 (0.022)	0.985 (0.008)	9.811 (4.571)	−0.70 (0.262)	−16.76
Ateke (TEK)	Gabon, southeast	54	0.945 (0.012)	0.985 (0.007)	9.088 (4.248)	−0.76 (0.231)	−21.96
Benga (BEN)	Gabon, northwest	50	0.931 (0.016)	0.952 (0.015)	9.922 (4.616)	−0.67 (0.307)	−4.53 (0.101)
Duma (DUM)	Gabon, east	47	0.925 (0.016)	0.973 (0.010)	9.258 (4.332)	−0.92 (0.193)	−9.09
Eshira (GIS)	Gabon, west	40	0.939 (0.016)	0.971 (0.012)	10.077 (4.703)	−0.68 (0.293)	−5.84 (0.060)
Eviya (EVI)	Gabon, center	38	0.898 (0.023)	0.932 (0.018)	9.135 (4.297)	−0.52 (0.299)	−0.08 (0.539)
Ewondo (EWD)	Cameroon, west	25	0.900 (0.023)	0.933 (0.023)	9.933 (4.702)	0.05 (0.571)	0.95 (0.692)
Fang (FAN-CM)	Cameroon, south	39	0.880 (0.028)	0.970 (0.014)	9.333 (4.381)	−0.44 (0.402)	−9.46
Fang (FAN-GB)	Gabon, north	66	0.930 (0.012)	0.971 (0.009)	8.849 (4.132)	−0.78 (0.235)	−12.99
Galoa (GAL)	Gabon, west	51	0.925 (0.019)	0.965 (0.011)	9.002 (4.214)	−0.96 (0.172)	−6.13 (0.047)
Kota (KOT)	Gabon, east	56	0.900 (0.023)	0.967 (0.010)	10.562 (4.885)	−0.61 (0.283)	−8.28 (0.021)
Makina (MAK)	Gabon, center	45	0.928 (0.017)	0.962 (0.016)	9.306 (4.356)	−0.71 (0.269)	−7.28 (0.027)
Mitsogo (TSO)	Gabon, center	64	0.898 (0.025)	0.961 (0.011)	9.058 (4.224)	−0.84 (0.219)	−9.50
Ndumu (NDU)	Gabon, southeast	39	0.953 (0.013)	0.973 (0.013)	9.417 (4.418)	−0.92 (0.178)	−8.01
Ngumba (NGU)	Cameroon, west	88	0.932 (0.010)	0.969 (0.007)	10.090 (4.655)	−0.35 (0.435)	−14.10
Nzebi (NZE)	Gabon, southeast	63	0.949 (0.010)	0.976 (0.010)	8.955 (4.181)	−1.16 (0.110)	−22.92
Obamba (OBA)	Gabon, southeast	47	0.942 (0.016)	0.988 (0.007)	9.741 (4.542)	−1.13 (0.108)	−17.49
Orungu (ORU)	Gabon, west	20	0.905 (0.041)	0.974 (0.025)	10.895 (5.173)	−0.13 (0.508)	−3.53 (0.090)
Punu (PUN)	Gabon, southwest	52	0.946 (0.014)	0.982 (0.007)	9.124 (4.266)	−1.24 (0.096)	−15.94
Shake (SHA)	Gabon, east	51	0.899 (0.022)	0.973 (0.011)	10.195 (4.733)	−0.68 (0.275)	−13.01
Eastern Pygmy							
Mbuti (MBU)	DRC	39	0.710 (0.041)	0.823 (0.034)	6.877 (3.307)	1.05 (0.886)	2.69 (0.851)
Western Pygmy							
Babongo (BAB)	Gabon, southeast	45	0.721 (0.052)	0.749 (0.058)	6.945 (3.327)	−0.27 (0.493)	1.75 (0.799)
Baka (BAK-CC)	Cameroon, center	30	0.540 (0.080)	0.830 (0.035)	5.425 (2.688)	0.26 (0.655)	3.34 (0.899)
Baka (BAK-CW)	Cameroon, southwest	58	0.654 (0.040)	0.786 (0.037)	5.667 (2.757)	−0.75 (0.256)	1.52 (0.766)
Baka (BAK-GB)	Gabon, northeast	39	0.533 (0.034)	0.757 (0.052)	4.124 (2.098)	0.08 (0.569)	2.76 (0.886)
Bakola (BAKO)	Cameroon, west	88	0.455 (0.033)	0.722 (0.024)	3.509 (1.805)	2.01 (0.971)	5.05 (0.942)
Bakoya (BKY)	Gabon, northeast	31	0.333 (0.096)	0.548 (0.087)	3.011 (1.614)	−0.99 (0.190)	4.26 (0.954)
Biaka (BIA)	CAR	56	0.724 (0.030)	0.823 (0.030)	6.006 (2.906)	0.05 (0.632)	2.42 (0.857)
Tikar (BEZ)	Cameroon, north	35	0.464 (0.054)	0.703 (0.027)	2.911 (1.565)	1.46 (0.923)	4.38 (0.955)

*Sample size.
†Gene diversity based on haplogroup profiles (Hg D) and standard error (*SE*).
‡Gene diversity based on HVS-I sequence-based haplotypes (Ht D).
§Average number of pairwise differences (Pi).
¶All *P* values are <0.02 (for Fu's F_S), unless otherwise stated.

limited number of directly sampled populations from CA and the low molecular resolution achieved with HVS-I and a few RFLP markers only.

This study provides insight into one of the most fundamental questions in human evolution: ancient and present-day genetic ties between AGR and PHG populations and the possible common ancestry between these two groups. We analyzed a large number of samples from CA, including Gabon, Cameroon, the Central African Republic (CAR) and the Democratic Republic of Congo (DRC), and present a population-based dataset of 1,404 samples, from 20 Bantu-speaking AGR populations (983 individuals) and nine PHG populations (421 individuals). We used a molecular approach with the highest resolution yet reported in CA, based on complete mtDNA sequences, to determine the phylogeny and phylogeography of mtDNAs from central African hunter–gatherers and farmers.

Results and Discussion

Impact of Lifestyle on Diversity and Demographic Patterns in Central African Belt. We characterized mtDNA variation in all samples by direct sequencing of the HVS-I (16024–16383) and by genotyping a set of 33 single-nucleotide polymorphisms (SNPs) from the coding region [supporting information (SI) Fig. 4] for accurate resolution into Hgs (SI Table 2). We used Hg profiles and the HVS-I sequence diversity of the entire collection of 1,404 samples (SI Table 3) to investigate the internal diversity and demography of the studied populations. AGR populations displayed higher levels of Hg diversity, sequence diversity, and mean numbers of pairwise differences than PHG populations (Table 1). Thirty-three subclades were identified, 32, 13, and 4 of which were present in Bantu-speaking AGR and western and eastern PHG populations, respectively (SI Table 2). Standard neutrality tests and mismatch distributions identified population expansion signatures among AGR, with negative values for Tajima's *D* and Fu's *Fs* tests (significance of most Fu's *Fs*, Table 1) and a clearly unimodal mismatch distribution (SI Fig. 5). These patterns contrasted with the nonsignificance of neutrality tests (Table 1) and the clearly multimodal mismatch distribution (SI Fig. 5) of PHG populations. All genetic diversity indices showed demographic differences between AGR and PHG, with AGR populations showing signs of population growth and PHG populations of small population sizes and strong genetic drift.

Dissecting L1c Phylogeny Based on Complete Mitochondrial Genomes. The L1c Hg predominated in CA, as reported (30, 32, 33, 35). The various L1c clades observed accounted for 35.7% (AGR populations) to 94% (western PHG populations) of the mtDNAs (SI Table 2). We investigated the internal structure of L1c, by complete genome sequencing of 27 mtDNA molecules covering the widest possible range of L1c variation, as inferred from HVS-I variation. The resulting mtDNA tree (Fig. 1) indicated an early split of L1c into L1c3 and L1c1′2′4′6, which then split into

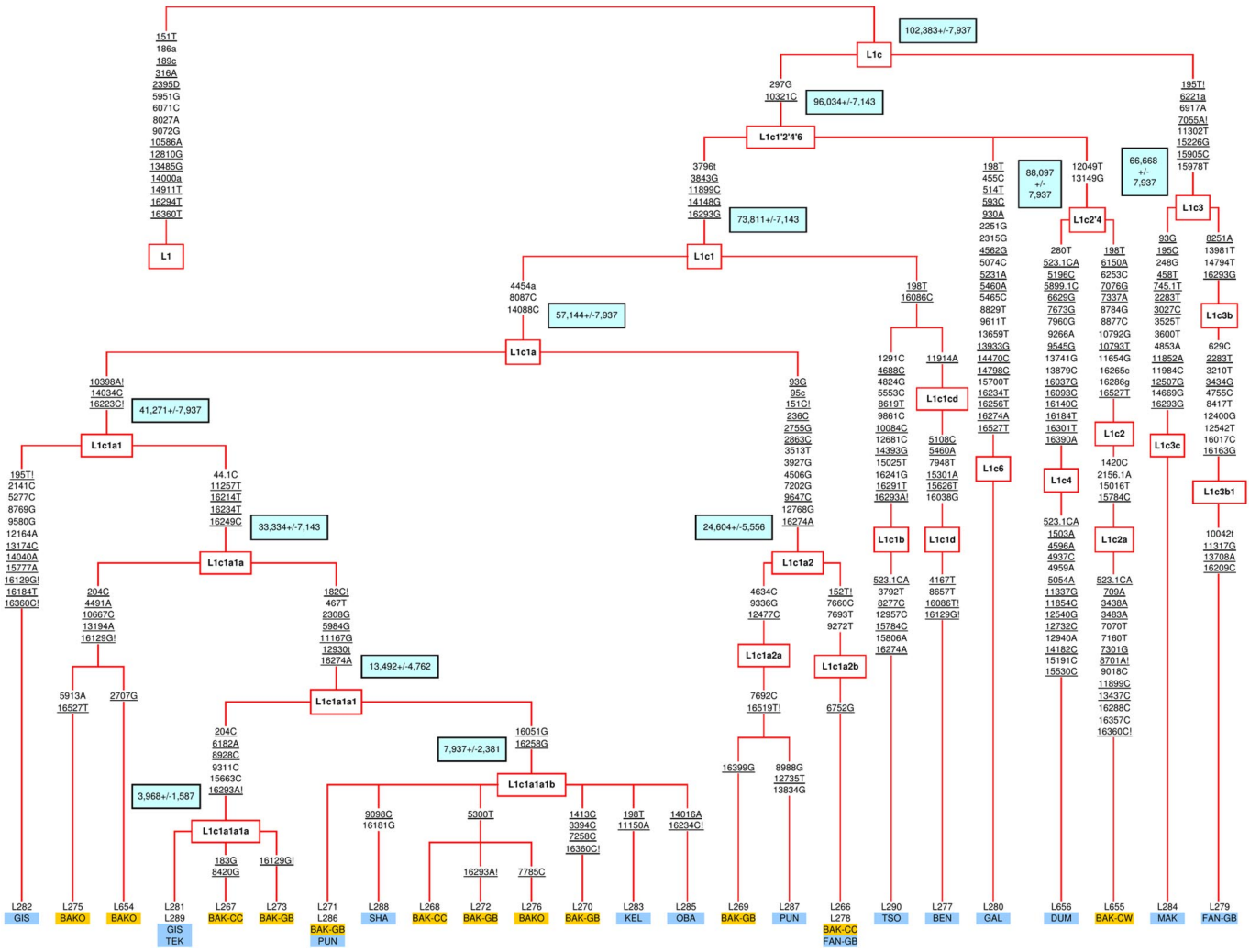

Fig. 1. Phylogenetic tree of complete mtDNA sequences belonging to haplogroup L1c. The tree is rooted on Hg L1 and shows subhaplogroup affiliations. Mutations are shown on the branches. Transitions are labeled in uppercase letters, transversions are indicated in lowercase letters, deletions are indicated by a ''d'' after the deleted nucleotide position, and insertions are indicated by a dot, followed by the number and type of inserted nucleotides. Underlined nucleotide positions occur at least twice in the tree. The exclamation mark (!) at the end of a nucleotide position denotes a reversion to the ancestral state in the relative pathway from the rCRS (36). Individuals highlighted in orange correspond to PHG and those in blue to Bantu-speaking AGR. Population codes for each individual are as in Table 1. Coalescence age estimates for the main subhaplogroups are also reported.

L1c1, L1c2′4, and L1c6. Coalescence time estimates for L1c and L1c1 were 102,600 ± 7,900 and 73,800 ± 7,100 YBP, respectively. Within L1c1, L1c1a coalesces at 57,100 ± 7,900 YBP and comprises the L1c1a1 and L1c1a2 sister lineages, which coalesce at 41,300 ± 7,900 and 24,600 ± 5,600 YBP, respectively. When naming the clades within the L1c topology (Fig. 1), we attempted to use the proposed HVS-I-based nomenclature (30, 33). We noted that the transition at 16293 alone is insufficient to define L1c1 as in ref. 30, although this control region site is the only one defining this branch. Based on complete genome information, we can now redefine the typical ''Pygmy'' clusters L1c1a and L1c1a1 (30) as L1c1a1a and L1c1a1a1, respectively. Finally, L1c2 retains its name but is now defined by nine coding region variants. Our complete mtDNA-based topology is not consistent with the recently proposed HVS-I-based clade L1c5 (33) (corresponding to L1c1a1 in ref. 30 and renamed L1c1a1a1 here). We did not use the L1c5 label in our topology, moving directly from L1c4 to L1c6 (represented by sample L280 in Fig. 1), to avoid confusion.

Maternal Diversity Is Homogeneous but Stratified by Time in Farming Populations. Despite the high levels of diversity of AGR populations from CA, the fraction of variation accounted for by interpopulation differences was very low (1.5%; $P < 0.0001$), indicating that almost all of the variation observed was within populations. Estimates of population differentiation (F_{ST}) based on Hg frequencies showed that ≈60% of interpopulation comparisons were not significant. This lack of population differentiation is illustrated in the scatterplot of the first two principal components (PC) on which the AGR populations are tightly clustered (Fig. 2). Thus, AGR populations from CA displayed a diverse homogeneous pattern with little apparent internal structure.

Almost all known subSaharan African Hgs (30) were represented among AGR populations, with the exception of L0d and L0k, which are typical of the Khoi and San peoples of South Africa (37). The most frequent Hgs (>5%) observed were L1c1a (20.3%), L2a1 (13.3%), L0a1 (6.5%), L1b (6.4%), L3e2 (5.8%), L3e1 (5.7%), and L3f1b (5.7%) (SI Table 2). We combined our data with a compiled database of >4,500 mtDNA profiles reported from the various African subregions (SI Table 4). L1c1a was the only Hg studied to show almost exclusive geographic clustering with CA (SI Table 5). The other Hgs displayed variable frequency and diversity patterns between West, East,

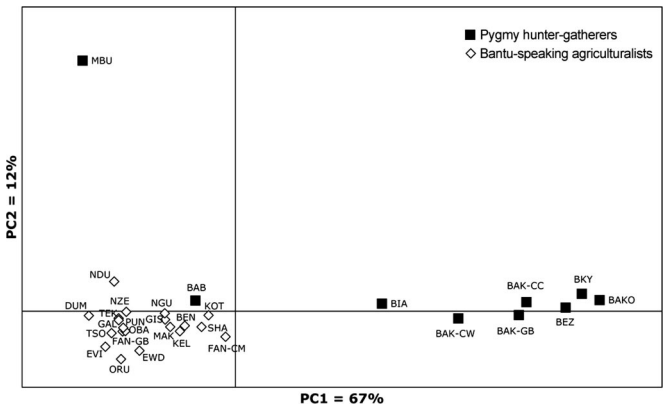

Fig. 2. PC plot based on haplogroup frequencies for the 29 population samples. For populations, codes, and geographic information, see Table 1.

and southeast Africa, indicating only that they originated in the equatorial zone. We investigated the geographic origin of L1c1a, by studying the frequency distribution and diversity of its ancestral types. L1c, with subclades L1c1-L1c6, had a frequency of 27% in CA, decreasing to ≈5% in both West and southeast Africa. The contemporary geographic distribution of L1c, including all its internal clades (Fig. 3), is consistent with the early arrival of L1c in CA (or a central African origin for this Hg, although this cannot be unambiguously proven), followed by a maturation phase in which L1c diversified (i.e., giving rise to its internal derivatives). The virtual restriction of L1c1a to CA provides evidence for an autochthonous origin of this lineage

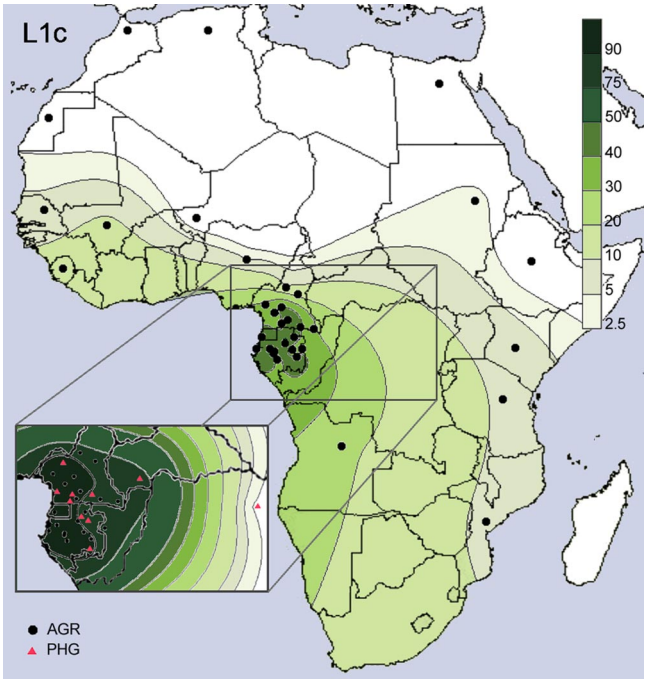

Fig. 3. Spatial frequency distribution of haplogroup L1c in Africa. The interpolation map was constructed by using the frequencies of L1c, as observed in our dataset, and those retrieved from all published data on the African continent (SI Tables 4 and 5) excluding PHG populations (therefore corresponding to the distribution of L1c among AGR populations). The map of the frequency of L1c in PHG populations and the geographic location of these populations are presented in the projected portion of CA, as presented in the lower lefthand corner of the figure. The PHG interpolation map is shown separately because of the unique demographic nature of these populations.

within this region. The coalescence age of L1c1a, estimated at 57,100 ± 7,900 YBP, is also ≈25,000 years greater than those estimated for the other Hgs frequent in central African AGR populations (15,800 to 29,400 YBP, SI Table 5). Our data therefore suggest that the contemporary maternal gene pool of AGR populations from CA has resulted from a gradual process in which an initial Central African gene pool dominated by L1c clades was subsequently enriched by the introgression (and/or expansion) of the L0a, L2, and L3 Hgs and their derivatives.

A Single Maternal Ancestor of Contemporary Western Pygmies. Our data showed a net separation between the eastern and western PHG of CA (all F_{ST} values for pairwise comparisons of eastern Mbuti with western PHG populations were very high (0.27–0.47) and highly significant), the Mbuti being an outlier on the PC plot (Fig. 2). We analyzed the Hg profile of the Mbuti based on coding-region sites. The Mbuti Hg profile comprised L0a2, L2*, L2a2, and L5 (SI Table 2). This maternal profile is qualitatively similar to that of eastern African populations (30) and very different from that of western PHG, indicating a lack of common maternal ancestry between eastern and western PHG. However, the extent to which this observation reflects a genuine general lack of common ancestry remains to be determined. Data for 80 autosomal loci from different African populations showed the Mbuti and Biaka PHG to be closely related, with both these groups more similar to West African than to East African populations (38). Further analyses of the paternally inherited Y-chromosome and sequence-based data from genome-wide autosomal regions are required to determine the mode and timing of divergence between western and eastern PHG.

The various populations of western PHG had similar Hg profiles and close affinities (Fig. 2, SI Table 2). The Babongo were an exception, clustering within the AGR group, suggesting greater gene flow between this population and AGR populations. Western PHG constituted a more heterogeneous group of populations than the AGR group in CA; the proportion of variation because of differences between populations was much higher for western PHG (11%; $P < 0.0001$) than for AGR (1.5%; $P < 0.0001$). Western PHG populations are dominated by two sister clades within L1c1a: L1c1a1 (53.4%) and L1c1a2 (30.1%) (SI Table 2). Thus, based on the genotyping of our large population sample, aided by complete mtDNA sequences, we can conclude that ≈83.5% of contemporary western PHG are descended from a single maternal ancestor of the autochthonous Central African Hg L1c1a.

Deep Common Ancestry and Asymmetric Gene Flow Between AGR and PHG. Despite differences in diversity and demography between AGR and PHG, our data clearly show that both groups have the same most common haplogroup. Hg L1c1a, in the form of the sister clades L1c1a1 and L1c1a2, has a frequency of 20.3% in AGR populations and 83.5% in PHG populations. This common heritage is also evident from HVS-I based lineages within L1c1a1 and L1c1a2. The three most common HVS-I haplotypes among AGR populations are the root HVS-I haplotype 16129 16187 16189 16223 16274 16278 16293 16294 16311 16360 of L1c1a2 (8.3%), the root HVS-I haplotype 16051 16129 16187 16189 16214 16234 16249 16258 16274 16278 16293 16294 16311 16360 of L1c1a1a1b (5.7%), and the haplotype 16129 16187 16189 16214 16234 16249 16274 16278 16294 16311 16360 of L1c1a1 (2.6%). These haplotypes were also the most frequent among western PHG, reaching frequencies of 29.6%, 22.3% and 13.9%, respectively (Fig. 1 and SI Table 3). The sharing of the most frequent Hg and HVS-I-based haplotypes between AGR and PHG clearly indicates common maternal ancestry between the two groups and/or high levels of gene flow between them.

The occurrence of various L1c clades in both AGR and PHG, the particularly high frequencies of this Hg among all western

EVOLUTION

PHG and the early coalescence age of the autochthonous L1c1a in CA (≈57,100 YBP) suggest that the maternal gene pool of the ancestors of contemporary AGR and PHG was dominated by the various L1c clades (probably including Hgs now extinct). Two populations arose from this presumed ancestral population: the modern AGR population, which includes various L1c clades (L1c1a, L1c1b, L1c1c, L1c2–6, etc.), and the western PHG population, in which L1c1a is the only surviving clade. The Pygmies must have split from this ancestral population no more than ≈73,800 years ago, when L1c1a began to diverge from L1c1 (Fig. 1). A long period of isolation (i.e., genetic and/or cultural) must then have occurred, accounting for the phenotypic differences characterizing PHG groups (5, 39). However, a common maternal ancestry and isolation alone cannot account for the current intimate sharing of L1c1a lineages in AGR and PHG populations. The isolation period must therefore have been interrupted at a certain point by gene flow. The L1c1a clade appears to have evolved within PHG, given (*i*) the very high frequency of this Hg in the western PHG groups from various geographic locations and (*ii*) phylogeographic patterns for HVS-I, showing that L1c1a lineages are slightly more diverse among PHG (L1c1a haplotype diversity: 0.772 ± 0.014 in PHG vs. 0.738 ± 0.021 in AGR) and that almost all those present among AGR are shared with PHG. Moreover, the lack of PHG- or AGR-specific well differentiated L1c1a lineages is not consistent with long isolation alone and suggests that subsequent gene flow must have occurred.

When considering the mode and timing of putative maternal gene flow between the ancestors of modern-day AGR and PHG populations, we must contemplate the different genetic, demographic, and cultural aspects of these groups. It seems unlikely that gene flow occurred in the AGR-to-PHG direction (actually, from the ancestors of farmers to Pygmies if gene flow occurred more than ≈4,000 YBP). First, gene flow from the diverse AGR populations, with their much greater diversity of Hgs, would have resulted in a much more assorted PHG gene pool than that currently observed (dominated by a single clade, L1c1a). Second, detailed ethnological data indicate that official marriages between AGR women and PHG men are forbidden in most societies. An exception to this cultural practice is found in the Babongo PHG. Intercultural marriages in both directions are more common in this population, accounting for the greater diversity of this population and its position on the PC plot (Fig. 2). By contrast, independent lines of evidence support instead long-standing maternal gene flow in the direction PHG-to-AGR, leading to enrichment of the AGR gene pool with L1c1a lineages. First, PHG women sometimes marry AGR men, and their children are integrated into the agricultural population (39). Second, contemporary demographic data estimate western PHG populations to consist of tens of thousands of individuals, whereas AGR populations are estimated to comprise tens of millions of individuals. The occurrence of L1c1a in ≈20% of AGR individuals is therefore unlikely to result from PHG-to-AGR gene flow in recent times. PHG-to-AGR gene flow must have occurred much earlier (from Pygmies to the ancestors of farmers), when the two populations were probably smaller and comparable, as would have been the case until the end of the Pleistocene (1). Third, our survey, based on complete mtDNA sequencing, shows that the same topographical structure of L1c1a is retrieved from both PHG and AGR, indicating substantial PHG-to-AGR gene flow since these populations first came into contact. The coalescence times of the two L1c1a lineages shared by AGR and western PHG are 41,200 for L1c1a1 and 24,600 for L1c1a2, indicating that PHG-to-AGR gene flow did not begin until ≈40,000 years ago and then continued until a few thousand years ago, as indicated by the sharing of minor and recent subclades of L1c1a between the two groups (e.g., the coalescence time of L1c1a1a1a is 3,960 ± 1,600, Fig. 1). These data thus provide insight into the long history of interactions between the ancestors of present-day AGR and PHG, which has been characterized by episodes of isolation followed by recurrent, long-term asymmetric gene flow between the two groups.

Conclusions

The mtDNA data presented here suggest that the ancestral population in CA that eventually gave rise to modern-day AGR and PHG populations, consisted principally of L1c clades that have survived to give the diverse forms observed among AGR, and essentially a single lineage among western PHG. The maternal gene pool composition of modern western PHG suggests a small number of ancestors that started to diverge from an ancestral Central African population no more than ≈70,000 YBP. After a period of isolation, accounting for current phenotypic differences between AGR and PHG, gene flow between the ancestors of the two groups began to occur no more than ≈40,000 YBP. Our data are consistent with continuous maternal gene flow from PHG-to-(proto)AGR over a long period. Unlike that of PHG, the proto-AGR maternal gene pool was enriched by the more recent arrival of L0a, L2, and L3 carriers, coinciding with the introduction of Late Stone Age technologies in the region and paving the way for the most important demographic, linguistic, and technological event in subSaharan Africa: the Bantu expansions.

Materials and Methods

Samples. We collected data for 1,404 individuals from different populations of Bantu-speaking AGR and PHG (Table 1). The AGR dataset corresponds to 983 individuals from 20 different populations, and the PHG dataset corresponds to 421 individuals from one eastern and eight western PHG populations. All individuals were unrelated healthy donors who gave appropriate informed consent.

We compared the mtDNA diversity in this dataset with 4,547 mtDNA profiles from various African subregions summarized in the mtDNA comprehensive database MURKA (SI Table 4). Because most of these studies have a lower resolution (Hg definition) than ours, we limited Hg definitions to the deepest common denominator available. For the HVS-I, we considered positions 16090–16365, which were common to all studies. Previously reported samples lacking Hg definition or with information for HVS-I not encompassing the 16090–16365 sequence range or containing ambiguities were eliminated from the analysis. For comparisons of AGR populations only, PHG samples were excluded from the analysis. The definition of subregions within Africa (SI Table 4) was as described (30).

mtDNA Sequencing. The first hypervariable segment (HVS-I) of the control region was sequenced in all samples, and variable positions were determined from position 16024 to 16383. The cytosine-track length variation at positions 16182 and 16183 in HVS-I was excluded from the analysis (SI Table 3). For complete mtDNA sequencing, 18 primers were used to yield nine overlapping fragments, as reported (40). The nine fragments were purified and sequenced, by using 56 internal primers to obtain the complete mtDNA genome. The complete mtDNA sequences reported here have been submitted to GenBank (accession nos. EU273476–EU273502). Sequence quality was ensured as follows: each base pair was determined once with a forward and once with a reverse primer, any ambiguous base call was checked by additional and independent PCR and sequencing reactions, and all sequences were examined by two independent investigators.

Hg Assignment. Based on the complete mtDNA sequences reported here and in the most recent mtDNA phylogenies (refs. 41 and 42 and D.M.B., unpublished work), 33 SNPs were tested in a hierarchical order (SI Fig. 4). Ten SNPs were initially genotyped in all samples, to identify the major Hgs to which they belonged (SI Fig. 4, in red). Within each Hg, we then genotyped a number of SNPs, to determine their exact location within the Hg (SI Fig. 4, in blue). These 33 diagnostic SNPs were genotyped by fluorescence polarization (VICTOR-2TM Technology; PerkinElmer) or by direct sequencing of the genomic region flanking the corresponding SNP. Finally, the polymorphic site at position 16241 in the HVS-I was used to identify L1c1b within L1c1; 16265c and 16286g were used to differentiate between L1c2 and L1c4 within L1c2′4; 16362 and

16274 were used to differentiate between L2b1 and L2b2 within L2b; 16264 was used to identify L2c2 within L2c; 16265t and 16264 were used to differentiate between L3e3 and L3e4 within L3e3′4; 16184 and 16325d were used to differentiate between L3e1a and L3e1b within L3e1; 16172 and 16189 were used to identify L3e2b within L3e2; and 16292 was used to identify L3f1b within L3f (SI Fig. 4).

L1c Phylogeny. The mtDNA tree of complete L1c sequences was drawn by hand, and its branches were subsequently validated by networks (21) constructed with Network 4.2.0.1 (www.fluxus-engineering.com). The hypervariable indels around positions 309, 315, and 16189 were excluded from the topology map. The average sequence divergence for each of the internal L1c clades was obtained by applying PAML (43) to the coding region polymorphisms, excluding indels, and by using the HKY85 substitution model. The calculation of the coalescence time in years followed (44) and was based on the addition of our samples to the complete L phylogeny based on 629 complete mtDNA sequences belonging to Hg L (D.M.B., unpublished data).

Statistical Analysis. Haplogroup and haplotype counts and diversity and population differentiation indices were calculated and sequence-based neutrality tests carried out with Arlequin 3.1 (45). DnaSP v. 4.1 (46) was used to calculate sequence mismatch distributions within PHG and AGR. The PC plot was obtained with GENALEX v. 6 software (47). The interpolation frequency map of L1c was obtained by using Surfer v. 6.04 (Golden Software), with the Kriging procedure, and estimates at each grid node were inferred from the entire dataset.

ACKNOWLEDGMENTS. We warmly thank all participants in this study. This work was supported by Institut Pasteur, a Language, Languages, and Human Origins (OHLL) grant from the Centre National de la Recherche Scientifique (CNRS), the European Science Foundation EUROCORES Origins of Man, Language, and Languages (OMLL) program ''Language, culture and genes in Bantu: A multidisciplinary approach to the Bantu-speaking populations of Africa,'' and the Ministry of Research project Action Concertée Incitative-Prosodie ''Histoire et diversité génétique des Pygmées d'Afrique Centrale et de leurs voisins.''

1. Diamond J, Bellwood P (2003) Farmers and their languages: The first expansions. *Science* 300:597–603.
2. Greenberg J (1972) Linguistic evidence regarding Bantu origins. *J African Hist* 17:189–216.
3. Newman J (1995) *The Peopling of Africa* (Yale Univ Press, New Haven, CT).
4. Phillipson D (1993) *African Archaeology* (Cambridge Univ Press, Cambridge, UK).
5. Cavalli-Sforza LL (1986) in *African Pygmies*, ed Cavalli-Sforza LL (Academic, Orlando, FL), pp 361–426.
6. Bahuchet S (1999) in *The Cambridge Encyclopedia of Hunters and Gatherers*, eds Richard B, Daly R (Cambridge Univ Press, Cambridge UK), pp 190–194.
7. Diamond JM (1991) Anthropology. Why are pygmies small? *Nature* 354:111–112.
8. Hiernaux J (1974) *The People of Africa* (Weidenfeld and Nicolson, London).
9. Froment A (1993) Adaptation biologique et variation dans l'espèce humaine: Le cas des Pygmées d'Afrique. *Bull et Mém Soc Anthropol Paris* 5:417–448.
10. Garrigan D, Hammer MF (2006) Reconstructing human origins in the genomic era. *Nat Rev Genet* 7:669–680.
11. Cavalli-Sforza LL, Feldman MW (2003) The application of molecular genetic approaches to the study of human evolution. *Nat Genet* 33 Suppl:266–275.
12. Cavalli-Sforza LL, Menozzi P, Piazza A (1994) *The History and Geography of Human Genes* (Princeton Univ Press, Princeton).
13. Cann RL, Stoneking M, Wilson AC (1987) Mitochondrial DNA, human evolution. *Nature* 325:31–36.
14. Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC (1991) African populations and the evolution of human mitochondrial DNA. *Science* 253:1503–1507.
15. Ingman M, Kaessmann H, Pääbo S, Gyllensten U (2000) Mitochondrial genome variation and the origin of modern humans. *Nature* 408:708–713.
16. Quintana-Murci L, *et al.* (1999) Genetic evidence of an early exit of *Homo sapiens sapiens* from Africa through eastern Africa. *Nat Genet* 23:437–441.
17. Macaulay V, *et al.* (2005) Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* 308:1034–1036.
18. Olivieri A, *et al.* (2006) The mtDNA legacy of the Levantine early Upper Palaeolithic in Africa. *Science* 314:1767–1770.
19. Mellars P (2006) Going east: New genetic and archaeological perspectives on the modern human colonization of Eurasia. *Science* 313:796–800.
20. Pakendorf B, Stoneking M (2005) Mitochondrial DNA, human evolution. *Annu Rev Genomics Hum Genet* 6:165–183.
21. Bandelt H-J, Forster P, Sykes BC, Richards MB (1995) Mitochondrial portraits of human populations using median networks. *Genetics* 141:743–753.
22. Coia V, *et al.* (2005) Brief communication: mtDNA variation in North Cameroon: Lack of Asian lineages and implications for back migration from Asia to sub-Saharan Africa. *Am J Phys Anthropol* 128:678–681.
23. Beleza S, Gusmão L, Amorim A, Carracedo A, Salas A (2005) The genetic legacy of western Bantu migrations. *Hum Genet* 117:366–375.
24. 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.
25. Soodyall H, Vigilant L, Hill AV, Stoneking M, Jenkins T (1996) mtDNA control-region sequence variation suggests multiple independent origins of an ''Asian-specific'' 9-bp deletion in sub-Saharan Africans. *Am J Hum Genet* 58:595–608.
26. Watson E, Forster P, Richards M, Bandelt H-J (1997) Mitochondrial footprints of human expansions in Africa. *Am J Hum Genet* 61:691–704.
27. Alves-Silva J, *et al.* (2000) The ancestry of Brazilian mtDNA lineages. *Am J Hum Genet* 67:444–461.
28. Bandelt H-J, *et al.* (2001) Phylogeography of the human mitochondrial haplogroup L3e: A snapshot of African prehistory and Atlantic slave trade. *Ann Hum Genet* 65:549–563.
29. Pereira L, *et al.* (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.
30. Salas A, *et al.* (2002) The making of the African mtDNA landscape. *Am J Hum Genet* 71:1082–1111.
31. Destro-Bisol G, *et al.* (2004) Variation of female and male lineages in sub-Saharan populations: the importance of sociocultural factors. *Mol Biol Evol* 21:1673–1682.
32. Rando JC, *et al.* (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.
33. Batini C, *et al.* (2007) Phylogeography of the human mitochondrial L1c haplogroup: genetic signatures of the prehistory of Central Africa. *Mol Phylogenet Evol* 43:635–644.
34. Destro-Bisol G, *et al.* (2004) The analysis of variation of mtDNA hypervariable region 1 suggests that Eastern and Western Pygmies diverged before the Bantu expansion. *Am Nat* 163:212–226.
35. Plaza S, *et al.* (2004) Insights into the western Bantu dispersal: mtDNA lineage analysis in Angola. *Hum Genet* 115:439–447.
36. Andrews RM, *et al.* (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA *Nat Genet* 23:147.
37. Knight A, *et al.* (2003) African Y chromosome and mtDNA divergence provides insight into the history of click languages. *Curr Biol* 13:464–473.
38. Tishkoff SA, Kidd KK (2004) Implications of biogeography of human populations for 'race' and medicine. *Nat Genet* 36:S21–27.
39. Bahuchet S (1993) in *Tropical forests, people and food. Biocultural interactions and applications to development*, eds Hladik CM, *et al.* (Unesco/Parthenon, Paris/Lancashire, UK), pp 37–54.
40. Taylor RW, Taylor GA, Durham SE, Turnbull DM (2001) The determination of complete human mitochondrial DNA sequences in single cells: Implications for the study of somatic mitochondrial DNA point mutations. *Nucleic Acids Res* 29:E74.
41. Kivisild T, *et al.* (2006) The role of selection in the evolution of human mitochondrial genomes. *Genetics* 172:373–387.
42. Torroni A, Achilli A, Macaulay V, Richards M, Bandelt H-J (2006) Harvesting the fruit of the human mtDNA tree. *Trends Genet* 22:339–345.
43. Yang Z (1997) PAML: A program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* 13:555–556.
44. Mishmar D, *et al.* (2003) Natural selection shaped regional mtDNA variation in humans. *Proc Natl Acad Sci USA* 100:171–176.
45. Excoffier L, Laval L, Schneider S (2005) Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evol Bioinformatics* 1:47–50.
46. Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19:2496–2497.
47. Peakall R, Smouse PE (2006) GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes* 6:288–295.