Insights into the Demographic History of African Pygmies from Complete Mitochondrial Genomes

Chiara Batini,†,[1,2] Joao Lopes,[3] Doron M. Behar,[4] Francesc Calafell,[1,5] Lynn B. Jorde,[6] Lolke van der Veen,[7] Lluis Quintana-Murci,[8] Gabriella Spedini,[2,9] Giovanni Destro-Bisol,[2,9] and David Comas*,[1,5]

[1]Institut de Biologia Evolutiva (CSIC-UPF), Department de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Barcelona, Spain

[2]Dipartimento di Biologia Animale e dell'Uomo, Sapienza Universita' di Roma, Roma, Italy

[3]Centro de Ciencias do Mar, Faculdade de Ciências e Tecnologia, Universidade do Algarve, Faro, Portugal

[4]Molecular Medicine Laboratory, Rambam Health Care Campus, Haifa,, Israel

[5]Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública, Barcelona, Spain

[6]Department of Human Genetics, University of Utah Health Sciences Center

[7]Laboratoire Dynamique du Language, UMR 5596, Institut des Sciences de l'Homme, Lyon, France

[8]Human Evolutionary Genetics, CNRS URA3012, Institut Pasteur, Paris, France

[9]Istituto Italiano di Antropologia, Roma, Italy

†Present address: Department of Genetics, University of Leicester, Leicester, United Kingdom.

*Corresponding author: E-mail: david.comas@upf.edu.

Associate editor: Anne Stone

Abstract

Pygmy populations are among the few hunter-gatherers currently living in sub-Saharan Africa and are mainly represented by two groups, Eastern and Western, according to their current geographical distribution. They are scattered across the Central African belt and surrounded by Bantu-speaking farmers, with whom they have complex social and economic interactions. To investigate the demographic history of Pygmy groups, a population approach was applied to the analysis of 205 complete mitochondrial DNA (mtDNA) sequences from ten central African populations. No sharing of maternal lineages was observed between the two Pygmy groups, with haplogroup L1c being characteristic of the Western group but most of Eastern Pygmy lineages falling into subclades of L0a, L2a, and L5. Demographic inferences based on Bayesian coalescent simulations point to an early split among the maternal ancestors of Pygmies and those of Bantu-speaking farmers (~70,000 years ago [ya]). Evidence for population growth in the ancestors of Bantu-speaking farmers has been observed, starting ~65,000 ya, well before the diffusion of Bantu languages. Subsequently, the effective population size of the ancestors of Pygmies remained constant over time and ~27,000 ya, coincident with the Last Glacial Maximum, Eastern and Western Pygmies diverged, with evidence of subsequent migration only among the Western group and the Bantu-speaking farmers. Western Pygmies show signs of a recent bottleneck 4,000–650 ya, coincident with the diffusion of Bantu languages, whereas Eastern Pygmies seem to have experienced a more ancient decrease in population size (20,000–4,000 ya). In conclusion, the results of this first attempt at analyzing complete mtDNA sequences at the population level in sub-Saharan Africa not only support previous findings but also offer new insights into the demographic history of Pygmy populations, shedding new light on the ancient peopling of the African continent.

Key words: mitochondrial genome, African Pygmies, coalescent simulations, demography, phylogeography.

Introduction

African Pygmy populations are one of the few human groups identified by their physical appearance rather than ethnographic, cultural, geographical, or linguistic criteria. Their height, among the smallest recorded for human populations (Cavalli-Sforza 1986b; Hitchcock 1999), has been interpreted as the consequence of different selective pressures (reviewed in Perry and Dominy 2009). These include thermoregulatory adaptation to the environment of the tropical forest (Cavalli-Sforza et al. 1993), reduction of the total caloric intake in a food-limited environment (Hart and Hart 1986), improved mobility in the dense forest (Diamond 1991), or advantageous earlier reproductive age in

short lifespan conditions (Migliano et al. 2007). Nowadays, these populations live scattered in the Central African rainforest and are clustered into two main groups, Western and Eastern Pygmies. The former is estimated to include 55,000 individuals inhabiting the Western Congo basin, across Cameroon, Republic of Congo, Gabon, and Central African Republic, and its subgroups are identified by different names, such as Binga, Baka, Biaka, and Aka (Cavalli-Sforza 1986c; Hitchcock 1999). Eastern Pygmies number approximately 30,000 individuals living in the Ituri forest (in the north-east of the Democratic Republic of Congo) and are usually referred to as Mbuti (Cavalli-Sforza 1986c; Hitchcock 1999). Other minor and scattered groups of Pygmies

are found in the Democratic Republic of Congo, Rwanda, and Burundi and are identified as Twa (Cavalli-Sforza 1986c). Most Pygmy groups live as hunter-gatherers, but none base their subsistence exclusively on forest products because they trade with neighboring farmers, creating a complex network of economic and social exchange. Intermarriage exists but seems to be mostly limited to unions between Pygmy females and farmer males (Cavalli-Sforza 1986a; Sayer et al. 1992). Pygmies speak languages that belong to Central Sudanic, Adamawa-Ubangian, or Bantu groups, mirroring those of their farmer neighbors. Several studies have attempted to identify remnants of an ancient Pygmy language, which might have been lost after contact with farmers (Letouzey 1976; Bahuchet 1993; Demolin 1996).

The issue of the origin of African Pygmies has stimulated a great deal of research because of their particular physical characteristics and their possible continuity with the first communities inhabiting Central Africa. However, this issue is still controversial, and different scenarios have been proposed. One hypothesis suggests that Pygmy ancestors occupied the equatorial forest since ancient times as a single group and diverged into Eastern and Western branches recently, around 5 thousands of years ago (Kya), when Bantu-speaking farmers expanded from the current Nigeria/Cameroon border and migrated southward through the tropical forest (Cavalli-Sforza 1986a). Pygmies from the west probably admixed with Bantu-speaking agriculturalists to a greater extent than those from the east, who are therefore regarded as the "purer" Pygmy group. Alternatively, the differences found between the Pygmy groups have been explained on the basis of independent and older origins, as ancient as the divergence between Pygmies' and farmers' ancestors (Hiernaux 1974, 1977). Finally, a third hypothesis suggests that Pygmy groups are the descendants of a specialized hunting-gathering subcaste of Bantu and Adamawa-Ubangian–speaking populations, which evolved to seasonally exploit the tropical forest. In this case, the divergence of Pygmies would trace back to 4–5 Kya, the time when Bantu-speaking agriculturalists started their expansion through the forest environment (Blench 1999).

Genetic studies have provided useful insights into the origin and relationships of Pygmies with other sub-Saharan populations. In general, autosomal data have highlighted a substantial homogeneity among Niger-Congo–speaking groups (including Bantu) and a deep structure among hunter-gatherer communities in sub-Saharan Africa, although the data are still scanty and contradictory (Rosenberg et al. 2002; Watkins et al. 2003; Jakobsson et al. 2008; Li et al. 2008; Tishkoff et al. 2009). Whole-genome analyses in sub-Saharan Africa have shown that hunter-gatherers, including Pygmies, are located near the root of African diversity (Jakobsson et al. 2008; Li et al. 2008; Tishkoff et al. 2009), suggesting a common origin for hunter-gatherers with an ancient divergence. The reconstruction of migration patterns and the estimation of population sizes and divergence times was first attempted in two studies of autosomal variation using coalescent simulations. Analysis of 28 autosomal STRs in Western Pygmies and neighboring

populations pointed to an ancient origin of Western Pygmies followed by a recent separation within this group (~3 Kya), coincident with the expansion of farmers (Verdu et al. 2009). Even more recently, the sequencing of ~33 kb of autosomal neutral regions in Pygmies and neighboring populations (Patin et al. 2009) suggested a common origin of Eastern and Western Pygmies (with a separation time of ~20 Kya) and their early divergence from the ancestors of neighboring farmers (~60 Kya), with differential migration patterns and effective population sizes.

Phylogeographic analyses of uniparental genomes, Y chromosome, and mitochondrial DNA (mtDNA) have provided evidence for asymmetrical gene flow between Pygmies and Bantu-speaking farmers, identified Pygmy-specific mtDNA lineages, and highlighted the different mtDNA haplogroup composition among the two Pygmy groups (Destro-Bisol, Coia, et al. 2004; Destro-Bisol, Donati, et al. 2004; Wood et al. 2005; Batini et al. 2007; Behar et al. 2008; Quintana-Murci et al. 2008; Berniell-Lee et al. 2009), although the paucity of data for Eastern Pygmies makes further sampling necessary for more robust inferences to be made. The two Pygmy groups share a high frequency of the Y chromosome B2b lineage (Wood et al. 2005; Berniell-Lee et al. 2009), which is also found in Khoisan-speaking populations, suggesting a possible common root among African hunter-gatherers. By contrast, Western Pygmies are distinguished by the very high frequencies (up to 100%) of specific subclades of the L1c mtDNA haplogroup (Batini et al. 2007; Quintana-Murci et al. 2008) while such lineages have not been detected in Eastern Pygmies, suggesting a possible ancient maternal separation between the two groups of Pygmies (Destro-Bisol, Coia et al. 2004; Destro-Bisol et al. 2006).

Analysis of complete mtDNA genomes has provided a refined phylogeny of maternal lineages (Ingman et al. 2000; Finnila et al. 2001; Maca-Meyer et al. 2001; Herrnstadt et al. 2002; Roostalu et al. 2007; Achilli et al. 2008; Behar et al. 2008; Quintana-Murci et al. 2008). These studies have focused on individuals, whose mtDNAs belong to specific haplogroups, making it possible to estimate their diversity and time depth. However, no inferences directly dealing with demography and history of human populations were drawn.

The present investigation tackles for the first time the analysis of complete mtDNA genomes at the population level in sub-Saharan Africa, with the aim of unraveling the history and evolution of Pygmies. Our results shed new light on the origin of maternal lineages of African Pygmies and their relationships with neighboring populations, providing estimates of divergence times, changes in effective population sizes, and female migration rates.

Materials and Methods

Population samples and database
The whole-mitochondrial genome was sequenced in a total of 205 individuals from ten Central African populations (see fig. 1 for population location). A population-based
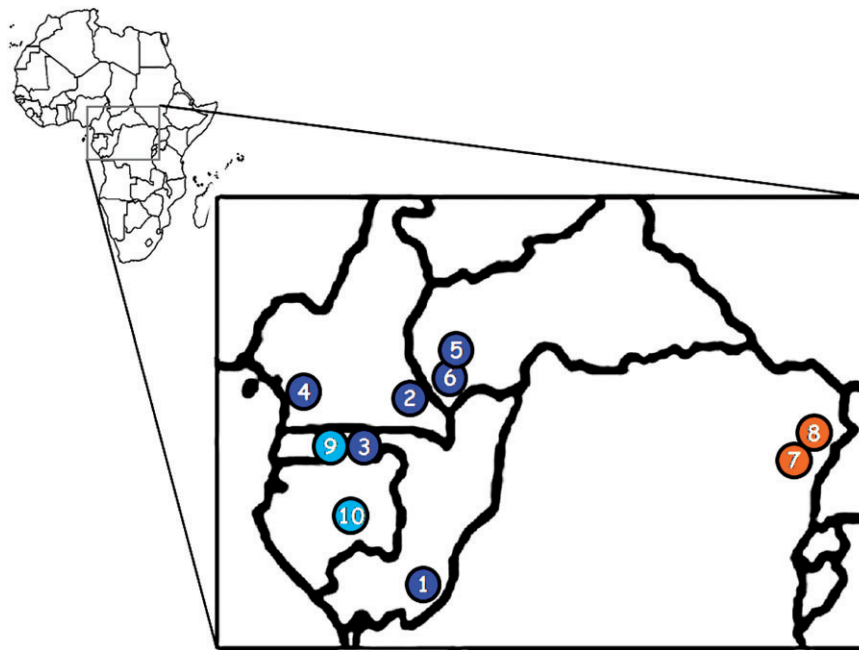
Fig. 1. Geographical location of populations analyzed in the present study. 1, Babinga; 2, BakaC (from Cameroon); 3, BakaG (from Gabon); 4, Bakola; 5, Biaka; 6, Mbenzele; 7, MbutiCEPH; 8, Mbuti; 9, Fang; 10, Nzebi. Blue, Western Pygmies; purple, Eastern Pygmies; light blue, Bantu-speaking farmers.

approach was applied for the selection of the samples, ignoring any previous information about the haplogroup classification derived from hypervariable regions. The data set includes 169 individuals from eight Pygmy populations: six Western (WPYG: 23 Babinga from the Republic of Congo, 27 Baka from Cameroon, 20 Baka from Gabon, 11 Bakola from Cameroon, 23 Biaka from Central African Republic, 30 Mbenzele from Central African Republic) and two Eastern (EPYG: 2 different Mbuti samples from the Democratic Republic of Congo with 14 and 21 individuals, respectively). Furthermore, we analyzed 36 individuals from two Bantu-speaking farming populations (WAGR: 17 Fang and 19 Nzebi from Gabon). All DNA samples were obtained from blood or buccal swabs and collected from unrelated healthy individuals who gave appropriate informed consent.

A database of 768 additional sequences was built for phylogeographic comparison (supplementary table S1, Supplementary Material online; Behar et al. 2008; Just et al. 2008).

Complete mtDNA Sequencing and Haplogroup Classification

The complete mitochondrial genome was amplified in four overlapping fragments (ranging from 4 to 5 kb) using the four primer pairs reported in supplementary table S2 (Supplementary Material online). Polymerase chain reactions (PCRs) (23 µl) contained $1\times$ EcoTaq buffer, 220 µM dNTPs, 1 mM MgCl$_2$, primers at 0.45 mM each, 2 U of EcoTaq polymerase, and 10–50 ng of DNA. Samples were denatured for 2 min at 94 °C; amplified for 14 cycles at 94 °C for 20 s, 60 °C for 30 s, and 68 °C for 5.5 min; for 16 cycles at 94 °C for 20 s, 60 °C for 30 s, and 68 °C for 5.5 min + 15 s per cycle; and given a final extension at 68 °C for 10 min. PCR prod-

ucts were purified using a MultiScreen PCRµ96 Filter Plate through a size-exclusion membrane and vacuum filtration.

The four resulting fragments were sequenced in a total of 32 reactions on the light chain of the mtDNA using the forward primers (L primers) described in Maca-Meyer et al. (2001). An alternative primer was designed: L10403 was replaced by L10396 (5′CTACAAAAAGGATTAGACTG3′) due to the common presence in African populations of a polymorphism at position 10398. In some samples, a poly-cytosine length polymorphism at positions 303–315, 567–573, 5894–5899, 8272–8278, and 16184–16193 prevented reading of the final tract of the sequence. In these cases, the heavy chain was sequenced using reverse primers (H408, H1232, H6460, H8416, and H16401). The heavy chain was also sequenced in cases of ambiguity or possible phantom mutations. Except for primers H1232 (5′CTGAGCAAGAGGTGGTGAGG3′), H6460 (5′TGCT-GTGATTAGGACGGATC3′), and H8416 (5′TGATGAGGA-ATAGTGTAAGG3′), all reverse primers were previously published in Maca-Meyer et al. (2001). The sequence reaction was performed with the BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems) and sequencing products purified using a Montage SEQ96 Sequencing Reaction Cleanup Kit (Millipore) through a size-exclusion membrane and vacuum filtration. Sequence products were run on an ABI PRISM 3100 sequencer (Applied Biosystems).

Sequences were assembled using SeqMan 5.05 software contained in the Lasergene 5.0 package (DNASTAR, Inc.) and annotated according to the revised Cambridge Reference Sequence (rCRS; Andrews et al. 1999). Because four independent overlapping PCR amplicons were sequenced, a total of 1,600 bp (~10%) for each individual have been sequenced twice from independent fragments, which

allowed us to use them as an internal sequencing control. In addition, quality control was carried out through a phylogenetic approach (see Behar et al. 2008), and previously unobserved mutations, as well as unexpected patterns, were rechecked through resequencing. The 121 different sequences found in this study have been deposited in Gen-Bank (accession numbers HM771113–HM771233). Haplotypes and their absolute frequencies in the populations analyzed are also reported in supplementary table S3 (Supplementary Material online). Length polymorphisms at positions 303–315, 567–573, 5894–5899, 8272–8278, and 16184–16193 were excluded from all analyses. Sequences were assigned to previously described haplogroups according to Behar et al. (2008).

Data Analysis

Intra-population diversity parameters (number of sequences, number of polymorphic sites, sequence diversity, mean number of pairwise differences, and nucleotide diversity) and neutrality tests (Tajima's D and Fu's Fs) were calculated with the Arlequin 3.11 package (Excoffier et al. 2005) for three different data sets: complete sequences, coding region (positions 577–16023), and control region (positions 16024–576).

Pairwise difference genetic distances between populations were calculated with the Arlequin 3.11 package. The distance matrix, corrected for nonsignificant values to 0, was represented in a multidimensional scaling (MDS) plot using the SPSS 15.0 software. In order to test whether the difference between two genetic distances was significantly larger (or smaller) than zero, a permutation test was performed. Population samples were created by sampling with replacement from each original sample; genetic distances between these samples were computed, and the sign of the difference was recorded. This was repeated 100,000 times, and the real difference was deemed significant if the original sign was found in at least 95% of the iterations.

A total of 973 complete sequences belonging to haplogroup L (present data and data from Behar et al. 2008 and Just et al. 2008) was represented in a median-joining network (Bandelt et al. 1999) using Network 4.5 (available at www.fluxus-engineering.com). The control region was excluded to allow comparisons with all available data (e.g., those published in Kivisild et al. 2004) and to avoid reticulations arising from recurrent mutations. Therefore, the resulting range was between nucleotide positions 435 and 16023 with respect to rCRS. In order to weight each position according to its evolutionary rate, the parameters suggested in Kong et al. (2008) were used. The time depth of different haplogroups (time to the most recent common ancestor [TMRCA]) was estimated from the coding region using the software BEAST 1.5.3 (Drummond et al. 2005; Drummond and Rambaut 2007). Markov chain Monte Carlo (MCMC) samples were based on 100,000,000 generations, logging every 1,000 steps, with the first 10,000,000 generations discarded as the burn-in. We used a constant size coalescent tree prior, HKY substitution model, and a strict clock with a mean substitution rate of 1.16649E-8 substitutions/site/year (adapted from Soares et al. 2009 to the sequence range considered in the present analysis).

Three main evolutionary scenarios were tested through an Approximate Bayesian Computation (ABC) approach (Beaumont et al. 2002). The scenarios considered were a common recent origin of Pygmy populations (Cavalli-Sforza 1986a), an independent origin of Pygmies and a shared history among populations of the same geographical region (Hiernaux 1974, 1977), and an external scenario that assumes a common origin of Eastern Pygmies and agriculturalists who diverged only after the separation from the Western Pygmies (see supplementary fig. S1, Supplementary Material online). Each of these scenarios was tested with five different patterns of migration among populations (absent, among all three populations, and pairwise), resulting in 15 different topologies (see supplementary fig. S1, Supplementary Material online). In order to focus on neutrally evolving sites and to avoid the confounding effects of common homoplasy in the control region, we built a data set with only the sequence of the 13 protein-coding genes (with the ND6 gene reverse complemented to give the same reading direction as the other genes), in which nonsynonymous mutations were considered nonpolymorphic, as previously reported (Soares et al. 2009). Priors for divergence times (t) were defined on the basis of archaeological evidence (Phillipson 1993; Cornelissen 2002; Marean and Assefa 2005) (table 1). Priors for migration rates (m) and effective population sizes (Ne) were set according to previous simulation-based studies (Patin et al. 2009; Verdu et al. 2009) (table 1).

The model selection was performed using an ABC-regression method (e.g., Beaumont et al. 2002; Fagundes et al. 2007). Each model was given a prior probability of 1/15 so that their prior distribution was uniform. The popABC program (Lopes et al. 2009) was used to perform the ABC algorithm by simulating 500,000 data sets per model, thus obtaining 7,500,000 total simulated data. The tolerance for the rejection step was set to 0.12% (9,000 data accepted). Prior to model selection, the ABC-regression method was tested against synthetic data (i.e., simulated data sets for which we knew the true values of the parameters) under the same conditions that were applied to the experimental data set. We ran the method on 500 synthetic data for each model considered (a total of 7,500 data sets). The overall performance of the ABC method was good (supplementary fig. S2, Supplementary Material online), in that the true model was estimated in up to 75% of cases and interestingly, when grouping the models according to migration pattern, this proportion reached almost 85% in all groups.

After model choice, a standard ABC-regression method was used to estimate the historical demographic parameters. A total of 3,000,000 data sets were simulated, from which the closest 0.3% (9,000 simulations) were accepted. The point estimates were taken as the mode of the posterior distributions. The prior probability for the mutation rate was set as a lognormal distribution of base 10 with a mean of −2.57 and a standard deviation (SD) of 0.055

Table 1. Prior and Posterior Distributions for Demographic Parameters Estimated on Model 1 Through an ABC Approach.

Parameter	Description	Prior distribution	Posterior distribution: mode (CI)
Ne1	EPYG effective population size	Uniform (10,10000)	4,547 (2,246–8,387)
Ne2	WPYG effective population size	Uniform (10,10000)	1,431 (653–2,820)
Ne3	WAGR effective population size	Uniform (10,100000)	35,850 (18,650–66,640)
NeA1	Effective population size of ancestral A1	Uniform (10,10000)	5,720 (1,181–10,000)
NeA2	Effective population size of ancestral A2	Uniform (10,10000)	4,922 (1,487–9,556)
t1	First splitting time	Uniform (4000,t2-39000)	27,305 (10,249–56,936)
t2	Second splitting time	Uniform (40000,110000)	70,866 (51,789–106,388)
m1	Migrants into EPYG	Uniform (0,0.005)	—
m2	Migrants into WPYG	Uniform (0,0.005)	1.30×10^{-03} ($0–3.69 \times 10^{-03}$)
m3	Migrants into WAGR	Uniform (0,0.0005)	2.93×10^{-04} ($0–4.75 \times 10^{-04}$)
mA1	Migrants into A1	Uniform (0,0.005)	2.35×10^{-03} ($0–4.36 \times 10^{-03}$)

NOTE.—CI, credible intervals.

(following Soares et al. 2009). Again, prior to the estimation of the parameters, the standard ABC-regression method was tested against synthetic data under the same conditions that were applied to the experimental data set. We considered 10,000 synthetic data simulated using parameter values sampled from the chosen prior distributions. For all the parameters, about 95% of the time, the true values were within the 95% credible interval. However, the size of a typical 95% credible interval varied considerably between parameters. In fact, posterior distributions obtained for NeA1, m3, and mA (see table 1) were mostly too flat for their point estimate to be considered reliable, whereas overall the other parameters (t1, t2, Ne1, Ne2, Ne3, and m2; table 1) had quite accurate posterior distributions. These results suggest that we can be confident of our analysis regarding the latter set of parameters.

The summary statistics chosen for the calculation of the distances were the number of haplotypes, number of segregating sites, and the average number of pairwise differences. These were computed for each population individually and for both populations pooled together. Thus, the distances were calculated from a total of 18 summary statistics. These were normalized, subtracting their mean and dividing by their SD, so that the different nature of their units would not lead to biases in the calculation of the Euclidean distances.

For each of the three population groups (WPYG, EPYG, and WAGR), a Bayesian Skyline Plot (BSP) of effective population size over time was generated using the software BEAST 1.5.3 (Drummond et al. 2005; Drummond and Rambaut 2007). The data set used for these analyses was the same as that for ABC simulations. MCMC samples were based on 100,000,000 generations, logging every 1,000 steps, with the first 10,000,000 generations discarded as the burn-in. We used a HKY substitution model and a strict clock with a mean substitution rate of 1.1131E-8 substitutions/site/year (Soares et al. 2009) and a generation time of 25 years.

Results

Intra- and Interpopulation Variation in Central African Complete Maternal Lineages

A total of 121 different complete mtDNA sequences were obtained from 205 individuals from ten Central African populations (six Western Pygmy, WPYG; two Eastern Pygmy, EPYG; and two Bantu-speaking agriculturalist, WAGR; fig. 1). A total of 585 transitions and 27 transversions, of which 282 were synonymous and 120 nonsynonymous mutations, were observed over the whole data set—values which are similar to previous studies (Soares et al. 2009).

Intra-population diversity parameters are shown in table 2. The frequencies of different (k/N) and private (kp/k) complete mtDNA sequences are lower in WPYG (mean values of 0.522 and 0.385, respectively) compared with EPYG (0.821 and 0.871, respectively) and WAGR (0.895 and 0.837, respectively). Notably, no private sequences were observed among the Baka from Gabon. Variation at the intra-population level was also calculated separately for coding and control regions (supplementary table S4, Supplementary Material online). As expected, due to the larger number of nucleotides considered, an increase in sequence diversity is observed when complete (up to 25%) or coding (up to 18.5%) data sets are compared with the control region, which is the portion of the mtDNA sequence usually analyzed in population studies. Larger values are also obtained for other diversity parameters, in particular the number of polymorphic sites and the mean number of pairwise differences. The only diversity parameter that is reduced for the coding region is the nucleotide diversity, as expected because of more intense functional constraints. Neutrality tests showed no significant value in any data set (table 2 and supplementary table S4, Supplementary Material online), with the exception of the signal of the Fs test in the agriculturalist Fang for the control region.

In order to explore interpopulation variation, an MDS plot based on pairwise genetic distances (supplementary table S5, Supplementary Material online) was drawn. The 2D plot (fig. 2) presented a low stress value (0.013), which is lower than the 1% cutoff value of 0.133 ascertained in Sturrock and Rocha (2000). Pygmy populations cluster in two main groups according to their geographical origin and are located at opposite ends of the first MDS dimension (mean distance value, 0.525). WAGR show an intermediate position between the two groups of Pygmies, but their average genetic distance from EPYG (0.224) is smaller than from WPYG (0.294). After running a one-tailed permutation tests with 100,000 iterations, we found that the

Table 2. Internal Variation Parameters Calculated on the Complete Data Set.

Population	N	k	S	k/N	kp/k	HD (SD)	MNPD (SD)	π (SD)	D (P value)	Fs (P value)
Babinga	23	10	122	0.435	0.500	0.889 (0.037)	37.968 (17.135)	0.002 (0.001)	0.561 (0.748)	11.203 (0.996)
BakaC	27	12	178	0.444	0.417	0.855 (0.048)	50.855 (22.709)	0.003 (0.001)	0.262 (0.672)	13.397 (0.999)
BakaG	20	9	53	0.450	0.000	0.889 (0.042)	19.579 (9.044)	0.001 (0.001)	1.253 (0.928)	5.876 (0.978)
Bakola	11	7	54	0.636	0.429	0.909 (0.066)	21.564 (10.316)	0.001 (0.001)	0.918 (0.857)	3.652 (0.938)
Biaka	23	18	167	0.783	0.556	0.968 (0.026)	42.356 (19.078)	0.003 (0.001)	−0.178 (0.495)	0.981 (0.657)
Mbenzele	30	17	87	0.567	0.412	0.954 (0.019)	20.170 (9.170)	0.001 (0.001)	−0.292 (0.433)	1.453 (0.729)
MbutiCEPH	14	11	170	0.786	0.909	0.934 (0.061)	63.253 (29.070)	0.004 (0.002)	0.685 (0.801)	3.516 (0.924)
Mbuti	21	18	187	0.857	0.833	0.981 (0.022)	56.967 (25.642)	0.003 (0.002)	0.318 (0.672)	0.625 (0.615)
Fang	17	17	249	1.000	0.941	1.000 (0.020)	62.250 (28.279)	0.004 (0.002)	−0.654 (0.289)	−1.870 (0.102)
Nzebi	19	15	287	0.789	0.733	0.977 (0.023)	69.743 (31.463)	0.004 (0.002)	−0.646 (0.267)	3.024 (0.905)

NOTE.—For population names, refer to the legend of figure 1. π, nucleotide diversity; D, Tajima's D neutrality test; Fs, Fu's Fs neutrality test; HD, sequence diversity; k, number of sequences; kp, number of private sequences; MNPD, mean number of pairwise differences; N, sample size; S, number of polymorphic sites; SD, standard deviation.

distance between EPYG and WPYG is significantly larger than that between each Pygmy group and the WAGR ($P < 10^{-5}$ in both cases). By contrast, the distance between EPYG and WAGR is not significantly larger than that between WPYG and WAGR ($P = 0.376$). Within the WPYG group, all populations cluster together, with the exception of the Babinga that exhibit large distances from the other populations, the lowest with the WAGR (0.223) and the highest with EPYG (0.464), in accordance with the rest of the WPYG group.

Phylogenetic Reconstruction of Complete Mitochondrial Sequences

In order to establish the relationships between our complete sequences (supplementary table S3, Supplementary Material online) and to provide a broad view of the L haplogroup phylogeny, a median-joining network was built (fig. 3). The resulting data set includes 973 individuals, representing 746 different sequences. Haplogroups L0 and L1 are mostly present in specific population groups and geographic areas (see supplementary fig. S3, Supplementary Material online). Sequences found in Khoisan-speaking groups cluster within clades L0d and L0k as previously reported (Salas et al. 2002; Behar et al. 2008), and most Western Pygmy sequences fall into the L1c haplogroup (Batini et al. 2007; Quintana-Murci et al. 2008).
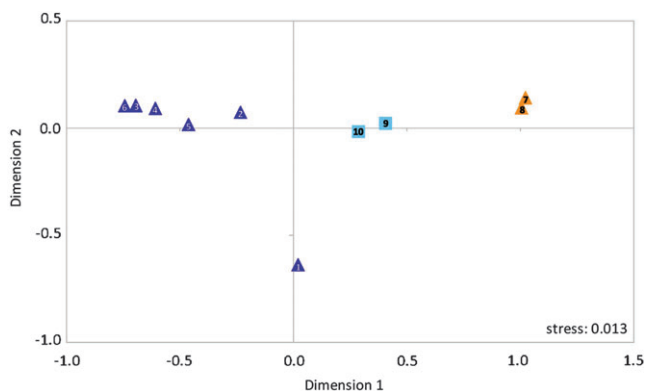
On the other hand, no clear geographical structure is found within haplogroups L2–L6 with a few exceptions. Eastern Pygmy sequences belong to haplogroups L0a, L2a, and L5, where they form specific subclades (see supplementary fig. S3, Supplementary Material online, for the geographical structure of the network). L0a also contains one sequence from WPYG (present in three Biaka from CAR), although this sequence is not shared with EPYG and is located in a different subclade. Within haplogroup L2a, EPYG are found in two different subclades (L2a2 and L2a4), which contain only five non-Pygmy sequences: two Sara (Chad), one Laal (Chad), one Nuba (Sudan), and one San (South Africa). L2a4 is a previously undescribed subclade, found exclusively in Eastern Pygmies and characterized by mutations 513A, 593a, 5147A, 6959T, 7897A, 8614C, 9438A, 12480T, 13125T, and 15812A (see supplementary fig. S4, Supplementary Material online). The other main branch of this lineage (L2a1) does not show population or geographical structure (see supplementary fig. S3, Supplementary Material online), including numerous African American (87; 57.6% of the total) and Western–Central African (19; 12.6%) sequences but only a few individuals from Eastern Africa (7; 4.6%). L5 presents long branches and an ancient TMRCA (table 3), although these inferences should be taken with caution given the low number of sequences in our data set (14; 1.4% of the total). Interestingly, five Eastern Pygmy sequences are found in this clade with other Eastern African individuals.

Besides the above-mentioned L0a and L1c haplogroups, two Western Pygmy sequences were observed within the L3e clade, represented, respectively, by four Babinga within L3e2b and seven Baka from Cameroon within L3e1.

Table 3 reports TMRCA estimates for the lineages described above. The mean value of the TMRCAs of Western Pygmy–specific clades ranges from 20 to 34 Kya, in agreement with previous studies (Batini et al. 2007; Quintana-Murci et al. 2008), and all of them belong to the ancient L1c clade (mean TMRCA: 85 Kya). Eastern Pygmy–specific clades show mean TMRCAs ranging from 6.5 to 25 Kya and belong to the L0a and L2a clades, respectively, dated at 48 and 46.5 Kya, as well as to L5. The latter is estimated to be very ancient (mean value: 106 Kya), although more data from this haplogroup are necessary for robust dating.
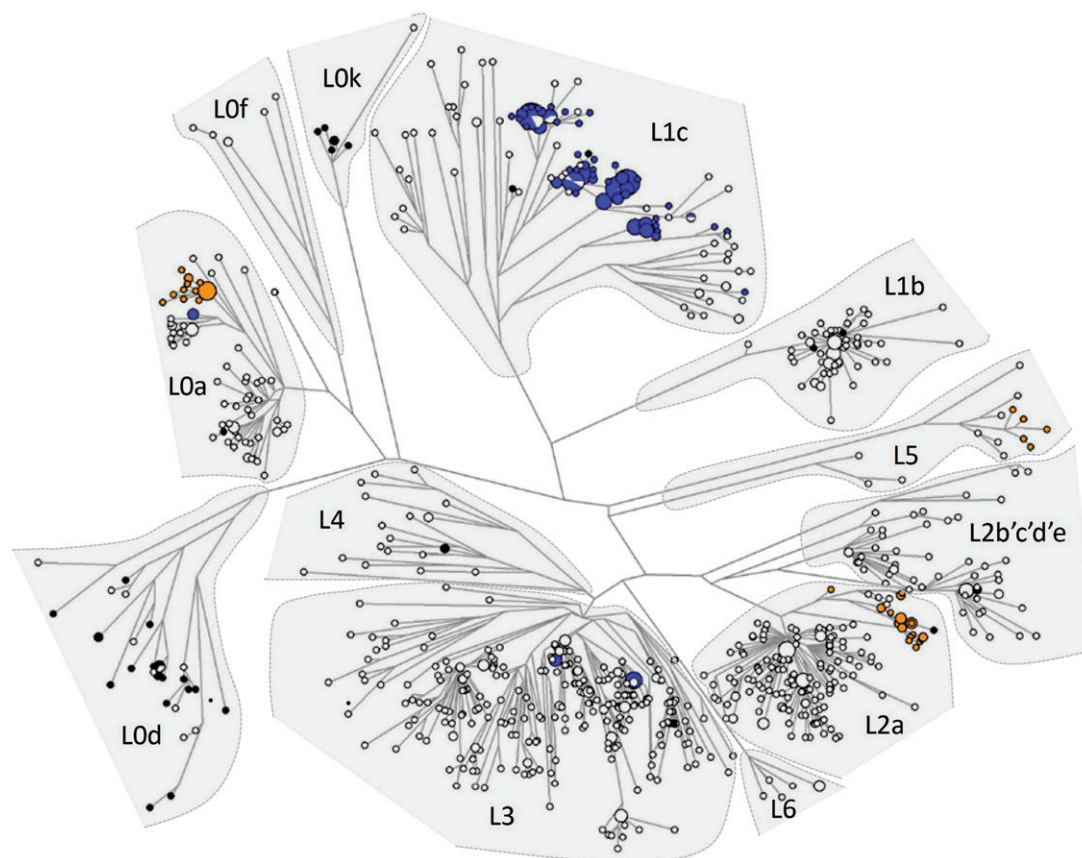

FIG. 2. MDS plot. Blue triangles, Western Pygmies; orange triangles, Eastern Pygmies; light blue squares, Western Bantu-speaking farmers. For number legend, refer to figure 1.

Fig. 3. Median-joining network of 973 mitochondrial-coding region sequences (corresponding to positions 435–16023). Blue, Western Pygmies; orange, Eastern Pygmies; black, Khoisan speakers.

Demographic Inferences from Complete mtDNA Lineages

Fifteen different scenarios (see supplementary fig. S1, Supplementary Material online) have been tested by means of ABC coalescent simulations in order to shed light on the origin of African Pygmies and their relationships with Western farmers from a maternal point of view. The results obtained from the model-choice test are presented in figure 4. Model 1, assuming a common and recent origin of Pygmies with gene flow between Western populations (Pygmies and non-Pygmies), was the most supported, with a probability of 48%. The total of the posterior probabilities (100%) is accounted by the first three scenarios, which present the same migration scheme, with gene flow allowed only between WPYG and WAGR.

Demographic parameters were estimated using the model best supported by the data. Their modes and credible intervals are shown in table 1, and their posterior distributions are presented in supplementary figure S5 (Supplementary Material online). EPYG show an effective population size, which is roughly three times that of WPYG, and both of them are much lower than that of WAGR. The last is roughly seven times higher than the ancestral Ne on the same branch of the topology (NeA2), suggesting population growth in WAGR, whereas Pygmy Ne values are compatible with ancestral Ne, which, however, shows a very wide posterior distribution. Population size parameters

obtained through the ABC simulations were in agreement with the BSP (supplementary fig. S6, Supplementary Material online). Pygmy groups present a lower Ne than the farmer group as well as a minimum value of Ne in the

Table 3. Coalescent-Based TMRCA Estimates for Major L Clades and Selected Subclades.

Clade	TMRCA (95% HPD)
L0	123,500 (103,600–142,400)
L0a	47,700 (34,900–61,100)
L0a2b	9,700 (4,110–16,140)
L0k	30,110 (17,360–44,270)
L0d	79,020 (61,750–97,260)
L1	110,500 (89,220–132,400)
L1c	85,300 (71,800–99,430)
L1c1a1	29,350 (18,500–41,300)
L1c1a2	20,310 (12,050–29,620)
L1c4	34,390 (20,690–48,790)
L2	99,300 (80,130–119,400)
L2a	46,550 (34,930–59,430)
L2a1	36,630 (27,580–47,085)
L2a2	24,840 (15,025–35,380)
L2a4	6,520 (1,910–11,320)
L3	100,700 (78,400–123,800)
L3e	44,160 (32,465–56,535)
L4	85,700 (67,195–105,500)
L5	106,000 (81,220–131,700)
L5a1c	11,430 (4,650–19,030)
L6	19,510 (10,601–29,290)

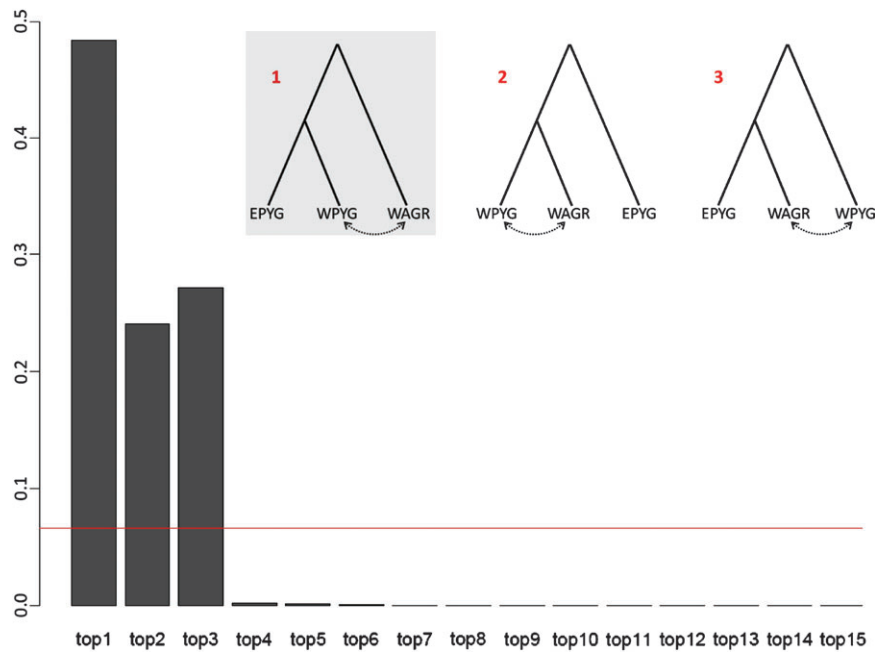Note.—95% HPD, 95% highest posterior density interval.

Fig. 4. Prior (red line) and posterior probabilities (gray bars) of the 15 evolutionary models tested. Only the three more supported scenarios are represented. All scenarios tested are shown in supplementary figure S1 (Supplementary Material) online. EPYG, Eastern Pygmies; WPYG, Western Pygmies; WAGR, Western Bantu-speaking farmers.

recent past (at 0.65 Kya for WPYG and at 4.3 Kya for EPYG), whereas the minimum Ne of WAGR matches the most ancient Ne. The reduction of Ne is similar among the two Pygmy groups, being up to 40% of the initial population size. However, the duration of this process differs between the two groups. WPYG seem to have experienced a reduction of population size beginning around 2.5 Kya and lasting until 0.65 Kya, with a very rapid subsequent recovery. On the other hand, EPYG show signals of a more ancient and gradual decrease from around 20 Kya, reaching a minimum around 4.3 Kya, and expanding again to a larger size (up to 400%). For WAGR, an increase of 700% is observed between the initial and present-day Ne. However, this population growth seems to be a gradual rather than a sudden process, with a slight acceleration between 65 and 30 Kya, when population size reached a level similar to that of the present.

In our method, migration rate is expressed as the proportion of immigrants in a population per generation and is assumed, together with Ne, to be constant through time. In order to obtain this rate in the units of absolute number of immigrants per generation (mig/gen), the rate must be multiplied by the effective population size. The estimated value for immigration from WAGR to WPYG is 1.9 mig/gen, whereas that for the opposite direction is 10.5 mig/gen, representing a 5-fold difference. As for the migration rate into the ancestral Pygmy population, prior to the separation of the two groups, the observed value is 13.4 mig/gen. However, the confidence intervals of both the m3 and mA1 are quite broad, thus leaving this issue to be further clarified, possibly through the analysis of a wider population sample.

The split between the Pygmy and the non-Pygmy populations seems to have occurred 71 (52–106) Kya. According to our estimates, a more recent event, around 27 (10–57) Kya, led to the separation between Western and Eastern Pygmy populations. Finally, our results suggest that after their separation, the two Pygmy groups remained isolated from each other, with little or no evidence for gene flow.

Discussion

Hunter-gatherers' Lineages in the Context of Sub-Saharan African Mitochondrial Variation

Phylogeographic approaches have been extensively used to study the distribution of genetic variation in human populations, especially for uniparentally transmitted markers. The present comprehensive analysis of the phylogeographical pattern of complete mitochondrial L lineages shows a general lack of geographical structure within sub-Saharan African lineages, with a few exceptions in haplogroups L0, L1, L2, and L5. Interestingly, these clades are those observed in hunter-gatherer populations. Indeed, the L0d and L0k haplogroups are mainly found among the Khoisan speakers from Southern Africa; most Western Pygmy mitochondrial genomes belong to L1c haplogroup; and finally, specific subclades of L0a, L2a, and L5 haplogroups have so far been detected mostly in Eastern Pygmies (see also Salas et al. 2002; Batini et al. 2007; Gonder et al. 2007; Behar et al. 2008; Quintana-Murci et al. 2008). This suggests that the genetic structure of maternal lineages across sub-Saharan Africa has been mainly shaped by differences in lifestyle and demographic history rather than by geography. The
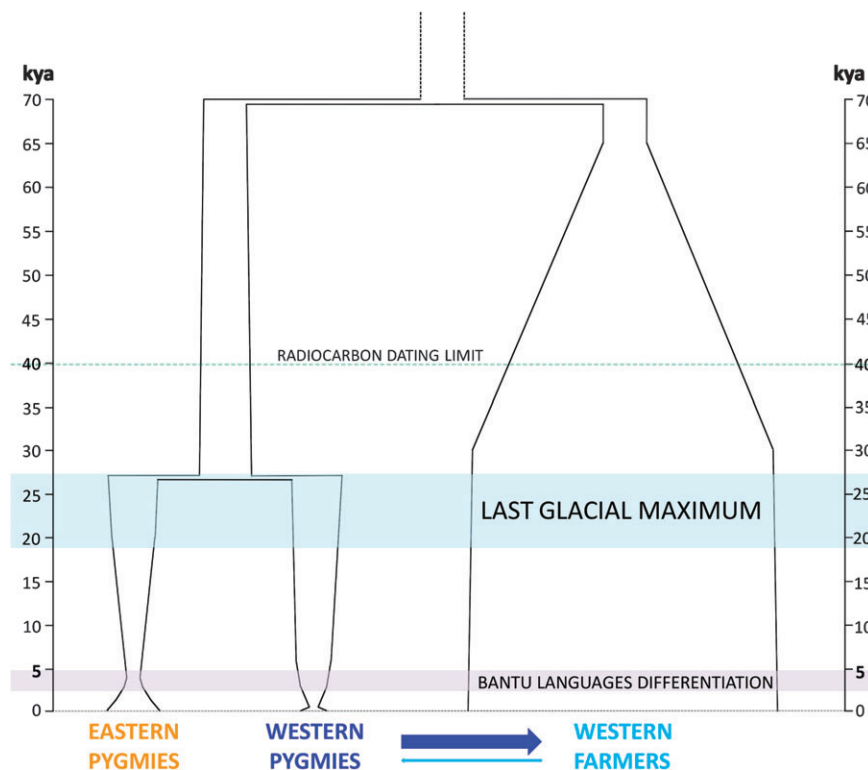
Fig. 5. Evolutionary scenario proposed for the origin of Pygmies and their relations with the present-day Bantu-speaking farmers. Blue, Western Pygmies; orange, Eastern Pygmies; light blue, Western Bantu-speaking farmers.

distinctive mitochondrial heritage of hunter-gatherers is also supported by the prevalence (frequencies from 0.48 to 0.60) of haplogroup L4g (absent from this data set) among Hadza and Sandawe from Tanzania, which contrasts with its rarity across the African continent (Kivisild et al. 2004; Gonder et al. 2007; Tishkoff et al. 2007).

This deep structure, together with ancient ancestries of most of these lineages (see table 3), suggests an ancient separation of the ancestors of present-day hunter-gatherer communities, followed by isolation. This pattern contrasts with results of previous studies based on Y chromosome and genome-wide variation in African samples (Wood et al. 2005; Hellenthal et al. 2008; Tishkoff et al. 2009), which point instead to a shared common ancestry between Central African Pygmies and Khoisan-speaking populations as a result of a common origin (>30 Kya) of hunter-gatherer populations (Tishkoff et al. 2009; Scheinfeldt et al. 2010). The discrepancy between these two lines of evidence might be explained by different factors. It has been previously proposed that gene flow of mtDNA lineages from non-Pygmies to Pygmies should be virtually absent or at least considerably less than that for Y chromosome and autosomes, as a result of social constraints (Cavalli-Sforza 1986a; Destro-Bisol, Donati, et al. 2004). Consequently, a higher degree of continuity between the ancestral and extant gene pool and a less marked homogenizing effect of gene flow from neighboring farmers are expected for mtDNA than for other genetic systems. However, the shared common ancestry found for other loci, both among Pygmy groups and

between them and Khoisan-speaking populations, could be explained on the basis of recent demographic male-mediated direct or indirect contact among these groups (this is corroborated by the very recent dating of the lineages involved; Batini C, Comas D, Capelli C, personal communication).

Common Origin of Pygmies and Their Relationship with Bantu Speakers

The issue of the origin of African Pygmies has been intensively discussed and three main evolutionary scenarios have been proposed. It has been suggested that the evolution of phenotypic features in Pygmies (e.g., stature and pigmentation) occurred in a single population that diverged only recently to form present-day groups (Cavalli-Sforza 1986a). Other authors have argued that the differences observed among Pygmy populations could be explained by an ancient and separate origin with convergent evolution for short stature as an adaptation to the hot and wet environment of the equatorial forest (Hiernaux 1974, 1977). Finally, Pygmies have been proposed to have originated as independent subgroups of expanding farmers in the last few millennia (Blench 1999).

Our data support a unique origin of African Pygmies in an ancient phase of the peopling of the Central African belt (see fig. 5). In addition, the genetic differences found between the two Pygmy groups (EPYG and WPYG) might be explained by an ancient divergence, compatible with the Last Glacial Maximum (LGM; 19–26.5 Kya; Clark

et al. 2009), and largely predating the later spread of various forms of agriculture in the area, which started ~5 Kya (Phillipson 1993). The distribution of mitochondrial genome variation and haplogroup composition shows a clear-cut difference between EPYG and WPYG, and our simulations suggest that their large genetic distances could be explained by an ancient separation around 27 (10–57) Kya, followed by complete isolation. The divergence between Pygmy groups occurred after the separation of the proto-Pygmy group from the ancestors of present-day Bantu-speaking farmers, which took place around 71 (52–106) Kya (see fig. 5). The two groups could have exploited different ecological niches, wooded environments, or savannah and open spaces, respectively (see Thomas 2000; Cornelissen 2002; Mercader 2003). Indeed, evidence for different types of Middle Stone Age industries in Central Africa, mainly Lupemban and quartz microlithics, preceding the limit of radiocarbon dating (~40 Kya) has been reported (see Cornelissen 2002; Marean and Assefa 2005).

Our estimates of population divergence times are in agreement with recent studies based on 20 autosomal 1-kb regions (Patin et al. 2009) and on autosomal STRs (Verdu et al. 2009), as well as previous studies based on mtDNA phylogenetic-based dating of specific clades in Western Pygmy populations (Destro-Bisol, Coia, et al. 2004; Batini et al. 2007; Quintana-Murci et al. 2008). Interestingly, the TMRCA estimates for the coding mitochondrial regions in Western Pygmy–specific clades (recalibrated using the latest revised mtDNA mutation rate presented in Soares et al. 2009) are compatible with the divergence date between the two groups estimated by coalescent simulations, whereas Eastern Pygmy clades show younger coalescence dates, with the sole exception of L2a2 (see table 3). This suggests that some of the extant phylogenetic variation could have been shaped in recent demographic events, whereas most of it seems to be the result of a separation during the LGM, which could have been the ecological cause of the split (see fig. 5). In fact, it has been observed that, during this climatic phase, the tropical forest in Africa suffered a dramatic reduction in size (Sayer et al. 1992). Despite controversy about the intensity of this process and the continuity and location of what have been defined as refuge areas (see Brook et al. 1990; Mercader et al. 2000; Thomas 2000), it is intriguing that the current distribution of the two Pygmy groups considered here mirrors the location of those areas in which concentration peaks of both animal and plant endemisms are observed (Hamilton 1982; Sayer et al. 1992). The same ecological isolation mechanisms could have acted on Pygmy populations, whose separation would therefore have been independent from the putative expansion of Bantu-speaking populations through sub-Saharan Africa since 5 Kya (Bahuchet 1993; Phillipson 1993; Nurse and Philippson 2003). Finally, we detected signatures of gene flow only among the two Western (Pygmy and non-Pygmy) groups with a ratio of 5:1 female migrants from Western Pygmies to Bantu-speaking farmers. However, the posterior distribution of the migration rate into

the the WAGR group is wide, thus leaving this issue to be further clarified.

Genetic Signatures of Differential Recent Events in Pygmy Demographic Histories

The effective population size estimated for Pygmies is lower than that for Western farmers but not very different from that of the ancestral population, suggesting that Pygmies may have maintained population sizes similar to early *Homo sapiens* communities. This would be in agreement with the demographic expansion thought to be associated with the rapid diffusion of Bantu languages throughout sub-Saharan Africa in the last 3–5 Kya (Ehret and Posnansky 1982; Phillipson 1993; Vansina 1995; Nurse and Philippson 2003). However, our demographic estimates suggest a more gradual increase in population size during the last millennia. The acceleration in the population growth rate of Western farmers is observed long before the expected demographic expansion at 3–5 Kya, between 65 and 30 Kya, leading to a present-day population size seven times higher than that of the ancestral one. By contrast, both Pygmy groups show signals of decrease in population size during the last 70 Kya, with the ratio of Western and Eastern Pygmies compared with the ancestral population of 0.25:1 and 0.80:1, respectively. Both Pygmy groups show evidence of recent bottlenecks with similar intensity (up to 40% reduction of the original Ne), although their timing is different. Eastern Pygmies show a gradual decrease of Ne starting 20 Kya and reaching its minimum 4 Kya, Western Pygmies show a sudden population reduction between 4 and 0.65 Kya, which overlaps with the putative dates of the diffusion of Bantu languages (see fig. 5).

Comparing our scenario with that obtained from autosomal markers (Patin et al. 2009), two important differences can be noticed. First, the strength of the reduction of Ne is greater with autosomal data (up to 80% for WPYG and to 95% for EPYG). Second, Eastern Pygmies showed signals of a more recent bottleneck than Western Pygmies. Differential social and demographical dynamics acting on the female and male component of Pygmy populations during their evolutionary history may help to explain this discrepancy, although only a larger sampling, as well as data from Eastern African Central Sudanic and Bantu speakers, could make any comparison and subsequent interpretation more robust. Also, we observe a greater heterogeneity in Eastern Pygmy samples compared with Western Pygmies, which is compatible with a more ancient bottleneck and almost complete restoration of population size. However, gene flow is expected to increase effective population size and internal diversity, and its detection could have been limited by the lack of key neighboring population samples. With a more comprehensive sampling of Pygmy groups and especially of Eastern Pygmy neighboring populations, issues like these may be further clarified.

In conclusion, we have presented the first population study of complete mtDNA variation among African Pygmies and have drawn a detailed demographic scenario for their evolutionary history. This investigation marks

a substantial difference from previous studies, where the genomic approach was applied to dissect specific lineages for phylogeographic purposes. We are aware that mtDNA offers a partial—maternal—view of population history, whereas lineage loss may confound results in small-sized populations. Nonetheless, our results support and complement previous findings, contributing to a more complete picture of the evolutionary history of African Pygmies and highlighting the importance of complete mtDNA sequencing at the population level for deciphering the prehistory of human populations.

Supplementary Material

Supplementary figures S1 to S6 and supplementary tables S1 to S5 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

References

Achilli A, Perego UA, Bravi CM, Coble MD, Kong QP, Woodward SR, Salas A, Torroni A, Bandelt HJ. 2008. The phylogeny of the four pan-American MtDNA haplogroups: implications for evolutionary and disease studies. *PLoS One*. 3:e1764.

Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet*. 23:147.

Bahuchet S. 1993. History of the inhabitants of the central African rain forest: perspectives from comparative linguistics. In: Hladik CM, Hladik A, Linares OF, Pagezy H, Semple A, Hadley M, editors. Tropical forests, people and food. New York: UNESCO p. 37–54.

Bandelt HJ, Forster P, Rohl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol*. 16:37–48.

Batini C, Coia V, Battaggia C, Rocha J, Pilkington MM, Spedini G, Comas D, Destro-Bisol G, Calafell F. 2007. Phylogeography of the human mitochondrial L1c haplogroup: genetic signatures of the prehistory of central Africa. *Mol Phylogenet Evol*. 43:635–644.

Beaumont MA, Zhang W, Balding DJ. 2002. Approximate Bayesian computation in population genetics. *Genetics* 162:2025–2035.

Behar DM, Villems R, Soodyall H, et al. 15 co-authors. 2008. The dawn of human matrilineal diversity. *Am J Hum Genet*. 82:1130–1140.

Berniell-Lee G, Calafell F, Bosch E, Heyer E, Sica L, Mouguiama-Daouda P, van der Veen L, Hombert JM, Quintana-Murci L, Comas D. 2009. Genetic and demographic implications of the Bantu expansion: insights from human paternal lineages. *Mol Biol Evol*. 26:1581–1589.

Blench R. 1999. Are the African pygmies an ethnographic fiction?. In: Biesbrouck K, Elders S, Rossel G, editors. Central African hunter-gatherers in a multidisciplinary perspective: challenging elusiveness. Leiden (The Netherlands): CNWS, Universiteit Leiden. p. 41–60.

Brook GA, Burney DA, Cowart JB. 1990. Paleoenvironmental data for Ituri, Zaire, from sediments in Matupi Cave, Mt Hoyo. In: Boaz NT, editor. Evolution of environments and Hominidae in the African Western Rift Valley. Martinsville: Virginia Museum of Natural History Memoir 1. p. 49–70.

Cavalli-Sforza LL. 1986a. African pygmies: an evaluation of the state of research. In: Cavalli-Sforza LL, editor. African pygmies. Orlando (FL): Orlando Academic Press. p. 361–426.

Cavalli-Sforza LL. 1986b. Anthropometric data. In: Cavalli-Sforza LL, editor. African pygmies. Orlando (FL): Orlando Academic Press. p. 81–93.

Cavalli-Sforza LL. 1986c. Demographic data. In: Cavalli-Sforza LL, editor. African pygmies. Orlando (FL): Orlando Academic Press. p. 23–44.

Cavalli-Sforza LL, Menozzi P, Piazza A. 1993. The history and geography of human genes. Princeton (NJ): Princeton University Press.

Clark PU, Dyke AS, Shakun JD, Carlson AE, Clark J, Wohlfarth B, Mitrovica JX, Hostetler SW, McCabe AM. 2009. The last glacial maximum. *Science* 325:710–714.

Cornelissen E. 2002. Human responses to changing environments in central Africa between 40,000 and 12,000 BP. *J World Prehist*. 16:197–235.

Demolin D. 1996. The languages spoken by the pygmies in the Ituri: comparative and historical perspectives. Colloquium on hunter-gatherers of Equatorial Africa; October 1996; Leiden, Netherlands.

Destro-Bisol G, Battaggia C, Coia V, Batini C, Spedini G. 2006. The western pygmies from central African republic: new data on autosomal loci. *J Anthropol Sci*. 84:161–164.

Destro-Bisol G, Coia V, Boschi I, Verginelli F, Caglia A, Pascali V, Spedini G, Calafell F. 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.

Destro-Bisol G, Donati F, Coia V, Boschi I, Verginelli F, Caglia A, Tofanelli S, Spedini G, Capelli C. 2004. Variation of female and male lineages in sub-Saharan populations: the importance of sociocultural factors. *Mol Biol Evol*. 21:1673–1682.

Diamond JM. 1991. Why are pygmies small? *Nature*. 354:111–112.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol*. 7:214.

Drummond AJ, Rambaut A, Shapiro B, Pybus OG. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol*. 22:1185–1192.

Ehret C, Posnansky M. 1982. The archaeological and linguistic reconstruction of African history. Berkley: University of California Press.

Excoffier L, Laval G, Schneider S. 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform Online*. 1:47–50.

Fagundes NJ, Ray N, Beaumont M, Neuenschwander S, Salzano FM, Bonatto SL, Excoffier L. 2007. Statistical evaluation of alternative

models of human evolution. *Proc Natl Acad Sci U S A.* 104: 17614–17619.

Finnila S, Lehtonen MS, Majamaa K. 2001. Phylogenetic network for European mtDNA. *Am J Hum Genet.* 68:1475–1484.

Gonder MK, Mortensen HM, Reed FA, de Sousa A, Tishkoff SA. 2007. Whole-mtDNA genome sequence analysis of ancient African lineages. *Mol Biol Evol.* 24:757–768.

Hamilton AC. 1982. Environmental history of east Africa: a study of the quaternary. London: Academic Press.

Hart TB, Hart JA. 1986. The ecological basis of hunter-gatherer subsistence in African rain forests: the Mbuti of eastern Zaire. *Hum Ecol.* 14:29–55.

Hellenthal G, Auton A, Falush D. 2008. Inferring human colonization history using a copying model. *PLoS Genet.* 4:e1000078.

Herrnstadt C, Elson JL, Fahy E, et al. (11 co-authors). 2002. Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups. *Am J Hum Genet.* 70:1152–1171.

Hiernaux J. 1974. The people of Africa. New York: Charles Scribner's Sons.

Hiernaux J. 1977. Long-term biological effects of human migration from the African savanna to the equatorial forest: a case study of human adaptation to a hot and wet climate. In: Harrison GA, editor. Population structure and human variation. Cambridge: Cambridge University Press. p. 187–217.

Hitchcock RK. 1999. Introduction: Africa. In: Lee RB, Daly RH, editors. The Cambridge encyclopedia of hunters and gatherers. Cambridge: Cambridge University Press. p. 175–184.

Ingman M, Kaessmann H, Paabo S, Gyllensten U. 2000. Mitochondrial genome variation and the origin of modern humans. *Nature* 408:708–713.

Jakobsson M, Scholz SW, Scheet P, Gibbs JR, VanLiere JM, Fung HC, Szpiech ZA, Degnan JH, Wang K, Guerreiro R. 2008. Genotype, haplotype and copy-number variation in worldwide human populations. *Nature* 451:998–1003.

Just RS, Diegoli TM, Saunier JL, Irwin JA, Parsons TJ. 2008. Complete mitochondrial genome sequences for 265 African American and U.S. "Hispanic" individuals. *Forensic Sci Int Genet.* 2:e45–e48.

Kivisild T, Reidla M, Metspalu E, Rosa A, Brehm A, Pennarun E, Parik J, Geberhiwot T, Usanga E, Villems R. 2004. Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears. *Am J Hum Genet.* 75:752–770.

Kong QP, Salas A, Sun C, Fuku N, Tanaka M, Zhong L, Wang CY, Yao YG, Bandelt HJ. 2008. Distilling artificial recombinants from large sets of complete mtDNA genomes. *PLoS One.* 3:e3016.

Letouzey R. 1976. Contribution de la botanique au problème d'une éventuelle langue pygmée. Paris: SELAF.

Li JZ, Absher DM, Tang H, et al. 11 (co-authors). 2008. Worldwide human relationships inferred from genome-wide patterns of variation. *Science* 319:1100–1104.

Lopes JS, Balding D, Beaumont MA. 2009. PopABC: a program to infer historical demographic parameters. *Bioinformatics* 25:2747–2749.

Maca-Meyer N, Gonzalez AM, Larruga JM, Flores C, Cabrera VM. 2001. Major genomic mitochondrial lineages delineate early human expansions. *BMC Genet.* 2:13.

Marean CW, Assefa Z. 2005. The middle and upper Pleistocene African record for the biological and behavioral origins of modern humans. In: Sthal AB, editor. African archaeology. Oxford: Blackwell Publishing. p. 93–129.

Mercader J. 2003. Introduction: the Paleolithic settlement of rain forest. In: Mercader J, editor. Under the canopy: the archaeology of tropical rain forests. London: Rutgers University Press. p. 1–31.

Mercader J, Runge F, Vrydaghs L, Doutrelepont H, Ewango CEN, Juan-Tresseras J. 2000. Phytoliths from archaeological sites in the tropical forest of Ituri, democratic republic of Congo. *Quat Res.* 54:102–112.

Migliano AB, Vinicius L, Lahr MM. 2007. Life history trade-offs explain the evolution of human pygmies. *Proc Natl Acad Sci U S A.* 104:20216–20219.

Nurse D, Philippson G. 2003. The Bantu languages. London: Routledge.

Patin E, Laval G, Barreiro LB, et al. (15 co-authors). 2009. Inferring the demographic history of African farmers and pygmy hunter-gatherers using a multilocus resequencing data set. *PLoS Genet.* 5:e1000448.

Perry GH, Dominy NJ. 2009. Evolution of the human pygmy phenotype. *Trends Ecol Evol.* 24:218–225.

Phillipson DW. 1993. African archaeology. Cambridge: Cambridge University Press.

Quintana-Murci L, Quach H, Harmant C, et al. 23 co-authors. 2008. Maternal traces of deep common ancestry and asymmetric gene flow between pygmy hunter-gatherers and Bantu-speaking farmers. *Proc Natl Acad Sci U S A.* 105:1596–1601.

Roostalu U, Kutuev I, Loogvali EL, Metspalu E, Tambets K, Reidla M, Khusnutdinova EK, Usanga E, Kivisild T, Villems R. 2007. Origin and expansion of haplogroup H, the dominant human mitochondrial DNA lineage in west Eurasia: the near eastern and Caucasian perspective. *Mol Biol Evol.* 24:436–448.

Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW. 2002. Genetic structure of human populations. *Science* 298:2381–2385.

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

Sayer JA, Harcourt CS, Collins NM. 1992. The conservation atlas of tropical forests: Africa. London: UCN, WCMC, Macmillan.

Scheinfeldt LB, Soi S, Tishkoff SA. 2010. Colloquium paper: working toward a synthesis of archaeological, linguistic, and genetic data for inferring African population history. *Proc Natl Acad Sci U S A.* 107(Suppl 2):8931–8938.

Soares P, Ermini L, Thomson N, Mormina M, Rito T, Rohl A, Salas A, Oppenheimer S, Macaulay V, Richards MB. 2009. Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am J Hum Genet.* 84:740–759.

Sturrock K, Rocha J. 2000. A multidimensional scaling stress evaluation table. *Field Methods.* 12:49–60.

Thomas MF. 2000. Late quaternary environmental changes and the alluvial record in humid tropical environments. *Quat Int.* 72:23–36.

Tishkoff SA, Gonder MK, Henn BM, et al. 12 co-authors. 2007. History of click-speaking populations of Africa inferred from mtDNA and Y chromosome genetic variation. *Mol Biol Evol.* 24:2180–2195.

Tishkoff SA, Reed FA, Friedlaender FR, et al. 25 co-authors. 2009. The genetic structure and history of Africans and African Americans. *Science.* 324:1035–1044.

Vansina J. 1995. New linguistic evidence and 'the Bantu expansion'. *J Afr Hist.* 36:173–195.

Verdu P, Austerlitz F, Estoup A, et al. 14 co-authors. 2009. Origins and genetic diversity of pygmy hunter-gatherers from western central Africa. *Curr Biol.* 19:312–318.

Watkins WS, Rogers AR, Ostler CT, Wooding S, Bamshad MJ, Brassington AME, Carroll ML, Nguyen SV, Walker JA, Prasad B. 2003. Genetic variation among world populations: inferences from 100 Alu insertion polymorphisms. *Genome Res.* 13:1607–1618.

Wood ET, Stover DA, Ehret C, et al. (11 co-authors). 2005. Contrasting patterns of Y chromosome and mtDNA variation in Africa: evidence for sex-biased demographic processes. *Eur J Hum Genet.* 13:867–876.