Contrasting Signatures of Population Growth for Mitochondrial DNA and Y Chromosomes among Human Populations in Africa

Maya Metni Pilkington, Jason A. Wilder,†‡ Fernando L. Mendez,§ Murray P. Cox,† August Woerner,† Thiep Angui,† Sarah Kingan,† Zahra Mobasher,† Chiara Batini,‖*** Giovanni Destro-Bisol,‖¶ Himla Soodyall,# Beverly I. Strassmann,** and Michael F. Hammer*†§*

*Department of Anthropology, University of Arizona; †Arizona Research Labs Division of Biotechnology, University of Arizona; ‡Department of Biology, Williams College; §Department of Ecology and Evolutionary Biology, University of Arizona; ‖Department of Animal and Human Biology, University La Sapienza, Rome, Italy; ¶Instituto Italiano di Antropologia, Rome, Italy; #Human Genomic Diversity and Disease Research Unit, National Health Laboratory Service and University of the Witwatersrand, Johannesburg, South Africa; **Department of Anthropology, University of Michigan; and ***Unitat de Biologia Evolutiva, Facultat de Ciències de la Salut i de la Vida, Universitat Pompeu Fabra, Barcelona, Spain

A history of Pleistocene population expansion has been inferred from the frequency spectrum of polymorphism in the mitochondrial DNA (mtDNA) of many human populations. Similar patterns are not typically observed for autosomal and X-linked loci. One explanation for this discrepancy is a recent population bottleneck, with different rates of recovery for haploid and autosomal loci as a result of their different effective population sizes. This hypothesis predicts that mitochondrial and Y chromosomal DNA will show a similar skew in the frequency spectrum in populations that have experienced a recent increase in effective population size. We test this hypothesis by resequencing 6.6 kb of noncoding Y chromosomal DNA and 780 basepairs of the mtDNA *cytochrome c oxidase subunit III* (*COIII*) gene in 172 males from 5 African populations. Four tests of population expansion are employed for each locus in each population: Fu's *Fs* statistic, the R_2 statistic, coalescent simulations, and the mismatch distribution. Consistent with previous results, patterns of mtDNA polymorphism better fit a model of constant population size for food-gathering populations and a model of population expansion for food-producing populations. In contrast, none of the tests reveal evidence of Y chromosome growth for either food-gatherers or food-producers. The distinct mtDNA and Y chromosome polymorphism patterns most likely reflect sex-biased demographic processes in the recent history of African populations. We hypothesize that males experienced smaller effective population sizes and/or lower rates of migration during the Bantu expansion, which occurred over the last 5,000 years.

Introduction

Several studies in the early 1990s concluded that the excess of low-frequency polymorphisms in human (mitochondrial DNA) mtDNA over the "expected" distribution under the standard neutral model was the signal of rapid population expansion (Di Rienzo and Wilson 1991; Slatkin and Hudson 1991; Lundstrom et al. 1992; Aris-Brosou and Excoffier 1996). Many populations also exhibit unimodal peaks in the distribution of the number of pairwise differences (mismatch distributions) in mtDNA sequences, suggesting that populations expanded in size beginning ~30–130 kya (Di Rienzo and Wilson 1991; Rogers and Harpending 1992; Sherry et al. 1994; Rogers and Jorde 1995). This pattern of rapid expansion from small ancestral effective population size (N_e) has not been generally observed for autosomal and X chromosomal DNA sequencing data sets (Hey 1997; Harpending and Rogers 2000; Wall and Przeworski 2000; Excoffier 2002). X-linked and autosomal sequences typically show no excess of low-frequency polymorphisms, instead there is a tendency for non-African populations to have positive Tajima's *D* values and African populations to have only slightly negative values (Przeworski et al. 2000; Garrigan and Hammer 2006). There is disagreement on the causes of the discrepancy between mtDNA and nuclear data sets: some investigators favor models involving differential natural selection (Hey 1997; Harpending and Rogers 2000; Excoffier 2002), whereas others favor a demographic model featuring a recent bottleneck (Fay and Wu 1999; Hammer et al. 2004). For example, Fay and Wu (1999) pointed out that for a period of time following a population bottleneck, there is expected to be an excess of low-frequency polymorphisms in mtDNA and an excess of intermediate frequency autosomal polymorphisms simply because of the different effective population sizes of these loci.

Under a simple bottleneck model with an equal breeding sex ratio, both the nonrecombining portion of the Y chromosome (NRY) and mtDNA are expected to experience a similar reduction in N_e. As the population recovers from this bottleneck, the 2 haploid loci should respond similarly with respect to changes in the frequency distribution of polymorphisms over time (Fay and Wu 1999). To date, there are no published studies designed specifically to test the bottleneck hypothesis. Although some studies of single nucleotide polymorphisms (SNPs) and short tandem repeats (STRs) on the NRY have supported models of demographic expansion (Pritchard et al. 1999; Thomson et al. 2000), the grid-sampling schemes employed have reduced power to distinguish between the effects of population expansion and population structure (Ptak and Przeworski 2002). For instance, Hammer et al. (2003) showed that sampling a few individuals from many global populations may lead to an upward bias in the number of singleton variants. They concluded that it is necessary to thoroughly sample within demes to obtain a robust estimate of the frequency spectrum. Moreover, inferences of population growth based on Y-STR data are difficult to interpret as a result of uncertainty associated with microsatellite mutation models (Pritchard et al. 1999; Wall and Przeworski 2000).

In this study, we use a population-based sampling strategy and DNA resequencing data and compare patterns of DNA polymorphism in the mtDNA and on the NRY in the same sample of 172 males from 5 sub-Saharan African

Key words: population growth, *Homo sapiens*, sub-Saharan Africa, mtDNA, NRY, hunter-gatherer.

E-mail: mfh@u.arizona.edu.

Mol. Biol. Evol. 25(3):517–525. 2008
doi:10.1093/molbev/msm279
Advance Access publication December 18, 2007

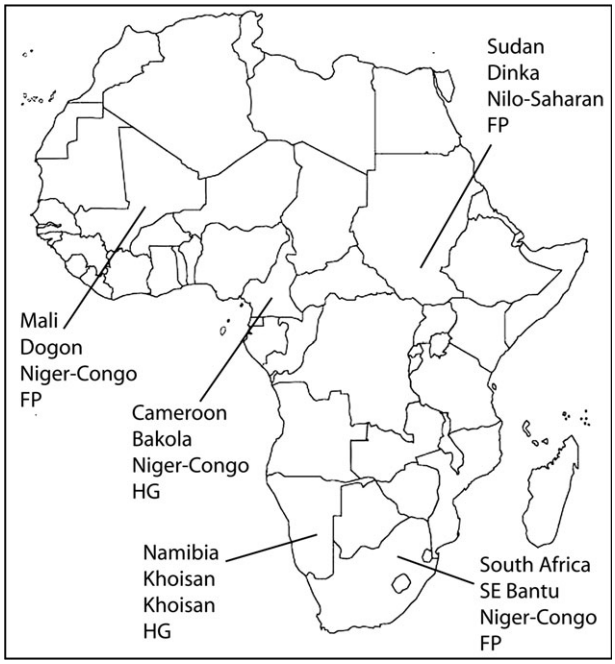

FIG. 1.—Map showing the country of origin, the population name, abbreviation, and the subsistence strategy (FP = food-producer and HG = hunter-gatherer) of the 5 sub-Saharan African populations.

populations. We chose to sequence approximately 780 basepairs (bp) of the mtDNA *cytochrome c oxidase subunit III* (*COIII*) gene and 6.6 kb of noncoding Y chromosomal DNA encompassing 13 *Alu* elements because the former has lower levels of homoplasy than the mtDNA control region and the latter has a 3-fold higher SNP density than other noncoding regions on the NRY (Wilder et al. 2004). The extent of mutation rate heterogeneity in a given region is a concern as this phenomenon may confound the signal of population growth (Aris-Brosou and Excoffier 1996). Previous DNA sequence surveys of the mtDNA control region indicated dramatically different signals of population expansion among African populations: Many food-producing populations exhibited unimodal mismatch distributions consistent with past demographic expansions, whereas most food-gatherers showed ragged distributions. This was interpreted to be the consequence of a recent contraction in effective population size as Bantu-speaking farmers infringed upon hunter-gatherer territory (Excoffier and Schneider 1999). Our samples comprise both food-producing (the Dogon, Dinka, and a group of Southeast [SE] Bantu speakers) and food-gathering groups (the Khoisan and the Bakola). We ask whether mtDNA and Y chromosomal genealogies reflect similar population histories: that is, do the food-producing populations better fit a model of population growth and food-gathering populations better fit a model of constant size for both haploid loci.

Materials and Methods

Populations and Loci Surveyed

We surveyed DNA sequence variation in 5 sub-Saharan African populations (fig. 1) including the Dogon of central Mali (n = 40–49), the Dinka of southern Sudan (n = 23), the Bakola of southern Cameroon (n = 24–25), the Khoisan from Namibia (n = 25), and a group of Southeast Bantu (SE Bantu) speakers from southern Africa (n = 46–50) (table 1). The SE Bantu samples comprise a collection of Zulu, Ndebele, Khosa, Sotho, Swazi, Tswana, Pedi, and Tsonga (Lane et al. 2002). Additionally, we analyzed orthologous DNA sequences from one common chimpanzee (*Pan troglodytes*) to determine ancestral states at each site. The Dinka DNA samples were obtained with written informed consent, and the protocols were approved by the Human Subjects Committee at the University of Arizona. The SE Bantu and Khoisan samples were obtained with either verbal or written consent with approval from the Committee for Research on Human Subjects, University of the Witwatersrand (protocol number M980553). The Dogon samples were collected with verbal consent with approval from the University of Michigan Health Sciences Institutional Review Board. The Bakola samples were collected by Gabriella Spedini and Giovanni Destro-Bisol with verbal informed consent and approval from the University of Rome "La Sapienza."

We generated DNA sequence data for 780 bp of the mitochondrial *cytochrome c oxidase subunit III* (*COIII*) (supplementary material 1, Supplementary Material online) and for 6.6 kb of noncoding NRY encompassing 13 *Alu* elements in the Yα5 subfamily (see Wilder et al. 2004; primers and protocols available upon request, supplementary material 2 [Supplementary Material online]). mtDNA and NRY resequence data were collected from the same individuals; however, sample numbers varied slightly between loci. Nucleotide sequences were obtained using an ABI Prism 3730 automated sequencer and aligned using Sequencher version 4.1 (Genecodes Corporation, Ann Arbor, MI). Insertion/deletion polymorphisms were excluded from all analyses.

Population Genetic Analyses and Tests of Population Growth

Population parameters such as nucleotide diversity (π and θ) and Tajima's D were calculated using DnaSP version 4.0 (Rozas et al. 2003). Inferences of population expansion were made using 4 different methods. First, we calculated Fs statistic of Fu (1997), which is based on the probability of having a number of haplotypes greater or equal to the observed number of samples drawn from a constant-sized population. To complement this method, we calculated the R_2 statistic (Ramos-Onsins and Rozas 2002), which is based on the difference between the number of singleton mutations and the average number of nucleotide differences. Ramos-Onsins and Rozas (2002) demonstrated that these statistics have the greatest power to detect population expansion for nonrecombining regions of the genome under a variety of different circumstances, especially when population sample sizes are large (~50, Fu's Fs) or when sample sizes are small (~10, R_2). They also found that the power of the R_2 statistic is relatively high when the number of segregating sites is low (e.g., <20). The significance of Fu's Fs and R_2 was obtained by examining the null distribution of 5,000 coalescent simulations of these statistics using DnaSP. Significantly large negative Fu's Fs values and

Table 1
mtDNA and NRY Polymorphism Data for Each Sub-Saharan African Population

	Population	N[a]	Sites[b]	Hp[c]	π[d]	θ[d]	Tajima's D	Number of Singletons	Fu's Fs	R_2	P SSD[e]	Fu's Fs[f]
mtDNA	Khoisan	25	10	8	2.00	2.70	−0.800	2	−1.651	0.096	0.469	−1.737
	Bakola	24	6	6	1.00	1.60	−1.135	2	−1.956	0.084	0.892[c]	−1.478
	SE Bantu	50	16	18	2.60	3.60	−0.857	6	−8.768**	0.077	0.328	−4.304*
	Dinka	23	15	12	1.90	4.10	−1.887*	12	−7.135**	0.065**	0.141	−6.735**
	Dogon	49	8	8	0.80	1.80	−1.472	6	−3.842**	0.060	0.011*	−1.038
NRY	Khoisan	25	10	7	2.50	2.40	0.089	1	0.183	0.133	0.514[g]	−0.108
	Bakola	25	14	9	3.70	3.40	0.286	3	−0.127	0.137	0.428[g]	−0.673
	SE Bantu	46	10	7	2.70	2.10	0.762	3	1.525	0.141	0.288[g]	1.826
	Dinka	23	8	5	2.40	2.00	−0.696	2	1.771	0.162	0.000**	2.188
	Dogon	40	12	8	2.30	2.60	−0.402	4	−0.071	0.102	0.025*	1.300

[a] Number of individuals.
[b] Number of polymorphic sites.
[c] Number of haplotypes.
[d] Calculated as % per gene per year.
[e] Probability of the SSDs statistic from the mismatch distribution.
[f] Average Fu's Fs value calculated from 100 random subsample of 20 individuals.
[g] Populations for which 95% confidence intervals for Θ_0 and Θ_1 overlap, thereby rejecting population growth.
* $P < 0.05$; ** $P < 0.001$.

significantly positive R_2 values were taken as evidence of a population expansion.

To further test for evidence of population expansion, we used coalescent simulations to generate maximum likelihood estimates of population parameters for mtDNA and the NRY under the infinite sites model using the program GENETREE version 9.0 (R. C. Griffiths, http://www.stats.ox.ac.uk/~griff/ software.html). GENETREE uses a coalescent Markov chain Monte Carlo approach to search the state space of DNA sequence evolution (Bahlo and Griffiths 2000). Unique gene trees describing both mtDNA and NRY mutational histories for each of the 5 African populations were constructed using Seq2tr (Wilder et al. 2004). For each population, we generated maximum likelihood estimates of the population mutation rate (θ_{ml}) under a model of constant population size and a model of exponential population growth, where we specified the range of growth parameters (β). The maximum likelihood estimate of θ_{ml} was then used to calculate N_e (or N_o) under each of the demographic models. Likelihood ratio tests of the log likelihood of the θ_{ml} produced under models of constant population size versus population growth were performed. The significance of θ_{ml} was evaluated using a simple chi-square test with one degree of freedom. To calculate effective population size, we used previously published mutation rate estimates (1.58×10^{-8} mutations per site per year for the mtDNA *COIII* region and 4.19×10^{-9} mutations per site per year for the NRY) with a generation time of 25 years (Wilder et al. 2004; Wilder and Hammer 2007).

The mismatch distributions were also examined (Rogers and Harpending 1992) using the program, ARLEQUIN version 2.000 (Excoffier et al. 2005). The number of observed differences between pairs of mtDNA or NRY haplotypes was compared with the expected distribution of differences under a specified demographic model (i.e., constant population size or population growth). The mismatch distribution method uses estimated parameters of the expansion (τ, θ_0, and θ_1) to perform coalescent simulations of stepwise expansions and create new estimates of the same parameters (τ*, θ_0*, and θ_1*). The estimated demographic model is tested by obtaining the sum of the squared differences (SSDs) between the observed and the estimated mismatch distributions (Schneider and Excoffier 1999). A significant SSD P value is interpreted here as departure from the estimated demographic model, which is population expansion when $\tau > 0$ and $\theta_1 > \theta_0$ and population stationarity when $\hat{\tau} = 0$ or $\hat{\theta}_1 = \theta_0$. Population stationarity was also inferred if the 95% confidence intervals for θ_1 and θ_0 overlapped, even if the P value of the SSD was not significant (Excoffier and Schneider 1999). To take into account known deviation from the infinite sites model for the mtDNA, the Kimura 2-parameter model was used (Kimura 1980) and mutation rates followed a gamma distribution with a shape parameter $\alpha = 0.22$ (Schneider and Excoffier 1999).

Results

Patterns of Nucleotide Diversity

Table 1 summarizes nucleotide diversity in the sub-Saharan African populations studied here. The number of segregating sites (s) identified at the 2 loci is similar: for mtDNA *COIII* s ranges from 6 in the Bakola to 16 in the SE Bantu samples, whereas s ranges from 8 in the Dinka to 14 in the Bakola samples for the NRY. The average value of π per locus is 1.6 for mtDNA and 2.7 for the NRY, whereas the average value of θ per locus is similar for both loci (2.6 and 2.5, respectively). In contrast, the average number of singleton substitutions is almost twice as high for mtDNA (5.0) compared with the NRY (2.6), although this difference is not statistically significant. The number of mtDNA singleton mutations ranges from 2 in the food-gatherer populations to 12 in the Dinka (mean of 7 in the food-producers), whereas the number of NRY singletons only ranges from 1 to 4 (mean of 3 in the food-producers).

For mtDNA, Fu's Fs values for all populations are negative, with the 3 food-producing populations (SE Bantu and Dinka) having statistically significant negative Fs

Table 2
Comparison of Observed and Simulated Fu's *Fs* Values for mtDNA and the NRY

	mtDNA–NRY *Fs*[a]	5% Quantile[b]	Variance[c]	*P*[d]	mtDNA *s*[e]	NRY *s*[e]
Khoisan	−1.83	−5.15	9.85	0.263	11.3	9.60
Bakola	−1.83	−5.03	9.92	0.250	7.10	15.8
SE Bantu	−10.2	−6.49	15.7	0.009*	22.5	10.1
Dinka	−8.91	−5.17	10.2	0.008*	20.6	7.05
Dogon	−3.77	−5.82	12.7	0.131	9.86	12.6

[a] The observed difference between the mtDNA and NRY Fu's *Fs* values. Fu's *Fs* was calculated in DnaSP.
[b] The 5% lower bound of the distribution of simulated mtDNA–NRY Fu's *Fs* values.
[c] The variance of the distribution.
[d] The probability of the observed mtDNA–NRY Fu's *Fs* value.
[e] The average number of segregating sites generated from 10,000 coalescent simulations (which can be compared with the observed values in table 1).
* $P < 0.05$.

values ($P < 0.001$). For the NRY, 2 populations have positive *Fs* values (not statistically significant) and the others have values close to zero. Only a single population had a statistically significant R_2 value (the Dinka for mtDNA; $P < 0.001$). To test for the effect of differing sample sizes on Fu's *Fs* statistic, we randomly subsampled 20 individuals from each population 100 times for both mtDNA and the NRY (table 1). The overall trend observed from reducing the sample sizes to 20 was for Fu's *Fs* to become less negative, except for situations where sample size was already small. For mtDNA, the only noticeable effect of subsampling was for the Dogon, which no longer had a significant Fu's *Fs* value. For the NRY, subsampling did not change any of the inferences (i.e., no population has a statistically significant Fu's *Fs* value).

To test whether the mtDNA and NRY Fu's *Fs* values were significantly different from each other, we conducted coalescent simulations to create a sampling distribution of the difference between the mtDNA and NRY Fu's *Fs*. We simulated neutral genealogies, conditioned on the θ values estimated from GENETREE under the null model of constant population size and using the sample size of each population (which ranged from 23 to 50). We calculated Fu's *Fs* for each of 10,000 simulated data sets after assuring that the average number of segregating sites generated was comparable with those observed (table 2). We arbitrarily subtracted the NRY Fu's *Fs* values from the mtDNA Fu's *Fs* values and asked whether the observed mtDNA–NRY Fu's *Fs* was significantly small under the distribution ($\alpha = 0.05$). We find that the mtDNA and NRY Fu's *Fs* values are significantly different for the SE Bantu and the Dinka (both food-producers), whereas they were not significantly different for the Dogon (a food-producer), Khoisan, and Bakola (both hunter-gatherers).

GENETREE Simulations

Table 3 lists the maximum likelihood estimates of the population mutation rate, growth rates, effective population sizes, and likelihood ratio tests for each population for each locus. For mtDNA, 3 food-producing populations better fit a model of exponential growth (likelihood ratio score [LRS] = 6.494, 6.720, 4.783 for the SE Bantu, Dinka, and Dogon, respectively), whereas the 2 hunter-gatherer populations better fit a model of constant size (Khoisan and Bakola). In contrast, all NRY populations better fit a model of constant population size.

Mismatch Distributions

Figure 2 shows mismatch distributions for mtDNA *COIII* and the NRY. For mtDNA, the SE Bantu and the Dinka fit a model of population growth. The Khoisan mtDNA mismatch distribution also appears to fit a model of growth; however, the θ_0 and θ_1 95% confidence intervals are very close to overlapping ($\theta_0 = 0.000$–1.416 and $\theta_1 = 2.269$–602.722) and in fact do overlap for their 99% confidence intervals ($\theta_0 = 0.000$–2.455 and $\theta_1 = 0.913$–6818.972). Thus, similar to Excoffier and Schneider (1999), the Khoisan mtDNA mismatch distribution is interpreted as being consistent with population stationarity. SSD was significant for the Dogon and the Bakola mtDNA mismatch distributions, indicating a poor fit to the stepwise growth model. In contrast to mtDNA, all populations reject a model of population expansion for the NRY (fig. 2).

Although there is a great deal of consistency among the methods used to infer changes in population size, some differences are noteworthy. For example, the SE Bantu mtDNA Fu's *Fs*, GENETREE, and mismatch distributions are consistent in supporting a model of growth, whereas the R_2 statistic is not statistically significant (i.e., does not reject a model of constant population size). This apparent contradiction may be explained by the finding the R_2 statistic has less power than Fu's *Fs* to detect expansion when sample sizes are large, as is the case here ($n = 50$). The mismatch distribution test provided the least consistent results as judged from the results of the other tests.

Discussion

Here we present the first comparative analysis of DNA resequence data from both the NRY and mtDNA in the same set of sub-Saharan African population samples and ask whether the genealogical histories of these 2 loci are concordant in their signals of population expansion. Similar to previous observations, we find that the mtDNA results indicate that, in general, the food-producers better fit a model of population expansion, whereas the hunter-gatherers better fit a model of population stationarity. In contrast, the NRY resequence data indicate that all 5 populations better fit

Table 3
Population Parameters Estimated Using GENETREE for Constant Size (Upper Row for Each Population) and Exponential Growth (Lower Row) Demographic Models

Locus	Pop	θ_{ml}[a]	β_{ml}[b]	N_e or N_o[c]	Likelihood Score (SE)	LRS[d]
mtDNA	Khoisan	2.97	—	4,950	2.12×10^{-10} (9.82×10^{-12})	0.704
		3.90	1.12	6,500	3.02×10^{-10} (8.68×10^{-13})	
	Bakola	1.90	—	3,167	5.31×10^{-07} (6.92×10^{-09})	1.910
		4.50	6.00	7,500	1.38×10^{-06} (1.02×10^{-08})	
	SE Bantu	5.00	—	8,183	1.25×10^{-16} (1.37×10^{-18})	6.494**
		10.50	9.05	17,500	3.22×10^{-15} (4.32×10^{-17})	
	Dinka	5.54	—	9,233	1.01×10^{-07} (5.35×10^{-09})	6.720**
		17.50	14.4	29,167	2.90×10^{-06} (2.50×10^{-07})	
	Dogon	2.21	—	3,683	3.64×10^{-07} (1.62×10^{-09})	4.783*
		7.70	15.15	12,833	3.46×10^{-06} (1.04×10^{-08})	
NRY	Khoisan	2.54	—	2,318	9.27×10^{-11} (2.26×10^{-12})	0.361
		2.50	0.10	2,282	1.11×10^{-10} (6.27×10^{-13})	
	Bakola	4.16	—	3,797	7.77×10^{-13} (1.15×10^{-14})	0.193
		5.00	0.80	4,564	8.56×10^{-13} (2.10×10^{-14})	
	SE Bantu	2.30	—	2,099	2.72×10^{-11} (5.96×10^{-13})	−0.231
		2.50	0.10	2,282	2.42×10^{-11} (6.28×10^{-14})	
	Dinka	1.91	—	1,743	1.55×10^{-6} (2.34×10^{-8})	0.057
		2.00	0.10	1,826	1.59×10^{-6} (2.19×10^{-9})	
	Dogon	2.96	—	2,702	1.02×10^{-11} (1.98×10^{-13})	−0.053
		3.00	0.10	2,738	9.94×10^{-12} (3.04×10^{-13})	

[a] Maximum likelihood estimate of θ.
[b] Maximum likelihood estimate of the growth parameter, β.
[c] N_e for constant population size model, N_o for present day population size, exponential growth model.
[d] Likelihood ratio score.
* $P < 0.05$, ** $P < 0.01$.

a model of constant population size. In the following sections, we explore 3 possible explanations for why different patterns of population growth were inferred from these mtDNA and NRY resequencing data sets: 1) variation in mutation rate and mode, 2) natural selection, and 3) sex-specific demographic processes.

Variation in Mutation Rate and Mode

Two concerns with comparing patterns of mtDNA and NRY sequence variation are that mtDNA has much higher mutation and homoplasy rates than nuclear DNA (Horai et al. 1995; Takahata and Satta 1997). As pointed out by Fay and Wu (1999), after passing through a stepwise bottleneck, the more rapid and severe fluctuations in the frequency spectrum of mitochondrial versus autosomal DNA are due only to the mitochondrial genome's smaller population size and have nothing to do with its higher mutation rate. However, the higher mutation rate for mtDNA compared with the NRY may contribute to a difference in power to detect population growth. For this reason, we sequenced much longer segments of the NRY to ensure that

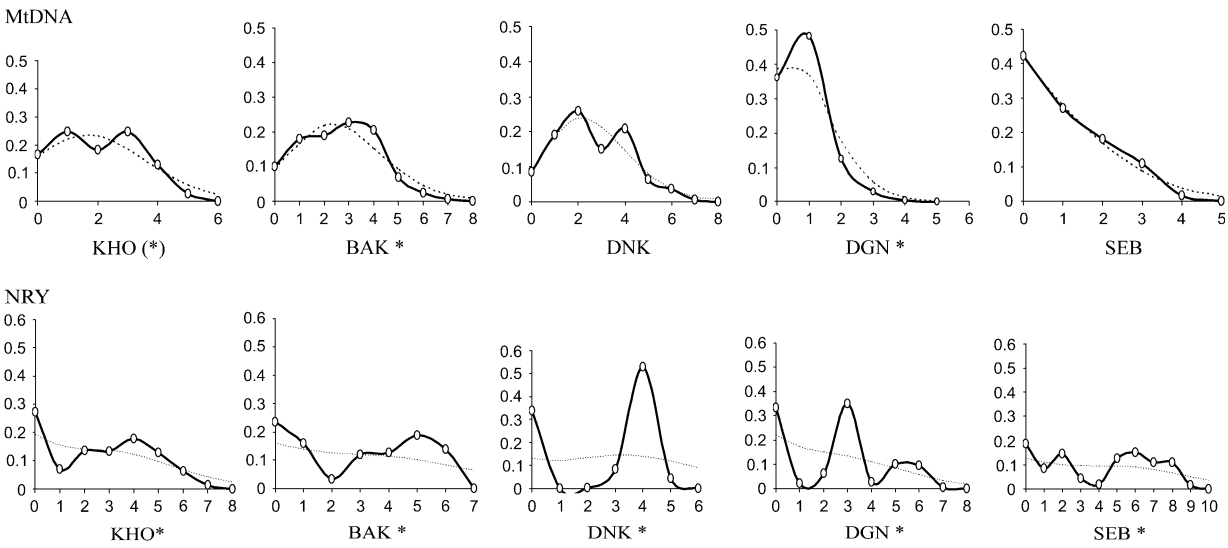

FIG. 2.—Mismatch distributions for mtDNA and NRY. * Denotes rejection of growth model (by either significant SSD values or overlap of 95% confidence intervals for Q_0 and Q_1). (*) Denotes the rejection of the growth model by the overlap of the 90% confidence intervals for θ_0 and θ_1.

a comparable number of segregating sites were recovered from the 2 haploid systems. Indeed, π and θ (calculated per sequence) are actually consistently higher for the NRY than for mtDNA (table 1). As such, it does not appear that mutation rate differences between the NRY and mtDNA are contributing to the differing patterns of growth observed here. This is also illustrated by the fact that the Dogon mtDNA shows a signal of growth, whereas the NRY does not, despite a greater number of segregating sites on the NRY. We incorporated a shape parameter (α) for the gamma distribution of the mutation rate. This minimizes the influence of mutation rate heterogeneity, which can lead to erroneous inference of population growth. (Slatkin and Hudson 1991; Rogers and Jorde 1995; Aris-Brosou and Excoffier 1996). Importantly, different patterns of population growth for food-producers and food-gatherers were initially inferred from D-loop sequences, a region of the mitochondrial genome suffering the highest levels of rate heterogeneity (Meyer et al. 1999; Ingman et al. 2000). Therefore, we do not believe that either mutation rate or rate heterogeneity are important factors contributing to the observed discrepancy between our mtDNA and NRY results.

Natural Selection

Positive directional selection is another process that may mimic the effects of population growth. Both haploid regions are particularly prone to the effects of periodic selective sweeps (genetic hitchhiking; Maynard Smith and Haigh 1974), which are expected to lead to transient phases when there is an excess of rare variants over neutral expectations (Kaplan et al. 1989; Braverman et al. 1995). Although positive selection is thought to have reduced variation on the NRY (Malaspina et al. 1990; Dorit et al.1995; Jaruzelska et al. 1999; Pritchard et al. 1999) and to have influenced mtDNA variation in some geographic regions (Mishmar et al. 2003; Ruiz-Pesini et al. 2004), there is still no strong evidence that differential selection underlies contrasting patterns of mtDNA and NRY variation in human populations (Wilder et al. 2004). None of the 5 populations examined here has a significant excess of rare NRY variants; and although it remains possible that directional selection is the underlying cause of an excess of rare mtDNA variants over neutral expectations, this explanation requires a selective agent that affects food-producers but not food-gatherers.

Another form of selection that could mimic the signal of population growth is weak purifying selection, which has been implicated as the factor underlying an observed excess of low-frequency nonsynonymous polymorphism in mtDNA (Nachman et al. 1994; Nachman 1998; Ballard and Dean 2001). However, the excess of rare mtDNA polymorphism in some African populations is not caused solely by replacement polymorphisms: When the mtDNA data are reanalyzed using only third position sites, results for Fu's *Fs* and the R_2 statistics are similar to results of analyses including all sites (data not shown). Interestingly, the mtDNA data for the food-producers do exhibit an excess number of polymorphic nonsynonymous sites, whereas those of the hunter-gatherers do not (McDonald-Kreitman test Fisher's exact test P value = 0.0001 and 0.070, respectively). This suggests weaker purifying selection on the food-producer versus the food-gatherer populations, which may in turn suggest larger effective population sizes for the food-gatherers in the past.

Sex-Specific Demographic Processes

In this section, we turn to demographic explanations for the observed discrepancy between mtDNA and NRY results. The different mtDNA patterns for food-producers and food-gatherers were initially explained by a model of Pleistocene demographic expansions followed by more recent population crashes for marginalized hunter-gatherers, effectively erasing any signature of population growth (Excoffier and Schneider 1999). Later work by Ray et al. (2003) examined the effects of a spatial expansion in a subdivided population. Their computer simulations demonstrated that a population with low levels of gene flow among neighboring demes (N_em) during a spatial expansion usually do not exhibit the signature of expansion, whereas demes experiencing high N_em values tend to show statistically significant negative values of Fu's *Fs* statistic (Ray et al. 2003). When range expansions are simulated with small deme sizes followed by recent demographic growth (i.e., deme sizes and levels of gene flow with surrounding demes increase well after the spatial expansion), the outcome is similar to simulations of spatial expansions with demes of constant size that always exchanged large numbers of individuals with their neighbors (Ray et al. 2003). In other words, spatial expansions followed by relatively recent population growth with differing rates of gene flow among neighboring demes (N_em) can also produce different signatures of population growth at the molecular level.

Drawing upon the simulation results of Ray et al. (2003), we suggest that a similar difference between effective population sizes and/or rates of gene flow (N_em) for males versus females may explain the discrepancy observed between our mtDNA and NRY results. For example, food-producing males may have experienced either a smaller effective population size (N_e) or a lower rate of migration (m) than females during a phase of expansion. What are the possible causes of lower N_e or m for males of food-producing populations? Two common cultural practices may lead to lower values of N_em for the NRY: polygyny and patrilocality. Polygyny, a marriage practice that allows males (but not females) more than one spouse, is widespread in many parts of the world, especially in Africa (Dorjahn 1959; Konotey-Ahulu 1980; Spencer 1980; Strassmann 2003). When males father children with more females than females do with males, the result is an increase in the variance in reproductive success among males, which lowers their N_e relative to females (Low 1988). Interestingly, food-producers are generally described as substantially more polygynous than hunter-gatherers (Cavalli-Sforza 1986; Biesele and Royal 1999). It is also well established that the widespread practice of patrilocality (defined as the tendency of females to migrate and males to remain in their natal groups) can result in lower rates of male migration (Murdock 1981). Although most agricultural societies are patrilocal (Murdock 1967),

hunter-gatherer groups are typically referred to as bilocal, that is, as spending time living with both the father's and the mother's families (Marlowe 2004). The importance of polygyny and patrilocality in shaping patterns of maternally and paternally inherited variation has been suggested previously for African populations (Destro-Bisol, Donati, et al. 2004; Wood et al. 2005). Assuming that these processes evolved as populations shifted from foraging to food-producing lifestyles, they may have played an important role in shaping the distinctive patterns of mtDNA and NRY polymorphism observed here.

Conclusions

This work constitutes the first direct comparison of mtDNA and NRY resequence data from the same African samples for the purpose of examining patterns of population expansion. The 4 tests that we employed (especially the Fu's *Fs*, the R_2 statistic and coalescent simulations methods) were very consistent in revealing different patterns of population size change for the 2 haploid compartments of the genome. Although all tests may be sensitive to unknown structure within populations (Ptak and Przeworski 2002; Hammer et al. 2003), we emphasize that the mismatch distribution test actually has the weakest power to detect growth (Ramos-Onsins and Rozas 2002). We find that mtDNA *COIII* data from African food-producers better fit a model of population growth, whereas those for food-gathering populations better fit a model of stationarity. In striking contrast, NRY data from both the food-producers and the food-gatherers sampled here better fit a model of constant population size. We hypothesize that these patterns are the result of differences in sex-specific demographic processes—in particular, asymmetrical migration and/or reduced male effective population sizes. These sex-biased demographic processes are expected to significantly alter the frequency spectrum of mtDNA and NRY polymorphisms during large spatial/demographic expansions (Ray et al. 2003). Such an expansion is known to have occurred recently as farmers speaking Niger-Congo Bantu languages expanded from a southern Cameroonian homeland over most of subequatorial Africa beginning ~4,000 years ago (Holden 2002; Lane et al. 2002; Destro-Bisol, Coia, et al. 2004; Wood et al. 2005). It still remains to be resolved whether the Bantu expansions and the spread of farming in Africa, or changes in population size that took place much earlier in the Pleistocene (Excoffier and Schneider 1999; Harpending and Rogers 2000), are responsible for the differential patterns observed here. Along theses lines, further analysis of populations with known differences in marriage customs or migration patterns would be extremely valuable. For example, the finding of similar signals of strong growth for both mtDNA and the NRY in a nonpolygynous (or bilocal) food-producing population (i.e., thereby mitigating a reduction in male effective population size) would support the hypothesis that recent changes in cultural practices made an impact on patterns of genetic diversity (Wilkins and Marlowe 2006). The collection of additional genetic data sets that minimize ascertainment biases, along with the development of more realistic models of human demography that incorporate nonequilibrium processes, will aid future analyses of the effects of sex-biased processes on patterns of genetic variation.

Supplementary Material

Supplementary materials 1 and 2 are available at *Molecular Biology Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We would like to thank Amy Russell, Heather Norton, Stephen Zegura, and Michael Nachman for helpful discussions. Publication of this work was made possible by grant GM-53566 from the National Institute of General Medical Sciences (to M.F.H.). Its contents are solely the responsibility of the authors and do not necessarily reflect the official views of the National Institutes of Health. Funding was also provided by National Science Foundation (NSF) Doctoral Dissertation Improvement Grant (to M.M.P.), and an NSF Integrative Graduate Education and Research Traineeships grant (to M.M.P.).

Literature Cited

Aris-Brosou S, Excoffier L. 1996. The impact of population expansion and mutation rate heterogeneity on DNA sequence polymorphism. Mol Biol Evol. 13:494–504.

Bahlo M, Griffiths RC. 2000. Inference from gene trees in a subdivided population. Theor Popul Biol. 57:79–95.

Ballard JW, Dean MD. 2001. The mitochondrial genome: mutation, selection and recombination. Curr Opin Genet Dev. 11:667–672.

Biesele M, Royal K. 1999. Africa; Mbuti. In: Richard B, Daly R, editors. The Cambridge encyclopedia of hunters and gatherers. Cambridge: Cambridge University Press. p. 210–214.

Braverman JM, Hudson RR, Kaplan NL, Langley CH, Stephan W. 1995. The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. Genetics. 140:783–796.

Cavalli-Sforza L. 1986. African Pygmies. Orlando (FL): Academic Press.

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.

Di Rienzo A, Wilson AC. 1991. Branching pattern in the evolutionary tree for human mitochondrial DNA. Proc Natl Acad Sci USA. 88:1597–1601.

Dorit RL, Akashi H, Gilbert W. 1995. Absence of polymorphism at the ZFY locus on the human Y chromosome. Science. 268:1183–1185.

Dorjahn VR. 1959. The factor of polygyny in African demography. In: Bascom WR, Melville JH, editors. Continuity and change in African cultures. Chicago (IL): University of Chicago Press. p. 87–112.

Excoffier L. 2002. Human demographic history: refining the recent African origin model. Curr Opin Genet Dev. 12:675–682.
Excoffier L, Laval G, Schneider S. 2005. Arlequin: an integrated software package for population genetics data analysis. Version 3. Evol Bioinform Online. 1:47–50.
Excoffier L, Schneider S. 1999. Why hunter-gatherer populations do not show signs of Pleistocene demographic expansions. Proc Natl Acad Sci USA. 96:10597–10602.
Fay JC, Wu CI. 1999. A human population bottleneck can account for the discordance between patterns of mitochondrial versus nuclear DNA variation. Mol Biol Evol. 16:1003–1005.
Fu YX. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics. 147:915–925.
Garrigan D, Hammer MF. 2006. Reconstructing human origins in the genomic era. Nat Rev Genet. 7:669–680.
Hammer MF, Bigham A, Garrigan D, Krenz J, Mobasher Z, Nachman MW, Wilder JA, Wood E. 2004. Heterogeneous patterns of variation among multiple X-linked loci in humans: the possible role of diversity reducing selection. Genetics. 167:1841–1853.
Hammer MF, Blackmer F, Garrigan D, Nachman MW, Wilder JA. 2003. Human population structure and its effects on sampling Y chromosome sequence variation. Genetics. 164:1495–1509.
Harpending H, Rogers A. 2000. Genetic perspectives on human origins and differentiation. Annu Rev Genomics Hum Genet. 1:361–385.
Hey J. 1997. Mitochondrial and nuclear genes present conflicting portraits of human origins. Mol Biol Evol. 14:166–172.
Holden CJ. 2002. Bantu language trees reflect the spread of farming across sub-Saharan Africa: a maximum-parsimony analysis. Proc Biol Sci. 269:793–799.
Horai S, Hayasaka K, Kondo R, Tsugane K, Takahata N. 1995. Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. Proc Natl Acad Sci USA. 92:532–536.
Ingman M, Kaessmann H, Paabo S, Gyllensten U. 2000. Mitochondrial genome variation and the origin of modern humans. Nature. 408:708–713.
Jaruzelska J, Zietkiewicz E, Labuda D. 1999. Is selection responsible for the low level of variation in the last intron of the *ZFY* locus? Mol Biol Evol. 16:1633–1640.
Kaplan NL, Hudson RR, Langley CH. 1989. The "hitchhiking effect" revisited. Genetics. 123:887–899.
Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol. 16:111–120.
Konotey-Ahulu F. 1980. Procreative superiority index (MPSI): the missing coefficient in African anthropogenetics. Br Med J. 281:1700–1702.
Lane AB, Soodyall H, Arndt S, Ratshikhopha ME, Jonker E, Freeman C, Young L, Morar B, Toffie L. 2002. Genetic substructure in South African Bantu-speakers: evidence from autosomal DNA and Y-chromosome studies. Am J Phys Anthropol. 119:175–185.
Low BS. 1988. Measures of polygyny in humans. Curr Anthropol. 29:189–194.
Lundstrom R, Tavare S, Ward RH. 1992. Modeling the evolution of the human mitochondrial genome. Math Biosci. 112:319–335.
Malaspina P, Persichetti F, Novelletto A, Iodice C, Terrenato L, Wolfe J, Ferraro M, Prantera G. 1990. The human Y chromosome shows a low level of DNA polymorphism. Ann Hum Genet. 54(Pt 4):297–305.
Marlowe FW. 2004. Is human ovulation concealed? Evidence from conception beliefs in a hunter-gatherer society. Arch Sex Behav. 33:427–432.
Maynard Smith J, Haigh J. 1974. The hitch-hiking effect of a favourable gene. Genet Res. 23:23–35.
Meyer S, Weiss G, Von Haeseler A. 1999. Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. Genetics. 152:1103–1110.
Mishmar D, Ruiz-Pesini E, Golik P, et al. (13 co-authors). 2003. Natural selection shaped regional mtDNA variation in humans. Proc Natl Acad Sci USA. 100:171–176.
Murdock GP. 1967. Ethnographic atlas. Pittsburgh (PA): University of Pittsburgh Press.
Murdock GP. 1981. Atlas of World Cultures. Pittsburgh (PA): University of Pittsburgh Press.
Nachman MW. 1998. Deleterious mutations in animal mitochondrial DNA. Genetica. 102–103:61–69.
Nachman MW, Boyer SN, Aquadro CF. 1994. Nonneutral evolution at the mitochondrial NADH dehydrogenase subunit 3 gene in mice. Proc Natl Acad Sci USA. 91:6364–6368.
Pritchard JK, Seielstad MT, Perez-Lezaun A, Feldman MW. 1999. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. Mol Biol Evol. 16:1791–1798.
Przeworski M, Hudson RR, Di Rienzo A. 2000. Adjusting the focus on human variation. Trends Genet. 16:296–302.
Ptak SE, Przeworski M. 2002. Evidence for population growth in humans is confounded by fine-scale population structure. Trends Genet. 18:559–563.
Ramos-Onsins SE, Rozas J. 2002. Statistical properties of new neutrality tests against population growth. Mol Biol Evol. 19:2092–2100.
Ray N, Currat M, Excoffier L. 2003. Intra-deme molecular diversity in spatially expanding populations. Mol Biol Evol. 20:76–86.
Rogers AR, Harpending H. 1992. Population growth makes waves in the distribution of pairwise genetic differences. Mol Biol Evol. 9:552–569.
Rogers AR, Jorde LB. 1995. Genetic evidence on modern human origins. Hum Biol. 67:1–36.
Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics. 19:2496–2497.
Ruiz-Pesini E, Mishmar D, Brandon M, Procaccio V, Wallace DC. 2004. Effects of purifying and adaptive selection on regional variation in human mtDNA. Science. 303:223–226.
Schneider S, Excoffier L. 1999. Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. Genetics. 152:1079–1089.
Sherry ST, Rogers AR, Harpending H, Soodyall H, Jenkins T, Stoneking M. 1994. Mismatch distributions of mtDNA reveal recent human population expansions. Hum Biol. 66:761–775.
Slatkin M, Hudson RR. 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. Genetics. 129:555–562.
Spencer P. 1980. Polygyny as a measure of social differentiation in Africa. In: Mitchell JC, editor. Numerical techniques in Social Anthropology. Philadelphia (PA): ISHI Press. p. 117–160.
Strassmann BI. 2003. Social monogamy in human society. In: Reichard H, Boesch C, editors. Mating strategies and partnerships in birds, humans and other mammals. Cambridge: Cambridge University Press. p. 177–189.
Takahata N, Satta Y. 1997. Evolution of the primate lineage leading to modern humans: phylogenetic and demographic inferences from DNA sequences. Proc Natl Acad Sci USA. 94:4811–4815.
Thomson R, Pritchard JK, Shen P, Oefner PJ, Feldman MW. 2000. Recent common ancestry of human Y chromosomes: evidence from DNA sequence data. Proc Natl Acad Sci USA. 97:7360–7365.
Wall JD, Przeworski M. 2000. When did the human population size start increasing? Genetics. 155:1865–1874.
Wilder JA, Hammer MF. 2007. Extraordinary population structure among the Baining of New Britain. In: Friedlaender JS, editor.

Genes, Language, and Culture History in the Southwest Pacific. Oxford: Oxford University Press. p. 199–207.
Wilder JA, Mobasher Z, Hammer MF. 2004. Genetic evidence for unequal effective population sizes of human females and males. Mol Biol Evol. 21:2047–2057.
Wilkins JF, Marlowe FW. 2006. Sex-biased migration in humans: what should we expect from genetic data? Bioessays. 28:290–300.
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.

Yoko Satta, Associate Editor

Accepted December 7, 2007