Genetic Evidence for Unequal Effective Population Sizes of Human Females and Males

Jason A. Wilder,† Zahra Mobasher,* and Michael F. Hammer*†*

*Division of Biotechnology, †Department of Ecology and Evolutionary Biology, University of Arizona, Tucson

The time to the most recent common ancestor (TMRCA) of the human mitochondria (mtDNA) is estimated to be older than that of the nonrecombining portion of the Y chromosome (NRY). Surveys of variation in globally distributed humans typically result in mtDNA TMRCA values just under 200 thousand years ago (kya), whereas those for the NRY range between 46 and 110 kya. A favored hypothesis for this finding is that natural selection has acted on the NRY, leading to a recent selective sweep. An alternate hypothesis is that sex-biased demographic processes are responsible. Here, we re-examine the disparity between NRY and mtDNA TMRCAs using data collected from individual human populations—a sampling strategy that minimizes the confounding influence of population subdivision in global data sets. We survey variation at 782 bp of the mitochondrial *cytochrome c oxidase subunit 3* gene as well as at 26.5 kb of noncoding DNA from the NRY in a sample of 25 Khoisan, 24 Mongolians, and 24 Papua New Guineans. Data from both loci in all populations are best described by a model of constant population size, with the exception of Mongolian mtDNA, which appears to be experiencing rapid population growth. Taking these demographic models into account, we estimate the TMRCAs for each locus in each population. A pattern that is remarkably consistent across all three populations is an approximately twofold deeper coalescence for mtDNA than for the NRY. The oldest TMRCAs are observed for the Khoisan (73.6 kya for the NRY and 176.5 kya for mtDNA), whereas those in the non-African populations are consistently lower (averaging 47.7 kya for the NRY and 92.8 kya for mtDNA). Our data do not suggest that differential natural selection is the cause of this difference in TMRCAs. Rather, these results are most consistent with a higher female effective population size.

Introduction

Our knowledge of patterns of genetic variation in the human genome is disproportionately shaped by two loci, the mitochondrial DNA (mtDNA) and the nonrecombining portion of the Y chromosome (NRY). Despite this, relatively few studies have directly compared patterns of DNA sequence variation between these two genomic compartments within human populations. In part, this is because the human NRY has extraordinarily low levels of sequence diversity (Malaspina et al. 1990; Dorit, Akashi, and Gilbert 1995; Hammer 1995; Whitfield, Sulston, and Goodfellow 1995; Jaruzelska, Zietkiewicz, and Labuda 1999; Shen et al. 2000; Sachidanandam et al. 2001), which has made characterization of variation difficult and labor intensive, even at global scales (Underhill et al. 1997). In contrast, mtDNA has proved to be a prolific source of DNA variation, even among very local populations (e.g., Vigilant et al. 1991). The primary cause for these disparate patterns is variation in the spontaneous mutation rate, with base substitutions in mtDNA accumulating approximately an order of magnitude faster than in the NRY (Ingman et al. 2000; Thomson et al. 2000). However, once this difference in mutation rate is taken into account, mtDNA and the NRY should reveal similar evolutionary histories, assuming they are evolving neutrally in a panmictic population with an equal breeding sex ratio. The degree to which these conditions are satisfied, and the extent to which evolutionary forces equally influence mtDNA and the NRY, remain open questions in human evolutionary genetics.

One of the most intriguing observations regarding the evolutionary histories of human mtDNA and Y chromosomes is that they are estimated to have very different times to the most recent common ancestor (TMRCA), with that of mtDNA estimated at 171.5 to 238 thousand years ago (kya) Ingman et al. 2000; Tang et al. 2002) and estimates for the NRY ranging between 46 and 109 kya in recent studies (Pritchard et al. 1999; Thomson et al. 2000; Hammer and Zegura 2002; Tang et al. 2002). Because the TMRCA of a selectively neutral locus is influenced primarily by its effective population size (N_e), the observed disparity between mtDNA and the NRY is somewhat unexpected. These loci are typically assumed to have equal N_e values in neutral evolutionary models and are, therefore, also expected to have similar TMRCAs. Taken alone, however, it is difficult to determine whether the observed difference between mtDNA and the NRY reflects anything more than simple stochasticity in the coalescent process (e.g., Hudson and Turelli 2003). Multilocus comparisons with other portions of the genome, however, indicate that the NRY has significantly less diversity (and, thus, a shorter genealogy) than expected under a standard neutral model (Shen et al. 2000). Although the reasons for this reduction in variation remain unclear, these findings suggest that mtDNA and the NRY may be influenced differently by natural selection or sex-specific demographic processes.

A leading hypothesis to explain the comparatively recent ancestry of the human NRY is that positive directional selection has played a strong role in shaping nucleotide diversity in this compartment of the genome (Malaspina et al. 1990; Dorit, Akashi, and Gilbert 1995; Whitfield, Sulston, and Goodfellow 1995; Jaruzelska, Zietkiewicz, and Labuda 1999; Pritchard et al. 1999). Because it is nonrecombining and haploid, the NRY acts as a single locus that may be particularly prone to periodic diversity-reducing selective sweeps (e.g., Maynard Smith and Haigh 1974; Begun and Aquadro 1992). Global surveys of nucleotide polymorphism on the NRY have typically shown low levels of variation and a significant excess of rare variants over

neutral expectations (Pritchard et al. 1999; Shen et al. 2000). Although this pattern is consistent with recent positive directional selection, it may also be caused by recent population growth, fine-scale population structure, or a combination of these factors (Tajima 1989a,1989b; Slatkin and Hudson 1991; Fu and Li 1993; Braverman et al. 1995; Ptak and Przeworski 2002; Hammer et al. 2003). Thus, directional selection acting on the NRY has been difficult to confirm or exclude because it is confounded by a number of demographic processes that are likely to have shaped human history. Interestingly, however, Hammer et al. (2003) observed that by sampling local NRY variation from discrete populations, in contrast to the "grid-sampling" strategy (in which a few individuals are sampled from many different populations) typically adopted in global surveys, they no longer recovered an excess of low-frequency mutations over neutral expectations. It was only by pooling samples across populations that rare variants began to exceed expected frequencies in their sample—a pattern compatible with population structure, rather than growth or selection, as the cause of the observed skew in the frequency spectrum in global NRY surveys (Hammer et al. 2003).

While the importance in accounting for population structure when making inferences regarding human evolutionary history has been demonstrated for many portions of the human genome (Ptak and Przeworski 2002; Hammer et al. 2003), relatively few studies have directly assessed variation across multiple loci using sampling schemes that minimize the confounding influence of subdivision. Here, we re-examine the apparent discrepancy in mtDNA and NRY TMRCAs using variation sampled from three discrete human populations: the Khoisan of southern Africa, Khalks of Mongolia, and highland Papua New Guineans. For each population sample, we compare mtDNA and NRY polymorphism ascertained through direct sequencing to obtain unbiased estimates of diversity at each locus. This represents the first population-based analysis of mtDNA and NRY TMRCAs from the same samples and also the first to use a uniform coalescent-based approach to estimate population parameters and the fit of observed data to alternative demographic models. Our results indicate a remarkably consistent trend across populations whereby the TMRCA of mtDNA is approximately twice as old as the NRY, despite variation in the apparent demographic histories of loci in some populations. We see no evidence that recent positive directional selection acting on the NRY is the cause of this disparity in TMRCAs, and we instead hypothesize that there is a widespread skew in the effective breeding ratio toward an excess of females over males among human populations.

Materials and Methods
DNA Survey Panel

DNA sequence variation at both mtDNA and the NRY was surveyed in the same panel of 73 unrelated males, originating from three populations. These populations are the Khoisan of southern Africa (n = 25, hereafter abbreviated SAN), Mongolian Khalks (n = 24, MNG), and highland Papua New Guineans (n = 24, PNG). All DNA was collected with informed consent according to protocols approved by the Human Subjects Committee at the University of Arizona. In addition, orthologous sequences were determined for a single common chimpanzee (*Pan trogolodytes*).

DNA Regions Surveyed

From the mitochondria, we examined 782 bp of the mitochondrial *cytochrome c oxidase subunit 3* (*Cox3*) gene. The surveyed region encompasses the entire coding sequence, excluding the first base of the first codon and including two bases 3′ of the gene. All of the NRY data analyzed in this work were previously reported in Hammer et al. (2003). This NRY survey region includes a total of 26.5 kb of noncoding DNA, comprising 11.3 kb of the *ARSD* pseudogene, 941 bp upstream of the *SRY* gene, 994 bp encompassing two Yα5 *Alu* elements, 2.7 kb of the YAP region, and 10.5 kb of anonymous noncoding DNA.

DNA Sequence Analysis

Summary statistics describing the NRY population data have been reported in Hammer et al. (2003). For the *Cox3* data set, we used DnaSP version 3.53 (Rozas and Rozas 1999) to estimate parameters, including nucleotide diversity, π (Nei and Li 1979), Watterson's θ (Watterson 1975), Tajima's D (TD [Tajima 1989a]), and Fu and Li's D* (FLD [Fu and Li 1993]). At mutation-drift equilibrium, both π and Watterson's θ estimate the quantity $2N_e\mu$, where N_e is the effective population size and μ is the mutation rate. Both TD and FLD are measures of the degree to which the frequency spectrum of observed mutations conform to equilibrium expectations. In this analysis, only single-nucleotide polymorphisms were considered (i.e., insertion/deletions and length polymorphisms were excluded).

To estimate population parameters such as N_e and the TMRCA, it is necessary to generate locus-specific estimates of the per generation mutation rate. Our specific methodology for estimating this quantity differed for the NRY and mtDNA, although in each case our estimate was based on the degree of sequence divergence between humans and chimpanzee. We assumed the split between humans and chimpanzee to have occurred 6 MYA (Haile-Selassie 2001; Glazko and Nei 2003) and also assumed a 25-year generation time for both males and females. For the NRY, the mutation rate was estimated based on the average number of nucleotide differences between humans and chimpanzee over all available orthologous sequence. For the mtDNA, we estimated the mutation rate following the methodology of Ingman et al. (2000), who calculated divergence from chimpanzee using a Tamura-Nei (1993) model of nucleotide substitution (which accommodates multiple substitutions per site) with gamma distributed rates of substitution among sites. We generated a maximum-likelihood estimate of the shape parameter of the gamma distribution (α) using PAUP* version 4.0b10 (Swofford 1998). To provide a robust estimate of both α and divergence, we assembled all available human *Cox3*

sequences from the Human Mitochondrial Genome Database (M. Ingman, http://www.genpat.uu.se/mtDB/index.html), as well as unpublished sequences from the Hammer Lab, a total of 779 human *Cox3* sequences, and estimated parameters from this data set. The net Tamura-Nei (1993) divergence between humans and chimpanzee, incorporating the estimated value of α, was then calculated using MEGA version 2.1 (Kumar et al. 2001).

Maximum-likelihood estimates of population parameters were generated with the program GENETREE version 9.0 (R.C. Griffiths, http://www.stats.ox.ac.uk/~griff/software.html). This method uses the standard coalescent model to evaluate the probability of obtaining a sample of DNA sequences from an equilibrium Wright-Fisher population. An infinite-sites mutation model is assumed, making this method sensitive to both recombination and recurrent mutation. Nonreticulating gene trees were created for both the mtDNA and NRY data sets using the program Seq2tr, which is distributed as part of the GENETREE package. Adjustments to the data sets were necessary to accommodate several observed violations of the infinite-sites model, as discussed in the *Results* section.

The null demographic model incorporated by GENETREE is one of panmixia and constant population size. In this analysis, we treated the SAN, PNG, and MNG samples separately, as each is likely to represent individuals from a single breeding population. We evaluated each population data set with respect to both a constant sized demographic model and a model incorporating exponential growth, as follows:

$$N(t) = N_0 \exp(-\beta t / N_0) \qquad (1)$$

where N_0 represents the present day population size, $N(t)$ represents the population size t generations in the past, and β represents the intrinsic population growth parameter. The fit of the data to exponential growth versus constant size models was evaluated using likelihood ratio tests.

The results of GENETREE analyses are conditional on population parameters provided by the user. Analysis of a panmictic constant-sized population requires estimation of only a single parameter, θ_{ml} ($= 2N_e\mu$, where N_e is the effective population size and μ is the mutation rate). We estimated θ_{ml} by generating a single-likelihood curve covering a wide range of possible values. The exponential growth model requires the estimation of an additional parameter, β, as shown above. Likelihood estimates of multiple parameters are not independent using this coalescent framework. Thus, to assess the fit of the growth model we estimated β over a wide range of θ_{ml} values to create a likelihood surface. Using the joint maximum-likelihood estimate of θ_{ml} and β generated from this surface as a starting point, we then iteratively estimated the likelihood of these parameters across a narrow range of values until a local maximum was reached.

Coalescent Simulations

To better understand the evolutionary forces influencing mtDNA and the NRY, we performed coalescent simulations conditioned on our observed data using the program "ms" (Hudson 2002; http://home.uchicago.edu/~rhudson1). This program allowed us to generate distributions of expected population genetic summary statistics under a wide range of population histories. Specifically, we used this program to test whether observed patterns of NRY variability were compatible with a historical bottleneck, such as might occur if the NRY were subject to strong positive selection. Specific run parameters are described in the *Results* section below.

Results
DNA Sequence Variation

Complete details regarding patterns of nucleotide variability in the NRY data set are reported in Hammer et al. (2003 [figure 5 therein contains the table of polymorphism used in the present study]). For the mtDNA, a complete table of polymorphism is shown in figure 1. Summary statistics describing NRY and mtDNA variation are shown in table 1. As expected, based on its higher rate of mutation, observed levels of variation for the mtDNA are much higher than for the NRY (π = 0.299% for the pooled mtDNA data versus 0.014% for the NRY data). Summary statistics describing the frequency spectrum of polymorphisms (TD and FLD) are negative in the pooled samples of mtDNA and NRY variation, although only FLD for the mtDNA differs significantly from neutral expectations (−3.244, two-tailed $P = 0.015$). In individual populations, TD and FLD deviate significantly from neutral equilibrium expectations in only two cases. First, the observed FLD for the mtDNA in the MNG population is more negative than expected (−3.018, two-tailed $P = 0.012$). Second, FLD for the NRY in the SAN population is greater than expected (1.512, two-tailed $P = 0.022$).

Mutation Rate
NRY

Our estimate of the mutation rate is based on the observed sequence divergence of 1.59% between humans and chimpanzee divided by twice the divergence time in years since the split of the human and chimpanzee lineages (which we assume to have occurred 6 MYA). Based on this calculation, we estimate the mutation rate for the NRY to be 1.33×10^{-9} mutations per site per year. It should be kept in mind that errors in estimates of the mutation rate will have a linear influence on estimates of the TMRCA. Our present estimate for the NRY is somewhat faster than the 1.03×10^{-9} value (modified here to reflect a 6-Myr human-chimpanzee divergence) of Thomson et al. (2000) from three genic regions. However, their estimate was averaged over both coding and noncoding segments and may have been more highly constrained than the noncoding regions we examine here.

mtDNA

The mtDNA has a high level of mutation rate heterogeneity among sites. From the expanded *Cox3* data set, we estimate the α parameter of a gamma distribution

SNP Position

Population	Sample	48	60	91	136	141	171	232	262	270	271	272	319	330	334	339	369	549a	549b	570	612	618	630	649	681	744	750	760
Consensus		A	G	C	G	A	A	G	A	A	G	T	G	C	C	A	G	G	G	G	C	C	T	T	A	T	T	A
MNG	257	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	A
MNG	259	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.
MNG	260	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.
MNG	261	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.
MNG	262	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.
MNG	263	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.
MNG	266	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.
MNG	267	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.
MNG	268	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.
MNG	269	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.
MNG	270	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.
MNG	271	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.
MNG	272	.	.	A	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
MNG	273	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.
MNG	276	.	.	.	.	.	.	.	.	A	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
MNG	277	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
MNG	278	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
MNG	279	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	C	.	.	.	.	.	.
MNG	281	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.
MNG	282	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
MNG	283	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.
MNG	284	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.
MNG	285	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.
MNG	286	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.
PNG	4	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
PNG	6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.
PNG	7	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.
PNG	13	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.
PNG	14	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
PNG	15	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.
PNG	17	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.
PNG	18	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.
PNG	20	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.
PNG	25	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.
PNG	26	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.
PNG	29	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.
PNG	30	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	A	.	.	.	.	.	.	.	.	.
PNG	34	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.
PNG	51	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
PNG	57	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.
PNG	62	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
PNG	67	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.
PNG	72	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
PNG	82	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.
PNG	84	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
PNG	85	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.
PNG	91	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
PNG	93	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.
SAN	13	.	.	.	G	.	A	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	C	.	.
SAN	16	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	C	.	.
SAN	20	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.
SAN	20	.	.	.	G	.	A	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.
SAN	23	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.
SAN	24	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.
SAN	25	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.
SAN	25	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	C	.	.
SAN	33	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	C	.	.
SAN	45	G	.	.	G	.	.	G	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.
SAN	50	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.
SAN	54	.	.	.	G	.	.	.	.	.	A	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.
SAN	60	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.
SAN	65	G	.	.	G	.	.	G	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.
SAN	77	.	.	.	G	.	A	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.
SAN	78	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.
SAN	118	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.
SAN	127	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.
SAN	140	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.
SAN	301	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	C	.	.
SAN	305	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.
SAN	321	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.
SAN	323	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.
SAN	354	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.
SAN	3043	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	C	.	.
Pan troglodytes		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	A	T	.	C	.	.	C	C	.	.

FIG. 1.—Polymorphic sites in 782 bp of the mitochondrial *Cox3* locus. Numbering refers to the position in the coding sequence of the *Cox3* gene.

describing this variation to be 0.22. Using this α value, the Tamura-Nei (1993) genetic distance between humans and chimpanzee is 18.4%. Dividing this divergence by twice the time since the human-chimpanzee split yields a *Cox3* mutation rate of 1.58×10^{-8} mutations per site per year. In comparison, Ingman et al. (2000) estimate a very similar rate of evolution for the entire mtDNA coding region (1.70×10^{-8} mutations per base pair per year).

Gene Trees

MtDNA and NRY gene trees for the pooled data set are shown in figure 2. For the NRY, the ancestral state of each polymorphic site was determined from comparison with chimpanzee. For cases where no chimpanzee sequence was available, the consensus human sequence was assumed to represent the ancestral state. Two instances of recurrent mutation were identified in the NRY data set based on their presence on different NRY haplogroup backgrounds (see Hammer et al. [2003]). Recurrent mutations were treated as separate sites in our analysis. For the mtDNA, multiple substitutions at a number of sites made determination of ancestry based on chimpanzee sequence uncertain. Accordingly, we assigned ancestral states based on the human consensus sequence. Using this procedure, a gene tree with a single reticulation was produced. This reticulation joined a polymorphic site shared by the SAN and PNG populations (site 549). If we assume that this shared polymorphism has independent origins in each population (i.e., a recurrent mutation), a single nonreticulating gene tree describes our mtDNA data set (fig. 2).

GENETREE Analyses

Using the gene trees produced above for each population and locus, we generated maximum-likelihood estimates of population parameters using the program GENETREE. In each case, we first estimated the value of θ_{ml} describing the population sample for each locus, assuming a model of constant population size. Based on these values of θ_{ml}, we then estimated the N_e and TMRCA of each locus in each population, as well as the likelihood of the observed genealogy (shown in table 2).

Based on the assumption of constant population size, there appears to be substantial heterogeneity in these parameters across populations and loci. For the NRY, the SAN have the deepest TMRCA (78.1 kya), as well as the highest values of θ_{ml} (4.10) and N_e (2,327). The quantities of each of these parameters are approximately 1.5 times higher than respective values in the non-African populations. In contrast, for the mtDNA, the deepest TMRCA is also observed in the SAN (176.5 kya), but the highest values of θ_{ml} and N_e are observed in the MNG population (5.08 and 8,223, respectively). All of the TMRCAs estimated for the mtDNA are considerably older than even the deepest TMRCA from the NRY.

Table 2 shows the joint estimates of θ_{ml} and the growth parameter β obtained when we fit our observed data to a model of exponential growth, as well as the TMRCAs and likelihood values obtained when coalescent simulations were performed using these parameter estimates. The influence of growth is mixed between our populations and loci. For the SAN NRY data set, the maximum-likelihood estimate of β is quite low (0.62), and the resulting TMRCA values are very similar to constant-sized estimates. In contrast, the largest effect of the growth model is observed in the MNG mtDNA data set, where the estimate of β is 69.3, indicating rapid exponential growth. In this data set, the observed TMRCA drops by 44% compared with the constant-sized model (i.e., from 172.6 kya to 96.2 kya).

Table 1
NRY and mtDNA Polymorphism from Each Population

Locus	Population	n	S	π (%)	θ (%)	Tajima's D	Fu and Li's D*
NRY	SAN	25	14	0.014	0.014	−0.007	1.512*
	PNG	24	9	0.008	0.009	−0.592	−0.811
	MNG	24	8	0.006	0.008	−0.701	−1.053
	Pooled	73	30	0.014	0.023	−1.208	−0.528
mtDNA	SAN	25	11	0.272	0.373	−0.906	0.016
	PNG	24	5	0.120	0.171	−0.895	−1.338
	MNG	24	13	0.228	0.445	−1.693	−3.018*
	Pooled	73	26	0.299	0.684	−1.752	−3.244*

NOTE.—Asterisk (*) indicates $0.05 > P > 0.01$.

To assess the fit of our observed data to the constant-sized and exponential growth cases, we performed likelihood ratio tests comparing likelihood estimates for each model (table 2). In only a single case does the more complex growth model provide a significantly better fit to our data than the constant-sized model: the MNG mtDNA data set (LRS = 7.56, $P = 0.006$, df = 1). Thus, we cannot reject a constant-size demographic model to describe our NRY data in any population, nor can we for the mtDNA data sets surveyed in the SAN and PNG populations.

Scaled genealogies from each locus and population are presented in figure 3. TMRCAs and the ages of individual mutations are estimated in each case using GENETREE with maximum-likelihood parameters estimated from the best-fit demographic model (constant-size for all data sets other than the mtDNA MNG). This figure illustrates two robust patterns. First, the genealogy of our African population is substantially deeper than those of the non-African populations at both mtDNA and the NRY. Specifically, the mean non-African mtDNA TMRCA is 52.6% that of the SAN, and the mean non-African TMRCA for the NRY is 64.3% that of the SAN. Second, in all populations, the TMRCA of mtDNA is approximately twice as old as that of the NRY. The ratio of these values is 2.32:1 for the SAN, 1.76:1 for the PNG, and 2.06:1 for the MNG.

Coalescent Simulations

To better understand whether observed patterns of population genetic variation fall within the distributions expected if natural selection had shaped differences in the mtDNA and NRY genealogies, we generated simulated genealogies under a variety of scenarios. Because the SAN population is thought to represent one of the most ancestral extant human lineages (Chen et al. 2000; Semino et al. 2002), and because it would not be influenced by a putative out-of-Africa bottleneck, we conditioned our coalescent analyses on data observed in this population. For our initial simulations, we generated genealogies using the value of θ_{ml} taken from the SAN mtDNA data set, corrected to reflect the mutation rate of the NRY (in other words, setting the N_e of the NRY to equal to that of mtDNA for the purposes of the simulations). We included in these simulations a population bottleneck occurring approximately 78 kya (specifically, a 100-fold reduction in population size for 100 generations), which was meant to
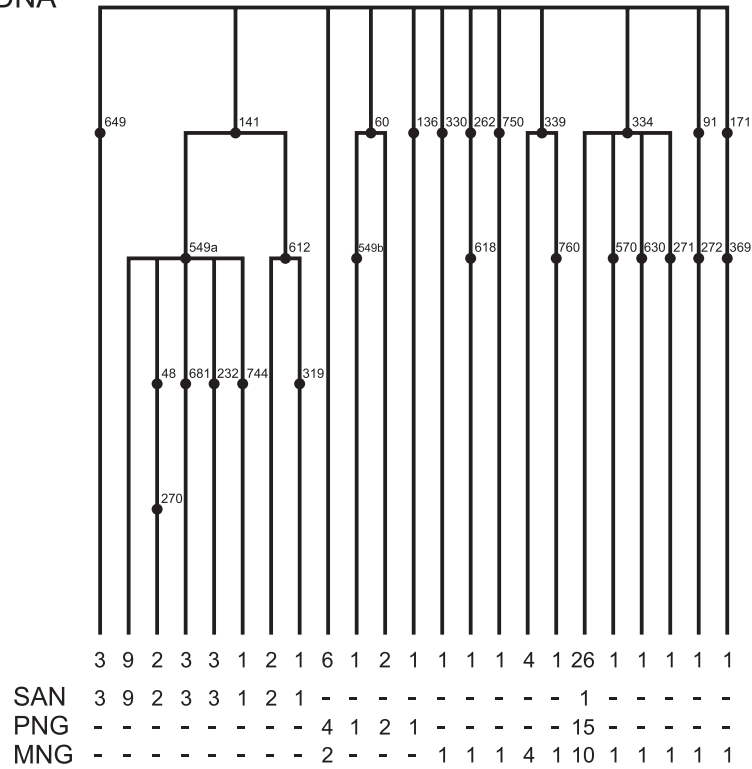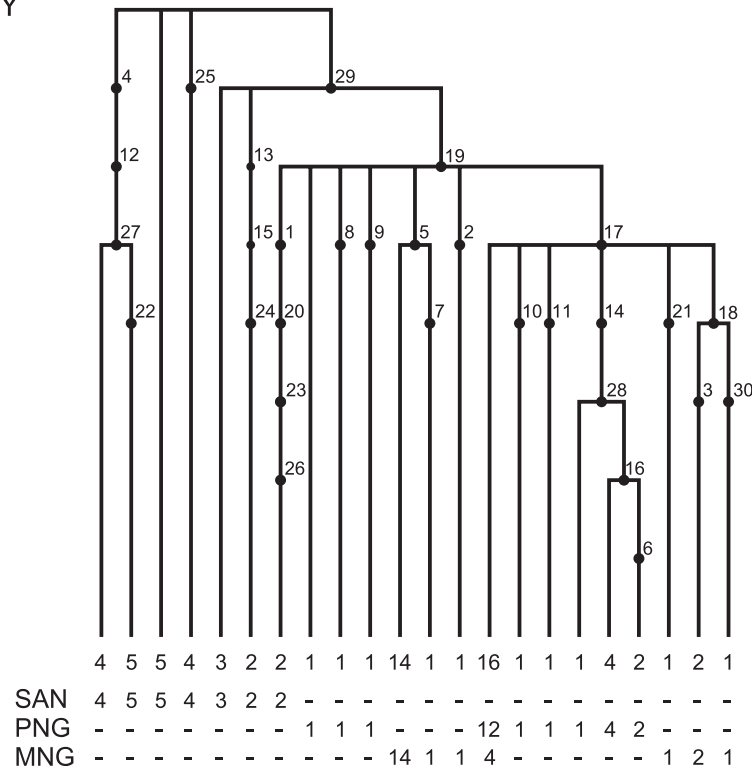
FIG. 2.—Gene trees for both (*A*) and (*B*) are shown. Polymorphic sites for the mtDNA are numbered as shown in figure 1. Sites in the NRY tree are numbered based on their positions (from left to right) in figure 5 of Hammer et al. (2003). The distribution of haplotypes among populations is shown at the base of each tree. Trees are unscaled with respect to genealogy depth.

Table 2
Population Parameters Estimated Using GENETREE for Constant Size (Upper Row for Each Locus/Population) Versus Exponential Growth (Lower Row) Demographic Models

Locus	Population	θ_{ml}	β_{ml}	N_e or N_0	TMRCA (SD) in kya	Likelihood Score (SE)	LRS
NRY	SAN	4.10	—	2327	78.1 (21.8)	1.09×10^{-14} (6.9×10^{-17})	0.167
		4.92	0.64	2792	74.8 (18.6)	1.18×10^{-14} (6.6×10^{-17})	
	PNG	2.98	—	1691	52.3 (18.0)	2.24×10^{-10} (6.5×10^{-13})	2.170
		6.17	5.33	3501	38.7 (7.1)	6.63×10^{-10} (3.5×10^{-12})	
	MNG	2.50	—	1419	48.1 (16.3)	3.28×10^{-9} (4.0×10^{-12})	1.209
		4.14	2.55	2349	39.0 (8.4)	5.99×10^{-9} (6.1×10^{-12})	
mtDNA	SAN	3.53	—	5714	176.5 (54.4)	2.23×10^{-11} (4.4×10^{-14})	1.543
		5.85	2.89	9469	147.0 (29.6)	4.83×10^{-11} (2.5×10^{-13})	
	PNG	1.54	—	2493	89.5 (35.7)	3.23×10^{-6} (1.8×10^{-9})	1.412
		4.02	7.43	8134	72.6 (12.6)	6.54×10^{-6} (9.5×10^{-9})	
	MNG	5.08	—	8223	172.6 (53.3)	5.13×10^{-9} (2.21×10^{-11})	7.560**
		34.7	69.3	56167	96.2 (10.1)	2.25×10^{-7} (1.8×10^{-9})	

NOTE.—N_e for constant size population model; N_0, the present day population size, for exponential growth model. **$0.01 > P > 0.001$.

mimic a near-selective sweep at the estimated TMRCA of the NRY. The distributions of summary statistics describing the genealogies created through this "strong bottleneck" scenario are shown in table 3. The observed NRY data from the SAN population (table 1) lie outside of the central 95% confidence interval of these simulations in two respects: there are too few observed segregating sites (observed 14, simulated CI: 15 to 44), and the observed FLD was too high (observed: 1.512, simulated CI: −2.307 to 0.918). We also performed a "weak bottleneck" scenario, which incorporated a 10-fold size reduction for 100 generations at the same time point. The observed data were similarly incompatible with the "weak bottleneck" genealogies (see table 3).

Several lines of evidence suggest that the Khoisan may have experienced a dramatic reduction in population size in the last several thousand years as agricultural Bantu speakers expanded through southern Africa (Bandelt and Forster 1997; Excoffier and Schneider 1999; Salas et al. 2002). Accordingly, we performed additional coalescent simulations as above incorporating a recent population contraction to check whether this recent size change could mask an earlier bottleneck event. We generated simulated genealogies under several simple population contraction scenarios (six scenarios: historical populations that were twofold, fivefold, and 10-fold larger than the present day with contractions occurring either 2 kya or 10 kya), in conjunction with the "strong bottleneck" described above. Results of the 2-kya contraction scenarios are shown in table 3 (the 10-kya contraction results are very similar and are not shown). Patterns of polymorphism generated by these simulations are inconsistent with the observed data from the SAN (table 1) in all cases. Every simulated data set produces genealogies with significantly more segregating sites than we observe, and none can account for the observed value of FLD. Therefore, we are unable to simulate any bottlenecked genealogies that can account for the observed patterns of NRY variation, even when we incorporate a recent population contraction into the coalescent simulations. Interestingly, simulations of a simple recent population contraction (incorporating a twofold contraction either 2 kya or 10 kya), conditioned on the estimated value of θ_{ml} from our observed NRY data set,

are fully compatible with the observed patterns of polymorphism in the SAN population (data not shown).

Discussion

Like previous studies that have used a global grid-based sampling design, our population-based survey indicates that the TMRCA of the mtDNA is older than that of the NRY. In fact, in each of the three populations we examined, the ultimate coalescence time for the mtDNA is approximately twice as old as that for the NRY. This ratio is remarkably robust considering that that there is heterogeneity among populations and loci with regard to the best-fit demographic model for the observed data. We found that only one of the six locus/population combinations that we examined was better described by a growth than a constant-sized population model (the mtDNA in the MNG population [tables 1 and 2]). This is in sharp contrast to global grid-sampling studies, which have inferred a high growth rate for both mtDNA and the NRY (Pritchard et al. 1999; Ingman et al. 2000; Thomson et al. 2000). In the case of the NRY, the incorporation of exponential growth has led to global estimates of the TMRCA that are substantially younger than our estimate for the SAN population alone (e.g., Pritchard et al. 1999; Thomson et al. 2000). This is a clear indication that TMRCA estimates that incorporate strong exponential growth models based on pooled population samples may produce unreasonably recent coalescent times. Although our sampling design may not be appropriate for generating a robust estimate of the global TMRCA, treatment of our NRY data set as a single panmictic population yields a TMRCA estimate of 100.3 kya, and the incorporation of an island model of population subdivision (following the methodology of Bahlo and Griffiths [2000]) increases this time to 117.2 kya (data not shown). Future studies that explicitly incorporate population subdivision may help to refine global estimates of the TMRCA, although it is not clear whether existing models of genetic structure (e.g., Wright's island model) adequately describe relationships among human populations (Hammer et al. 2003). Despite this, our result confirms that the TMRCA for the NRY is much younger than that of mtDNA, even at very local scales.
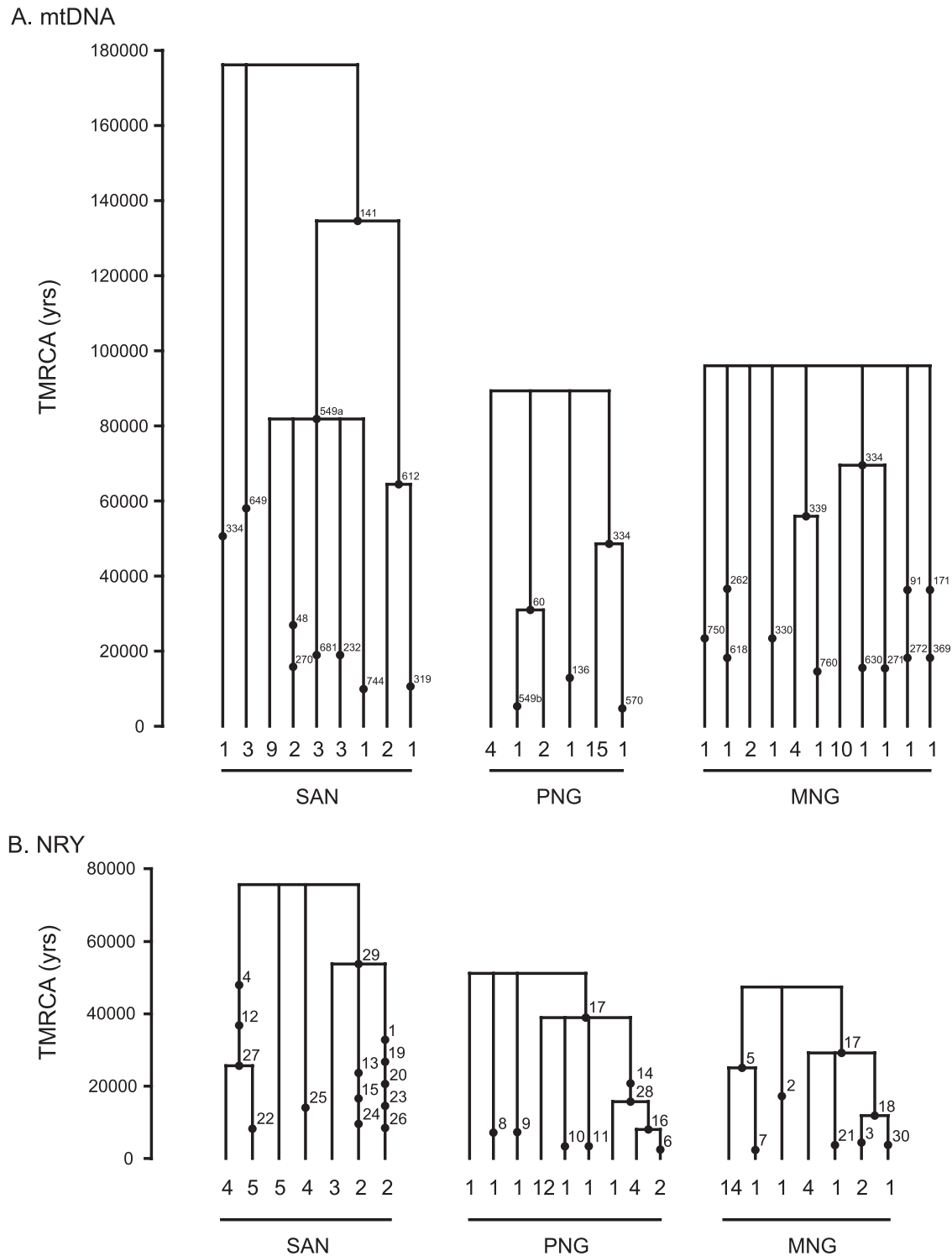
A. mtDNA

B. NRY

FIG. 3.—Gene trees for both mtDNA (*A*) and the NRY (*B*), scaled according to their TMRCAs. These estimates incorporate the best-fit demographic models and parameter estimates indicated in table 2.

Beginning with early studies of the NRY, it has been hypothesized that positive directional selection has reduced variation on the human NRY (Malaspina et al. 1990; Dorit, Akashi, and Gilbert 1995; Whitfield, Sulston, and Good-fellow 1995; Jaruzelska, Zietkiewicz, and Labuda 1999;

Pritchard et al. 1999). All else being equal, a recent selective sweep on the NRY would indeed cause it to have a more recent TMRCA than that of mtDNA in accordance with our results. However, no population exhibits a significant excess of rare variants in our NRY data sets, which is thought to be

one of the genetic signatures of a recent selective sweep (Tajima 1989b; Fu and Li 1993; Braverman et al. 1995). Furthermore, after performing coalescent simulations of bottlenecked genealogies meant to mimic the affects of a selective sweep at the base of the NRY tree, we were unable to produce results that are compatible with the observed levels of variation and the frequency spectrum of mutations (although a simple demographic scenario with no strong bottleneck was compatible with our data). Thus, our results provide no additional support for the hypothesis that differential positive selection causes the disparate TMRCAs of the NRY and mtDNA, although we cannot rule out selective sweeps (acting either globally or locally) that may have occurred before the coalescence of the observed genealogies. It is also notable that positive selection acting on mtDNA, which has recently been hypothesized to influence variation in some geographic regions (Mishmar et al. 2003; Ruiz-Pesini et al. 2004), does not appear to have caused a reduction in TMRCAs relative to those of the NRY in any of the three populations we survey here.

Before considering potential demographic explanations for the consistently older TMRCAs for mtDNA than the NRY, we will consider one additional nonneutral force that may underlie the observed difference: purifying selection. For both genetic systems, diversity (and, hence, the TMRCA) may be reduced by selection acting against deleterious mutations at linked sites (Charlesworth, Morgan, and Charlesworth 1993). Given an equal N_e for mtDNA and the NRY, we expect purifying selection to affect both loci similarly, assuming that deleterious mutations occur at the same rate and have the same average fitness effects. In fact, several lines of evidence suggest that mtDNA is likely to be subject to stronger purifying selection than the NRY. First, mtDNA has a background mutation rate that is nearly an order of magnitude higher than that of the NRY, indicating that the absolute frequency of mutation of any given fitness effect is expected to be higher for mtDNA. Second, mtDNA has much less redundancy with regard to gene content than the NRY. Whereas all mtDNA genes are single copy (and play a vital role in cell metabolism), nearly all genes on the NRY are multicopy or have closely related autosomal or X-linked homologs (Skaletsky et al. 2003). Consequently, we do not expect purifying selection to differentially reduce the effective size of the NRY relative to that of mtDNA.

These lines of reasoning lead us to postulate that sex-specific demographic processes are the most likely causes of the observed discrepancy in the TMRCAs of mtDNA and the NRY. To date, processes that could generate this pattern in humans have received relatively little attention. Based on one of the few direct comparisons of patterns of mtDNA and NRY sequence variability, Tang et al. (2002) suggested that the difference in coalescence times for mtDNA and the NRY (which were given in numbers of generations) could be minimized to some degree (and hence explained) by taking account of evidence (see below) that males have a longer generation time than females (i.e., 30 versus 25 years). However, a caveat that is unmentioned in this work is that the difference in generation times between males and females must be a recent evolutionary phenomenon, such that it is not reflected in estimates of the per generation

Table 3
Summary Statistics for Coalescent Simulations Examining Possible NRY Histories

Model[a]	Statistic	Mean	Lower 95% CI	Upper 95% CI
Strong bottleneck	S	27.240	15	44
	TD	−0.635	−1.642	0.596
	FLD	−0.636	−2.307	0.918
Weak bottleneck	S	37.069	17	71
	TD	−0.208	−1.725	1.660
	FLD	−0.209	−2.085	1.251
Strong bottleneck plus recent twofold contraction (2.5 kya)	S	38.442	22	71
	TD	−1.017	−1.190	0.470
	FLD	−0.962	−2.623	0.709
Strong bottleneck plus recent fivefold contraction (2.5 kya)	S	57.965	33	148
	TD	−1.504	−2.176	0.871
	FLD	−1.545	−3.186	0.677
Strong bottleneck plus recent 10-fold contraction (2.5 kya)	S	75.625	41	261
	TD	−1.712	−2.389	1.414
	FLD	−1.948	−3.638	0.994

NOTE.—Population contractions refer to a reduction from a larger ancestral population to the present day size, which is held constant across simulations.

[a] See text for model details.

mutation rate (in other words, utilizing a longer male than female generation time when estimating the degree of divergence with chimpanzee will nullify the effect of using different generation times when estimating TMRCAs). Furthermore, it is unclear whether the twofold greater TMRCA observed for mtDNA relative to the NRY could be explained by a recent increase in the male generation time. Although a number of studies have indicated that contemporary human males do have slightly longer generation times than females, this difference has never been estimated to approach the magnitude necessary to cause a twofold discrepancy. For instance, a recent study of Icelandic genealogies showed a mean generation interval of approximately 28 years for matrilines and 31 years for patrilines, or a difference of approximately 10% (Helgason et al. 2003). A similar study of a Quebec population produced mean maternal generation times of 29 years and paternal estimates of 35 years (Tremblay and Vézina 2000).

As an explanation for the observed differences in the genealogies of the human NRY and mtDNA, we favor a model in which the human effective population size is skewed toward an excess of females by sex-biased demographic processes. The human mating system has often been considered to be moderately polygynous, based on both surveys of world populations (Murdock 1981; Low 1988) and on characteristics of human reproductive physiology (Harcourt et al. 1981; Anderson and Dixson 2002; Dixson and Anderson 2002). The practice of polygyny, in both the traditional sense and via "effective polygyny" (whereby males tend to father children with more females than females do with males—a common practice in many contemporary western cultures [Low 2000]), would tend to increase the variance in reproductive success among males, thereby lowering their N_e relative to females. This effect will have an influence on the N_e of the NRY, even when practiced sporadically, but can have extraordinary consequences if male mating success is inherited patrilineally. An example of this phenomenon

was recently described in central Asia, where Y chromosomes likely to be descendents of Genghis Khan and his male relatives can be found at exceptionally high frequencies (Zerjal et al. 2003), indicating a vastly disproportionate contribution of male members of this family to the contemporary gene pool.

Another sex-specific demographic process that may cause the observed differences in the genealogies of mtDNA and the NRY is a sex bias in rates of migration among human populations. For instance, the widespread phenomenon of patrilocality (defined anthropologically as the tendency for a wife to move to her husband's natal domicile) could contribute to the observed pattern if it resulted in higher rates of mtDNA than NRY gene flow between genetically distinct populations (Seielstad, Minch, and Cavalli-Sforza 1998). However, until we learn more about the scale of human genetic population structure and how this may differ between males and females, it is difficult to assess whether sex-biased migration alone could cause the observed differences in the genealogies of mtDNA and the NRY. In reality, both polygyny and patrilocality are common occurrences in human cultures, and it is, therefore, not surprising to see patterns in population genetic data that are congruent with these phenomena.

In conclusion, our results indicate that the human NRY tends to have an approximately twofold smaller N_e and TMRCA than mtDNA within human populations. There is no indication from our data that this difference is caused by different forms or intensities of natural selection acting on mtDNA and the NRY. Instead, we favor a hypothesis whereby sex-specific demographic processes act to reduce the male breeding population size. Further studies examining the scale of female and male population structure and focusing on populations with variable mating systems, will help to clarify the degree to which human population genetic variation is shaped by these processes. Regardless of the evolutionary cause of the skew in effective population sizes among the sexes, it is important to take this phenomenon into account in the development of null models describing the distribution of neutral genomic variation. For instance, Shen at al. (2000) observed an approximately 5:1 ratio of autosomal to NRY variability in their global survey of nucleotide variation. This observation differs significantly from neutral expectations based on a one-to-one breeding ratio but is extremely close to the expected results given a breeding ratio of two females per male (Hedrick 2000). Thus, a simple skew in the human breeding ratio may be sufficient to explain the low levels of variation and recent TMRCAs that have been observed for the human NRY.

Acknowledgments

Literature Cited

Anderson, M. J., and A. F. Dixson. 2002. Sperm competition: motility and the midpiece in primates. Nature **416**:496.

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

Bandelt, H. J., and P. Forster. 1997. The myth of bumpy hunter-gatherer mismatch distributions. Am. J. Hum. Genet. **61**:980–983.

Begun, D. J., and C. F. Aquadro. 1992. Levels of naturally occurring DNA polymorphism correlate with recombination rates in D. melanogaster. Nature **356**:519–520.

Braverman, J. M., R. R. Hudson, N. L. Kaplan, C. H. Langley, and W. Stephan. 1995. The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. Genetics **140**:783–796.

Charlesworth, B., M. T. Morgan, and D. Charlesworth. 1993. The effect of deleterious mutations on neutral molecular variation. Genetics **134**:1289–1303.

Chen, Y. S., A. Olckers, T. G. Schurr, A. M. Kogelnik, K. Huoponen, and D. C. Wallace. 2000. mtDNA variation in the South African Kung and Khwe and their genetic relationships to other African populations. Am. J. Hum. Genet. **66**:1362–1383.

Dixson, A. L., and M. J. Anderson. 2002. Sexual selection, seminal coagulation and copulatory plug formation in primates. Folia Primatol. **73**:63–69.

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

Excoffier, L., and S. Schneider. 1999. Why hunter-gatherer populations do not show signs of pleistocene demographic expansions. Proc. Natl. Acad. Sci. USA **96**:10597–10602.

Fu, Y. X., and W. H. Li. 1993. Statistical tests of neutrality of mutations. Genetics **133**:693–709.

Glazko, G. V., and M. Nei. 2003. Estimation of divergence times for major lineages of primate species. Mol. Biol. Evol. **20**:424–434.

Haile-Selassie, Y. 2001. Late Miocene hominids from the Middle Awash, Ethiopia. Nature **412**:178–181.

Hammer, M. F. 1995. A recent common ancestry for human Y chromosomes. Nature **378**:376–378.

Hammer, M., F. Blackmer, D. Garrigan, M. Nachman, and J. Wilder. 2003. Human population structures and its effects on sampling Y chromosome variation. Genetics **164**:1495–1509.

Hammer, M. F., and S. L. Zegura. 2002. The human Y chromosome haplogroup tree: nomenclature and phylogeography of its major divisions. Ann. Rev. Anthropol. **31**:303–321.

Harcourt, A. H., P. H. Harvey, S. G. Larson, and R. V. Short. 1981. Testis weight, body weight and breeding system in primates. Nature **293**:55–57.

Hedrick, P., 2000. Genetics of populations. Jones and Bartlett, Sudbury, Mass.

Helgason, A., B. Hrafnkelsson, J. R. Gulcher, R. Ward, and K. Stefansson. 2003. A populationwide coalescent analysis of icelandic matrilineal and patrilineal genealogies: evidence for a faster evolutionary rate of mtDNA lineages than Y chromosomes. Am. J. Hum. Genet. **72**:1370–1388.

Hudson, R. R. 2002. Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics **18**:337–338.

Hudson, R. R., and M. Turelli. 2003. Stochasticity overrules the "three-times rule": genetic drift, genetic draft, and coalescence times for nuclear loci versus mitochondrial DNA. Evolution **57**:182–190.

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

Jaruzelska, J., E. Zietkiewicz, and D. Labuda. 1999. Is selection responsible for the low level of variation in the last intron of the ZFY locus? Mol. Biol. Evol. **16**:1633–1640.

Kumar, S., K. Tamura, I. B. Jakobsen, and M. Nei. 2001. MEGA2: molecular evolutionary genetics analysis software. Bioinformatics **17**:1244–1245.

Low, B. S. 1988. Measures of polygyny in humans. Curr. Anthropol. **29**:189–194.

———. 2000. Why sex matters: a Darwinian look at human behavior. Princeton University Press, Princeton, NJ.

Malaspina, P., F. Persichetti, A. Novelletto, C. Iodice, L. Terrenato, J. Wolfe, M. Ferraro, and G. Prantera. 1990. The human Y chromosome shows a low level of DNA polymorphism. Ann. Hum. Genet. **54** (pt 4):297–305.

Maynard Smith, J., and J. Haigh. 1974. The hitch-hiking effect of a favorable gene. Genet. Res. **23**:23–35.

Mishmar, D., E. Ruiz-Pesini, P. Golik et al. (13 co-authors). 2003. Natural selection shaped regional mtDNA variation in humans. Proc. Natl. Acad. Sci. USA **100**:171–176.

Murdock, G. P., 1981. Atlas of world cultures. University of Pittsburgh Press, Pittsburgh, Pa.

Nei, M., and W. H. Li. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. USA **76**:5269–5273.

Pritchard, J. K., M. T. Seielstad, A. Perez-Lezaun, and M. W. Feldman. 1999. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. Mol. Biol. Evol. **16**:1791–1798.

Ptak, S. E., and M. Przeworski. 2002. Evidence for population growth in humans is confounded by fine-scale population structure. Trends Genet. **18**:559–563.

Rozas, J., and R. Rozas. 1999. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics **15**:174–175.

Ruiz-Pesini, E., D. Mishmar, M. Brandon, V. Procaccio, and D. C. Wallace. 2004. Effects of purifying and adaptive selection on regional variation in human mtDNA. Science **303**:223–226.

Sachidanandam, R., D. Weissman, S. C. Schmidt et al. (38 co-authors). 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature **409**:928–933.

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

Seielstad, M. T., E. Minch, and L. L. Cavalli-Sforza. 1998. Genetic evidence for a higher female migration rate in humans. Nat. Genet. **20**:278–280.

Semino, O., A. S. Santachiara-Benerecetti, F. Falaschi, L. L. Cavalli-Sforza, and P. A. Underhill. 2002. Ethiopians and Khoisan share the deepest clades of the human Y-chromosome phylogeny. Am. J. Hum. Genet. **70**:265–268.

Shen, P., F. Wang, P. A. Underhill et al. (13 co-authors). 2000. Population genetic implications from sequence variation in four Y chromosome genes. Proc. Natl. Acad. Sci. USA **97**: 7354–7359.

Skaletsky, H., T. Kuroda-Kawaguchi, P. J. Minx et al. (40 co-authors). 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature **423**:825–837.

Slatkin, M., and R. R. Hudson. 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. Genetics **129**:555–562.

Swofford, D., 1998. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4.0b10. Sinauer Associates, Sunderland, Mass.

Tajima, F. 1989a. The effect of change in population size on DNA polymorphism. Genetics **123**:597–601.

———. 1989b. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123**: 585–595.

Tamura, K., and M. Nei. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. **10**:512–526.

Tang, H., D. O. Siegmund, P. Shen, P. J. Oefner, and M. W. Feldman. 2002. Frequentist estimation of coalescence times from nucleotide sequence data using a tree-based partition. Genetics **161**:447–459.

Thomson, R., J. K. Pritchard, P. Shen, P. J. Oefner, and M. W. Feldman. 2000. Recent common ancestry of human Y chromosomes: evidence from DNA sequence data. Proc. Natl. Acad. Sci. USA **97**:7360–7365.

Tremblay, M., and H. Vézina. 2000. New estimates of intergenerational time intervals for the calculation of age and origins of mutations. Am. J. Hum. Genet. **66**:651–658.

Underhill, P. A., L. Jin, A. A. Lin, S. Q. Mehdi, T. Jenkins, D. Vollrath, R. W. Davis, L. L. Cavalli-Sforza, and P. J. Oefner. 1997. Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. Genome Res. **7**:996–1005.

Vigilant, L., M. Stoneking, H. Harpending, K. Hawkes, and A. C. Wilson. 1991. African populations and the evolution of human mitochondrial DNA. Science **253**:1503–1507.

Watterson, G. A. 1975. On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. **7**:256–276.

Whitfield, L. S., J. E. Sulston, and P. N. Goodfellow. 1995. Sequence variation of the human Y chromosome. Nature **378**: 379–380.

Zerjal, T., Y. Xue, G. Bertorelle et al. (23 co-authors). 2003. The genetic legacy of the mongols. Am. J. Hum. Genet. **72**:717–721.