Current Biology, Vol. 13, 464–473, March 18, 2003, ©2003 Elsevier Science Ltd. All rights reserved. DOI 10.1016/S0960-9822(03)00130-1

African Y Chromosome and mtDNA Divergence Provides Insight into the History of Click Languages

Alec Knight,[1,*] Peter A. Underhill,[2] Holly M. Mortensen,[1,4] Lev A. Zhivotovsky,[3] Alice A. Lin,[2] Brenna M. Henn,[1] Dorothy Louis,[1] Merritt Ruhlen,[1] and Joanna L. Mountain[1,2]
[1]Department of Anthropological Sciences
[2]Department of Genetics
Stanford University
Stanford, California 94305
[3]N.I. Vavilov Institute of General Genetics
Russian Academy of Sciences
Moscow 117809
Russia

Summary

Background: About 30 languages of southern Africa, spoken by Khwe and San, are characterized by a repertoire of click consonants and phonetic accompaniments. The Ju|'hoansi (!Kung) San carry multiple deeply coalescing gene lineages. The deep genetic diversity of the San parallels the diversity among the languages they speak. Intriguingly, the language of the Hadzabe of eastern Africa, although not closely related to any other language, shares click consonants and accompaniments with languages of Khwe and San.
Results: We present original Y chromosome and mtDNA variation of Hadzabe and other ethnic groups of Tanzania and Y chromosome variation of San and peoples of the central African forests: Biaka, Mbuti, and Lisongo. In the context of comparable published data for other African populations, analyses of each of these independently inherited DNA segments indicate that click-speaking Hadzabe and Ju|'hoansi are separated by genetic distance as great or greater than that between any other pair of African populations. Phylogenetic tree topology indicates a basal separation of the ancient ancestors of these click-speaking peoples. That genetic divergence does not appear to be the result of recent gene flow from neighboring groups.
Conclusions: The deep genetic divergence among click-speaking peoples of Africa and mounting linguistic evidence suggest that click consonants date to early in the history of modern humans. At least two explanations remain viable. Clicks may have persisted for tens of thousands of years, independently in multiple populations, as a neutral trait. Alternatively, clicks may have been retained, because they confer an advantage during hunting in certain environments.

Introduction

In the early 1960's, Joseph Greenberg grouped all languages characterized by a repertoire of click consonants and phonetic accompaniments (hereafter referred to simply as clicks) within the Khoisan phylum [1]. Greenberg's Khoisan includes languages of southern Africa that are spoken by Khwe (traditional herders) and San (traditional foragers). Two linguistic isolates, the Hadzane and Sandawe languages of Tanzania, were also included in Khoisan by Greenberg. Hadzane is the language spoken by the Hadzabe of the Lake Eyasi region in north-central Tanzania. Many Hadzabe continue to rely on hunting and gathering for subsistence. Currently a relatively small population [2], the Hadzabe may have descended in situ from among the earliest fully modern human inhabitants of the region [3]. Languages among San are so divergent from one another that their relationships remain controversial [4, 5]. Such diversity suggests ancient population divergences.

A small sample of Ju|'hoansi (previously identified in the literature as !Kung) San has been tested extensively for both mitochondrial DNA (mtDNA) and nonrecombining Y chromosome (NRY) haplotypic variation. Other Khwe and San populations have also been tested at these genetic segments, but those results must be interpreted in light of long association and gene flow with Bantu speakers [6]. Several genetic systems, and nongenetic evidence, indicate long-term isolation of Ju|'hoansi [7, 8]. The NRY biallelic mutations they share with other groups appear to be tens of thousands of years old [9]. MtDNAs of Ju|'hoansi are distinct and form an ancient, separate cluster of lineages [10, 11].

Correspondence within Africa between genetic differentiation and linguistic classification has been recognized for some time [7, 12–14]. Hadzabe DNA variation has the potential to play a key role in furthering our understanding of the history of click languages in Africa. Their language, with the exception of the repertoire of click consonants and accompaniments, is dissimilar to every other known language [15]. Classical genetic markers indicated similarity of Hadzabe to Bantu speakers [7]. A small number of individuals studied for mtDNA hypervariable regions 1 and 2 (HV1 and HV2) haplotypes revealed no recent shared ancestry with Ju|'hoansi [10]. Those limited genetic data led to the suggestion of recent gene flow between Hadzabe and neighboring peoples, but did not indicate genetic affinity with other click-speaking groups [7, 16].

To further elucidate ancestral contributions to the current Hadzabe gene pool and the relationship of Hadzabe to San, we examined mtDNA and NRY variation of Hadzabe individuals in the context of regional and continental African genetic diversity, and we compared those data with data for Ju|'hoansi. Our goal was to distinguish among a number of possibilities. Are the Hadzabe descendents of click speakers who arrived relatively recently from southern Africa? Did ancestors of the Hadzabe migrate to southern Africa relatively recently and give rise to San? Are most Hadzabe descendents of neighboring Bantu-, Cushitic-, or Nilotic-speaking peoples? Alternatively, do Hadzabe share only ancient genetic traits with San, revealing great divergence from San at rapidly evolving loci and thereby

*Correspondence: aknight@stanford.edu
[4]Present address: Department of Biology, University of Maryland, College Park, Maryland 20742.

indicating a very ancient separation? Our observations of extensive genetic divergence between click-speaking peoples of eastern and southern Africa, in the context of nongenetic evidence, suggest that click consonants date to early in the history of modern humans.

Results

Mitochondrial DNA Variation

We sequenced HV1 and HV2 mtDNA for Tanzanian individuals who self-identified as Hadzabe (n = 49), Datoga (n = 18), Iraqw (n = 12), and Bantu speakers, primarily of Sukuma ancestry (n = 21). We also typed these individuals for a set of restriction sites outside of HV1 and HV2, and these restriction sites are diagnostic for haplogroups L1*, L2, and L3* and M*. Sukuma, Iraqw, and Datoga population samples revealed similar control region nucleotide diversity (0.023– 0.024). Nucleotide diversity of Hadzabe was lower (0.012).

To examine the relationship between Hadzabe and San, we included published Ju|'hoansi data [10] in mtDNA analyses (Figure 1). For inference of mtDNA phylogeny, we took into consideration the highly conserved nature of markers such as those that define major haplogroups and the extent of homoplasy at many control region sites. Ju|'hoansi mtDNA lineages were previously found to have separated early from other mtDNA haplotypes observed in Africa, and they have retained considerable diversity within the L1d haplogroup [11]. All Ju|'hoansi mtDNAs studied carry the ancestral states 247A and 16230G. Polarity was determined by outgroup comparison to Neandertal and chimpanzee [17, 18]. Sites are numbered according to the Cambridge Reference Sequence (CRS) [19]. Ju|'hoansi mtDNA haplotypes have been assigned to haplogroups L1d and L1k [20] within the set of isolated, divergent L1i [11] lineages of Africa. Of 49 Hadzabe mtDNAs studied, 1 was in haplogroup L1a [20], and 48 were within the clade defined by the stable mutations 247G and 16230A, which include haplogroups L2 and L3*. L2 and L3* are further characterized by 16390A and 3594C, respectively. Two groups of haplotypes (group 5 of haplogroup L2 and group 10 of haplogroup L3*) were particularly frequent among Hadzabe and rare among other populations studied. No mtDNA haplotypes were shared between San and Hadzabe.

We also compared Tanzanian HV1 mtDNA sequences with published sequences for other Africans. We excluded HV2 in this analysis, as far more populations are represented by HV1 published data. That comparison, in terms of genetic distance among populations, is summarized in Figure 2 and Table S1. Ju|'hoansi, Mbuti, Biaka, and Hadzabe samples are the most genetically distant from other studied populations. Genetic distance was greater between San and Hadzabe than between any other pair of sub-Saharan African populations.

Overall, mtDNA analyses revealed that the Hadzabe population is primarily represented in haplogroups L2 and L3*, includes both haplotypes unique to Hadzabe and haplotypes shared with neighboring populations, and is genetically very distant from the Ju|'hoansi population. All Ju|'hoansi mtDNA haplotypes diverge from all

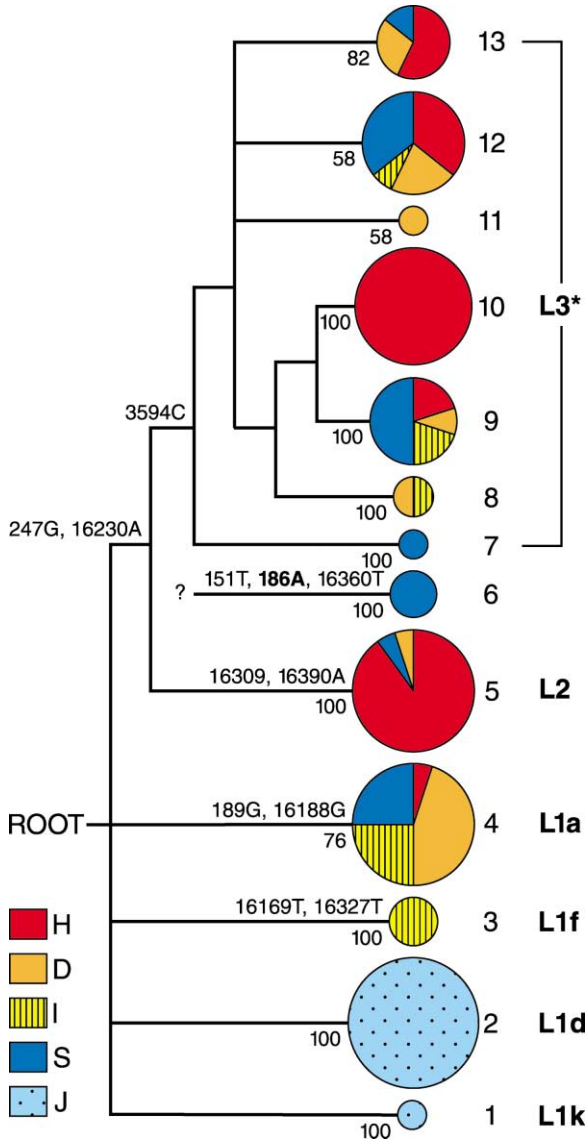

Figure 1. Pylogenetic Relationships and Population Distributions among Mitochondrial Lineages within and among Study Populations

Tanzanian populations are H, Hadzabe (n = 49); D, Datoga (n = 18); I, Iraqw (n = 12); and S, Sukuma (n = 21). Ju|'hoansi San of Botswana and Namibia are indicated as J, n = 24. Ju|'hoansi sequences with internal missing data were excluded. Terminal clades were derived from aligned mitochondrial control region sequence haplotypes spanning 768 nucleotide positions comprising HV1 and HV2. Rooted with Neandertal [17, 18], both neighbor-joining and maximum parsimony resulted in consistent topology. The tree is a skeleton phylogeny that defines principal clades and pronounced subclades. Groups 1–13 are the highest level clusters with strong support. The stable nucleotide positions 189, 247, 3594; 16169, 16188, 16230, 16327, 16309, and 16390 (numbered according to the CRS [46]) determine basal topology of the phylogenetic tree and provide resolution of haplogroups L1*, L2, and L3* (see [20] and references therein), rooted with Neandertal and chimpanzee. Ju|'hoansi L1d and L1k haplotypes have been characterized previously [11, 20]. Diagnostic character states are shown above the branches (transversion bold). The numbers below the terminal branches are the percentages found among 374.5 million TBR branch swaps. Group 6 is unusual in that all four haplotypes were distinct from other haplotypes in control region sequence, including a unique transversion, and all four typed (outside the control region) to four different haplogroups (L1*, L2, L3*, and M*). In group 12, one Sukuma and one Datoga typed to M* (a derivative of L3*).

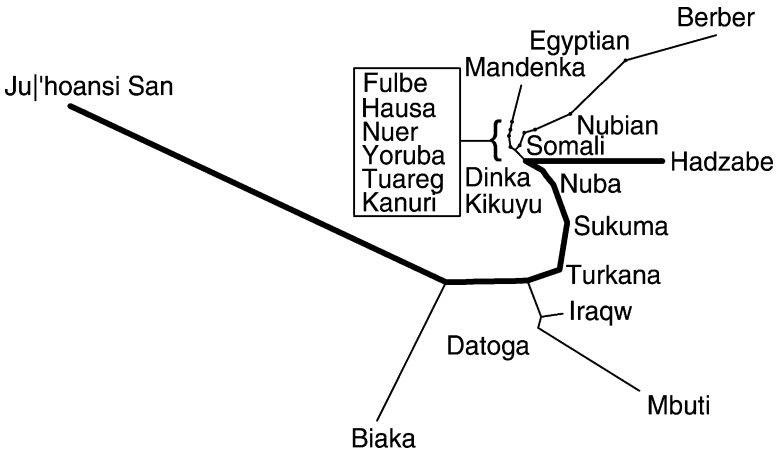

Figure 2. Population Relationships as a Summary of Genetic Distances Derived from HV1 mtDNA Sequences of African Origin

Genetic distances (Table S1) were derived from mutation and drift among 740 HV1 mtDNA sequences. A tree derived from F_{ST} was essentially identical in topology and branch lengths. Negative branches were collapsed to zero. Sequences for Hadzabe, Datoga, Iraqw, and Sukuma are original. Others are from the literature [10, 11, 49–51]. Published sequences with internal missing data were excluded. Sample sizes in this analysis were: 49 Hadzabe, 12 Iraqw, 19 Datoga, 32 Sukuma, 24 Ju|'hoansi, 27 Somali, 61 Fulbe, 37 Turkana, 81 Nubian, 44 Dinka, 17 Biaka, 11 Mbuti, 26 Tuareg, 25 Kikuyu, 17 Hausa, 14 Kanuri, 33 Yoruba, 77 Mandenka, 91 Egyptian, 11 Nuer, 14 Nuba, and 18 Berber. The relatively large distance (heavy line) between San and Hadzabe is consistent with independent retention of click consonants over much of modern human prehistory.

those of the Hadzabe at the root of the human mtDNA phylogenetic tree.

Tanzanian Y Chromosome UEP Variation

We typed all Tanzanian male individuals for a set of basal NRY unique event polymorphisms (UEPs; Figure S1) that define haplogroups [21]. Data are presented in Table 1 with previously published data for other African populations. Tanzanian populations harbor representatives of haplogroups A (one Iraqw), B (primarily Hadzabe), E (all populations), and C + F (primarily Iraqw and Datoga). All Hadzabe fell within haplogroups B and E, with the exception of one individual whose haplotype placed him in haplogroup C or F. Haplogroup A, frequent within San [22, 23], was not observed in Hadzabe. Over 50% of Hadzabe fell within haplogroup B2b, also frequent in San. Remaining Hadzabe fall within haplogroups E3a and E3b.

Distributions of M2, M35, and M112 Mutations

Given the occurrence of haplogroups B2b (M112+), E3a (M2+), and E3b (M35+) within Hadzabe, we explored

Table 1. Y Chromosome Haplotype Frequencies as Percentages in 18 African Groups

			Haplotype and Defining Mutation[a]								
Language Family; Subfamily[b]	Population	N[c]	A (91)	B2a (150)	B2b (112)	B1,B (60)	E3a (2)	E,D[d] (35)	C,F (YAP)		Study
Kh; isolate	Hadzabe	23			52		30	13		4	Present
Kh; C Khoisan	Khwe	26	12				54	31	4		[23]
Kh; N Khoisan	!Kung (Sekele)[e]	64	36		8		39	11	6		[23]
Kh; N Khoisan	Ju|'hoansi/Sekele[f]	39	44		28		18	10			[22]
NK; C Bantu	Sukuma	32		9	6		63	6	16		Present
NK; Adamawa	Mixed	72	1	13			54		4	28	[23]
NK; Bantoid	Bamileke	48				4	96				[23]
NK; NW Bantu	Biaka	20		5	30		65				[22]
NK; NW Bantu	Lissongo	4			25		75				[22]
NK; NW Bantu	Mixed	41		7			90			2	[23]
NK; NW Oti-Volt.	Mossi	49				2	90	2	6		[23]
NK; W Atlantic	Mixed	74	3				57		34	7	[23]
NS; E Sudanic	Datoga	8					13	63		25	Present
NS; C Sudanic	Mixed	9	22		22		33		11	11	[23]
NS; C Sudanic	Mbuti	12		8	25		42		25		[22]
AA; S Cushitic	Iraqw	6	17					33	17	33	Present
AA; Chadic	Mixed	54	2	4			13	4	7	70	[23]
AA; Semitic	Mixed	135	8				2	70	4	16	[23]

[a] Nomenclature as outlined by the Y Chromosome Consortium [21].

[b] Language family; subfamily, according to [52]. For language families: AA, Afro-Asiatic; Kh, Khoisan; NK, Niger-Kordofanian; NS, Nilo-Saharan. For language subfamilys: C, Central; N, Northern; S, Southern; E, Eastern; NW, Northwest.

[c] Number of NRYs studied.

[d] This column includes the frequency of individuals in haplogroups E1, E2, E* plus D only.

[e] Individuals identified in [23] as !Kung are more accurately identified as Sekele.

[f] Individuals identified in [22] as "Khoisan" are more accurately identified as a mixed sample of Ju|'hoansi and Sekele (approximately 50% each).

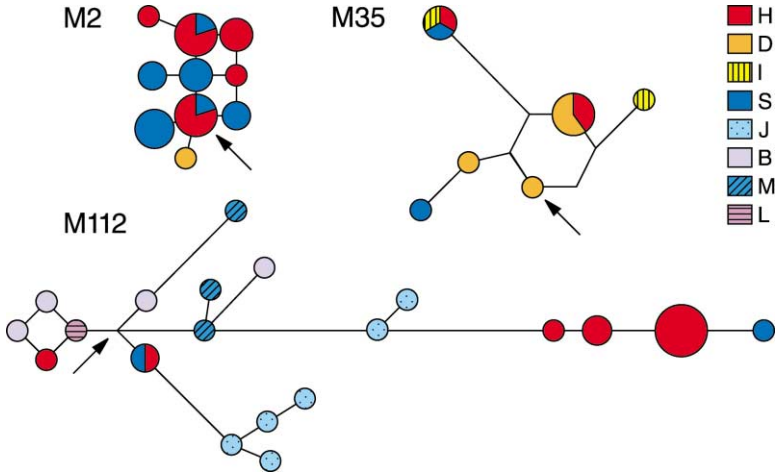

Figure 3. Median-Joining Networks for Each of Three SNP-Defined NRY Clades, Inferred from Variation at Eight STR Loci

Arrows indicate the ancestral nodes of median-joining networks [45] used for ρ estimates. M2 and M35 were rooted with each other (as sister clades) and additional YAP+ haplotypes. M112 was rooted with M150 (its sister clade). M2 and M35 were typed for Tanzanians only. The population occurrence of M112 (see Table 1) allows estimation of relative divergence of Hadzabe and Ju|'hoansi, as this mutation occurs in high frequency only within forest peoples (Biaka, Mbuti, Lisongo), Ju|'hoansi San, and Hadzabe. Ju|'hoansi were included to gain insight into Hadzabe and San divergence. Forest peoples were included for context. Branch lengths are proportional to the number of mutations. Branch lengths of M112 STR diversity suggest a relatively great age for this mutation, concordant with published results based on a different set of STRs in [23]. The area of each pie chart is proportional to the observed number of individuals. H, Hadzabe; D, Datoga; I, Iraqw; S, Sukuma; J, Ju|'hoansi San; B, Biaka; M, Mbuti; L, Lisongo.

geographic distributions of the relevant characteristic mutations. M2 is frequent in most African populations, with the exception of Afro-Asiatic-speaking groups (Table 1). The highest frequencies of M2 are observed across Bantu-speaking groups. M35 is rare within Bantu speakers and is widely though nonuniformly dispersed throughout Africa (Table 1). M112 has been observed at highest frequencies within San and peoples of the central African forests (i.e., Biaka, Mbuti, and Lisongo; hereafter referred to collectively as forest peoples). Among Khwe and San, M112 is particularly frequent within Ju|'hoansi, a San population known to have experienced relatively little gene flow from Bantu speakers [8, 11]. M112 was observed in two Sukuma NRYs and a single Lisongo NRY. Sukuma are Bantu speakers who coexist with Hadzabe. Lisongo are Bantu speakers who coexist with Biaka.

Y Chromosome STR Variation, Ages of UEP Mutations, and San-Hadzabe Divergence

We analyzed variation at eight NRY STR loci (Figure 3 and Table S2). Two or more Tanzanian populations shared 5 of the 33 resulting haplotypes. One M35 haplotype, for example, was observed in three of four Tanzanian groups (Hadzabe, Iraqw, and Sukuma). Two of four Hadzabe M2 haplotypes were shared with Sukuma. Hadzabe also shared M35 STR haplotypes with neighboring populations. The single Sukuma individual with M112 shared his STR haplotype with a Hadzabe M112 individual.

Observing the high frequency of M112 in the two click-speaking groups, Ju|'hoansi and Hadzabe, we tested variation among all available carriers of this mutation at eight STR loci. This set of M112+ individuals included Ju|'hoansi, forest peoples, Hadzabe, and one Sukuma (Figure 3 and Table S2), revealing a total of 20 STR haplotypes. No M112 STR haplotypes were shared between Ju|'hoansi and Hadzabe. Furthermore, most Ju|'hoansi haplotypes differed by a large number of repeat units from Hadzabe haplotypes. We conducted analysis of molecular variance (AMOVA) [24] based on the eight STR loci. For San, Hadzabe, Biaka, and Mbuti, 43% of variation was among populations.

One goal in analyzing STR variation on a UEP background was to determine whether sharing among populations of a characteristic mutation reflects recent or ancient shared ancestry. STR haplotypes provide indirect information regarding the age of each UEP-generating mutation [25]. To be precise, STR variation carries information regarding the age of the most recent common ancestor (MRCA) of the group of haplotypes. We took both a phylogenetic and a population genetic approach to estimating relative ages of UEPs M2, M35, YAP, and M112. Specifically, we calculated ρ, a phylogenetic estimator of allele age, and we also used a Bayesian population genetic approach (Table 2). In both cases, we considered two sets of mutation rates: those estimated from pedigree data [26] and the evolutionary rate obtained in [27]. The four mutations under consideration differ in relative age. Age estimates indicate that M2 is roughly half as old as M35, while M35 is 1/2 to 2/3 as old as M112. M112 is estimated to be older than YAP.

Age estimates yield an upper bound for the date of common ancestry of populations harboring a particular UEP. The T_D estimator [28] yields a more accurate estimate of the length of time since two populations diverged. Although STR haplotypes without context of UEPs would yield an estimate, the analysis of STRs within a UEP-defined haplogroup reduces interclade homoplasy [29, 30]. M2, while observed within both the Hadzabe and San, has been observed within San primarily in Sekele (also referred to as Vassekele or Omega) originally from Angola. That population has associated extensively with Ovambo Bantu speakers for at least several hundred years [6]. Hadzabe have extensive contact with Sukuma Bantu speakers. The presence of M2 in both of these click-speaking groups likely represents relatively recent gene flow from respective neighboring populations, given the high frequency of M2 among Bantu speakers across Africa (Table 1) and nongenetic evidence for extensive contact. M35, which is older than M2, is too rare within Ju|'hoansi to allow for estimation of time of common ancestry. M112, observed at high

Table 2. Relative Ages of Mutations Frequent in Tanzanian Populations Estimated from Associated STR Variation

Mutation	n	ρ	SE	Phylogenetic Estimates		Posterior Probabilities[a]	
				T_{lower}	T_{upper}	T_{lower}	T_{upper}
M2	27	1.41	0.23	7.6	19.0	7.0 (1.9, 27.7)	67.4 (19.3, 244.3)
M35	12	3.33	0.52	17.9	45.0	10.4 (2.5, 47.6)	78.8 (22.4, 273.9)
YAP	45	3.96	0.30	21.3	53.4	17.2 (5.1, 63.2)	88.8 (24.7, 401.5)
M112 (Tz)[b]	14	5.57	0.63	30.0	75.2	15.4 (3.4, 71.5)	125.2 (34.2, 409.5)
M112	29	6.79	0.48	36.5	91.6	18.7 (5.3, 73.6)	95.8 (29.5, 315.5)

Estimates are of time, T, (in 1000 years) to the most recent common ancestor of representatives of haplogroups and are therefore the minimum estimates of the age of the allele. Phylogenetic estimates were generated according to [46]. Posterior probabilities were generated according to [48]. Upper estimates were calculated by using mutation rate(s) estimated in [26]. Lower estimates were calculated by using mutation rate(s) estimated in [27] from pedigree data.

[a] Median (and 95% equal-tailed intervals).

[b] Estimate included Tanzanian individuals only.

frequency in both Hadzabe and Ju|'hoansi San, thereby lends itself to relative date estimation with the goal of inferring the history of ancient San and Hadzabe ancestors without the influence of the relatively recent Bantu expansion. This approach to studying STR variation of selected UEP clades has been applied previously and has been used, for example, to date the Bantu expansion [31].

Because of uncertainty regarding mutation rates and generation time, we report ratios of T_D estimates to the upper bound, taken as 1.00, for separation of San and other populations (forest peoples and Hadzabe) rather than their absolute values (dates). The lower bound ratio for their divergence, 0.75 ± 0.20, was estimated by using variation in forest peoples as a reference to obtain V_0, with a correction of 0.661 [28]. The upper bound for separation of San and Hadzabe was found to be 1.09 ± 0.22. Upper and lower bounds for ratios of separation of ancestors of forest peoples and Hadzabe were found to be 0.89 ± 0.32 and 0.65 ± 0.31, respectively. The separation ratio for non-Africans and forest peoples, based on eight NRY STRs (data for calculation were from [32]), is bounded by 0.81 ± 0.33 and 0.72 ± 0.30 (assuming the same reference for V_0). The separation ratio estimate of 0.51 ± 0.20 derived from 131 autosomal STR loci [28] may be regarded as an alternative lower bound for separation of non-Africans and forest peoples.

Consistent with mtDNA analyses, NRY analyses reveal that the Hadzabe population is primarily represented in haplogroups B2 and E3 (phylogenetically equivalent to L2 and L3* mtDNA haplogroups), includes both haplotypes unique to Hadzabe and haplotypes shared with neighboring populations, and is genetically very distant from the Ju|'hoansi population.

Discussion

Genetic Variation in Tanzanian Populations

We observed extensive mtDNA and NRY diversity within the set of four Tanzanian linguistic groups. Only one individual with the basal NRY haplogroup A was observed, probably reflecting the small number of male Iraqw tested. A linguistically diverse set of Tanzanians fall into the basal mtDNA haplogroup L1a, and L1f haplotypes were observed in Iraqw. Those haplotypes are distinct, however, from those of Ju|'hoansi L1d and L1k lineages (Figure 1), and the differences have been previously characterized [11, 20]. For both mtDNA and NRY, the second most basal haplogroups (mtDNA L2 and NRY B) were observed most frequently within Hadzabe. A total of 35% of Hadzabe mtDNAs were L2, and 52% of Hadzabe NRYs were B. Most other Hadzabe fell within more derived mtDNA and NRY superclades, which occur both within and outside of Africa. A total of 59% of Hadzabe were mtDNA haplogroup L3*, and 48% of Hadzabe NRYs carried M168 (see Figures 1 and S1). Therefore, the phylogenetic positions of the largest proportions of Hadzabe lineages (in the global tree topology) are remarkably equivalent for mtDNA and NRY.

As indicated in Figures 1 and 3, there is clear mtDNA evidence for limited gene flow among the different Tanzanian linguistic groups. Populations represented in low frequency within individual haplogroups are consistent with a hypothesis of recent gene flow (e.g., Hadzabe in L1a group 4, Figure 1). NRY data are also consistent with recent gene flow. The observation of a single Sukuma individual with M112 and sharing of NRY STR-defined haplotypes by Hadzabe and Sukuma is expected given the association of the two populations [2]. The relatively high frequency of M2 (30%) in Hadzabe probably reflects substantial gene flow from Sukuma, given the high frequency of that mutation within Bantu speakers throughout sub-Saharan Africa (Table 1). Sharing of STR-defined haplotypes within the E3b haplogroup suggests that there may have been gene flow between the Hadzabe and other populations as well.

Along with evidence for recent gene flow, there is evidence for earlier isolation of Hadzabe. The Hadzabe, with high frequencies of mtDNA haplogroups L2 and group 10 (Figure 1) of L3* and NRY haplogroup B2b, are distinct from their neighbors. Their relatively small population size likely explains lower mtDNA nucleotide diversity. Yet, Hadzabe have maintained genetic distinctiveness for both paternally and maternally derived genetic lineages and are clearly differentiated from other eastern Africans (Figures 1–3).

Southern African Click Speakers

Extensive linguistic and genetic diversity among the southern African click speakers remains largely unchar-

acterized. Appellations in the genetic anthropological literature have, at times, obscured heterogeneity of "Khoisan" populations. "Khoisan" groups include both those who self-identify as "Khwe" and those who self-identify as "San." The term "San" refers to a set of groups of traditional hunter-gatherers with linguistic diversity scattered from Angola to the Cape. Khwe are a broadly distributed, culturally defined group, traditionally subsisting as pastoralists. Clarification of precise ethnic or linguistic identity of populations, their possibly diverse origins, and their recent or prehistorical associations with other populations is necessary for accurate reconstruction of phylogenetic relationships, inference of migrations, and estimation of dates.

In selecting southern African click speakers for genetic comparisons with Hadzabe, we took into consideration nongenetic information concerning history and identity. While we recognize that gene flow from Bantu speakers is part of the history of different Khwe and San peoples, in seeking information about the ancient origins of these groups, we aimed to focus on groups that were least impacted by gene flow within the last 2000–3000 years. Ju|'hoansi, a group of San living in the vicinity of the Namibian/Botswanan border, have had little contact with Bantu-speaking populations, intermarriage has been infrequent, and when such intermarriage occurred, the children tended to identify with Bantu speakers [8]. On the other hand, Khwe and Sekele San populations have had extensive associations with Bantu speakers; in the former case, this occured for at least 1000 years, and in the latter case, this occurred for at least several hundred years [6]. These observations are consistent with previous findings that gene lineages of Ju|'hoansi appear to have distinct haplotypes and early divergence from other human populations [10, 11]. DNA variation of Khwe and Sekele (e.g., [23, 33, 34]), however, reveals many haplotypes that cluster with those of Bantu-speaking populations. The 39 "Khoisan" individuals studied in [22] represented both Sekele (20 individuals) and Ju|'hoansi (19 individuals) San. The 64 !Kung individuals studied in [23] included only Sekele San. In [23], M112 was observed in only 8 of 64 Sekele San and 0 of 26 Khwe (both highly admixed with Bantu speakers). In [22], M112 was observed in 28% of the "Khoisan" sample, reflecting the Ju|'hoansi component that has less admixture. In fact, within the "Khoisan" sample in [22], the Sekele San account for 6 of the 7 M2 NRYs (in high frequency among Bantu speakers) and only 3 of the 11 M112 NRYs. Given the evidence for recent gene flow from Bantu speakers to the Sekele, we focused our study on the Ju|'hoansi.

Ages of Y Chromosome UEP Mutations

While ascertainment of mutations on the NRY has been inconsistent, phylogenetic position [22] indicates the relative ages of UEPs considered herein [9]. A recent analysis of a different set of STRs on M2, M35, and M112 backgrounds provides additional insight into the relative ages of these mutations [23]. The number of mutations separating M112 haplotypes (Figure 4B in [23]) was much higher than that for both M35 (Figure 4E in [23]) and M2 (Figure 4D in [23]). Note that in [23] the M112 network (their Figure 4B) is reduced in size by 1/2 relative to the other haplogroups. Those data suggest that M112 is the oldest mutation, M35 is of intermediate age, and M2 is the youngest. There is a clear nonuniform geographic distribution of STR haplotypes for M35 and M112, with little sharing across broad geographic regions within Africa [23].

Table 2 presents relative age estimates for the mutations based on the independent STR data of this paper. While mutation rate influences absolute dates, relative ages are generally consistent across the two estimation methods. M112 was the oldest of the mutations considered, including YAP, even when only Tanzanian M112 haplotypes were considered. The relative age estimates and the deep position in the UEP tree [22] suggest that M112 arose early in the history of modern humans, prior to out-of-Africa expansions. If the mutation is indeed that old, although a derived state among hominoids, it may have been present in the population that gave rise to extant humans. In that case, sharing of M112 would not provide evidence for recent shared ancestry of San, Hadzabe, and forest peoples. Indeed, the geographic subdivision and divergence within the set of M112 STR haplotypes is apparent from both [23] and this study. The AMOVA estimate of 43% variation among the San, Hadzabe, and forest populations also suggests a high degree of differentiation and is consistent with early population divergence.

M112 has been observed only very rarely outside of Khwe and San, forest, and Hadzabe populations. Two exceptions considered here likely reflect recent gene flow from foraging Hadzabe and Biaka to neighboring agricultural peoples.

Timing of Shared Ancestry of Hadzabe and San

Phylogenetic relationships among African mitochondrial lineages illustrate the time depth of the separation of Hadzabe and San. All 24 Ju|'hoansi haplotypes diverge from all 49 Hadzabe haplotypes at the root (the most ancient split) among known extant human mtDNA lineages (the root has been established by outgroup comparison to Neandertal and chimpanzee). A single Hadzabe has a L1a haplotype (possibly acquired through recent gene flow, see Figure 1). L1a is not observed among Ju|'hoansi and splits from Ju|'hoansi L1d and L1k lineages at the root [20]. All other Hadzabe mtDNAs are in the L2 + L3* superclade defined by 247G and 16230A (plus 16390A or 3594C), also divergent from Ju|'hoansi L1d and L1k at the root [20]. The observed population frequency difference across the root of 100% versus 0% is highly significant. Considering haplotype diversity within Ju|'hoansi L1d [11] and Hadzabe L2 + L3* (Figure 1), the distribution across the root is unlikely to be an artifact of sampling. As L2 and L3* were not observed in Ju|'hoansi but are present in many other populations, the mtDNA data provide one independent body of compelling evidence that the Hadzabe are more closely related to other populations than are the Ju|'hoansi. The separation of the ancestors of click-speaking Hadzabe of Tanzania and click-speaking San of Botswana and Namibia appears to be among the earliest of human population divergences.

As indicated in Table 1, the sample of 39 individuals identified as Ju|'hoansi and Sekele San include NRYs in haplogroups A (M91+), B2b (M112+), E3a (M2+), and E3b (M35+). Hadzabe overlap with individuals in haplogroups B2b, E3a, and E3b. As discussed above, M2 is relatively young, is frequent in Bantu speakers across Africa, and so likely represents recent gene flow from neighbors. M112 likely reflects an ancient connection between San and Hadzabe that possibly dates as far back as the common ancestor of extant humans. Noting low frequencies of M35 in both the San sample in [22] (four individuals) and Hadzabe (three individuals), we did not examine M35 STR diversity. We compared STR diversity of all available M112+ individuals. No M112 STR haplotypes were shared between Ju|'hoansi and Hadzabe. The observation that Ju|'hoansi haplotypes differ by a relatively large number of repeat units from Hadzabe haplotypes is consistent with ancient separation of these populations.

Time of divergence (T_D) [28] estimates for M112 NRYs also suggest great antiquity for separation of Ju|'hoansi from other populations, including Hadzabe. With V_0 set to zero, the estimate of 112,200 $\pm$ 41,800 years serves as an upper bound for the time of separation of ancestors of Ju|'hoansi from other populations, based on a mutation rate of 0.00026/20 years per locus [27]. Mutation rate at NRY loci is a controversial issue. No matter which mutation rate (or generation time) is used, relative T_D estimates are concordant with STR network branch lengths (Figure 3 and [23]) and mtDNA estimates [11]. This finding indicates that gene lineages that make up a major component of the Ju|'hoansi mtDNA and NRY gene pool diverged early from other studied human gene lineages.

History of Click Languages

The two independently inherited DNA segments each reveal variation that provides evidence that San and Hadzabe are among the most highly divergent of African (and therefore global) population pairs. Considered without population genetic and linguistic context, such divergence might be consistent with a number of scenarios, including separate, independent invention of clicks by ancestors of San and Hadzabe; gene replacement without language replacement; borrowing of clicks by one group from the other; or independent retention of clicks since early in human prehistory.

Two lines of evidence, rarity of clicks in human languages and complexity of the shared repertoire of clicks and accompaniments, suggest that independent invention of clicks in San and Hadzabe populations is an unlikely explanation for the observed genetic pattern. With regards to complexity of click repertoires, each click language includes a particular set of clicks and accompaniments. Some languages include larger sets than others do, but these sets do overlap. The clicks integral to Hadzane largely overlap with those clicks integral to Khwe and San languages. The hypothesis of independent invention, as it applies to the languages of the Hadzabe and San, lacks linguistic support.

Another a priori explanation for genetic differentiation within the context of linguistic overlap is gene replacement without language replacement. That is, Hadzabe and San might have diverged recently (hence sharing of clicks), but the genes of one group might have been replaced through gene flow from non-click-speaking neighbors. With the exception of clicks, however, Hadzane is only very distantly related to San languages [15]. Furthermore, as both San and Hadzabe have large unique components to their maternal and paternal gene pools, no potential source of gene replacement is known.

A third a priori explanation of sharing of clicks by San and Hadzabe in the context of genetic differentiation is linguistic borrowing. Xhosa, for instance, while uncontestedly a Bantu language, incorporates some clicks borrowed from Khwe or San languages. The extensive population contact required for such click borrowing, however, leaves a genetic signature through gene flow, as has been well documented [16, 35, 36]. The minimal genetic similarity between San and Hadzabe consists of sharing the NRY M2 mutation. Data herein and elsewhere strongly suggest that M2 has been introduced into click-speaking groups by non-click-speaking neighbors. In addition, gene flow leads to short, central branches for admixed populations, contrary to Ju|'hoansi and Hadzabe differentiation. Finally, distortions of the tongue required to produce click consonants [37, 38] inhibit borrowing of the full repertoire of clicks by adult nonnative speakers. The Nguni language, for instance, includes a click system that is far less deeply integrated and complex than the systems of Hadzabe and San languages [39]. Deep mtDNA and NRY divergence between San and Hadzabe is contrary to expectations under a scenario of borrowing of clicks by Hadzabe from San. Current genetic and nongenetic data are inconsistent with three of four a priori explanations for sharing of clicks without genetic similarity.

Conclusions

The remaining explanation involves independent retention of clicks, possibly for tens of thousands of years, in separate populations leading to present day San and Hadzabe. Indeed, the molecular data are consistent with the most recent shared ancestry of these two populations coinciding with the earliest divergence among extant human populations. Although confirmation or refutation of this suggestion awaits further interdisciplinary investigation, that scenario has implications for our understanding of the history of click languages. If, in fact, San-Hadzabe separation dates back to a time prior to out-of-Africa expansions of modern humans, clicks may be more than 40,000 years old. Under that scenario, clicks would have been lost subsequently in most other populations.

While estimates of dates are too imprecise to draw a robust conclusion regarding the use of clicks by our earliest common ancestors, the estimates and tree topology do imply a depth of tens of thousands of years. With that conclusion in mind, we consider likely causes of the present widely disjunct geographic distribution of southern African and eastern African populations of click speakers. Genetic and archaeological data have been interpreted as possible evidence for an ancient

San presence in eastern Africa [16, 35, 36], yet scrutiny of some of that evidence suggests long-term differentiation between eastern and southern Africa [40], as does the differentiation of L1i lineages [11]. Although the archaeological history of Africa is complex, the expansion of Bantu-speaking agriculturalists from West Africa about 2,500 years ago has often been viewed as having caused a split separating a somewhat homogeneous population of click-speaking peoples [16, 35, 36]. Indeed, the Bantu expansion has left a signature in the genes of most, if not all, click-speaking populations [16, 35, 36]. Another possibility, as follows, deserves consideration given the data presented here. Perhaps an early population of modern humans, speaking a click language, increased in number, dispersed, and came to occupy most of southern and eastern Africa. Via geographic isolation, these peoples came to form small regional populations across millions of square kilometers, until a time was reached when gene flow essentially ceased between many populations. Such isolation among groups is especially plausible given that population sizes in Africa appear to have been reduced between about 40,000 and 20,000 years ago [41]. Under this scenario, eastern African and southern African click speakers had already been isolated from one another for tens of thousands of years by the time Bantu speakers entered their range.

So far, we have discussed clicks as if assuming their cultural neutrality. We cannot rule out the possibility, however, that clicks may have persisted because they confer, in particular environments, an advantage. Click systems may impact hunting success. During stalking of prey, Ju|'hoansi revert to a hushed whisper-like communication. Speech is devoiced and consists almost entirely of clicks. Such behavior has been documented on film, as, for example, in hunting scenes in the films of the Marshalls. Devoicing has often been observed during stalking (J. Marshall, personal communication). Click density of Ju|'hoan allows devoiced communication. While there is little precedence for phonetic elements conferring a functional advantage, we hesitate to rule out this possibility without further study.

Our estimate suggests that a substantial portion of the current Hadzabe gene pool reflects the ancestral Hadzabe population, and gene flow within the last few generations from neighbors has contributed another substantial portion. In the broader African and global context, the Hadzabe are more closely related to other populations than to San. The deep genetic divergence between the click-speaking groups is consistent with the hypothesis that clicks are an ancient element of human language.

Experimental Procedures

DNA Samples

The Commission for Science and Technology (COSTECH) of Tanzania granted permission for sample collection. Samples from Tanzania, including Hadzabe, Sukuma, Iraqw, and Datoga, were collected by using buccal swabs (Epicentre). The Stanford University Institutional Review Board and the National Institute for Medical Research of Tanzania approved the protocol. Informed consent was obtained from all donors. Note that there is a slight possibility of overlap in our 49 Hadzabe samples and the 17 in [10]. For NRY STR analysis, Ju|'hoansi San, Biaka, Lisongo, and Mbuti samples were obtained from the archival collection of L. Luca Cavalli-Sforza.

Genotyping and DNA Sequencing

Sequence, STR, and most SNP data were obtained by fluorescent sequencing of PCR products (primers are given in Table S3 and [9]). NRY biallelic markers M91, M94, M60, M150, M112, M168, YAP, M2, and M35 were typed by fragment analysis or sequencing [9] or were inferred by hierarchy. In the case of buccal cell samples, nested PCR was used. Single amplifications were used for archival DNAs. DNA cycle sequencing was performed by using BigDye chemistry and was detected on a 310 Genetic Analyzer (Applied Biosystems). For HV1 and HV2 mtDNA, all samples were amplified at a 55°C annealing temperature with 2.5 mM $MgCl_2$. Nested PCR was performed by adding 5 μl initial PCR product to 250 μl water and then using that solution as template in a nested PCR at a 58°C annealing temperature. NRY regions were amplified at 58°C with 2.5 mM $MgCl_2$. For mtDNA and NRY sequencing, PCR primers (Table S3) were used. An internal primer was used for the HV1 complementary strand (Table S3).

Statistical Analyses

We analyzed both HV1 and HV2 sequences for 49 Hadzabe, 18 Datoga, 12 Iraqw, and 21 Sukuma and compared those sequences with 24 Ju|'hoansi [10]. We analyzed HV1 sequences for 49 Hadzabe individuals in the context of comparable published data for 24 Ju|'hoansi, original data for 12 Iraqw, 19 Datoga, and 32 Sukuma individuals, and published data for representatives of 17 other African populations (Figure 2). We conducted neighbor-joining and parsimony analyses [42] of individual sequences. Nucleotide diversity, HV1 genetic distances based on mutation and drift, and F_{ST} were estimated with Arlequin [43]. Genetic distances among populations were summarized with neighbor joining [44]. Pairwise HV1 genetic distances are given in Table S1.

Sample sizes for NRY STR analysis are provided in Table S2. AMOVA of NRY STRs was performed with Arlequin [43]. To estimate the relative time (as ratios) of the divergence of populations, we used the estimator T_D [28] and compared NRY STR variation in two groups of individuals that diverged from a common ancestor evolving via both mutation and genetic drift. T_D is affected by assumptions of mutation rate and generation time. T_D does not assume mutation drift equilibrium, is robust to population dynamics and gene flow between diverging populations, and requires knowledge of variance in the number of repeats, V_0, at the beginning of population separation. Variance in the number of repeats at the time of origin of any UEP is zero. Therefore, T_D, when calculated with zero V_0, provides a theoretical, though unlikely, upper bound for population separation. When calculated with V_0 set to a value that exceeds the variance prior to separation, T_D provides a lower bound for the time of separation. We used this approach to estimate the upper and lower bounds for time of San-Hadzabe population separation, relative to other population divergences.

For each UEP-defined NRY haplogroup, we conducted phylogenetic analysis of variation at the eight STR loci according to the Median Joining (MJ) network algorithm [45] by using the Network 3.1.1.1 program (www.fluxus-engineering.com). ϵ was set to 0, generating networks closest to maximum parsimony trees. Differential and equal weighting of the eight STR loci, by using weights calibrated according to [27] with some modifications, yielded identical networks.

Relative ages of NRY SNP mutations were estimated by using both a phylogenetic (via the ρ statistic) and a population genetic (via a Bayesian-based coalescence analysis) approach. In both cases, an estimate refers to the MRCA. ρ compares a set of selected haplotypes to an ancestral node, as measured in single differences [46]. The root for each haplogroup was inferred by incorporating the sister haplogroup into the MJ analysis. The sampling error of ρ was approximated by $\sqrt{\rho/n}$, where n denotes sample size [47]. Absolute time estimates were obtained by multiplying ρ by both pedigree [26] and evolutionary [27] estimates of mutation rates specific to STRs used in this study.

The Bayesian approach to estimates of relative ages of NRY UEP mutations was carried out by using BATWING [48]. BATWING incor-

porates a Markov-chain Monte Carlo algorithm, deriving posterior distributions for all parameters of a specified model. The model considered here incorporates the possibility of exponential population growth after a period of constant size, N. We specified a gamma(2,400) distribution as the prior for the growth rate and a gamma(5,1) distribution as the prior for the log of the ratio of the current and original population sizes [48]. We considered two sets of priors for the mutation rate that were based upon the pedigree-based [26] and evolutionary [27] estimates of mutation rates at STR loci. The prior for the initial population size was a gamma distribution with a mean of 2,000 and a mode of 1,900. Priors for N with higher means also resulted in posterior estimates of N that were less than 2,000 (details not given). A total of 10,000 initial rearrangements were discarded, and posterior distributions were estimated from the subsequent 50,000 rearrangements. The median and equal-tailed 95% interval limits were calculated for each parameter of interest. The estimated coalescence time, T, is measured in terms of N X generation time; the postdata values of N and T along with a generation time of 25 years were used to generate absolute coalescence times.

Supplemental Data

Supplemental Data including genetic distances between African populations estimated from HV1 mtDNA sequence data, Y chromosome STR haplotypes, within UEP-defined haplogroups, and PCR and sequencing primers used in this study are available at http://images.cellpress.com/supmat/supmatin.htm. Tanzanian mtDNA control region HV1 and HV2 sequences are available at http://www.stanford.edu/~aknight/Click/.

Acknowledgments

This study was supported by a L.S.B. Leakey Foundation grant, National Science Foundation grant BCS9905574, and National Institutes of Health (NIH) grant GM28428 to J.L.M.; NIH grant TW05540 to M.W. Feldman; and NIH grant GM55273 to L.L. Cavalli-Sforza. We thank A. Mabulla, C. Magori, P. Lufungulo, C. O'Halloran, F. Marlowe, G. Gidasaid, R. Matiya, and E. Lyimo for research assistance in Tanzania and 130 Tanzanians who provided buccal cell samples. We thank L.L. Cavalli-Sforza for samples, helpful comments, and for initially suggesting study of the Hadzabe. We thank N. Blurton-Jones, D. Bygott, N. Crawhall, R. Edwards, J. Hanby, R.G. Klein, G. Passarino, A.B. Smith, and A. Traill for helpful discussion. We thank M. Bramlage, M. Jobin, M. Keli'ihanapule, T. Kivisild, and A. Miller for technical assistance. We thank three anonymous reviewers for helpful comments.

Received: July 10, 2002
Revised: December 6, 2002
Accepted: January 13, 2003
Published: March 18, 2003

References

1. Greenberg, J. (1963). The Languages of Africa (Bloomington: Indiana University Press).
2. Blurton Jones, N.G., Smith, L.C., O'Connell, J.F., Hawkes, K., and Kamuzora, C.L. (1992). Demography of the Hadza, an increasing and high density population of savanna foragers. Am. J. Phys. Anthropol. *89*, 159–181.
3. Ambrose, S.H. (1982). Archaeology and linguistic reconstructions of history in Eastern Africa. In The Archaeological and Linguistic Reconstruction of African History, C. Ehret and M. Posnansky, eds. (Berkeley: University of California Press), pp. 104–157.
4. Güldemann, T., and Vossen, R. (2000). Khoisan. In African Languages. B. Heine and D. Nurse, eds. (Cambridge: Cambridge University Press), pp. 99–122.
5. Traill, A., and Nakagawa, H. (2000). A historical !Xóõ-|Gui contact zone: linguistic and other relations. In The State of Khoesan Languages in Botswana. H.M. Batibo and J. Tsonope, eds. (Tasalls: Mogoditshane), pp. 1–17.
6. Barnard, A. (1992). Hunters and Herders of Southern Africa (Cambridge: Cambridge University Press).
7. Excoffier, L., Pellegrini, B., Sanchez-Mazas, A., Simon, C., and Langaney, A. (1987). Genetics and history of sub-Saharan Africa. Yearbook Phys. Anthropol. *30*, 151–194.
8. Lee, R.B. (1993). The Dobe Ju|'hoansi. (Orlando: Harcourt Brace Publishers).
9. Underhill, P.A., Passarino, G., Lin, A.A., Shen, P., Lahr, M.M., Foley, R.A., Oefner, P.J., and Cavalli-Sforza, L.L. (2001). The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. Ann. Hum. Genet. *65*, 43–62.
10. Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K., and Wilson, A.C. (1991). African populations and the evolution of human mitochondrial DNA. Science *253*, 1503–1507.
11. Watson, E., Forster, P., Richards, M., and Bandelt, H.-J. (1997). Mitochondrial footprints of human expansions in Africa. Am. J. Hum. Genet. *61*, 691–704.
12. Poloni, E., Semino, O., Passarino, G., Santachiara-Benerecetti, A., Dupanloup, I., Langaney, A., and Excoffier, L. (1997). Human genetic affinities for Y-chromosome P49a,f/TaqI haplotypes show strong correspondence with linguistics. Am. J. Hum. Genet. *61*, 1015–1035.
13. Cavalli-Sforza, L.L., Minch, E., and Mountain, J.L. (1992). Coevolution of genes and languages revisited. Proc. Natl. Acad. Sci. USA *89*, 5620–5624.
14. Scozzari, R., Cruciani, F., Santolamazza, P., Malaspina, P., Torroni, A., Sellitto, D., Arredi, B., Destro-Bisol, G., De Stefano, G., Rickards, O., et al. (1999). Combined use of biallelic and microsatellite Y-chromosome polymorphisms to infer affinities among African populations. Am. J. Hum. Genet. *65*, 829–846.
15. Sands, B. (1998). The linguistic relationship between Hadza and Khoisan. In Language, Identity, and Conceptualization among the Khoisan, M. Schladt, ed. (Cologne: Quellen zur Khoisan-Forschung 15. Köppe), pp. 265–283.
16. Cavalli-Sforza, L.L., Menozzi, P., and Piazza, A. (1994). The History and Geography of Human Genes (Princeton: Princeton University Press).
17. Krings, M., Stone, A., Schmitz, R.W., Krainitzki, H., Stoneking, M., and Pääbo, S. (1997). Neandertal DNA sequences and the origin of modern humans. Cell *90*, 19–30.
18. Krings, M., Geisert, H., Schmitz, R.W., Krainitzki, H., and Pääbo, S. (1999). DNA sequence of the mitochondrial hypervariable region II from the Neandertal type specimen. Proc. Natl. Acad. Sci. USA *96*, 5581–5585.
19. Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., et al. (1981). Sequence and organization of the human mitochondrial genome. Nature *290*, 457–465.
20. Salas, A., Richards, M., De la Fe, T., Laeu, M.-V., Sobrino, B., Sánchez-Diz, P., Macaulay, V., and Carracedo, Á. (2002). The making of the African mtDNA landscape. Am. J. Hum. Genet. *71*, 1082–1111.
21. Y-Chromosome Consortium. (2002). A nomenclature system for the tree of human Y-chromosomal binary haplogroups. Genome Res. *12*, 339–348.
22. Underhill, P.A., Shen, P., Lin, A.A., Jin, L., Passarino, G., Yang, W.H., Kauffman, E., Bonné-Tamir, B., Bertranpetit, J., Francalacci, P., et al. (2000). Y-chromosome sequence variation and the history of human populations. Nature Genet. *26*, 358–361.
23. Cruciani, F., Santolamazza, P., Peidong, S., Macaulay, V., Moral, P., Olckers, A., Modiano, D., Holmes, S., Destro-Bisol, G., Coia, V., et al. (2002). A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes. Am. J. Hum. Genet. *70*, 1197–1214.
24. Excoffier, L., Smouse, P., and Quattro, J. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes. Genetics *131*, 479–491.
25. Rannala, B., and Bertorelle, G. (2001). Using linked markers to infer the age of a mutation. Hum. Mutat. *18*, 87–100.
26. Kayser, M., Roewer, L., Hedman, M., Henke, L., Henke, J., Brauer, S., Kruger, C., Krawczak, M., Nagy, M., Dobosz, T., et al. (2000). Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as re-

vealed by direct observation in father/son pairs. Am. J. Hum. Genet. *66*, 1580–1588.
27. Forster, P., Röhl, A., Lünnemann, P., Brinkmann, C., Zerjal, T., Tyler-Smith, C., and Brinkmann, B. (2000). A short tandem repeat-based phylogeny for the human Y-chromosome. Am. J. Hum. Genet. *67*, 182–196.
28. Zhivotovsky, L.A. (2001). Estimating divergence time with the use of microsatellite genetic distances: impacts of population growth and gene flow. Mol. Biol. Evol. *18*, 700–709.
29. de Knijff, P. (2000). Messages through bottlenecks: on the combined use of slow and fast evolving polymorphic markers on the human Y chromosome. Am. J. Hum. Genet. *67*, 1055–1061.
30. Bosch, E., Calafell, F., Santos, F.R., Pérez-Lezaun, A., Comas, D., Benchemsi, N., Tyler-Smith, C., and Bertranpetit, J. (1999). Variation in short tandem repeats is deeply structured by genetic background on the human Y chromosome. Am. J. Hum. Genet. *65*, 1623–1638.
31. Thomas, M.G., Tudor, P., Weiss, D.A., Skorecki, K., Wilson, J.F., le Roux, M., Bradman, N., and Goldstein, D.B. (2000). Y chromosomes travelling south: the Cohen modal haplotype and the origins of the Lemba—the "Black Jews of southern Africa". Am. J. Hum. Genet. *66*, 674–686.
32. Seielstad, M., Bekele, E., Ibrahim, M., Toure, A., and Traore, M. (1999). A view of modern human origins from Y-chromosome microsatellite variation. Genome Res. *9*, 558–567.
33. Soodyall, H., and Jenkins, T. (1992). Mitochondrial DNA polymorphisms in Khoisan populations from Southern Africa. Ann. Hum. Genet. *56*, 315–324.
34. Chen, Y.-S., Olkers, A., Schurr, T.G., Kogelnik, A.M., Huoponen, K., and Wallace, D.C. (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.
35. Nurse, G.T., Weiner, J.S., and Jenkins, T. (1985). The Peoples of Southern Africa and Their Affinities (Oxford: Clarendon Press).
36. Cavalli-Sforza, L.L. (2000). Genes, Peoples, and Languages (Berkeley: University of California Press).
37. Ladefoged, P., and Maddieson, I. (1995). The Sounds of the World's Languages (Oxford: Blackwell).
38. Traill, A., and Vossen, R.J. (1997). Sound changes in the Khoisan languages: new data on click loss and click replacement. JALL *18*, 21–56.
39. Lass, R. (1997). Historical Linguistics and Language Change (Cambridge: Cambridge University Press).
40. Morris, A.G. (2003). Isolation and the origin of the Khoisan: late Pleistocene and early Holocene human evolution at the southern end of Africa. Hum. Evol., *17*, 105–114, in press.
41. Klein, R.G. (1992). The archaeology of modern human origins. Evol. Anthropol. *1*, 5–14.
42. Swofford, D.L. 1998. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods), Version 4. (Sunderland, Massachusetts: Sinauer Associates).
43. Excoffier, L. (2001). Arlequin: A Software for Population Genetic Analysis (available at http://lgb.unige.ch/arlequin) (Switzerland: University of Geneva).
44. Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. *4*, 406–425.
45. Bandelt, H.-J., Forster, P., and Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. *16*, 37–48.
46. Forster, P., Harding, R., Torroni, A., and Bandelt, B. (1996). Origin and evolution of Native American mtDNA variation: a reappraisal. Am. J. Hum. Genet. *59*, 935–945.
47. Torroni, A., Bandelt, H., D'Urbano, L., Lahermo, P., Moral, P., Sellitto, D., Rengo, C., Forster, P., Savontaus, M., Bonne-Tamir, B., et al. (1998). mtDNA anaylsis reveals a major Late Paleolithic population expansion from southwestern to northeastern Europe. Am. J. Hum. Genet. *62*, 1137–1152.
48. Wilson, I., Weale, M., and Balding, D. (2000). BATWING: Bayesian Analysis of Trees with Internal Node Generation (http://www.maths.abdn.ac.uk/~ijw) (Aberdeen, UK: Department of Mathematical Sciences, University of Aberdeen).
49. Graven, L., Passarino, G., Semino, O., Boursot, P., Santachiara-Benerecetti, S., Langaney, A., and Excoffier, L. (1995). Evolutionary correlation between control region sequence and restriction polymorphisms in the mitochondrial genome of a large Senegalese Mandenka sample. Mol. Biol. Evol. *12*, 334–345.
50. Pinto, F., Gonzalez, A.M., Hernandez, M., Larruga, J.M., and Cabrera, V.M. (1996). Genetic relationship between the Canary Islanders and their African and Spanish ancestors inferred from mitochondrial DNA sequences. Ann. Hum. Genet. *60*, 321–330.
51. Krings, M., Salem, A.H., Bauer, K., Geisert, H., Malek, A.K., Chaix, L., Simon, C., Welsby, D., Di Rienzo, A., Utermann, G., et al. (1999). mtDNA analysis of Nile River Valley populations: a corridor or a barrier to migration? Am. J. Hum. Genet. *64*, 1166–1176.
52. Ruhlen, M.A. (1991). Guide to the World's Languages (Stanford: Stanford University Press).

Accession Numbers

MtDNA sequences are deposited in GenBank as accession numbers AY217550–AY217649.