History of Click-Speaking Populations of Africa Inferred from mtDNA and Y Chromosome Genetic Variation

Sarah A. Tishkoff,[1] Mary Katherine Gonder,*[2] Brenna M. Henn,† Holly Mortensen,* Alec Knight,† Christopher Gignoux,† Neil Fernandopulle,† Godfrey Lema,‡ Thomas B. Nyambo,‡ Uma Ramakrishnan,‖ Floyd A. Reed,* and Joanna L. Mountain†[1]*

*Department of Biology, University of Maryland; †Department of Anthropological Sciences, Stanford University; ‡Department of Biochemistry, Muhimbili University College of Health Sciences, Dar es Salaam, Tanzania; and ‖National Centre for Biological Sciences, Bangalore, India

Little is known about the history of click-speaking populations in Africa. Prior genetic studies revealed that the click-speaking Hadza of eastern Africa are as distantly related to click speakers of southern Africa as are most other African populations. The Sandawe, who currently live within 150 km of the Hadza, are the only other population in eastern Africa whose language has been classified as part of the Khoisan language family. Linguists disagree on whether there is any detectable relationship between the Hadza and Sandawe click languages. We characterized both mtDNA and Y chromosome variation of the Sandawe, Hadza, and neighboring Tanzanian populations. New genetic data show that the Sandawe and southern African click speakers share rare mtDNA and Y chromosome haplogroups; however, common ancestry of the 2 populations dates back >35,000 years. These data also indicate that common ancestry of the Hadza and Sandawe populations dates back >15,000 years. These findings suggest that at the time of the spread of agriculture and pastoralism, the click-speaking populations were already isolated from one another and are consistent with relatively deep linguistic divergence among the respective click languages.

Introduction

Comparison of linguistic similarity, geographic proximity, and genetic similarity among populations can provide insights into both human population history and the events and processes underlying language change. Here we examine the genetic diversity of speakers of one set of languages, those with a repertoire of phonemes called "click" consonants. Click languages, spoken only in Africa with the exception of the extinct Damin ritual language of Australia (Hale 1992), are among the richest of all human languages in terms of the number of distinct phonemes (Güldemann and Vossen 2000). Greenberg included all languages with click consonants in the Khoisan (or "Khoe-San") language family. Although they share the element of click consonants, African click languages are highly divergent in other respects, leading some linguists to suggest that the languages do not constitute a single language family (Sands 1998; Güldemann and Vossen 2000) and others to suggest that the Khoisan language family dates back to at least 20,000 years (Ehret 2000). African click languages have been classified into 5 groups: Ju, !Ui-Taa, Khoe, Sandawe, and Hadza (Ehret 2000; Güldemann and Vossen 2000; fig. 1). Linguists often group the Ju, Khoe, and !Ui-Taa languages into a southern African Khoisan (SAK) branch and consider the eastern African Hadza and Sandawe click languages to be more distantly related (Heine and Nurse 2000; fig. 1).

The presence of Khoisan linguistic groups in Tanzania was earlier considered to support a paleobiological-based model, indicating that Khoisan populations inhabited all southern Africa and much of eastern Africa (as far north as Egypt; Tobias 1964; Bräuer 1978). This model of a continuous distribution of Khoisan populations in southern and eastern Africa has been criticized (Stringer et al. 1985; Morris 2002). Instead, the Hadza and Sandawe are thought either to be population isolates or to resemble their Tanzanian neighbors genetically (Cavalli-Sforza et al. 1994).

Scholars suggest that the ancestors of present day Hadza and Sandawe speakers resided in the region now known as Tanzania long before the arrival of neighboring agricultural and pastoral populations (Iliffe 1979; Newman 1995). Herding and cultivating Cushitic (Afro-Asiatic) speakers, originating from Ethiopia, first reached northern Tanzania roughly 4000 years ago, followed by largely pastoral Nilotic (Nilo-Saharan) speakers, originating from southern Sudan (Newman 1995). Agricultural Bantu (Niger-Kordofanian) speakers, originating from West Africa, reached northwestern Tanzania ~2,500 years ago (Newman 1995). According to Iliffe (1979), the linguistically and culturally divergent groups of Tanzania interacted extensively following their arrival in the region, and ethnic labels have been highly fluid, suggesting that there might be little genetic divergence among any of the Tanzanian populations.

Today, the Hadza and Sandawe live only ~150 km apart. The Hadza comprise a relatively small number of individuals (~1,000) living near Lake Eyasi in the Arusha district of north-central Tanzania (Blurton Jones 1992). The Sandawe, living primarily in the Kondoa district southeast of Arusha, are more numerous (~30,000 individuals; Newman 1970). Traditionally, both populations subsisted through hunting and gathering; many Hadza continue to do so, whereas the Sandawe currently subsist via agriculture, which was recently introduced by the neighboring Bantu-speaking Turu (Newman 1970).

Despite the long history of anatomically modern humans in Africa, our knowledge of human population processes within Africa, particularly prior to the spread of agriculture within the past ~5,000 years, remains relatively patchy because of poor preservation of archaeological remains in some areas and because genetic studies of certain

[1] These authors contributed equally to this work.

[2] Department of Biological Sciences, University at Albany, State University of New York, 1400 Washington Ave., Albany, NY 12222.

Key words: mtDNA evolution, Y chromosome evolution, human evolution, genetic variation, African diversity, language evolution.

E-mail: Tishkoff@umd.edu.

Mol. Biol. Evol. 24(10):2180–2195. 2007
doi:10.1093/molbev/msm155
Advance Access publication July 26, 2007

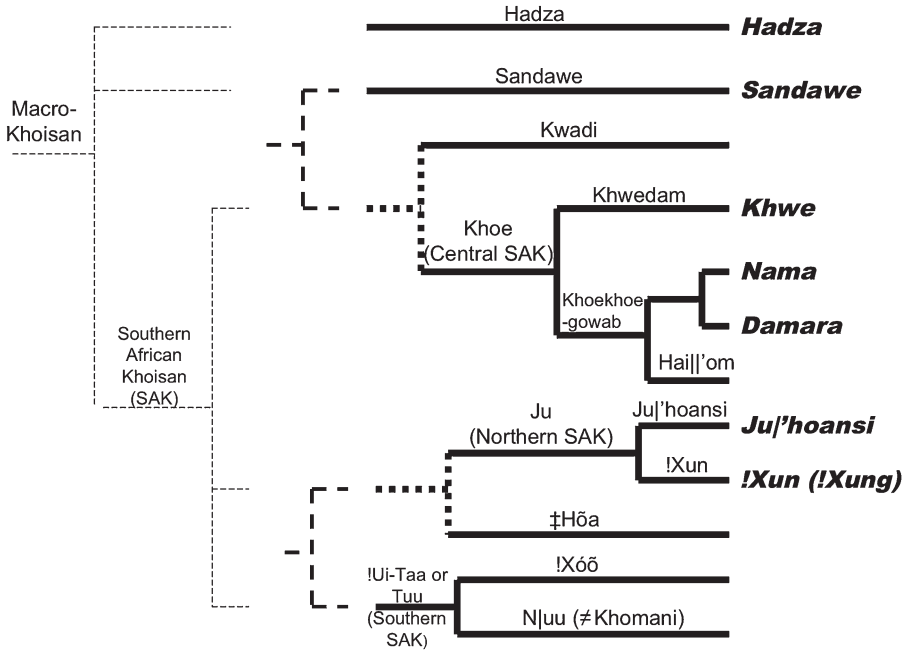


FIG. 1.—Click-language speaking groups considered in this study (italics) and suggested historical relationships among click languages, including several controversial groupings. Solid lines indicate well-accepted linguistic relationships. Thick dashed lines indicate controversial/possible relationships between Kwadi and Khoe (central SAK) languages and between ‡Hõa and Ju (northern SAK) languages (Güldemann and Vossen 2000). Intermediate dashed lines indicate possible relationships between Sandawe and Khoe languages (Güldemann and Vossen 2000) and between Ju and !Ui-Taa (southern SAK) languages. Fine dashed lines indicate suggested deeper links (Greenberg 1963; Wood et al. 2005). Spelling of linguistic terms reflects published and unpublished proposals by several linguists. Inclusion of the Hadza language within Macro Khoisan remains particularly controversial (Sands 1998), as does the relationship among the SAK languages (Güldemann and Vossen 2000).

regions, such as eastern Africa, have been limited (Tishkoff and Williams 2002; Reed and Tishkoff 2006). Although few click-speaking populations have been studied for mitochondrial (mt) DNA (Chen et al. 2000) and Y chromosome (Cruciani et al. 2002; Semino et al. 2002) variation, available data indicate consistently that genetic lineages found among the Ju (including groups called San, !Kung, Zhu|'twasi, or Ju|'hoansi in previous publications) and Khoe (including Khwe, Dama, and Nama) speakers comprise the most basal clades in global phylogenetic trees. Prior studies of mtDNA (Vigilant et al. 1991; Knight et al. 2003) and Y chromosome variation (Knight et al. 2003) in the Hadza and Ju|'hoansi San indicated a deep separation between the 2 groups (>40 kya; Knight et al. 2003). The currently small size of the Hadza population raises the possibility, however, that the population divergence estimate reflects a high level of genetic drift and lineage loss.

The primary focus of the current study is the genetic relationship among individuals from 3 click-speaking groups: the Sandawe, Hadza, and SAK. Several linguists argue for a genealogical relationship between the Sandawe language and some SAK languages (Ambrose 1982; Elderkin 1982; Güldemann and Vossen 2000). The Hadza language is now considered by some linguists to be a linguistic isolate, genealogically unrelated to other click languages (Ruhlen 1991; Sands 1998), although others have suggested that Hadza may have similarities with Afro-Asiatic languages (Elderkin 1982). Overall, proposed linguistic relationships among the click languages (fig. 1) predict deep genetic divergence between these 3 groups (Cavalli-Sforza et al. 1988).

Given the linguistic similarities between the Sandawe and SAK, the Sandawe constitute a key population for reconstructing the history of African click-speaking populations. Previously, the Sandawe were studied at the population level only for a small number of classical protein polymorphisms, which revealed no particularly close relationship with any other population (Cavalli-Sforza et al. 1994). Here, we present mtDNA and Y chromosome genetic data for the Sandawe, in addition to novel data for neighboring Tanzanian populations (including the Hadza) and from SAK-speaking populations.

We evaluate 2 models for the relationships among these populations. The first model is suggested by the linguistic relationships described above, which predict a closer genetic relationship between the Sandawe and SAK speakers than between either of these groups and the Hadza. The second model, based on the observation that genetic distances often correlate with geographic distances among populations (Cavalli-Sforza et al. 1994; Rosenberg et al. 2002), predicts that the Hadza and Sandawe populations, who reside within 150 km of one another, will be genetically more similar to each other and to neighboring Tanzanian populations than either is to the SAK, who reside >2,000 km away. By evaluating the consistency between the genetic data and each of the above models, we can assess the relationships between geographic distance, linguistic differentiation, and genetic diversity as well as the extent of isolation of click-speaking populations of Africa.

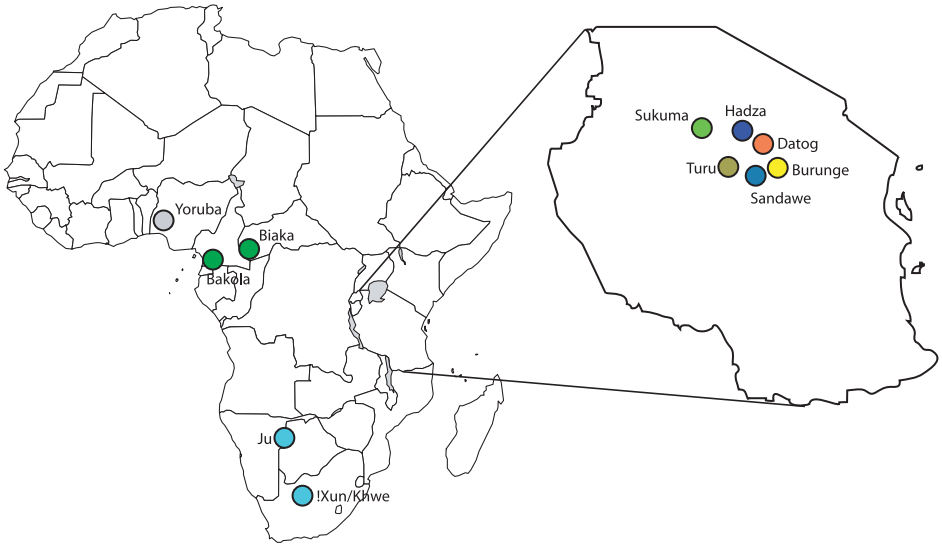


FIG. 2.—Maps indicating locations of populations considered in this study.

Methods

Population Samples

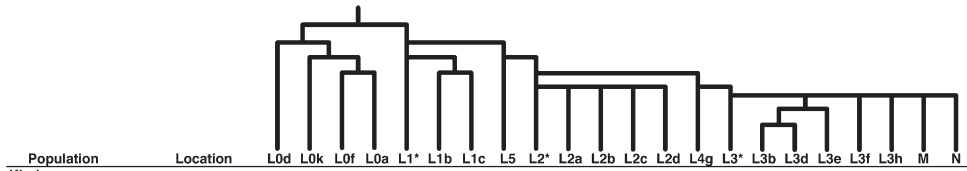
We generated Y chromosome and mtDNA data for the Sandawe, Hadza, and neighboring Tanzanian populations and for SAK speakers represented by Ju speakers (Jul'hoansi) and a mixed sample of Ju (!Xun) and Khoe (Khwe) speakers, referred to here as !Xun/Khwe (fig. 2). DNA samples were collected from the following populations (language classification in parentheses): Hadza and Sandawe (Khoisan), Burunge (Afro-Asiatic: Cushitic), Datog (Nilo-Saharan: Nilotic), Turu (who reside near the Sandawe; Niger-Kordofanian: Bantu), and Sukuma (who reside near the Hadza; Niger-Kordofanian: Bantu) in the Arusha and Dodoma provinces of Tanzania. The Ju-speaking !Xun (also known as Vasekela) and Khoe-speaking Khwe samples were collected from individuals in the area of Schmidtsdrift in the North-West Cape region of South Africa and were provided by Dr M. J. Kotze. Individuals were grouped according to self-identified ethnicity, and only samples from unrelated individuals who could trace ancestry to the same ethnic group as far back as the grandparents were included in the study. Written informed consent was obtained from all donors, and Institutional Review Board approval and permits from Commission for Science and Technology and National Institute for Medical Research in Tanzania were obtained prior to sample collection. In the field, a red cell lysis buffer (1 mM NH_4HCO_3, 115 mM NH_4CL) was added to 9 ml of whole blood DNA. White blood cells were isolated via centrifugation and suspended in a white cell lysis buffer (100 mM Tris–HCl [pH 7.6], 40 mM ethylenediaminetetraacetic acid [pH 8.0], 50 mM NaCl, 0.2% sodium dodecyl sulfate, 0.05% sodium azide) and were stored at ambient temperature. DNA was isolated in the laboratory using a Purgene kit (Gentra, Minneapolis, MN). Additional DNA samples used in some comparative analyses (Yoruba: Niger-Kordofanian, Defoid, and SAK speakers) were obtained from the Centre d'Étude du Polymorphisme Humain–Human Genome Diversity Cell Line Panel collection (Cann et al. 2002). The geographic distribution of populations and total number of individuals included in this study are shown in figures 2–4.

Y Chromosome and mtDNA Data Collection

We generated nucleotide sequence data for 649 bp of the mtDNA control region, including hypervariable region I (HVRI; position 16027–16363) and hypervariable region II (HVRII; position 073–379), as well as single-nucleotide polymorphism (SNP) genotypes at sites outside the control region that were used to define haplogroups (fig. 3; supplementary Table 1, Supplementary Material online). All sequences have been deposited in GenBank (accession numbers EF999432–EF999757). For the nonrecombining portion of the Y chromosome, we genotyped 15 unique event polymorphisms (UEPs; fig. 4) and 12 short tandem repeat (STR) polymorphisms. Y-STR markers include: DYS391, 389I and II, 439, 438, 437, 19, 392, 392, 390, and 385a/b using the Promega PowerPlex Y System. DYS385a/b consists of a duplicated tetranucleotide STR region (Kayser et al. 2001) and was omitted from some analyses. Y chromosome genotype data for all individuals are given in supplementary Table 1 (Supplementary Material online).

Analyses

Haplotype networks were generated for human mtDNA haplogroups L0, L1, L2, and L3 and for the Y chromosome STRs on UEP backgrounds via the median-joining algorithm of Network 4.1.1.1 (http://www.fluxus-engineering.com). Because it allows for reticulation, the median-joining approach to the inference of haplotype relationships is appropriate for the analyses of human mtDNA control region sequences and Y chromosome short tandem repeat polymorphism haplotypes, which exhibit high levels of homoplasy (Bandelt et al. 1999; Posada and Crandall 2001). For the mtDNA data, hypermutable sites were identified by



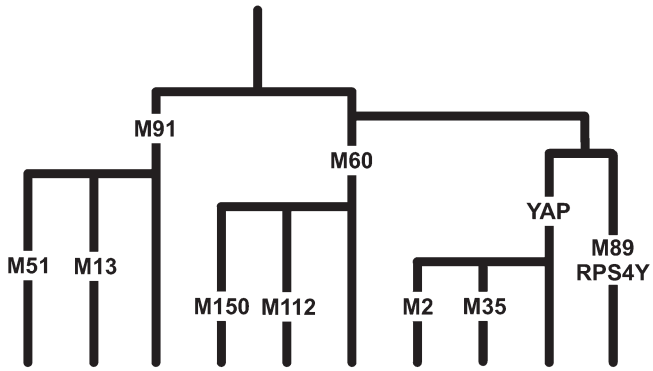
Population	Location	L0d	L0k	L0f	L0a	L1*	L1b	L1c	L5	L2*	L2a	L2b	L2c	L2d	L4g	L3*	L3b	L3d	L3e	L3f	L3h	M	N	EA+	Total	Study
Khoisan																										
Hadza II	**Tz**				**5**	**1**			**1**	**1**	**10**				**60**	**8**	**4**	**1**		**5**	**1**	**1**	**1**		**79**	**Present**
Sandawe	**Tz**	**5**		**4**	**17**				**5**		**1**				**48**				**13**	**2**		**4**			**82**	**Present**
Hadza I	**Tz**									17†					83†										12	Knight (2003)
!Xun/Khwe	**SA**	**61**	**22**							**17**															**18**	**Present**
!Xun	Nm	51	26		2					16									5						43	Vigilant (1991)
!Kung	Bo	96	4																						26	Vigilant (1991)
Niger-Kordofanian																										
Sukuma II	**Tz**											**9**			**18**	**9**	**9**	**27**	**9**	**18**					**11**	**Present**
Turu	**Tz**			**14**	**10**	**3**		**3**			**7**				**28**	**10**	**7**	**3**	**3**	**3**	**3**	**3**	**3**		**29**	**Present**
Mozambicans	Mz	5			25		1	5	1	1	33	1	1			1	3	6	15	2					416	Pereira (2001), Salas (2002)
Sukuma I	Tz					29†				9†						62†									21	Knight (2003)
Mbenzele	CA				4			97																	57	Destro-Bisol (2004)
Biaka	CA				24			77																	17	Vigilant (1991)
Bakola	CA							**100**																	**30**	**Present**
Nigerians	WA				1	1	15	3		1	20	1	4				13	12	15	6	1		3	4	137	Watson (1997), Vigilant (1991)
Guineans	WA				5		11	5		2	17	8	16	3			8	9	7	2	4	1	2	3	372	Rosa (2004)
Nilo-Saharan																										
Datoga	**Tz**			**5**	**28**	**3**								**8**	**13**	**16**					**11**	**8**	**8**		**39**	**Present**
Mbuti	CA				30				15	55															20	Vigilant (1991)
Afro-Asiatic																										
Burunge	**Tz**	**3**		**30**	**25**					**3**					**8**	**18**		**3**		**8**		**3**	**3**		**38**	**Present**
Ethiopia	Et	<1		<1	7		2				13	20						3		5	1	17	3	28	270	Rosa (2004)
Mixed																										
Kenya/Sudan	Ky/Su	1			11				16		16				7	25			3	8		5	2	6	88	Watson (1997)

FIG. 3.—Schematogram of phylogency of major mtDNA haplotype lineages based on Gonder et al. (2007) and frequencies (%) of major mtDNA haplogroups in a set of African populations. mtDNA haplotype fequencies determined within the current study are shown in bold: Burunge, Datog, Hadza, Sandawe, Sukuma, Turu, !Xun/Khwe, and Bakola Pygmies. Locations of populations are abbreviated as: Bo, Botswana; CA, Central Africa; Et, Ethiopia; Mz, Mozambique; Nm, Namibia; SA, South Africa; Tz, Tanzania; and WA, western Africa. Haplogroup designations for samples produced for this study follow Salas et al. (2002; 2004) and Kivilsild et al. (2004). Samples classified as EA (column heading) were defined as Eurasian by Rosa et al. (2004); these sequences are all non-L's, M1, or U6 sequences. L1*, L2*, and L3* from previous studies indicate samples that were not further subdivided into subhaplogroups. L2* and L3* from this study indicate samples that were tested for SNP variation but did not fit into known haplogroup classifications. Samples labeled Sukuma I (Knight et al. 2003) were combined with Sukuma II samples for additional analyses by MDIV.

postprocessing using the Steiner maximum parsimony (MP) algorithm within Network 4.1.1.1 (Polzin and Daneschmand 2003). These sites were, in most cases, confirmed to be hypermutable in previous studies (Hasegawa et al. 1993; Wakeley 1993; Meyer et al. 1999) and were excluded from the network analyses. When estimating the age of mtDNA haplotypes and times of population divergence, we assumed a mtDNA mutation rate of $\mu = 2.5 \times 10^{-6}$ per nucleotide per generation over 649 bp (based on a substitution rate of 11%/Myr [Ward et al. 1991], bracketed by higher [Forster et al. 1996] and lower [Horai et al. 1995] rate estimates, a 6 Myr divergence between humans and chimpanzees, and assuming a female generation time of 25 years). Genetree 9.0 (http://www.stats.ox.ac.uk/~griff/software.html) was used to estimate the time to most recent common ancestor (T_{MRCA}) of haplogroups (L4g, L0d) via Markov chain coalescent simulation (a constant population size was assumed with 10 million step chains to estimate the likelihood surface; Griffiths and Tavaré 1997).

For the Y chromosome data, we generated median-joining networks for Y chromosome STR haplotypes on each of the following UEP backgrounds: E3a-M2, B2-M60, B2b-M112, and E3b1-M35 (Bandelt et al. 1999). Epsilon (reticulation permissivity) was set to zero in order to generate the most parsimonious networks. Because of the high level of reticulation in the E3a-M2 sample, data were preprocessed using the star contraction option in Network 4.1.1 (Forster et al. 2001). In order to assign an ancestral haplotype for ρ estimates, we rooted networks using haplotypes from closely related UEP-defined clades. When available, published mutation rates specific to a locus were assumed (Forster et al. 2000). STR loci were subdivided into 3 mutation rate classes based on observed STR allelic variance (low: DYS437, DYS438, DYS391, DYS392; intermediate: DYS389I, DYS389II, DYS439, DYS19, DYS392, DYS390; or high: DYS385a/b). When no published mutation rate was available, we assumed a mutation rate of 0.0009/generation for a rapidly mutating locus, 0.0008/generation for an intermediate-rate locus (Zhivotovsky et al. 2004), and 0.00018/generation for a slowly mutating locus (Forster et al. 2000). In generating networks, loci were weighted as follows: low (4):intermediate (2):high (1). Across 11 loci, the average mutation rate was assumed to be 0.00059/locus/generation, with an average generation time of 30 years. The ages of Y chromosome haplogroups and subhaplogroups were estimated via the ρ statistic (i.e., the average number of STR mutations from derived haplotypes to a haplotype designated as ancestral for the haplogroup or subhaplogroup), using the software package Network 4.1.1 (Forster et al. 1996).

A Markov Chain Monte Carlo approach was used to determine the likelihood of values of population divergence times and migration rates for the mtDNA HVRI/HVRII sequence data and linked Y-SNPs and Y-STRs, in the context of the isolation with migration (IM) model (Nielsen and Wakeley 2001). For pairs of populations, we conducted the analysis via the software packages MDIV for mtDNA and IM for the Y chromosome, estimating the scaled mutation rate ($\theta = N_e\mu$, where N_e is the effective population size, assuming an equal sex ratio, and μ is the per generation locus mutation rate), the migration rate ($M = N_e m$, where m is the per generation migration fraction), and the time of population divergence (t, in units of N_e generations). For the mtDNA analyses, runtime parameters included



Population	Location	A3b1	A3b2	A*	B2a	B2b	B1,B*	E3a	E3b	E*,D	C,F-R	Total	Study
Khoisan													
Hadza I	Tz					52		30	13		4	23	Knight (2003)
Hadza II	**Tz**					**51**	**9**	**24**	**16**			**57**	**Present**
Sandawe	**Tz**		**4**		**1**	**13**		**43**	**34**	**5**		**68**	**Present**
Khwe	SA	12						54	31	4		26	Cruciani (2004)
Dama	Nm	6		6			6	55		18	12	18	Wood (2005)
Nama	Nm	55		9				18	9		9	11	Wood (2005)
!Kung	SA	28		8		8		39	11	6		64	Cruciani (2004)
Khoisan-mixed	SA	28		15		28		18	10			39	Underhill (2001)
Niger-Kordofanian													
Yoruba	**Ng**						**8**	**92**				**13**	**Present**
Turu	**Tz**					**5**		**50**	**15**	**30**		**20**	**Present**
Mbugwe	**Tz**				**7**			**33**	**20**	**40**		**15**	**Present**
Wairak	Tz		7		2	5		42	37	2		43	Luis (2004)
Hutu	Rw				3	1		83	3	8		69	Luis (2004)
Tutsi	Rw				1	14		80	1	4		94	Luis (2004)
Sukuma	Tz				9	6		63	6	16		32	Knight (2003)
Biaka	CA				5	30		65				20	Underhill (2001)
Lissongo	CA					25		75				4	Underhill (2001)
Herero	Nm							71			29	24	Wood (2005)
Ambo	Nm							82	5	10	5	22	Wood (2005)
Xhosa	SA	5					5	54	5	32		80	Wood (2005)
Bamileke	Cm						4	96				48	Cruciani (2004)
Mossi	BF						2	90	2	6		49	Cruciani (2004)
Nilo-Saharan													
Datog I	Tz							13	63		25	8	Knight (2003)
Datog II	**Tz**		**3**			**3**		**11**	**54**	**23**	**6**	**35**	**Present**
Maasai	Ky		27		8			16	50			26	Wood (2005)
Mbuti	Co				8	25		42		25		12	Underhill (2001)
Afro-Asiatic													
Burunge	**Tz**					**25**		**4**	**33**	**29**	**8**	**24**	**Present**
Iraqw	Tz		17						33	17	33	6	Knight (2003)
Mixed	Cm		2		4			13	4	7	70	54	Cruciani (2004)
Mixed	Mo		2					2	78	1	17	113	Cruciani (2004)

FIG. 4.—Y chromosome UEP-defined haplogroup frequencies (%) in 30 African population samples, with haplogroup nomenclature as outlined by Y Chromosome Consortium (International Human Genome Sequencing Consortium 2002). Language family and subfamily are assigned according to Nurse et al. (1985). Locations of populations are abbreviated as: BF, Burkina Faso; CA, Central African Republic; Cm, Cameroon; Co, Democratic Republic of Congo; Ky, Kenya; Mo, Morocco; Ng, Nigeria; Nm, Namibia; Rw, Rwanda; SA, South Africa; Se, Senegal; Tz, Tanzania. "*N*" indicates the number of Y chromosomes studied. Column "A*" includes the frequencies of those Y chromosomes that are M91 positive, M13 negative, and M51 negative. Column "B1, B*" includes all individuals of haplogroup B except for those of B2a and B2b haplogroups. Column "E*, D" includes individuals in haplogroups E1, E2, E* plus D only. Individuals identified in Underhill et al. (2000) as Khoisan are more accurately identified as a mixed sample of Jul'hoansi and !Xun (~50% each). Both groups belong to the Northern Khoisan "Ju" language family. Individuals described here as "!Xun" are also known by the Bantu derived term "Sekele" or "Vasekele." None of the E3b mutations (M78, M81, M123, V6) or B2b mutations (P6 or P7) were observed among the Tanzanian samples, with the following exceptions—Datog: one M78+ individual, one V6+ individual; Mbugwe: one M78+ individual. Previously published data originate from Cruciani et al. (2002), Knight et al. (2003), Luis et al. 2004, Underhill et al. (2001), and Wood et al. (2005).

a maximum migration rate of $Nm = 10$, a maximum divergence time of 5 N_e generations, a burn in time of 1,000,000 steps followed by a run length of 10,000,000 steps, and an HKY mutation model. A high correlation among runs was observed (mode parameter value estimates were correlated with an $r^2 = 0.90$ or higher). Confidence intervals (CIs) were estimated assuming a standard chi-square/2 approximation (i.e., −2 log likelihood units from the mode). In cases where the upper tail of the posterior distribution does not decline greater than 2 log likelihood units from the mode, we presumed an upper bound of 100 kya for population divergence (see supplementary figure 1, Supplementary Material online). For the Y chromosome analyses, run-time parameters included a maximum migration

rate of $Nm = 10$, a maximum divergence time of 10 N_e generations, and a burn in of 200,000 steps followed by a run of 2,000,000 steps with 5 genealogy updates per step under the HapSTR mutation model for 13 SNPs and 10 STRs (Hey and Nielsen 2004). Y chromosome runs used 5 coupled chains with a linear heating scheme of increment 0.1. All pairwise population sample comparisons were replicated with different random number seeds. The mode of each marginal posterior distribution was considered a point estimate of the corresponding parameter value. Note that the publicly available MDIV and IM-type computer simulation package is limited to models with 1) two populations only 2) a one population split occurring at time, *t*, in the past, and 3) migration occurring at a constant rate following the split (Nielsen and Wakeley 2001; Hey and Nielsen 2004). These simplified models do not take into account the effect of shared third-party migration, which can result in overly high estimates of migration and overly recent estimates of divergence time if both the 2 populations under consideration have exchanged genes with the same neighboring populations. Nor do these models take into account the possibility of changing migration rates. For example, high levels of ancient gene flow between populations followed by population divergence could result in an estimate of *t* that reflects the time at which gene flow stopped occurring (Hey J, personal observation). To complete the large number of MDIV simulations required for this study, we used Grid computing (Myers and Cummings 2003; Cummings and Huskamp 2005) through The Lattice Project (Bazinet and Cummings, forthcoming). A Grid service for the MDIV analyses was developed using a special programming library and associated tools (Bazinet et al. 2007).

Genetic distances between populations were estimated from Y chromosome STR variation according to Goldstein et al. (1995). Distances were summarized graphically according to the Neighbor-Joining algorithm (Saitou and Nei 1987).

The presence of the Y chromosome haplogroup E3a-M2, very frequent among Bantu speakers of western Africa, in both the Hadza and Sandawe populations suggested that gene flow from Bantu-speaking groups into both the Hadza and the Sandawe populations might obscure the historical relationships between these 2 populations. Therefore, for the Y chromosome data, we investigated models incorporating 3 populations with gene flow: 2 click-speaking populations (representing Hadza and Sandawe) and 1 Niger-Kordofanian-speaking population (represented in this analysis by the Yoruba to eliminate the impact of recent gene flow into Tanzanian Bantu-speaking populations). Because no standard approaches allow for evaluation of 3-population models and multiple migration rates, we used a rejection-sampling approach, generating simulated likelihoods (Pritchard et al. 1999). We used an extended version of the software, SIMCOAL (Excoffier et al. 2000), that allows the user to simulate mutations generating UEPs and STRs on the same gene genealogy, as is appropriate for the Y chromosome.

Specifically, we generated distributions of STR-based summary statistics conditioned on UEP frequencies that result from simulating a particular demographic history. We then estimated the probability of the observed summary statistics given these distributions. The product of these probabilities for all summary statistics, combined with the frequency of simulations with UEP frequencies near those observed, provides a simulated likelihood for a given model of population history. Simulated data were accepted if observed UEP frequencies (f) met the following ascertainment criteria: Hadza: f(M112) $>$ 0, f(M2) $>$ 0, f(M35) $>$ 0, f(M112)-f(M2)-f(M35) $>$ 0, f(M112) $>$ f(M2), and f(M112) $>$ f(M35); Sandawe: f(M112) $>$ 0, f(M2) $>$ 0, f(M35) $>$ 0, f(M112)-f(M2)-f(M35) $>$ 0, f(M2) $>$ f(M112), and f(M2) $>$ f(M35); Yoruba: f(M2) $>$ 0, f(M112)-f(M2)-f(M35) $>$ 0, and f(M2) $>$ 50%. For each accepted simulation, 12 STR-based summary statistics were calculated—E3a-M2 background: $(\delta\mu)^2$ between Hadza and Sandawe, $(\delta\mu)^2$ between Hadza and Yoruba, $(\delta\mu)^2$ between Yoruba and Sandawe, Hadza allelic variance, Sandawe allelic variance, and Yoruba allelic variance; B2b-M112 background: $(\delta\mu)^2$ between Hadza and Sandawe, Hadza allelic variance, and Sandawe allelic variance. E3b1-M35 background: $(\delta\mu)^2$ between Hadza and Sandawe, Hadza allelic variance, and Sandawe allelic variance. We evaluated a set of 24 population history models via this simulation-based approach by considering all ascertained models that generated summary statistics ($[\delta\mu]^2$ and allelic variance) within $\pm 10\%$ of the observed statistics. Specifically, the simulated likelihood (L_{sim}) for any given population history model was estimated via the product of the frequency of ascertainment (f_a) and the frequency of ascertained simulations with acceptable summary statistics (f_{ss}). Models were then ranked according to L_{sim} values.

Results

Mitochondrial DNA

Frequencies of the major mtDNA haplogroups in each population, plus previously published data for comparison, are shown in figure 3. Networks indicating genealogical relationships of haplotypes are shown in figure 5. The L0 plus L1 clades, which are basal in the global mtDNA tree and are specific to Africa (Salas 2002; Chen 2000; Gonder 2007), are common in our data set (figs. 3 and 5*a*). We observe the L0d haplogroup at low frequency in the Sandawe (and in one individual from the neighboring Burunge population) but not in the Hadza (figs. 3 and 5*a*,*b*). Prior to the current study and a recent study of complete mtDNA genomes in Tanzania (Gonder 2007), the mtDNA L0d haplogroup was detected at very high frequencies only in SAK-speaking populations (61–96%); the haplogroup was detected at low frequencies in neighboring Mozambique populations that are likely to have exchanged individuals with the SAK speakers (Salas et al. 2002) and in one Turkana individual from Kenya (Watson et al. 1997; Salas et al. 2002). Network analysis of the HVRI region, for which there is a comparative data set, indicates that the Tanzanian and Kenyan L0d lineages form a monophyletic subclade of the L0d haplogroup (fig. 5*b*). Based on analysis of the concatenated HVRI/HVRII data, Tanzanian and SAK L0d lineages differ by a minimum of 6 mutations, and coalescent analysis indicates that the youngest clade to include both

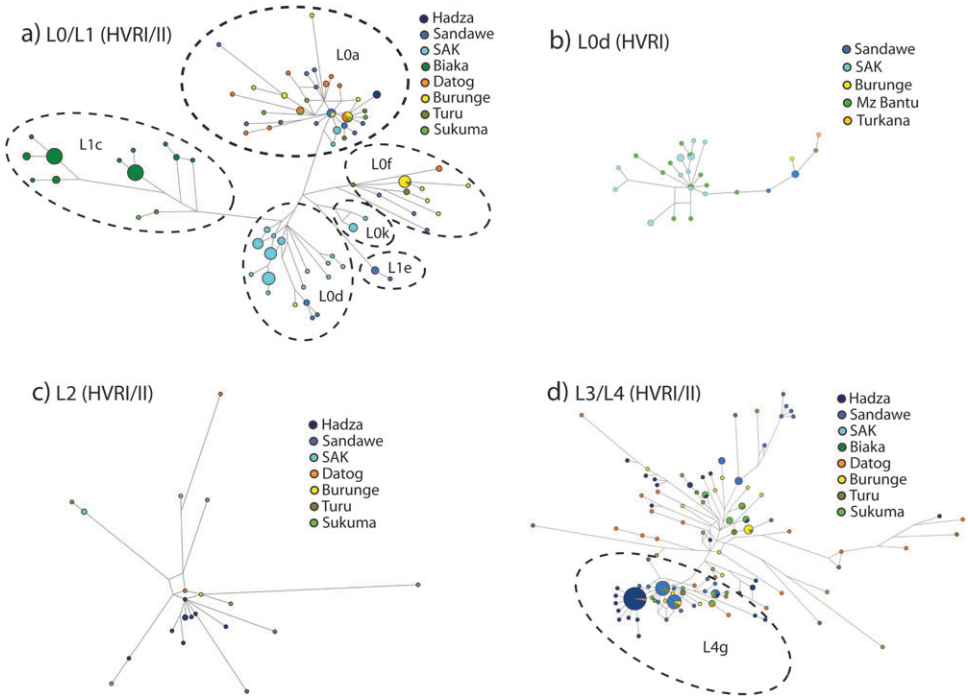


FIG. 5.—Median-joining networks (Bandelt et al. 1999) derived from mtDNA sequences of (*a*) L0 plus L1, (*b*) L0d, (*c*) L2, and (*d*) L3/L4 haplogroups. L0/L1/L2/L3/L4 haplogroups were defined based on SNPs shown in supplementary Table 1 (Supplementary Material online; hypervariable sites are excluded).

SAK and Tanzanian L0d lineages has a coalescent-based T_{MRCA} estimate of ~58 kya (95% CI 33.7–90.1 kya). The T_{MRCA} of the Tanzanian L0d lineages is ~23.4 kya (95% CI 9.4–50.7 kya), whereas the T_{MRCA} of the SAK L0d lineages is ~59 kya (95% CI 18.5–145 kya).

Haplogroup L4g (previously designated L3g) is present in both Tanzanian click-speaking populations at high frequencies (60% Hadza, 48% Sandawe) but is absent in the SAK. The L4g haplogroup is most frequent in eastern and northeastern Africa and was previously dated to ~40–45 kya (Salas et al. 2002; Kivisild et al. 2004). We observe very little HVRI/HVRII nucleotide diversity within the L4g haplogroup in the Hadza sample, consistent with previous studies of this population (fig. 5*d*; Vigilant et al. 1991; Knight et al. 2003). The "star-like" pattern of variation in the Hadza (with identical HVRI/HVRII haplotypes in 36 of 46 individuals) is consistent with an expansion of this subhaplogroup ~4 kya. Our sample of the Sandawe also reveals a high frequency of L4g but with somewhat greater HVRI/HVRII nucleotide variation than in the Hadza. The most common Sandawe and Hadza L4g lineages differ by 3 mutations (fig. 5*d*), and the estimated T_{MRCA} for this lineage in these populations is 24.5 kya (95% CI of 11.2–47.6 kya). The Hadza and Datog share several L4g haplotypes (after removal of hypervariable sites); the Sandawe, Turu, and Burunge share other L4g haplotypes (fig. 5*d*).

Both the Hadza and Sandawe have a high frequency of the mtDNA L3 haplogroup that is common in eastern African populations (figs. 3 and 5*d*; Salas et al. 2002). The Hadza, unlike the Sandawe, have a high frequency of the L2 haplogroup that is most common in western African populations, although not uncommon in eastern Africa (Salas et al. 2002; figs. 3 and 5*c*). We observed few shared haplotypes between the Hadza and Sandawe (and none that are identical when hypervariable sites are included), suggesting limited recent gene flow between the Hadza and Sandawe (fig. 5).

Maximum likelihood estimates of migration rates (m) and dates of population divergence (t), inferred from the mtDNA data (Nielsen and Wakeley 2001), are given in table 1. Figure 6 provides graphical summaries of these estimates. These joint estimates of t and m indicate that the Hadza and Sandawe diverged ~23 kya (CI 13–37 kya) and have had relatively low levels of genetic exchange with each other ($M = 1.15$ [CI 0.08–2.5 M], suggesting approximately one migrant per generation) and moderate levels of genetic exchange with neighboring Tanzanian populations (mode estimates range from $1.53 \leq M \leq 3.88$). The Sandawe sample shows evidence of very low levels of migration with the Jul'hoansi ($M = 0.32$ [CI 0.02–1.12 M] or an average of one migrant every 3 generations) and with the !Xun/Khwe ($M = 0.49$ [CI 0.04–1.68 M]). The Sandawe and SAK-speaking populations share no identical mtDNA haplotypes, suggesting that any genetic exchange was not recent. Neither the Jul'hoansi sample nor the!Xun/Khwe sample reveals mtDNA evidence for any migration between the SAK speakers and the Hadza ($M = 0.02$ [CI 0–0.66 M and 0–0.40 M, respectively]; table 1 and fig. 6). A neighbor-joining tree constructed from the pairwise matrix of t values is shown in figure 6*b*. The Pygmy populations show a relatively recent time of divergence, as do the 2 SAK populations, and both groups of populations are highly divergent from all the Tanzanian populations, which cluster together. However, the Hadza and Sandawe cluster more closely with

Table 1
Estimates of Migration Rates (*M*) and Time of Divergence between Population Pairs Based on MDIV Analysis for mtDNA HVRI/HVRII Sequences

Population[a]	Ju	!X./K.[b]	Sandawe	Hadza	Turu	Mbugwe	Sukuma	Burunge	Datog
Ju	—	**1.04**	0.32	0.02	0.02	0.02	0.05	0.29	0.15
!X./K.	**5.2**	—	0.49	0.02	0.02	0.34	0.41	0.38	0.32
Sandawe	50	44	—	**1.15**	**3.88**	**1.53**	**3.04**	**2.86**	**2.58**
Hadza	56	56	**23**	—	**2.3**	**2.06**	**2.91**	**2.29**	**3.19**
Turu	43	45	**8.2**	**15**	—	**2.75**	**9.95**	**9.87**	**4.92**
Mbugwe	55	58	**27**	**21**	40	—	**9.99**	**3.16**	**2.24**
Sukuma	64	69	**19**	**14**	**2.6**	**2.9**	—	**7.13**	**9.94**
Burunge	49	48	**23**	**23**	**2.9**	**14**	**9.6**	—	**6**
Datog	57	53	**23**	31	**23**	52	**3.4**	**15**	—

[a] Estimates of *M* are shown above the diagonal (values larger than 1 are in bold). Estimates of time of population divergence in thousands of years is shown below the diagonal (values less than 30 kya are in bold).
[b] !Xun/Khwe.

each other than with other Tanzanian populations and more closely with the SAK-speaking populations relative to any other Tanzanian populations. The estimated time of divergence between the Sandawe and SAK speakers is ~44–50 kya (CI 21–100 kya for the Jul'hoansi and 29–100 kya for the !Xun/Khwe), whereas the Hadza are estimated to have diverged from the SAK-speaking populations ~56 kya (CI 33–100 kya for the Jul'hoansi and 40–100 kya for the !Xun/Khwe; table 1).

Y Chromosome

Frequencies of the major Y chromosome UEP-defined haplogroups are presented in figure 4; STR-based networks are presented in figure 7. Although both the Sandawe and SAK populations have Y chromosome haplogroup A-M91 lineages, the Sandawe share subhaplogroup A3b2-M13 with other eastern Africans to their north but lack any subhaplogroup A3b1-M51 lineages observed in the SAK-speaking populations (Scozzari et al. 1999; Cruciani et al. 2002; Wood et al. 2005). Therefore, the sharing of the A haplogroups by the Sandawe and SAK speakers cannot be taken as evidence of a particularly close relationship.

In the Hadza, Sandawe, and SAK populations, we observe 3 relatively basal Y chromosome haplogroups: B2b (defined by the M112 mutation), E3b1 (defined by the M35 mutation), and E3a (defined by the M2 mutation; figs. 4 and 7). Estimates of Y UEP haplogroup ages derived from STR variation are presented in table 2. The relatively old (>55 kya, table 2 and fig. 7*c*) B2b-M112 haplogroup is unreported outside of sub-Saharan Africa and is most common in hunter/gatherer populations across sub-Saharan Africa, notably the central African Pygmies and the SAK-speaking Jul'hoansi (fig. 4; Knight et al. 2003; Wood et al. 2005). We detected the B2b haplogroup at a high frequency in the Hadza and at a moderate frequency in the Sandawe (figs. 4 and 7*c*). Tanzanian populations do not, however, harbor the P6 and P7 mutations that define the subhaplogroups of B2b that are found among SAK speakers and central African Pygmy populations (Wood et al. 2005). The frequency of B2b in the Hadza (51%) is higher than reported in any other population (fig. 4). In addition, the B2b lineages in the Hadza have higher STR diversity than any of the other surveyed populations, and STR haplotypes are shared with only one Datog individual (fig. 7*c*). Sandawe and SAK B2b lineages appear to be similarly distinct from those of other populations (fig. 7*c*). STR variation of B2b subclusters that include primarily SAK speakers, Hadza, and Sandawe indicates that these groups shared a common ancestor ~35 (±4) kya (table 2 and fig. 7*c*).

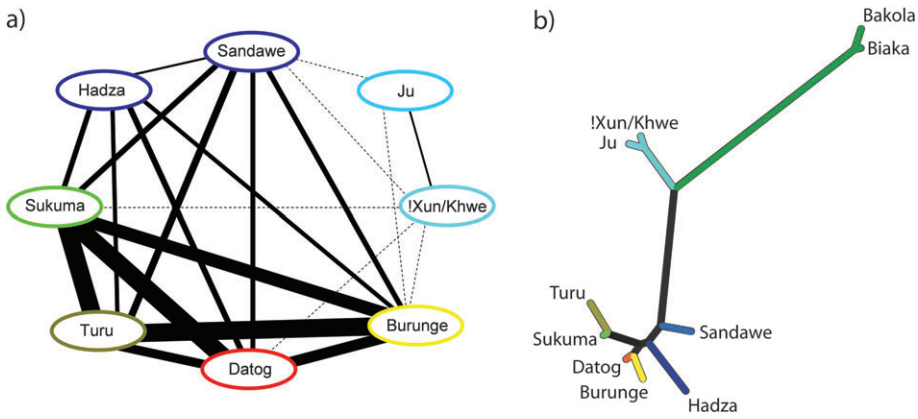


FIG. 6.—Graphical representations of pairwise population estimates of migration and time of population divergence from table 1 for mtDNA. (*a*) Migration rates among the populations (*M*) are represented by a web, where $M > 1$ line width is proportional to *M* estimates. Migration rate estimates of $0.3 < M < 1$ are plotted as dashed lines, and estimates of $M < 0.3$ are not plotted. (*b*) A neighbor-joining tree derived from estimates of the time of population divergence (table 1).

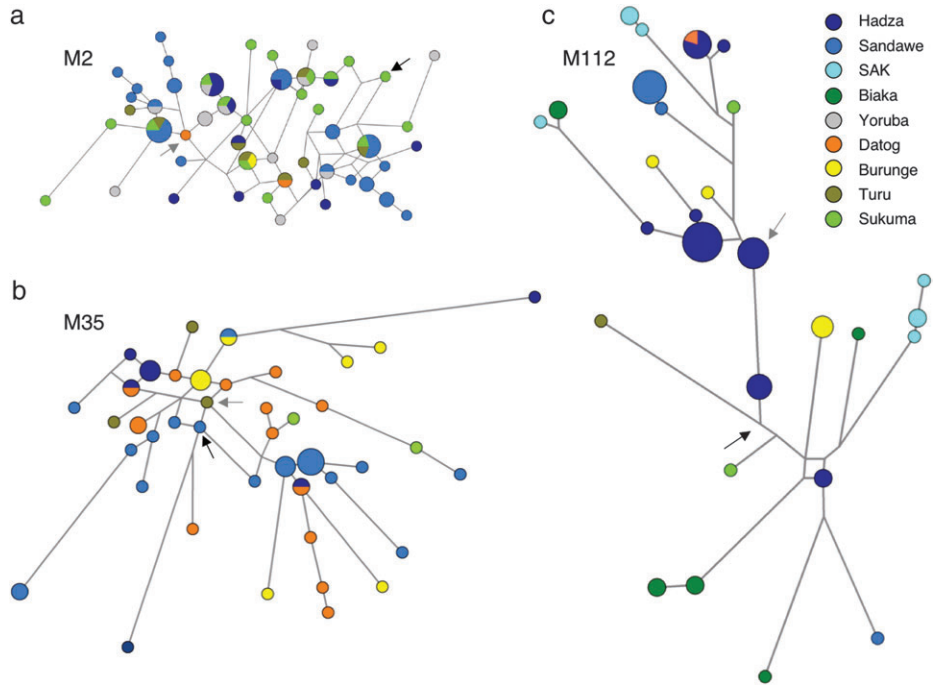


FIG. 7.—Median-joining networks (Bandelt et al. 1999) for 3 SNP-defined NRY clades generated on the basis of variation at 12 Y chromosome STR loci. (*a*) E3a: M2 positive; (*b*) E3b: M35 positive; (*c*) B2b: M112 positive. Black arrows indicate the ancestral root of each network inferred from adding STR data from respective NRY sister clades. Gray arrows indicate the haplotype in each network that led to the minimum ρ estimate. Jul'hoansi haplotypes (CEPH set sample) were available only for B2b (M112)-positive individuals.

The younger (>27 kya, table 2 and fig. 7*b*) E3b1 haplogroup is most frequent in eastern Africa but is also found in northern Africa and southern Europe (fig. 4; Cruciani et al. 2004). E3b1-M35 is found at relatively high frequency in the Sandawe (34%) and SAK/Khwe speakers (31%) but is less common (~10–15%) in the Hadza and in the other SAK speakers (fig. 4). Most of the Tanzanian E3b1-M35 lineages are undifferentiated (indicated by M35*) and do not contain mutations which define subhaplogroups common in Afro-Asiatic and Nilo-Saharan populations from Ethiopia and Kenya (i.e., -M78, -M81, -M123, -M281, -V6; Cruciani et al. 2004). The Datog population from Tanzania carries exceptionally high frequencies of E3b1-M35 (54–63%) and the Burunge population has moderate frequencies of this haplogroup (21%) (fig. 4), the vast majority of which are M35*. Given the high frequency of E3b1-M35* in these populations, it is possible that the Hadza and Sandawe have acquired M35* through admixture with neighboring Nilotic- and Cushitic-speaking populations. This is supported by 2 shared Y-STR haplotypes between the Hadza and Datog (fig. 7*b*). However, as indicated in the network shown in figure 7*b*, there are several E3b1-M35* STR haplotypes in the Hadza and Sandawe that are highly divergent and separated by a large number of mutations from the rest of the E3b1-M35* cluster.

Although STR-based networks for Y chromosome haplogroups A-M91, B2b-M112, and E3b1-M35 reveal no haplotype sharing between the Hadza and Sandawe, these 2 populations share one STR haplotype on the E3a-M2 background (fig. 7*a*). The E3a haplogroup appears to be relatively young (>21 kya, table 2 and fig. 6*a*) and is most frequent in Bantu-speaking populations of western Africa (Underhill et al. 2001; Wood et al. 2005).

Maximum likelihood estimates of migration rates (*m*) and dates of population divergence (*t*) were inferred from the Y chromosome data (Hey and Nielsen 2004). In several cases, estimation of divergence time consistently yielded multimodal distributions, and hence, estimates of *t* are not presented here. Estimates of migration rate appeared

Table 2
Estimates of Dates (*T*) of Y-Chromosome Nodes Derived from Associated STR variation on Each UEP background via ρ estimates (Forster et al. 1996)

UEP	Nodes	*N*	ρ	Standard Error	T^a	Bound
M112	Sister clade root	64	14.22[b]	0.47	69,900	Upper
	Minimizing root	64	11.25[c]	0.42	55,300	Upper
	Hadza versus Sandawe versus SAK[d]	25	7.042	0.53	34,600	Upper
M35	Sister clade root	60	6.263[b]	0.32	30,800	Upper
	Minimizing root	60	5.439[c]	0.30	26,700	Upper
M2	Sister clade root	89	7.402[b]	0.29	36,400	Upper
	Minimizing root	89	4.256[c]	0.22	20,900	Upper

NOTE.—Bounds refer to divergence time for all study populations that include individuals with the UEP, unless otherwise specified.

[a] Date in years assuming a mutation rate of 0.0061 mutations per 11 STR loci and a 30-year male generation length.

[b] ρ estimate based on rooting the UEP network with a sister clade.

[c] Minimum ρ estimate.

[d] Average ρ for 2 clusters in the M112 network; ρ indicates the average number of mutations from SAK, Sandawe, and Hadza haplotypes to the MRCA for all 3 populations.

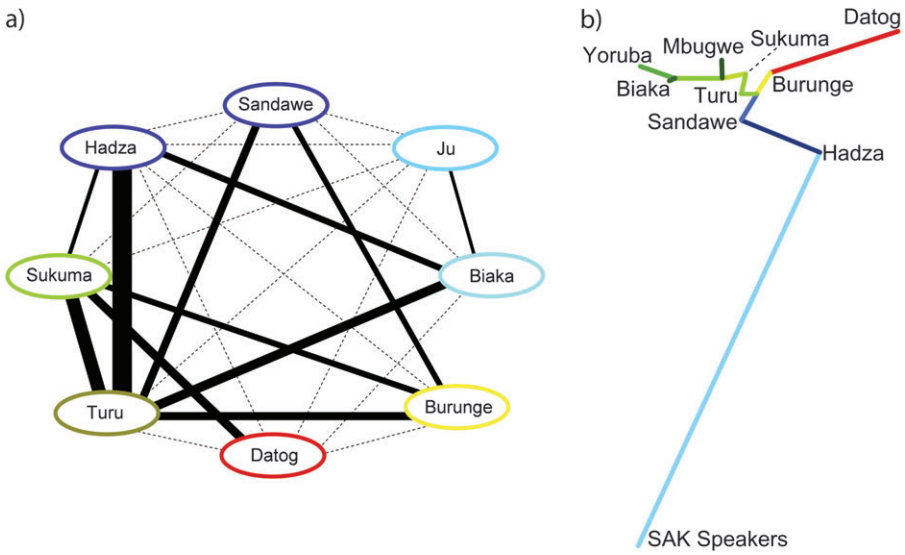


FIG. 8.—Graphical representations of pairwise population estimates of migration and genetic distance. (*a*) Migration rates (from IM analysis) among the populations (*M*) are represented by a network, with $M < 0.3$ not shown, $0.3 < M < 1$ indicated with dashed lines, and $M > 1$ indicated with a line whose thickness is proportional to the *M* estimates. (b) Neighbor-joining tree generated from $\delta\mu^2$ estimates of population divergence derived from Y chromosome STR data (Supplementary Table 2).

more robust and are presented in figure 8. Those estimates indicate low levels of exchange between the 2 Tanzanian click-speaking groups but relatively high levels of genetic exchange between each of the click-speaking populations and the Bantu-speaking populations.

In order to investigate the possibility that the sharing of an E3a lineage by Hadza and Sandawe reflects gene flow from Bantu speakers into each of these populations, we conducted a coalescent-based simulation study, generating and comparing simulated likelihoods for a range of demographic models (Pritchard et al. 1999). Table 3 provides simulated likelihoods given Y chromosome UEP and STR data for 24 models with varying population divergence times and migration rates. This analysis allowed us to rule out the models involving complete isolation of the 3 populations, as well as models involving a constant rate of migration between the populations since divergence. The models with the highest likelihoods are those involving separation of the Hadza and Sandawe ~15 kya followed by recent (last 5 kya) unidirectional gene flow from Bantu-speaking populations into each click-speaking population. Note that the small male sample size for the SAK-speaking population precluded consideration of that population in the simulated likelihood analysis.

A neighbor-joining tree derived from genetic distances ($\delta\mu^2$) inferred from the Y-STR data is consistent with linguistic relationships in that all Bantu-speaking populations are united by a single branch as shown in figure 8. Divergence among the click-speaking populations is far higher than the divergence among the Bantu-speaking populations, with the branch leading to the SAK-speaking group particularly long. As observed for mtDNA, the Hadza and Sandawe cluster more closely with the other Tanzanian populations than with the SAK-speaking group but are located nearer in the tree to the SAK-speaking group relative to the other Tanzanian populations.

Discussion

Relationships among the Click-Speaking Populations of Africa

The SAK speakers and the 2 eastern African click-speaking populations share ancient mtDNA and Y chromosome haplogroups; all 3 populations share an ancient Y chromosome haplogroup (B2b) that is rare elsewhere, and the SAK speakers and Sandawe share additional ancient haplogroups (Y chromosome A-M91, mtDNA L0d) not found among the Hadza (figs. 3 and 4). Although the Sandawe have a low frequency of the ancient Y chromosome A-M91 haplogroup, the Sandawe Y chromosomes belong to the M13-defined subhaplogroup of A present in other eastern/northeastern African populations rather than to any of the subhaplogroups of A specific to SAK speakers (e.g., defined by the M51 or M6 mutations; Underhill et al. 2001; Semino et al. 2002; fig. 4). Additionally, neither the Hadza nor Sandawe samples include the Y chromosome B2b subhaplogroups defined by mutations P6 and P7 that are frequent among the SAK speakers and Pygmy populations, respectively.

The mt L0d haplogroup, relatively frequent in most SAK-speaking populations, is present in our Tanzanian sample at low frequency in the Sandawe and in one individual from the neighboring Burunge population, with whom the Sandawe have exchanged individuals (fig. 3). The L0d lineages in Tanzania form a monophyletic subclade (fig. 5) with a T_{MRCA} of 23.4 kya (95% CI 9.4–50.7 kya), suggesting (coupled with monophyly), a minimum time of population divergence from the SAK-speaking populations. This estimate is slightly more recent than that obtained from whole-mtDNA genome sequences in these same individuals (T_{MRCA} of L0d in Tanzania of 30.6 ± 17.8 kya; Gonder et al. 2007). The T_{MRCA} of the 2 most similar Tanzanian and SAK L0d lineages is

Table 3
Simulated Likelihoods for a Range of 3-Population Models of Population History

Model[a]	H → S[b]		H → Y		S → Y		f_a[c]	f_{ss} ($\times 10^{-15}$)[d]	L_{sim} ($\times 10^{-19}$)[e]	Rank
CI	0	0	0	0	0	0	<0.00001	—	—	—
IM1	1	1	1	1	1	1	0.00012	—	—	—
IM2	2	2	2	2	2	2	0.00016	—	—	—
IM3	2	2	0	2	0	2	0.00054	0.010	0.052	8
IM4	1	1	0	2	0	2	0.00053	0	0	—
IM5	2	2	0	2	0	2	0.00020	0	0	—
IM6	5	5	0	2	0	2	0.00010	—	—	—
IM7	2	2	0	5	0	5	0.00004	—	—	—
IM8	0	0	0	5	0	5	0.00002	—	—	—
CIRM1	2	2	0	2	0	2	0.00002	—	—	—
CIRM2	1	1	0	2	0	2	0.00004	—	—	—
CIRM3	0	0	0	2	0	2	<0.00001	—	—	—
CIRM4	5	5	0	5	0	5	0.00030	0.040	0.119	6
CIRM5	2	2	0	5	0	5	0.00040	0.088	0.356	3
CIRM6	0	0	0	5	0	5	0.00016	0	0	—
CIRM7[f]	0	0	0	5	0	5	0.00022	4.460	9.380	1
CIRM8[g]	0	0	0	5	0	5	0.00020	0.075	0.151	5
CIRM9	0	2	0	5	0	5	0.00022	0.021	0.046	9
CIRM10	1	1	0	5	0	5	0.00026	0	0	—
CIRM11	0	1	0	5	0	5	0.00014	0	0	—
CIRM/IM1	1	1	0	5	0	5	0.00044	0	0	—
CIRM/IM2	2	2	0	5	0	5	0.00040	0.017	0.068	7
CIRM/IM3	0	1	0	5	0	5	0.00022	0.126	0.279	4
CIRM/IM4	0	2	0	5	0	5	0.00024	0.436	1.050	2

NOTE.—H, Hadza; S, Sandawe; Y, Bantu-speaking Yoruban population.

[a] Models are abbreviated as follows: CI, complete isolation of all populations after divergence; IM, isolation followed by continuous migration; CIRM, period of complete isolation of all populations followed by recent migration among all populations; CIRM/IM, complete isolation of Bantu-speaking population followed by recent migration into click-speaking populations (3 kya); continuous migration between click-speaking populations.

[b] Unless otherwise indicated, the population ancestral to H and S split from Y at 60 kya; H and S split at 20 kya. Columns 2–7 indicate migration rate from source population (left column) to sink population (right column).

[c] Frequency of ascertainment.

[d] For a given summary statistic, the frequency of ascertained runs within ±10% of the observed statistic.

[e] Product of f_a and f_{ss}.

[f] Hadza–Sandawe split at 15 kya.

[g] Hadza–Sandawe split at 10 kya.

estimated at ~58 kya, suggesting an upper bound for the time of population divergence between the SAK speakers and the Sandawe. Although the presence of L0d in the Sandawe and SAK establishes a unique connection between these populations, it is possible that L0d could be a shared ancestral trait, or symplesiomorphy, rather than a shared, derived character. This pattern is similar to that observed for the Y chromosome A-M91 haplogroup. The absence of the L0d haplogroup in the Hadza suggests a lack of contact and gene flow between the Hadza and the SAK-speaking populations; however, the absence of L0d may also reflect a recent population bottleneck in the Hadza suggested by demographic data (Blurton Jones et al. 1992). Because of their antiquity, these haplogroups (mtDNA L0d and Y chromosome A) provide no evidence of recent exchange or recent common ancestry (prior to ~35 kya) of these southern and eastern African populations.

The point estimates from maximum likelihood analyses of mtDNA variation indicate more migration between the Sandawe and SAK-speaking populations than between the Hadza and SAK speakers (M = 0.32–0.49 for Sandawe, M = 0.02 for Hadza), although the estimated time of divergence is quite ancient for all pairs of populations (t = ~44–50 kya for the Sandawe and 2 SAK-speaking populations and ~56 kya for the Hadza and each SAK-speaking population; table 1 and fig. 6). Interestingly, the neighbor-joining tree inferred from genetic distances based on Y chromosome data indicates that the Hadza are more genetically similar to the SAK-speaking group than are the Sandawe, although the Y chromosome sample is particularly small for the SAK-speaking population (fig. 8).

Both mtDNA and Y chromosome likelihood analyses are consistent with a closer genetic relationship between the Hadza and Sandawe than between either Tanzanian click-speaking group and the SAK click speakers (figs. 6 and 8). The mtDNA-based maximum likelihood estimate for the population divergence time of the Hadza and Sandawe is 21 kya (table 1), similar to the estimate of a divergence ~15 kya from the Y chromosome–simulated likelihood analysis (table 3). However, estimates of migration rates between the Hadza and Sandawe for both the mtDNA and Y chromosome data are quite low compared with other Tanzanian populations (figs. 6 and 8). Estimates of T_{MRCA} for specific mtDNA and Y chromosome haplogroups such as mtDNA L4g and Y chromosome B2b/E3b1* are also consistent with divergence of the Sandawe and Hadza >15 kya with very little recent migration.

The genetic data presented here suggest consideration of 3 scenarios of population history for the 3 click-speaking groups (fig. 9), none of which reflects the effect of gene

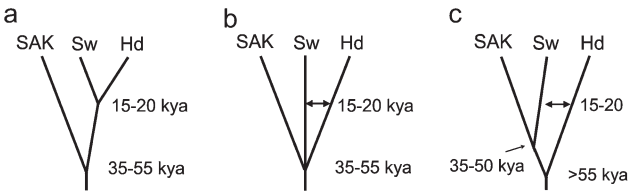


FIG. 9.—Simplified diagrams of population relationships suggested by mtDNA and Y chromosome data. Diagrams do not indicate migration that is estimated to have occurred within the last 5,000 years between click-speaking and nonclick-speaking populations. SAK, populations speaking languages classified as belonging to the Southern African Khoisan language families (see fig. 1); Sw, Sandawe; and Hd, Hadza. (*a*) Model in which initial divergence (~35–55 kya) is between the ancestor of SAK-speaking populations and a population ancestral to both Hadza and Sandawe. The Hadza and Sandawe diverge ~15–20 kya with little subsequent genetic exchange. (*b*) Model in which divergence among click-speaking populations is so ancient (~35–55 kya) that the sequence of initial divergence events is uncertain. In both models (*a*) and (*b*), sharing of L0d lineages by Sandawe and SAK speakers may reflect the loss of this lineage in the Hadza due to genetic drift. After a period of divergence, Hadza and Sandawe populations come into contact and experienced gene flow ~15–20 kya and then diverged again with little subsequent genetic exchange. (*c*) Model in which the Sandawe and SAK share a common ancestor ~35–50 kya. The Hadza originate from a lineage that diverged from the SAK/Sandawe lineage at least 55 kya, but came into contact with the Sandawe and experienced gene flow ~15–21 kya, and then diverged with little subsequent genetic exchange.

flow from neighboring populations, discussed below. Our results (specifically mtDNA MDIV analyses, L4g- and E3b1-M35*-based date estimates $[\delta\mu]^2$, and simulated likelihood analyses for the Y chromosome) are consistent with the possibility that the SAK speakers diverged from a population ancestral to both Hadza and Sandawe populations over 35 kya, and then the latter population split to form the Hadza and Sandawe population lineages ~15–21 kya, with little subsequent gene flow (fig. 9*a*). The results are also consistent with scenarios that consider the possibility of changing rates of migration between the Sandawe and Hadza, a possibility neglected by the IM analysis models. For example, the 3 populations may have diverged from one another >35 kya (where population divergences occurred so close to one another that the events can be summarized as a trifurcation), then the Hadza and Sandawe came into contact with one another and exchanged alleles roughly 15–20 kya, with little subsequent gene flow (fig. 9*b*). In both these scenarios, the mtDNA L0d lineages would have existed in the populations ancestral to all 3 groups and were lost subsequently in the Hadza. The latter lineage loss is quite likely given the estimated time depth of L0d common ancestry between the Sandawe and SAK (>58 kya) and the likelihood of a recent bottleneck in the Hadza population (Blurton Jones et al. 1992). Additionally, the observation in a Turkana population from northern Kenya of an L0d lineage that is phylogenetically close to the Tanzanian L0d lineages (fig. 5*b*) suggests that at one time the L0d haplogroup may have been more widespread across eastern Africa (Watson et al. 1997). A third possibility (fig. 9*c*; consistent with the sharing of the L0d haplogroup by the Sandawe and SAK speakers and the IM analyses for mtDNA) is that the SAK-speaking and Sandawe populations diverged from one another more recently, although still >35 kya, than either split from the Hadza population (>55 kya). As in model 9b, our estimates of common ancestry of the Hadza and Sandawe at 15–20 kya may reflect longer term isolation of these 2 populations with high levels of interaction (gene flow) around 15–20 kya, and little subsequent gene flow. Under this scenario, either the Sandawe and SAK speakers share the L0d haplogroup through a common ancestral population >35 kya (after separation from the Hadza) or they share the L0d haplogroup through a population ancestral to all click-speaking populations and the L0d haplogroups was lost subsequently in the Hadza population. These data provide no insight into whether populations ancestral to present day click-speaking populations originated in eastern Africa and migrated south, or vice versa. Although these 3 scenarios (*a–c*) differ, particularly in terms of the nature of the relationship between the Hadza and Sandawe, under any of the scenarios divergence of these 3 click-speaking populations occurred very deep in time.

The genetic pattern for the 3 click-speaking groups corresponds loosely with their geographic distribution, in that the 2 most geographically proximate populations (Hadza and Sandawe) are most genetically similar. However, the divergence time estimates for the 2 Tanzanian click-speaking populations are remarkably high given their geographic proximity; despite currently living only ~150 km apart, the Hadza and Sandawe appear to have had very low levels of genetic exchange for ~15–20 kya. Our data suggest that the Sandawe and Hadza hunter-gatherer populations were isolated prior to the arrival of agriculturalist and pastoralist populations into Tanzania within the last 4,000 years. The latter finding is consistent with conclusions reached by Destro-Bisol, Coia, et al. (2004) who proposed that western and eastern Pygmy populations were separated >18 kya, well before the arrival of Bantu-speaking groups. The difference between the Pygmy and Tanzanian cases is that the geographic distance separating the Hadza and Sandawe is far smaller than that separating the 2 Pygmy populations. Combined, these studies suggest that hunter-gatherer populations in sub-Saharan Africa became isolated from one another at some point between 15 and 60 kya, when the region was far more sparsely populated than today. A number of factors could have contributed to isolation of the southern and eastern African click-speaking groups, including a marked dry period in southern Africa at the height of the last glacial maximum, ~17–24 kya (Stokes et al. 1997; Lahr and Foley 1998; Mitchell 2002).

Impact of Recent Migration

Phylogeographic analyses of mtDNA and Y chromosome lineages indicate recent gene flow between both the Hadza and Sandawe and their neighboring populations. For example, we observe mtDNA L0a lineages shared by the Sandawe and Burunge (fig. 5*a*), and we observe several L3 lineages shared by the Hadza and Sukuma (fig. 5*d*). We also observe L4g lineages shared by the Hadza and Datog and other L4g lineages shared by the Sandawe, Turu, and Burunge (fig. 5*d*). Given the relatively high haplotype frequencies of the L4g haplogroup in the Hadza and Sandawe, it is possible that these lineages originated in

the Hadza and Sandawe populations and then were introduced via females into the neighboring agriculturalist and pastoralist populations. By contrast, the Y chromosome data suggest gene flow from the neighboring groups into the Hadza/Sandawe (e.g., E3a-M2). For example, the Hadza and Sandawe populations appear to have absorbed Y chromosome E3a lineages associated with Bantu speakers, and yet have contributed little in terms of the B2b lineages to their neighbors (fig. 7). These data suggest the possibility of a biased migration pattern, with higher female migration from hunter-gatherer groups into neighboring agriculturalist/pastoralist populations and higher male migration from agriculturalists/pastoralist populations into the hunter-gatherer populations. This pattern is consistent with other studies of hunter-gatherer populations in central Africa (Destro-Bisol, Donati, et al. 2004).

Likelihood analyses of both the mtDNA and Y chromosome data also indicate moderate levels of gene flow between the Hadza and Sandawe and their neighboring populations (table 1 and figs. 6 and 8). The Bantu-speaking populations are known to have had a geographically broad impact: archaeological remains record the spread of ancestors of today's Bantu-speaking peoples from western Africa throughout sub-Saharan Africa within the past 4,000 years (Newman 1995). Both mtDNA and Y chromosome data have been taken as evidence of the extensive genetic impact of these migrations (Salas et al. 2002; Destro-Bisol, Coia, et al. 2004). Indeed, the Hadza and Sandawe have mtDNA lineages (within L2a, L3b, L3e haplogroups) as well as Y chromosome lineages (within the E3a-M2 haplogroup) that likely reflect recent gene flow from Bantu-speaking populations into these groups. In addition, simulated likelihood analysis indicates that the Y chromosome data are consistent with gene flow from Bantu-speaking populations into both the Hadza and Sandawe population during the last few thousand years.

The SAK speakers are also likely to have experienced gene flow from neighboring Bantu-speaking populations. Barnard (1992) notes that the Khwe have morphological similarities with Bantu speakers, suggesting gene flow between these groups at some point in the past. The Ju|'hoansi sampled from the Kalahari Desert have been more isolated; previous Y chromosome and mtDNA studies revealed little admixture with Bantu-speaking populations (Lee 1993; Underhill et al. 2001). This difference between the Ju|'hoansi and Khwe suggests that SAK-speaking populations vary in the extent of genetic exchange with neighboring Bantu-speaking populations. Within the last few thousand years, click-speaking populations of both southern and eastern Africa have been the recipients of gene flow from neighboring Bantu-, Cushitic-, and/or Nilotic-speaking populations. Such gene flow may have obscured the relationships among the click-speaking groups, at least partially. For example, the mtDNA-based estimate of a low level of migration ($M < 0.40$) between the !Xun/Khwe and several Tanzanian populations, including the Sandawe (table 1 and fig. 7), may reflect gene flow from Bantu-speaking groups into both eastern and southern African populations.

Despite genetic exchange with neighboring populations, the Hadza and Sandawe populations each have maintained not only their respective click languages but also indigenous genetic lineages and, in the case of the Hadza, a hunting and gathering subsistence pattern. The Sandawe and Hadza are also fairly divergent from each other, especially in light of their geographic proximity. This genetic divergence is consistent with their deep linguistic divergence and may reflect greater geographic separation of the 2 populations in the more distant past.

History of African Click Languages

This study, based on 2 highly informative and independently inherited genetic regions (the mtDNA and Y chromosome), indicates that any connections between African click-speaking populations, in the form of common ancestry and/or migration, were quite ancient: >15 kya for the Sandawe and Hadza and between 35 and 55 kya for the Sandawe/Hadza and SAK. In addition, the Hadza and Sandawe are genetically more similar to their Nilotic-, Cushitic-, and Bantu-speaking neighbors than they are to the SAK-speaking population. However, they are also genetically more similar to the SAK speakers than are any of the other Tanzanian populations (figs. 6 and 8).

The Hadza sample introduced here is larger than, and largely independent of, the Hadza sample of Knight et al. (2003). This new Hadza data set supports the previous finding, based on mtDNA and Y chromosome analyses, that the Hadza are highly divergent from the SAK speakers with no evidence of genetic exchange over the past 40 kya (Knight et al. 2003). On the basis of the high level of genetic divergence between the Hadza and SAK speakers, Knight et al. (2003) concluded that if clicks arose only once, then they arose tens of thousands of years ago. However, the possibility remained that the Hadza acquired clicks through relatively recent interaction with a neighboring click-speaking population such as the Sandawe, who had never previously been studied at the DNA level.

Our analyses of mtDNA and Y chromosome variation of the Sandawe, and of a larger set of Hadza individuals, indicate very little recent genetic exchange between the Hadza and Sandawe or between either of these groups and the SAK-speaking population. These results are consistent with several scenarios for the relationships among click-speaking populations, as summarized above and in figure 9. If populations can influence each other linguistically without significant genetic exchange, then population history is not expected to correlate with models of linguistic history. However, the coupling of linguistic and genetic history is expected under a demographic–subsistence model of linguistic and genetic exchange (Renfrew 1992; Cavalli-Sforza et al. 1994). For example, in well-accepted cases of click borrowing, the southern African Bantu languages have borrowed click phonemes and have experienced accompanying gene flow from Khoisan populations, reflected by the presence of the A3b1-M51 haplogroup among the Zulu and Xhosa (Wood et al. 2005) and the L0d haplogroup among the Ronga and Tswa (Salas et al. 2002).

If we assume that click phoneme acquisition is unlikely to have occurred without common ancestry and/or genetic exchange, then inferences of population genetic

relationships have implications for linguistic history. Under any of the population history models described in figure 9, if clicks arose only once, then they are likely to have arisen over 35 kya. Under model 9c, clicks could have arisen in a population ancestral to the Sandawe and SAK, and subsequent genetic and linguistic exchange between the Hadza and Sandawe, on the order of 15–20 kya, then led to the current existence of click phonemes in both the Hadza and Sandawe languages. A second possibility (fig. 9*a* and *c*), although less likely given that click languages have arisen only once outside of Africa, is that clicks arose twice: once in the ancestral SAK population and once in the language of either the ancestral Sandawe/Hadza (fig. 9*a*) or Hadza (fig. 9*c*) populations sometime before 15 kya. More recent origins of clicks require at least 3 independent origin events in languages of each of the 3 populations. The most parsimonious interpretation of these data is that click phonemes arose on the order of tens of thousands of years ago in sub-Saharan Africa.

Interestingly, recent linguistic analyses suggest that the Sandawe language is distantly related to the Khoe (Central SAK) language family (Ambrose 1982; Elderkin 1982; Güldemann and Vossen 2000; Güldemann and Elderkin, forthcoming). Sharing of the mtDNA L0d haplogroup by the Sandawe and SAK speakers, as well as maximum likelihood estimates of time of population divergence and migration based on mtDNA, are consistent with a genetic connection, albeit deep, between these 2 groups (fig. 9*c*). The model presented in figure 9*c* is also consistent with linguistic data, indicating that the Hadza language is highly divergent from both the Sandawe and SAK languages (fig. 1; Heine and Nurse 2000). Note that the estimated time of divergence between the Sandawe and SAK based on genetic data, ~35–50 kya, is much older than the time depth at which most linguists are comfortable making language connections (i.e., <10 kya; Comrie 2000).

In summary, our data indicate that the southern and eastern African click-speaking populations share relatively rare Y chromosome and mtDNA haplogroups. However, our analyses suggest that population divergence and/or genetic exchange among SAK-speaking and eastern African click-speaking populations was quite ancient (>35 kya). Even within Tanzania, divergence of click-speaking populations is inferred to be remarkably old (~15–20 kya), consistent with linguists' conclusions (fig. 1) that any relationship between the Hadza and Sandawe languages is very deep, if detectable at all.

Supplementary Materials

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

Acknowledgments

We thank African participants who generously donated DNA samples so that we might learn more about their population history. We thank Karuna Panchapakesan, Megan Baumgartner, Ruth Emerson, Ann Horsburgh, Alice Lin, Dorothy Louis, Trini Miguel, and Wesley Sutton for technical assistance. We thank Dr M.J. Kotze for contributing the Khwe DNA samples. We thank Kweli Powell and Alain Froment for assistance with DNA sample collection and Nigel Crawhall, Christopher Ehret, and Alison Brooks for helpful discussion. We thank Salum Juma Deo, Paschal Lufungulo, Waja Ntandu, Dr Audax Mabulla, Jeannette Hanby, and David Bygott for their assistance with field work in Tanzania. We thank Michael Cummings and Adam Bazinet for assistance with running MDIV analyses on the grid computing system at University of Maryland. This study was funded by L.S.B. Leakey Foundation grants to S.A.T. and J.L.M.; National Science Foundation (NSF) grant BCS-9905574 to J.L.M.; NSF grant DEB-0108541 to E. Hadly and J.L.M.; and NSF grants BCS-0196183 and BCS-0552486 and National Institutes of Health (NIH) grant R01GMS076637 to S.A.T.; a Wenner Gren Foundation grant, Packard and Burroughs Wellcome Foundation Career Awards to S.A.T.; NIH grant GM28428 to J.L.M; F.A.R. is supported by NIH grant F32HG03801; H.M. is supported by NSF grant IGERT-9987590.

Literature Cited

Ambrose SH. 1982. Archaeology and linguistic reconstructions of history in Eastern Africa. In: Ehret C, Posnansy M, editors. Archaeological and linguistic reconstruction of African history. Berkeley (CA): University of California Press. p. 104–157.

Barnard A. 1992. Hunters and herders of southern Africa. Cambridge: Cambridge University Press.

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

Bazinet AL, Cummings MP. Forthcoming. The Lattice Project: a grid research and production environment combining multiple grid computing models in distributed & Grid Computing. In: Weber MHW, editor. Science made transparent for everyone. Marburg (Germany): Tectum.

Bazinet AL, Myers DS, Fuetsch J, Cummings MP. 2007. Grid services base library: a high-level, procedural application program interface for writing Globus-based grid services. Future Generation Computer Systems. 23:517–522.

Blurton Jones NG, Smith LC, O'Connell JF, Hawkes K, Kamuzora CL. 1992. Demography of the Hadza, an increasing and high density population of savanna foragers. Am J Phys Anthropol. 89:159–181.

Bräuer G. 1978. The morphological differentiation of anatomically modern man in Africa, with special regard to recent finds from East Africa. Z Morphol Anthropol. 69:266–292.

Cann HM, de Toma C, Cazes L, et al. (41 co-authors). 2002. A human genome diversity cell line panel. Science. 296: 261–262.

Cavalli-Sforza LL, Piazza A, Menozzi P. 1994. History and geography of human genes. Princeton (NJ): Princeton University Press.

Cavalli-Sforza LL, Piazza A, Menozzi P, Mountain J. 1988. Reconstruction of human evolution: bringing together genetic, archaeological, and linguistic data. Proc Natl Acad Sci USA. 85:6002–6006.

Chen YS, Olckers A, Schurr TG, Kogelnik AM, Huoponen K, Wallace DC. 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.

Comrie B. 2000. Is there a single time depth cut-off point in historical linguistics?. In: Renfrew C, McMahon A, Trask L, editors. Time depth in historical linguistics. Cambridge: McDonald Institute for Archeological Research. p. 33–43.

Cruciani F, La Fratta R, Santolamazza P, et al. (19 co-authors). 2004. Phylogeographic analysis of haplogroup E3b (E-M215) y chromosomes reveals multiple migratory events within and out of Africa. Am J Hum Genet. 74:1014–1022.

Cruciani F, Santolamazza P, Shen P, et al. (16 co-authors). 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.

Cummings MP, Huskamp JC. 2005. Grid computing. Educause Rev. 40:116–117.

Destro-Bisol G, Coia V, Boschi I, Verginelli F, Caglia A, Pascali V, Spedini G, Calafell F. 2004. The analysis of variation of mtDNA hypervariable region 1 suggests that Eastern and Western Pygmies diverged before the Bantu expansion. Am Nat. 163:212–226.

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

Ehret C. 2000. Language and history. In: Heine B, Nurse D, editors. African languages: an introduction. Cambridge: Cambridge University Press. p. 272–297.

Elderkin ED. 1982. On the classification of Hadza. Sprach Gesch Afr. 4:67–82.

Excoffier L, Novembre J, Schneider S. 2000. SIMCOAL: a general coalescent program for the simulation of molecular data in interconnected populations with arbitrary demography. J Hered. 91:506–509.

Forster P, Harding R, Torroni A, Bandelt HJ. 1996. Origin and evolution of native American mtDNA variation: a reappraisal. Am J Hum Genet. 59:935–945.

Forster P, Rohl A, Lunnemann P, Brinkmann C, Zerjal T, Tyler-Smith C, Brinkmann B. 2000. A short tandem repeat-based phylogeny for the human Y chromosome. Am J Hum Genet. 67:182–196.

Forster P, Torroni A, Renfrew C, Rohl A. 2001. Phylogenetic star contraction applied to Asian and Papuan mtDNA evolution. Mol Biol Evol. 18:1864–1881.

Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW. 1995. Genetic absolute dating based on microsatellites and the origin of modern humans. Proc Natl Acad Sci USA. 92: 6723–6727.

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

Greenberg J. 1963. The languages of Africa. Bloomington: Indiana University Press.

Griffiths RC, Tavaré S. 1997. Computational methods for the coalescent. In: Donnelly P, Tavaré S, editors. Progress in population genetics and human evolution. Berlin (Germany): Springer Verlag. p. 165–182.

Güldemann T, Elderkin ED. Forthcoming. On external genealogical relationships of the Khoe family. In: Brenzinger M, König C, editors. Khoisan languages and linguistics: the Riezlern symposium 2003. Köln (Germany): Rüdiger Köppe.

Güldemann T, Vossen R. 2000. Khoisan. In: Heine B, Nurse D, editors. African languages. Cambridge: Cambridge University Press. p. 99–122.

Hale K. 1992. Language endangerment and the human value of linguistic diversity. Language. 68:35–42.

Hasegawa M, Di Rienzo A, Kocher TD, Wilson AC. 1993. Toward a more accurate time scale for the human mitochondrial DNA tree. J Mol Evol. 37:347–354.

Heine B, Nurse D. 2000. African languages: an introduction. Cambridge: Cambridge University Press.

Hey JD, Nielsen R. 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. perimilis*. Genetics. 167:747–760.

Horai S, Hayasaka K, Kondo R, Tsugane K, Takahata N. 1995. Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. Proc Natl Acad Sci USA. 92:532–536.

Iliffe J. 1979. A modern history of Tanganyika. Cambridge: Cambridge Univiversity Press.

International Human Genome Sequencing Consortium. 2004. YCC: A nomenclature system for the tree of Y chromosomal binary haplogroups. Genome Research. 12:339–348.

Kayser M, Krawczak M, Excoffier L, et al. (12 co-authors). 2001. An extensive analysis of Y-chromosomal microsatellite haplotypes in globally dispersed human populations. Am J Hum Genet. 68:990–1018.

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

Knight A, Underhill PA, Mortensen HM, Zhivotovsky LA, Lin AA, Henn BM, Louis D, Ruhlen M, Mountain JL. 2003. African Y chromosome and mtDNA divergence provides insight into the history of click languages. Curr Biol. 13:464–473.

Lahr MM, Foley RA. 1998. Towards a theory of modern human origins: geography, demography, and diversity in recent human evolution. Am J Phys Anthropol. (Suppl 27):137–176.

Lee RB. 1993. The Dobe Jul'hoansi. Orlando (FL): Harcourt Brace Publishers.

Luis JR, Rowold DJ, Regueiro M, Caeiro B, Cinnio Iu C, Roseman C, Underhill PA, Cavalli-Sforza LL, Herrera RJ. 2004. The Levant versus the Horn of Africa: evidence for bidirectional corridors of human migrations. Am J Hum Genet. 74:532–544.

Meyer S, Weiss G, von Haesler A. 1999. Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. Genetics. 152:1103–1110.

Mitchell P. 2002. The archaeology of Southern Africa. Cambridge: Cambridge University Press.

Morris AG. 2002. Isolation and the origin of the Khoisan: late Pleistocene and early Holocene human evolution at the southern end of Africa. Hum Evol. 17:231–240.

Myers DS, Cummings MP. 2003. Necessity is the mother of invention: a simple grid computing system using commodity tools. J Parallel Distrib Comput. 63:578–589.

Newman J. 1970. The ecological basis for subsistence change among the Sandawe of Tanzania. Washington (DC): National Academy of Sciences-National Research Council.

Newman J. 1995. The Peopling of Africa. New Haven: Yale University Press.

Nielsen R, Wakeley J. 2001. Distinguishing migration from isolation: a Markov chain Monte Carlo approach. Genetics. 158:885–896.

Nurse GT, Weiner JS, Jenkins T. 1985. The peoples of southern Africa and their affinities. Oxford: Clarendon Press.

Pereira L, Macaulay V, Torroni A, Scozzari R, Prata M-J, Amorim A. 2001. Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade. Ann Hum Genet. 65:439–458.

Polzin T, Daneschmand SV. 2003. On Steiner trees and minimum spanning trees in hypergraphs. Operations Res Lett. 31:12–20.
Posada D, Crandall KA. 2001. Intraspecific gene genealogies: trees grafting into networks. Trends Ecol Evol. 16:37–45.
Pritchard JK, Seielstad MT, Perez-Lezaun A, Feldman MW. 1999. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. Mol Biol Evol. 16:1791–1798.
Reed FA, Tishkoff SA. 2006. African human diversity, origins and migrations. Curr Opin Genet Dev. 16:597–605.
Renfrew C. 1992. Archaeology, genetics and linguistic diversity. Man. 7:445–478.
Renfrew A. 1992. Archaeology, genetics and linguistic diversity. Man. 7:445–478.
Rosa A, Brehm A, Kivisild T, Metspalu E, Villems R. 2004. MtDNA profile of West Africa Guineans: towards a better understanding of the Senegambia region. Ann Hum Genet. 68:340–352.
Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW. 2002. Genetic structure of human populations. Science. 298:2381–2385.
Ruhlen MA. 1991. Guide to the world's languages. Stanford (CA): Stanford University Press.
Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 4:406–425.
Salas A, Richards M, De la Fe T, Lareu MV, Sobrino B, Sanchez-Diz P, Macaulay V, Carracedo A. 2002. The making of the African mtDNA landscape. Am J Hum Genet. 71:1082–1111.
Salas A, Richards M, Lareu MV, Scozzari R, Coppa A, Torroni A, Macaulay V, Carracedo A. 2004. The African diaspora: mitochondrial DNA and the Atlantic slave trade. Am J Hum Genet. 74:454–465.
Sands B. 1998. The linguistic relationship between Hadza and Khoisan. In: Schladt M, editor. Language, identity and conceptualization among the Khoisan. Köln (Germany): Rudiger Kupper Verlag. p. 266–283.
Scozzari R, Cruciani F, Santolamazza P, et al. (17 co-authors). 1999. Combined use of biallelic and microsatellite Y-chromosome polymorphisms to infer affinities among African populations. Am J Hum Genet. 65:829–846.
Semino O, Magri C, Benuzzi G, et al. (16 co-authors). 2004. Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the neolithization of Europe and later migratory events in the Mediterranean area. Am J Hum Genet. 74:1023–1034.
Semino O, Santachiara-Benerecetti AS, Falaschi F, Cavalli-Sforza LL, Underhill PA. 2002. Ethiopians and Khoisan share the deepest clades of the human Y-chromosome phylogeny. Am J Hum Genet. 70:265–268.
Stokes S, Thosmas DSG, Washington R. 1997. Multiple episodes of aridity in southern Africa since the last interglacial period. Nature. 388:154–158.
Stringer CB, Cornish L, Stuart-Macadam P. 1985. Preparation and further study of the Singa skull from Sudan. Bull Br Mus Nat Hist Geol. 38:347–358.
Tishkoff SA, Williams SM. 2002. Genetic analysis of African populations: human evolution and complex disease. Nat Rev Genet. 3:611–621.
Tobias PV. 1964. Bushman hunter-gatherers: a study in human ecology. In: Davis DH, editor. Ecological studies in Southern Africa. Den Haag (The Netherlands): W. Junk. p. 67–86.
Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonne-Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, Kidd JR, Mehdi SQ, Seielstad MT, Wells RS, Piazza A, Davis RW, Feldman MW, Cavalli-Sforza LL, Oefner PJ. 2000. Y chromosome sequence variation and the history of human populations. Nat Genet. 26:358–61.
Underhill PA, Passarino G, Lin AA, Shen P, Mirazon Lahr M, Foley RA, Oefner PJ, Cavalli-Sforza LL. 2001. The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. Ann Hum Genet. 65:43–62.
Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC. 1991. African populations and the evolution of human mitochondrial DNA. Science. 253:1503–1507.
Wakeley J. 1993. Substitution rate variation among sites in hypervariable region 1 of human mitochondrial DNA. J Mol Evol. 37:613–623.
Ward RH, Frazier BL, Dew-Jager K, Paabo S. 1991. Extensive mitochondrial diversity within a single Amerindian tribe. Proc Natl Acad Sci USA. 88:8720–8724.
Watson E, Forster P, Richards M, Bandelt HJ. 1997. Mitochondrial footprints of human expansions in Africa. Am J Hum Genet. 61:691–704.
Wood ET, Stover DA, Ehret C, et al. (11 co-authors). 2005. Contrasting patterns of Y chromosome and mtDNA variation in Africa: evidence for sex-biased demographic processes. Eur J Hum Genet. 13:867–876.
Zhivotovsky LA, Underhill PA, Cinnioglu C, et al. (17 co-authors). 2004. The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. Am J Hum Genet. 74:50–61.

Lisa Matisoo-Smith, Associate Editor

Accepted July 3, 2007