

Genetic evidence for a higher female migration rate in humans

Mark T. Seielstad¹, Eric Minch² & L. Luca Cavalli-Sforza²

Mitochondrial DNA and the Y chromosome have been used extensively in the study of modern human origins and other phylogenetic questions, but not in the context of their sex-specific modes of transmission. mtDNA is transmitted exclusively by females, whereas the Y chromosome is passed only among males. As a result, differences in the reproductive output or migration rate of males and females will influence the geographic patterns and relative level of genetic diversity on the Y chromosome, autosomes and mtDNA (ref. 1). We have found that Y chromosome variants tend to be more localized geographically than those of mtDNA and the autosomes²⁻⁵. The fraction of variation within human populations for Y chromosome single nucleotide polymorphisms (SNPs) is 35.5%, versus 80-85% for the autosomes and mtDNA (refs 6-8). A higher female than male migration rate (*via* patrilocality, the tendency for a wife to move into her husband's natal household) explains most of this discrepancy, because diverse Y chromosomes would enter a population at a lower rate than mtDNA or the autosomes. Polygyny may also contribute, but the reduction of variation within populations that we measure for the Y chromosome, relative to the autosomes and mitochondrial DNA, is of such magnitude that differences in the effective population sizes of the sexes alone are insufficient to produce the observation.

The geographic distribution of Y chromosome SNPs was compared with previously published studies of autosomal and mitochondrial DNA variation⁶⁻⁸ (Table 1). Theory based on the island model of migration indicates that F_{ST} (approximated as the variation among populations) is equal to $1/(1+4Nv)$ for diploid systems, and to $1/(1+Nv)$ for haploid systems (that is, mitochondrial and Y chromosome data), where N is the effective population size and v is the sum of migration (m) and mutation (μ ; more exactly, $v=m+\mu-m\mu$; ref. 9). The mutation rate of 3.5×10^{-8} per site, per year for mtDNA is approximately 10-20 times higher than the autosomal or Y-chromosomal rates of 1.9×10^{-9} - 5.4×10^{-9} (refs 10,11). The mutation rates for all three systems, however, are substantially lower than any estimates of the migration rate¹², and are unlikely to contribute to major differences among these three genetic systems. We assume that v is equal to m .

Estimates of Nv from F_{ST} between world populations are shown (Table 2). If N is the same for males and females, the female migration rate would exceed the male migration rate by a factor of 7.96. As expected, the autosomes displayed a value of Nv , which is intermediate between the values calculated from mtDNA and Y chromosomes (and close to their geometric mean).

We compared data for autosomal and Y-chromosomal microsatellites (short tandem repeat polymorphisms) in 14 African populations (Table 3). The fraction of the total variance found within populations was higher for the autosomes than for the Y chromosome. An analogue of F_{ST} for microsatellites (R_{ST} ; refs 13,14) indicated an Nv for the autosomes of 9.37, versus 4.26 for the Y. Their ratio of 2.20 was nearly the same as the ratio reported for SNPs (Table 2).

Another approach, using the regression of genetic distance onto geographic distance (Fig. 1), leads to the same conclusion. The slope for the Y chromosome data was higher than for the mtDNA data by a factor of 8.22 ($P < 0.05$), indicating a more rapid increase of genetic distance for the Y chromosome. This, too, was in close agreement with the F_{ST} analysis presented (Table 2).

Although a difference in the female versus male migration rate appears to be the most likely explanation for these results, several other factors may contribute. Natural selection could produce this pattern of geographically localized Y chromosome variation by eliminating variation from within populations. Evidence has been presented against recent Y chromosome selective sweeps in humans¹¹, and our own data fail to indicate the significant involvement of natural selection.

Polygyny or higher rates of male mortality might also reduce the level of Y chromosome variation within populations relative to the autosomes and mtDNA. In strictly monogamous populations, the effective population size of autosomal genes is four times greater than for the haploid Y chromosome and mtDNA, because the autosomes are diploid and transmitted by both sexes. Polygyny decreases the effective population size of the Y chromosome (N_Y), although the autosomal N is also reduced by polygyny, to $4MF/(M+F)$, where M and F are the number of males and females, respectively. It can be shown that the ratio of the autosomal to Y-chromosomal N is approximately equal to $\frac{2+2R}{R}$ where R is the ratio of married men to married women in a population.

For an extreme value of $R=0.54$, where 53% of the men of a population are married and have 2.6 wives on average¹⁵, N_Y differs from N_{au} only by a factor of 5.7, instead of the expected factor of 4. We have calculated a more realistic global value of N_{au} to N_Y of 4.27 from data on a sample of 66 populations¹⁶. This is a relatively minor, and probably undetectable, deviation from the expected ratio of 4. Although some two-thirds of traditional societies may allow or encourage polygyny, there are several reasons to expect our estimates to exaggerate its effects. First, very few males (typically fewer than 20%) control sufficient

Table 1 • Single nucleotide polymorphism diversity

Genetic system	within populations	within continents	between continents
autosomes	85.6 (0.20)	5.7 (1.60)	8.8 (1.42)
mtDNA	81.4 (8.84)	6.1 (4.32)	12.5 (11.01)
Y chromosome	35.5 (10.43)	11.8 (2.82)	52.7 (10.88)

Apportionment of diversity for autosomal, mitochondrial and Y-chromosomal SNPs. Autosomal figures result from the average of several studies, as described⁷. The values for mitochondrial DNA are from re-analysis of the RFLP data as described⁸. The Y chromosome analysis was performed on previously published data⁵. All values are expressed as percentages. Standard errors derived from standard deviations (for autosomal data) or from bootstrapping of haplotypes of mtDNA and Y chromosome data are shown in parentheses. The Behrens-Fisher test was used to test for significance.

¹Program for Population Genetics, Harvard School of Public Health, 665 Huntington Avenue, Boston, Massachusetts 02115, USA. ²Department of Genetics, Stanford University School of Medicine, Stanford, California 94305-5120, USA. Correspondence should be addressed to M.T.S. (e-mail: mark@ppg.harvard.edu).

Table 2 • F_{ST} , N_v and their ratios for SNP data

Genetic system	F_{ST}	N_v	N_v ratios
mtDNA	0.186	4.38	2.94 (mtDNA:autosomes)
autosomes	0.144	1.49	7.96 (mtDNA:Y)
Y chromosome	0.645	0.550	2.71 (autosomes:Y)

Table 3 • Microsatellite diversity

Genetic system	within populations	within groups	between groups
29 autosomal microsatellite loci	97.8	0.19	2.03
10 Y chromosome microsatellite loci	83.5	13.95	2.52

Apportionment of diversity for autosomal and Y-chromosomal microsatellite polymorphisms in 14 African populations. Values are expressed as percentages, and population groups are as defined in Methods.

resources to support more than one wife. Second, approximately one-third of human societies practice sororal polygyny, in which additional wives must be sisters of the first wife. Sisters share identical mitochondrial genomes, and one-half of their autosomal genes (on average). In such cases, the Y chromosome should experience no reduction in effective population size relative to mtDNA, and its reduction relative to the autosomes should be much lower than if wives were unrelated. Third, we assume all women are equally fertile, when, in fact, women in polygynous marriages tend to be less fertile than their monogamous counterparts^{17–19}.

The conclusion that males migrate less than females is not immediately obvious, as men do travel more than women in many societies. Among the Biaka Pygmies, the exploration range of men is 1.8 times greater than that of women²⁰. The most important parameter from a genetic perspective is the distance between the birthplace of parents and children, and this is largely determined when one spouse leaves the natal household to join the other. Marriages in agricultural economies (and probably earlier foraging economies as well²¹), where the land is inherited by sons, are typically patrilocal. Thus women tend to be more mobile than men with respect to transgenerational movements. Precise estimates of the magnitude of the difference in migration rates produced by patrilocality are difficult to obtain, but approximately 70% of societies are patrilocal^{22,23}.

There seem to be no comparable data among primates, although some may be forthcoming. The closest data available today are perhaps in macaques, where a contrasting pattern has been observed. Female macaques generally remain in their natal groups, whereas most (if not all) males disperse to join others. Although Y-chromosomal data from macaques are not available, a much smaller fraction of mitochondrial variation (approx-

mately 9% of the total) is found within species when compared with the distribution of autosomal variation (where approximately 91% of the total is distributed within species²⁴). By some measures, macaques are polygynous, with most offspring in a social group fathered by a single male in any particular year²⁵. However, females rarely bear offspring with the same male on more than one occasion²⁵, so the relative effective population sizes of males and females are not precisely known. If the effective population size of males is in fact smaller than for females, as some expect²⁵, this provides further evidence against an extensive contribution of polygyny to our observations from humans.

Our data on humans are consistent in showing that a significantly greater fraction of the total variance is distributed between populations for the Y chromosome relative to both the autosomes and mitochondrial DNA. Our analysis indicates that a lower transgenerational migration rate of males compared with that of females must be the most significant factor in producing this observation. Polygyny makes a modest contribution, but it appears to account for less than 10% of the discrepancy.

Methods

Population samples. DNA was analysed from 14 African populations. Informed consent was obtained from all DNA donors. The number of complete Y chromosome haplotypes is indicated in parentheses for each population. Sudan: samples from the Hadendowah and Beni Amer sections of the Beja (6) were collected in and around Kassala. Eleven populations from Ethiopia were sampled: Konso (8) from Konso town; Tsamako (5) and Ongota (9) near Weyto; Hamar (5) near Turmi; Dasenech (5) in Omorate; Dizi (4) and Surma (11) near Maji; Bume (or Nyangatom; 11) at Kibish; Bench (8) in Mizan Teferi; and Majangir (10) near Tepi. In Mali, collections from three populations were made: Dogon (7) and Peulh near Bandiagara (14.361 °N, 3.647 °W); Tuareg (9) near Timbuktu (16.661 °N, 3.240 °W); and Songhai (5) near Gao (16.287 °N, 0.044 °W). Three groups were defined in calculating the apportionment of diversity presented (Table 3): group 1, Dizi, Dasenech, Hamar, Bench, Ongota, Beja, Tsamako and Konso; group 2, Nyangatom, Majangir and Surma; group 3, Dogon, Tuareg and Songhai.

DNA extraction. DNA was extracted from whole blood (5 ml) drawn in EDTA anticoagulant. Extraction was begun in the field as quickly as possible (but always less than 2 d after collection) using a modified 'salting out' procedure²⁶.

Y microsatellites. Eight sets of primers were used to amplify 10 microsatellite loci on the Y chromosome: *DYS19*, *DYS385I/II*, *DYS388*, *DYS389I/II*, *DYS390*, *DYS391*, *DYS392* and *DYS393*. Two of the primer sets typically produced two bands (*DYS385* and *DYS389*). In many of the analyses, the two bands of *DYS385* and *DYS389* were treated as if from separate loci.

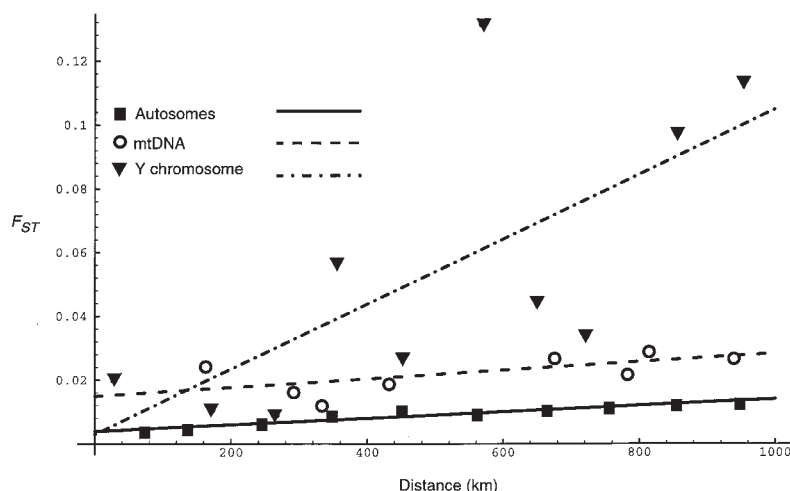


Fig. 1 Regressions of genetic distance (F_{ST}) onto geographic distance (km) in Europe. Three independent data sets were used (Y chromosome²⁸, mtDNA (ref. 29) and classical markers³⁰). The slope for the Y chromosome (1.108×10^{-4}) is significantly greater than for the autosomes (1.014×10^{-5} ; $P=0.0095$) and mtDNA (1.348×10^{-5} ; $P=0.0269$).



Autosomal microsatellites. Primers for the following 29 loci were used (ABI Prism linkage mapping set). Primers for all members of a panel were multiplexed in the PCR and run on a single gel. Panel 11, *D7S513*, *D7S550*, *D7S640*, *D7S657*, *D8S260* and *D8S514*; panel 12, *D7S519*, *D7S531*, *D8S270*, *D8S277*, *D8S284* and *D8S550*; panel 21, *D15S117*, *D15S130*, *D15S131*, *D15S153*, *D15S165*, *D16S411*, *D16S503*, *D16S515* and *D16S520*; panel 25, *D19S210*, *D19S220*, *D19S420*, *D20S100*, *D20S115*, *D20S117*, *D20S171* and *D20S196*.

PCR amplification and allele size determination. Microsatellite loci were amplified by PCR in reactions (5 μ l) according to the manufacturer's protocols. Primers were multiplexed as noted above. Reactions included the Taqstart (Clontech) anti-Taq antibody.

We used a 'touchdown' cycling regime, beginning with 14 cycles of successive 0.5 °C decreases in annealing temperature (63–56.5 °C). Twenty cycles with a constant 56 °C annealing temperature followed. Each step lasted 30 s, and a denaturing temperature of 94 °C and extension temperature of 72 °C were used throughout.

We ran PCR products on an automated DNA sequencer (ABI373a) with GS-350 or GS-500 dye-labelled size standard. Allele sizes were determined with GS analysis software (ABI).

The apportionment of diversity was calculated using the analysis of molecular variance (AMOVA), implemented in the Arlequin program^{8,27}. We calculated all genetic distances, F_{ST} and R_{ST} using Microsat (E. Minch, available on request).

Acknowledgements

We thank all the DNA donors who participated in this project. Laboratory work was funded by NIH grant GM28428 to L.L.C.-S. The collection of DNA samples in Ethiopia, Sudan and Mali was supported by the Arthur Green Fund of Harvard University and the L.S.B. Leakey Foundation, with the assistance of E. Bekele, M. Ibrahim, M. Traoré and A. Touré. M.T.S. was a U.S. National Science Foundation Predoctoral Fellow. We thank M. Feldman for helpful advice and discussion.

Received 22 May; accepted 21 September 1998.

- Salem, A.H., Badr, F.M., Gaballah, M.F. & Paabo, S. The genetics of traditional living: Y-chromosomal and mitochondrial lineages in the Sinai Peninsula. *Am. J. Hum. Genet.* **59**, 741–743 (1996).
- Seielstad, M.T. *et al.* Construction of human Y-chromosomal haplotypes using a new polymorphic A to G transition. *Hum. Mol. Genet.* **3**, 2159–2161 (1994).
- Ruiz-Linares, A. *et al.* Geographic clustering of human Y-chromosome haplotypes. *Ann. Hum. Genet.* **60**, 401–408 (1996).
- Underhill, P.A., Jin, L., Zemans, R., Oefner, P.J. & Cavalli-Sforza, L.L. A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history. *Proc. Natl Acad. Sci. USA* **93**, 196–200 (1996).
- Underhill, P.A. *et al.* Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res.* **7**, 996–1005 (1997).
- Lewontin, R.C. The apportionment of human diversity. *Evol. Biol.* **6**, 381–398 (1972).
- Barbujani, G., Magagni, A., Minch, E. & Cavalli-Sforza, L.L. An apportionment of human DNA diversity. *Proc. Natl Acad. Sci. USA* **94**, 4516–4519 (1997).
- Excoffier, L., Smouse, P.E. & Quattro, J.M. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**, 479–491 (1992).
- Cavalli-Sforza, L.L. & Bodmer, W.F. *The Genetics of Human Populations* (Freeman, San Francisco, 1971).
- Miyata, T. *et al.* Molecular clock of silent substitution: at least six-fold preponderance of silent changes in mitochondrial genes over those in nuclear genes. *J. Mol. Evol.* **19**, 28–35 (1982).
- Hammer, M.F. A recent common ancestry for human Y chromosomes. *Nature* **378**, 376–378 (1995).
- Wijsman, E.M. & Cavalli-Sforza, L.L. Migration and genetic population structure with special reference to humans. *Annu. Rev. Ecol. Syst.* **15**, 279–301 (1984).
- Slatkin, M. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**, 457–462 (1995).
- Feldman, M.W., Kumm, J. & Pritchard, J.K. Mutation and migration in models of microsatellite evolution. In *Microsatellites, Evolution and Applications* (eds Goldstein, D.J. & Schloetterer, C.) (Oxford University Press, Oxford, in press).
- Thomas, J.M.C. *Les Ngbaka de la Lobaye* (Mouton, The Hague, 1963).
- White, D.R. Rethinking polygyny. *Curr. Anthropol.* **29**, 529–572 (1988).
- Dorjahn, V. R. in *Continuity and Change in African Culture* (eds Herskovits, M.J. & Bascomb, W.R.) 87–112 (University of Chicago Press, Chicago, 1959).
- Hrdlicka, A. Fecundity in Sioux women. *Am. J. Phys. Anthropol.* **16**, 81–90 (1931).
- Isaac, B. Female fertility and marital form among the Mende of rural Upper Bambara Chiefdom, Sierra Leone. *Ethnology* **19**, 297–313 (1980).
- Hewlett, B., van de Koppel, J.M.H. & Cavalli-Sforza, L.L. Exploration ranges of Aka Pygmies of the Central African Republic. *Man* **17**, 418–430 (1982).
- Ember, C.R. Myths about hunter-gatherers. *Ethnology* **17**, 439–448 (1978).
- Burton, M.L., Moore, C.C., Whiting, J.W.M. & Romney, A.K. Regions based on social structure. *Curr. Anthropol.* **37**, 87–123 (1996).
- Murdock, G. P. *Ethnographic Atlas* (University of Pittsburgh Press, Pittsburgh, 1967).
- Melnick, D.J. & Hoelzer, G.A. Differences in male and female macaque dispersal lead to contrasting distributions of nuclear and mitochondrial DNA variation. *Int. J. Primatol.* **13**, 379–393 (1992).
- Keane, B., Dittus, W.P.J. & Melnick, D.J. Paternity assessment in wild groups of toque macaques *Macaca sinica* at Polonnaruwa, Sri Lanka using molecular markers. *Mol. Ecol.* **6**, 267–282 (1997).
- Miller, S.A., Dykes, D.D. & Polesky, H.F. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* **16**, 1215 (1988).
- Schneider, S., Kueffer, J. M., Roessler, D. & Excoffier, L. Arlequin ver. 1.1, University of Geneva (1997).
- Semino, O., Passarino, G., Brega, A., Fellous, M. & Santachiara-Benerecetti, A.S. A view of the Neolithic demic diffusion in Europe through two Y chromosome-specific markers. *Am. J. Hum. Genet.* **59**, 964–968 (1996).
- Richards, M. *et al.* Paleolithic and Neolithic lineages in the European mitochondrial gene pool. *Am. J. Hum. Genet.* **59**, 185–203 (1996).
- Cavalli-Sforza, L.L., Menozzi, P. & Piazza, A. *The History and Geography of Human Genes* (Princeton University Press, Princeton, New Jersey, 1994).

