

# Global patterns of human mitochondrial DNA and Y-chromosome structure are not influenced by higher migration rates of females versus males

Jason A Wilder<sup>1,2</sup>, Sarah B Kingan<sup>1</sup>, Zahra Mobasher<sup>1</sup>, Maya Metni Pilkington<sup>3</sup> & Michael F Hammer<sup>1,2</sup>

Global-scale patterns of human population structure may be influenced by the rate of migration among populations that is nearly eight times higher for females than for males. This difference is attributed mainly to the widespread practice of patrilocality, in which women move into their mates' residences after marriage<sup>1</sup>. Here we directly test this hypothesis by comparing global patterns of DNA sequence variation on the Y chromosome and mitochondrial DNA (mtDNA) in the same panel of 389 individuals from ten populations (four from Africa and two each from Europe, Asia and Oceania). We introduce a new strategy to assay Y-chromosome variation that identifies a high density of single-nucleotide polymorphisms, allows complete sequencing of all individuals rather than relying on predetermined markers and provides direct sequence comparisons with mtDNA. We found the overall proportion of between-group variation ( $\Phi_{ST}$ ) to be 0.334 for the Y chromosome and 0.382 for mtDNA. Genetic differentiation between populations was similar for the Y chromosome and mtDNA at all geographic scales that we tested. Although patrilocality may be important at the local scale<sup>2,3</sup>, patterns of genetic structure on the continental and global scales are not shaped by the higher rate of migration among females than among males.

Differences in migration rates between males and females can be inferred from analyses of within- and between-group variation (e.g.,  $F_{ST}$ , a measure of genetic differentiation among populations) for the maternally inherited mtDNA and paternally inherited Y chromosome. Using this approach, one study pieced together data sets from published mtDNA, Y-chromosome and autosomal studies and found that the Y chromosome had larger between-group differences than did other portions of the genome<sup>1</sup>. But these data sets varied considerably with regard to sample sizes, populations represented and method used to assay genetic variation<sup>4</sup>, making any comparison of the extent of genetic differentiation between populations difficult. This problem is exacerbated for Y chromosomes because they lack sequence diversity<sup>5</sup>. To accommodate this difficulty, Y-chromosome researchers

have adopted a strategy<sup>6–9</sup> to estimate levels of between-group variation using single-nucleotide polymorphisms (SNPs) discovered in small panels of globally diverse males and then genotyped in much larger population samples (e.g., the global data set<sup>10</sup> analyzed in ref. 1). This nonrandom sampling of SNPs can result in an ascertainment bias that has an unknown effect on estimates of  $F_{ST}$  values. Several researchers have suggested that an extensive study of mtDNA and Y-chromosome diversity in the same samples, using the same method to assay variation, is necessary before one can make firm conclusions regarding the relative degree of mtDNA and Y-chromosome differentiation<sup>4,11</sup>.

Here, we directly compare 6.7 kb of Y-chromosome sequence and 770 bp of the gene mitochondrial cytochrome c oxidase 3 (*MTCO3*) in a sample of 389 individuals representing ten globally distributed human populations. We assayed Y-chromosome variation using DNA sequences that encompass recently inserted *Alu* retrotransposons

**Table 1** Characteristics of individual *Alu* elements sequenced on the Y chromosome

<i>Alu</i> name	<i>Alu</i> family	Gene region (intron number)	Length, including flanking DNA (bp)	Segregating sites
16e4	Y $\alpha$ 5	SMCY (14)	422	6
486	Y $\alpha$ 5	DFFRY (31)	290	4
DBYc2	Yc2	DBY (2)	340	2
DBYd1	Yd2	DBY (2)	588	2
AMELY	Y	AMELY (2)	734	6
DFFRY30	Y	DFFRY (11)	663	5
DFFRY40	Y	DFFRY (13)	546	6
DFFRY50	Y	DFFRY (17)	413	1
SMCY2	Y	SMCY (5)	406	1
UTY62	Y	UTY (10)	525	3
UTY87	Y	UTY (19)	440	1
DFFRY04	Y	DFFRY (1)	520	2
UTY83	Y	UTY (19)	796	4
Total			6,683	43

<sup>1</sup>Division of Biotechnology, <sup>2</sup>Department of Ecology and Evolutionary Biology and <sup>3</sup>Department of Anthropology, University of Arizona, Tucson, Arizona 85721, USA. Correspondence should be addressed to M.F.H. (mfh@u.arizona.edu).

Published online 19 September 2004; doi:10.1038/ng1428

**Table 2 AMOVA results for the Y chromosome and mtDNA at global and continental scales**

Geographic region	Variation within groups (%)	Variation between groups (%)
Global (10 populations)		
Y chromosome	66.5	33.4
mtDNA	62.0	38.2
Africa (4 populations)		
Y chromosome	79.2	20.8
mtDNA	68.0	32.0
Non-Africa (6 populations)		
Y chromosome	76.2	23.8
mtDNA	73.0	27.0
Asia (2 populations)		
Y chromosome	97.4	2.6
mtDNA	97.7	2.3
Oceania (2 populations)		
Y chromosome	79.5	20.5
mtDNA	60.9	39.1
Europe (2 populations)		
Y chromosome	101.1	-1.1
mtDNA	100.5	-0.5

(e.g., the Y family of *Alu* elements and its subfamilies<sup>12</sup>; **Table 1**), because these elements may have a higher mutation rate than other noncoding DNA as a result of their high density of CpG dinucleotides<sup>13</sup>. Focusing on these regions, we uncovered a much higher density of SNPs than reported in previous surveys. SNP density was 3.15 times higher in our data set than in a study of a pseudogene and other noncoding regions on the Y chromosome<sup>5</sup> (comparing data from 73 individuals that overlap between the two studies). Given the value of Watterson's  $\theta$  (an estimator of the quantity  $2N_e\mu$ , where  $N_e$  is the effective population size and  $\mu$  is the mutation rate) in our present *Alu* data set, we estimated the probability of observing a SNP density equal to or lower than that observed in the non-*Alu* data set to be  $P = 0.007$  using coalescent simulations. Thus, there is a significant increase in DNA sequence variation in regions encompassing *Alu* elements over the background level of noncoding diversity on the Y chromosome. Using this rich source of polymorphism, we can directly compare sequence variation between the Y chromosome and mtDNA in a manner that is free of ascertainment bias for both loci. We observed 43 Y-chromosome SNPs and 68 mtDNA SNPs (**Supplementary Tables 1 and 2** online).

We found no evidence that the Y chromosome has a higher level of differentiation between populations than does mtDNA. Using an analysis of molecular variance (AMOVA), we calculated the overall value of  $\Phi_{ST}$  (which approximates the quantity  $1/(1 + N_e m)$  assuming an equilibrium island model of population structure<sup>14</sup>, where  $m$  is the migration rate between populations) to be 0.334 for the Y chromosome and 0.382 for mtDNA. We examined distinct geographic regions individually and observed the same pattern of slightly higher  $\Phi_{ST}$

values for mtDNA than for the Y chromosome (**Table 2**). The between-groups component of genetic variation was higher for mtDNA than for the Y chromosome in every region except Asia, where the Y-chromosome  $\Phi_{ST}$  value slightly exceeded that of mtDNA (by 0.002). An AMOVA that incorporated a hierarchical grouping of populations within continents yielded similar results (**Table 3**). Values of  $\Phi_{SC}$  (between-group, within-continent variation) and  $\Phi_{CT}$  (between-continent variation) were similar for mtDNA and the Y chromosome, though slightly higher for mtDNA.

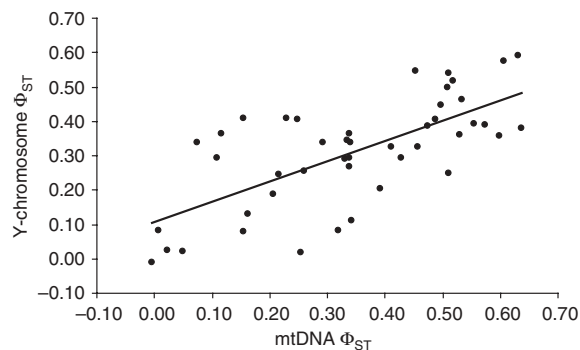
In addition to the overall similarity of between-group components of variation for the Y chromosome and mtDNA, there was a strong and statistically significant correlation of  $\Phi_{ST}$  values between pairs of populations for these loci (**Fig. 1**; Mantel correlation coefficient = 0.688;  $P < 0.001$ ). This result has several important implications. First, it indicates that putative gene flow among populations is relatively symmetrical for females and males. The fact that  $\Phi_{ST}$  values for these loci covary so strongly suggests that population-specific processes, such as variation in rates of migration for females versus males, do not influence our divergence data. Second, it indicates that there is no obvious trend towards different rates of divergence for mtDNA versus the Y chromosome. Although the nonindependence of the data points in **Figure 1** precludes conventional statistical analyses, we noted that the slope of the regression line suggests a faster increase in divergence between populations for mtDNA than for the Y chromosome, contrary to the pattern that would be expected if females had a higher rate of migration. Finally, the strong correlation between  $\Phi_{ST}$  values for these two compartments of the genome indicates that demographic, rather than locus-specific, evolutionary forces are the primary determinants of genetic distance between the populations we surveyed. Positive directional selection, for instance, operating in a subset of populations on either the Y chromosome or mtDNA would tend to uncouple  $\Phi_{ST}$  values between loci. We see no evidence for such a process in our data.

Our survey of the Y chromosome and mtDNA found markedly different between-group components of variation than have been reported in previous global studies. Some Y-chromosome studies rely on predetermined SNPs and find that between-population components of genetic variation are much higher than we estimated<sup>1,11</sup>. Our data suggest that ascertainment biases associated with the use of particular SNPs may result in overestimates of genetic distance between populations. In contrast, previous studies of mtDNA often focus on hypervariable portions of the control region (where high mutation rates may cause a downward bias in estimates of between-group variation<sup>15</sup>) rather than coding DNA<sup>11</sup>. We obtained much higher estimates of between-group variation when we compared mtDNA coding sequences with hypervariable regions in the same panel of individuals (**Supplementary Note** online).

Our interpretation of the between-group components of genetic variation for the Y chromosome and mtDNA in terms of rates of migration relies on the assumption that the effective population sizes of the sexes (and thus of the Y chromosome and mtDNA) are equal. Among human populations, forces that skew the breeding sex ratio

**Table 3 AMOVA grouping populations by continent of origin**

Locus	$\Phi_{ST}$ (within-groups %)	$\Phi_{SC}$ (between-groups, within-continent %)	$\Phi_{CT}$ (between-continent %)
Y chromosome	0.357 (64.3)	0.209 (17.0)	0.187 (18.7)
mtDNA	0.401 (59.9)	0.261 (21.1)	0.189 (18.9)



**Figure 1** Pairwise mtDNA and Y-chromosome  $\Phi_{ST}$  values. A Mantel test indicates that there is a significant correlation between pairwise mtDNA and Y-chromosome  $\Phi_{ST}$  values (Mantel correlation coefficient = 0.688;  $P < 0.001$ ). Although nonindependence of data points precludes statistical analysis of the regression line, mtDNA  $\Phi_{ST}$  values seem to increase more quickly between populations than do those for the Y chromosome (regression equation,  $\Phi_{ST(Y)} = 0.589 \times \Phi_{ST(mtDNA)} + 0.107$ ).

probably do so by increasing the number of females relative to males (e.g., owing to the widespread practice of polygyny and rarity of polyandry among cultures<sup>16,17</sup>, a higher variance in male lifetime reproductive success<sup>18</sup> or higher rates of male mortality<sup>19</sup>); the magnitude of this skew is not known. If the effective size of the human female population is indeed somewhat larger than that of males, then our observation of roughly equal between-group components of variation for the Y chromosome and mtDNA implies a lower rate of migration for females than for males among the widely spaced populations we surveyed.

We did not detect the signature of a higher migration rate among populations for females than for males in our global survey, but this does not contradict the evidence for patrilocality effects at local scales. For instance, in a comparison of genetic variation in Northern Thailand, patrilocality villages were characterized by lower levels of variation for the Y chromosome than for mtDNA and higher Y-chromosome genetic distances between villages, whereas the opposite was true among matrilocality groups. Similar patterns were observed among patrilocality Bedouin tribes from the Sinai Peninsula<sup>3</sup>. One of the outstanding questions raised by studies such as these is the extent to which local cultural practices influence genetic patterns at the regional and global scale<sup>4</sup>. At present, there are too few studies that specifically examine these issues of scale with respect to Y-chromosome versus mtDNA differentiation to draw firm conclusions. But our results, taken together with several regional-scale studies that did not detect a genetic signal of increased migration among females versus males<sup>20–22</sup>, suggest that broader-scale genetic patterns may not always reflect the sum of local cultural processes. This may be because other demographic events (e.g., long-distance migrations) become proportionately more important at larger geographic scales, or because behavioral customs of individual populations do not have the temporal or geographic stability necessary to influence global patterning. Although we are unable to distinguish among these hypotheses, our results suggest that the role of female migration is no more important than that of male migration at the continental and global scales.

## METHODS

**Population samples.** We examined mtDNA and Y-chromosome variation in the same panel of individuals from ten globally distributed populations, as follows (the number of individuals sampled is indicated in parentheses): Africa: Bakola

from Cameroon (25), Dogon from Mali (37), Bantu speakers from South Africa (47), Khoisan from Namibia and South Africa (25); Europe: Dutch (47) and Italians (47); Asia: Mongolian Khalks (46) and Sri Lankans (43); Oceania: highland Papua New Guineans (24) and Baining from New Britain (48). All samples were obtained with informed consent using protocols approved by the Human Subjects Protection Program at the University of Arizona.

**DNA sequencing.** Our study focused on a 770-bp region in the gene *MTCO3* of the mtDNA. We chose *MTCO3* rather than the hypervariable portions of mtDNA to mitigate to the greatest extent possible the degree to which homoplasy would downwardly bias our estimates of population differentiation<sup>23</sup>. Our survey spanned 13 separate regions from the nonrecombining portion of the Y chromosome. We chose these regions on the basis of three criteria: (i) they fall within introns of single-copy genes<sup>24</sup>; (ii) they contain at least one element from the Y family (including subfamilies) of *Alu* retrotransposons<sup>12</sup>; and (iii) *Alu* insertions were fixed in our sample. We determined the family affiliation of *Alu* elements using RepeatMasker. To ensure priming specificity, we located amplification primers in unique regions flanking *Alu* elements. We then directly sequenced both flanking and *Alu* DNA. Sequences of amplification and sequencing primers, as well as reaction conditions, are available on request.

**Data analysis.** We apportioned diversity within and between populations using an AMOVA, implemented in the program Arlequin v. 2.000 (ref. 25). The resulting  $\Phi_{ST}$  values are especially sensitive to differences in mutation rate<sup>23</sup>. To minimize biases associated with a higher mutation rate for mtDNA, we calculated genetic distances using a Tamura-Nei distance with high among-site rate heterogeneity ( $\gamma = 0.22$ ; data not shown). This measure accommodates homoplasy that may differentially occur in the mtDNA data set. For the Y chromosome, we calculated genetic distance using a Jukes-Cantor model of nucleotide substitution. All results for both loci are insensitive to choice of substitution model. There was evidence for a single recurrent mutation in our Y-chromosome SNP data. Parsimony analysis of the 40 haplotypes observed in our sample of 389 chromosomes resulted in a single tree of 44 steps with a consistency index of 0.977 (Supplementary Fig. 1 online). An A→C transversion occurs twice on the tree; however, because it occurs on separate branches, we were able to identify both mutational events. For the mtDNA *MTCO3* locus, we analyzed the entire data set ( $S = 68$ ), as well as a subset of the data including only synonymous sites ( $S = 49$ ). Analyses of synonymous sites, which are presumably under less selective constraint than coding sites, produced similar results (data not shown) to analyses of the entire data set. We implemented a Mantel test (100,000 permutations) comparing pairwise distances between populations for mtDNA and the Y chromosome in Arlequin.

**URLs.** RepeatMasker is available at <http://www.repeatmasker.org/>. Arlequin v. 2.000 is available at <http://lgb.unige.ch/arlequin/>.

**GenBank accession numbers.** mtDNA and Y-chromosomal sequences from the 389 individuals in our study, AY714986–AY720431.

*Note: Supplementary information is available on the Nature Genetics website.*

## ACKNOWLEDGMENTS

We thank D. Garrigan and A. Indap for assistance with data analysis; D. Meltzer and S. Zegura for comments; and P. de Knijff, G. Destro-Bisol, J. Friedlaender, T. Jenkins, H. Soodyall and B. Strassmann for samples. This manuscript was made possible by a grant from the National Institute of General Medical Sciences (to M.H.). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

## COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 19 May; accepted 17 August 2004

Published online at <http://www.nature.com/naturegenetics/>

- Seielstad, M.T., Minch, E. & Cavalli-Sforza, L.L. Genetic evidence for a higher female migration rate in humans. *Nat. Genet.* **20**, 278–280 (1998).
- Oota, H. *et al.* Human mtDNA and Y-chromosome variation is correlated with matrilocality versus patrilocality residence. *Nat. Genet.* **29**, 20–21 (2001).

3. Salem, A.H., Badr, F.M., Gaballah, M.F. & Pääbo, S. The genetics of traditional living: Y-chromosomal and mitochondrial lineages in the Sinai Peninsula. *Am. J. Hum. Genet.* **59**, 741–743 (1996).
4. Stoneking, M. Women on the move. *Nat. Genet.* **20**, 219–220 (1998).
5. Hammer, M.F. *et al.* Human population structures and its effects on sampling Y chromosome variation. *Genetics* **164**, 1495–1509 (2003).
6. Hammer, M.F. *et al.* The geographic distribution of human Y chromosome variation. *Genetics* **145**, 787–805 (1997).
7. Hammer, M.F. *et al.* Hierarchical patterns of global human Y-chromosome diversity. *Mol. Biol. Evol.* **18**, 1189–1203 (2001).
8. Romualdi, C. *et al.* Patterns of human diversity, within and among continents, inferred from biallelic DNA polymorphisms. *Genome Res.* **12**, 602–612 (2002).
9. Kayser, M. *et al.* Reduced Y-chromosome but not mitochondrial DNA, diversity in human populations from west new Guinea. *Am. J. Hum. Genet.* **72**, 281–302 (2003).
10. Underhill, P.A. *et al.* Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res.* **7**, 996–1005 (1997).
11. Jorde, L.B. *et al.* The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. *Am. J. Hum. Genet.* **66**, 979–988 (2000).
12. Batzer, M.A. & Deininger, P.L. Alu repeats and human genomic diversity. *Nat. Rev. Genet.* **3**, 370–379 (2002).
13. Labuda, D. & Striker, G. Sequence conservation in Alu evolution. *Nucleic Acids Res.* **17**, 2477–2491 (1989).
14. Wright, S. Evolution in Mendelian populations. *Genetics* **16**, 97–159 (1931).
15. Jin, L. & Chakraborty, R. Population structure, stepwise mutations, heterozygote deficiency and their implications in DNA forensics. *Heredity* **74**, 274–285 (1995).
16. Low, B.S. Measures of polygyny in humans. *Curr. Anthropol.* **29**, 189–194 (1988).
17. Murdock, G.P. *Atlas of World Cultures* (University of Pittsburgh Press, Pittsburgh, 1981).
18. Chagnon, N.A. Is reproductive success equal in egalitarian societies? in *Evolutionary Biology and Human Social Behavior: An Anthropological Perspective* (eds. Chagnon, N.A. & Irons, W.) 374–401 (Duxbury Press, North Scituate, Massachusetts, 1979).
19. Alexander, R.D. *et al.* Sexual dimorphisms and breeding systems in pinnipeds, ungulates, primates, and humans. in *Evolutionary Biology and Human Social Behavior: An Anthropological Perspective* (eds. Chagnon, N.A. & Irons, W.) 402–435 (Duxbury Press, North Scituate, Massachusetts, 1979).
20. Fuselli, S. *et al.* Mitochondrial DNA diversity in South America and the genetic history of Andean highlanders. *Mol. Biol. Evol.* **20**, 1682–1691 (2003).
21. Mesa, N.R. *et al.* Autosomal, mtDNA, and Y-chromosome diversity in Amerinds: pre- and post-Columbian patterns of gene flow in South America. *Am. J. Hum. Genet.* **67**, 1277–1286 (2000).
22. Al-Zahery, N. *et al.* Y-chromosome and mtDNA polymorphisms in Iraq, a crossroad of the early human dispersal and of post-Neolithic migrations. *Mol. Phylogenet. Evol.* **28**, 458–472 (2003).
23. Urbanek, M., Goldman, D. & Long, J.C. The apportionment of dinucleotide repeat diversity in Native Americans and Europeans: a new approach to measuring gene identity reveals asymmetric patterns of divergence. *Mol. Biol. Evol.* **13**, 943–953 (1996).
24. Skaletsky, H. *et al.* The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423**, 825–837 (2003).
25. 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).