Variation of Female and Male Lineages in Sub-Saharan Populations: the Importance of Sociocultural Factors

Giovanni Destro-Bisol,* Francesco Donati,* Valentina Coia,* Ilaria Boschi,† Fabio Verginelli,‡ Alessandra Caglià,† Sergio Tofanelli,§ Gabriella Spedini,* and Cristian Capelli†

*Department of Animal and Human Biology, University “La Sapienza,” Rome, Italy; †Institute of Legal Medicine, Catholic University, Rome, Italy; ‡Department of Oncology and Neurosciences and Center for Research and Training on Cancer in Sub-Saharan Africa, University G.d’Annunzio, Chieti, Italy; §Department of Ethology, Ecology and Evolution, University of Pisa, Italy

In this paper, we present a study of genetic variation in sub-Saharan Africa, which is based on published and unpublished data on fast-evolving (hypervariable region 1 of mitochondrial DNA and six microsatellites of Y chromosome) and slow-evolving (haplogroup frequencies) polymorphisms of mtDNA and Y chromosome. Our study reveals a striking difference in the genetic structure of food-producer (Bantu and Sudanic speakers) and hunter-gatherer populations (Pygmies, !Kung, and Hadza). In fact, the ratio of mtDNA to Y-chromosome N is substantially higher in food producers than in hunter-gatherers as determined by fast-evolving polymorphisms (1.76 versus 0.11). This finding indicates that the two population groups differ substantially in female and male migration rate and/or effective size. The difference also persists when linguistically homogeneous populations are used and outlier populations are eliminated (1.78 vs 0.19) or when the jackknife procedure is applied to a paired population data set (1.32 to 7.84 versus 0.14 to 0.66). The higher ratio of mtDNA to Y-chromosome N in food producers than in hunter-gatherers is further confirmed by the use of slow-evolving polymorphisms (1.59 to 7.91 versus 0.12 to 0.35). To explain these results, we propose a model that integrates demographic and genetic aspects and incorporates ethnographic knowledge. In such a model, the asymmetric gene flow, polygamy, and patrilocality play an important role in differentiating the genetic structure of sub-Saharan populations. The existence of an asymmetric gene flow is supported by the phylogeographic features of mtDNA and Y-chromosome haplogroups found in the two population groups. The role of polygyny and patrilocality is sustained by the evidence of a differential pressure of genetic drift and gene flow on maternal and paternal lineages of food producers and hunter-gatherers that is revealed through the analysis of mitochondrial and Y-chromosomal intrapopulational variation.

Introduction

The polymorphisms of mitochondrial DNA (mtDNA) and Y chromosome have opened up new perspectives in the study of genetic variation. This finding is made possible through three key features shared by the two genetic systems. First, they allow us to study maternal and paternal lineages separately because of their unilinear transmission. Second, the substantial lack of recombination allows mutations to accumulate along lines of descent. This feature makes mtDNA and Y chromosome much more powerful tools for phylogeographical analyses than autosomes. Third, polymorphisms can be used with comparable modes and rates of evolution. The fast evolving sites of the hypervariable region 1 (HVR-1) of mtDNA and the Y-chromosomal microsatellites on one side and the slow-evolving mtDNA and Y-chromosome sites that define haplogroups on the other side, indeed have comparable evolutionary rates. This condition makes minimizing the confounding effect of a differential mutational dynamic between the two genetic systems possible.

The study by Seielstad, Minch, and Cavalli-Sforza (1998) may be regarded as the pioneer study on the relations between genetic variation of unilinearly transmitted polymorphisms and sociocultural factors. In this worldwide analysis of mtDNA and Y-chromosome polymorphisms, the authors observed that the interpopulation differentiation was much higher for the Y chromosome than for mtDNA. Applying an island migration model (Cavalli-Sforza and Bodmer 1971), they obtained a ratio between female and male No—which approximates to the product of the effective population size per migration rate—close to 8 (7.96). According to Seielstad, Minch, and Cavalli-Sforza (1998), the higher female to male migration rate associated with the widespread habit of patrilocality may account for most of this difference. The study paved the way for further investigations that substantially confirmed the importance of the social structure in shaping the genetic variation of human populations (e.g. Oota et al. 2001).

In a study recently published in Molecular Biology and Evolution, Hammer et al. (2001) analyzed 43 biallelic polymorphisms of the nonrecombining portion of the Y chromosome in a total of 50 human populations that included five populations from sub-Saharan Africa (Bagandans, East Bantus, Gambians, Khoisan, and Pygmies). These authors obtained a $Fst$ of 0.251 for sub-Saharan Africans, a value that is slightly smaller than that obtained for Asians (0.271). They also noted that this finding contrasts with previous estimates for mtDNA (Melton et al. 1997), in which the $Fst$ value for sub-Saharan Africans (0.339) largely exceeds those of Europeans (0.045) and Asians (0.009). Hammer et al. (2001) suggested that the relatively small interpopulation variation of Y chromosome among sub-Saharan could be caused by a male-biased gene flow during the expansion of populations speaking Bantu languages. This interpretation is supported by the widespread distribution of haplotype 15 in Bantu-speaking populations (Hg E3a according to the nomenclature suggested by the Y Chromosome Consortium.

Key words: sub-Saharan Africa, food producers, hunter-gatherers, mtDNA, Y chromosome, sociocultural factors. 

E-mail: destrobisol@uniroma1.it.

Mod. Biol. Evol. 21(9):1673–1682. 2004
doi:10.1093/molbev/msi186
Advance Access publication June 9, 2004

Molecular Biology and Evolution vol. 21 no. 9 © Society for Molecular Biology and Evolution 2004; all rights reserved.
and by nested cladistic analysis. Thus, according to Hammer et al. (2001), sub-Saharan Africa might represent another case in which the genetic structure of human populations has been shaped by the greater mobility of males (Mesa et al. 2000; Carvajal-Carmona et al. 2000; Oota et al. 2001), in contrast with what was observed at the worldwide level (Seielstad, Minch, and Cavalli-Sforza 1998; Stoneking 1998). Interestingly, the conclusion drawn by Hammer et al. (2001) also differs from what was observed by Seielstad, Minch, and Cavalli-Sforza (1998) in another group of African populations. These authors analyzed 14 populations from Eastern and Central Africa and observed that within-population variance for autosomal microsatellite was higher than variance for Y-chromosomal microsatellites. The data set includes populations from Sudan (Beja), Ethiopia (Konso, Tsamako, Ongota), Hamar, Dasenech, Surma, Nyangatom, Bench, and Majangir), and Mali (Dogon, Peuhl, Tuareg, and Songhai). According to Seielstad, Minch, and Cavalli-Sforza (1998), this evidence fits the expectation of a higher female than male migration rate and is consistent with the patrilocal habit of most sub-Saharan populations.

The discrepancy between the conclusions drawn by Seielstad, Minch, and Cavalli-Sforza (1998) and Hammer et al. (2001) suggests that sub-Saharan populations analyzed in these two studies differ for variation of maternal and paternal lineages and/or for sex-linked migration rates. A substantial difference between the data sets used by the two studies lies in the fact that Seielstad, Minch, and Cavalli-Sforza (1998) analyzed only populations that can be considered food producers (food-producer populations [FPPs]), that is, agriculturalists and pastoralists, whereas Hammer et al. (2001) also surveyed populations with a traditional economy based on hunting and gathering (hunter-gatherer populations [HGPS], Pygmies and !Kung). Although based primarily on the type of subsistence economy, the distinction between FPPs and HGPS is important from a genetic point of view. In fact, FPPs and HGPS differ in their social structure in three specific aspects that may have important effects on variation of male and female lineages. First, the level of polygyny is higher among FPPs than among HGPS (Cavalli-Sforza 1986a). Second, exceptions to patrilocality have been described for HGPS but not for FPPs (Bahuchet 1999; Biesele and Royal 1999). Third, sociocultural barriers strongly influence the way in which unions between individuals of the two groups take place (see below). Therefore, an investigation of the relations between sociocultural factors and genetic variation may be very useful to better understand the mechanisms driving the genetic structure of sub-Saharan populations and to define more precisely the role of culture in determining the diversity within and between groups.

Materials and Methods

The Database

A total of 40 populations from sub-Saharan Africa was used in this study (see tables S1–S3 in Supplementary Material online). Our database also comprises unpublished results from three populations from Cameroon (Bakaka, Bassa, and Fulbe; data available at http://www.scienzemfn.uniroma1.it/labantro/index.html). FPPs include a total of 32 populations, mainly speakers of languages of the Niger-Kordofanian phylum (Greenberg 1963), from central and southern Africa (see tables S1–S3 in Supplementary Material online, for more detailed information on population location and sample size). HGPS (eight populations) include pygmies from western and eastern Africa, Hadza and Hadzabe from Tanzania, and !Kung from southern Africa.

Laboratory and Data Analyses

The genetic systems considered include the hypervariable region-1 (HVR-1, from np 16024 to 16384) of mtDNA, six Y-chromosomal microsatellites (DYS19, 389I, 390, 391, 392, and 393), and haplogroups of the mtDNA and Y chromosome built by use of unique evolutionary events. Sequencing of the HVR-1 (between positions 16024 and 16384) and Y-chromosomal microsatellite typing for Bakaka, Bassa, and Fulbe were carried out as previously described (Caglià et al. 2003; Destro-Bisol et al. 2004). MtDNA sequences were assigned to the phylogenetic tree of Salas et al. (2002).

Parameters of within-population (haplotype diversity and mean number of pairwise comparisons), between-population, and among-population diversity (Fst and qst) were calculated by the Arlequin software (Schneider et al. 1997). The Nv parameter (which incorporates effective size, migration, and mutation) was calculated by application of the formula Nv = (1/Fst) − 1, according to the island model of migration for haploid systems (Cavalli-Sforza and Bodmer 1971). Because the contribution of mutation rate to the Nv parameter in our genetic systems may be considered negligible, the fluctuations of Nv values have been assumed to be the result of differences in migration rate and/or effective population size among populations (see also Wijisman [1984] and Seielstad, Minch, and Cavalli-Sforza [1998]). The Kimura two-parameter (Kimura 1980) and Rst (Slatkin 1995) methods were used to calculate the genetic distances for HVR-1 of mtDNA and Y-chromosome microsatellites, respectively. To visualize the genetic relationships among the groups examined, we analyzed the genetic distance matrices by the nonmetric multidimensional scaling method (MDS [Kruskal 1964]) in the Statistica version 5.0 software.

Results

From all the data available for the hypervariable region-1 of mtDNA (28 populations) and Y-chromosomal microsatellites (22 populations), we obtained a qst of 0.172 for mtDNA and of 0.100 for Y chromosome (table 1). When an island model of migration is applied (Cavalli-Sforza and Bodmer 1971), these values produce a ratio of mtDNA to Y-chromosome Nv of 0.54, which may be consequence of a migration rate and/or an effective population size that is nearly double for males. When the populations are divided into FPPs and HGPS, a striking difference emerges. In fact, among FPPs the No estimated
for mtDNA is 1.76 times that obtained for Y chromosome, whereas the ratio of mtDNA to Y-chromosome Nu is only 0.11 among HGPs. Before attempting any interpretation of these initial results, we tested their robustness by three further analyses.

First, we analyzed pairwise genetic distances to establish whether the divergent ratio of mtDNA to Y-chromosome Nu between HGPs and FPPs reflects trends shared by populations within each group or, alternatively, whether it is the result of any detectable confounding factor. We focused on two possible confounding factors: the heterogeneity among linguistically different populations and the presence of outlier populations. Previous studies of African populations have revealed important correlations between genetic and linguistic distances (Cavalli-Sforza, Menozzi, Piazza 1994; Poloni et al. 1997; Lane et al. 2002; Sanchez-Mazas 2001). Thus, the reliability of estimates obtained from large African data sets by use of linguistically homogeneous populations must be tested. Poloni et al. (1997) compared Fst values for Y-chromosome p49, f/Taq1 haplotypes and mtDNA low-resolution Restriction Fragment Length Polymorphisms (RFLPs) from populations that belonged to the Niger-Congo branch of the Niger-Kordofanian phylum (Greenberg 1963) and obtained slightly higher Fst for mtDNA than for Y chromosome. Because the Niger-Congo branch include populations heterogeneous both linguistically and historically, we preferred to use a more narrow linguistic group and selected the FPPs who speak languages of the Benue-Congo subfamily of the Niger-Congo branch (see tables S1–S3 in Supplementary Material online). These populations, also referred to as Bantu-speaking populations, are thought to be in genetic continuity with the farmers who migrated from southwestern Cameroon to all of Central and South Africa some 3,000 years ago (Vansina 1984). The MDS plot of genetic distances calculated for the HVR-1 of mtDNA shows that FPPs (22 populations) tend to cluster together, whereas the HGPs (six populations) are widely dispersed throughout the plot (fig. 1A). This difference between the two groups was already predicted by their substantially different Fst values (table 1). The Bantu-speaking populations are less widely dispersed than FPPs on the whole. In the MDS plot based on Rst genetic distances (Slatkin 1995) for Y chromosome, no appreciable difference exists between the dispersion of FPPs (17 populations) and HGPs (five populations) (fig. 1C), and Bantu-speaking FPPs again seem to be less heterogeneous than HGPs as a whole. Further important information is provided by the MDS plots concerning the Mbenzele pygmys and the Bakaka, who cluster separately from HGPs (in the mtDNA plot), and Bantu-speaking FPPs (in the Y-chromosome plot), respectively. The genetic distinctiveness of these two populations is confirmed by the inspection of the distribution of genetic distances. In fact, the Mbenzele are responsible for all four mtDNA genetic distance values that are greater than 0.5 (figure 1B and genetic distance matrix in table S4 in Supplementary Material online). Considering the Y-chromosomal genetic distances between FPPs, the only genetic distance greater than 0.4 and four out of the six values greater than 0.3 are the results of comparisons between the Bakaka and other populations (figure 1D and the genetic distance matrix in table S5 in Supplementary Material online). The Fst values for the HVR-1 of mtDNA and the six microsatellites of the Y chromosome recalculate for the Bantu-speaking FPPs and after the elimination of the Mbenzele and Bakaka from the data set are reported in table 1. The most important changes concern the mitochondrial Fst of FPPs, which decreases by 40% (from 0.063 to 0.038) in the subset of Bantu-speakers, and the Y-chromosomal Fst of FPPs, which decreases by 30% (from 0.106 to 0.074) in the subset of Bantu-speakers without the Bakaka. However, the newly estimated ratio of mtDNA to Y-chromosome Nv for Bantu-speaking populations with (2.50) and without (1.78) Bakaka remains greater than 1 and much larger than for HGPs after the elimination of the Mbenzele (0.19).

Second we calculated Fst values for only the populations analyzed for both mtDNA and Y chromosome (table 2). Apart from the greater reliability of estimates obtained from paired mtDNA and Y-chromosomal population data sets, this step has another less evident but equally important advantage. The island model assumes an equilibrium between migration and drift. This condition is probably not met by some of the populations included in our data set, because past demographic events have probably left an important signature in their present genetic variation (e.g., the neolithic and Bantu expansion in most FPPs and bottleneck events in HGPs [Excoffier and Schneider 1999; Salas et al. 2002]). However, departures from equilibrium conditions can reasonably be expected to affect estimates for mtDNA and Y chromosome in a comparable way if paired data sets are used. Consequently, substantial bias should occur in the resulting ratio of mtDNA to Y-chromosome Nv. The use of paired data sets reconfirmed the difference between the two population groups observed when the complete data set is used. In fact, among FPPs, the Nv estimated for mtDNA is 3.83 times that obtained for Y chromosome, whereas the ratio of mtDNA to Y-chromosome Nu is of only 0.10 among HGPs. To test the robustness of this

| Table 1 Estimates of Fst and Nv of Sub-Saharan Populations Obtained from the Hypervariable Region-1 of Mitochondrial DNA and Microsatellite Haplotypes of the Y Chromosome |
|-----------------|-----------------|-----------------|
|                | mtDNA           | Y Chromosome    | mtDNA/Y Chromosome |
| All populations| 0.172 (28)       | 0.100 (22)      | 0.54              |
| FPPs           | 0.063 (22)       | 0.106 (17)      | 1.76              |
| FPPs (only Bantu speakers) | 0.038 (9)       | 0.090 (9)       | 2.50              |
| FPPs (Bantu speakers, Bakaka excluded) | 0.043 (8)       | 0.074 (6)       | 1.78              |
| HGPs           | 0.428 (6)        | 0.074 (5)       | 0.11              |
| HGPs (Mbenzele excluded) | 0.302 (5)       | 0.076 (4)       | 0.19              |

Note.—The number of populations analyzed is in parentheses. The abbreviations are explained in the text. The Nv values were calculated by application of an island model of migration (Cavalli-Sforza and Bodmer 1971). FPPs = food-producer populations. HGPs = hunter-gatherer populations.
difference, we applied a jackknife procedure by excluding one population at a time from the mtDNA and Y-chromosome data sets. This method accounts for the variance caused by interpopulation differentiation, although we cannot assume this procedure has the same effect in FPPs and HGPs, because the former are probably more historically correlated than the latter as a result of their more recent expansion events. No overlap occurred between the \( N_m \) interval estimates obtained for the two groups when a jackknife procedure is used, even taking into consideration the values obtained for Bantu-speakers only and after the elimination of the Bakaka Bantus and Mbenzele pygmies (table 2). This finding indicates that the discrepancy between the two population groups is not caused by an unbalanced composition of the database, the effect of single populations, or the small size of some population samples (as in the case of the eastern pygmies with only five individuals examined for Y-chromosomal microsatellites).

Third, we reestimated \( N_v \) from frequencies of haplogroups. Haplogroups of both mtDNA and Y-chromosome are mostly the result of unique evolutionary events created by mutations that occur at a relatively low rate (on the order of \( 10^{-9} \) for Y-chromosome single nucleotide polymorphisms [SNPs] [Thomson et al. 2000]). This rate contrasts with those of the HVR-1 and Y-chromosome microsatellites whose mutation rates are much higher (on the order of \( 10^{-6} \) per site per generation for the HVR-1 of mtDNA and \( 10^{-3} \) for Y-chromosome microsatellites [Vigilant et al. 1989; Kayser et al. 2002]) and part of their variation may be caused by recurrent mutation. Therefore, haplogroups may provide information different from HVR-1 and Y-chromosome microsatellites. The values obtained by use of haplogroup frequencies (table 3) confirm and even enlarge the previously observed differences between FPPs and HGPs. In fact, the ratio of mtDNA to Y-chromosome \( N_v \) largely increases in both FPPs and HGPs, but the value obtained for the former group (8.57) largely exceeds that for the latter (0.26). Also, in this case, the interval estimates for FPPs and HGPs obtained by use of the jackknife procedure do not overlap (see table 3).

**Table 2**

<p>| ( N_v ) Obtained from the Subset of Populations Analyzed for both the Hypervariable Region-1 of Mitochondrial DNA and the Microsatellite Haplotypes of the Y Chromosome |
|---|---|---|</p>
<table>
<thead>
<tr>
<th>( N_v )</th>
<th>mtDNA</th>
<th>Y Chromosome</th>
</tr>
</thead>
<tbody>
<tr>
<td>All populations (11)</td>
<td>0.211</td>
<td>0.106</td>
</tr>
<tr>
<td>FPPs (7)</td>
<td>0.038</td>
<td>0.130</td>
</tr>
<tr>
<td>FPPs (only Bantu speakers) (5)</td>
<td>0.036</td>
<td>0.114</td>
</tr>
<tr>
<td>FPPs (only Bantu speakers, Bakaka excluded) (4)</td>
<td>0.040</td>
<td>0.083</td>
</tr>
<tr>
<td>HGPs (4)</td>
<td>0.431</td>
<td>0.072</td>
</tr>
<tr>
<td>HGPs (Mbenzele excluded) (3)</td>
<td>0.278</td>
<td>0.071</td>
</tr>
</tbody>
</table>

*Note.—The number of populations analyzed is in parentheses. The \( N_v \) values were calculated by application of an island model of migration (Cavalli-Sforza and Bodmer 1971). Interval estimates of \( N_v \) parameter (in brackets) were obtained by use of the jackknife procedure. FPPs = food-producer populations. HGPs = hunter-gatherer populations.*

**Fig. 1.**—Multidimensional scaling plots and distributions of genetic distances among sub-Saharan populations for mitochondrial DNA hypervariable region-1 (A and B) and Y-chromosome microsatellite haplotypes built from loci DYS19, 389I, 390, 391, 392, and 393 (C and D). Genetic distances between FPPs are indicated by black histograms, and those between HGPs are indicated by gray histograms. HGPs are indicated by arrows. Bantu-speaking populations are in italic. Population abbreviations are reported in Table S1 in Supplementary Material online. The stress index value is of 0.114 and 0.130 for the mtDNA and the Y-chromosome plots, respectively.

**Discussion**

The results obtained in the course of this study points out to the existence of a clearcut difference between the genetic structure of FPPs and HGPs. In fact, the ratios of
mtDNA to Y-chromosome Nv of FPPs and HGPs constantly remain above and below 1, respectively, when the complete data set for fast-evolving polymorphisms is used (table 1). Furthermore, no overlap between the two groups exists, even when the selected data sets (obtained by using linguistically homogeneous populations and eliminating outlier populations) and the interval estimates produced by the jackknife procedure are considered (tables 1 and 2). A further confirmation of the difference in Nv between FPPs and HGPs is provided by the use of slow-evolving polymorphisms (table 3). At the same time, a certain difference between the estimates obtained by use of fast-evolving (HVR-1 and microsatellites) and slow-evolving (mtDNA and Y-chromosome haplogroups) polymorphisms can be noted. In fact, the ratio of mtDNA to Y-chromosome Nv for FPPs ranges from 1.76 to 4.85 for fast-evolving polymorphisms and from 6.24 to 14.79 for slow-evolving polymorphisms. Also, in the case of HGPs, the estimates based on slow-evolving polymorphisms (0.12 to 0.35) are higher than those based on fast-evolving polymorphisms (0.09 to 0.19). Interestingly, the haplogroup-based estimate for Bantu-speaking FPPs (6.24) is not far from the estimate obtained by Seielstad et al. (1998) for world populations by means of SNP data. Because the estimates for the two kinds of polymorphisms were obtained by use of different populations, differential sampling could explain at least part of the discrepancy. This possibility is clearly indicated by the marked variation that can also be observed among the estimates obtained by use of either fast-evolving (see and compare the values reported in tables 1 and 2) or slow-evolving polymorphisms (see table 3). The comparison between the results of fast-evolving and slow-evolving polymorphisms is further complicated by the fact that mtDNA haplogroups were defined on the basis of a subset of variable nucleotides of HVR-1, whereas Y-chromosomal haplogroups were assigned on the basis of loci different from those used to build microsatellite haplotypes. Comparisons based on data from homogeneous data sets are required to shed more light on the important issue of the comparison of Fst estimates based on slow and fast-evolving polymorphisms.

Once the robustness of the initial results had been tested, our next logical step was to discuss the microevolutionary processes underlying the different patterns of mitochondrial and Y-chromosomal variation of FPPs and HGPs. To this end, we built a model (fig. 2) that integrates demographic and genetic aspects and incorporates ethnographic knowledge, especially knowledge of African pygmies.

The possibility has been suggested that HGPs could have maintained relatively large effective population sizes through high migration rates before the arrival of FPPs, namely Bantu-speaking farmers. In fact, until that time HGPs had no competitors and could occupy a wider territory (Cavalli-Sforza 1986b). The arrival of farmers, who penetrated the equatorial belt in the course of the expansion of Bantu-speakers peoples, around 2,000 to 3,000 years ago (Vansina 1980) caused an important change in the demography of HGPs. Because of their higher growth rate, the Bantu-speaking farmers progressively occupied most of the habitat previously available to HGPs and pushed the HGPs into less favorable areas. The fragmentation of HGPs inevitably reduced the gene flow between subpopulations and, consequently, their effective size decreased. This scenario of competition between preexisting HGPs and expanding Neolithic populations has been proposed by Excoffier and Schneider (1999) to explain the lack of signs of Pleistocene expansion in both African and non-African HGPs. In the case of African HGPs, the explanation receives support from the study by Weiss and Von Haeseler (1998, who detected traces of a recent decrease of population size in Biaka pygmies from the Central African Republic by use of mtDNA variation. However, with this model, we do not want to imply that the marginalization of HGPs by Bantu-speakers was the only cause underlying their peculiar genetic structure. In fact, the tendency of HGPs to conform to a stable demographic model must be taken into account. In this circumstance, genetic drift may be more effective than in agriculturalists, who tend to be more fertile than non-agriculturalists (Sellen and Mace 1997; von Haeseler, Sajantila, and Paabo 1996).

Another important consequence of the meeting between HGPs and FPPs is their reciprocal gene flow. No substantial taboos or social barriers existed between HGPs and FPPs during the early stages of contact. The two groups to probably exchanged genes symmetrically, an assumption supported by the similarity of anthropometric characters between the Twa Konda (pygmies) and Oto Konda (farmers) of Central Africa (Cavalli-Sforza 1986b). However, the present-day situation indicates that this initial condition changed considerably, and an asymmetric gene flow progressively developed between HGPs and FPPs because of the establishment of sociocultural inequalities (Cavalli-Sforza 1986b). In fact, pygmy women are accepted as wives by Bantu communities, first, because they are famed for their great fertility, and, secondly, because the future husband must pay a relatively low “bride price” to the wife’s family to gain the right to

### Table 3

| Estimate of **Fst** and Nv of Sub-Saharan Populations Obtained from Haplogroups of the mtDNA and Y Chromosome |
|--------------------------------------------------|-----------------|-------------|-----------------|
| **mtDNA** | **Y Chromosome** | **mtDNA/Y Chromosome** |
| All populations (11) | 0.173 | 0.277 | 1.81 |
| FPPs (7) | 0.070 | 0.392 | 5.57 [6.91–14.79] |
| FPPs (only Bantu speakers) (4) | 0.021 | 0.118 | 6.24 [1.59–7.51] |
| HGPs (4) | 0.373 | 0.134 | 0.26 [0.12–0.35] |

**Note:** The number of populations analyzed is in parentheses. The Nv values were calculated by application of an island model of migration (Cavalli-Sforza and Bodmer 1971). Interval estimates of Nv parameter (in brackets) were obtained by use of the jackknife procedure. For mtDNA we considered frequencies of haplogroups L1a, L1b, L1c, L1d, L1e, L1f, L2a, L2b, L2c, L2d, L3a, L3b, L3d, L3e, L3f, and L3g (see table S1 in Supplementary Material online for further details). For Y chromosome we considered frequencies of haplogroups A, B2a, B2b, B2c, B3, B4, E3a, E3d, and C/F (YCC 2002; Knight et al. 2003). FPPs = food-producer populations. HGPs = hunter-gatherer populations.
marriage. These conditions make a pygmy-to-Bantu flow of maternal lineages possible. A Bantu-to-pygmy flow of paternal lineages is also expected through three mechanisms: first, extramarital unions between pygmy females and Bantu-speaking males, second, adoption of orphans born from mixed unions, and, third, a return to the HGP of Pygmy women and their children after the divorce from Bantu-speaking FPP males. On the other hand, both the pygmy-to-Bantu flow of paternal lineages and the Bantu-to-pygmy flow of maternal lineages are inhibited by sociocultural taboos against unions between Bantu-speaking females and Pygmy males (Cavalli-Sforza 1986b). The resulting asymmetric gene flow has left a signature in both pygmies and Bantu speakers, but it must have had a deeper impact on the genetic structure of the latter because of their smaller population size. This asymmetric gene flow between FPPs and HGPs probably affected the gene flow between HGP subpopulations as well. In fact, the females of pygmy groups in which mixed unions with Bantu speakers occur more frequently are less available for marriages with males from other pygmy subgroups (Biasutti 1967). This circumstance is another factor that could have contributed to the high level of genetic differentiation observed for mtDNA among HGPs.

Different levels of polygyny and patrilocality in FPPs and HGPs are other factors probably involved in the differences observed between the two groups. By decreasing the male effective size, polygyny decreases diversity of paternal lineages within populations and increases that among populations. Polygyny is known to be substantially higher among FPPs than among HGPs, both in terms of proportion of polygamists and in terms of average number of wives per polygamist (Cavalli-Sforza 1986a; Biesele and Royal 1999). Furthermore, whereas all FPPs practice a rigid patrilocality, some HGPs do not exclusively follow such social behavior. Among the Aka pygmies, a group that includes the Biaka and Mbenzele pygmies analyzed here, the young couple generally settles in the husband’s camp after the birth of the first child. However, the husband may remain in the wife’s community, where he may be joined by one of his brothers or sisters (Bahuchet 1999). Among some !Kung groups of Botswana, males often join the wife’s family (Biesele and Royal 1999). Interestingly, a high level of polygyny and extreme patrilocality have been proposed as probable causes of the low Y-chromosome and high mtDNA diversity observed in West New Guinea populations (Kayser et al. 2003).

Apart from being consistent with the results described above and providing an explanation for the discrepancy between the conclusions of Seielstad, Minch, and Cavalli-Sforza (1998) and Hammer et al. (2001), this model predicts the existence of phylogeographic traces of a sex-biased gene flow between HGPs and FPPs. The existence of a sex-biased gene flow is supported by the distribution of the most-common haplogroups in FPPs and HGPs. Haplogroup E3a is the modal Y-chromosome type in FPP neighbors of HGPs; frequencies range from 42% (Wairak from Tanzania [Luis et al. 2004]) to 96% (Bamileke from Cameroon [Cruciani et al. 2002]). E3a has been indicated as a signature of Bantu expansion (Underhill et al. 2001). It is present in all HGPs, where it ranges from 5% among the Ju/'hoansi !Kung (recalculated from Underhill et al. [2001] by Knight et al. [2003]), who are known to have intermarried with Bantu-speakers to a low degree, to 65% among the Biaka pygmies (Cruciani et al. 2002). The haplogroup E2b1x (E2b1a) is also probably associated with the Bantu expansion (Cruciani et al. 2002). It reaches

![FIG. 2.—Model describing the role of asymmetric gene flow, polygyny, and patrilocality in determining the different pattern of mtDNA and Y-chromosome variation between FPPs and HGPs.](image-url)
a frequency of 15% among southern African Bantu and 6% in the !Kung (Underhill et al. 2001; Cruciani et al. 2002). Furthermore, most of the other Y-chromosome types observed among HGPs are absent among FPPs ([A3b1, A2, B2a*, B2b2, B2b4b, B2b* (B2b3*), B2b3a] [Cruciani et al. 2002]), with two exceptions. The first exception is represented by haplogroups B2a1 and B2b3x(B2b3a) (both occur at a frequency of 5% among Biaka pygmies [Cruciani et al. 2002]), which, together with the other haplogroups that belong to the group II of Underhill et al. (2001), is thought to be the remnant of early diversification and dispersal processes within Africa (Underhill et al. 2001; Cruciani et al. 2002). The second exception is provided by the haplogroup E3b*x(E3b1, E3b2, and E3b3), which is found only among !Kung from southern Africa (with a frequency of 11%) and in geographically distant populations such as Ethiopian Jews and Mossi from Burkina Faso (Cruciani et al. 2002). Therefore, the instances of haplogroup sharing mentioned above seem to be the result of the maintenance of ancestral characteristics diluted elsewhere by more recent demographic events rather than reverse gene flow (from HGPs to FPPs). On the other hand, the analysis of mtDNA haplogroup distribution show quite a different pattern. Western pygmies and !Kung have two different mtDNA modal types, L1c1a1 (60% among the Mbenzele pygmies and 30% among the Biaka pygmies [Destro-Bisol et al. 2004]) and L1d (frequency of 51% to 96% among !Kung: recalculated by Destro-Bisol et al. [2004] from Chen et al. [2000]). This substantial heterogeneity between pygmies and !Kung illustrates their long reciprocal isolation. These two mtDNA types have probably originated among pygmies and !Kung (Destro-Bisol et al. 2004; Salas et al. 2002) and are also present in some FPP neighbors of HGPs. The L1c1a1 is found among Ewondo (frequency of 13% [Destro-Bisol et al. 2004]), whereas the L1d is present among Mozambicans (frequency of 5% [Salas et al. 2002]). Some signatures of reverse gene flow (from FPPs to HGPs for maternal lineages) are also detectable in the mtDNA haplogroups of probable Bantu origin present both in the HGPs and in their FPP neighbors (L1a2, L1a1a, L3d3, and L3e2b [Pereira et al. 2001; Salas et al. 2002]). This finding is indicated by the cumulative frequency of the haplogroups mentioned above, which ranges from 0% (Botswana !Kung [Vigilant et al. 1991]) to 19% (South African !Kung [Chen et al. 2000]), with intermediate values of 4% among the Mbenzele (Destro-Bisol et al. 2004) and 18% among the Biaka (Vigilant et al. 1991). This indication of introgression of FPP maternal lineages into HGPs is not in contrast with the asymmetric gene flow predicted by our model. In fact, the cumulative frequency of the Bantu mtDNA haplogroups is substantially lower than the frequency of the E3a haplogroups in the same HGPs (from 39% in the !Kung to 65% in the Biaka [Cruciani et al. 2002]).

The evidence that the differential gene flow of paternal lineages has left a stronger signature than the differential gene flow of maternal lineages merits some further considerations. This difference may be seen on two different levels. First, haplogroups bearing the M2 mutation have been observed among all the HGPs analyzed so far, whereas the L1c1a1 and L1d haplogroups have only been found in some FPPs. Second, at present, no clear signs of reverse gene flow exists for Y chromosome (from HGPs to FPPs), whereas such signs do exist for mtDNA (from FPPs to HGPs). This discrepancy can be explained by the substantial difference in size between FPPs and HGPs. In the case of paternal lineages, the signs of the FPP-to-HGP gene flow are more evident and persistent because of the smaller size of recipient populations. Furthermore, the smaller size of HGPs could have facilitated the retention of FPP maternal lineages acquired during the initial period of symmetric gene flow. On the other hand, the larger size of the FPPs has probably diluted the signs of the HGP-to-FPP gene flow of Y chromosomes that probably occurred in the initial phase of contact between the two population groups.

Another important implication of our model is in the differential pressure of microevolutionary forces on maternal and paternal lineages of HGPs and FPPs. In fact, because of the combined effect of asymmetric gene flow and different levels of polygyny and patri locality, the model predicts that genetic drift had been more effective on maternal than on paternal lineages of HGPs, whereas gene flow is expected to be the prevailing microevolutionary force on their paternal lineages. The opposite situation is expected for FPPs. Consequently, HGPs should show a higher intrapopulational diversity for paternal than for maternal lineages, whereas the opposite should be valid for FPPs. To test these expectations, we compared HVR-1 mtDNA and Y-chromosome microsatellite haplotype diversity in the same populations for both genetic systems. From this comparison, we found that HGPs show the lowest level of haplotype diversity for mtDNA but nearly the highest for the Y chromosome (fig. 3). Furthermore, the difference in the ratio of mtDNA HVR-1 to Y-chromosome microsatellite haplotype diversity between HGPs and FPPs is statistically significant by the Mann-Whitney U test (P = 0.011). Therefore, our results seem to reflect a substantial difference between FPPs and HGPs concerning the degree of patri locality and polygyny, which is so far suggested by only a few, nonsystematic anthropological studies (Cavalli-Sforza 1986a; Bahuchet 1999; Biesele and Royal 1999).

Conclusion

In this study, we have analyzed variation of maternally and paternally transmitted morphisms by use of loci with different evolutionary rates and a large data set of sub-Saharan populations. A final picture of genetic variation in this area requires ad hoc studies on the populations of the northern fringes of sub-Saharan Africa (e.g., Sudan, Niger, Chad, Somalia, and Ethiopia). Nonetheless, the results obtained so far allow us to draw four main conclusions. First, our study provides evidence for a marked heterogeneity between FPPs and HGPs in terms of distribution of unilinearly transmitted markers. This result demonstrates the danger of making excessive generalizations regarding the genetics of the populations living south of the Sahara desert (see also Tishkoff and Williams [2002]). Second, the results obtained in the
course of our study suggest that the inferences drawn by both Seielstad, Minch, and Cavalli-Sforza (1998) and Hammer et al. (2001) are both valid. In fact, the genetic structure of present-day HGPs seems to maintain a strong signature of the male-biased gene flow of Bantu origin proposed by Hammer et al. (2001). Furthermore, the higher female migration rate suggested by Seielstad, Minch, and Cavalli-Sforza (1998) for human populations on a global scale and for East and Central Africa is supported for FPPs. Third, our study introduces one important novelty relative to the two studies mentioned above. We claim that the values of Fst obtained for mitochondrial and Y-chromosomal polymorphisms reflect not only the mobility of males and females but also the pressure of genetic drift on maternal and paternal lineages. Fourth, our research reconfirms the importance of the multidisciplinary approach for studies on human genetic variation. In our model, the asymmetric gene flow, polygyny, and patriolocality, and, hence, the sociocultural factors underlying them, have an important role in determining and differentiating the genetic structure of sub-Saharan populations.

Supplementary Material

The following tables are available online: table S1 (mtDNA hypervariable region-1 data used in this study), table S2 (Y-chromosome microsatellite data used in this study), table S3 (Y-chromosome haplogroup data used in this study), table S4 (genetic distance matrix obtained using mtDNA hypervariable region-1 data), and table S5 (Genetic distance matrix obtained using Y-chromosomal microsatellite data).

Acknowledgments

This study develops the work initiated with our unforgettable friend and colleague Michele Belledi, to whom the paper is dedicated. We thank the Equipe du Dispensaire de Santé Intégrée, Mission Catholique du Belimboké (Central African Republic) and the blood donors, whose availability made this study possible. This research was supported by grants from the M.U.R.S.T. (Cofin Projects 2003054059 “Dna e Biodemografia: approccio integrato allo studio della mobilità umana,” and 2002063871 “La struttura genetica del cromosoma Y in Italia”) and the University of Rome “La Sapienza” (Ateneo project C26A034352 “Analisi della variabilità in geni sottoposti a selezione in popolazioni dell’Africa sub-Sahariana”). We thank Alec Knight and Elizabeth Wood for their useful comments on an earlier version of this paper.

Literature Cited


Cruciani, F., P. Santolamazza, P. Shen et al. (16 co-authors) 2002. An Asia to sub-Saharan Africa back migration is...


David Goldstein, Associate Editor

Accepted May 17, 2004