Sex-Specific Migration Patterns in Central Asian Populations, Revealed by Analysis of Y-Chromosome Short Tandem Repeats and mtDNA

Anna Pérez-Lezaun,[1] Francesc Calafell,[1] David Comas,[1] Eva Mateu,[1] Elena Bosch,[1] Rosa Martínez-Arias,[1] Jordi Clarimón,[1] Giovanni Fiori,[2] Donata Luiselli,[2] Fiorenzo Facchini,[2] Davide Pettener,[2] and Jaume Bertranpetit[1]

[1]Unitat de Biologia Evolutiva, Facultat de Ciències de la Salut i de la Vida, Universitat Pompeu Fabra, Barcelona; and [2]Dipartimento di Biologia evoluzionistica sperimentale, Unità di Antropologia, Università di Bologna, Bologna

Summary

Eight Y-linked short-tandem-repeat polymorphisms (DYS19, DYS388, DYS389I, DYS389II, DYS390, DYS391, DYS392, and DYS393) were analyzed in four populations of Central Asia, comprising two lowland samples—Uighurs and lowland Kirghiz—and two highland samples—namely, the Kazakhs (altitude 2,500 m above sea level) and highland Kirghiz (altitude 3,200 m above sea level). The results were compared with mtDNA sequence data on the same individuals, to study possible differences in male versus female genetic-variation patterns in these Central Asian populations. Analysis of molecular variance (AMOVA) showed a very high degree of genetic differentiation among the populations tested, in discordance with the results obtained with mtDNA sequences, which showed high homogeneity. Moreover, a dramatic reduction of the haplotype genetic diversity was observed in the villages at high altitude, especially in the highland Kirghiz, when compared with the villages at low altitude, which suggests a male founder effect in the settlement of high-altitude lands. Nonetheless, mtDNA genetic diversity in these highland populations is equivalent to that in the lowland populations. The present results suggest a very different migration pattern in males versus females, in an extended historical frame, with a higher migration rate for females.

Received December 7, 1998; accepted for publication April 23, 1999; electronically published May 24, 1999.

Address for correspondence and reprints: Dr. Jaume Bertranpetit, Unitat de Biologia Evolutiva, Facultat de Ciències de la Salut i de la Vida, Universitat Pompeu Fabra Doctor Aiguader 80, 08003 Barcelona, Catalonia, Spain. E-mail: jaume.bertranpetit@cexs.upf.es

0002-9297/99/6501-0027$02.00

Introduction

Central Asia is a vast extension of land often regarded as a borderland between East and West. It comprises the republics of Uzbekistan, Tajikistan, Turkmenistan, Kirghizstan, and part of Kazakhstan, along the regions of the Pamir, the Hindu Kush, and farther to the northeast.

Genetically, Central Asia is one of the least-studied major regions of the world. The scarce studies on the region, based on classical genetic markers (Cavalli-Sforza et al. 1994) and mtDNA (Comas et al. 1998), seem to indicate that Central Asia has a genetic composition intermediate between those of Asian and European populations, probably because of admixture of already differentiated Eastern and Western populations rather than because of an intermediate position in a general Eurasian cline.

The value of Y-chromosome polymorphisms for human evolutionary studies has largely been recognized. Recently, a large amount of population variation in the Y chromosome has been described (Underhill et al. 1997; Kayser et al. 1997), which seems to be highly relevant in molecular studies of human evolution (Pena et al. 1995; Hammer and Zegura 1996; Underhill et al. 1996; Hammer et al. 1997). Other studies (Cooper et al. 1996; Deka et al. 1996) have shown, in particular, the applicability of Y-chromosome short tandem repeat (STR) haplotypes to human population genetics. Because of the smaller effective population size of Y chromosomes compared with the autosomes, Y-chromosome polymorphisms are more affected by genetic-drift processes and thus could be very useful to point out genetic differences between closely related populations whose time of divergence has been relatively short.

In a previous study (Pérez-Lezaun et al. 1997), the degree of polymorphism in Y-chromosome STRs and their autosomal counterparts was analyzed and assessed to be equivalent, when corrections for the smaller effective population size of the Y chromosome with respect to the autosomes were applied. The fact that the majority of the Y chromosome does not recombine offers the possibility of tracing back male lineages in time. In the present study, eight Y-specific STRs were typed in four population samples of Central Asia—Uighurs, Kazakhs, highland Kirghiz, and lowland Kirghiz (fig. 1)—to investigate the male-mediated relations among these populations and their genetic structure. The results were compared with mtDNA sequence data on the same individuals (Comas et al. 1998), to study possible differences in male versus female migration patterns in Central Asian populations.

A global comparison of mtDNA and Y chromosome–polymorphism data (Seielstad et al. 1998) has shown that a higher female than male migration rate can explain the discrepancy in the patterns of genetic variation between the maternally and the paternally transmitted genome regions. Nonetheless, the sample sizes, populations represented, and methods used to assay genetic variation differed considerably between the two types of markers, which makes a strict comparison problematic (Stoneking 1998). In order to address the differential sex-related migration patterns, the different markers should be studied in the same populations—and even in the same individuals (Stoneking 1998)—and the analysis should be restricted to a well-defined region, since there may have been heterogeneity in migration patterns in different human societies (Hassan 1981; Kelly 1995).

Material and Methods

Blood samples were collected by G.F., D.L., F.F., and D.P. within the CAHAP (Central Asia High Altitude People) research program, in collaboration with the Laboratory of Anthropology of the Academy of Science of Kazakhstan.

The Uighurs were sampled in the village of Penjim in the Panfilov district, Taldy-Corgan region, in the easternmost section of Kazakhstan. This region is inhabited mostly by Uighurs, who emigrated from Sinkiang (Chinese Uighur autonomous region) during recent decades; both Tardy-Colgan and Sinkiang are lowlands. The Kazakh samples were collected in the village of Aktasty (Almaty region, Kazakhstan, altitude 2,100 m above sea level). Two samples from Kirghizstan were analyzed: one from the high-altitude village of Sary Tash, in the Pamir region (3,200 m above sea level), and one from the lowland village of Bakai Ata in the Talas Valley (900 m above sea level). Samples were collected, after appropriate informed consent had been obtained, directly from healthy male donors in the villages, and special care was taken to avoid related individuals.

Six tetranucleotide Y-linked polymorphisms—DYS19, DYS389I, DYS389II, DYS390, DYS391, and DYS393—and two trinucleotide Y-chromosome polymorphisms—namely, DYS388 and DYS392—were analyzed in all four populations. A total of 43–56 males were tested for each marker and population.

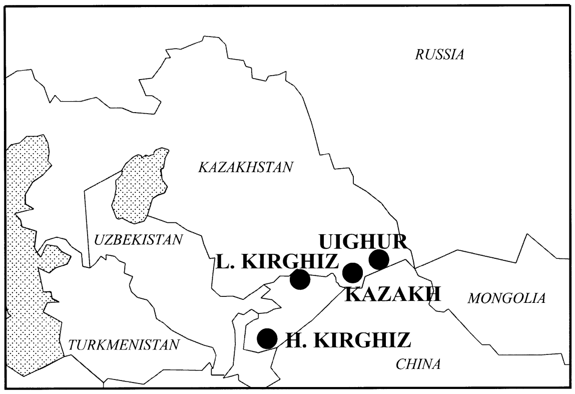

Figure 1 Map illustrating geographic location of Central Asian populations sampled. L. Kirghiz = Lowland Kirghiz (Talas Valley); H. Kirghiz = Highland Kirghiz (Sary-Tash).

Markers DYS19, DYS388, DYS389I, DYS389II, DYS391, and DYS393 were amplified by means of denaturing steps of 94°C for 20 s and elongation steps of 72°C for 1 min. Annealing temperatures were decreased from 63°C, by 0.5°C intervals, within each of the 14 initial cycles, followed by 20 cycles at 56°C; annealing time was 1 min.

DYS390 and DYS392 were amplified by means of denaturing steps of 94°C for 20 s and elongation steps of 72°C for 30 s. Annealing temperatures were decreased from 58°C, by 0.5°C intervals, within each of the 8 initial cycles, followed by 27 cycles at 54°C; annealing time was 30 s.

An initial denaturing step at 94°C for 1 min and a final extension step at 72°C for 5 min were included in both protocols. All PCR reactions were performed in a 10-μl final reaction volume, by means of a Perkin Elmer 9600 thermal cycler and under standard reagent conditions. Fluorescently labeled primers were used for all the amplifications.

PCR products were run in an ABI 377™ sequencer. Genescan 672™ and Genotyper 1.1™ software packages were used to analyze and to size allele bands. All the alleles (save for those of DYS388; see below) have been designated according to their composition, in number of repeats, as reported by Kayser et al. (1997); the correspondence between fragment size and repeat length was established through the use of sequenced allele ladders provided by P. de Knijff (Leiden). Alleles 12, 13, 14, 15, 16, and 17 for locus DYS19 correspond, respectively, to alleles Z and A, B, C, D, and E in the study by Santos et al. (1993). For marker DYS388, three individuals presenting different allele sizes were sequenced by means of

Table 1

Allele Frequencies of Eight Y-Chromosome STRs in Four Central Asian Populations

Allele	Uighurs	Kazakhs	Highland Kirghiz	Lowland Kirghiz
	Frequency			
DYS19:				
10	0	0	0	.023
11	0	0	0	0
12	0	0	0	0
13	.070	0	0	0
14	.186	.080	.070	.136
15	.209	.080	.140	.182
16	.465	.840	.790	.659
17	.070	0	0	0
	Summary Statistics for DYS19			
	n = 43; D = .712	n = 50; D = .287	n = 43; D = .360	n = 44; D = .526
	Frequency			
DYS388:				
10	.082	0	0	0
11	.041	.018	0	0
12	.551	.164	.958	.617
13	.041	.055	0	.170
14	.286	.745	.042	.064
15	0	0	0	0
16	0	0	0	0
17	0	0	0	.043
18	0	.018	0	.106
	Summary Statistics for DYS388			
	n = 49; D = .619	n = 55; D = .414	n = 48; D = .082	n = 47; D = .586
	Frequency			
DYS389I:				
9	.245	.018	.021	.106
10	.428	.839	.125	.213
11	.327	.143	.854	.681
	Summary Statistics for DYS389I			
	n = 49; D = .663	n = 56; D = .280	n = 48; D = .260	n = 47; D = .490
	Frequency			
DYS389II:				
24	.043	0	0	0
25	.065	.058	0	.128
26	.348	.731	.125	.106
27	.348	.115	.021	.128
28	.174	.096	.041	.170
29	.022	0	.792	.447
30	0	0	.021	.021
	Summary Statistics for DYS389II			
	n = 46; D = .737	n = 52; D = .440	n = 48; D = .362	n = 47; D = .743

(*continued*)

the DNA Sequencing Kit™ (Perkin-Elmer), to determine the number of repeats and the repeat structure.

Primers for the DYS389 locus amplify a partially duplicated region and generate two PCR products, DYS389I and DYS389II. It has been shown that the DYS389II fragment contains the length variation in DYS389I, as well as three additional stretches of tetranucleotide repeats (Pestoni et al. 1999; Rolf et al. 1998). Therefore, we have used only DYS389I in the joint analysis with the other six Y-linked STRs.

Gene diversity at each locus was estimated as $D = 1 - [(N/N - 1)(\Sigma f_i^2)]$, where N is sample size and f_i is

Table 1 (continued)

Allele	Uighurs	Kazakhs	Highland Kirghiz	Lowland Kirghiz
	Frequency			
DYS390:				
19	.022	0	0	0
20	0	.018	0	0
21	.043	.018	0	.022
22	.130	.018	.021	0
23	.261	.071	.063	.156
24	.174	.179	.042	.111
25	.326	.607	.854	.644
26	.022	.089	.021	.067
27	.022	0	0	0
	Summary Statistics for DYS390			
	n = 46; *D* = .792	*n* = 56; *D* = .596	*n* = 48; *D* = .270	*n* = 45; *D* = .556
	Frequency			
DYS391:				
8	0	0	0	.021
9	.080	.071	0	.146
10	.660	.893	.188	.354
11	.200	.018	.812	.479
12	.060	.018	0	0
	Summary Statistics for DYS391			
	n = 50; *D* = .525	*n* = 56; *D* = .200	*n* = 48; *D* = .312	*n* = 48; *D* = .637
	Frequency			
DYS392:				
9	0	0	0	.021
10	.123	.018	0	.043
11	.571	.928	.874	.745
12	.082	.036	.021	.021
13	.041	0	0	.085
14	.163	.018	.063	.085
15	.020	0	.042	0
	Summary Statistics for DYS392			
	n = 49; *D* = .636	*n* = 56; *D* = .139	*n* = 48; *D* = .235	*n* = 47; *D* = .437
	Frequency			
DYS393:				
11	.073	0	0	.125
12	.145	.018	.021	.021
13	.564	.946	.958	.750
14	.145	.036	.021	.104
15	.055	0	0	0
16	.018	0	0	0
	Summary Statistics for DYS393			
	n = 55; *D* = .643	*n* = 56; *D* = .105	*n* = 48; *D* = .083	*n* = 48; *D* = .419

allele frequency. This is formally equivalent to unbiased expected heterozygosity in an autosomal locus. Interpopulation variability was estimated for each locus through χ^2 contingency tests and F_{ST} (Wright 1951).

Haplotypes consisting of seven Y-chromosome STRs were constructed for each individual. Haplotype diversity was calculated as gene diversity, with allele frequency being substituted for haplotype frequency. The mean pairwise difference, in number of repeats, across all seven loci was computed within each population. This provides a relative measure of how closely related are STR haplotypes within each population, although it is not necessarily an accurate measure of the phylogenetic distance between Y chromosomes when specific populations are being compared. The detailed phylogeny of the Y chromosome (which is not the object of this report) would be better understood if account were taken of the haplotypes defined by the unique-event polymorphisms such as single-nucleotide polymorphisms or *Alu* insertions (Hammer et al. 1997; Jobling et al. 1997; Underhill et al. 1997).

Genetic homogeneity among populations was tested through analysis of molecular variance (AMOVA [Excoffier et al. 1992]), by means of the Arlequin package (Schneider et al. 1996). Genetic variance was estimated with the "sum of size differences" option, which takes into account the difference in allele size when the difference between two STR alleles is being estimated.

Results

Allele Definition and Sequence Variability of DYS388 Polymorphism

Sequence analysis of several alleles of the trinucleotide DYS388 showed that this marker comprises a simple repetitive structure of the core unit (ATA). Alleles designated as "129," "132," and "135" by Kayser et al. (1997) and Pérez-Lezaun et al. (1997) correspond, respectively, to 12, 13, and 14 repetitions of the core unit (table 1). In a previous study (Pérez-Lezaun et al. 1997), six alleles of this marker had been described. Two new alleles, one containing 10 repeats and the other containing 18 repeats, are reported in the present study's survey of samples from Central Asia. The allele frequency distribution is irregular in the populations studied to date, and an allele containing 16 repeats has not yet been found.

Allele Frequencies and Locus Informativeness

Table 1 shows the distribution of allele frequencies found. Allele frequency differences among Central Asian populations, as shown by χ^2 tests for each STR, are statistically significant in all eight loci (all values are $P < .005$).

Table 2 shows the number of alleles, gene diversity, and F_{ST} values for each of the STRs used in the construction of haplotypes. The mean gene-diversity value found in four Central Asian populations ($D = .432 \pm .088$) is very similar to that found in two Western European population samples (Catalans and Basques, $D = .435$ [Pérez-Lezaun et al. 1997]). For the four populations under study, the mean F_{ST} value for markers DYS19, DYS389I, DYS390, DYS391, DYS392, and DYS393 was .224 ± .079, which is an extremely high value compared with the mean F_{ST} value, .039 ± .015, found, for the same markers, between two European populations (Pérez-Lezaun et al. 1997). The difference is statistically significant ($P = .017$, by signed-rank test).

Table 2

Frequency of Alleles, *D,* and F_{ST} in Eight Y-Chromosome STR Loci in Four Central Asian Samples

Locus	No. of Alleles Found	D	F_{ST}
DYS19	6	.460	.075
DYS388	7	.418	.381
DYS389I	5	.415	.377
DYS389II	7	.561	.338
DYS390	9	.542	.110
DYS391	5	.410	.387
DYS392	7	.354	.092
DYS393	6	.298	.146
Mean	...	.432	.238

Thus, Central Asian populations seem to be particularly heterogeneous. Although F_{ST} values vary among markers, all of them are high. Even the lowest F_{ST} value found, that for DYS19, is clearly higher than the values found on a worldwide scale for classic genetic markers (Cavalli-Sforza et al. 1994), although this was expected, given the smaller effective population size of Y chromosomes compared with autosomes. If a correction for the smaller effective population size of the Y chromosome were to be applied (Pérez-Lezaun et al. 1997), then a Y-chromosome F_{ST} of .224, such as that found, would correspond to an autosomal F_{ST} of .067, which is much closer to the values actually found for autosomal markers. No significant relationship between F_{ST} and the number of different alleles of each locus was found ($r = -.437$, $P = .279$).

Genetic Structure of Central Asian Populations

AMOVA showed that, when all four Central Asian populations were treated as a single group, 79.40% ($P < .0001$) of the genetic variance was found within populations, whereas 20.60% ($P < .0001$) was attributable to differences among populations. When we separated highland populations (highland Kirghiz and Kazakhs) from lowland populations (Uighurs and lowland Kirghiz), 81.00% ($P < .0001$) of the variance was due to differences within populations, 25.04% ($P < .0001$) was due to differences between populations within altitude groups, and −6.04% ($P \approx 1$) was due to differences between altitude groups. When we grouped populations according to their language, three different groups resulted: Kirghiz, Kazakhs, and Uighurs. Then, 77.35% ($P < .0001$) of the variance could be ascribed to differences within populations, 6.49% ($P = .172$) was found to be due to differences between populations within groups, and 16.16% ($P < .0001$) was attributable to differences between groups. When we distinguished between Kirghiz and non-Kirghiz groups, we observed that 77.26% ($P < .0001$) of the variance was due to differences within populations, 14.63% ($P < .0001$) of the variance was due to differences between populations within groups, and 8.11% ($P = .666$) was due to differences between groups. When we took one population and compared it with the other three, the results were similar for all four possibilities: the fraction of variation found between populations within groups was relatively large and highly significant ($P < .0001$), whereas the variance between groups remained smaller and not statistically significant ($P > .05$). Thus, although the fraction of variation found among populations was high, it did not appear to be distributed according to a clear-cut pattern. The significance of this apportionment of the genetic variance, especially when compared with the results of a similar analysis conducted with mtDNA sequences in the same samples, will be discussed below.

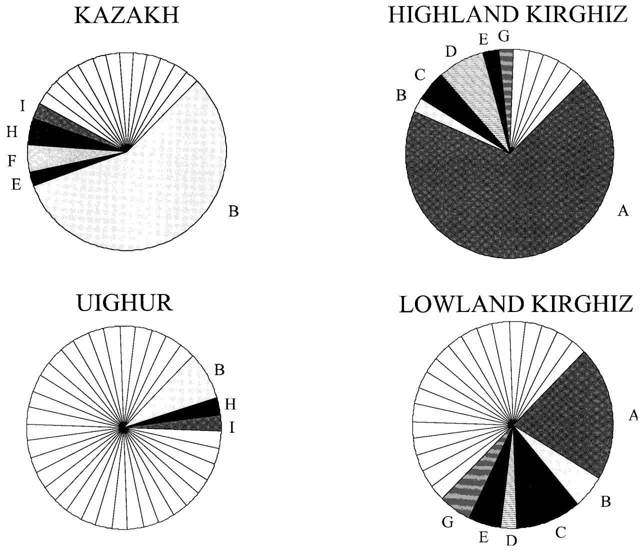

Figure 2 Y-chromosome STR haplotype frequencies in four Central Asian samples. Only haplotypes shared by more than one population have been labeled. Haplotypes for the DYS19-DYS388-DYS389I-DYS390-DYS391-DYS392-DYS393 loci are as follows: A, 16-12-11-25-11-11-13; B, 16-14-10-25-10-11-13; C, 16-12-11-25-10-11-13; D, 15-12-11-25-11-11-13; E, 14-12-11-23-10-14-14; F, 15-14-10-25-10-11-13; G, 16-12-11-26-11-11-13; H, 16-14-10-24-10-11-13; I, 16-14-10-25-11-11-13.

Haplotype Analysis

In a total of 172 complete seven-locus haplotypes in the four Central Asian populations, we observed 90 different haplotypes (listed in the Appendix), only one of which—namely, 16-14-10-25-10-11-13 for DYS19-DYS388-DYS389I-DYS390-DYS391-DYS392-DYS393 (haplotype B in fig. 2)—was found in all Central Asian samples. The four populations tested had dramatic differences in their haplotype diversities (fig. 2): the Uighurs presented 36 different haplotypes in 39 individuals (haplotype diversity [D] .995 ± .008), and the lowland Kirghiz had 25 haplotypes in 41 individuals ($D = .955 \pm .022$). In contrast, the two highland populations presented a much lower haplotype diversity: the Kazakhs had 18 different haplotypes in 49 individuals ($D = .738 \pm .069$), and the highland Kirghiz had 11 haplotypes in 43 individuals ($D = .545 \pm .091$). Moreover, both the Kazakhs and the highland Kirghiz each presented one haplotype at high frequency (haplotype 16-14-10-25-10-11-13 in the Kazakhs and haplotype 16-12-11-25-11-11-13 in the highland Kirghiz; haplotypes B and A, respectively, in fig. 2), and, with only one or a few putative mutation steps, most of the remaining haplotypes in those two populations could be derived from the most frequent haplotype. No such reduction of variability was observed in 360-bp mtDNA sequences of control region I in the same individuals and populations.

The average difference between different haplotypes within each population was measured as the mean pairwise difference in the number of repeats, across all seven loci. On average, two seven-locus haplotypes in the Uighurs have a difference of 7.81 repeats; these pairwise averages are 8.32 in the lowland Kirghiz, 5.52 in the Kazakhs, and 6.22 in the highland Kirghiz. Thus, haplotypes in the highland populations are, on average, more related to each other than are haplotypes in the lowland populations.

Discussion

The present study contains detailed information on the allele frequencies of eight trinucleotide and tetranucleotide STRs on the Y chromosome in four populations from Central Asia, a region previously poorly studied at the genetic level. This report gives information on the sequence structure of marker DYS388, which has been shown to be a quite polymorphic, trinucleotide-repeat polymorphisms on the Y chromosome. Population information on that marker contributes to its validation for forensic analysis and demonstrates its informativeness for human population genetics.

In this study we have been able to compare (*a*) the information provided by eight STRs for as many as 207 individuals from four populations in Central Asia with (*b*) the information on mtDNA control-region sequence data on the same individuals (Comas et al. 1998). This will allow us to discuss differences and similarities between results obtained from those nonrecombining regions of the human genome, and it underscores the importance of analyzing jointly both data on mtDNA and data on the Y chromosome, to help us unravel the genetic history of human populations (Salem et al. 1996;

Huoponen et al. 1997; Poloni et al. 1997; Bamshad et al. 1998; Passarino et al. 1998; Seielstad et al. 1998). It should be stressed that it is important to analyze the same populations and individuals, in order to obtain a clear picture of male-lineage versus female-lineage differentiation (Stoneking 1998).

Before we proceed to compare sequence mtDNA data with Y-chromosome STRs, we will argue that such a comparison is legitimate. It may be argued that sequence polymorphism and STRs evolve at very different rates and with different patterns, thus preventing any meaningful comparison. STRs have higher mutation rates than do nucleotide changes, and they mutate in a stepwise fashion (Weber and Wong 1993), with repeated mutations resulting in homoplasy. However, the latest empirical estimate of mutation rates in Y-chromosome STRs, derived from the overall observation of 1,088 father-son transmissions (Bianchi et al. 1998), is 1.2×10^{-3}, or a combined 8.4×10^{-3} for the seven loci. Phylogenetic mutation rate estimates for the hypervariable region I (HVR-I) of mtDNA are on the order of 10^{-3} (Forster et al. 1996), with genealogical estimates as high as $1.1–1.8 \times 10^{-2}$ (Parsons et al. 1997, although this frequency has been questioned by Jazin et al. 1998). Thus, mutation rates in the hypervariable regions of mtDNA do not seem to be significantly slower than those of Y-chromosome STRs, and, in populations that have split during relatively recent times, a differential accumulation of genetic variation due to the different mutation rates does not seem likely. In the differentiation of human populations, drift has had a deeper effect than mutation, as has been clearly demonstrated for a wide range of autosomal STRs (Pérez-Lezaun et al. 1997). Owing to the lower effective population sizes of both mtDNA and the Y chromosome relative to autosomes, mutation is likely to have been even less important than drift in the differentiation of human groups, especially in those groups that have a recent common origin, such as the Central Asian populations that we have analyzed. Recurrent mutation leading to homoplasy is also a common occurrence in the mtDNA control region: in HVR-I, Wakeley (1993) identified 29 nucleotide positions that had undergone repeated mutations in very different sequence backgrounds.

Both mtDNA and the Y chromosome contain markers that mutate more slowly than those that we used—namely, nucleotide variation outside the control region in mtDNA (Torroni et al. 1992; Macaulay et al. 1999) and biallelic polymorphisms in the Y chromosome (Jobling et al. 1997). Both sets of markers have been used to define haplogroups in their respective molecules. When the variation in the faster loci has been analyzed in a haplogroup framework, a good correlation has been found: RFLP haplogroups in mtDNA tend to bear specific control-region motifs (Macaulay et al. 1999), and Malaspina et al. (1998) have found that 80% of haplotypes defined by just four Y-chromosome STRs were not shared between the haplogroups defined by two biallelic polymorphisms (i.e., the YAP Alu insertion and the alphoid *Hin*dIII site). Thus, the faster markers retain most of the phylogenetic information on the genealogy of the slower sites, and, on average, two related Y-chromosome STR haplotypes will be borne by chromosomes close to each other in the Y-chromosome phylogeny. In summary, a relatively high mutation rate and homoplasy happen *both* in the Y-chromosome STRs *and* in the mtDNA control-region sequences, and it is not to be expected that they would differentially affect the two genetic systems. Y-chromosome STRs and mtDNA control-region sequences have been compared successfully in the recent work, by Bamshad et al. (1998), on gene flow among Indian castes.

The apportionment of genetic diversity in Y-chromosome STR haplotypes showed two main features: (1) extremely high levels of interpopulation diversity when compared with mtDNA sequences and (2) a dramatic reduction of internal genetic diversity in the two highland populations. Interpopulation diversity was extreme whether measured by F_{ST} or by AMOVA. Average F_{ST} by locus was .224, which reduces to .067 when a correction for the smaller effective population size of Y chromosomes compared with autosomes is applied, as has been suggested by Pérez-Lezaun et al. (1997). This is a relatively high value for neighboring populations speaking closely related languages: it is roughly half of the worldwide variation for autosomal *classic* markers (i.e., blood groups and protein electromorphs) and is six times larger than F_{ST}, for the same Y-chromosome STRs, between Basques and Catalans, who speak languages belonging to two separate families. AMOVA shows that ~20% of the haplotype genetic variation is found among populations (equivalent to $F_{ST} \cong .2$), whereas 15% is the *worldwide* average for autosomal STRs (Barbujani et al. 1997). Moreover, a 360-bp stretch of the hypervariable segment I of the mtDNA control region sequenced in the same individuals showed that only 0.5% of the genetic variation could be attributed to interpopulation differences. Three types of reasons could account for this differential behavior of Y-chromosome STRs: (i) intrinsic properties of Y-chromosome STRs; (ii) a different effective population size for males versus females (in which case, high levels of polygyny could account for a lower effective population size for males); and (iii) a differential pattern of male versus female migration.

A high mutation rate of Y-chromosome STRs could account for a rapid differentiation of Y-chromosome haplotypes in populations that have split recently. However, pedigree analysis has shown that the mutation rate of Y-chromosome STRs (Heyer et al. 1997; Bianchi et al. 1998) is consistent with the average mutation rate in autosomal STRs. Moreover, Pérez-Lezaun et al. (1997) have shown that, for two European populations, the levels of differentiation of Y-chromosome and autosomal STRs are almost identical. Thus, it does not seem likely that the high degree of interpopulation diversity shown by the Y-chromosome STRs should be attributed to any intrinsic properties.

As we will show below, the relative reduction of effective population size for males that is generated by polygyny or other causes made a very small contribution to the observed pattern in Y-chromosome intra- and interpopulation differentiation. The levels of polygyny vary greatly across cultures, although there are important similarities within regional areas (i.e., distances <1,500 km; White 1988). That allows us to consider the existing ethnographic data for Kazakhs to be representative for the other populations in our study. The Kazakhs, Kirghiz, and Uighurs are Moslems, although religious practice was suppressed during the Soviet period, and polygyny was forbidden. As reported in White's (1988) compilation of data on polygyny, the Kazakhs were ethnographically surveyed in 1885, and, at that time, they practiced polygyny, although the practice was limited to the top ~10% of men of the highest social rank. The number and distribution of wives per husband is not known, but harems were absent. Sororal polygyny was not practiced. When the worldwide average of wives per polygynous husband (i.e., 2.6) is taken into account, the ratio *R*—that is, male:female effective population size—is .862. When the equation suggested by Pérez-Lezaun et al. (1997) is generalized, the relation between male-lineage F_{ST} (F_{STm}) and female-lineage F_{ST} (F_{STf}) is given by the formula $F_{\mathrm{STm}} = F_{\mathrm{STf}}/[F_{\mathrm{STf}} + R(1 - F_{\mathrm{STf}})]$. For $F_{STf} = .005$, as observed for mtDNA sequences in Central Asia, the corresponding male-mediated F_{ST} is .0058, far from the observed .2. A similar correction can be applied to haplotype diversity, with the same result: after correction for polygyny, the mtDNA haplotype diversity in the highland Kirghiz changes from .984 to .981, still far from the observed Y-chromosome diversity ($D = .545$). In summary, polygyny had an almost negligible contribution to both the reduction of intrapopulation Y-chromosome diversity and the increase of Y-chromosome interpopulation differentiation.

An additional factor that could reduce the relative effective population size for males is a higher male prereproductive mortality, caused by hunting or warfare. A distribution of ages at death, by sex, could be obtained from burial remains, but such data are not available for our populations. However, ethnographic observation during recent decades does not show evidence of a higher male prereproductive mortality (O. Ismagulov, personal communication), and this seems to be a general pattern for most of ancient human populations (Kelly 1995, and references therein).

A traditional structure of patrilocal marriages—that is, those in which the custom is for the bride to move to the groom's village—would result in a very high female:male ratio of short-distance migration. Patrilocal marriages are indeed practiced by the Kazakhs and the Kirghiz (O. Ismagulov, personal communication), as is a strong paternal-clan exogamy, and our observations confirm that the prescribed matrimonial pattern has a strong impact on the genetic structure of the populations (Bamshad et al. 1998). This mechanism could explain the observed disparity, in interpopulation differentiation, between Y-chromosome and mitochondrial markers, and it also has been invoked by Salem et al. (1996), to explain similar patterns of Y-chromosome and mtDNA diversity in Sinai Bedouins, and by Seielstad et al. (1998), in a global survey. The rates of male and female migration may be very different because of the cultural practices in different world populations, and it is to be expected that the relative levels of Y-chromosome and mtDNA interpopulation differentiation would vary accordingly. Moreover, genetic analysis may be able to detect ancient, extinct cultural practices not observed by social scholars.

When we grouped the four Central Asian populations in higher-level hierarchies, we found that altitude contributed a negative fraction of genetic variability; that is, there was more variability between populations at the same altitude than between the two altitude groups. It seems unlikely, then, that altitude has exerted a common selective pressure on the Y-chromosome genes of the highland Kazakhs and the Kirghiz; a similar pattern was observed in the mtDNA of the same populations (Comas et al. 1998). When populations were grouped according to language, 16% of Y-chromosome genetic diversity could be attributed to linguistic groups, whereas this fraction was negligible (−0.5%) for mtDNA sequences. Then, it seems that female migration may overcome linguistic barriers with greater ease than does male migration. Zones of sharp *autosomal* genetic change in Europe appear to overlay linguistic boundaries (Barbujani and Sokal 1990), and Poloni et al. (1997) have shown that, in populations worldwide, linguistic distances present a higher correlation with Y-chromosome than with mtDNA markers. Both this finding and our results seem to indicate that genetic isolation at linguistic boundaries would be largely contributed by male, rather than female, isolation.

The level of haplotype diversity detected in the Y chromosomes of Central Asia appears to be linked to the altitudinal habitat of the populations: the highland dwellers have reduced haplotype diversities when compared with the lowland Kirghiz and Uighurs. The highland Kirghiz and the Kazakhs each present both a different haplotype at very high frequency and a reduced number of other haplotypes that may derive from the main haplotypes by one mutation event. Haplotype 16-12-11-25-11-11-13 (haplotype A in fig. 2) was found at high frequencies in the highland Kirghiz and was also found at lower frequencies in the lowland Kirghiz. The most common haplotype in the Kazakhs, 16-14-10-25-10-11-13 (haplotype B in fig. 2), is also the only haplotype present in all four samples. The most frequent alleles in both Kirghiz samples for three loci (DYS389I, DYS389II, and DYS390) are the same, and they are different from the most frequent allele elsewhere in Europe and Asia. It seems likely, then, that the Y-chromosome pool of the high-altitude populations reflects a relatively recent foundation event. The Keghen Valley, where the highland Kazakhs live, was colonized 200–400 years ago by the lowland Kazakhs (O. Ismagulov, personal communication), and the present inhabitants consider themselves to be direct descendants of the original colonizers. The Kirghiz seem to have come to the high valleys of the Tien Shan, Pamir, Alay, and Qara Qorum no earlier than the 16th or 17th centuries (Menges 1994), whereas for the Talas valley an archaeological record dating back to the Bronze Age is available.

The reduction in Y-chromosome STR haplotype diversity in the highland dwellers could have been caused by selection operating on some gene in the Y chromosome and thereby reducing the diversity on the whole chromosome. This has not been observed in other populations or in general surveys. However, AMOVA showed that altitude contributed a nonsignificant, negative fraction of the genetic variation, and this makes it unlikely that altitude exerted a common selective pressure on highland Kirghiz and Kazakh Y chromosomes. Therefore, a founder effect could explain better the low levels of Y-chromosome genetic diversity in highlanders.

In contrast with what is seen for the Y chromosome, the same highland populations do not show a reduction in the diversity of mtDNA sequences (Comas et al. 1998). A historically higher effective population size for females could be achieved in a number of not necessarily exclusive ways. It is possible that the highland populations were founded by groups containing a few, possibly related males and a higher number of females. Alternatively or concomitantly, exogamic patrilocal marriages could have resulted in an inflow of females that could have restored the mitochondrial genetic diversity.

The lowland populations present a very high internal genetic diversity for *both* the mtDNA *and* the Y chromosome. The mtDNA diversity is among the highest in Eurasia, and the same is true for the Y-chromosome diversity. Comas et al. (1998) have concluded that the high mtDNA diversity, along with other features of the Central Asian mtDNA sequence pool, could be a result of admixture between the already differentiated populations of Europe and eastern Asia. It is likely that the Central Asian gene pool has been shaped by high levels of female-mediated migration, possibly related to patrilocal marriages within and between linguistic and altitudinal populations, a situation that would have maintained high levels of mtDNA diversity in all populations, whereas the male-mediated migration has been less intense and has occurred mainly in high-altitude populations, where a male founding effect seems clear. In conclusion, the joint analysis of Y-chromosome and mtDNA genetic material has allowed us to observe the genetic outcome of sex-specific migration patterns and has revealed itself as a powerful tool in the analysis of human population history.

Acknowledgments

This research was supported mainly by Dirección General de Investigación Científico Técnica (Spain) grant PB95-0267-C02-01. Additional support was provided by Human Capital and Mobility contracts (network ERCHRXCT92-0032 and ERB-CHRX-CT920090) to J.B., by Direcció General de Recerca, Generalitat de Catalunya, grants 1995SGR00205 and 1996SGR00041, and by the Institut d'Estudis Catalans. The CAHAP research project was partially supported by MURST grants to F.F. and D.P. and by private grants from "Fondazione della Cassa di Risparmio di Bologna." This work was also made possible thanks to Spanish Ministry of Education and Science fellowships to A.P.-L. (FP93), F.C. (postdoctoral return contract), and R.M.-A. (AP96); Universitat de Barcelona fellowships to E.M. and J.C.; and Comissionat per a Universitats i Recerca, Generalitat de Catalunya fellowships to E.B. (FI/96-1153) and F.C. Allele ladders and invaluable advice were provided by Peter de Knijff. We also acknowledge the help and technical assistance offered by the Unitat de Seqüenciació, Servei Científico-Tècnic, Universitat de Barcelona. Mark Seielstad reviewed an earlier version of the manuscript.

Appendix

STR Haplotypes in 172 Central Asian Individuals

In the list below, the data are absolute frequencies of the haplotypes.

DYS19	DYS388	DYS389I	DYS390	DYS391	DYS392	DYS393	Kazakhs	Highland Kirghiz	Lowland Kirghiz	Uighurs	All Four Populations
16	12	11	25	11	11	13	0	29	8	0	37
16	14	10	25	10	11	13	25	1	2	3	31
16	12	11	25	10	11	13	0	2	4	0	6
15	12	11	25	11	11	13	0	3	1	0	4
16	14	10	26	10	11	13	3	0	0	0	3
15	14	10	25	10	11	13	2	1	0	0	3
16	14	10	24	10	11	13	2	0	0	1	3
14	12	11	23	10	14	14	1	0	2	0	3
16	12	11	26	11	11	13	0	1	2	0	3
15	13	10	24	9	11	13	0	0	3	0	3
15	14	10	26	10	11	13	2	0	0	0	2
16	12	11	24	10	11	13	2	0	0	0	2
16	14	11	25	10	11	13	2	0	0	0	2
16	14	10	25	11	11	13	1	0	0	1	2
14	12	10	23	10	15	13	0	2	0	0	2
16	13	9	25	10	13	11	0	0	2	0	2
16	12	10	25	11	11	13	0	0	0	2	2
14	13	10	23	10	12	12	1	0	0	0	1
14	13	10	23	10	12	13	1	0	0	0	1
14	13	11	24	10	11	13	1	0	0	0	1
16	11	11	24	9	11	13	1	0	0	0	1
16	12	10	24	9	11	13	1	0	0	0	1
16	12	11	24	9	11	13	1	0	0	0	1
16	14	10	20	10	11	13	1	0	0	0	1
16	14	10	22	10	11	13	1	0	0	0	1
16	18	10	25	10	11	13	1	0	0	0	1
14	12	10	23	10	14	13	0	1	0	0	1
15	12	10	22	11	14	12	0	1	0	0	1
15	12	9	24	10	14	14	0	1	0	0	1
16	12	11	25	11	12	13	0	1	0	0	1
10	12	11	25	8	11	13	0	0	1	0	1
14	12	11	23	9	14	14	0	0	1	0	1
14	13	11	23	9	10	14	0	0	1	0	1
14	17	10	23	11	11	12	0	0	1	0	1
15	12	9	23	10	10	11	0	0	1	0	1
15	12	11	25	10	11	13	0	0	1	0	1
15	12	10	25	11	11	13	0	0	1	0	1
15	18	11	25	11	11	13	0	0	1	0	1
16	12	11	25	9	11	13	0	0	1	0	1
16	12	11	25	10	9	11	0	0	1	0	1
16	12	11	25	11	11	14	0	0	1	0	1
16	12	11	25	11	13	13	0	0	1	0	1
16	13	9	25	10	12	11	0	0	1	0	1
16	14	10	21	10	11	13	0	0	1	0	1
16	17	11	25	11	11	13	0	0	1	0	1
16	18	10	24	11	11	13	0	0	1	0	1
16	18	9	26	10	13	11	0	0	1	0	1
13	12	9	25	10	13	13	0	0	0	1	1
13	12	10	25	10	14	13	0	0	0	1	1
13	14	10	24	10	11	13	0	0	0	1	1
14	10	11	24	10	11	13	0	0	0	1	1
14	10	9	24	10	14	12	0	0	0	1	1
14	12	10	22	10	14	14	0	0	0	1	1
14	12	11	23	9	10	14	0	0	0	1	1
14	12	11	23	10	12	14	0	0	0	1	1

14	12	10	25	11	13	12	0	0	0	1	1
14	13	10	23	11	14	11	0	0	0	1	1
15	12	11	19	10	11	12	0	0	0	1	1
15	12	11	21	10	10	13	0	0	0	1	1
15	12	11	22	10	14	11	0	0	0	1	1
15	12	9	23	10	12	13	0	0	0	1	1
15	12	11	23	10	14	11	0	0	0	1	1
15	12	10	26	11	11	13	0	0	0	1	1
15	12	9	27	10	11	13	0	0	0	1	1
16	12	10	22	10	11	12	0	0	0	1	1
16	12	11	23	9	10	14	0	0	0	1	1
16	12	11	23	9	11	13	0	0	0	1	1
16	12	10	23	10	11	13	0	0	0	1	1
16	12	9	23	10	12	12	0	0	0	1	1
16	12	10	23	10	14	12	0	0	0	1	1
16	12	10	23	11	11	13	0	0	0	1	1
16	12	9	25	10	10	15	0	0	0	1	1
16	12	10	25	11	15	14	0	0	0	1	1
16	14	9	24	10	14	14	0	0	0	1	1
16	14	9	25	10	11	13	0	0	0	1	1
16	14	10	25	11	11	15	0	0	0	1	1
17	11	10	24	10	11	13	0	0	0	1	1
17	12	10	24	11	11	13	0	0	0	1	1
17	13	11	22	10	10	12	0	0	0	1	1

References

Bamshad MJ, Watkins WS, Dixon ME, Jorde LB, Rao BB, Naidu JM, Prasad BVR, et al (1998) Female gene flow stratifies Hindu castes. Nature 395:651-652

Barbujani G, Magagni A, Minch E, Cavalli-Sforza LL (1997) An apportionment of human DNA diversity. Proc Natl Acad Sci USA 94:4516-4519

Barbujani G, Sokal RR (1990) Zones of sharp genetic change in Europe are also linguistic boundaries. Proc Natl Acad Sci USA 87:1816-1819

Bianchi NO, Catanesi CI, Bailliet G, Martinez-Marignac VL, Bravi CM, Vidal-Rioja LB, Herrera RJ, et al (1998) Characterization of ancestral and derived Y-chromosome haplotypes of New World native populations. Am J Hum Genet 63:1862-1871

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

Comas D, Calafell F, Mateu E, Pérez-Lezaun A, Bosch E, Martínez-Arias R, Clarimon J, et al (1998) Trading genes along the Silk Road: mtDNA sequences and the origin of Central Asian populations. Am J Hum Genet 63:1824-1838

Cooper G, Amos W, Hoffman D, Rubinsztein D (1996) Network analysis of human Y microsatellite haplotypes. Hum Mol Genet 5:1759-1766

Deka R, Jin L, Shriver MD, Mei Yu L, Saha N, Barrantes R, Chakraborty R, et al (1996) Dispersion of human Y chromosome haplotypes based on five microsatellites in global populations. Genome Res 6:1177-1184

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479-491

Forster P, Harding R, Torroni A, Bandelt H-J (1996) Origin and evolution of Native American mtDNA variation: a reappraisal. Am J Hum Genet 59:935-945

Hammer MF, Spurdle AB, Karafet T, Bonner MR, Wood ET, Novelletto A, Malaspina P, et al (1997) The geographic distribution of human Y chromosome variation. Genetics 145: 787-805

Hammer M, Zegura SL (1996) The role of the Y chromosome in human evolutionary studies. Evol Anthropol 5:116-134

Hassan FA (1981) Demographic archaeology. Academic Press, New York

Heyer E, Puymirat J, Dieltjes P, Bakker E, de Knijff P (1997) Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees. Hum Mol Genet 6:799-803

Huoponen K, Torroni A, Wickman P, Sellito D, Gurley DS, Scozzari R, Wallace D (1997) Mitochondrial DNA and Y chromosome-specific polymorphisms in the Seminole tribe of Florida. Eur J Hum Genet 5:25-34

Jazin E, Soodyall H, Jalonen P, Lindholm E, Stoneking M, Gyllensten U (1998) Mitochondrial mutation rate revisited: hot spots and polymorphism. Nat Genet 18:109-110

Jobling MA, Pandya A, Tyler-Smith C (1997) The Y chromosome in forensic analysis and paternity testing. Int J Legal Med 110:118-124

Kayser M, Caglià A, Corach D, Fretwell N, Gehrig C, Graziosi G, Heidorn F, et al (1997) Evaluation of Y-chromosomal STRs: a multicenter study. Int J Legal Med 110:125-133, 141-149

Kelly RL (1995) The foraging spectrum: diversity in hunter-gatherer lifeways. Prentice Hall, New York

Macaulay V, Richards M, Hickey E, Vega E, Cruciani F, Guida V, Scozzari R, et al (1999) The emerging tree of west Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. Am J Hum Genet 64:232-249

Malaspina P, Cruciani F, Ciminelli BM, Terrenato L, Santolamazza P, Alonso A, Banyko J, et al (1998) Network analyses of Y-chromosomal types in Europe, northern Africa, and western Asia reveal specific patterns of geographic distribution. Am J Hum Genet 63:847-860

Menges KH (1994) People, languages and migrations. In: Allworth E (ed) Central Asia: 130 years of Russian dominance, a historical overview, 3d edition. Duke University Press, Durham, NC, pp 60-91

Parsons TJ, Muniec DS, Sullivan K, Woodyatt N, Alliston-Greiner R, Wilson MR, Berry DL, et al (1997) A high observed substitution rate in the human mitochondrial DNA control region. Nat Genet 15:363-368

Passarino G, Semino O, Quintana-Murci L, Excoffier L, Hammer M, Santachiara Benericetti AS (1998) Different genetic components in the Ethiopian population, identified by mtDNA and Y-chromosome polymorphisms. Am J Hum Genet 62:420-434

Pena SD, Santos FR, Bianchi NO, Bravi CM, Carnese FR, Rothhammer F, Gerelsaikhan T, et al (1995) A major founder Y-chromosome haplotype in Amerindians. Nat Genet 11:15-16

Pérez-Lezaun A, Calafell F, Seielstad M, Mateu E, Comas D, Bosch E, Bertranpetit J (1997) Population genetics of Y-chromosome short tandem repeats in humans. J Mol Evol 45:265-270

Pestoni C, Cal ML, Lareu MV, Rodríguez-Calvo MS, Carracedo A (1999) Y chromosome STR haplotypes: genetic and sequencing data of the Galician population (NW Spain). Int J Legal Med 112:15-21

Poloni ES, Semino O, Passarino G, Santachiara-Benerecetti AS, Dupanloup I, Langaney A, Excoffier L (1997) Human genetic affinities for Y-chromosome p49a,f/*Taq*I haplotypes show strong correspondence with linguistics. Am J Hum Genet 61:1015-1035

Rolf B, Meyer E, Brinkmann B, de Knijff P (1998) Polymorphism at the tetranucleotide repeat locus DYS389 in 10 populations reveals strong geographic clustering. Eur J Hum Genet 6:583-588

Salem A-H, Badr FM, Gaballah MF, Pääbo S (1996) The genetics of traditional living: Y-chromosomal and mitochondrial lineages in the Sinai Peninsula. Am J Hum Genet 59: 741-743

Santos FR, Pena SDJ, Epplen JT (1993) Genetic and population study of a Y-linked tetranucleotide repeat DNA polymorphism with a simple non-isotopic technique. Hum Genet 90:655-656

Schneider S, Kueffer JM, Roessli D, Excoffier L (1996) Arlekin (version 1.0): a software environment for the analysis of population genetics data. Genetics and Biometry Lab, University of Geneva, Geneva

Seielstad M, Minch E, Cavalli-Sforza LL (1998) Genetic evidence for a higher female migration rate in humans. Nat Genet 20:278-280

Stoneking M (1998) Women on the move. Nat Genet 20:219-220

Torroni A, Schurr TG, Yang C-C, Szathmary EJE, Williams RC, Schanfield MS, Tromp GA, et al (1992) Native mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. Genetics 130:153-162

Underhill PA, Jin L, Lin A, Medhdi SQ, Jenkins T, Vollrath D, Davis RW, et al (1997) Detection of numerous Y chromosome biallelic polymorphisms by denaturing high performance liquid chromatography (DHPLC). Genome Res 7: 996-1005

Underhill PA, Jin L, Zemmans R, Oefner PJ, Cavalli-Sforza LL (1996) A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history. Proc Natl Acad Sci USA 93:196-200

Wakeley J (1993) Substitution rate variation among sites in hypervariable region I of human mitochondrial DNA. J Mol Evol 37:613-623

Weber JL, Wong C (1993) Mutation of human short tandem repeats. Hum Mol Genet 2:1123–1128

White DR (1988) Rethinking polygyny: co-wives, codes, and cultural systems. Curr Anthropol 29:529-558

Wright S (1951) The genetical structure of populations. Ann Eugenics 15:323-354