Am. J. Hum. Genet. 74:50–61, 2004

The Effective Mutation Rate at Y Chromosome Short Tandem Repeats, with Application to Human Population-Divergence Time

Lev A. Zhivotovsky,[1] Peter A. Underhill,[2] Cengiz Cinnioğlu,[2] Manfred Kayser,[4] Bharti Morar,[5] Toomas Kivisild,[6] Rosaria Scozzari,[7] Fulvio Cruciani,[7] Giovanni Destro-Bisol,[8] Gabriella Spedini,[8] Geoffrey K. Chambers,[9] Rene J. Herrera,[10] Kiau Kiun Yong,[5] David Gresham,[5] Ivailo Tournev,[11,12] Marcus W. Feldman,[3] and Luba Kalaydjieva[5]

[1]N. I. Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow; Departments of [2]Genetics and [3]Biological Sciences, Stanford University, Stanford, CA; [4]Max Planck Institute for Evolutionary Anthropology, Leipzig; [5]Western Australian Institute for Medical Research and Centre for Medical Research, University of Western Australia, Perth; [6]The Estonian Biocentre, Tartu, Estonia; Departments of [7]Genetics and Molecular Biology and [8]Human and Animal Biology, Section of Anthropology, University of Rome "La Sapienza," Rome; [9]Institute for Molecular Systematics, School of Biological Sciences, Victoria University, Wellington, New Zealand; [10]Department of Biological Sciences, Florida International University, Miami; and [11]Department of Neurology, Medical University, and [12]Foundation for Health Problems of Ethnic Minorities, Sofia, Bulgaria

We estimate an effective mutation rate at an average Y chromosome short-tandem repeat locus as 6.9×10^{-4} per 25 years, with a standard deviation across loci of 5.7×10^{-4}, using data on microsatellite variation within Y chromosome haplogroups defined by unique-event polymorphisms in populations with documented short-term histories, as well as comparative data on worldwide populations at both the Y chromosome and various autosomal loci. This value is used to estimate the times of the African Bantu expansion, the divergence of Polynesian populations (the Maoris, Cook Islanders, and Samoans), and the origin of Gypsy populations from Bulgaria.

Introduction

Microsatellites, or STR polymorphisms, are abundant in the human genome and can be easily genotyped and scored; they have thus become a useful tool for the elucidation of human population history and for forensic purposes. Knowledge of the mutation rate at STR loci is important, both for calibration of the molecular clock in evolutionary studies and for forensic probabilistic calculations. Increasing attention has recently been paid to microsatellite variation within Y chromosome haplogroups defined by binary polymorphisms, such as SNPs or, as a general term, "unique-event polymorphisms" (UEPs) (Underhill et al. 1996; Zerjal et al. 1997; de Knijff 2000; Kayser et al. 2000*a*), many of which are specific to populations related through their recent or distant history (Underhill et al. 2000; Hammer et al. 2001; Y Chromosome Consortium 2002). Although the Y chromosome locus is the ultimate SNP-STR system, similar linked SNP and STR haplotypes are also available in autosomes (Mountain et al. 2002).

A mutation rate of 2×10^{-3} per generation has been estimated for Y chromosome microsatellites by direct count in deep-rooted pedigrees (Heyer et al. 1997). A similar average mutation rate value of 3×10^{-3} per locus per generation was estimated by studying Y chromosome STRs (Y-STRs) in father/son pairs of confirmed paternity, although locus-specific values varied from 0 to 8×10^{-3} (Kayser et al. 2000*b*). Y-STR analysis in sperm revealed an average rate of repeat gains of 2×10^{-3} for two Y-STRs (Holtkemper et al. 2001); however, mutations that included repeat losses could not be considered owing to limitations of the methodology used. No germline mutation was observed in Y-STRs from cell line DNA (Bianchi et al. 1998), but this result does not differ in a statistically significant way from the mutation rate estimates mentioned above.

By counting the number of mutations in the branches of a haplotype network from samples of Native American populations, Forster et al. (2000) found a striking difference between their "evolutionary" estimate (2.6×10^{-4} per 20 years) and the "pedigree" estimate described above. It is unclear which rate should be used; for evolutionary studies, we need to know those mutations that are involved in differences between lineages or populations. An inappropriate choice of the mutation rate value may produce a 10-fold deviation from the true age of past population events. The discordance between the two kinds of estimate needs to be addressed.

The estimate by Forster et al. (2000) refers to a median network constructed from the Y chromosome haplotypes found in the combined data from Native Amer-

Received August 1, 2003; accepted for publication October 15, 2003; electronically published December 19, 2003.

Address for correspondence and reprints: Dr. Lev Zhivotovsky, Institute of General Genetics, 3 Gubkin Street, Moscow 119991, Russia. E-mail: lev@zhivotovsky.net

0002-9297/2004/7401-0006$15.00

ican populations. However, haplotype history might not represent the population history. In addition, such a network assumes single-repeat–unit mutational changes. Multistep Y-STR mutations, which have been observed recently (Forster et al. 1998; Kayser et al. 2000*b;* Nebel et al. 2001), can contribute significantly to the *effective* mutation rate (w, the product of the mutation rate and the variance of mutational changes in repeat scores), which determines the rate of microsatellite evolution (Slatkin 1995; Zhivotovsky and Feldman 1995). Furthermore, Forster et al. (2000) used for calibration an estimate of the time of population expansion in North America of 20,000 before the present (BP)—an estimate upon which there is no general agreement.

In the present study, we estimate the effective mutation rate, using data on microsatellite variation within Y chromosome haplogroups defined by SNPs in populations with documented short-term histories, as well as comparative data on worldwide populations at both autosomal and Y chromosome loci. Then we apply our finding to estimate the time of expansion of Bantu-speaking populations in sub-Saharan Africa, the time of differentiation of some Polynesian populations, and the time of origin of Bulgarian Gypsy populations.

Material and Methods

Samples and DNA Data

To estimate the effective mutation rate, we use here three data sets. The first includes two Polynesian populations: Maoris and Cook Islanders. The Maori are a tribe of Polynesians who arrived in New Zealand during the Polynesian expansion, not later than 800 years BP (Diamond and Bellwood 2003); this date is used here as the time of colonization of New Zealand by the Maoris. Although it has been shown elsewhere that some Polynesian men carry European Y chromosomes (Hurles et al. 1998; Underhill et al. 2001*a*), the majority of Polynesian Y chromosome lineages reflect Polynesian origin in Melanesia and eastern/southeastern Asia (Su et al. 2000; Kayser et al. 2000*a;* Underhill et al. 2001*a*). In the present study, we have used lineage C2 (characterized by mutations at RPS4Y711 and M38), marked additionally with the mutation M208, originally described in 42 Polynesian individuals (Kayser et al. 2000*a;* Underhill et al. 2001*a*). Our sample included 22 Maori and 23 Cook Islander men whose Y chromosomes carry M208 (table 1; fig. 1). The M208 mutation is most likely of Melanesian origin (Kayser et al. 2003), and individuals carrying the M208 mutation have been observed, so far, only in Melanesia (West New Guinea highlands, Papua New Guinea coast, and Trobriand Islands) and Polynesia (Cook, Maori, and Samoa). M208 is the most frequent Y chromosomal haplogroup in Polynesia (Kayser et al. 2003; Underhill et al. 2001), although it is fixed or nearly fixed in two linguistically closely related groups from the highlands of West New Guinea (Kayser et al. 2003). M208-associated Y-STR diversity was observed to be relatively low (Kayser et al. 2003), suggesting the recent origin of M208.

The second data set contains samples from Bulgarian Gypsy (Roma) populations. The Gypsies are of Indian origin, and the population split into numerous endogamous groups after arrival in Europe 900–1,000 years BP (Fraser 1992). Gypsy presence in Bulgaria was recorded ~700 years BP (Marushiakova and Popov 1997). Here, we have analyzed 179 individuals from 12 Gypsy groups from Bulgaria (table 2 and 3) who represent a Y chromosome lineage, defined by mutation M82, that is derived from the Indian subcontinent and is exceedingly rare in Europe (Semino et al. 2000; Underhill et al. 2000; Gresham et al. 2001). The populations that we have analyzed include the Lom, Koshnichari (south-central Bulgaria), Koshnichari (southwestern Bulgaria), Turgovzi, and Feredjelli (see table 2 of Gresham et al. [2001]). For the present study, we analyzed data from previously reported populations (Gresham et al. 2001), and we expanded the sample size of the following populations: the Rudari (encompassing the Lingurari North, Lingurari South, Intreni, and Monteni populations), Kalderash, Kalaidjii (southern Bulgaria), and Kalaidjii (northern Bulgaria). Three additional populations—namely the Blacksmiths, Darakchii, and Musicians—are analyzed for the first time in the present study.

The third data set includes variation at 58 tri- and 274 tetranucleotide autosomal microsatellites and at 2 tri- and 5 tetranucleotide-repeat STRs on the Y chromosome in 52 worldwide populations. Information on the genotypes and populations used in this study is available at the Human Diversity Panel Genotypes Web site (Weber and Broman 2001). The populations are described by Cann et al. (2002), and details about the genetic variation at autosomal microsatellite loci in those populations are reported by Rosenberg et al. (2002) and Zhivotovsky et al. (2003).

For estimation of population-divergence time, we investigate the E3a7-M191 haplogroup, which occurs at high frequency in the Bantu populations, with traces in other, non-Bantu-speaking groups from sub-Saharan Africa. A total of 148 individuals with the M191 mutation were analyzed. They included the following sub-Saharan African populations: Mossi (Burkina Faso, $N = 11$), Biaka Pygmy (Central African Republic, $N = 6$), Mbuti Pygmy (Zaire, $N = 4$), Fali (North Cameroon, $N = 13$), Bakaka (South Cameroon, $N = 4$), Bamikele (South Cameroon, $N = 27$), Ewondo (South Cameroon, $N = 6$), and !Kung San (South Africa, $N = 10$) (as described by Cruciani et al. [2002]), Zulu (South Africa, $N = 4$), and, in addition, 16 males from various

Table 1

The Distribution of Haplotypes with Mutation M208 in Samples from the Maori, Cook Islander, and Samoan Populations

	No. of Carriers at Locus										No. of Carriers among		
Haplotype	DYS19	DYS388	DYS389-CD	DYS389-AB	DYS390	DYS391	DYS392	DYS393	DYS439	DYSA7.2	Maori	Cook Islander	Samoan
PA1	16	15	12	17	20	10	12	14	13	9	8	0	0
PA2	16	15	12	17	20	10	12	14	14	9	6	8	0
PA3	16	15	12	17	20	10	12	14	15	9	1	0	0
PB	15	15[a]	12	17	20	10	12	14	13[a]	9[a]	1	1	0
PC	16	15[a]	12	18	20	10	12	14[b]	13[a]	9[a]	1	1	0
PD	16	15[a]	13	17	20	11	12	14	12[a]	10[a]	1	1	0
PE	16	15	13	17	21	11	12	14	11	10	1	0	0
PF	16	…	13	17	20	10	12	14	…	…	0	6	0
PG	16	…	12	17	20	10	13	14	…	…	0	2	0
PH	15	…	12	17	21	10	12	13	…	…	0	1	0
PI	14	…	13	17	20	10	12	14	…	…	0	2	0
PJ	16	…	12	17	21	10	12	14	…	…	0	1	0
PK	16	15	13	17	19	10	12	14	13	9	0	0	1
PL	16	15	14	17	20	10	12	14	13	9	0	0	1
PM	16	15	13	17	20	?[c]	12	14	12	9	0	0	1
PN	16	15	13	17	21	10	12	14	13	9	0	0	1
PO[d]	15	15	13	17	20	11	12	14	12	10	1	0	0
PP[d]	16	15	12	18	20	10	12	14	13	9	1	0	0
PQ[d]	16	15	13	17	20	10	12	14	14	9	1	0	0

[a] Not determined in the Cook Islanders.
[b] Not determined in the Maori
[c] Not determined in Samoan.
[d] Individuals of mixed (European/Maori) origin were removed from the analysis; empty spaces indicate that the Cook Islanders have not been typed at loci DYS388, DYS439, and DYSA7.2.

other Bantu-speaking South African populations represented by <4 sampled individuals, described in Underhill et al. (2000). Also included were unpublished samples from Kenya ($N = 9$) and Rwanda ($N = 38$). The individuals were genotyped at 10 STR loci; the data are available in table A (online only). We also consider the same SNP lineage C2 as used above (marked with mutation M208), to estimate divergence times of the Polynesian Maoris, Cook Islanders, and Samoans; the latter sample includes four individuals (table 1).

Genotyping Methods

The binary Y chromosome mutation M208 (Underhill et al. 2001*b*) was genotyped in relevant Maori, Cook Islander, and Samoan samples that were reported elsewhere to carry the M38 mutation (Underhill et al. 2001*a*); genotyping was performed by use of denaturing high-performance liquid chromatography (DHPLC) (Oefner and Underhill 1998) or PCR-RFLP (Kayser et al. 2003). The M191 transversion was also typed by DHPLC. The M82 indel mutation was detected using a 5′-labeled fluorescent primer and length separation on an ABI 377 DNA analyzer.

The following Y chromosome microsatellite markers were analyzed using PCR amplification with 5′ fluorescent labeling of one of the primers, length separation on ABI 3100 or 377 DNA Analyzers, and GeneScan fragment-analysis software: DYS19, DYS388, DYS389I, DYS389II, DYS390, DYS391, DYS392, and DYS393 (Kayser et al. 1997), as well as DYS439 (Ayub at al. 2000) and DYSA7.2 (White et al. 1999). DYSA7.2 is referred to as "DYS461" by Bosch et al. (2002).

Markers DYS388, DYS439, and DYSA7.2 were not analyzed in the Cook Islanders, and DYS439 and DYSA7.2 were not analyzed in the Gypsies.

Statistical Analysis

For estimation of w from the first two data sets, we use the $(\delta\mu)^2$ distance (Goldstein et al. 1995), whose expected value is $2\mu t$ if mutations are single-step and is $2wt$ for a general mutation scheme (Zhivotovsky and Feldman 1995); here, w is the effective mutation rate, and t is the time since populations separated. Another measure used is the average squared difference in the number of repeats between all current chromosomes

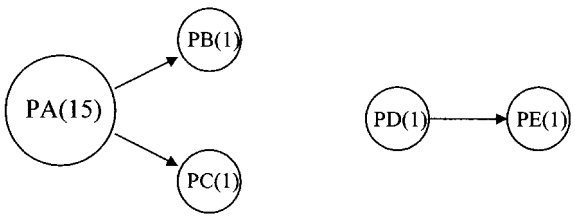

Figure 1 Network of Maori haplotypes (without loci DYS439, DYS388, and DYSA7.2). The number of chromosomes is shown in parentheses (see table 1).

Table 2

List of Haplotypes with Mutation M82 in the Gypsy Populations

	No. of Carriers at Locus							
Haplotype	DYS19	DYS388	DYS389I (CD)	DYS389II (AB)	DYS390	DYS391	DYS392	DYS393
A	15	12	14	16	22	10	11	12
B	14	12	14	16	22	10	11	12
C	15	12	14	16	23	10	11	12
D	15	12	14	16	22	10	11	13
E	14	12	14	16	22	9	11	12
F	15	12	14	17	22	10	11	12
G	15	12	13	16	22	10	11	12
H	15	12	14	16	21	10	11	12
I	15	12	15	16	22	10	11	12
J	15	12	14	15	22	10	11	12
K	15	10	14	16	22	10	11	12
L	14	12	14	17	22	10	11	12
M	15	12	14	16	22	11	11	12
N	16	12	14	16	22	10	11	12
O	15	12	13	17	22	10	11	12
P	15	12	14	17	23	10	11	12
Q	15	12	16	17	22	10	11	12

of a sample and the founder haplotype, denoted by "ASD_0," which has an expected value of μt for single-step mutations (Thomas et al. 1998), and, as follows from the properties of the $(\delta\mu)^2$ distance (Zhivotovsky and Feldman 1995), wt for a general mutation scheme. Therefore, knowing the time of population divergence or the time when the population was founded with an ancestral haplotype, one can estimate w.

For estimation of w from the third data set of 52 worldwide populations, we compare variation at Y chromosome STRs to that at autosomal STRs. For autosomes, the expected variance at a microsatellite locus is $\sim 2Nw$, where $2N$ is the number of chromosomes in the population. Zhivotovsky et al. (2003) estimated the effective mutation rates at autosomal microsatellite loci as 0.71×10^{-3} per 25 years and 0.70×10^{-3} per 25 years for tri- and tetranucleotide repeats, respectively. In each of the 52 populations, we computed (1) the variance in repeat scores for each Y chromosome locus and (2) the average variance over autosomal loci of trinucleotides and of tetranucleotides. Then, for each Y-STR, we computed the ratio of the variance at this locus and the average variance at the corresponding (with tri- or tetranucleotide repeats) autosomal loci. Finally, the effective mutation rate at each Y chromosome locus was computed as four times the product of this ratio and the average effective mutation rates at the corresponding autosomal loci, to take account of the fact that there are four autosomes for every Y chromosome in the population.

Table 3

Distribution of Haplotypes in the Gypsy Populations

	No. of Carriers in											
Haplotype	Rudari	Kalderash	Lom	Kalaidjii (South)	Koshnichar (South- West)	Koshnichari (South-Central)	Kalaidjii (North)	Turgovzi	Darakchii	Feredjelli	Blacksmiths	Musicians
A	62	12	9	11	3	2	9	4	1	4	5	7
B			15				2					
C	1	1										
D	2											
F			1								3	1
G							5				1	2
H	1											
I	1											
J						1						
K				1								
L			1									
M											1	2
N							1					1
O												1
P												4
Q												1
Total	67	13	26	12	3	3	17	4	1	4	10	19
w	.000333	.000343	.000343	.001488	0	.001488	.002101	0	0	0	.002232	.00235

For the purposes of estimation of the time since two populations split from a common ancestor, we use the T_D estimator: $T_D = (D_1–2V_0)/2w$ (Zhivotovsky 2001). Here D_1 is the average squared difference between two alleles sampled from two populations (Goldstein et al. 1995), corrected for bias (Zhivotovsky 2001, p. 708). V_0 is the within-population variance in the number of repeats in the ancestral population prior to its subdivision. The estimator is robust to population dynamics and weak gene flow (Zhivotovsky 2001).

For mutation rates and divergence times at Y chromosome loci, an average value was calculated as the arithmetic mean of estimates across loci, and the SE was formally computed as their SD divided by $\sqrt{k}$, where k is the number of loci (that is, simply the usual error of the mean). Although the SE defined in this way must be biased because of genetic linkage, the bias should not be significant because mutations at different loci are independent; it therefore gives an approximate order of magnitude of statistical error introduced by both genetic sampling and variation in mutation rates.

Results

Estimates of Effective Mutation Rates on the Basis of Genetic Distances

Comparison of the Maoris and Cook Islanders gave an average value (over the seven loci; see table 1; fig. 1) for $(\delta\mu)^2/2$ of 0.00998, which suggests an average effective mutation rate of 0.000312 per 25 years ($25 \times 0.00998/800$). Pairwise comparisons of the 11 Bulgarian Gypsy populations (without the Darakchii sample, in which only one M82 individual was found) gave $(\delta\mu)^2/2$ of 0.01272 (averaged across population pairs and loci) or 0.000454 for the average effective mutation rate. However, these are most probably underestimates, because the $(\delta\mu)^2$ distance assumes constant size for each SNP lineage over time, and it also assumes the same within-lineage variation in an ancestral population prior to its split as at the present generation. It is more likely that each of those populations was founded by a small number of haplotypes and, thus, had lower STR variation prior to divergence; this can lead to an underestimate of the rate of divergence (Zhivotovsky 2001). Therefore, we apply the second estimator, the average squared difference, to the Maori and the Gypsy populations.

The haplotype network shown in table 1 and figure 1 suggests two founder haplotypes for the seven loci in the Maori population, PA and PD (both present in the Cook Islanders), because these haplotypes differ at two loci with no connection by single mutations (see the network of Maori haplotypes in table 1 and fig. 1). By using the ASD_0 estimator for each of the haplotype networks in table 1 and figure 1 and then averaging them with weights proportional to sample sizes, we obtain a mean ± SE effective mutation rate of 0.000705 ± 0.000332, with SD = 0.00078 across loci.

Each of the Gypsy populations contains haplotype A at high frequency (table 3), which suggests that it is the ancestral type. The Lom population is the only one that contains a different haplotype, B, at the highest frequency; therefore, it is likely that both A and B were founder haplotypes in this population. The Musicians are extremely heterogeneous compared with the other populations: of 19 Y chromosomes, 6 carry haplotypes that differ from haplotype A by two alleles. No other Gypsy population displays chromosomes that diverge to this extent from the ancestral haplotype (except for the Lom, in which only 1 of 26 chromosomes differed by two alleles from haplotype A). Moreover, the distribution of chromosomes in the Musicians (with 0, 1, and 2 differences relative to haplotype A) has a mean of 0.237 and a variance of 0.417, thus deviating significantly from a Poisson distribution—that is, $t = [(N-1)/2]^{0.5} \times (s^2/m-1)] =\sim 2.28$, where N is the sample size, and t follows a t distribution, with df = $N-1 = 18$ and a one-tailed P value of .018. This is not the case for the other populations, which suggests that the genetic structure of the Musicians differs from that of other Gypsy groups in Bulgaria. The population of the Musicians could have been founded with multiple haplotypes and/or could have been subject to admixture; therefore, we do not include it in the analysis. After removing the Musicians, we compute w for each population and then weight its values with the sample sizes; this gives $w = 0.000725 \pm 0.000187$ (SD = 0.00053) across loci. For the two sets of comparisons, we use the estimates 0.000705 and 0.000725 in the subsequent analysis.

Estimates of Effective Mutation Rates Based on Comparative Variation

The variances in the number of repeats were computed for each Y chromosome locus in each of 52 worldwide populations. The variances were then converted into estimates of effective mutation rates, as described in the "Material and Methods" section. Averaging over populations gives a w estimate of 0.000638 ± 0.000109 (SD = 0.00029) across loci.

Overall Estimate

On the basis of the arithmetic mean of the above three figures (i.e., 0.000705, 0.000725, and 0.000638), we suggest the following estimate of the effective mutation rate at the average Y chromosome locus: $w = (6.9 \pm 1.3) \times 10^{-4}$ per locus per 25 years (or approximately $[2.8 \pm 0.5] \times 10^{-5}$ per locus per year). The SD across

loci, obtained by averaging the three variances, is 5.7×10^{-4}. We should note that another way of obtaining an overall estimate would be to average the above three figures with some weights; however, it is not clear what kind of weights should be applied in this particular case.

Discussion

Comparison with Autosomal Loci

Our estimate of the average effective mutation rate at Y chromosome STR loci (6.9×10^{-4} per 25 years) is close to those at autosomal microsatellites with tri- and tetranucleotide repeats, 8.5×10^{-4} and 9.3×10^{-4} (Zhivotovsky et al. 2000) and 7.1×10^{-4} and 7.0×10^{-4} (Zhivotovsky et al. 2003), which probably reflect the same slippage machinery that underlies STR mutations. It should be kept in mind that our estimate of effective mutation rate was based on STRs with three- and four-nucleotide motifs; inclusion of loci with dinucleotide repeats may increase this value, because they generally have a higher (effective) mutation rate (Chakraborty et al. 1997; Zhivotovsky et al. 2000).

Dependence of the Estimate on Nongenetic Information

Estimating mutation rates for the SNP/STR data from populations with available archaeological/historical records relies heavily on those records. For example, in the present study, we used 800 years BP as the time of arrival of the Maoris in New Zealand. This may be a lower bound for the time of colonization, and 800–1,000 years BP seems to be an appropriate range for that event (Irwin 1992; Sutton 1994; Diamond and Bellwood 2003); the latter date would give a mutation rate of 0.00056. Therefore, the above mutation rate, 0.000705, which was inferred from the Maori data, might be an overestimate. Other proposed dates include 650–700 years BP (McFadgen et al. 1994), and 1,200 years BP (Bellwood 1989), leading to mutation rate estimates of ~0.00087–0.00081 and ~0.00047, respectively.

The same argument can be applied to the Gypsy data. Historical records suggest that the Gypsies arrived in Bulgaria ~700 years BP. This may be an underestimate, since small groups are not historically "visible" until they become numerous or involved in an important event. If an actual divergence occurred 800 years BP, this would give an effective mutation rate of 0.000634 instead of 0.000725.

An additional problem is uncertainty about founding Y-STR variation at time of arrival in a given geographic region. We assumed, on the basis of our observation of present Y-STR variation, that the Maori founders carried two Y-STR haplotypes (PA and PD [table 1; fig. 1]). Except for the Lom and the Musicians, the founders of Bulgarian Gypsies were assumed to have carried just one Y-STR haplotype (table 2). If the initial Y-STR variation was lower than was assumed, the estimate of effective mutation rate would increase, whereas higher variation would decrease it. However, this uncertainty is almost unavoidable, as is often the case with such data.

The comparative data on worldwide variation at Y chromosome and autosomal STR loci may underestimate the average effective mutation rate. Indeed, the factor four used here assumes that the effective population size of males equals that of females. However, the ratio of variance at Y-STRs to that at autosomal loci varies from population to population. Averaged over regional populations and over Y chromosome loci, the ratio equals 1.14, 0.51, 0.97, 0.93, 1.06, and 0.61, respectively, for sub-Saharan African hunter-gatherers (the number of populations, k, is 3), sub-Saharan African "farmers" ($k = 3$), Eurasia ($k = 21$), East Asia ($k = 18$), Oceania ($k = 2$), and America ($k = 5$). Outlying values for American and sub-Saharan African farming populations, 0.61 and 0.51, may indicate lower effective population size for males than for females in these populations. Removal of these two regions (eight populations) from the analysis gives a point estimate of 6.8×10^{-4}, which is higher than that obtained from the whole data set of 52 populations, 6.38×10^{-4}.

Additional uncertainties in our estimates of mutation rates might be caused by migration. Gene flow can increase within-population variation and thus lead to overestimation of accumulated STR variation. To account for migrants of European origins, which is a major source of admixture for the Bulgarian Gypsy and Polynesian populations, we have applied the ASD_0 method to STR variation within a specific Y haplogroup. In the case of the Gypsies, this is a haplogroup that occurs in the Indian subcontinent but is extremely rare in Europe (Semino et al. 2000; Underhill et al. 2000; Gresham et al. 2001) and is therefore unlikely to have been brought in by admixture. The same approach can be applied to the Maori, Samoan, and Cook Islander populations, because no European males carrying the M208 have been observed so far (Underhill et al. 2001; Kayser et al. 2003). Nevertheless, migrants with the same haplogroup from a different population could not be distinguished if they did not carry any specific labels. In this case, the corresponding mutation rates obtained would be overestimates. Migration would not influence the estimate of w obtained from comparison of within-population variation at Y chromosome STRs with that at autosomal STRs if male and female migration rates were identical; if a migration rate were higher for females than for males, the effective population size would be higher for females, and vice versa. This situation has been discussed in the preceding paragraph.

All this demonstrates that, despite variation in estimates of average Y-STR effective mutation rate (variations due to uncertainties in archaeological/historical data and in male/female population dynamics), these estimates are close to the overall point estimate (6.9×10^{-4} per 25 years) and lie within the interval defined by SE = 1.3×10^{-4}, which is attributable to differences between loci. Doubling the SE, we obtain 4.3×10^{-4} and 9.5×10^{-4} as heuristic confidence limits for w. Potential errors in estimates of w attributable to uncertainties in archaeological/historical records (see the first two paragraphs of the present section) lie within these limits. Therefore, variation among loci in effective mutation rate of various loci may be a major source of deviation of an average estimate of w from a true value for Y chromosome STRs.

Between-Locus Variation in the Effective Mutation Rate

Mutation rates are reported to vary substantially among autosomal microsatellites (Di Rienzo et al. 1998; Zhivotovsky et al. 2001); the same is expected for Y chromosome STRs (Forster et al. 1998; Kayser et al. 2000*b;* Nebel et al. 2001). On the basis of our data, we calculate that the coefficient of between-locus variation in effective mutation rate (100 × 0.00057/0.00069) is >80%; a similar level of between-locus variation in effective mutation rate has been observed for autosomal loci (Zhivotovsky et al. 2001). Although sampling errors contribute to this variation, the differences in w between loci are nevertheless important. Indeed, mutation rates can vary from locus to locus, depending on their structure. For example, DYS389 is a complex locus consisting of four tetranucleotide-repeat subloci (Cooper et al. 1996; Rolf et al. 1998) that yield two distinctive fragments when genotyped using conventional protocols, since the forward primer anneals twice. One fragment contains all four repeat motifs (A, B, C, and D), and the other fragment, which contains just two (C and D), is often denoted by "I". The shorter CD fragment is subtracted from the larger to yield the AB ("II") allele. It is important to note that the C motif is almost always only three repeats and thus is monomorphic, whereas the longer combined AB motifs are both polymorphic, thus making the AB region more mutable than the CD region. The sublocus DYS389AB can be treated as a separate microsatellite locus that has an inherently higher mutation rate than the CD sublocus. This genomic complexity and the consequent differential mutation properties of the subloci are expected to increase the overall mutation rate for DYS389. Removing DYS389 from the analysis gives $w = 0.00061$. Counting that locus twice produces the same value, 0.00061. However, it is difficult to conclude that DYS389AB or other such loci will always behave—in UEP lineages or entire populations—as loci with high mutation rates, and more data will be needed to distinguish loci with different effective mutation rates. Another source of apparent between-locus variation may be different mutation rates for alleles with different numbers of repeats (Brinkmann et al. 1998). This variation actually occurs within a locus and can greatly confound between- and within-locus variation. Probably, our estimate of SD, 5.7×10^{-4}, includes both kinds of variability and therefore encompasses an entire range of "between-allele" variation.

Variation in mutation rates should be kept in mind, because it might be a major source of uncertainty when a small number of loci are used. The large SE of the average mutation rate obtained here and the large SE of divergence time estimates (see below) reflect such variation. (Note that highly variable Y chromosome haplotypes cause very big CIs for coalescent times based on microsatellites [Pritchard et al. 1999]). Therefore, dating historical events on the basis of a small number of Y-STR loci might disagree with historical/archaeological records, although the latter might also have large "SEs." Theoretically, hundreds of loci may be needed for precise dating of ancient demographic events (Zhivotovsky and Feldman 1995; Goldstein et al. 1996; Jorde et al. 1997), and different subsets of loci may give different estimates because of different mutation rates (Zhivotovsky et al. 2003). Analysis of population divergence within UEP lineages should require fewer microsatellite loci for precise dating, because STR variation within a UEP lineage must be smaller than that in the entire UEP-heterogeneous population. The sample of Y chromosome STR loci (no more than 10 were used here) still seems too small, and a larger number of loci need to be analyzed (e.g., Seielstad et al. 2003), and >150 new Y-STRs will be available in the near future (M. Kayser, M. A. Jobling, A. Sajantila, C. Tyler-Smith, unpublished data). Furthermore, we cannot exclude the possibility that mutation rates at the same STR locus vary among haplogroups because of differences in allele repeat scores, repetitive structures, or other factors (Nebel et al. 2001); mutation rates might also be population specific, because of variation in genes that encode proteins involved in DNA replication and repair mechanisms or proteins that cause associated selection, if these exist (see Jobling and Tyler-Smith 2000). A large sample of loci might decrease these possible effects, but, in the absence of hard information, it seems reasonable to use the same overall average mutation rate for all instances.

The estimates of average effective mutation rate (6.9×10^{-4}) and the SD (5.7×10^{-4}) can be used to obtain a two-parameter prior distribution for Y chromosome effective mutation rates for use in coalescent models.

Comparison with Pedigree and Familial Studies

Mutation rates estimated from pedigree and familial studies (Heyer et al. 1997; Bianchi et al. 1998; Kayser et al. 2000*b*) are substantially higher—three times higher—than our "evolutionary" estimate: an average of 18 events in 8,659 meioses, or 0.0021 per generation (exclusion of the entire DYS389 locus reduces this figure to ~0.0017). This figure significantly exceeds our estimate of the effective mutation rate (one-tailed $P = .003$). When we consider the same nine Y-STRs typed in these three cited studies (one tri- and eight tetranucleotide loci), the average mutation rate per locus per generation is 0.0028, on the basis of 8,169 meioses, or 0.0022, on the basis of 7,292 meioses, at just the eight tetranucleotide Y-STRs. Note that an *effective* mutation rate based on pedigree/familial segregations must be even higher because of multistep mutations, as observed by Kayser et al. (2000*b*) in 1 of 14 Y-STR mutation events. The distribution of mutation rates from the cited studies (fig. 2) highlights the pattern of discordance between pedigree/familial and evolutionary estimates: the former has a long tail of loci with high mutation rates.

The discordance may not be limited to the Y chromosome. A 10-fold difference has been found for mtDNA, in which the "fast" mutations are mostly observed in familial studies, whereas "slow" rates are estimated from evolutionary calculations (Heyer et al. 2001). The same argument can be invoked to explain the discordance at Y chromosome STRs. Indeed, we have shown in the "Between-Locus Variation in the Effective Mutation Rate" section that there is variation in mutation rates across loci. Therefore, frequent mutations might be more likely to occur within the few generations used in familial/pedigree studies, whereas the contribution of slowly mutating loci can become significant only within a longer time interval. Association of haplotypes with functional genes, if it exists, can produce differential selection among haplotypes (Jobling and Tyler-Smith 2000) and thus further increase this discordance.

An additional explanation for the discrepancy between the two kinds of mutation rate estimates for microsatellites is that the evolutionary rate estimates are based on statistics of current variation, which has undoubtedly been influenced by reverse mutation of old alleles as well as forward mutation to new alleles. This reverse mutation would actually reduce the standing number of alleles. By contrast, in pedigree-based calculations, mutations are detected on a per-meiosis basis. Thus, what would count as reversals of standing alleles at the population level would be identified as new alleles at the individual level.

Comparison with Data Obtained by Forster and Colleagues

Forster et al. (2000) estimated a mutation rate of 2.6×10^{-4} per 20 years (3.3×10^{-4} per 25 years), which is about half our estimate. Several factors could account for this discrepancy: (1) Forster et al. (2000) estimated the rate of one-step mutations, which must be smaller than the *effective* mutation rate; (2) "fast" locus DYS392 and part of DYS389 were removed from their analysis, whereas we considered all available STRs; and (3) the Native American populations used in the study reported by Forster et al. (2000) were assumed to have diverged 20,000 years BP; a more recent time of divergence would increase their estimate for mutation rate.

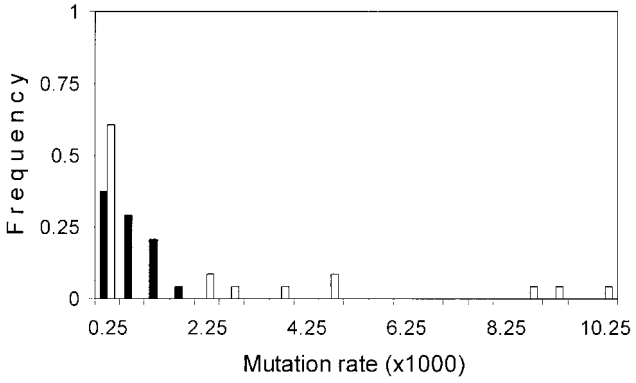

Figure 2 Distribution of mutation rates at Y chromosome STRs. Black columns indicate estimates inferred from the present study; white columns indicate estimates from pedigree and familial studies.

Application

We apply the estimated value of effective mutation rate at an average Y chromosome STR, 6.9×10^{-4} per 25 years, to population data, to see how they correspond to archaeological, linguistic, and historical data.

The Bantu Expansion Time

The Bantu-speaking group is the most populous in sub-Saharan Africa and is probably descended from a "proto-Bantu" population, whose center of origin is between Nigeria and Cameroon (Vansina 1984; Cavalli-Sforza et al. 1994). Two paths have been suggested for the Bantu expansion toward southern Africa. One includes initial movement to East Africa with subsequent migration south (an eastern stream), and the other involves a direct, earlier, western stream from the Nigeria/Cameroon region (Cavalli-Sforza et al. 1994, pp. 163–167).

If we assume that the Bantu populations expanded from a single protopopulation, and therefore average T_D estimates over the 11 populations listed in the "Material and Methods" section, with V_0 computed as the present average within-population variance in repeat scores (among the M191 individuals), we obtain 3,400 ± 1,100 years BP as the Bantu expansion time. Archaeo-

logical and linguistic data support this estimate, because they date the Bantu presence in West Africa to Neolithic times, ~1000 BC (Cavalli-Sforza et al. 1994, pp. 163–167), or even 2000 BC (Vansina 1984; Diamond and Bellwood 2003). Therefore, our estimate supports the view of a pre–Iron Age Bantu expansion and supports the hypothesis of the western stream for this expansion.

The estimate of 3,400 years BP can be viewed as a lower bound for the time of Bantu expansion estimated with genetic data. Indeed, assuming V_0 equal to the present variance means that the within–M191 lineage STR variation at the time of Bantu expansion had already reached the current value. However, if we take the most frequent haplotype among the 148 chromosomes (with the repeat scores at the loci listed in the "Material and Methods" section [i.e., 16, 10,12, 12, 21, 11, 17, 13, 15, and 11, in that order]) as ancestral for the M191 mutation and use $w = 0.00069$, the estimated age of M191 is 14,700 years. This figure (~600 generations) is not a long time to approach an equilibrium, especially if the population (actually, the lineage) was growing in size. Therefore, V_0 must be smaller than the present variation if there were fewer M191 individuals prior to divergence than in contemporary populations. In this case, the Bantu expansion might have occurred even earlier than 3,500 years BP. However, this does not necessarily entail physical movement: earlier genetic divergence of a proto-Bantu population into local subpopulations within the Nigeria/Cameroon area may have preceded actual geographic expansion. Indeed, the early expansion of western proto-Bantus went through many diverse environments (river shores, ocean fringes, forest/savanna mosaics), with subsequent specialization in such activities as river and deep sea fishing and farming (Vansina 1984). The variety of environments could have fostered isolation of local groups, followed by divergence of dialects into languages and by genetic differentiation.

Relationship between the Polynesian Populations

We performed the T_D analysis on the Maori, Cook Island, and Samoan samples. The Samoans differ greatly from the Maori and from Cook Islanders (table 1; fig. 1), mainly because of the DYS389-CD locus: three of four individuals carry haplotypes with 13 repeats, and there is 1 haplotype with 14 repeats, whereas 17 of 19 Maori haplotypes carry 12 repeats, and the two remaining haplotypes have 13 repeats at this locus. Also, there is quite a large difference between them at the DYS439 locus. Although four individuals do not constitute a representative sample, it is unlikely that this sample would carry alleles with 13 and 14 repeats only at the DYS389-CD locus if they occurred at low frequency throughout the Samoan population.

The origin and time of arrival of the Samoans are in question. The tree in figure 3 shows that the Samoans split much earlier than the ancestral population that gave rise to the Maoris and Cook Islanders. Recall that the Maoris arrived in New Zealand 800–1,000 years BP; together with the time estimate between the two separation events in figure 3, this implies ~3,500 years BP for divergence of the Samoans from a common ancestral population. This can be compared with the time of East Polynesian settlement, estimated to have occurred by 500–1000 BC (Irwin 1992, p. 81), and to the estimate of the early peopling of Polynesia, 3,000–4,000 years BP (Underhill et al. 2001*a*).

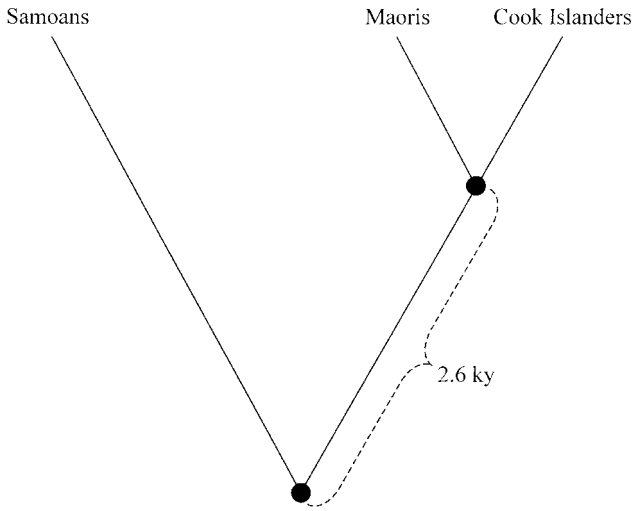

Figure 3 Genetic differentiation among the Polynesian populations. The time between the nodes was computed as $\Delta T_D = (D'_1 - D''_1)/2w$ (Zhivotovsky 2001), where D'_1 is the average value of D_1 for population pairs (Samoans/Maoris and Samoans/Cook Islanders), and D''_1 is for the Maoris/Cook Islanders comparison.

Divergence between the Gypsy Populations

Compution of T_D values averaged over all possible pairs of the 10 Gypsy populations (omitting both the Darakchii, with one M82 individual only, and the Musicians) and use of $V_0 = 0$, gives us an estimate of the time of founding of an ancestral population of related males sharing the same Y haplotype that gave rise to the contemporary Bulgarian Gypsy populations—namely, 1,500 ± 700 years ago (mean ± SE). This estimate gives an upper bound for the divergence time and is compatible with the formation of the proto-Gypsies in India, predating their entry into the Byzantine Empire 900–1,000 years BP (Fraser 1992). If diversity was already substantial in the founder male population, the value of V_0 would be positive, and, thus, the estimate of divergence time would be smaller.

The genetic composition of the Musicians differs greatly from that of the other studied Bulgarian Gypsy populations, a fact that points to possible differences in

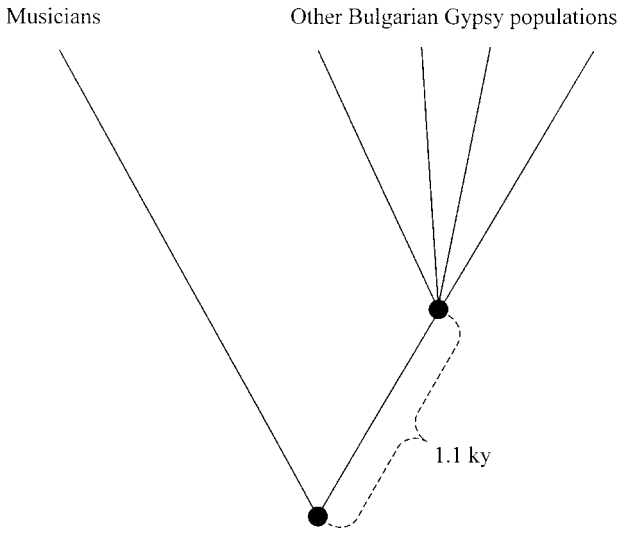

Figure 4 Genetic differentiation among the Bulgarian Gypsy populations. The time between the nodes was computed as for that in figure 3.

their evolutionary history. There are two possible explanations for this: the Musicians share the same origin but were greatly admixed with populations from South Asia that carried the M82 mutation, or they descended from an ancestral population different from that of the other Bulgarian Gypsy populations studied. The origins of the proto-Gypsies, as well as the time and number of migrations out of India, are still disputed among cultural anthropologists and linguists (Fraser 1992; Marushiakova and Popov 1997; Hancock 2000). Our previous study (Gresham et al. 2001) suggested a common origin from a small group of ancestors. One should note, however, that the Musicians were not included in that study and that they are the sole representatives of a particular Balkan dialect of the Romanes language. In addition to the unusual distribution of M82 haplotypes, they display a generally higher diversity of Y chromosome lineages, including other uncommon types, that are unlikely to result from European admixture. If we follow the "different origins" scenario, the T_D estimator gives an upper bound of 2,600 years BP for the separation of the Musicians from a population ancestral to the other studied populations of Bulgarian Gypsies. The difference, ΔT_D, of 1,100 years between the two splits (fig. 4) allows not only for heterogeneous origins but also for the possibility of different proto-Gypsy migrations from the Indian subcontinent.

Acknowledgments

We thank Dr. David Modiano (University "La Sapienza," Rome), for providing the Mossi samples; Dr. Antonel Olckers (Potchefstrum University for Christian Higher Education, Pretoria, South Africa), for the !Kung samples; and Dr. Erika Hagelberg (University of Oslo), for providing the Cook samples. We thank two anonymous reviewers for their helpful comments. Research was supported, in part, by National Institutes of Health grants R03 TW005540, GM 28016, and GM 28428 and by the Australian Research Council, the Wellcome Trust, the Max Planck Society (Germany), and Russian Foundation for Basic Research grant 01-04-48441. We thank Ms. Lisa Diamond for technical assistance in preparing this manuscript.

References

Ayub Q, Mohyuddin A, Qamar R, Mazhar K, Zerjal T, Mehdi SQ, Tyler-Smith C (2000) Identification and characterisation of novel human Y-chromosomal microsatellites from sequence database information. Nucleic Acids Res 28:e8

Bellwood PS (1989) The colonization of the Pacific: some current hypotheses. In: AVS Hill, SW Serjeantson (eds) The colonization of the Pacific: a genetic trial. Clarendon Press, Oxford, pp 1–90

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

Bosch E, Lee AC, Calafell F, Arroyo E, Henneman P, de Knijff P, Jobling MA (2002) High resolution Y chromosome typing: 19 STRs amplified in three multiplex reactions. Forensic Sci Int 125:42–51

Brinkmann B, Klintchar M, Neuhuber F, Hühne J, Rolf B (1998) Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. Am J Hum Genet 62:1408–1415

Cann HM, de Toma C, Cazes L, Legrand MF, Morel V, Piouffre L, Bodmer J, et al (2002) A human genome diversity cell line panel. Science 296:261–262

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

Chakraborty R, Kimmel M, Stivers DN, Davidson J, Deka R (1997) Relative mutation rate at di-, tri-, and tetranucleotide microsatellite loci. Proc Natl Acad Sci USA 94:1041–1046

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

Cruciani F, Santolamazza P, Shen P, Macaulay V, Moral P, Olckers A, Modiano D, Holmes S, Destro-Bisol G, Coia V, Wallace DC, Oefner PJ, Torroni A, Cavalli-Sforza LL, Scozzari R, Underhill PA (2002) A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes. Am J Hum Genet 70: 1197–1214

de Knijff P (2000) Messages through bottlenecks: on the combined use of slow and fast evolving polymorphic markers on the human Y chromosome. Am J Hum Genet 67:1055–1061

Diamond J, Bellwood P (2003) Farmers and their languages: the first expansions. Science 300:597–603

Di Rienzo A, Donnelly P, Toomajian CH, Sisk B, Hill A, Petzl-Erler ML, Haines GK, Barch DH (1998) Heterogeneity of microsatellite mutations within and between loci, and im-

plications for human demographic histories. Genetics 148: 1269–1281

Forster P, Kayser M, Meyer E, Roewer L, Pfeiffer H, Benkmann H, Brinkmann B (1998) Phylogenetic resolution of complex mutational features at Y-STR DYS390 in aboriginal Australians and Papuans. Mol Biol Evol 15:1108–1114

Forster P, Röhl A, Lünnemann P, Brinkmann C, Zerijal T, Tyler-Smith CH, Brinkmann B (2000) A short tandem repeat–based phylogeny for the human Y chromosome. Am J Hum Genet 67:182–196

Fraser AM (1992) The Gypsies. Blackwell, Oxford

Goldstein DB, Linares AR, Cavalli-Sforza LL, Feldman MW (1995) An evaluation of genetic distances for use with microsatellite loci. Genetics 139:463–471

Goldstein DB, Zhivotovsky LA, Nayar K, Linares AR, Cavalli-Sforza LL, Feldman MW (1996) Statistical properties of the variation at linked microsatellite loci: implications for the history of human Y chromosomes. Mol Biol Evol 13:1213–1218

Gresham D, Morar B, Underhill PA, Passarino G, Lin AA, Wise Ch, Angelicheva D, Calafell F, Oefner PJ, Shen P, Tournev I, de Pablo R, Kuinskas V, Perez-Lezaun A, Marushiakova E, Popov V, Kalaydjieva L (2001) Origins and divergence of the Roma (Gypsies). Am J Hum Genet 69:1314–1331

Hancock I (2000) The emergence of Romani as a koine outside of India. In: Acton T (ed) Scholarship and Gypsy struggle: commitment in Romani studies (essays in honour of Donald Kenrick on the occasion of his seventieth birthday). University of Hertfordshire Press, Hatfield, UK

Hurles ME, Irven C, Nicholson J, Taylor PG, Santos FR, Loughlin J, Jobling MA, Sykes BC (1998) European Y-chromosomal lineages in Polynesians: a contrast to the population structure revealed by mtDNA. Am J Hum Genet 63: 1793–1806

Jobling MA, Tyler-Smith C (2000) New uses for new haplotypes: the human Y chromosome, disease and selection. Trend Genet 6:356–362

Jorde LB, Rogers AR, Bamshad M, Watkins WS, Krakowiak P, Sung S, Kere J, Harpending H (1997) Microsatellite diversity and the demographic history of modern humans. Proc Natl Acad Sci USA 94:3100–3103

Hammer MF, Karafet TM, Redd AJ, Jarjanazi H, Santachiara-Benerecetti S, Soodyall H, Zegura SL (2001) Hierarchical patterns of global human Y-chromosome diversity. Mol Biol Evol 18:1189–1203

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

Heyer E, Zietkiewicz E, Rochowski A, Yotova V, Puymirat J, Labuda D (2001) Phylogenetic and familial estimates of mitochondrial substitution rates: study of control region mutations in deep-rooting pedigrees. Am J Hum Genet 69: 1113–1126

Holtkemper U, Rolf B, Hohoff C, Forster P, Brinkmann B (2001) Mutation rates at two human Y-chromosomal microsatellite loci using small pool PCR techniques. Hum Mol Genet 10:629-633

Irwin G (1992) The prehistoric exploration and colonization of the Pacific. Cambridge University Press, Cambridge

Kayser M, Brauer S, Weiss G, Schiefenhövel W, Underhill P, Shen P, Oefner P, Tommaseo-Ponzetta M, Stoneking M (2003) Reduced Y-chromosome, but not mtDNA, diversity in human populations from West New Guinea. Am J Hum Genet 72:281–302

Kayser M, Brauer S, Weiss G, Underhill PA, Roewer L, Schiefenhovel W, Stoneking M (2000*a*) Melanesian origin of Polynesian Y chromosomes. Curr Biol 10:1237–1246

Kayser M, Caglia 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

Kayser M, Roewer L, Hedman M, Henke L, Henke J, Brauer S, Krüger K, Krawczak M, Nagy M, Dobosz T, Szibor R, de Knijff P, Sajantila A (2000*b*) Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. Am J Hum Genet 66:1580–1588

Marushiakova E, Popov V (1997) Gypsies (Roma) in Bulgaria. Studium zur Tsiganologie und Folkloristik (band 18). Peter Lang, Frankfurt

McFadgen BG, Knox FB, Cole TRL (1994) Radiocarbon calibration curve variations and their implications for the interpretation of New Zealand prehistory. Radiocarbon 36: 221–236

Mountain JL, Knight A, Jobin M, Gignoux C, Miller A, Lin AA, Underhill PA (2002) SNPSTRs: empirically derived, rapidly typed, autosomal haplotypes for inference of population history and mutational processes. Genome Res 12:1766–1722

Nebel A, Filon D, Hohoff C, Faerman M, Brinkmann B, Oppenheim A (2001) Haplogroup-specific deviation from the stepwise mutation model at the microsatellite loci DYS388 and DYS392. Eur J Hum Genet 9:22–26

Oefner PJ, Underhill PA (1998) DNA mutation detection using denaturing high performance liquid chromatography (DHPLC). Current protocols in human genetics, suppl 19. Wiley & Sons, New York, 7.10.1–7.10.12

Pritchard JK, Seielstad MT, Pérez-Lezaun A, Feldman MW (1999) Population growth of human Y chromosomes: a study of Y chromosome microsatellites. Mol Biol Evol 16: 1791–1798

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

Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW (2002) Genetic structure of human populations. Science 298:2381–2385

Seielstad M, Yuldasheva N, Singh N, Underhill P, Oefner P, Shen P, Wells RS (2003) A novel Y chromosome variant puts an upper limit on the timing of first entry into the Americas. Am J Hum Genet 73:700–705

Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, De Benedictis G, Francalacci P, Kouvatsi A, Limborska S, Marcikiae M, Mika A, Mika B, Primorac D, Santachiara-Benerecetti AS, Cavalli-Sforza LL, Underhill PA (2000) The genetic legacy of Paleolithic *Homo sapiens sap-*

iens in extant Europeans: a Y chromosome perspective. Science 290:1155–1159

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457–462

Su B, Jin L, Underhill PA, Martinson J, Saha N, McGarvey ST, Shriver MD, Chu J, Oefner P, Chakraborty R, Deka R (2000) Polynesian origins: insights from the Y chromosome. Proc Natl Acad Sci USA 97:8225–8228

Sutton DG (1994) Conclusion: origins. In: Sutton DG (ed) The origins of the first New Zealanders. Auckland University Press, Auckland, pp 243–258

Thomas MG, Skorecki K, Ben-Ami H, Parfitt T, Bradman N, Goldstein DB (1998) Origins of Old Testament priests. Nature 394:138–140

Underhill PA, Jin L, Zemans 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

Underhill PA, Passarino G, Lin AA, Marzuki S, Oefner PJ, Cavalli-Sforza LL, Chambers GK (2001*a*) Maori origins, Y chromosome haplotypes and implications for human history in the Pacific. Hum Mutat 17:271–280

Underhill PA, Passarino G, Lin AA, Shen P, Lahr MM, Folewy RA, Oefner PJ, Cavalli-Sforza LL (2001*b*) The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. Ann Hum Genet 65:43–62

Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonne-Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, Kidd JR, Mehdi SQ, Seielstad MT, Wells RS, Piazza A, Davis RW, Feldman MW, Cavalli-Sforza LL, Oefner PJ (2000) Y chromosome sequence variation and the history of human populations. Nat Genet 26:358–361

Vansina J (1984) Western Bantu expansion. J Afr Hist 25:129–145

Weber JL, Broman KW (2001) Genotyping for human whole-genome scans: past, present, and future. Adv Genet 42:77–96

White PS, Tatum OL, Deaven LL, Longmire JL (1999) New male-specific microsatellite markers from the human Y chromosome. Genomics 57:433–437

Y Chromosome Consortium (2002) A nomenclature system for the tree of human Y-chromosomal binary haplogroups. Genome Res 12:339–348

Zerjal T, Dashnyam B, Pandya A, Kayser M, Roewer L, Santos F, Schiefenhövel W, Fretwell N, Jobling MA, Harihara S, Shimizu K, Semjidmaa D, Sajantila A, Salo P, Crawford MH, Evgrafov O, Tyler-Smith C (1997) Genetic relationships of Asians and northern Europeans revealed by Y-chromosomal DNA analysis. Am J Hum Genet 60:1174–1183

Zhivotovsky LA (2001) Estimating divergence time with the use of microsatellite genetic distances: impacts of population growth and gene flow. Mol Biol Evol 18:700–709

Zhivotovsky LA, Feldman MW (1995) Microsatellite variability and genetic distances. Proc Natl Acad Sci USA 92: 11549–11552

Zhivotovsky LA, Goldstein DB, Feldman MW (2001) Genetic sampling error of distance $(\delta\mu)^2$ and variation in mutation rate among microsatellite loci. Mol Biol Evol 18:2141–2145

Zhivotovsky LA, Rosenberg NA, Feldman MW (2003) Features of evolution and expansion of modern humans inferred from genomewide microsatellite markers. Am J Hum Genet 72:1171–1186