Am. J. Hum. Genet. 66:979–988, 2000

The Distribution of Human Genetic Diversity: A Comparison of Mitochondrial, Autosomal, and Y-Chromosome Data

L. B. Jorde,[1] W. S. Watkins,[1] M. J. Bamshad,[2] M. E. Dixon,[1] C. E. Ricker,[1] M. T. Seielstad,[3] and M. A. Batzer[4]

Departments of [1]Human Genetics and [2]Pediatrics, University of Utah Health Sciences Center, Salt Lake City; [3]Program for Population Genetics, Harvard School of Public Health, Boston; and [4]Departments of Pathology, Biometry and Genetics, Biochemistry and Molecular Biology, Stanley S. Scott Cancer Center, Louisiana State University Health Sciences Center, New Orleans

We report a comparison of worldwide genetic variation among 255 individuals by using autosomal, mitochondrial, and Y-chromosome polymorphisms. Variation is assessed by use of 30 autosomal restriction-site polymorphisms (RSPs), 60 autosomal short-tandem-repeat polymorphisms (STRPs), 13 *Alu*-insertion polymorphisms and one *LINE-1* element, 611 bp of mitochondrial control-region sequence, and 10 Y-chromosome polymorphisms. Analysis of these data reveals substantial congruity among this diverse array of genetic systems. With the exception of the autosomal RSPs, in which an ascertainment bias exists, all systems show greater gene diversity in Africans than in either Europeans or Asians. Africans also have the largest total number of alleles, as well as the largest number of unique alleles, for most systems. G_{ST} values are 11%–18% for the autosomal systems and are two to three times higher for the mtDNA sequence and Y-chromosome RSPs. This difference is expected because of the lower effective population size of mtDNA and Y chromosomes. A lower value is seen for Y-chromosome STRs, reflecting a relative lack of continental population structure, as a result of rapid mutation and genetic drift. Africa has higher G_{ST} values than does either Europe or Asia for all systems except the Y-chromosome STRs and *Alu*s. All systems except the Y-chromosome STRs show less variation between populations within continents than between continents. These results are reassuring in their consistency and offer broad support for an African origin of modern human populations.

Introduction

The distribution of human genetic diversity has long been a subject of interest, and it has important implications for human evolution, forensics, and the distribution of genetic diseases in populations. Genetic diversity in human populations is low relative to that in many other species, attesting to the recent origin and small size of the ancestral human population (Li and Sadler 1991; Crouau-Roy et al. 1996; Kaessmann et al. 1999*b*). The proportion of diversity that exists between human populations is also relatively low. An early study, based on protein polymorphisms, arrived at a between-groups diversity estimate of 15% (Lewontin 1972). Other studies, based on protein polymorphisms as well as on blood groups and craniometrics, have yielded similar results (Nei and Livshits 1990; Relethford and Harpending 1994). Recently, surveys of mitochondrial (Merriwether et al. 1991), Y-chromosome (Hammer et al. 1997), and various types of autosomal polymorphisms (Bowcock et al. 1991; Batzer et al. 1994; Deka et al. 1995*a,* 1999; Jorde et al. 1995; Watkins et al. 1995; Barbujani et al. 1997; Stoneking et al. 1997) have all shown that most human genetic diversity is found within, rather than between, populations.

Although most assessments of genetic diversity have been based on a single type of genetic system, some of the most informative diversity studies have involved the comparison of estimates based on different types of systems. Such comparisons have led to important conclusions about human origins and about sex-specific differences in population size and gene flow (Jorde et al. 1995, 1998; Sajantila et al. 1996; Spurdle and Jenkins 1996; Merriwether et al. 1997; Poloni et al. 1997; Scozzari et al. 1997; Bamshad et al. 1998; Lum et al. 1998; Passarino et al. 1998; Seielstad et al. 1998; Kittles et al. 1999; Perez-Lezaun et al. 1999). A drawback of many of these studies, however, is that they are based on limited types of genetic systems, and they often do not compare the same polymorphisms in the same populations. Here, we present the first published comparison of within- and between-population genetic diversity in autosomal, mtDNA, and Y-chromosome loci in the same set of individuals. Variation is assessed in autosomal short-tandem-repeat polymorphisms (STRPs),

Received October 20, 1999; accepted for publication December 6, 1999; electronically published February 25, 2000.

Address for correspondence and reprints: Dr. Lynn B. Jorde, Department of Human Genetics, University of Utah Health Sciences Center, Salt lake City, UT 84112. E-mail: lbj@genetics.utah.edu

0002-9297/2000/6603-0021$02.00

autosomal restriction-site polymorphisms (RSPs), autosomal *Alu* polymorphisms, mtDNA control-region sequences (hypervariable sequences 1 and 2), and Y-chromosome polymorphisms. Substantial congruity is seen among various types of systems for most comparisons, but illuminating differences are also seen.

Methods

The study population, which includes 72 Africans, 63 Asians, and 120 Europeans, has been described elsewhere (Jorde et al. 1995, 1997). These three continental groups were further subdivided into six African populations (Biaka Pygmy, Mbuti Pygmy, Nguni, San, Sotho/Tswana, and Tsonga), five Asian populations (Cambodian, Chinese, Japanese, Malay, and Vietnamese), and four European populations (Finnish, French, Polish, and northern European). Informed consent was obtained from all subjects whose blood was drawn at the University of Utah.

The 60 autosomal STRP and 30 RSPs, as well as 200 bp of mitochondrial hypervariable sequence 2 (HVS2) data, were obtained by methods described elsewhere (Jorde et al. 1995, 1997). HVS1 sequences (411 bp) were PCR amplified in a 1.1- or 0.45-kb fragment in 1 × PCR buffer (10 mM Tris pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$) by use of 20 ng of template, 200 μM dNTPs, 50 pmol of each primer (UPL15996, RPH408, and H16401 [Bamshad et al. 1998]), and 1 U of *Taq* DNA polymerase, in a total reaction volume of 50 μl. Samples were cycled by use of a standard three-step PCR profile, with the annealing temperature for the first five cycles set to 58°C and then lowered to 54°C for an additional 25 cycles. Sequence for HVS1 was generated from the UPL15996 and H16401 primer sites by use of ABI Dye-primer or dRhodamine sequencing reagents and an ABI 377 automated DNA sequencer. Sequence data were compared and edited by use of the SEQUENCHER software package (Genecodes).

Multiplex genotyping of six Y-chromosome STRPs (Y STRPs)—*DYS19, DYS288, DYS388, DYS389, DYS390,* and *DYS393*—was done by use of an ABI 377 automated DNA sequencer. The PCR reaction contained one fluorescent end-labeled primer for each locus. DNA samples were amplified by PCR in 1 × buffer (10 mM Tris pH 8.3, 50 mM KCl) by use of 25 ng of genomic template DNA reaction product, 50 μM each dNTP, and 0.2 U of *Taq* DNA polymerase complexed with *Taq*Start antibody (Clontech). Primer concentrations were optimized for each multiplex PCR panel. Thermal cycling was done in a Perkin-Elmer 9600 PCR machine by use of a modified touchdown protocol in which the first five cycles are done with annealing temperatures 2°C above the predicted average melting temperature (T_m), followed by 25 additional cycles with annealing temperatures 4°–6°C lower. The Y-chromosome genotyping used four-color fluorescent-dye chemistry; six STRPs were multiplexed in two PCR reactions and were run in a single ABI lane. Products were resolved with urea denaturing polyacrylamide gels on the ABI sequencer by use of internal size standard in each lane. Raw genotype data were collected by use of GENESCAN software (ABI), and gel files were analyzed by use of GENOTYPER software package (ABI). Y-chromosome polymorphisms for the southern-African populations were obtained from the Y Chromosome Microsatellites Web site and are more fully described by Seielstad et al. (1999). In addition to the Y STRPs, the Y-specific *Alu* insertion (YAP, *DYS287*) and three Y-specific RSPs (*DYS257, DYZ3,* and *SRY10831*) were assayed manually by use of PCR amplification, electrophoresis on 2.5% agarose gels, and ethidium bromide staining for visualization.

Thirteen *Alu* insertion polymorphisms (HS2.43, HS4.14, HS4.65, HS4.75, Sb19.3, Sb19.12, APO, B65, COL3A1, D1, PV92, TPA25, and HS4.32) (Arcot et al. 1996, 1997, 1998; Milewicz et al. 1996; Stoneking et al. 1997) and one *LINE-1* element (*DV1.9*) were genotyped manually. Regions containing polymorphic *Alu* inserts were amplified by PCR; 25 ng of genomic DNA was amplified with locus-specific flanking primers in 1 × PCR buffer (10 mM Tris pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$) by use of 200 μM each dNTP, 10 pmol each of both flanking primers, and 1 U of *Taq* DNA polymerase, in a total reaction volume of 25 μl. Samples were cycled in a standard three-step PCR profile with the annealing temperatures standardized for the 58°C/54°C PCR protocol described above for most primer sets. To reduce nonspecific amplification products for systems HS2.43, HS4.65, Sb19.12, B65, and PV92, dimethyl sulfoxide was added to a final concentration of 10%. In addition, the cycling temperatures were increased 4°C for COL3A1. PCR products were resolved on 3% Nusieve agarose gels in 0.5 × Tris-borate EDTA and were visualized by ethidium bromide staining.

For mtDNA sequence data, gene diversity was estimated for each population as $n/(n-1)\Sigma x_i x_j d_{ij}$, where n is the number of DNA sequences examined, x_i and x_j are the population frequencies of the ith and jth type of DNA sequences, respectively, and d_{ij} is the proportion of nucleotides that differ between the ith and jth types of DNA sequence. For the Y-chromosome and autosomal systems, gene diversity was estimated as $n/(n-1)(1-\Sigma x_i^2)$, where x_i is the estimated frequency of the ith allele in the system. For diploid loci, this approach provides an estimate of the heterozygosity level expected under random mating.

To take into account the information provided by multiple alleles, allele size variance was estimated for the Y-chromosome and autosomal STRPs. This ap-

Table 1

Gene-Diversity Estimates for Continental Populations

	Gene-Diversity Estimate (Average Variance Ratio)					
Continent	STRP	RSP	*Alu*	HVS1	HVS2	Y STRP
Africa	.679 (1.13)	.293	.276	.022	.030	.576 (1.05)
Asia	.638 (.92)	.350	.233	.015	.011	.472 (1.00)
Europe	.675 (.94)	.401	.243	.009	.010	.498 (.95)

proach assumes a stepwise mutation model for STRPs, which appears to be consistent with the distribution of allele sizes (Di Rienzo et al. 1994; Shriver et al. 1995). Population comparisons of average variances across loci can be misleading, however, because the average variance tends to be dominated by a small number of highly variable, rapidly mutating STRPs. To control for this effect, the average variance of each locus across the three continental populations was estimated, and then a variance ratio was constructed by dividing each individual continental variance by the average variance across continental populations. This variance ratio was then averaged across all loci for each of the three continental populations (Jorde et al. 1997).

Gene-diversity levels within and between populations were used to estimate the proportion of genetic variance due to subdivision, termed "F_{ST}" or "G_{ST}" (Wright 1965). The grouping of populations into major continents (Africa, Asia, and Europe) allowed the analysis of variation (analysis of molecular variance, or AMOVA) at three levels: within individual populations, between populations within continents, and between continents. All calculations, including random-permutation procedures to assess statistical significance, were performed by use of the ARLEQUIN package (Excoffier et al. 1992; Schneider et al. 1997).

Results

Gene-diversity levels for each continental population are given in table 1. The most notable pattern is that Africans have the highest level of diversity for all types of systems except the nuclear RSPs, for which Europeans have the highest level of diversity. The latter finding is explained by a pronounced ascertainment bias because RSPs were originally identified as a result of heterozygosity in European subjects (Mountain and Cavalli-Sforza 1994). STRPs, because of higher levels of polymorphism, are less subject to this bias (Rogers and Jorde 1996) and thus reveal higher levels of African diversity. This trend is more evident when allele-size-variance ratios are examined. In this case, the autosomal STRPs exhibit variance that is ~20% higher in African populations than in non-African populations. A slightly smaller difference (10%) is seen when variance ratios are estimated for the Y STRPs. Another trend seen in table 1 is that African mtDNA diversity is two to three times higher than that of non-African populations, whereas, for the nuclear *Alu* and STR polymorphisms, African diversity is only 5%–30% higher than non-African diversity.

Table 2 provides a breakdown of gene-diversity estimates for each of the 15 individual populations. Again, the individual African populations tend to have the highest levels of diversity for most systems. For the Y STRPs, two European populations (northern Europeans and Finns) have strikingly low levels of diversity, especially in comparison with those in the other European populations. This contrasts with their diversity estimates for other systems, in which they tend to be quite similar to those in other European populations.

Table 3 summarizes gene diversity in terms of both the number of alleles per continental population and the number of alleles unique to each population (i.e., present in one population but not in the other two). This tabulation was not done for the autosomal RSP and *Alu* systems because nearly all loci were polymorphic in all three continental populations. A total of 635 different autosomal STRP alleles are seen in the 60 mi-

Table 2

Gene-Diversity Estimates for Individual Populations

	Gene-Diversity Estimate					
Population	STRP	RSP	*Alu*	HVS1	HVS2	Y STRP
Biaka	.714	.319	.281	.019	.030	...[a]
Mbuti	.739	.195	.298	.021	.027	.564
Nguni	.692	.311	.315	.022	.030	.501
San	.648	.240	.322	.018	.014	.594
Sotho/Tswana	.697	.310	.308	.025	.035	.581
Tsonga	.690	.307	.316	.017	.030	...[a]
Cambodian	.669	.395	.252	.013	.010	.413
Chinese	.660	.365	.232	.015	.013	.552
Japanese	.629	.313	.147	.014	.007	.434
Malay	.625	.398	.299	.017	.010	.560
Vietnamese	.664	.337	.222	.016	.010	.567
Finnish	.660	.391	.227	.010	.014	.312
French	.649	.406	.275	.006	.009	.494
Northern European	.684	.400	.244	.009	.010	.343
Polish	.685	.415	.234	.006	.010	.700

[a] Because an insufficient number of male subjects were available, variation could not be estimated reliably.

Table 3

Total Number of Alleles and Number of Unique Alleles, per Continental Population

Population	Total No. of Alleles (No. of Unique Alleles)			
	STRP	HVS1	HVS2	Y STRP
Africa	544 (62)	62 (28)	25 (10)	36 (8)
Asia	474 (16)	73 (28)	17 (4)	28 (2)
Europe	526 (34)	67 (27)	24 (7)	32 (6)
Overall	635	128	44	47

crosatellite loci, and a total of 47 Y-STRP alleles are seen in 6 loci. Africans have the largest number of alleles, as well as the largest number of unique alleles, for autosomal and Y STRPs. The same holds true for the HVS2 sequence data, in which an allele is defined as a variant nucleotide at a given position along the sequence.

Although HVS1 gene diversity is nearly twice as high in Africans as in non-Africans (table 1), Africans have a slightly lower number of HVS1 alleles, and the number of unique HVS1 alleles is nearly identical in the three populations. This apparent discrepancy is explained by the pattern seen in figure 1, in which the frequencies of the minor alleles (i.e., deviations from the most common nucleotide) are plotted for each population. Figure 1 shows that approximately half of the minor alleles in Asians and Europeans occur only once in the population. In contrast, fewer than one third of the African minor alleles occur only once, in spite of the fact that the African sample size is substantially smaller than the European sample size. Europeans and Asians also have an excess of minor alleles that are seen in only two or three copies. Africans have more minor alleles seen in greater copy number, especially in the ">10" category. Each of these relatively common minor alleles is found in most or all of the six African populations; they are thus likely to represent relatively old mutations. The presence of many minor alleles of low frequency in non-Africans is consistent with the appearance of new mutations in these populations after their exit from Africa and subsequent population expansion. Although a similar pattern of minor-allele frequencies is seen for the HVS2 sequence (fig. 2), Africans have a larger number of HVS2 alleles than do non-Africans. This difference between the HVS1 and HVS2 allele counts can be explained by the fact that the mutation rate for HVS1 sequence is, on average, twice as high as that for HVS2 sequence (Meyer et al. 1999). One would thus expect that, in non-African populations, the accumulation of new mtDNA mutations would be more rapid for HVS1 than for HVS2.

G_{ST} values estimated for the worldwide sample by use of the three continents as subdivisions are given in table 4. All three autosomal systems (STRPs, RSPs, and *Alu*s) yield overall G_{ST} estimates of 11%–18%, in accord with the results of previous studies. The G_{ST} estimates for the two mtDNA sequences are substantially higher, reflecting the fact that the effective size for mtDNA is only one-fourth that for nuclear DNA. Consequently, differentiation due to random drift tends to proceed more rapidly for mtDNA. The G_{ST} for Y STRPs is surprisingly low and may reflect some degree of convergence in allele sizes for these rapidly mutating systems when major continental populations are compared. When the 15 individual populations, rather than the three continental populations, are used as units of subdivision, the Y-chromosome estimate increases to 18%.

Table 4 also shows G_{ST} estimates for each continent separately when the individual populations are used as subdivisions. For the mtDNA sequences, African populations are by far the most highly differentiated populations, again reflecting high rates of drift for mtDNA. Africans also have the highest G_{ST} values for the autosomal RSP and STRP systems but not for the *Alu*s and the Y STRPs. For Y STRPs, Europeans have by far the greatest level of differentiation. A genetic-distance analysis (data not shown) indicates that this reflects extreme divergence of the Finnish and northern-European populations, which also have the lowest levels of Y-chromosome diversity.

A hierarchical analysis of genetic variation (i.e., AMOVA) is presented in table 5. Consistent with the G_{ST} results presented in table 4, the results of this analysis show that the great majority of genetic variation occurs within populations. Notably, for all systems except the Y STRPs, the differentiation of individual populations within continents is several times lower than the differentiation between continental populations. For the Y STRPs, there is a very high level of differentiation between populations within continents (especially within Europe). Because the estimate for variation between continents is obtained by subtraction, it becomes slightly negative for these systems. When the divergent Finnish and northern-European populations are omitted from this analysis, the percentage of variation between

Table 4

G_{ST} Values, by Continent and for All Populations

Continent	G_{ST} Value for					
	STRP	RSP	*Alu*	HVS1	HVS2	Y STRP
Africa	.024	.027	.017	.088	.092	.026
Asia	.007	.017	.022	.032	.017	.092
Europe	.023	.013	.009	.045	.013	.602
Overall:						
Three continents	.109	.142	.179	.237	.267	.044
Fifteen populations	.097	.113	.151	.233	.261	.178

Note.—All G_{ST} values differ significantly ($P < .0001$) from 0.

Table 5

Hierarchical AMOVA Analysis, Showing the Percentage of Variation at Each of Three Levels of Population Hierarchy

Comparison	Variation[a] (%)					
	STRP	RSP	*Alu*	HVS1	HVS2	Y STRP
Within populations	87.9	85.5	80.9	72.0	68.9	83.3
Between populations within continents	1.7	1.3	1.8	6.0	6.2	18.5[b]
Between continents	10.4	13.2	17.4	22.0	24.9	−1.8[b]

[a] All values, except for the negative value for Y STRPs, are significantly ($P < .0001$) >0.

[b] When the highly divergent Finnish and northern-European populations are omitted, the percentage of variation becomes 5.1%. between populations within continents and 7.8% between continents.

populations within continents becomes much smaller (5.1%), and the percentage of variation between continents becomes positive (7.8%). The overall G_{ST} estimate for Y STRPs increases to 13%. Thus, a possible male-founder effect in these two populations accounts for much of the discrepancy between the Y STRPs and other systems.

Discussion

The gene-diversity results presented here are consistent with one another and with those of many previous studies in showing higher levels of diversity in African populations than in non-African populations (Vigilant et al. 1991; Nei et al. 1993; Bowcock et al. 1994; Deka et al. 1995*b;* Jorde et al. 1997). A study of eight *Alu* polymorphisms (Stoneking et al. 1997) showed that western Asia has a slightly higher level of gene diversity than does Africa. The present study, incorporating 13 *Alu* polymorphisms and a *LINE-1* element, indicates higher variation in Africa than elsewhere. A higher level of African diversity supports the hypothesis that modern humans first arose in Africa and then colonized other parts of the world (Stoneking 1993), but genetic diversity is related not just to a population's "age" but also to demographic events in a population's history, such as bottlenecks and effective population size (Relethford 1995; Rogers and Jorde 1995; Stoneking et al. 1997; Relethford and Jorde 1999).

These results are also consistent with other those of reports in showing that the ratio of African genetic diversity to non-African genetic diversity is much higher for mtDNA than for nuclear DNA (Hey 1997; Cavalli-Sforza 1998). This is sometimes attributed to natural selection on the mtDNA genome in non-African populations (Hey 1997), possibly as a result of adaptation to new climates as modern humans radiated out of Africa. The difference could also derive from the lower effective size of the mtDNA genome, which makes it more responsive to population bottleneck effects (Jorde et al. 1997; Fay and Wu 1999). Still another explanation is that a bottleneck effect is caused by mtDNA heteroplasmy (Cavalli-Sforza 1998). Arguing against the natural selection hypothesis is the fact that autosomal and Y-chromosome analyses tend to confirm the broad pattern of a rapid Pleistocene population expansion first suggested by mtDNA data (Rogers 1995; Shriver et al. 1997; Harpending et al. 1998; Kimmel et al. 1998; Reich and Goldstein 1998). It is unlikely that natural selection would operate in a similar fashion on autosomal, Y-chromosome, and mtDNA.

Another possible explanation for the higher ratio of African mtDNA diversity to non-African mtDNA diversity is that the autosomal systems, including STRPs, are subject to some degree of ascertainment bias and thus underestimate African diversity. Such a bias would not apply to mtDNA sequence because the latter is ascertained uniformly in all populations. This argument is now defused by several recent studies of nuclear-DNA sequence, all of which show excess African diversity, but generally at a level of 10%–30% or so (Nickerson et al. 1998; Zietkiewicz et al. 1998; Halushka et al. 1999; Kaessmann et al. 1999*a;* Rieder et al. 1999). An exception is the *PDHA1* locus, which shows much higher African than non-African diversity (Harris and Hey 1999). However, this is attributed to natural selection acting on this locus in non-Africans (Harris and Hey 1999). Considering these findings, it is probable that excess African mtDNA diversity is the result of lower mtDNA effective population size rather than natural selection.

Although the gene-diversity levels of each population are mostly similar for different types of genetic systems, there are some intriguing exceptions. In particular, the Finnish and northern-European populations show markedly reduced Y-chromosome variation relative to that in the other European populations, but they do not show such a reduction for other systems. This result confirms other studies that have compared Y-chromosome, autosomal, and mtDNA variation in Finland and

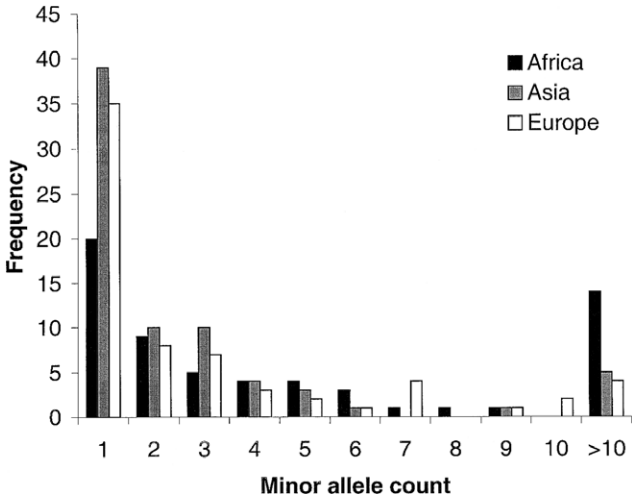

Figure 1 Distribution of minor-allele counts for HVS1 nucleotides in Africans, Asians, and Europeans. The *X*-axis indicates the copy number of each minor allele in each population (i.e., whether the allele is seen once, twice, etc.), and the *Y*-axis indicates the number of alleles that fall into each *X*-axis category.

that have concluded that a founder effect has been much more pronounced for Finnish males than for Finnish females (Sajantila et al. 1996; Kittles et al. 1999). The northern-European population, which consists of subjects of British and Scandinavian ancestry, shows a similar reduction of Y-chromosome diversity, possibly indicating a male-specific founder effect in these populations.

The worldwide G_{ST} values observed in this study are remarkably consistent with those obtained elsewhere. Other studies of autosomal RSPs have obtained G_{ST} values of 10%–15% (Bowcock et al. 1991; Barbujani et al. 1997). Similar values have been derived in analyses of autosomal STRP variation (Bowcock et al. 1994; Deka et al. 1995*a;* Barbujani et al. 1997; Calafell et al. 1998), although the G_{ST} values for microsatellite systems tend to be slightly lower than those based on RSPs. This reflects the higher mutation rate of microsatellite systems, which increases within-groups variation relative to between-groups variation and thus decreases G_{ST} (Jin and Chakraborty 1995). The G_{ST} value of 18% for the *Alu* systems is somewhat higher than the value of 13% that was obtained by an earlier study of eight *Alu* polymorphisms (Stoneking et al. 1997). Using individual populations as the unit of subdivision, as was done in the earlier study, we obtained a G_{ST} value of 15%.

The G_{ST} values observed for mtDNA are also similar to other published values. Two large surveys based on mitochondrial RSPs obtained G_{ST} values of 31% (Stoneking et al. 1990) and 35% (Merriwether et al. 1991). These estimates, which are slightly higher than the control-region estimates reported here, could be elevated, in part, because of the lower mutation rate outside the control region. As discussed above, this would tend to increase the relative level of between-group variation. Similarly, the slightly higher G_{ST} values observed here for HVS2 relative to HVS1 are consistent with a higher mutation rate for the latter sequence (Meyer et al. 1999).

As discussed above, the low G_{ST} value observed for the Y STRPs is caused mainly by a high degree of differentiation between some populations within continents (particularly the Finnish and northern-European populations). An examination of Y-STRP haplotypes shows that nearly all haplotype sharing occurs within populations and, occasionally, between populations that are located close to one another. A relative lack of large-scale geographic pattern in Y STRPs has been observed in other studies (Deka et al. 1996; de Knijff et al. 1997; Kayser et al. 1997). This can be attributed to a combination of high mutation rate in these microsatellites (Heyer et al. 1997) and to the fact that the effective population size of Y-chromosome polymorphisms, like that of mtDNA, is one-fourth that of autosomal DNA. A combination of high mutation rate and high drift rate is likely to erase long-term population history quickly. Thus, relationships at the continental level may become blurred, as was seen in a previous analysis of Y STRPs (Deka et al. 1996).

In this context, it is interesting that a preliminary analysis of three Y-chromosome RSPs and YAP in our data set yields worldwide G_{ST} estimates of 36.3% and 53.9% when continents and individual populations, respectively, are used as the unit of subdivision. These

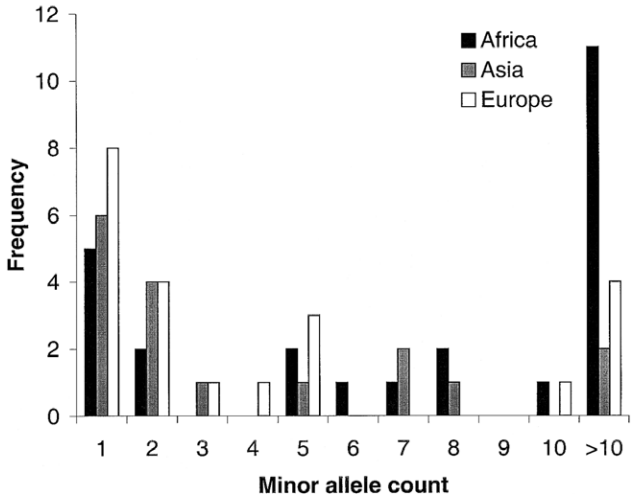

Figure 2 Distribution of minor-allele counts for HVS2 nucleotides in Africans, Asians, and Europeans. The *X*- and *Y*-axes are as described for figure 1.

estimates are substantially higher than the microsatellite estimates and are more similar to the estimate of 64% obtained by Seielstad et al. (1998). Also consistent with the RSP-STRP difference is a recent analysis of worldwide Y-chromosome variation which obtained a G_{ST} value of 40% for the YAP system and an average G_{ST} of only 8% for two Y STRPs (Quintana-Murci et al. 1999). These patterns suggest that Y STRPs may be useful for reconstruction of relatively recent population history but may be somewhat undependable, on their own, for reconstruction of ancient population history.

On the basis of a substantial elevation in worldwide Y-chromosome single-nucleotide polymorphism (SNP) variation (i.e., G_{ST}) relative to mtDNA G_{ST}, it has been suggested that, throughout much of human evolutionary history, females have experienced greater population movement than have males (Seielstad et al. 1998). An earlier study that compared Y-chromosome and mtDNA variation found a higher G_{ST} value for mtDNA variation than for Y-chromosome variation (Poloni et al. 1997), but the Y-chromosome RSP results reported here offer some support for the hypothesis. A large-scale comparison of Y-chromosome RSPs and SNPs with mtDNA in the same human populations is needed to help resolve this question.

Most systems show higher G_{ST} values for African populations than for other populations, but this pattern is not seen for either the *Alu* systems or the Y STRPs. Although the *Alu* G_{ST} is only slightly greater in Asia than in Africa, this result differs from that in a previous survey of eight *Alu* polymorphisms, in which African populations displayed the highest G_{ST} values (Stoneking et al. 1997). The Y-STRP pattern is influenced by the strong degree of local population differentiation. It is important to exercise caution in comparing G_{ST} values in these populations, because the number of populations within each continental group is relatively small and does not necessarily represent all of the variation within the continent. G_{ST} estimates are sensitive to the number and type of populations included in the sample (Jorde 1980), and they are also affected by the assumption that populations have differentiated to an equal degree at each level of hierarchy (Urbanek et al. 1996).

The hierarchical AMOVA analysis shows that, with the exception of Y STRPs, all systems show much less differentiation between populations within continents than between continents. This result is expected when there is greater gene flow between populations that are in close geographic proximity to one another. The autosomal values (table 5, row 2) are especially small, ranging from 1.3% for the RSPs to 1.8% for the *Alu* polymorphisms. This is in agreement with the small continental G_{ST} values shown in table 4. Although the sample sizes for individual populations in this analysis are relatively small, all of the values shown in table 5 differ significantly ($P < .0001$, by the permutation test) from 0. In addition, they are highly consistent both with one another and with previous analyses of worldwide variation in autosomal microsatellites and RFLPs, which also show considerably greater differentiation between continents than between populations within continents (Deka et al. 1995*b*; Barbujani et al. 1997). The fact that there is little differentiation between populations within continents has important implications in the forensic setting, in that it supports the current practice of grouping reference populations into broad ethnic categories when autosomal STRP data are used (National Research Council 1996).

In general, the results obtained here are encouraging in the broad congruency seen among different types of genetic systems. These systems portray similar accounts of the evolution of our species, and they support the general conclusions that humans show relatively little between-population diversity and that Africans have greater genetic diversity than do other populations. They thus provide further support for a relatively recent African origin of modern humans. The differences that are seen, for example, in Y-chromosome and mtDNA variation suggest intriguing phenomena in our history that need further testing with additional data from additional populations sampled throughout the world.

Acknowledgments

We are grateful to Trefor Jenkins, Kenneth Kidd, Judy Kidd, and Juha Kere for supplying some of the DNA samples used in this study. Alan Rogers provided useful comments. This work was supported by National Institutes of Health grants GM-52290 and RR-00064 and by National Science Foundation grants DBS-9310105, SBR-9514733, and SBR-9512178.

Electronic-Database Information

The URL for data in this article is as follows:

Y Chromosome Microsatellites, http://www.stats.ox.ac.uk/~pritch/data/ydata.html

References

Arcot SS, Adamson AW, Lamerdin JE, Kanagy B, Deininger PL, Carrano AV, Batzer MA (1996) Alu fossil relics—distribution and insertion polymorphism. Genome Res 6: 1084–1092

Arcot SS, Adamson AW, Risch G, LaFleur J, Lamerdin JE, Carrano AV, Batzer MA (1998) High-resolution cartography of recently integrated chromosome 19-specific Alu fossils. J Mol Biol 281:843–855

Arcot SS, DeAngelis MM, Sherry ST, Adamson AW, Lamerdin JE, Deininger PL, Carrano AV, et al (1997) Identification

and characterization of two polymorphic Ya5 Alu repeats. Mutat Res 382:5–11

Bamshad MJ, Watkins WS, Dixon ME, Rao BB, Naidu JM, Prasad BVR, Reddy PG, 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

Batzer MA, Stoneking M, Alegria-Hartman M, Bazan H, Kass DH, Shaikh TH, Novick GE, et al (1994) African origin of human-specific polymorphic Alu insertions. Proc Natl Acad Sci USA 91:12288–12292

Bowcock AM, Kidd JR, Mountain JL, Hebert JM, Carotenuto L, Kidd KK, Cavalli-Sforza LL (1991) Drift, admixture, and selection in human evolution: a study with DNA polymorphisms. Proc Natl Acad Sci USA 88:839–843

Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL (1994) High resolution of human evolutionary trees with polymorphic microsatellites. Nature 368:455–457

Calafell F, Shuster A, Speed WC, Kidd JR, Kidd KK (1998) Short tandem repeat polymorphism evolution in humans. Eur J Hum Genet 6:38–49

Cavalli-Sforza LL (1998) The DNA revolution in population genetics. Trends Genet 14:60–65

Crouau-Roy B, Service S, Slatkin M, Freimer N (1996) A fine-scale comparison of the human and chimpanzee genomes: linkage, linkage disequilibrium and sequence analysis. Hum Mol Genet 5:1131–1137

Deka R, Guangyun S, Smelser D, Zhong Y, Kimmel M, Chakraborty R (1999) Rate and directionality of mutations and effects of allele size constraints at anonymous, gene-associated, and disease-causing trinucleotide loci. Mol Biol Evol 16:1166–1177

Deka R, Jin L, Shriver MD, Yu LM, DeCroo S, Hundrieser J, Bunker CH, et al (1995*a*) Population genetics of dinucleotide (dC-dA)n.(dG-dT)n polymorphisms in world populations. Am J Hum Genet 56:461–474

Deka R, Jin L, Shriver MD, Yu LM, 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

Deka R, Shriver MD, Yu LM, Ferrell RE, Chakraborty R (1995*b*) Intra- and inter-population diversity at short tandem repeat loci in diverse populations of the world. Electrophoresis 16:1659–1664

de Knijff P, Kayser M, Caglia A, Corach D, Fretwell N, Gehrig C, Graziosi G, et al (1997) Chromosome Y microsatellites: population genetic and evolutionary aspects. Int J Legal Med 110:134–140

Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB (1994) Mutational processes of simple-sequence repeat loci in human populations. Proc Natl Acad Sci USA 91:3166–3170

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

Fay JC, Wu C-I (1999) A human population bottleneck can account for the discordance between patterns of mitochondrial versus nuclear DNA variation. Mol Biol Evol 16: 1003–1005

Halushka MK, Fan JB, Bentley K, Hsie L, Shen N, Weder A, Cooper R, et al (1999) Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. Nat Genet 22:239–247

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

Harpending HC, Batzer MA, Gurven M, Jorde LB, Rogers AR, Sherry ST (1998) Genetic traces of ancient demography. Proc Natl Acad Sci USA 95:1961–1967

Harris EE, Hey J (1999) X chromosome evidence for ancient human histories. Proc Natl Acad Sci USA 96:3320–3324

Hey J (1997) Mitochondrial and nuclear genes present conflicting portraits of human origins. Mol Biol Evol 14: 166–172

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

Jin L, Chakraborty R (1995) Population structure, stepwise mutations, heterozygote deficiency and their implications in DNA forensics. Heredity 74:274–285

Jorde LB (1980) The genetic structure of subdivided human populations: a review. In: Mielke JH, Crawford MH (eds) Current developments in anthropological genetics. Vol 1. Plenum Press, New York, pp 135–208

Jorde LB, Bamshad M, Rogers AR (1998) Using mitochondrial and nuclear DNA markers to reconstruct human evolution. Bioessays 20:126–136

Jorde LB, Bamshad MJ, Watkins WS, Zenger R, Fraley AE, Krakowiak PA, Carpenter KD, et al (1995) Origins and affinities of modern humans: a comparison of mitochondrial and nuclear genetic data. Am J Hum Genet 57:523–538

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

Kaessmann H, Heissig F, von Haeseler A, Pääbo S (1999*a*) DNA sequence variation in a non-coding region of low recombination on the human X chromosome. Nat Genet 22: 78–81

Kaessmann H, Wiebe V, Pääbo S (1999*b*) Extensive nuclear DNA sequence diversity among chimpanzees. Science 286: 1159–1162

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, 141–149

Kimmel M, Chakraborty R, King JP, Bamshad M, Watkins WS, Jorde LB (1998) Signatures of population expansion in microsatellite repeat data. Genetics 148:1921–1930

Kittles RA, Bergen AW, Urbanek M, Virkkunen M, Linnoila M, Goldman D, Long JC (1999) Autosomal, mitochondrial, and Y chromosome DNA variation in Finland: evidence for a male-specific bottleneck. Am J Phys Anthropol 108: 381–399

Lewontin RC (1972) The apportionment of human diversity. Evol Biol 6:381–398

Li W-H, Sadler LA (1991) Low nucleotide diversity in man. Genetics 129:513–523

Lum JK, Cann RL, Martinson JJ, Jorde LB (1998) Mitochondrial and nuclear genetic relationships among Pacific Island and Asian populations. Am J Hum Genet 63: 613–624

Merriwether DA, Clark AG, Ballinger SW, Schurr TG, Soodyall H, Jenkins T, Sherry ST, et al (1991) The structure of human mitochondrial DNA variation. J Mol Evol 33: 543–555

Merriwether DA, Huston S, Iyengar S, Hamman R, Norris JM, Shetterly SM, Kamboh MI, et al (1997) Mitochondrial versus nuclear admixture estimates demonstrate a past history of directional mating. Am J Phys Anthropol 102: 153–159

Meyer S, Weiss G, von Haeseler A (1999) Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. Genetics 152:1103–1110

Milewicz DM, Byers PH, Reveille J, Hughes AL, Duvic M (1996) A dimorphic Alu Sb-like insertion in COL3A1 is ethnic-specific. J Mol Evol 42:117–123

Mountain JL, Cavalli-Sforza LL (1994) Inference of human evolution through cladistic analysis of nuclear DNA restriction polymorphisms. Proc Natl Acad Sci USA 91:6515–6519

National Research Council (1996) The evaluation of forensic DNA evidence. National Academy Press, Washington, DC

Nei M, Livshits G (1990) Evolutionary relationships of Europeans, Asians, and Africans at the molecular level. In: Takahata N, Crow JF (eds) Population biology of genes and molecules. Baifukan, Tokyo, pp 251–265

Nei M, Livshits G, Ota T (1993) Genetic variation and evolution of human populations. In: Sing CF, Hanis CL (eds) Genetics of cellular, individual, family, and population variability. Oxford University Press, New York, pp 239–252

Nickerson DA, Taylor SL, Weiss KM, Clark AG, Hutchinson RG, Stengard J, Salomaa V, et al (1998) DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. Nat Genet 19:233–240

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

Perez-Lezaun A, Calafell F, Comas D, Mateu E, Bosch E, Martinez-Arias R, Clarimon J, et al (1999) Sex-specific migration patterns in Central Asian populations, revealed by analysis of Y-chromosome short tandem repeats and mtDNA. Am J Hum Genet 65:208–219

Poloni ES, Semino O, Passarino G, Santachiara-Beneceretti 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

Quintana-Murci L, Semino O, Poloni ES, Liu A, Van Gijn M, Passarino G, Brega A, et al (1999) Y-chromosome specific YCAII, DYS19 and YAP polymorphisms in human populations: a comparative study. Ann Hum Genet 63: 153–166

Reich DE, Goldstein DB (1998) Genetic evidence for a Paleolithic human population expansion in Africa. Proc Natl Acad Sci USA 95:8119–8123

Relethford JH (1995) Genetics and modern human origins. Evol Anthropol 4:53–63

Relethford JH, Harpending HC (1994) Craniometric variation, genetic theory, and modern human origins. Am J Phys Anthropol 95:249–270

Relethford JH, Jorde LB (1999) Genetic evidence for larger African population size during recent human evolution. Am J Phys Anthropol 108:251–260

Rieder MJ, Taylor SL, Clark AG, Nickerson DA (1999) Sequence variation in the human angiotensin converting enzyme. Nat Genet 22:59–62

Rogers AR (1995) Genetic evidence for a Pleistocene population explosion. Evolution 49:608–615

Rogers AR, Jorde LB (1995) Genetic evidence on the origin of modern humans. Hum Biol 67:1–36

——— (1996) Ascertainment bias in estimates of average heterozygosity. Am J Hum Genet 58:1033–1041

Sajantila A, Salem A, Savolainen P, Bauer K, Gierig C, Pääbo S (1996) Paternal and maternal DNA lineages reveal a bottleneck in the founding of the Finnish population. Proc Natl Acad Sci USA 93:12035–12039

Schneider S, Kueffer J-M, Roesslie D, Excoffier L (1997) Arlequin, version 1.1: a software for population genetic data analysis. Genetics and Biometry Laboratory, University of Geneva, Geneva

Scozzari R, Cruciani F, Malaspina P, Santolamazza P, Ciminelli BM, Torroni A, Modiano D, et al (1997) Differential structuring of human populations for homologous X and Y microsatellite loci. Am J Hum Genet 61:719–733

Seielstad M, Bekele D, Ibrahim M, Touré A, Traoré M (1999) A view of modern human origins from Y chromosome microsatellite variation. Genome Res 9:558–567

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

Shriver MD, Jin L, Boerwinkle E, Deka R, Ferrell RE, Chakraborty R (1995) A novel measure of genetic distance for highly polymorphic tandem repeat loci. Mol Biol Evol 12: 914–920

Shriver MD, Jin L, Ferrell RE, Deka R (1997) Microsatellite data support an early population expansion in Africa. Genome Res 7:586–591

Spurdle AB, Jenkins T (1996) The origins of the Lemba "Black Jews" of southern Africa: evidence from p12F2 and other Y-chromosome markers. Am J Hum Genet 59: 1126–1133

Stoneking M (1993) DNA and recent human evolution. Evol Anthropol 2:60–73

Stoneking M, Fontius JJ, Clifford SL, Soodyall H, Arcot SS, Saha N, Jenkins T, et al (1997) *Alu* insertion polymorphisms and human evolution: evidence for a larger population size in Africa. Genome Res 7:1061–1071

Stoneking M, Jorde LB, Bhatia K, Wilson AC (1990) Geographic variation of human mitochondrial DNA from Papua New Guinea. Genetics 124:717–733

Urbanek M, Goldman D, Long JC (1996) The apportionment of dinucleotide diversity in Native Americans and Europeans: a new approach to measuring gene identity reveals asymmetric patterns of divergence. Mol Biol Evol 13: 943–953

Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC (1991) African populations and the evolution of human mitochondrial DNA. Science 253:1503–1507

Watkins WS, Bamshad MJ, Jorde LB (1995) Population genetics of trinucleotide repeat polymorphisms. Hum Mol Genet 4:1485–1491

Wright S (1965) The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution 19:395–420

Zietkiewicz E, Yotova V, Jarnik M, Korab-Laskowska M, Kidd KK, Modiano D, Scozzari R, et al (1998) Genetic structure of the ancestral population of modern humans. J Mol Evol 47:146–155