
VNTR DNA Variation in Siberian Indigenous Populations

J. McCOMB,¹ N. BLAGITKO,² A.G. COMUZZIE,³ M.S. SCHANFIELD,⁴
R.I. SUKERNIK,² W.R. LEONARD,⁵ AND M.H. CRAWFORD¹

Abstract The VNTR loci D7S104, D11S129, D18S17, D20S15, and D21S112 in three indigenous Siberian populations were analyzed to determine the populations' genetic structure. Using the Kolmogorov-Smirnov test, we found that the Siberian indigenous populations of Surinda and Sulamai are separated at the D11S129 locus ($p < 0.05$). However, the population of Poligus is genetically homogeneous compared with the villages of Sulamai and Surinda. Principal component plots for the sets of VNTR loci cluster the Siberian groups together, reflecting the homogeneity of these populations. An analysis of mean per locus heterozygosity versus the distance from the centroid of distribution suggests gene flow into Sulamai but little genetic exchange with Surinda and Poligus. Ultimately, the VNTR data reflect the genetic distinctiveness of the Kets and the Evenki.

Over the past five years frequency distributions of variable-number tandem repeats (VNTRs) have been used to discriminate between ethnically defined populations (Chakraborty et al. 1992; Deka et al. 1992; Balazs et al. 1989). These studies have demonstrated the utility of these DNA polymorphisms in measuring population affinities and in reconstructing phylogenetic relationships. Although VNTRs appear to confirm the information provided by protein polymorphisms, they possess several unique characteristics: (1) VNTRs are associated with a high mutation rate (Jeffreys et al. 1988), which generates a wider range of polymorphisms than most proteins; and (2) because most VNTRs are noncoding segments of

¹Laboratory of Biological Anthropology, Department of Anthropology, 622 Fraser Hall, University of Kansas, Lawrence, KS 66045.

²Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia.

³Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio, TX 78228.

⁴Analytical Genetic Testing Center, Denver, CO 80231.

⁵School of Human Biology, Guelph University, Guelph, Ontario, Canada N1G 2W1.

Human Biology, April 1995, v. 67, no. 2, pp. 217-229.

Copyright © 1995 Wayne State University Press, Detroit, Michigan 48201-1309

KEY WORDS: VNTR, SIBERIA, EVENKI, KETS, POPULATION STRUCTURE

DNA, they are less affected by natural selection (Harding 1992). These characteristics make VNTRs particularly effective in examining recent evolutionary events, such as migration (Harding 1992).

Siberia is of evolutionary importance because it is the crossroads between Europe, Asia, and ultimately the New World. The genetic relationships among indigenous Siberian groups have been studied using traditional markers (Crawford and Enciso 1982), mitochondrial DNA (Torroni et al. 1993), and other DNA markers. In this study we have analyzed data from five VNTR loci to examine the phylogenetic relationships between three Siberian groups. In addition to using the VNTR frequencies to differentiate the Siberians, we also study the effectiveness of VNTR data to detect common genetic phenomena, such as gene flow. Last, we use VNTR frequencies of the Siberians and US populations to construct genetic maps that reflect genetic relationships between the Siberians and the other groups.

Materials and Methods

Populations. During the summers of 1991 and 1992 an international field research team sampled two adjacent Evenki settlements (Surinda and Poligus) and one Ket village (Sulamai), located along the Stony Tunguska River and its tributaries in the Evenki Autonomous Region of central Siberia (see Figure 1).

The two Evenki populations each number approximately 600 residents. The Evenki are a Tungusic-speaking ethnic group whose settlements are geographically distributed along the vast expanse of the boreal forest (taiga) of central Siberia to the Lower Amur region of eastern Asia. Although the cultural and biological origins of the Evenki contain some uncertainties, it is likely that they were once reindeer herders who eventually domesticated reindeer and adopted a breeding and herding form of subsistence (Vasilevich 1946). The Evenki were socially organized into patrilineal clans (family lineages) that held ownership of the reindeer herds and served as herding units until the 1930s, when they were forcibly collectivized into cooperative settlements and brigades (Forsyth 1992). Approximately 30,000 Evenki presently reside in Siberia; most of them (79.8%) continue to exist in isolated and rural areas (Hannigan 1991).

The Kets, primarily located along the Yenisey River and its effluents, practice a fishing and hunting form of subsistence. Their language is unintelligible to the surrounding populations and cannot be included in the three major linguistic phyla of Siberia. The origin of the Kets is somewhat enigmatic; their folklore dictates a southern homeland on the far side of a high impassable mountain range (Popov and Dolgikhi



Figure 1. Location of the central Siberian villages sampled in this study.

1956). The Kets sampled for this study came from the small village of Sulamai, located near the juncture of the Yenisey and Stony Tunguska rivers. This settlement of approximately 120 persons includes approximately equal numbers of Russians and Kets. Thus our small sample of 22 Kets represents more than one-third of their total in Sulamai and most of the adults.

The Evenki and Ket populations were compared with four composite population samples compiled by the Analytical Genetic Testing Center (Denver, Colorado): US whites, US blacks, Mexican Americans, and Native Americans. The sample numbers for these groups are given in Table 1.

DNA Analyses. Blood samples (10 ml) were collected in the 1991 and 1992 field seasons from the indigenous populations of central Siberia. The blood was iced and shipped to the Laboratory of Biological Anthropology at the University of Kansas for analysis. DNA was extracted from the buffy coats using a Super Quik Gene kit (Analytical Genetic Testing Center). Ten μg of DNA was digested overnight using

Table 1. Sample Sizes of the Reference Populations Used in the Analyses

<i>Reference Population</i>	<i>D7S104</i>	<i>D11S129</i>	<i>D18S17</i>	<i>D20S15</i>	<i>D21S112</i>
Native Americans	140	41	57	62	170
Mexican Americans	597	135	160	149	379
US blacks	695	284	253	250	553
US whites	1881	461	640	387	1367
Surinda	8	79	64	54	8
Poligus	18	18	18	18	18
Sulamai	22	22	22	22	22

*Pst*I at 37°C. The digests and several sizing ladders were run electrophoretically on a 0.8% agarose gel for 24 hr at 45 V. The DNA was transferred to Pall Biodyne B nylon membranes using Southern blotting. The membranes were hybridized overnight at 42°C with the probes CRI-PAT-pS194 (D7S104), CRI-PAT-pR365-1 (D11S129), CRI-PAT-pL159-1 (D18S17), CRI-PAT-pL355-8 (D20S15), and CRI-PAT-pL427-4 (D21S112) using HS-45 hybridization solution (Analytical Genetic Testing Center) containing 45% formamide. The probes and repeat sizes are described in Table 2. These probes have been described by Barker et al. (1987), Helms et al. (1992), Schumm et al. (1988), and Takaesu et al. (1992). After hybridization the VNTRs were detected nonisotopically using the Non-Isotopic Southern Detection System (Analytical Genetic Testing Center).

For the 1991 field season 75 individuals from Surinda were tested using probes for D11S129, D18S17, and D20S15. After the 1992 field season DNA was extracted from an additional 8 persons from Surinda, 18 from Poligus, and 22 from Sulamai. These specimens were probed for D7S104, D11S129, D18S17, D20S15, and D21S112.

Statistical Analyses. Allelic frequencies were established by placing the VNTR fragment sizes into 2% error bins. Expected heterozygosities

Table 2. Characteristics of the Probes and Loci Used in the Analyses

<i>Locus</i>	<i>Probe</i>	<i>Repeat Length (bp)</i>
D7S104	pS194	50
D11S129	pR365-1	28
D18S17	pL159-1	48
D20S15	pL355-8	33-35
D21S112	pL427-4	26-30

Probes provided by Collaborative Research Inc., Waltham, Massachusetts.

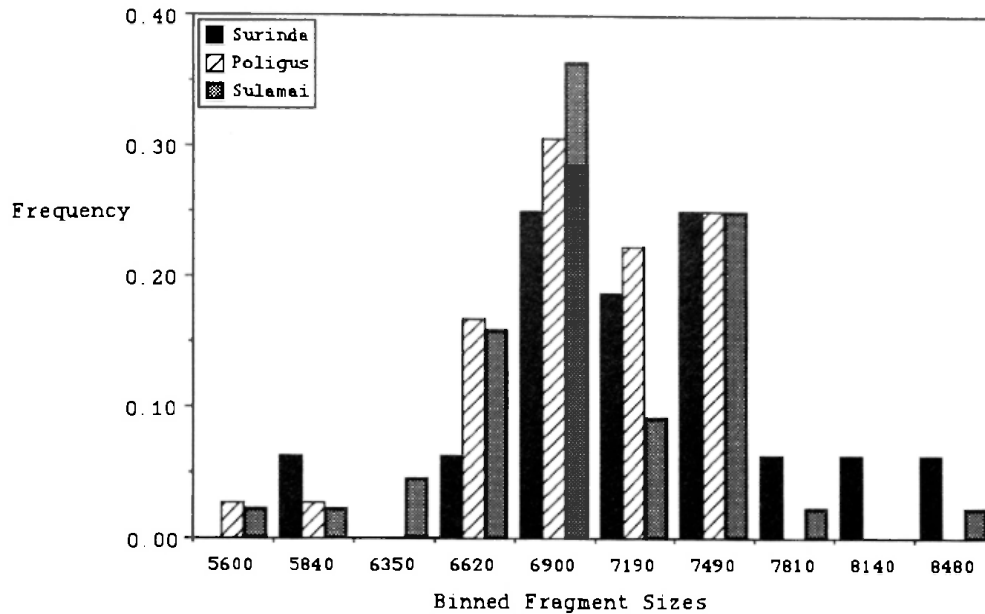


Figure 2. Distribution of alleles at the D7S104 locus.

were calculated using the binned frequencies by the method of Harpending and Chasko (1976). The magnitude of gene flow between the populations was estimated by the regression of the distance from the centroid (r_{ii}) versus the mean per locus heterozygosity (H) (Harpending and Ward 1982).

The statistical significance of VNTR fragment distributions between populations was measured using the Kolmogorov-Smirnov test with Bonferroni's protection (Sokal and Rohlf 1981). In addition, genetic maps for the Siberian populations were constructed using the R matrix method (Harpending and Jenkins 1973). Because census size was not available for the compiled populations, the R matrix was weighted by sample size. Because the sample sizes of the Siberian groups are small, genetic differentiation for each R matrix plot was measured by a G_{ST} statistic (Nei 1973), which was weighted by sample size and is independent of the number of subdivisions.

Results and Discussion

The initial histograms (Figures 2–5) revealed fairly continuous distributions of the DNA fragments in the Siberian populations in all the loci tested except D11S129 (Figure 6). Locus D11S129 has a bimodal distribution with a large gap starting at 2.1 kb and continuing to 2.7 kb. This void is not unique and appears in all the populations tested to date by probe pR365-1. Only 6 of the 2076 alleles detected by this probe in

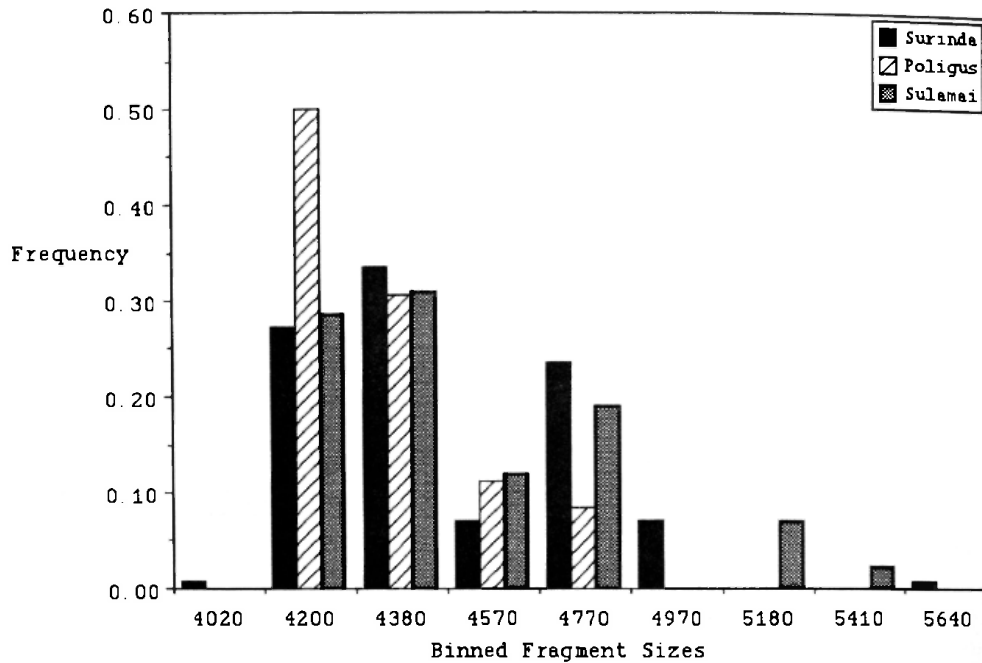


Figure 3. Distribution of alleles at the D18S17 locus.

7 populations have been detected in this 0.6-kb area, with 5 alleles being found among US whites.

The distribution of the D11S129 locus has at least two possible explanations: (1) A restriction site has been lost (or gained) so that some

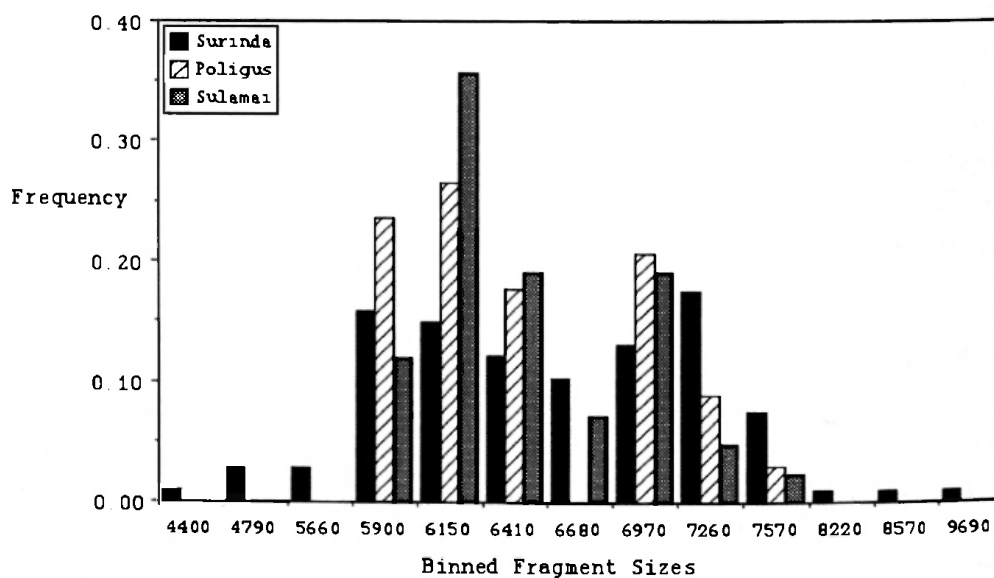


Figure 4. Distribution of alleles at the D20S15 locus.

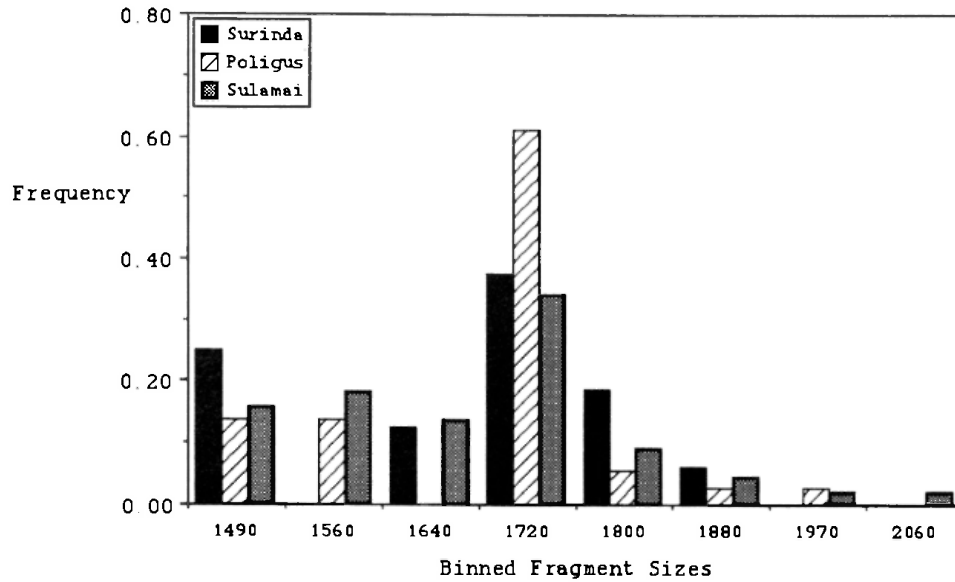


Figure 5. Distribution of alleles at the D21S112 locus.

fragments are cut about 0.6 kb longer; or (2) a 0.6-kb section of DNA has been inserted, duplicated, or deleted from the flanking region. As yet, neither of these hypotheses has been tested, and this should be the subject of future research.

In addition, D11S129 is the only locus that can be used to discriminate between the Siberian populations. The Kolmogorov-Smirnov test

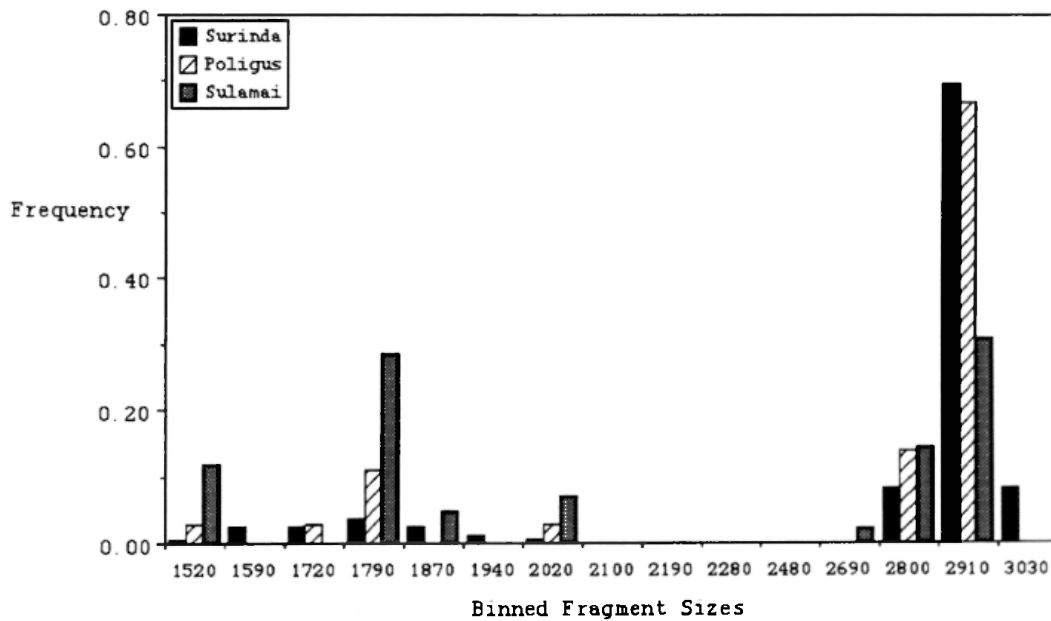


Figure 6. Distribution of alleles at the D11S129 locus.

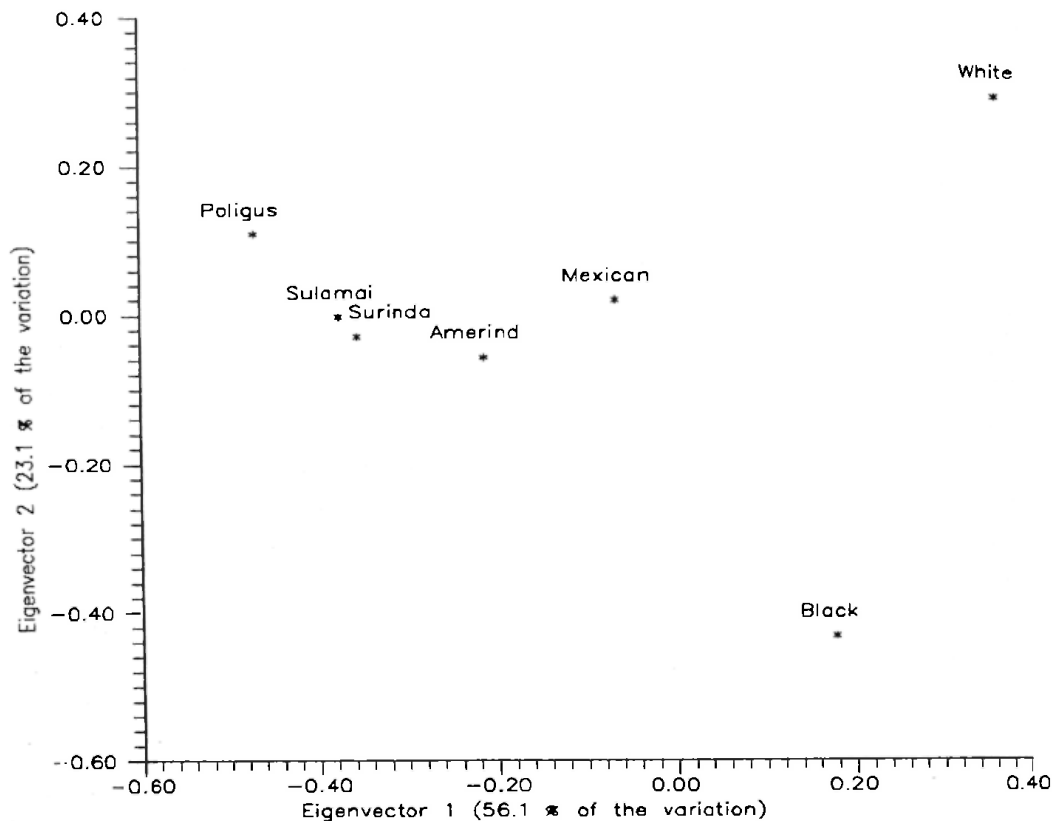


Figure 7. Genetic map based on a least-squares reduction of an R matrix for loci D11S129, D18S17, and D20S15 using the Siberian populations.

for locus D11S129 indicates that the allelic distributions of Sulamai and Surinda are significantly different ($D = 0.469$; $p < 0.05$). The tests of the other loci indicate that the Siberian groups are genetically homogeneous. The samples of locus D11S129 from Surinda (one-eighth of the village) and Sulamai (one-third of the Kets) are large relative to the number of people in each village, and thus this significant difference may be the result of stochastic processes acting at this locus.

As was demonstrated by Balazs et al. (1989), VNTRs can be used to discriminate between American ethnic groups: (1) American whites and blacks can be significantly separated by the five loci ($p < 0.05$, with Bonferroni's protection); (2) American blacks and Mexican Americans can be separated ($p < 0.05$) at loci D7S104, D11S129, D20S15, and D21S112; and (3) American whites and Mexican Americans can be separated ($p < 0.05$) with loci D7S104, D11S129, and D18S17. The Siberian populations and the compiled samples can be separated from each other significantly, depending on the loci used.

Figure 7 shows the results of the R matrix analysis using the frequencies of VNTRs from loci D11S129, D18S17, and D20S15. The first

eigenvector accounts for 56.1% of the variance, whereas 23.1% of the total variance is explained by the second scaled eigenvector. As expected, the Evenki and the Kets of Sulamai exhibit close genetic affinity and cluster together. Native Americans also cluster with the Siberians, showing a genetic relationship between Siberia and the New World. The Mexican American group, a tri-ethnic hybrid, probably clusters with the Siberians because its gene pool predominantly contains European (white) and Amerindian allelic frequencies with a small proportion of African (black) alleles. These results demonstrate the utility of VNTR analyses because the different loci reflect the expected evolutionary relationships between the populations. An overall G_{ST} value of 0.082 suggests a medium level of microdifferentiation between population subdivisions. This statistic complements the Kolmogorov-Smirnov results, which indicate that Mexican Americans and US whites and blacks (the populations that are most differentiated in the plot) are all significantly different.

The plot of the distance from the centroid (r_{ii}) versus mean per locus heterozygosity (H) using loci D11S129, D18S17, and D20S15 provides evidence of gene flow and isolation in the Siberian populations (Figure 8). The theoretical regression line suggests an isolation by distance model, and none of the populations are outliers. However, some trends are apparent: (1) The Sulamai population exhibits a heterozygosity level above the regression line and apparently has experienced a higher magnitude of gene flow than did the Evenki; (2) in contrast, both Poligus and Surinda have a large r_{ii} and a low heterozygosity, suggesting that these populations have experienced genetic isolation and little gene flow. This correlates well with the ethnographic data because, although only 12% of the Evenki report admixture with Russians or other ethnic groups, 27.3% of the Kets trace admixture with Russians in their genealogies.

Because of the small sample of Surinda loci D7S104 and D21S112 and because these loci are redundant, a plot of them is not presented here. Plots created using loci D7S104 and D21S112 (with the Evenki groups lumped together) appear to be identical to Figure 8.

Figure 9 is an R matrix plot of the Siberian populations in relation to the compiled US data using allelic frequencies from all the loci. Unfortunately, the sample size of Surinda for loci D7S104 and D21S112 is small and thus necessitates combining the Surinda data with the Poligus data to form an Evenki sample. The first two scaled eigenvectors account for 80.9% of the gene frequency variance. The first eigenvector accounts for 41.3% of the variance, and the second eigenvector accounts for 39.6% of the variance. Although the position of several of the populations has shifted, this plot mimics the results of the plot using three loci. The Siberians form a cluster with the Mexican Americans and Native Americans that is distinct from the American white and black populations.

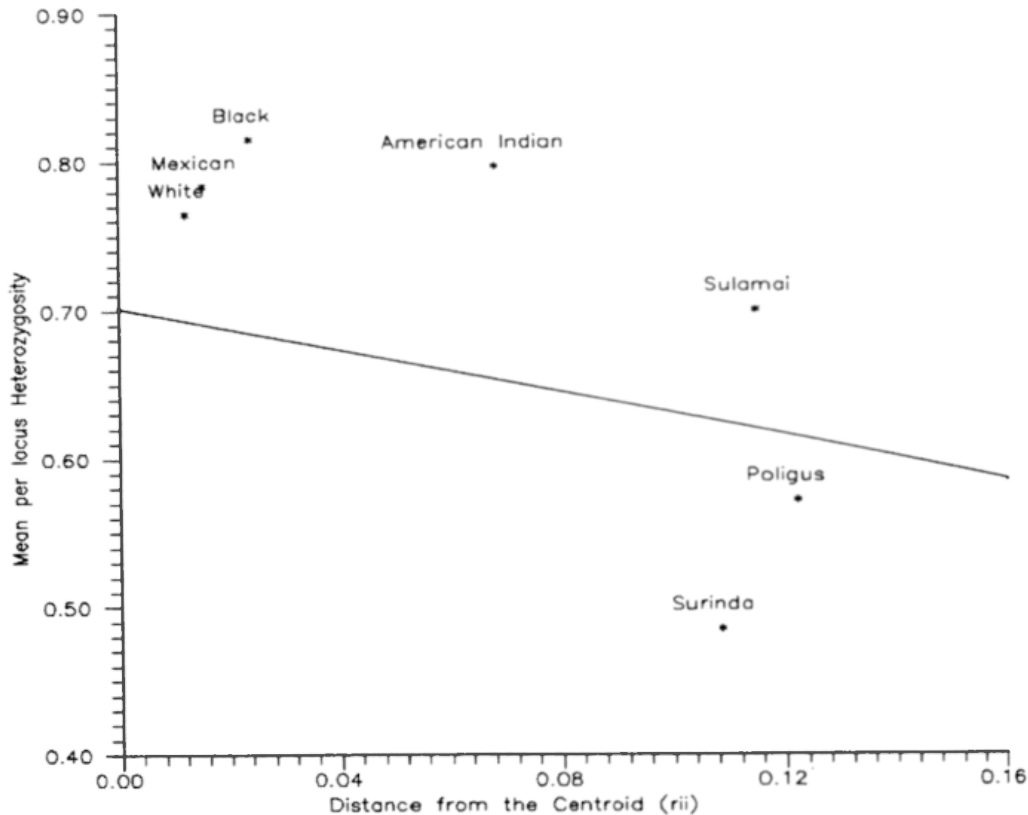


Figure 8. Regression between mean per locus heterozygosity and distance from the centroid of distribution for loci D11S129, D18S17, and D20S15.

The overall G_{ST} of this plot is 0.086, again indicating that the groups are relatively distinct.

Conclusion

This study reveals two key points about the indigenous Siberian populations and their VNTR distributions: (1) The Siberian groups are genetically similar to each other but are distinct from the compiled US aggregates of different ethnic origins; and (2) the level of heterozygosity in the Evenki indicates that they are genetically isolated, whereas apparently the Kets of Sulamai experienced higher rates of gene flow from Russian populations. The smaller, more isolated populations, such as those of Sulamai, Surinda, and Poligus, exhibit fewer alleles at the VNTR loci tested than do the synthetic groups such as American blacks and whites and Mexican Americans. This is probably the result of stochastic processes operating on small populations. In addition, this study demonstrates the general utility of VNTRs in the examination of population structure. Although different loci were used in the construction of each

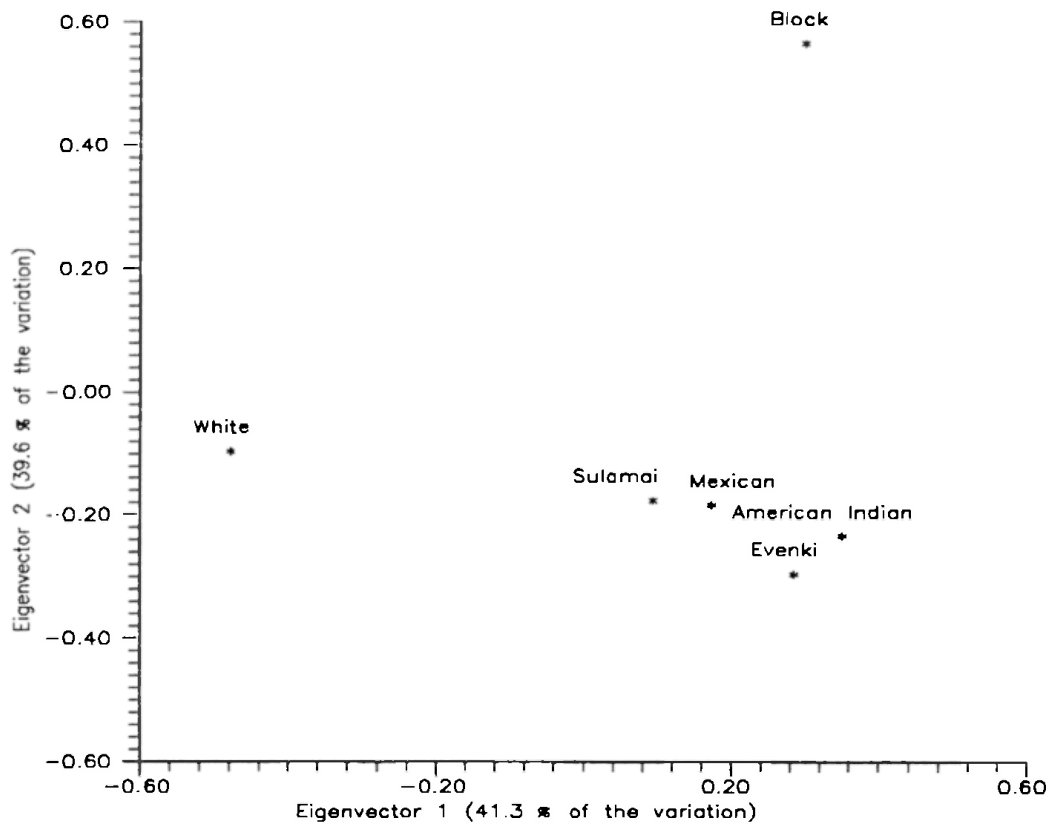


Figure 9. Least-squares reduction of an R matrix based on the allelic frequencies from five VNTR loci of Siberian and US populations.

R matrix, the results are similar. The Siberian populations tend to cluster together. The Mexican Americans and Native Americans also cluster near the Siberians, reflecting the phylogenetic relationships between these groups. Ultimately, the results of this study are in agreement with the documented ethnographic separation between the Kets and the Evenki. Also, these results reflect the Siberian origins of Native Americans, as these groups cluster closer to the Siberians than to other populations. In conclusion, these results demonstrate the usefulness of VNTR distributions in the reconstruction of the phylogenetic relationships between human populations.

Acknowledgments We thank the people of Sulamai, Surinda, and Poligus for their cooperation and goodwill. This research would not have been possible without the help of the nurses, physicians, and administrators from these three communities and from the administrative center in Baikit. We also thank the helicopter pilots who braved the wilds of the taiga on a daily basis. Our lives and the safety of our limbs depended on their judgment and skill.

In addition, we would like to acknowledge T. Keith, Collaborative Research Inc., Waltham, Massachusetts, for the size of the repeats detected by the various CRI probes. We also acknowledge the technical assistance of Lisa Dowler, Shirley Miller, Linda Netzel, and Clay Stern for training in DNA extraction and assistance in sizing DNA membranes.

This research was supported by the National Science Foundation (Man and the Biosphere Program) through grant BSR-99101571, by the Natural Sciences and Engineering Research Council of Canada through grant OGP-0116785, and by the General Research Fund of the University of Kansas.

Received 26 May 1994; revision received 29 September 1994.

Literature Cited

- Balazs, I., M. Baird, M. Clyne et al. 1989. Human population genetic studies of five hypervariable DNA loci. *Am. J. Hum. Genet.* 39:182–190.
- Barker, D., P. Green, R. Knowlton et al. 1987. Genetic linkage map of human chromosome 7 with 63 DNA markers. *Proc. Natl. Acad. Sci. USA* 84:8006–8010.
- Chakraborty, R., R. Deka, L. Jin et al. 1992. Allele sharing at six VNTR loci and genetic distances among three ethnically defined human populations. *Am. J. Hum. Biol.* 4:387–397.
- Crawford, M.H., and V.B. Enciso. 1982. Population structure of circumpolar groups of Siberia, Alaska, Canada, and Greenland. In *Current Developments in Anthropological Genetics*, M.H. Crawford and J.H. Mielke, eds. New York: Plenum Press, 51–91.
- Deka, R., S. Deroo, L.M. Yu et al. 1992. Variable number tandem repeat (VNTR) polymorphism at locus D17S5 (YNZ22) in four ethnically defined human populations. *Hum. Genet.* 90:86–90.
- Forsyth, J. 1992. *A History of the Peoples of Siberia*. Cambridge, England: Cambridge University Press.
- Hannigan, J. 1991. *Statistics on the Economic and Cultural Development of the Northern Aboriginal People of the USSR (for the Period of 1980–1989)*. Ottawa, Canada: Bureau of Indian and Northern Affairs.
- Harding, R.M. 1992. VNTRs in review. *Evol. Anthropol.* 1:62–71.
- Harpending, H., and W. Chasko. 1976. Heterozygosity and population structure in southern Africa. In *The Measures of Man*, E. Giles and J.S. Friedlaender, eds. Cambridge, MA: Peabody Museum Press, 214–229.
- Harpending, H., and T. Jenkins. 1973. Genetic distances among southern African populations. In *Methods and Theories of Anthropological Genetics*, M.H. Crawford and P.L. Workman, eds. Albuquerque, NM: University of New Mexico Press, 177–199.
- Harpending, H.C., and R.H. Ward. 1982. Chemical systematics and human populations. In *Biological Aspects of Evolutionary Biology*, M.H. Nitecki, ed. Chicago, IL: University of Chicago Press, 213–256.
- Helms, C., S.K. Mishra, H. Riethman et al. 1992. Closure of a genetic linkage map of human chromosome 7q with centromere and telomere polymorphisms. *Genomics* 14:1041–1054.

- Jeffreys, A.J., N.J. Royle, V. Wilson et al. 1988. Spontaneous mutation rates to new length alleles at tandem-repetitive hypervariable loci in human DNA. *Nature* 332:278–281.
- Nei, M. 1973. Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci. USA* 70(12):3321–3323.
- Popov, A.A., and B.O. Dolgikh. 1956. The Kets. In *The Peoples of Siberia*, M.G. Levin and L.P. Potapov, eds. Chicago, IL: University of Chicago Press, 607–619.
- Schumm, J.W., R.G. Knowlton, J.C. Braman et al. 1988. Identification of more than 500 RFLPs by screening random genomic clones. *Am. J. Hum. Genet.* 42:143–159.
- Sokal, R.R., and F.J. Rohlf. 1981. *Biometry*. New York: W.H. Freeman.
- Takaesu, N., S. Newfeld, and T. Hassold. 1992. Characterization of three VNTR systems at D21S112. *Genomics* 14:816–817.
- Torroni, A., R.I. Sukernik, T.G. Schurr et al. 1993. mtDNA variation of aboriginal Siberians reveals distinct genetic affinities with Native Americans. *Am. J. Hum. Genet.* 53:591–608.
- Vasilevich, G.M. 1946. Most ancient ethnic group of Asians, termed the tribes of Ev-enki. *Sov. Etnogr.*, special issue 4.