12-2005

Applications of Pooled DNA Samples to the Assessment of Population Affinities: Short Tandem Repeats

M. H. Crawford
University of Kansas

P. Banerjee
Columbia University

D. A. Demarchi
Universidad Nacional de Cordoba

M. Zlojutro
University of Kansas

J. McComb
University of Kansas

See next page for additional authors

Follow this and additional works at: https://cedar.wwu.edu/anthropology_facpubs

Part of the Biological and Physical Anthropology Commons, and the Medicine and Health Sciences Commons

Authors

M. H. Crawford, P. Banerjee, D. A. Demarchi, M. Zlojutro, J. McComb, G. Livshits, M. Henneberg, M.J. Mosher, M. S. Schanfield, and J. A. Knowles

REFERENCES

Linked references are available on JSTOR for this article:
https://www.jstor.org/stable/41466370?seq=1&cid=pdf-reference#references_tab_contents
You may need to log in to JSTOR to access the linked references.

Applications of Pooled DNA Samples to the Assessment of Population Affinities: Short Tandem Repeats

M. H. CRAWFORD,[1] P. BANERJEE,[2] D. A. DEMARCHI,[3] M. ZLOJUTRO,[1] J. MCCOMB,[1] G. LIVSHITS,[4] M. HENNEBERG,[5] M. J. MOSHER,[1,6] M. S. SCHANFIELD,[7] AND J. A. KNOWLES[2]

Abstract   Pooled DNA samples have been used in association studies of Mendelian disease genes. This method involves combining equal quantities of DNA from patients and control subjects into separate pools and comparing the pools for distributions of genetic markers. In this study identical quantities of DNA from 300 individuals representing 6 populations were pooled and amplified for 296 loci using the touchdown polymerase chain reaction (PCR) method. The purpose of this study is to test the efficacy of pooled DNA markers in the reconstruction of the genetic structure of human populations. The populations sampled included Chuvash, Buryats, Kizhi, Native Americans, South Africans, and New York City whites. To test the accuracy of the allele-frequency distributions, we genotyped the Buryats and New York samples individually for six microsatellite markers and compared their frequencies to the allele frequencies derived from the electropherogram peak heights for the pooled DNA, producing a correlation of 0.9811 with a variance of less than 0.04. Two-dimensional scaling of genetic distances among the six populations produced clusters that reflected known historical relationships. A distance matrix was created using all 296 loci, and matrices based on individual chromosomes were correlated against the total matrix. As expected, the largest chromosomes had the highest correlations with the total matrix, whereas one of the smallest chromosomes, chromosome 22, had the lowest correlation and differed most from the combined STR distance matrix.

Genetic markers and their allele frequencies in human populations reflect the evolutionary history and the genetic structure of these aggregates. However,

[1]Laboratory of Biological Anthropology, University of Kansas, Lawrence, KS 66045.
[2]Department of Psychiatry, College of Physicians and Surgeons, Columbia University, New York, NY.
[3]Museo de Antropologia, Universidad Nacional de Cordoba, 5000 Cordoba, Argentina.
[4]Department of Anatomy, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel.
[5]Department of Anatomical Sciences, University of Adelaide, Adelaide, Australia.
[6]Department of Epidemiology, University of North Carolina, Chapel Hill, NC.
[7]Department of Forensic Sciences, George Washington University, Washington, DC; and Analytical Genetic Testing, Denver, CO.

KEY WORDS: POOLED DNA, SHORT TANDEM REPEATS, GENETIC DIVERSITY, *R* MATRIX ANALYSIS, CHUVASH, BURYATS, KIZHI, NATIVE AMERICANS, SOUTH AFRICANS, NEW YORK CITY WHITES.

analyses of markers from different regions of the genome often reveal vastly different evolutionary relationships between and among populations. Mitochondrial (mt) DNA variation, which results from the absence of recombination and the maternal transmission of mtDNA, reflects the patterns of migration and gene flow of females and often provides population genetic affinities that are distinct from those revealed by Y-chromosome or short tandem repeat (STR) markers. For example, based on mtDNA haplogroup frequencies, the Aleuts of Alaska and Siberia display only Native American and Siberian haplogroups, despite the historically documented gene flow from Russians during previous centuries of contact (Rubicz et al. 2003). The use of both mtDNA and Y-chromosome markers (NRY) revealed sex-specific gene flow patterns in Indian castes of Andhra Pradesh (Bamshad et al. 1998). However, different regions of the genome used to reconstruct the peopling of the Americas yielded varying numbers of migrations and affinities with possible Siberian ancestral groups (Crawford 1998).

Nuclear DNA reflects the recombined effects of the evolutionary processes that may have had an impact on the population throughout its history. However, because coding regions of the genome are more likely to experience the action of natural selection, populations sharing similar disease loads or other environmental constraints may share identical coding genes because of convergence rather than recent common ancestry. Tandem repeats are scattered throughout the genome, and if all regions are sampled equally, the repeats should provide an unbiased measure (no gender-specific migration) of population phylogeny. However, there is some question about the number of loci that are necessary to adequately sample the entire genome. Jorde et al. (1997) proposed that a minimum of 30 STR loci are required to reliably reconstruct population phylogeny. A total of 300–400 short tandem repeats have been used to create a 10-centimorgan (cM) map for purposes of linkage analyses. In most laboratories the use of a large number of these anonymous markers, distributed throughout the genome, for studies involving numerous populations is not economically feasible. Furthermore, the cost-benefit ratio, the threshold where the added cost of analysis may not yield a sufficient amount of new information, needs to be considered.

The amplification and genotyping of pooled DNA samples from multiple individuals has been used to reduce labor and costs (Arnheim et al. 1985; Michelmore et al. 1991; Asada et al. 1994). This pooling technique has been successfully applied to mapping numerous Mendelian disease genes (Arbour et al. 1997; Brennan et al. 1998; Shaw et al. 1998; Barcellos et al. 1997). Pooling was used in a case-control association study of HLA-DR and HLA-DQ alleles in insulin-dependent diabetes (Arnheim et al. 1985). However, this method does not provide information about individual genotypes necessary for the detection of heterozygotes. Dubreuil et al. (1999) tested the reliability of this method for comparing maize populations, using pooled DNA versus individual genotyping. They observed that by using the pooled approach, allele frequencies could be

estimated with high precision. However, the overall genotype pattern of pooled samples may be distorted by PCR artifacts. Schnack et al. (2004) proposed a method for correcting the frequencies derived from stutter-corrected DNA pool patterns.

Shaw et al. (1998) investigated the accuracy of the allele-frequency estimates on human groups generated by PCR amplification of DNA pools using a 5′ fluorescently labeled oligonucleotide (quantified by an ABI 377 DNA sequencer) and found that the frequencies were accurate. According to this study, tetranucleotide markers provided more accurate allele-frequency estimates than did dinucleotide markers. There was some evidence that oligonucleotides labeled with FAM (5-carboxyl fluorescein) might provide better estimates than those labeled with TET (4,7,2′,7′-tetrachloro-6-carboxyfluorescein) or HEX (4,7,2′,4′,5′,7′-hexachloro-6-carboxyfluorescein) but this should be repeated on a larger sample.

In this paper we describe the amplification of 296 STR loci for pooled DNA obtained from 6 populations located in diverse regions of the world. The reliability of this pooling method is tested for a panel of loci that were individually genotyped, and the allele frequencies are compared for two populations. The results from the application of a variety of analytical methods were compared using different arrays of loci.

Methods

Equimolar pools of 50 DNA samples from each of the six populations were constructed and amplified using a touchdown PCR program with a first annealing step of 65°C, which decreased to 55°C in five 2°C steps on both the individual and pooled samples [for a discussion of the pooling methodology, see Sham et al. (2002)]. Preliminary experiments indicated that a small number of touchdown steps reduced the appearance of spurious amplification products compared to using no touchdown steps. Additional experiments confirmed that too many touchdown steps allowed preferential amplification of some alleles with a number of markers, making individuals appear to be homozygous when in fact they were heterozygous and decreasing the apparent heterozygosity of the pool.

To test the accuracy of the use of pools of DNA for the generation of allele-frequency distributions, two pools of 50 individuals (New York City residents of European ancestry and Buryats) were genotyped with six microsatellite markers. The allele frequencies in the pools were estimated using the proportion of peak height or the area within each peak (using the Genotyper 2.5 software package of ABI), divided by the total height or area for the amplification of each marker. The frequencies of DNA fragments of tandem repeats, based on pooled samples, were then compared to the frequencies computed by gene counting methods and based on the same samples genotyped individually. Correlations among the pooled and individually genotyped allele frequencies were computed.

Analytical Methods. R matrix analysis was used to compute the genetic distances based on 296 and 30 randomly selected STR loci for a two-dimensional scaling representation (Harpending and Jenkins 1973). Genetic diversity was estimated as

$$\frac{n}{n-1} \sum x_i^2,\tag{1}$$

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. This measure of population genetic diversity was plotted against the distance from the centroid of distribution to estimate the effects of gene flow between the populations (Harpending and Ward 1982). Mantel tests were used to compare matrices of genetic distances based on 30 STR loci versus the total data set (296 STR loci) between the six populations.

Populations. DNA was pooled from six populations, each with a sample of 50 individuals: Kizhi, Buryat, Chuvash, Native Americans, South Africans, and a New York City sample of European ancestry. Two of these indigenous populations, Buryats and Kizhi, were sampled from villages located in Siberia, Russia. The Kizhi sample is from Mendur-Sokhon, a small Kizhi village located in the southern region of the Gorno Altai Autonomous Republic (McComb et al. 1996). The Buryat community, Gakhani, is located in the Ust Orda Buryat Autonomous Okrug region of Siberia, west of Lake Baikal (Mosher 2002). The Buryats are closely related genetically to the Mongol populations (Novoradovsky et al. 1993). The Chuvash sample is a Turkish-speaking population from southern Russia; they are believed to be closest genetically to Finnish populations (Melvin 2001). The Native American sample comes from a paternity-exclusion forensic database in Denver, Colorado. The South African sample consists of a mixed Bantu and Khoisan population, termed "Cape Coloreds" (Henneberg et al. 1998). The New York City population is a sample of individuals of European ancestry.

Results

The results of the correlation analyses indicate that DNA pools can provide extremely accurate allele-frequency estimates of the individual alleles in the pool. Twenty-five STR loci from six populations were individually genotyped, and their allele frequencies were compared to the frequencies based on DNA pools of the same loci. The use of peak height, as measured by GeneScan software (ABI) and the Genotyper software package, provides a slightly more accurate estimation of frequencies than the use of peak area. The correlation between allele frequencies based on individual genotyping for the 25 loci and the frequencies computed using peak height is 0.9811 versus 0.97 for peak area; the variance in measurement error resulting from pooling is less than 4%.

Table 1. *R* Matrix Analysis of STRs in Pooled DNA Samples[a]

Population	Kizhi of Gorno Altai	Buryats	Chuvash	Amerindians	New York City Europeans	South Africans	H^b
Kizhi	**0.0253**	0.0457	0.0556	0.0576	0.0693	0.1054	0.727
Buryats	0.0046	**0.0296**	0.0599	0.0619	0.0794	0.1127	0.719
Chuvash	−0.0035	−0.0035	**0.0233**	0.0562	0.0509	0.0994	0.737
Amerindians	−0.0041	−0.0041	−0.0044	**0.0241**	0.0561	0.0948	0.735
New York City Europeans	−0.0086	−0.0115	−0.0004	−0.0026	**0.0268**	0.0869	0.746
South Africans	−0.0136	−0.0151	−0.0116	−0.0089	−0.0036	**0.0529**	0.774

a. The *R* matrix is given below the diagonal; d^2 distances are given above the diagonal. The bold-
   faced numbers making up the diagonal are the r_{ii} values (distances to the centroid).
b. Computed diversity.

Table 1 summarizes the *R* matrix analysis for 296 STR loci based on pooled DNA samples from 6 populations. The R_{ST} (an estimate of population differentiation equivalent to F_{ST}; Harpending and Jenkins 1973) for the entire data set (6 populations, 296 systems, 2,203 alleles) is only 0.03, an unexpectedly low estimate of population differentiation compared to the values found by Jorde et al. (2000) (F_{ST} = 0.097) or by Barbujani et al. (1997) (F_{ST} = 0.155) in 30 STRs for worldwide studies.

The lowest and highest diversity estimates were exhibited by the Buryat (0.719) and South African (0.774) samples, respectively. The comparatively high African diversity coincides with previous estimates reported by Jorde et al. (2000). The distance from the centroid of distribution (r_{ii}) is uniform in all populations (0.0233–0.0268), except for the South African sample, which has an r_{ii} value of 0.0529, or twice that of the other populations. The South African population exhibits the highest genetic distances compared to the other five groups and is closest genetically to Europeans and differs most from the Buryats and Kizhi. The Kizhi and Buryats are closest genetically (D^2 = 0.045), followed by the genetic affinity between the Chuvash and Europeans (D^2 = 0.050).

Figure 1 is a two-dimensional scaling representation of genetic distances measured by the *R* matrix and is based on 296 loci among the 6 populations. The matrix correlation between the original D^2 distance matrix and the two-dimensional representation is *r* = 0.983. The first axis separates the South African population from all the other groups. The second axis separates the Asian from the European populations, with the mixed Native American group being intermediate.

Few differences are observed when the two-dimensional plot based on 296 loci and 2,203 alleles is compared to a plot based on 30 randomly selected repeats (see Figure 2). In both plots the South African sample is separated from all the
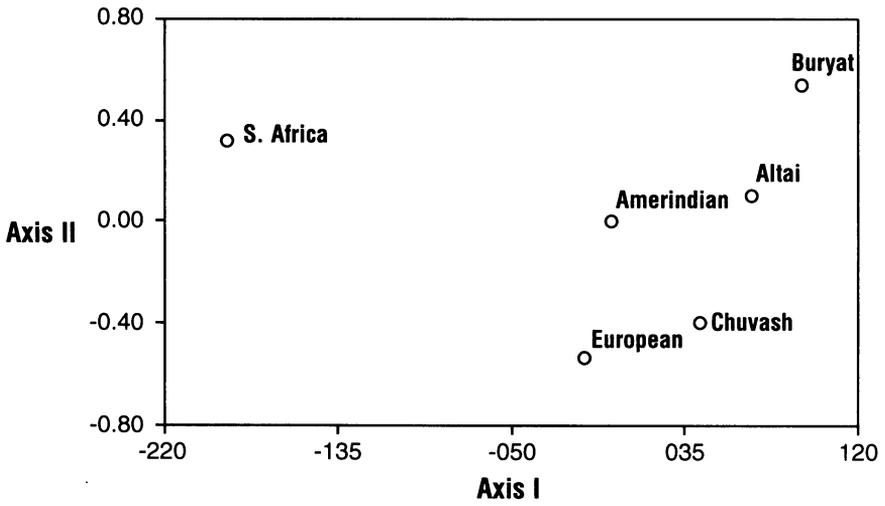
Figure 1. Two-dimensional scaling representation of genetic distances measured by R matrix and based on 296 loci and 2,203 alleles among 6 populations.
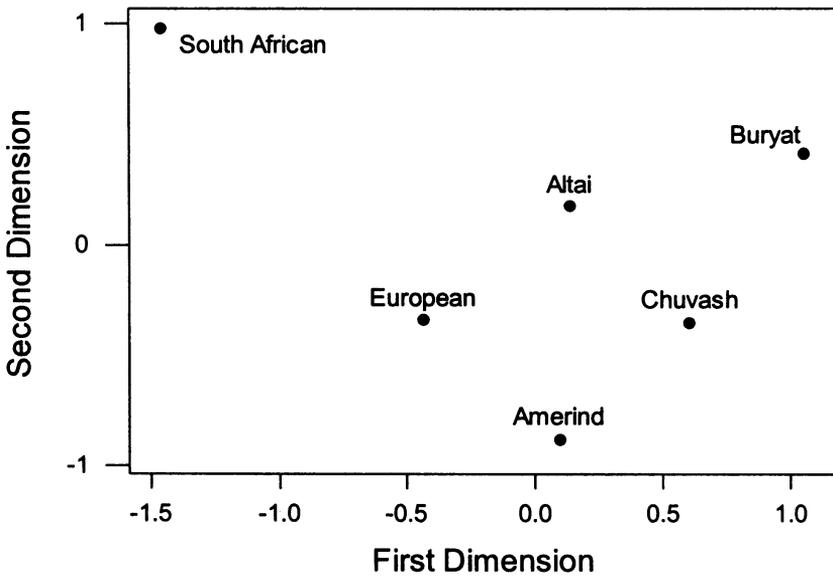
Figure 2. Two-dimensional scaling representation based on 30 randomly selected STR loci from the 6 populations characterized by DNA pools.
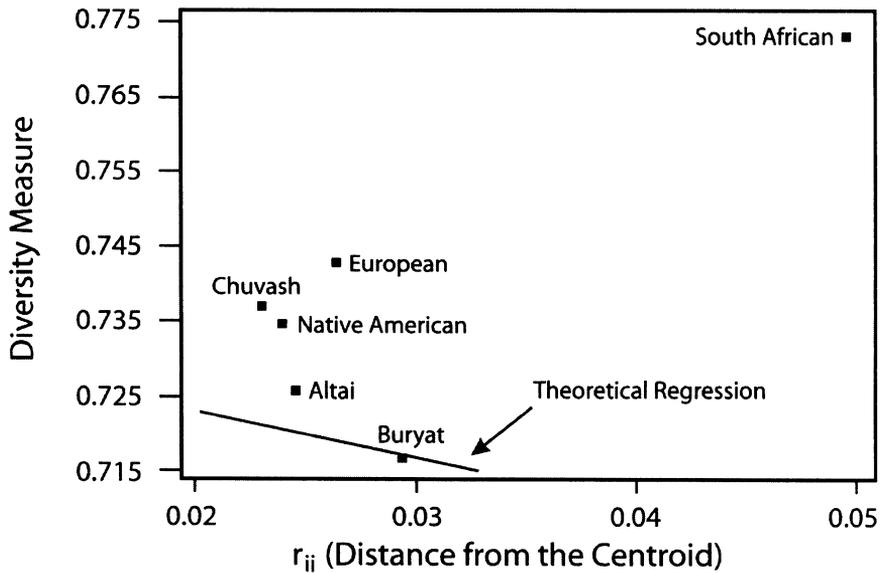
Figure 3. Plot of genetic diversity (*H*) against the distance from the centroid of distribution for 296 STR loci in 6 populations.

other populations along the first axis, whereas the Chuvash cluster with the European sample. In both plots the Buryats are closest to the Kizhi population. In the 30-loci plot the Amerindians do not cluster closely with the Buryats and Amerindians but are separated along the second axis. The two plots generally give similar patterns that reflect common ancestry with slight differences in detail. A normalized Mantel correlation between the two matrices (genetic distances between the 6 populations, 30 randomly selected STR loci versus the total data set of 296 loci) reveals a correlation of $r = 0.8295$.

The plot of heterozygosity estimates (genetic diversity) versus the distance from the centroid of distribution of each population reveals that both the South African and the New York City European populations lie outside the 95% confidence intervals of the theoretical regression line (see Figure 3). This suggests that the South African population, a mixed Bantu and Khoisan-speaking group, has by far the greatest amount of genetic diversity and the highest distance from the centroid—possibly reflecting higher gene flow. The European sample from New York City exhibits high genetic diversity but lower r_{ii} and may reflect heterogeneous origins from an assortment of European regions.

Correlation Analysis by Chromosome. Table 2 provides a breakdown of the numbers of markers, alleles, and R_{ST} values computed for each chromosome. The

Table 2. Comparison of Interpopulation Distance Matrices Computed by Individual Chromosome and R_{ST} Values by Chromosome Number

Chromosome	Number of Markers	Number of Alleles	R_{ST}	r
1	26	193	0.028	0.954
2	19	162	0.033	0.968
3	17	147	0.025	0.966
4	14	92	0.034	0.912
5	16	121	0.036	0.893
6	19	131	0.027	0.920
7	16	136	0.026	0.865
8	19	166	0.027	0.965
9	11	89	0.035	0.890
10	15	98	0.028	0.687
11	12	86	0.032	0.809
12	12	82	0.027	0.795
13	10	74	0.022	0.926
14	10	73	0.029	0.885
15	10	86	0.033	0.923
16	12	102	0.026	0.972
17	10	75	0.026	0.771
18	4	34	0.023	0.818
19	8	54	0.032	0.344
20	9	71	0.043	0.632
21	8	51	0.043	0.944
22	6	41	0.031	0.275
X	13	82	0.034	0.888
Dinucleotide markers		235	0.025	0.969
Trinucleotide markers		290	0.033	0.985
Tetranucleotide markers		1,676	0.031	0.983
Total	296	2,203	0.030	

R_{ST} values range from 0.022 for chromosome 13 to 0.043 for chromosomes 20 and 21. The r values summarized in Table 2 are the interpopulation distance matrices calculated for individual chromosomes correlated with the distance matrix for all systems combined. The r values range from 0.275 (chromosome 22) to 0.972 (chromosome 16). Although most of the chromosomes provide similar distance matrices, there are some exceptions: chromosome 19 ($r = 0.344$, ns) and chromosome 22 ($r = 0.275$, ns). The correlation magnitudes are positively correlated to the number of loci ($r = 0.501, p = 0.015$) and, principally, the number of alleles ($r = 0.540, p = 0.008$).

A separate analysis using the number of repeats at each locus (di-, tri-, and tetranucleotide repeats) revealed that the distance matrices correlate highly with the combined distance matrix constructed from all systems, irrespective of the number of systems or alleles (see Table 2). The correlations between the matrices

based on the number of repeats and the matrix based on the total ranged from 0.969 to 0.985.

Discussion

The use of pooled DNA samples provides an accurate measure of allele frequencies for the reconstruction of population phylogeny. A comparison of individually genotyped population estimates versus pooled DNA yields correlations of 0.98 ± 0.04 based on peak height and 0.97 ± 0.04 based on peak area. Both measures of the number of repeats provide highly accurate estimates of allele frequencies. One source of error in the use of pooled samples comes from a group of alleles that have low but nonzero frequencies that were assigned a frequency of 0 in the DNA pools. This measurement error can be corrected when using DNA pooling through linear regression and by accounting for error of prediction. This has the effect of raising the frequencies of alleles that are low and lowering those that are high.

The amount of genetic variation or diversity appears to be reduced in this study: $R_{ST} = 0.03$ versus 0.097 (Jorde et al. 2000) or 0.155 (Barbujani et al. 1997) compared to various worldwide distributions of populations using the equivalent F_{ST} statistic. It is unclear whether the relatively low R_{ST} values are due to the particular panels of STRs used or to the fact that the six populations being compared in this study are more similar genetically than the samples used by Jorde et al. (2000) or Barbujani et al. (1997). If forensic panels were used, there may have been a bias in favor of the more polymorphic or more variable markers compared to panels used for genomic scans for genetic linkage.

Shaw et al. (1998) observed that tetranucleotide markers provide more accurate allele-frequency estimates in pooled samples than do dinucleotide STR markers. However, in this study we constructed distance matrices based separately on di-, tri-, and tetranucleotide STRs and correlated them to matrices based on combined markers (296 loci) to determine which subset of markers provides the most accurate characterization of the gene pools of these populations. The dinucleotide loci provided a slightly lower correlation with the combined data set ($r = 0.970$) than the trinucleotide loci ($r = 0.985$) and tetranucleotide loci ($r = 0.983$) did. The statistical differences are so minor that in this data set it makes little difference if di-, tri-, or tetranucleotide loci are used to characterize the relationships between populations.

An examination of the correlations between specific chromosomes and distance matrices based on the combined databases shows considerable variation in population affinities, depending on which chromosome is used in the analysis. Chromosome 22 provides the poorest fit with the combined distance matrix ($r = 0.275$), whereas the largest chromosomes have higher correlations with the 296-loci matrix. The greater the number of loci present on a given chromosome,

the higher its correlation to the total array. Yet there are some exceptions to this observation: The small chromosome 21 (with only 8 loci used in this analysis) has a high correlation to the combined distance matrix of $r = 0.944$. Intermediate-size chromosome 10, with 15 loci, has a correlation of $r = 0.687$. In general, larger chromosomes are more informative than smaller ones with regard to population genetic affinities, and this is likely to be due to the larger number of markers and alleles assayed. The X chromosome in this study exhibits a relatively high correlation (0.888) to the combined matrix and better reflects the evolutionary history of the six populations than some of the smaller chromosomes.

This method of pooling DNA samples and estimating allele frequencies for the pools has a number of advantages and a few disadvantages in the study of population structure. Amplifying pooled DNA reduces labor and costs, but it is less informative than individual genotyping of STR markers. It is impossible to examine deviations from random mating within populations and to estimate genetic heterozygosities within the pools. When the level of genetic diversity in pooled samples is compared to other studies of widely distributed populations, it appears that pooled samples tend to underestimate diversity. On the other hand, because of the genome-wide distribution of the STRs (almost equivalent to a 10-cM scan), it is highly unlikely that linkage disequilibrium will be seen between markers, and hence the markers are effectively independent. Overall, the pooling of DNA in human populations, followed by amplification and estimation of allele frequencies, provides a useful, less costly method of providing an accurate reflection of population history and structure.

Literature Cited

Arbour, N. C., J. Zlotogora, R. G. Knowlton et al. 1997. Homozygosity mapping of achromatopsia to chromosome 2 using DNA pooling. *Hum. Mol. Genet.* 6:689–694.
Arnheim, N., C. Strange, and H. Erlich. 1985. Use of pooled DNA samples to detect linkage disequilibrium of polymorphic restriction fragments and human disease: Studies of the HLA class II loci. *Proc. Natl. Acad. Sci. USA* 82:6970–6974.
Asada, Y., D. S. Varnum, W. N. Frankel et al. 1994. A mutation in the Ter gene causing increased susceptibility to testicular teratomas maps to mouse chromosome 18. *Nat. Genet.* 6:363–368.

Bamshad, M. J., W. S. Watkins, M. E. Dixon et al. 1998. Female gene flow stratifies Hindu castes. *Nature* 395:651–652.

Barbujani, G., A. Magagni, E. Minch et al. 1997. An apportionment of human diversity. *Proc. Natl. Acad. Sci. USA* 94:4516–4519.

Barcellos, L. F., W. Klitz, L. L. Field et al. 1997. Association mapping of disease loci, by use of pooled DNA genomic screen. *Am. J. Hum. Genet.* 61:734–747.

Brennan, T. M., D. Landau, H. Shalev et al. 1998. Linkage of infantile Bartter syndrome with sensori-neural deafness to chromosome 1p. *Am. J. Hum. Genet.* 62:355–361.

Crawford, M. H. 1998. *Origins of Native Americans.* Cambridge: Cambridge University Press.

Dubreuil, P., C. Rebourg, M. Merlino et al. 1999. Evaluation of a DNA pooled-sampling strategy for estimating the RFLP diversity of maize populations. *Plant Mol. Rep.* 17:123–138.

Harpending, H., and T. Jenkins. 1973. Genetic distance among Southern African populations. In *Methods and Theories of Anthropological Genetics,* M. H. Crawford and P. L. Workman, eds. Albuquerque: University of New Mexico Press, 177–199.

Harpending, H., and R. H. Ward. 1982. Chemical systematics and human populations. In *Biochemical Aspects of Evolutionary Biology,* M. H. Nitecki, ed. Chicago: University of Chicago Press, 213–256.

Henneberg, M., G. A. Harrison, and G. Brush. 1998. The small child anthropometric and physical performance characteristics of short-for-age children growing in good and poor socio-economic conditions. *Eur. J. Clin. Nutr.* 52(4):286–291.

Jorde, L. B., A. R. Rogers, M. Bamshad et al. 1997. Microsatellite diversity and demographic history of modern humans. *Proc. Natl. Acad. Sci. USA* 94(7):3100–3103.

Jorde, L. B., W. S. Watkins, M. J. Bamshad et al. 2000. The distribution of human genetic diversity: A comparison of mitochondrial, autosomal, and Y-chromosome data. *Am. J. Hum. Genet.* 66:979–988.

McComb, J., M. H. Crawford, L. Osipova et al. 1996. DNA interpopulational variation in Siberian indigenous populations: The Mountain Altai. *Am. J. Hum. Biol.* 8:599–607.

Melvin, K. 2001. *Genetic Diversity Among the Chuvash Using Randomly Amplified Polymorphic DNA (RAPD) Markers.* Master's thesis. University of Kansas, Lawrence.

Michelmore, R. W., I. Paran, and R. V. Kesseli. 1991. Identification of markers linked to disease genes by bulked segregant analysis: A rapid method to detect markers in specific genomic regions by using segregating populations. *Proc. Natl. Acad. Sci. USA* 88:9828–9832.

Mosher, M. J. 2002. *The Genetic Architecture of Plasma Lipids in the Buryats: An Ecogenetic Approach.* Ph.D. dissertation. University of Kansas, Lawrence.

Novoradovsky, A., V. Spitsyn, R. Duggirala et al. 1993. Population genetics and structure of Buryats from the Lake Baikal region of Siberia. *Hum. Biol.* 65(5):689–710.

Rubicz, R., T. Schurr, P. Babb et al. 2003. Mitochondrial DNA and the origin of Aleuts. *Hum. Biol.* 75(6):809–835.

Schnack, H. G., S. C. Bakker, R. van't Slot et al. 2004. Accurate determination of microsatellite allele frequencies in pooled DNA samples. *Eur. J. Hum. Genet.* 12:925–934.

Sham, P., J. S. Bader, I. Craig et al. 2002. DNA pooling: A tool for large scale association studies. *Nat. Rev. Genet.* 3:862–871.

Shaw, S. H., M. M. Carrasquillo, C. Kashuk et al. 1998. Allelic frequency distributions in pooled DNA samples: Applications to mapping complex disease genes. *Genome Res.* 8:111–123.