MAX-PLANCK-INSTITUTE FOR EVOLUTIONARY ANTHROPOLOGY

Max-Planck-Institute for evolutionary Anthropology
Inselstraße 22, D-04103 Leipzig, Germany

Dr. Richard Rau
National Institute of Justice
810 Seventh Street, NW
Washington, DC 20531
USA

July, 07, 1999

Dear Dick,

Enclosed is a copy of the final report for NIJ #92-IJ-CX-K040 (S-1), "Mitochondrial DNA Variation in North American Populations of Forensic Interest."

Please let me know if you have any questions concerning this report.

Best wishes,

Mark Stoneking

Professor of Biological Anthropology

Final Report

NIJ 92-IJ-CX-K040 (S-1)

"Mitochondrial DNA Variation in North American Populations of Forensic Interest"

Mark Stoneking, P.I.

8 July 1999

Goals

The goals of this project were to produce a database of 2,000 mitochondrial DNA (mtDNA) types in North American populations, and use this database to assess the amount of variation and degree of subpopulation heterogeneity within the major racial groups (African-Americans, European-Americans, and Hispanics). These are important aspects of establishing the utility of mtDNA in forensic cases. A forensic DNA marker should exhibit high levels of variation within populations, so that it is likely that different individuals will have different DNA profiles. At the same time, it is desirable that different subpopulations should not exhibit large differences in the distribution of DNA profiles, otherwise to assess the likelihood that other individuals would have matching DNA profiles in a particular case, it will be necessary to use subpopulation-specific databases, rather than racial group-specific databases.

Results and Conclusions

Using a method we developed previously that involves detecting variation by hybridization with sequence-specific oligonucleotide (SSO) probes, we determined mtDNA types for 2,282 individuals (805 African-Americans from 10 populations, 922 European-Americans from 11 populations, and 555 Hispanics from 7 populations). The sample size, number of different mtDNA types, and amount of variation (genetic diversity) in each population is shown in Tables 1-3. In general the level of variation is high (diversity exceeds 0.95) in almost all populations. This means that mtDNA analysis is indeed likely to be informative in forensic casework, as the chance that unrelated individuals would have matching mtDNA types is less than 5%. Moreover, this is based on SSO-type determination, which captures only a fraction of the total variation; sequence analysis, which is the usual method employed in forensic mtDNA analysis, should reveal even higher levels of variation.

The amount of heterogeneity was determined by apportioning the total genetic variance for each racial group into between-population and within-population components. These results are shown in Table 4, along with their statistical significance. The between-population variance components are all significantly greater than 0 (p values are less than 0.05), indicating that there is statistically-significant heterogeneity within each major racial group. However, the magnitude of the between-population component is quite small, ranging from 1.4% for Hispanics to 0.5% for both African-Americans and European-Americans. Thus, the vast majority of the genetic variance (98.6-99.5%) is within populations, not between populations. Moreover, further inspection of the data reveals that the statistical significance of the between-population component is due primarily to a single population in each case (Californian African-Americans, Vermont European-Americans, Pennsylvanian Hispanics); removing each of these populations and repeating the analysis results in between-population components that are no longer statistically-significant (Table 4). The fact that the statistical significance of the heterogeneity within each racial group is due to a single population (which differs for each racial group) means that the statistical significance is due to random fluctuations in the mtDNA type frequencies for these single populations, and not any systematic deviations. This result, combined with the very small magnitude of the between-population component of the total genetic variance, means that the observed statistically-significant heterogeneity is not biologically significant, and hence it is appropriate to use mtDNA type databases for each racial group that combine all of the populations together.

Products

A manuscript describing the results of this study is being prepared for submission to the Journal of Forensic Sciences; copies of the published study will be sent to the NIJ administrators. The mtDNA SSO-type databases will be made publicly-available via the Internet.

Table 1. Sample size, number of mtDNA types, and genetic diversity for African-American populations.

Population	Sample Size	Number of Types	Diversity (± 1 S.E.)
California	119	62	0.975 ± 0.006
Illinois	68	55	0.993 ± 0.004
Louisiana	55	36	0.968 ± 0.012
Maryland	38	27	0.980 ± 0.011
Missouri	126	64	0.977 ± 0.005
Oregon	86	55	0.986 ± 0.004
Pennsylvania	93	64	0.988 ± 0.004
Texas	85	51	0.981 ± 0.005
Virginia	86	58	0.984 ± 0.006
Washington	49	33	0.974 ± 0.011

Table 2. Sample size, number of mtDNA types, and genetic diversity for European-American populations.

Population	Sample Size	Number of Types	Diversity (± 1 S.E.)
California	128	72	0.972 ± 0.008
Cajun (Louisiana)	58	30	0.953 ± 0.015
Illinois	42	30	0.974 ± 0.014
Louisiana	57	35	0.948 ± 0.021
Maryland	38	26	0.962 ± 0.021
Missouri	90	51	0.970 ± 0.009
Oregon	98	54	0.973 ± 0.007
Pennsylvania	105	53	0.958 ± 0.012
Texas	103	56	0.968 ± 0.010
Virginia	109	56	0.974 ± 0.006
Vermont	94	47	0.922 ± 0.023

Table 3. Sample size, number of mtDNA types, and genetic diversity for Hispanic populations.

Population	Sample Size	Number of Types	Diversity (± 1 S.E.)
Louisiana	58	27	0.929 ± 0.021
Mexico City	94	38	0.938 ± 0.009
Oregon	74	45	0.959 ± 0.014
Pennsylvania	115	40	0.925 ± 0.014
Texas	60	34	0.970 ± 0.009
Virginia	102	52	0.967 ± 0.006
Washington	52	36	0.978 ± 0.009

Table 4. Genetic variance components for the major racial groups.

Racial Group	Variance Component (% of Total) Within Populations	Between Populations	p-value
African-Americans	99.5	0.5	0.02
(-California)	99.7	0.3	0.10
European-Americans	99.5	0.5	<0.01
(-Vermont)	99.8	0.2	0.15
Hispanics	98.6	1.4	<0.01
(-Pennsylvania)	99.4	0.6	0.06