AEROSPACE REPORT NO.
ATR-79(7910)-1

FORENSIC IMPLICATIONS OF GENETIC POPULATION DATA COLLECTED IN DIFFERENT GEOGRAPHICAL REGIONS

FINAL REPORT
November 1978

Prepared for

The Law Enforcement Assistance Administration
U.S. DEPARTMENT OF JUSTICE

Prepared by

Eastern Technical Division
THE AEROSPACE CORPORATION

AEROSPACE REPORT NO.
ATR-79(7910)-1

FORENSIC IMPLICATIONS OF GENETIC POPULATION DATA COLLECTED IN DIFFERENT GEOGRAPHICAL REGIONS

FINAL REPORT

November 1978

Prepared for

The Law Enforcement Assistance Administration
U.S. DEPARTMENT OF JUSTICE

Prepared by

Eastern Technical Division
THE AEROSPACE CORPORATION

This project was supported by Contract Number J-LEAA-025-73 awarded by the Law Enforcement Assistance Administration, U.S. Department of Justice, under the Omnibus Crime Control and Safe Streets Act of 1968, as amended. Points of view or opinions stated in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Justice.

REPORT NO.
ATR-79(7910)-1

FORENSIC IMPLICATIONS OF GENETIC POPULATION DATA COLLECTED IN DIFFERENT GEOGRAPHICAL REGIONS

Prepared by

Robert Shaler
Group Leader
Explosives Control and
Forensics Directorate

Approved

Robert B. Moler
Director
Explosives Control and
Forensics Directorate

Joseph Meltzer
General Manager
Eastern Technical Division

ACKNOWLEDGMENTS

The author would like to acknowledge the support of several individuals whose efforts have made the completion of this work possible.

The genetic data were provided by Gary Siglar, Los Angeles County Sheriff's Office; Carl Stuver, Dade County Crime Laboratory; and Mark Stolorow, Michigan State Police Crime Laboratory.

Statistical analyses and interpretation (Chapters II and III) were provided by Dr. Peter Bright and computer support was provided by the University of California, Los Angeles, and Dr. Joseph Wertz, The Aerospace Corporation.

Significant consulting guidance was given by Dr. George Sensabaugh, University of California, Berkeley; Dr. Peter Bright and Carl Stuver, Dade County Crime Laboratory; and Mark Stolorow, Michigan State Police Crime Laboratory.

Genevieve Denault, The Aerospace Corporation, deserves special credit for her continued enthusiastic support and guidance throughout the program from its inception, through management changes, to its completion.

Dr. Quon Y. Kwan and Dr. Robert B. Moler made contributions in reviewing and revising portions of the text.

EXECUTIVE SUMMARY

The significance and usefulness of genetic markers to type blood can only be determined based on a knowledge of the frequency distribution of the genetic markers. The objectives of this study were to collect sufficient blood group frequency distribution data to determine whether variations exist: (1) among racial populations in the United States, (2) among geographic regions in the United States, and (3) between United States and British Caucasian populations. Originally, four representative areas--New York City, Miami, Detroit, and California--were selected as subject areas with the expectation that they would provide a combined total sample size of 12,700 individuals. However, unforeseen circumstances caused a substantial reduction in the sample size as well as in number of areas represented. Even with the contribution of unanticipated data, a sample consisting of only 2871 individuals was finally obtained. The consequences of these were reductions in the sensitivity with which phenotypic differences in populations could be determined and in the accuracy of the statistical analyses of the data. In spite of these limitations, the study still endeavored to carry out the objectives, albeit with a doubling of the statistical error in the results.

The institutions that contributed frequency distribution data to this study were Michigan State Police Crime Laboratory, Northville, Detroit; Dade County Department of Public Safety Crime Laboratory, Miami; Los Angeles County Sheriff's Department Crime Laboratory; University of California at Berkeley, Forensic Science Group; and The Aerospace Corporation. The sampling bases for the first two laboratories were essentially blood bank populations--blood donors. For the data, samplings were obtained from specific proportions of ethnic subpopulations of Orientals, Caucasians, Blacks, and Hispanics so that statistically valid

phenotypic frequencies as a function of ethnic groups could be derived. The sampling basis for the third laboratory was a forensic population--the set of all people, either suspects or victims, directly involved in a criminal investigation. The last two institutions sampled another type of restricted population--volunteers.

The specific information encoded onto standardized data sheets completed for each sample included: contributing laboratory, analysis date, gender, ethnic origin, birth place, names of persons completing the sheet and entering the data on the computer, and the results of blood phenotyping for three antigenic and eight or more enzyme/protein genetic marker systems. The antigenic systems tested for were the ABO (including subtypes of A), MNSs, and Rh groups; the enzyme/protein polymorphic systems tested for were adenosine deaminase (ADA), adenylate kinase (AK), erythrocyte acid phosphatase (EAP), esterase-D (EsD), glucose-6-phosphate dehydrogenase (G-6-PD), group-specific component (Gc), haptoglobin (Hp), and phosphoglucomutase-locus one (PGM_1). In some cases, additional enzyme/protein systems were tested: carbonic anhydrase-locus II (CA II), peptidase-A (Pep-A), and 6-phosphogluconate dehydrogenase (6-PGD). All of the samples were analyzed in the form of wet blood or "fresh" bloodstains. The methods used to analyze blood samples consisted of conventional hemagglutination or electrophoreses. In a few cases, a number of laboratories contributing the data did not perform the actual analyses themselves, especially for the ABO and MNSs antigens, but relied on the typing results of blood banks.

During the course of collecting blood phenotype frequencies from analyzed blood samples, an effort was made to check and control sources of errors. Errors in analyzing blood samples were checked by a referee laboratory and by exchanging samples between laboratories. Errors in transcribing the phenotypes from the laboratory notebooks into the computer

(there were two modes of data entry--interactive remote and keypunching) were checked by comparison of the notebook and entered values for 10 cases selected at random for each contributing laboratory. Discrepancies encountered were counted and corrected. An approximation for P, the probability of the average phenotype entry being erroneous, was calculated as being within the domain of 5.7×10^{-4} to 3.6×10^{-3}. These figures imply that errors due to transcription are negligible for all practical purposes.

Various statistical analyses were performed on the blood phenotype frequency data that were collected. The first statistical analysis was carried out to determine the homogeneity of phenotypic frequency over ethnicity while maintaining geographic location and sex constant. (Tests for significance were based on the Pearson chi square and Fisher Exact criteria.) As could be expected, phenotypic frequencies are not homogeneous but vary significantly from one ethnic group to another. The second statistical analysis was carried out (using the likelihood ratio chi square and Pearson chi square tests) to test whether the factors of sex, genetic marker, and geographic location are independent while maintaining ethnicity constant. The results show that for most genetic markers among Caucasians and Blacks, these factors cannot be considered to be independent. This means, for example, that the probability of being male, blood type A, in Detroit is not simply the product of the probabilities of being male times that of being blood type A, times that of being in Detroit, for either a Caucasian or Black.

In view of these results, the third statistical analysis was conducted to determine which of the interactions (i.e., between sex and genetic marker, sex and geographic location and genetic marker and geographic location) was significant. It was found (using likelihood ratio chi square and Pearson chi square statistics) that there is no interaction between sex and geographic location and between sex and

genetic marker, except for the genetic marker, G-6-PD, which is already known to correlate with sex. Under the assumption that sex and geographic location and sex and genetic marker do not interact (i.e., are independent), the interaction between genetic marker and geographic location was tested (using the likelihood ratio chi square and Pearson chi square tests). It was found that with few exceptions, there is some interaction between genetic marker and geographic location. This implies that phenotypic frequencies are affected by geographic location and signifies that it is inappropriate to take the phenotypic frequencies for a given area and apply them to a different area.

Because of the surprising and unexpected nature of this result and its discomforting impact on forensic serology, two attempts were made to verify it. One involved a further statistical analysis and a second involved purely phenotypic frequency data comparisons for different cities without statistical analysis; both tend to confirm this result.

In conclusion, there are variations in phenotypic frequency distributions of blood groups: (1) among racial populations in the United States, (2) among different geographic regions in the United States, and (3) between United States and British Caucasian populations. However, these conclusions are only tentative, and they must be qualified as such because the inadequacy of the sample size precludes a positive statement as to whether these variations can be regarded as significant. The nature of these results would suggest that a more detailed study involving a greatly expanded sample set is urgently needed.

CONTENTS

ILLUSTRATIONS

INTRODUCTION

A. OBJECTIVES

1. Background

The ability to connect a bloodstain at a crime scene with either a suspect or a victim is of immense value in criminal investigation. In principle, it is possible to identify a bloodstain as originating from a specific individual from the entire world population on the basis of existing genetic marker systems. A genetic marker system is a set of mutually exclusive characteristics, known as phenotypes, of which only one is inherited and expressed by an individual. There are two types of genetic marker systems: (1) antigenic and (2) isoenzyme or enzyme/protein polymorphic systems. To use genetic markers to their fullest extent for purposes of criminal investigation requires that the frequency-of-occurrence (or genetic population) data on each analyzed phenotype be known. In January 1973, The Aerospace Corporation in support of the National Institute of Law Enforcement and Criminal Justice (NILECJ) of the Law Enforcement Assistance Administration (LEAA) initiated a computerized data file for collecting and compiling frequency-of-occurrence data on genetic markers as a function of ethnic and geographic origin. The file includes data from crime laboratory casework analyses and data procured under subcontracts with crime laboratories located in designated cities. The file is intended to be a continuing program to which crime laboratories, blood banks, and population geneticists can add new data as obtained and retrieve statistical information as needed.

2. Purpose

The purpose of this study was to obtain sufficient genetic marker distribution frequencies, from geographic regions, that were representative of the country's population sectors, in order to evaluate the degree of

consistency of these data: (1) among the United States racial populations, (2) among various regions in the United States, and (3) between United States Caucasian and British Caucasian populations. The basic concept was to test the validity of the assumption that frequency-of-occurrence data derived from other geographic regions (or populations) are applicable to the geographic region (or population) at hand. Faced with a lack of frequency-of-occurrence data on phenotypic distributions in the United States, criminalists have routinely taken Cullifcrd's data[1] from Britain and applied them here, presuming that the populations are equivalent. Secondly, a very limited number of crime laboratories within the United States are beginning to collect their own frequency-of-occurrence data, and crime laboratories elsewhere, which do not possess the resources to collect such data, are taking these data and applying them to their own jurisdictions under the same presumption. The need for this study is quite apparent considering the widely varying nature of the racial mix within the United States.

3. Limitations

At the onset of this study, representative areas of the United States were selected to be sampled for frequency-of-occurrence data: New York City, Miami, and Detroit. California, a major population center, was intentionally omitted from this study because Dr. Benjamin Grunbaum had received a grant from that State's criminal justice planning agency to undertake a frequency-of-occurrence data collection study for the State and had offered to cooperate by providing the California data to Aerospace.

Based on a statistical analysis, it was concluded that a sample size of 12,700 individuals including the California data, was the minimum total size that would provide statistically acceptable frequency-of-occurrence data. The planned sample sizes were: New York City--1200 individuals; Miami--900; and

Detroit--600. The California frequency-of-occurrence data sampling of 10,000 individuals was expected to increase the total size to 12,700. However, as the program progressed, collection of frequency-of-occurrence data in the New York area was not carried out because of budgetary limitations and the high cost proposed for this study. In addition, Aerospace did not receive any of the frequency-of-occurrence data collected in Dr. Grunbaum's study. Consequently, only a substantially reduced sample size and reduced number of areas were possible. Even with frequency-of-occurrence data contributed by institutions (Los Angeles County Sheriff's Crime Laboratory) not originally anticipated, a sample size total of only 2871 individuals was achieved, one-fourth of the expected sample. Because the reduction in sample size introduced a doubling of the statistical error in the results, the conclusions based on this sample size must be qualified as tentative and interpreted with caution.

B. POPULATIONS

The definition of a typical population used for this study presented an unexpected problem. Within the jurisdiction of any crime laboratory is a population represented by many ethnic mixes. For example, Los Angeles County (excluding Los Angeles City and Long Beach, both not served by the Los Angeles County Sheriff's Office Laboratory) has a population of approximately 3.85 million composed of 71.5-percent Caucasian, 6.2-percent Black, 1.5-percent Oriental, 19.3-percent Mexican, etc.[2] This sampling population of interest to that laboratory is called the "jurisdictional population." Ideally, a genetic population study should include an appropriate number of each ethnic group so that statistically valid phenotypic frequencies can be derived. The problem of population selection for the crime laboratory is not solved by using the regional ethnic mix; the population, as defined by

census takers, may not represent the population mix which is routinely observed by law enforcement personnel, which for the purposes of this discussion will be called the "forensic population." It is this latter population that is the most important to the crime laboratory.

Another problem involves population changes and the need for a continuing reappraisal of population components. From a pragmatic point of view, it may not be adequate simply to perform one genetic population study and hold the results indefinitely; it is well known that both the forensic and total populations change.

Integral to the changing of population components is the problem of changing phenotypic frequencies within a given population subcomponent. For example, that the Caucasian population in Los Angeles County changes 50 percent between censuses is not as important as a change of 50 percent in the phenotypic frequencies used for genetic markers of interest within that Caucasian component.

Unfortunately, phenotypic information cannot be determined from a one-time survey such as has been conducted in this program. Ideally, each laboratory should have an ongoing data collection program in order to recognize population changes within its jurisdiction. It is possible to relieve the time and expense of periodic population studies in two ways. The first is to have a recordkeeping system that will retrieve and record statistical information as it is developed during casework. The second method is to assume that the phenotypic frequencies of certain ethnic groups as discussed above do not change. Although not proven, this is assumed by forensic organizations throughout the world. Thus, the burden on the crime laboratory is to stay informed of the changing population components within its jurisdictional area. It is expected that a combination of these two methods would prove most valuable to laboratories. By continually deriving statistics from casework, each laboratory can keep abreast of its forensic population, and, by reviewing

census statistics, it can be aware of population changes occurring within its jurisdiction.

In this particular study, different populations were sampled by different participating laboratories. Laboratories No. 1 and 2 did not have a population base considered large enough for statistical analysis. Laboratory No. 3 used a forensic population, and Laboratories No. 4 and 5 used blood bank populations. These latter populations may not necessarily be relevent to either the forensic or jurisdictional populations discussed above. In these instances, however, it was simpler and quicker to gather blood data from blood bank sampling than by criminal investigation casework sampling.

C. GENETIC CONSIDERATIONS OF COLLECTED DATA

In genetic data collection, several observations, both practical and philosophical, can be made. Practically speaking, it is of major importance that the interpretation of the data obtained provide valuable information to a criminal investigation. For example, genetic marker analysis of blood can provide information regarding the presence of certain phenotypes which may be characteristic of certain ethnic populations. Thus, a logical first investigative step is to analyze for the presence of these phenotypes. (Appendix A shows the electrophoretic results for enzyme and protein phenotypes.) If, after completion of such analysis, phenotypes of 2-1 of peptidase A, 2 of carbonic anhydrase II, and AS of hemoglobin were found, which are characteristic phenotypes of the Black population, the indication would be that the blood sample was from a Black. Depending upon the investigational circumstance, this may be valuable information in providing a clue or lead to a particular suspect.

It must be stressed that this biochemically derived information is extremely selective in nature. Although the phenotypes recorded in the

hypothetical example are characteristic of the Black population, the genetic makeup of this ethnic group is more inclusive. In fact, the individual perpetrator of the crime may, by all visible criteria, appear Caucasian, and the inclusion of these three "Black" genetic phenotypes may have resulted from a mixed marriage at some point in his ancestry.

In the phenotypic frequency data collected during this study, several examples of this type of inconsistency were noted. When donors were checked into the blood bank, the clerk (nurse) recorded, in addition to other information, their ethnic origins which were determined in several ways. In some cases, the donor provided the information about race during an interview. For Hispanics, the last name, birthplace, and the ability of an individual to speak Spanish was used to confirm Hispanic origin. Skin color and visible anatomical characteristics were used for determining the racial origin, e.g., Black, White, Oriental, etc. Thus, that a blood sample classified as Caucasian contains genetic phenotypes that are traditionally characteristic of another ethnic background is not surprising because of the subjective judgment used to classify race.

The philosophical aspect of this problem is whether or not to include these samples in the overall data base. Exclusion of these data will permit a more accurate or purified ethnic versus genetic marker set of statistics to be derived and evaluated. It can be reasoned, however, that this would not be a valid representation of the population sample and would not be useful to the forensic analyst or police investigator. Bloodstains from either unknown or known sources do not tend to originate solely from persons who are genetically pure or solely from those of mixed ancestry. When suspects or victims are racially classified, it is usually based on the subjective criteria discussed. Subsequent blood analysis that reveals genetic

components appearing to be at odds with the subjective classification should not be a source of alarm or controversy, but rather, an additional point of uniqueness of the individual. The number of such data compared to the whole of this study will not alter the statistics in a significant way, and their inclusion will be a reminder that all is not black or white in forensic genetic analysis.

The number of data sheets for each ethnic group contributed by each of the five laboratories is shown in Table 1-1. In specific instances, the classification of Hispanics became fairly complicated, and a more detailed system was employed. The Miami Hispanic populations were coded as double entries, e.g., Caucasian-Hispanic, Black-Hispanic, or Oriental-Hispanic.

Several other special situations arose during the study. In one laboratory, eight pairs of homozygotic twins and three family studies were examined. One of each of the twins was included and the other deleted from the file because of the intention of the sampling to be as random as possible. It was not feasible to delete all members of a family except for one.

D. DATA COLLECTION

The data collected for each person were transcribed from the laboratory notes to the blood data forms (Appendix B), and the information obtained is shown below. The data base contains three separate sets of data organized in a manner dependent upon the sophistication of the data management system, the available storage hardware, and the application logic. The data base content included census frequency data and genetic phenotype frequency data as outlined below:

- Census Frequency Data--Includes population distribution statistics defined in terms of: country, state (province), county, city; ethnic group, subgroup, tribe; sex; etc.

Table 1-1. Number of Coding Sheets From Participating Laboratories

Racial/Ethnic Origin	Participating Laboratory Number				
	Aerospace 1	Berkeley 2	Los Angeles 3	Miami 4	Detroit 5
Missing	0	0	256	9	
Caucasian	10	58	239	359	507
American Indian			4	3	
Oriental	2	4	6	6	
Black			104	334	507
Other			95	158	
Hispanic			11	0	
Black-Hispanic				12	
White-Hispanic				187	
Oriental-Hispanic					
Total	12	62	715	1068	1014

- Genetic Phenotype Frequency Data.
- Forensically Useful Genetic Markers (Phenotypes)--Includes blood antigens, enzymes, proteins, etc.

The information in this data base was maintained in a disaggregated form with each set of entries qualified by a reference to the reporting laboratory. Data entries also contained the name of the laboratory generating the information, the name of the person entering the data on the computer, and the date of entry.

The three antigenic systems for which sizeable amounts of data were collected were the ABO with A subgroups, Rh, and MNSs. For each of these systems, the absence of a mark in any result indicated that the test was inconclusive or not done. In the ABO system, one of the four phenotypes O, A, B, or AB was recorded when the system was tested conclusively. The A subgroups, A_1 and A_2, were tested for when the subject was either type A or AB. Other subgroups of A were not tested. In the Rh system, every sample tested was either D or D^u. When D was not confirmed, D^u was recorded. The four antigens, C, c, E, e, were always tested, but C^W was not. For the MN system, inconclusive tests were reported as missing data.

For the 13 enzyme/protein systems with data, the testing procedure was such that if the test was run conclusively, all alternate possible responses were examined. The absence of any phenotype was construed as a test not done, and the absence of a particular phenotype when another was present was construed as a negative result. For example, if for the adenylate kinase (AK) system, with three major phenotypes, no results were obtained, it was recorded that the test was not done. However, if the 2-1 and 1 phenotypes were not present but the 2 was, the former two were considered as negative. The original data collection forms were revised to accommodate additional phenotypes for the erythrocyte acid phosphatase and group specific component systems.

II. ERROR ANALYSES

There are many biochemical steps between the transcription by ribonucleic acid (RNA) of the genetic code and in the translation of RNA into proteins or enzymes. There are also many steps from the serological or electrophoretic stages of analysis to inputting data to the computer base where errors can be introduced at each step. Estimating the accuracy of phenotypic frequencies could depend markedly on an estimate of the errors. The errors discussed in this section are determined through replication and thus pertain to precision and not to accuracy (i.e., the random error and not the systematic error). The error sources are classed as biological, quality control, data transcription, and sample-size reduction.

A. BIOLOGICAL

The first source of error is biological variation. Although the genetic determinant for these marker systems does not change for any individual, at times extragenic factors such as medication may change the test results.[3] Blood bank donors are screened in order to determine the possible use of medication. Although duplicate samples should be taken from the same donor at different times under different conditions in sufficient number to give meaningful estimates of the error, adequate duplication was not economically feasible during the study.

B. QUALITY CONTROL

During the course of this work, different quality control procedures were used. The purpose of quality control checks was to ensure that the results obtained were correct and that all participants agreed on the interpretation of results on identical analysis of identical phenotypes. The procedures used by Laboratories 4 and 5 are given in Appendix C. Comments on Laboratory Procedures are given in Appendix D.

There are several methods that can be used to accomplish this. One

is to have the participating laboratory send the analytical samples to an outside or referee laboratory for verification of results. In one instance, this was done by having the participating laboratory perform analyses on 12 blood samples in 9 different genetic marker systems. The blood samples were sent to a referee university laboratory for verification. When the results were returned, eight discrepancies were found in five of the nine genetic marker systems. At that point, it was impossible to determine which laboratory performed the tests correctly. The participating laboratory repeated the analyses, obtained the original results, and sent additional samples to a second referee university laboratory that confirmed the results of the participating laboratory. The data illustrated in Table 2-1 show the discrepancies that occurred between the laboratories.

The results of this quality control test suggest that the first referee laboratory may not have been proficient in performing simple antigenic testing though this laboratory was known to be competent in electrophoretic genetic marker analyses; it was not possible to know this in advance, because the first referee laboratory did not make known its weaknesses until the discrepancies were found.

A second way of testing the laboratories to ensure consistency in interpretation is to exchange samples. Each laboratory analyzed 10 samples and sent them to another laboratory for confirmation. The results of this testing are illustrated in Tables 2-2 and 2-3. These results showed that the consistency between the program's participating laboratories was not acceptable. However, it was found that one of the participating laboratories was relying on results of the ABO, MNSs, and Rh typing done by an outside clinical laboratory before the same blood samples were received and analyzed for other genetic marker systems by that particular

Table 2-1. Interlaboratory Discrepancies in Interpretation

Genetic Marker System	Laboratory Sample No.	Laboratory Results		
		Participating Laboratory	Referee Laboratory Number 1	Referee Laboratory Number 2
ABO	3	A_2	A_1	A_2
MNSs	6	MNSs	MNs	MNSs
	8	NSs	Ns	NSs
Fy	3	a+b+	a-b+	a+b+
	6	a+b+	a-b+	a+b+
P	6	1	2	1
Jk (a)	6	+	-	+
	8	+	-	+

Table 2-2. Interlaboratory Discrepancies
Antigenic Genetic Markers

A. ABO System

REPORTED \ Actual	O	A_1	A_2	B	A_1B	A_2B
O	19					
A_1	1	10	1			
A_2			1			
B				4		
A_1B					0	
A_2B						2

B. MN System

REPORTED \ Actual	M	N	MN
M	11		1
N	1	4	
MN			21

C. Ss System

REPORTED \ Actual	S	s	Ss
S	5		
s		22	1
Ss	1		9

D. Rh - Cc

REPORTED \ Actual	C	c	Cc
C	2		
c		30	
Cc			6

E. Rh - Ee

REPORTED \ Actual	E	e	Ee
E	1		
e		28	
Ee			10

F. Rh - D, Ø, D^u

REPORTED \ Actual	D	Ø	D^u
D	26		
Ø		12	
D^u			0

Table 2-3. Interlaboratory Discrepancies - Isoenzyme/ Polymorphic Protein Genetic Markers

A. ADA

REPORTED \ Actual	1	2-1	2
1	37		
2-1	1	0	
2			0

B. AK

REPORTED \ Actual	1	2-1	2
1	38		
2-1		0	
2			0

C. EAP

REPORTED \ Actual	A	B	BA	CB
A	3			
B		22		
BA	1		10	
CB				2

D. EsD

REPORTED \ Actual	1	2-1	2
1	30		
2-1		2	
2			6

E. G-6-PD*

REPORTED \ Actual	A	B	AB
A	8		
B		28	
AB			0

F. PGM

REPORTED \ Actual	1	2-1	2
1	26		
2-1		10	1
2			1

* One not called.

Table 2-3. Interlaboratory Discrepancies - Isoenzyme/Polymorphic Protein Genetic Markers (Continued)

G. Hp*

REPORTED \ Actual	1	2-1	2	2-1M**
1	6			
2-1		13		
2		1	14	
2-1M				0

* One not called

** One 2-1M actual was changed to 2-1 by consensus

participating laboratory. It was found that the outside clinical laboratory was not washing the red cells completely prior to performing their analyses, especially important for the MN system.

The one discrepancy that occurred in the haptoglobin system (Hp) was resolved by agreeing it to be a 2-1M. The original interpretation of the sample as a 2-1 by Laboratory No. 2 did involve the question of whether it was a 2-1 or 2-1M.

In the group specific component (Gc) system, the staff at Laboratory No. 1 had never seen a 1-Y variant. Thus, this error became an educational experience and necessitated rerunning the previous Gc1 phenotypes obtained in order to check for 1-Y variants that had been missed.

C. DATA TRANSCRIPTION

Data were transcribed from coding sheets (Appendix B) onto computer punch cards by two different procedures. The first was an interactive remote entry procedure for entering data onto the Aerospace computer from a terminal originally developed with the intention of becoming an ongoing data bank collecting data from many different locations. Data were screened before entry. Interactive remote entry was discontinued when time became an important programmatic consideration, and data from the remaining individuals were entered via keypunching onto the University of California, Los Angeles (UCLA), computer. Entry errors were discovered in both procedures.

The different types of transcription errors between the laboratory notebook and the computer printout require definition. Coding sheet errors refer to transcription errors between laboratory notes and the coding sheet. Data entry errors refer to those occurring between the coding sheet and the printout of the computer file.

To estimate the size of the data transcription error, 10 case numbers were randomly selected for each participating laboratory. These were compared to the computer printouts which were provided and discrepancies noted. Although the cases were not completely random, and hence did not permit each possible combination of errors to have the same probability of being included, each individual entry within each laboratory had the same probability of being included. It was decided that this sampling procedure was sufficiently random for the purposes of estimating an overall error rate and had the advantage over a "completely" random sample in that each laboratory was ensured of the same size sample as a basis for intralaboratory estimates. Selection of individual entries randomly would not appreciably add to the validity while greatly increasing the amount of work. After encoding the data in a sequential file of card images in the computer, each column was screened for correctness of the entry. Incorrect entries were found and rectified. In columns where no entries were expected, a printout of the corresponding laboratory and facility identification was made. These were compared with the coding sheets to see if the marks had been intended for a neighboring marker system but were misplaced. All the data sheets were reviewed by a person understanding both the genetic and statistical aspects of the study, including the written comments. Over 640 changes were made after entering the 2871 original coding sheets into the computer.

The raw data for the 10 randomly selected cases from each laboratory were compared to those in the printout from the computer files as a check on transcription errors. Furthermore, a preliminary frequency count of entries for each system, overall and broken down by phenotypic and ethnic origin, was mailed to each laboratory with a listing of their data for visual inspection to ensure that the ratios agreed with their experience. In each case, the discrepancies were corrected.

In the interactive remote entry procedure, a sizeable number of errors were discovered; the greatest source of error was omission of data by excluding information contained in the comments section.

In the keypunch and verification procedure, only one direct transcription error was discovered. This involved a handwritten "7" which both keypunch operator and verifier interpreted as a "9." In addition, the keypunch operator was asked to take the data from the comment section on the glyoxalase (GLO)-I system and record it in the position for GR (Appendix B). This was not done in 25 cases and had to be entered later. Dates of birth were transcribed correctly.

No data entry errors were discovered for the 10 randomly selected cases, but errors were uncovered by the screening procedures. For example, test results were recorded for tests the laboratory claimed not to have run; it was found such marks were misplaced on the coding sheet.

From the random sample drawn from the data bank to check transcription errors, an upper limit may be computed using the relation $P = (1 - \alpha)^{1/N}$. P is defined as the average probability that a phenotype entry is erroneous; α is the confidence level (95 and 99 percent); and N is the number of entries. Using a lower value of N as the number of phenotypes gives a larger more conservative value for P. Because the number of entries for each phenotype differs, the value of P for each phenotype would be expected to differ. The P calculated here is an average over the phenotypes considered.

Table 2-4 shows the number of marker systems and the number of data transcriptions entered for each laboratory. The overall upper limit for the 95- and 99-percent confidence limits, respectively, are calculated as follows:

$$P \leq (1 - 0.05)^{1/590} = 5.0 \times 10^{-3}$$
$$P \leq (1 - 0.01)^{1/590} = 7.7 \times 10^{-3}$$

Table 2-4. Number of Data Entries Checked by Random Sample

Laboratory Number	Number of Marker Systems	Number of Transcriptions
1	15	150
2	13	130
3	6	60
4	10	90
5	16	160

On the basis of the randomly selected individuals, there is less than a 5-percent chance that the probability of the average phenotype entry being in error exceeds 0.005, and less than a 1-percent chance that it exceeds 0.0077.

Because some data entry errors were in fact found, a lower limit can be given for P which is greater than zero. This analysis is somewhat more complex. The errors due to omission of information in the comment section are of a different type and will not be discussed here. The effect will be to lower the estimate of the lower limit. Because it is ineffable that all the errors were found in the screening process, any estimate based on these data will be lower than the actual rate.

It was possible for more than one box to be checked for several marker systems during the study, namely in the Rh and MNSs systems; they could have a maximum of five and four boxes checked, respectively. Each box checked was regarded as a single data entry. Table 2-5 gives the number of boxes checked per coding sheet for each laboratory, the number of coding sheets processed by remote entry or by keypunch, and the value for the maximum number of data entries. Although numerous errors were found for the remote entry process, their number was not recorded separately from the changes made in the coding sheets.

For a comparison of the two procedures for data entry, however, note that of the 144 changes made in the data bank which could not be ascribed to omissions of data from the original coding sheet or known systematic technician error (e.g., the Gc data were thrown out for one laboratory), there were four for the data entered by keypunching (1 keypunch error and 3 changes to original coding sheets) as shown in Table 2-6.

Thus, a lower limit can be found for data entry errors caused by keypunching. The null hypothesis that P is less than P_o is rejected. The probability that zero errors occurred in N = 6979 trials would then be

Table 2-5. Maximum Number of Data Entries

Laboratory Number	Maximum Number of Boxes Checked	Number of Coding Sheets		Maximum Number of Data Entries	
		Remote Entry	Keypunch	Remote Entry	Keypunch
1	24	12	0	288	0
2	14	62	0	868	0
3	17	574	141	9,758	2,397
4	10	637	451	6,370	4,510
5	24	886	3	21,264	72
Total		2,171	595	38,548	6,979

$(1 - P)^{6979}$. If this occurred over $(1 - \alpha)$ (e.g., 95 percent of the time), errors would occur less than α (e.g., 5 percent of the time) in rejecting lower values of P when one or more errors occurred. In other words, with a critical region of 1 or more errors, the hypothesis that P is less than:

$$1 - (1-\alpha)^{1/N} = \frac{1}{N}\alpha - \frac{1}{2}\left(\frac{1}{N}\right)\left(\frac{1}{N} - 1\right)\alpha^2 + \frac{1}{6N}\left(\frac{1}{N} - 1\right)\left(\frac{1}{N} - 2\right)\alpha^3 - \text{etc.}$$

is rejected in favor of the hypothesis that $P>P_o$. (Note: Because N is large and α is small, the first few terms of the binomial expansion suffice. Calculations based on taking logarithms, done in some cases, may not be sufficiently accurate.)

For the 1 error in 6979 data entries, there is a 95 percent certainty in rejecting P less than 7×10^{-6} and 99-percent certainty in rejecting P less than 1.5×10^{-6}. Combining the upper and lower bounds results in $7 \times 10^{-6} < P < 5 \times 10^{-3}$ for the 95-percent confidence bounds and $1.5 \times 10^{-6} < P < 7.7 \times 10^{-3}$ for the 99-percent confidence bounds.

Table 2-6. Changes or Errors in Modes of Data Entry

Entry System	Data Entered	No. Changes*	P
Key Punching	6,979	4	5.7×10^{-4}
Interactive Remote	38,548	140	3.6×10^{-3}

* These changes include corrections of the original coding sheet. Only one error was uncovered for the keypunching process.

D. SAMPLE SIZE REDUCTION

The effect of reduction of the sample size should be elaborated upon. As noted, the sample size originally expected was 12,700 and the final sample was 2871. There are two ways of approaching the effect of sample size

reduction: (1) the probability of not observing a phenotype or a combination of phenotypes and (2) the statistical error or the standard deviation in the phenotypic frequency-of-occurrence data.

1. Probability of Not Observing a Phenotype

The frequency of occurrence f is the number of individuals with a given combination of blood genetic markers n divided by the total number of persons in the population n_T

$$f = n/n_T$$

An individual selected at random (i.e., each individual has the same opportunity of being selected) would have a probability p = f of having this combination. Estimates of the frequencies in the total population can be made by looking at the results from a reduced number of individuals selected at random. For multiple random samples of size N, it would be found that the number of individuals X in each sample with the given combination of genetic markers would have a binomial distribution with mean

$$\overline{X} = Np$$

and standard deviation

$$\sigma_X = \left[N_p (1-p)\right]^{\frac{1}{2}}$$

The frequency $\overline{X}/N$ will then have a binomial distribution with mean

$$\overline{X}/N = p$$

and standard deviation

$$\sigma = \left[p(1-p)/N\right]^{\frac{1}{2}}$$

Thus, $\overline{X}/N$ is an estimate of p. The larger N, the more reliable the estimate. Confidence intervals are easily obtained when N becomes sufficiently large that the binomial distribution of sample means can be approximated by the normal distribution. The 95- and 99-percent confidence intervals about the expected estimate p are approximately $\pm 2\sigma_p$ and $\pm 3\sigma_p$ respectively.

For any sample of size N there will be a non-zero probability $(1-p)^N$ that the event is not included in the sample. Therefore, in order to reduce the probability of not observing the combination at all to less than Z, one would need to increase N until

$$(1-p)^N < Z$$

Values for N are given for several values of Z in Table 2-7.

Table 2-7. Sample Size Requirements, p = 0.001

Z*	N**
0.06	2871
0.05	3000
0.01	4600
0.001	6900
0.000003	12,700

* Probability of not observing a rare phenotype of p = 0.001.
** Number of persons in the sample

Table 2-7 implies that reducing the sample size from 12,700 to 2871 substantially increases the probability of not observing a rare phenotype or a rare combination of phenotypes (whose frequency-of-occurrence estimate is p=0.001) from 0.000003 to 0.06. In other words, the one-fourth reduction of sample size increased the error of missing a phenotype by a factor of 20,000.

2. Statistical Error

The statistical error or standard deviation in the phenotypic frequency-of-occurrence data can be calculated if one simplifying but important assumption is made. Because there is no a priori knowledge of

what the averages or the statistical means for frequency-of-occurrence data would be if the sample size were 12,700, it is assumed that the averages or the statistical means would not differ from those for the frequency-of-occurrence data collected from a sample size of 2871. With this assumption in mind, the following formula given above for the standard deviation can be used as follows:

$$\frac{\sigma_{p(N = 12,700)}}{\sigma_{p(N = 2871)}} = \frac{[p(1-p)/12,700]^{\frac{1}{2}}}{[p(1-p)/2871]^{\frac{1}{2}}}$$

By eliminating p and (1-p), the ratio of the standard deviation for the sample size of 12,700 to that for the sample size of 2871 is simply the ratio of the square roots of the respective sample sizes:

$$\frac{\sigma_{p(N = 12,700)}}{\sigma_{p(N = 2871)}} = \frac{(12,700)^{\frac{1}{2}}}{(2871)^{\frac{1}{2}}} = 2.09$$

Therefore, by assuming that the statistical means would be the same in the phenotypic frequency-of-occurrence data for the original target as in the reduced sample, the one-fourth reduction in sample size increases the statistical error in the phenotypic frequency-of-occurrence data by a factor of two.

The two discussed effects resulting from the reduction in sample size should be borne in mind in interpretation of the results presented in the next section.

III. ANALYSES OF DATA

A. PHENOTYPIC FREQUENCIES

The frequency-of-occurrence data collected for genetic marker phenotypes investigated during this study are compiled in Tables 3-1 through 3-11. Frequency-of-occurrence data are broken down by city, and, when a reference is not given, they were contributed by the crime laboratory having jurisdiction in that city. These data are juxtaposed to frequency-of-occurrence data taken from the literature for other cities to provide comparisons. These comparisons are discussed in a later section. Rare variants are not shown in these tables for simplification.

The data presented in Tables 3-1 through 3-11 are used in the statistical analyses discussed in Section III. B. The forensic serologist may also find these data of value in evidence analysis or in court litigation. The proper use of the phenotypic frequency data requires calculations of the accuracy which can be done in terms of the standard error. As discussed at the end of Section II, phenotypic frequencies are only estimates of p, the true probability of occurrence of the phenotype. If one were to take all possible samples of size N from the population to arrive at the probability of occurrence of a phenotype, one would obtain N number of frequencies, f_1, ..., f_N, which would be distributed about a mean, $\bar{f}$. When N is large, the distribution is approximately a normal distribution with a standard deviation or standard error, σ_p, given as:

$$\sigma_p = \sqrt{\frac{p(1-p)}{N}} \qquad (1)$$

Because p is not known a priori, it is necessary to find approximate forms for the standard error. First, it is noted that p(1-p) is never greater than 0.25. Thus, an upper bound for σ_p is

$$\frac{1}{2\sqrt{N}}$$

Table 3-1. Phenotypic Frequencies for ABO Obtained from Different Geographical Locations

Location	Phenotypic Frequencies (%)												Sample Size	
	O		A_1		A_2		B		A_1B		A_2B			
	Cauc	Blk	Cauc	Blk	Cauc	Blk	Cauc	Blk	Cauc	Blk	Cauc	Blk	Cauc	Blk
Miami	48.8	52.4	22.8	18.3	7.5	3.0	13.4	22.2	4.7	1.8	2.8	2.4	359	334
Detroit	41.2	53.1	33.5	16.0	1.4	0.2	17.0	26.2	4.1	2.8	2.8	1.8	507	507
Los Angeles	57.9	54.8	17.9	14.4	6.7	5.8	15.4	21.2	0.8	0	1.3	3.8	240	104
Bullocks & Evans Counties, Georgia[5]	51.0	54.0	29.0	17.0	12.0	8.0	6.0	19.0	1.0	0	0	2.0	335	300
Tecumseh, Mich.[6]	43.7	--	34.0	--	9.5	--	9.1	--	2.6	--	0.01	--	8965	--
San Francisco[7]	45.0	49.0	32.0	19.0	9.0	8.0	11.0	19.0	3.0	3.0	1.0	1.0	8962	3146
San Francisco[7] (Mexicans)	56.0	--	22.0	--	5.0	--	13.0	--	4.0	--	0	--	335	--
Seattle[8]	42.0	--	35.0	--	9.0	--	11.0	--	3.0	--	1.0	--	5657	--
West Virginia[9]	45.9	--	41.3	--	10.0	--	9.1	--	3.7	--	1.0	--	1412	--

Cauc. denotes Caucasians
Blk. denotes Blacks

Table 3-2. Phenotypic Frequencies for MNSs Obtained from Different Geographical Locations (Sample size on last page of table)

Location	Phenotypic Frequencies (%)									
	MS		MSs		Ms		MNS		MNSs	
	Cauc	Blk	Cauc	Blk	Cauc	Blk	Cauc	Blk	Cauc	Blk
Miami	8.7	2.2	12.0	7.3	9.1	15.3	5.5	3.8	19.1	10.8
Detroit	6.3	3.1	17.6	6.2	13.4	13.4	2.4	2.7	24.3	11.0
San Francisco[7]	6.4	1.8	14.8	6.4	9.5	15.3	3.7	3.1	22.2	11.1
Tecumseh, Mich.[6]	6.95	--	14.9	--	8.25	--	3.3	--	23.2	--
West Virginia[9]	6.4	2.0	13.8	6.0	9.6	16.0	3.6	4.0	24.1	10.0
England[5]	5.7*		14.0*		10.1*		3.9*		22.4*	

*Caucasian and Black populations grouped together.

Table 3-2. Phenotypic Frequencies for MNSs (continued)

Location	Phenotypic Frequencies (%)								Sample Size	
	MNs		NS		NSs		Ns			
	Cauc	Blk	Cauc	Blk	Cauc	Blk	Cauc	Blk	Cauc	Blk
Miami	22.0	34.4	1.0	2.0	6.1	4.1	16.5	20.0	309	314
Detroit	22.3	38.1	1.7	2.3	3.0	4.0	10.0	19.4	461	485
San Francisco[7]	23.3	36.0	0.5	1.0	5.1	5.3	14.5	19.9	8962	3146
Tecumseh, Mich.[6]	23.1	--	0.35	--	5.0	--	14.9	--	not given	--
West Virginia[9]	24.1	33.0	0.6	1.0	4.5	4.0	13.4	24.0	1051	106
England[5]	22.6*		0.3*		5.4*		15.6*		1000*	

*Caucasian and Black populations grouped together.

Figure 3-3. Phenotypic Frequencies for Rh Obtained from Different Geographical Locations (Sample size on last page of table)

Location	Phenotypic Frequencies (%)											
	CCDE		CCDee		CCddee		CcDE		Ccdee		Ccddee	
	Caucasian	Black	Caucasian	Black	Caucasian	Black	Caucasian	Black	Caucasian	Black	Caucasian	Black
Miami	0.3	0	17.7	2.1	0	0	0.6	0	35.2	20.2	0.8	0.6
Detroit	0.2	0	13.5	1.6	0.6	0	0.4	0	32.3	20.8	1.2	1.2
Los Angeles	0.5	0	23.1	5.4	0.3	0	1.1	1.1	31.2	22.8	0.5	0
San Francisco[7]	0	0	19.2	2.2	0	0	0	0	34.4	22.7	0	0
San Francisco[7] (Mexicans)	0	--	24.2	--	0	--	0	--	30.8	--	0	--
Tecumseh, Mich.[6]	0	--	16.2	--	0.01	--	0.1	--	34.0	--	0.4	--
West Virginia[9]	0	0	17.4	2.0	0	0	0.3	0	33.4	26.0	0.3	1.0
London[5]	0.09	--	20.7	--	0	--	11.5	--	34.1	--	0.6	--

Table 3-3. Phenotypic Frequencies for Rh (continued)
(Sample size on last page of table)

Location	Phenotypic Frequencies (%)											
	ccDE		ccDee		ccddE		ccddee		ccD^uE		CCDEe	
	Caucasian	Black	Caucasian	Black	Caucasian	Black	Caucasian	Black	Caucasian	Black	Caucasian	Black
Miami	3.7	1.8	3.1	51.4	0	0	13.2	5.4	0	0	0.8	0
Detroit	2.4	1.6	2.6	50.3	0	0	23.2	7.3	0	0	0.2	0
Los Angeles	7.0	2.2	4.3	51.1	0	0	10.8	6.5	0	0	0	0
San Francisco[7]	2.4	1.1	2.3	48.1	0	0	15.2	6.4	0	0	0	0
San Francisco[7] (Mexicans)	3.9	--	2.4	--	0	--	6.0	--	0	--	0	--
Tecumseh, Mich.[6]	2.3	--	2.1	--	0.01	--	15.8	--	0.01	--	0.06	--
West Virginia[9]	3.9	2.0	2.1	47.0	0	0	14.3	7.0	0	0	0.2	0
London[5]	14.7	--	0.2	--	0.7	--	15.3	--	0	--	0	--

Table 3-3. Phenotypic Frequencies for Rh (continued)

Location	Phenotypic Frequencies (%)												Sample Size	
	CcDEe		ccDEe		CCD^{u}ee		ccD^{u}Ee		ccddEe		CcD^{u}ee			
	Caucasian	Black	Caucasian	Black	Caucasian	Black	Caucasian	Black	Caucasian	Black	Caucasian	Black	Caucasian	Black
Miami	11.8	3.9	11.3	13.9	0.3	0	0	0	1.1	0	0.3	0	355	331
Detroit	13.1	3.6	9.3	13.5	0	0	0	0	1.0	0	0.2	0	505	505
Los Angeles	9.7	5.4	11.3	5.4	0	0	0	0	0.5	0	0	0	186	92
*San Francisco[7]	13.3	3.4	11.5	14.6	0	0	0	0	0	0	0	0	8962	3142
San Francisco[7] (Mexicans)	17.6	--	9.9	--	0	--	0	--	0	--	0	--	335	--
Tecumseh, Mich.[6]	12.5	--	12.5	--	0.04	--	0.17	--	0.6	--	0.4	--	8963	--
West Virginia[9]	14.4	3.0	12.8	13.0	0	0	0.1	0	0.6	0	0.3	0	1412	135
London[5]	0	--	0	--	0	--	0	--	0	--	0	--	1038	--

* All other phenotypes not shown were expressed by 1.6% Caucasian, 1.6% Blacks, and 5.4% Mexicans.

Table 3-4. Phenotypic Frequencies for ADA Obtained for Different Geographical Locations

Location	Phenotypic Frequencies (%)						Sample Size	
	ADA 1		ADA 2-1		ADA 2			
	Caucasian	Black	Caucasian	Black	Caucasian	Black	Caucasian	Black
Miami	90.7	97.0	9.0	1.5	0.3	1.5	354	330
Detroit	88.7	98.4	11.1	1.6	0.2	0	506	504
Los Angeles	93.5	87.5	6.5	12.5	0	0	31	16
Seattle[10]	90.5	96.7	9.5	3.2	0	0	168	186
London[1]	90.40*		9.38*		0.22*		1353*	

* Caucasian and Black populations are grouped together.

Table 3-5. Phenotypic Frequencies for AK Obtained from Different Geographical Locations

Location	Phenotypic Frequencies (%)						Sample Size	
	AK 1		AK 2-1		AK 2			
	Caucasian	Black	Caucasian	Black	Caucasian	Black	Caucasian	Black
Miami	93.0	97.9	6.8	2.1	0.3	0	355	331
Detroit	94.7	99.6	5.3	0.3	0	0	506	506
Los Angeles	90.6	100	9.3	0	0	0	32	19
Ann Arbor, Michigan[11]	94.5	97.1	5.5	2.9	0	0	254	139
England[1]	91.15*		8.74*		0.10*		1887*	

* Caucasian and Black populations not separated.

Table 3-6. Phenotypic Frequencies for EAP Obtained from Different Geographical Locations

Location	Phenotypic Frequencies (%)																Sample Size	
	EAP A		EAP BA		EAP B		EAP CA		EAP CB		EAP C		EAP RB		EAP RA			
	Cauc	Blk	Cauc	Blk	Cauc	Blk	Cauc	Blk	Cauc	Blk	Cauc	Blk	Cauc	Blk	Cauc	Blk	Cauc	Blk
Miami	12.9	7.0	39.7	30.9	42.3	59.4	2.6	0.3	2.3	1.8	0.3	0	0	0.3	0	0.3	350	330
Detroit	13.1	6.0	37.0	34.5	39.0	55.9	4.2	0.8	6.8	1.2	0	0	0	0	0	0	503	501
Los Angeles	10.0	5.0	40.6	47.0	46.1	46.0	2.7	0	0.5	1.0	0	0	0	1.0	0	0	219	100
Seattle[8]	17.1	7.0	39.4	35.0	31.6	51.7	5.2	0.5	6.7	2.3	0	0	0	2.6	0	0.7	193	429*
Austin, Texas[12]	--	3.9	--	32.5	--	55.5	--	0.6	--	4.2	--	2.0	--	0.6	--	0.6	--	357**
Pittsburgh[13]	11.7	5.4	42.5	33.3	39.6	59.3	1.6	0.3	4.6	1.5	--	0	--	--	--	0.1	1239	718
England[5]	13.5[+]		43.1[+]		32.0[+]		4.4[+]		6.9[+]		0[+]		0[+]		0[+]		880[+]	

*also 0.2% DB

**also 0.3% RC and 0.6% DB

[+]Caucasian and Black populations grouped together

Table 3-7. Phenotypic Frequencies for EsD* Obtained from Different Geographical Locations

Location	Phenotypic Frequencies (%)						Sample Size	
	EsD 1		EsD 2-1		EsD 2			
	Caucasian	Black	Caucasian	Black	Caucasian	Black	Caucasian	Black
Miami	79.3	83.8	19.8	15.2	0.8	0.9	348	328
Detroit	78.1	84.0	20.9	15.2	1.0	0.8	507	505
Los Angeles	71.8	77.4	26.9	20.4	1.4	2.2	216	26

* EsD phenotypic frequencies could not be found for other areas.

Table 3-8. Phenotypic Frequencies for G6PD Obtained from Different Geographical Locations

Location	Phenotypic Frequencies (%)						Sample Size	
	G6PD A		G6PD BA		G6PD B			
	Caucasian	Black	Caucasian	Black	Caucasian	Black	Caucasian	Black
Miami	0	23.1	0	8.1	100	68.8	352	321
Detroit	0.8	21.7	0	11.4	99.2	66.9	506	498
England[5]	13.85*		52.64*		31.98*		397*	

* Caucasian and Black populations grouped together.

Table 3-9. Phenotypic Frequencies for Gc Obtained from Different Geographical Locations

Location	Phenotypic Frequencies (%)						Sample Size	
	Gc 1		Gc 2-1		Gc 2			
	Caucasian	Black	Caucasian	Black	Caucasian	Black	Caucasian	Black
Miami*	46.0	78.3	43.8	18.7	10.2	1.8	352	327
Detroit	49.1	75.6	42.3	23.3	8.6	1.0	501	484
Los Angeles	56.5	84.6	34.8	15.4	8.7	0	23	13
Tecumseh, Michigan[6]	51.1	--	40.9	--	7.9	--	7658	--
England[5]	48.98**		44.90**		6.10**		49	

* There were also 1.2% Blacks who were 1-Y.
** Caucasian and Black populations grouped together.

Table 3-10. Phenotypic Frequencies for Hp Obtained from Different Geographical Locations

Location	Phenotypic Frequencies (%)						Sample Size	
	Hp 1		Hp 2-1		Hp 2			
	Caucasian	Black	Caucasian	Black	Caucasian	Black	Caucasian	Black
Miami*	16.1	37.9	44.1	44.1	39.8	12.5	347	311
Detroit**	13.8	28.1	41.9	41.6	44.1	16.7	506	502
Los Angeles***	15.9	29.1	47.2	41.9	35.5	24.4	214	86
Tecumseh, Michigan[6]	17.1	--	48.8	--	33.6	--	7655	--
Pittsburgh[13]	14.6	28.6	46.0	49.8	39.2	20.5	1263	721
England[1]	13.85+		52.64+		31.98+		397+	

* There were also 5% Blacks who were Hp 2-1M and 0.3% Blacks Hp O.
** There were also 0.19% Caucasians who were Hp 2-1M, 11.3% Blacks Hp 2-1M, and 2% Blacks Hp O.
*** There were also 1.4% Caucasians who were Hp 2-1M, and 4.6% Blacks Hp 2-1M.
+ Caucasian and Black populations not separated.

Table 3-11. Phenotypic Frequencies for PGM_1 Obtained from Different Geographical Locations

Location	Phenotypic Frequencies (%)						Sample Size	
	PGM_1 1		PGM_1 2-1		PGM_1 2			
	Caucasian	Black	Caucasian	Black	Caucasian	Black	Caucasian	Black
Miami	60.1	62.7	33.1	32.7	6.8	4.5	353	330
Detroit*	57.6	60.8	35.0	34.6	7.3	3.8	505	503
Los Angeles**	57.5	68.3	36.8	25.7	5.7	5.0	228	101
Seattle[8]	56.6	65.5	37.3	30.9	--	--	508	654
Pittsburgh[13]	57.7	67.4	38.9	29.3	5.3	3.4	1253	714
England[1]	58.4	61.8	36.1	33.6	5.5	4.6	2115	103

* There were also 0.4% Blacks who were 1-AT and 0.4% Blacks 2-1AT.

** There were also 0.9% Blacks who were 3-2.

by substitution into equation (1). To exemplify, for a sample size of N = 100, the standard error is no greater than 0.05. When the sample size, N, is very large and f is a "good" estimate of p, f may be substituted for p in equation (1) to give equation (2):

$$\sigma_p \approx \sqrt{\frac{f(1-f)}{N}} \qquad (2)$$

Example applications of equations (1) and (2) to the phenotypic frequencies given in Tables 3-1 through 3-11 follow. If a Los Angeles Caucasian tested as AK type 2-1 is said to match that of a blood stain, what is the probability that a Caucasian picked at random from the Los Angeles population would also match the bloodstain? From Table 3-5, it can be seen that the phenotypic frequency for AK type 2-1 in Los Angeles for a Caucasian is 0.093. The sample size, N = 32, is considerably smaller and suggests that this frequency may not be a "good" estimate of the true probability of occurrence of this phenotype. Using equation (1), one obtains the standard deviation,

$$\sigma_p = \frac{1}{2\sqrt{32}} = 0.09$$

Because the normal distribution gives a 95-percent level of confidence that the true probability of occurrence is within two standard deviations of the estimated probability of occurrence of this phenotype, a conservative estimate of the probability that a person picked at random from the Los Angeles forensic population matching the bloodstain is

$$0.093 + 2(0.09) = 0.27.$$

For a second example, if a bloodstain is typed as Hp 2, what is the probability that a Black picked at random in Detroit is the source of the bloodstain? From Table 3-10, it can be seen that the phenotypic frequency for Hp type 2 in Detroit for a Black is 0.167. The sample size was N = 502.

Using equation (1), one obtains $\sigma_p = 0.022$ which implies the estimated probability of occurrence is fairly accurate and that f can be substituted for p. Using equation (2), one obtains the standard deviation as follows:

$$\sigma_p = \sqrt{\frac{(0.167)(1-0.167)}{502}} = 0.016$$

With over 95-percent confidence, one can state that the probability that a Black picked at random from the Detroit population is the source of the bloodstain is

$$0.167 + 2(0.016) = 0.199.$$

B. STATISTICAL ANALYSES

To reiterate, the objectives of this study were to determine whether variations exist in frequency distributions of blood groups: (1) among racial populations in the United States, (2) among regions in the United States, and (3) between United States and British Caucasian populations. This section is concerned with statistical analyses of the phenotypic frequency data collected during this study and is organized to address the stated objectives in that order. The statistical analyses conducted for this study were done at the University of California, Los Angeles (UCLA), with a standard biomedical package of statistical programs[4] which include the Pearson chi square, likelihood ratio chi square, and Fisher exact Tests.

1. Homogeneity Among Ethnic Groups

The first set of hypotheses to be tested is that the phenotypic frequencies are homogeneous among ethic groups. This is tested by considering each ethnic group to be a random sample from the same population. Although samples from the same population are not expected to give exactly the same estimates of a phenotypic frequency, they are expected to be "close." When only two ethnic groups (e.g., Blacks and Caucasians) are involved, the distance between the frequencies might well be taken as the difference

between the groups. When multiple ethnic groups are involved, however, more complex concepts of distance are required. Some of these concepts are more easily related to probabilistic statements than others and some are more easily calculable than others. One measure of distance for counting phenotypes in a population is the Fisher exact statistic.[4] Unless the sample size is small and the differences between phenotypic frequencies large, however, the computations involved become excessive. For samples which are so large that the distribution is approximately continuous and the subsamples are not too far apart, another measure of distance used is the Pearson chi square statistic.[4] If one theoretically considers all possible samples, the resulting statistics reflect the distance between ethnic group frequencies has a distribution that can be calculated (Fisher) or estimated (Pearson) theoretically. The fraction of the number of samples with distance statistics greater than an observed value is considered to be the fraction of the number of times (i.e., the probability) of being in error in rejecting the hypothesis when in fact it is true. When the probability is fairly small (small being typically taken as less than 5 percent), one is fairly safe in rejecting the underlying hypothesis of homogeneity of phenotypic frequencies among ethnic groups. The hypothesis is expressed as $P(M, E \mid S, C) = P(M \mid S, C)P(E \mid S, C)$. $P(M, E \mid S, C)$ is the probability of occurrence of a genetic marker M and an ethnic group E, both within sex S and city C; $P(M \mid S, C)$ is the probability of occurrence of a genetic marker within sex and city; and $P(E \mid S, C)$ is the probability of occurrence of an ethnic group within sex and city.

Nine ethnic groups were used as described in Table 1-1. The underlying hypothesis was tested for each genetic marker for each sex in each city. If any of the nine ethnic groups was not represented in a particular data set, it was not considered in computing the statistics. For example, the Los Angeles data used the one category of Hispanic (primarily Mexican) while

the Miami data subdivided Hispanics (primarily Cuban). The probabilities for finding a distance statistic (Pearson or chi square) as large or larger than that observed when taking multiple samples from the same population are given in Table 3-12. When the probability for the Fisher statistic was computed, it was more accurate than the probability for the Pearson statistics based on the large sample approximation. For that reason, the Fisher statistic should be given more weight when both statistics are available.

In interpreting these results, several comments should be kept in mind. In testing a large number of such hypotheses, one would expect 5 percent of the tests to show statistical significance at the 0.05 probability level due to random fluctuation when the hypothesis is in fact correct. For Laboratories 4 and 5, far more than 5 percent of the results have occurrence probabilities of less than 0.05.

In interpreting those probabilities greater than 0.05, it is very important to keep in mind that a high probability for a statistic does not mean that one can safely accept the hypothesis of homogeneity of phenotypic frequencies among ethnic groups. It may simply mean that the sample sizes were too small to detect the existing differences. Indeed, the total numbers of persons involved (see Table 1-1) show that Laboratories 1 and 2 have much smaller sample sizes than Laboratories 4 and 5 and hence, one would be less able to detect small differences.

In summary, the results in Table 3-12 show that many more phenotypes were significantly nonhomogeneous than would be expected by chance, and strong ethnic bias is indicated for several phenotypic frequencies.

2. Independence of Sex, Genetic Marker, and City Within Ethnic Groups

To determine whether sex, genetic marker, and city have bearing on the phenotypic frequencies obtained within an ethnic background, another statistical analyses was carried out. The question of independence is not

Table 3-12. Probabilities for Test of Homogeneity of Phenotypic Frequencies Over Ethnicity Within City and Sex, P(M, E | S, C) = P(M | S, C) P(E | S, C)

Type	Sex	LABORATORY									
		Number 1		Number 2		Number 3		Number 4		Number 5	
		Pearson	Fisher	Pearson	Fisher	Pearson	Fisher	Pearson	Fisher	Pearson	Fisher
ABO	M	0.917		0.9162		0.0105*		0.0001*		0.0000*	
	F			0.3998		0.2731		0.0004*		0.0000*	
ASUB	M					0.3241		0.3169		0.3766	0.3079
	F			0.6210	0.8182	0.5514		0.0000*		0.6356	0.4031
Rh-D	M					0.2561		0.0174*		0.0000*	
	F					0.2249		0.0314*		0.0000*	
Rh-C	M	0.4611		0.2361	0.3538	0.0000*		0.0000*		0.0000	
	F			0.4059	0.5714	0.0182*		0.0000*		0.0000	
Rh-e	M	0.0261*		0.0530	0.1923	0.0761		0.0000*		0.0000*	
	F			0.5991	0.5667	0.0006*		0.0009*		0.0000*	
Rh-E	M			0.2870	0.4167	0.0218*		0.1110		0.0790	
	F			0.5439	0.7206	0.5994		0.0166*		0.0136*	
Rh-e	M			0.8163	0.9500	0.4519		0.6096		0.2059	0.1727
	F					0.8849		0.3819		0.3952	0.2855
M	M	0.4250	0.4909	0.5422	0.7192			0.4584		0.0001*	
	F			0.5991	0.5667			0.9372		0.0028*	
N	M	0.4250	0.4909	0.3715	0.5205			0.4126		0.0001*	
	F			0.5991	0.5667			0.9278		0.028*	
S	M	0.6576	0.6182	0.1894	0.2962			0.0000*		0.0001*	
	F			0.0679	0.1474			0.0005*		0.0028	
s	M	0.4250	0.4908	0.6286	0.5077			0.0762		0.1289	0.0885
	F			0.5312	0.7158			0.6544		0.6362	
Rh	M	0.0889		0.4921		0.0000*		0.0000*		0.0000*	
	F			0.9263		0.0002*		0.0000*		0.0000*	
MN	M	0.1636		0.7901				0.0445*		0.0000*	
	F			0.0850				0.1187		0.0000*	
Lewis a	M							0.0117*			
	F							0.7975			

* Error is small (<5%) in rejecting hypothesis of homogeneity of phenotypic frequencies over ethnicity within city and sex.

Table 3-12. Probabilities for Test of Homogeneity of Phenotypic Frequencies Over Ethnicity Within City and Sex P(M, E | S, C) = P(M | S, C) P(E | S, C) (continued)

Type	Sex	LABORATORY									
		Number 1		Number 2		Number 3		Number 4		Number 5	
		Pearson	Fisher	Pearson	Fisher	Pearson	Fisher	Pearson	Fisher	Pearson	Fisher
ADA	M	0.4250	0.4909	0.4971	0.6694	0.0258*		0.1825		0.0000*	0.0000*
	F			0.7324	0.9000	0.4204		0.0090*		0.0000*	
AK	M			0.6751	0.8502	0.7529		0.8843		0.0001*	0.0000*
	F			0.7324	0.9000	0.5222		0.2524		0.0104*	0.0100*
EAP	M	0.8865	0.7273	0.2169		0.9363		0.5516		0.0000*	
	F			0.4375		0.6709		0.0002*		0.0000*	
FsD	M			0.5638		0.0178*		0.0005*		0.0710	
	F			0.3894	0.4474	0.0002*		0.0035*		0.4057	
G6PD	M			0.9446				0.0000*		0.0000*	
	F							0.0000*		0.0000*	
6PGD	M									0.0740	
	F									0.1482	
GPT	M	0.8214									
	F										
GLO	M					0.5991					
	F					0.7214					
PEP-A	M									0.0000*	
	F									0.0000*	
PGM-1	M	0.4250	0.4909	0.4469		0.3617		0.0106*		0.0116*	
	F			0.2735		0.6729		0.0171*		0.6741	
PGM-2	M					0.5167				0.3207	0.5040
	F									0.1565	0.2495
CA-2	M									0.0000*	
	F									0.0000*	
Hb	M									0.0000*	
	F									0.0000*	
Hp	M	0.2307		0.0298*		0.1888		0.0000*		0.0000*	
	F			0.4994		0.8430		0.0012*		0.0000*	
Gc	M			0.4255		0.8212		0.0000*			
	F										

*Error is small (<5%) in rejecting hypothesis of homogeneity of phenotypic frequencies over ethnicity within city and sex.

simple because there are several ways in which several factors can be independent. Events A and B are said to be independent if P(A and B) = P (A) P (B). The events A, B, and C, however, can be pair-wise independent without being "mutually" independent in the sense that P(A and B and C) $\neq$ P(A) P(B) P(C).

The first hypothesis tested was that the three factors (sex, genetic marker, and city) are mutually independent, i.e., P(S, M, C) = P(S) P(M) P(C). The likelihood ratio chi square and the Pearson chi square statistics were calculated comparing the hypothesis of mutual independence (as given above) with the hypothesis of mutual dependence. The results in Table 3-13 show that in most genetic markers among Caucasians and Blacks, the hypothesis of mutual independence is highly improbable and can be safely rejected. Higher probabilities for mutual independence were found among the Hispanics and may, in part, be due to the smaller numbers sampled and hence, giving rise to diminished sensitivity. Therefore, these results are not shown.

Because dependence between sex, genetic marker, and city within ethnic groups was shown, it becomes important to determine which of these interactions is significant. In order to determine this, the hypothesis that sex and genetic marker are independent within a given city and ethnic group, expressed as P(S, M | C, E) = P(S | C, E) x P(M | C, E), is compared with the hypothesis that sex and genetic marker within a given city and ethnic group are not independent.

In Table 3-14, the probabilities for the likelihood ratio chi squares and the Pearson chi squares for the analyzed genetic markers are given. It is seen that the probabilities are sufficiently large and that there is no evidence for interaction between sex and genetic marker within these populations which is significant at the 5-percent level, except for G6PD. This is in accordance with known results.[8]

Under the safe assumption that sex does not interact with genetic

Table 3-13. Probabilities for Tests of Independence of Sex, Genetic Marker, and City Within Ethnic Group, P(S, M, C) = P(S) P(M) P(C)

Genetic Marker System	Probabilities			
	Caucasian		Black	
	LR	P	LR	P
ABO	0.0000*	0.0000*	0.0000*	0.0000*
A Subgroups	0.0001*	0.0001*	0.0000*	0.0000*
Rh	0.0202*	0.0221*	0.2779	0.1486
MNSs	0.0166*	0.0164*	0.0009*	0.0011*
ADA	0.1766	0.2044	0.0000*	0.0000*
AK	0.1135	0.1248	0.0000*	0.0000*
EAP	0.0021*	0.0035*	0.0022*	0.0016*
EsD	0.0031*	0.0032*	0.0000*	0.0000*
PGM_1	0.0076*	0.0089*	0.0000*	0.0000*
Hp	0.0223*	0.0141*	0.0000*	0.0000*
Gc	0.1271	0.0491*	0.0001*	0.0001*

*Error is small (<5%) in rejecting hypothesis that sex, genetic marker, and city within ethnic group are independent.

LR = Likelihood Ratio
P = Pearson Chi Square

Table 3-14. Probabilities for Tests of Independence of Sex and Genetic Marker Given City and Ethnic Origin $P(S, M \mid C, E) = P(S \mid C, E)\, P(M \mid C, E)$

Genetic Marker System	Probabilities*			
	Caucasian		Black	
	LR	P	LR	P
ABO	0.4805	0.4864	0.2743	0.3029
A Subgroups	0.2429	0.2451	0.0750	0.0947
Rh	0.9047	0.9401	0.9976	0.9974
MNSs	0.4155	0.4398	0.2930	0.2913
ADA	0.8863	0.8881	0.5552	0.5268
AK	0.2790	0.2846	0.3306	0.4388
EAP	0.8497	0.8507	0.8824	0.8770
EsD	0.4759	0.5199	0.2011	0.1841
PGM_1	0.9991	0.9991	0.1618	0.1796
Hp	0.9558	0.9514	0.7914	0.7853
Gc	0.0630	0.0680	0.2362	0.2159
G6PD	0.4557**	0.5300**	0.0000	0.0000

*All probabilities in Table suggest significant error in rejecting hypothesis that sex and genetic marker, given city and ethnic origin, are independent.

**All except four persons reporting as Caucasians were tested as G6PD type B. These exceptions were male type A. It cannot be confirmed whether these four have any Black ancestry. Without these four male types A, the two sexes would express the same phenotypic frequencies, and there is no evidence in the data for interaction between sex and phenotype.

LR = Likelihood Patio P = Pearson Chi Square

marker and city within ethnic groups, the interaction between marker and city was tested. In this instance, the hypothesis, that genetic marker and city are independent within a given sex and ethnic group, $P(M, C \mid S, E) = P(M \mid S, E)\, P(C \mid S, E)$ was tested against the hypothesis that genetic marker and city are not independent within a given sex and ethnic group. The probabilities for the tests shown in Table 3-15 are generally not significant, but four notable exceptions can be seen where both the likelihood ratio and Pearson chi square statistics have a probability of occurrence of less than 0.05. This is unusual because the number of tests with probability less than 0.05 is expected to be less than 3.1 at the 95 percent confidence level. There are two ways in which one of the four exceptions can be accounted for. First, if one of the four exceptions, namely, that exhibited in MNSs phenotypic frequencies, was ascribed to differences in laboratory analysis between Laboratories 4 and 5 and if there was no actual difference in the two populations, then there would leave only three notable exceptions. The second way to account for one of four exceptions is to note that of the 22 tests, one would expect 5 percent or 1.1 tests to have a probability less than 0.05, by mere random fluctuations. With either consideration, then one could say that three remaining apparent geographic differences in phenotypic frequencies are expected by chance.

These results showed that the phenotypic frequencies differ in a significant manner in separate geographical locations and led to another statistical analysis to address the same question. Data from the two sexes were not distinguished, but phenotypes and city were distinguished within each ethnic group. The probabilities for the likelihood ratio chi square and Pearson chi squares for testing the hypothesis $P(M, C \mid E) = P(M \mid E)P(C \mid E)$ are given in Table 3-16. Of the 64 tests (2 per ethnic group and marker), 18 were significant at the 5-percent level. This is far greater than the number expected under the hypothesis of no effect due to city and

Table 3-15. Probabilities for Tests of Independence of Genetic Marker and City Given Sex and Ethnic Origin, $P(M, C \mid S, E) = P(M \mid S, E)\, P(C \mid S, E)$

Genetic Marker System	Probabilities			
	Caucasian		Black	
	LR	P	LR	P
ABO	0.0342*	0.0569	0.1969	0.1681
A Subgroups	0.0001*	0.0001*	0.0008*	0.0012*
Rh	0.0681	0.0905	0.8947	0.7622
MNSs	0.0088*	0.0090*	0.6542	0.6731
ADA	0.8863	0.8881	0.5552	0.5268
AK	0.2251	0.1917	0.1332	0.0379*
EAP	0.9497	0.8507	0.9824	0.8770
EsD	0.4689	0.3963	0.3046	0.1907
PGM_1	0.9926	0.9926	0.5900	0.5981
Hp	0.9558	0.9514	0.7914	0.7853
Gc	0.1001	0.0326*	0.3830	0.3151

* Error is small (<5%) in rejecting hypothesis that genetic marker and city, given sex and ethnic origin, are independent.

Table 3-16. Probabilities for Tests of Independence of Genetic Marker and City Given Ethnic Origin $P(M, C \mid E) = P(M \mid E)\, P(C \mid E)$

Genetic Marker System	Probabilities			
	Caucasian		Black	
	LR	P	LR	P
ABO	0.0014*	0.004*	0.1474	0.1399
A Subgroups	0.0000*	0.0000*	0.0010*	0.0006*
Rh	0.0025*	0.0026*	0.7708	0.5540
MN	0.0010*	0.0009*	0.9421	0.9420
ADA	0.6161	0.6386	0.0602	0.0106*
AK	0.5261	0.5057	0.0532	0.0545
EAP	0.0026*	0.0086*	0.1661	0.2049
EsD	0.3084	0.2906	0.6006	0.5315
PGM_1	0.8139	0.8184	0.4824	0.4992
Hp	0.1812	0.1316	0.0001*	0.0002*
Gc	0.8394	0.8819	0.2777	0.3363

*Error is small (<5%) in rejecting hypothesis that genetic marker and city, given ethnic origin, are independent.

genetic marker interaction. Thus, again, the hypothesis that city origin has no effect on phenotypic frequencies within ethnic groups was rejected. There is strong evidence that phenotypic frequencies for any ethnic group vary from one geographic location to another. This information could be quite discomforting to forensic scientists who use the data of others in the presentation of their evidence in court.

These differences can also be viewed by not using statistical parameters. Tables 3-1 through 3-11 compare the values obtained by other sources for 11 genetic marker systems. Tables 3-1 through 3-11 also show that there is variation for each phenotype from one geographical location to another. The variations are usually between 0 and 10 percent.

It is tempting to speculate the reasons for these variations in terms of population equilibrium, drift, ethnic origins, etc., but the critical question concerns the significance of these variations for the forensic scientist in the presentation of his data during litigation. In this light, does it make a difference whether the results of bloodstain phenotyping are reported as coming from blood occurring in 56.5 or 60.2 percent of the population? The terminology generally used for presentation by serologists in court is described as an approximation, which implies a tolerance in the absolute values being reported. It is not clear that the magnitude of these variations is important.

3. Comparison of U.S. and British Caucasian Populations

The phenotypic frequencies for United States Caucasians obtained from this study (i.e., pooled from the data contributed by Miami, Detroit, and Los Angeles) were compared with those for the British-population phenotypic frequencies as compiled by Culliford.[1] In order to make the comparison, a number of simplifications had to be made. First, the question of homogeneity of phenotypic frequencies among the sampled cities in the United States was ignored. Second, the question of the representa-

tiveness or randomness of the samples (Miami, Detroit, and Los Angeles) in terms of the over-all Unived States population was ignored. Third, the British population, as sampled by Culliford, was assumed to be essentially Caucasian. (Although other ethnic populations were included in his data, Caucasians constitute 90 percent or more of the British population.) Fourth, the statistical analysis was limited to a comparison of protein/enzyme genetic marker systems because Culliford's data did not include antigenic systems.

As a measure of similarity or difference between the two populations, the probability was calculated for obtaining the observed chi square value (or larger) that the two populations (U.S. and British) were samples randomly selected from the same population. The results are summarized in Table 3-17. Of the five genetic marker systems compared, half had probabilities of occurrence of less than 0.05. Although this suggests that differences may be detectable between the U.S. and British populations, the conclusion that the two populations are different is tenuous without additional data to substantiate it.

Table 3-17. Comparison of United States and British Populations

Genetic Marker System	χ^2	υ	Probability of obtaining the observed χ^2 value or larger that the two populations were randomly selected samples from the same population.
ADA	0.55	2	0.76
AK	6.07	2	0.048**
Hp	11.36	2	0.0034***
6PGD	0.50	2	0.78
PGM			
Caucasian	2.93	2	0.23
Black	6.10	2	0.047**

* Unless otherwise specified, populations are Caucasian. U.S. population data were obtained by pooling those of Miami, Detroit, and Los Angeles, acquired during this study. The source of the British population data is reference 1. υ is the product of the number of phenotypes being compared minus one times the number of populations being compared minus one.

** Probability of observed chi square is less than 0.05.

*** Probability of observed chi square is less than 0.005.

IV. PROBABILITY OF DISCRIMINATION

A. STATISTICAL CONCEPTS

The selection of a genetic marker system for use in bloodstain analysis to discriminate between blood samples of different individuals or to determine a particular sample's rarity is generally accomplished by checking its frequency-of-occurrence data in an appropriate source. The greater the number of phenotypes and/or the more uniform the distribution of phenotypes, the more useful is the genetic marker system. The selection of the specific system to use can be quantitatively based on a statistical concept known as the discrimination probability. The discrimination probability is the probability that any two individuals chosen at random from the population will have the same phenotype(s). The ability to discriminate between individuals based upon blood phenotypes depends geometrically upon the number of genetic markers used for phenotyping. The greater the number of genetic markers in which a particular bloodstain from a crime scene matches an individual's blood, the greater the likelihood that the unknown sample came from that person. These concepts are based on the assumption of independence of genetic marker systems.*

---

*A question of fundamental importance to forensic serologists which was not directly addressed in this study was that of statistical independence of population phenotype frequencies. For example, it is commonly assumed that the probability of a person being type A and type MN is the product of the probabilities of that person being type A and of being type MN. However, it can be shown that even if the ABO system is statistically independent from the MN system in each of two populations (e.g., ethnic groups) but the two systems have different frequencies of occurrence in the two populations (as shown to be the case for ethnic groups in the previous chapter), then the two systems will not be statistically independent in the combined population. The critical question that needs to be answered here is how large an error might be introduced, and heretofore in the literature, this has not been addressed.

CONTINUED

1 OF 2

During criminal investigations, the opportunity to match blood occurs when bloodstains found at a crime scene are compared to the suspect's blood. Thus, for unknown bloodstains, a match is sought from a relatively restricted population, i.e., suspects, victims, and perhaps, witnesses.

The probability of a person's blood chosen at random matching this unknown stain, P, is equal to the simple product of the probability of occurrence (frequency of occurrence) p_i of each mutually exclusive and independent characteristic or phenotype in each genetic marker system, i, for $i = 1, \ldots, N$, analyzed.

The product of the probability of occurrence of each phenotype, p_i

$$P = \prod_{i=1}^{N} p_i \qquad (1)$$

is called the probability of match or identity. It gives the probability that a person who is chosen at random from the entire population will match the phenotypes for all i, from 1 to N genetic marker systems in the bloodstain clue. From another point of view, it can also be seen as the percentage of the population that has blood which matches the bloodstain clue. Let us apply equation (1).

If a detective discovers a dried blood clue under the fingernails of a victim, he would then obtain blood samples from the victim and the suspect and have them analyzed. While the analysis proceeds, he is presented with what is known about blood composition such as that shown in Table 4-1. Assume that the bloodstain clue contains the most common phenotype in each genetic marker system and matches that of the suspect as shown in Table 4-2. Because the victim's blood differs from the clue in the genetic marker systems, MN, Hp, and PGM_1, it is certain that the blood under the victim's fingernail is not his own. The probability that the suspect's blood

Table 4-1. Known Phenotypic Frequencies for Genetic Markers and Discrimination Probability

Genetic Marker System	Phenotype (frequency)	Discrimination Probability
ABO	$p_{A_1} = 0.340$, $p_{A_2} = 0.080$, $p_B = 0.090$, $p_{A_1B} = 0.024$, $p_{A_2B} = 0.006$, $p_0 = 0.460$	0.3423
MN	$p_{MN} = 0.498$, $p_M = 0.288$, $p_N = 0.214$	0.3769
Hp	$p_1 = 0.139$, $p_{2-1} = 0.526$, $p_2 = 0.320$, $p_0 = 0.015$, $p_{2-1M} = 0.015$	0.3986
PGM_1	$p_1 = 0.584$, $p_{2-1} = 0.361$, $p_2 = 0.053$	0.4744
AK	$p_1 = 0.912$, $p_{2-1} = 0.087$, $p_2 = 0.001$	0.8393
6PGD	$p_A = 0.958$, $p_C = 0.041$, $p_{CA} = 0.001$	0.9194

Table 4-2. Frequency of Phenotypes Determined in Blood Samples

Genetic Marker System	Bloodstain Clue		Suspect's Blood		Victim's Blood	
	Phenotype	Frequency	Phenotype	Frequency	Phenotype	Frequency
ABO	O	0.460	O	0.460	O	0.460
MN	MN	0.498	MN	0.498	M	0.288
Hp	2-1	0.526	2-1	0.526	2	0.320
PGM_1	1	0.584	1	0.584	2-1	0.361
AK	1	0.912	1	0.912	1	0.912
6PGD	A	0.958	A	0.958	A	0.958

matches the clue merely by chance is simply the probability of match or identity, which by equation (1) is $0.460 \times 0.460 \times 0.498 \times 0.526 \times 0.584 \times 0.912 \times 0.958 \simeq 0.060$. This probability is not significant enough to provide reasonable certainty that the suspect was involved.

Now let us suppose that the clue contained the rarest phenotype in each genetic marker system listed. The probability of a person chosen at random from the population matching the clue is given by $0.006 \times 0.214 \times 0.015 \times 0.053 \times 0.001 \times 0.001 = 1 \times 10^{-13}$. If a suspect was found having this blood type, it would be difficult to argue his innocence. To reiterate qualitatively, equation (1) states that the probability of a match is dependent upon (a) the rarity of the phenotype and (b) the number of genetic marker systems analyzed for.

If every blood specimen contained one or more rare phenotypes, it would be very useful to always analyze for rare phenotypes. However, because the rare phenotypes are difficult to find, numerous genetic marker systems must be analyzed for. Such a task is not within the time or sample size constraints of forensic serologists. The optimal approach is to analyze a blood sample for between 7 and 10 genetic marker systems, each containing between 2 and 5 equally distributed phenotypes.

That this approach is ideal is demonstrated as follows: Given that the phenotypes are equally distributed within each genetic marker system, the probability P, that two blood samples merely match by chance is given by the reciprocal of the number of phenotypes M raised to the exponent number N of the genetic marker systems analyzed, as shown in equation (2):

$$P = \left(\frac{1}{M}\right)^N \qquad (2)$$

Probability of match is shown in Table 4-3.

Table 4-3. Probability of Match

M, No. of Phenotypes	N, No. of Genetic Marker Systems								
	1	2	3	4	5	6	7	8	9
2	.500	.250	.125	.062	.031	.016	.008	.004	.002
3	.333	.111	.037	.012	.004	.001	.0003	.0001	.00005
4	.250	.063	.016	.004	.001	.0002	.00006	.00002	.000004
5	.200	.040	.008	.002	.0003	.00006	.00001	.000003	.0000005

This table shows the probability of a random match with a blood sample when N genetic marker systems are analyzed for with M equally distributed phenotypes in each system. Note that three genetic marker systems with five phenotypes in each provide the same probability of match as seven genetic marker systems with only two phenotypes each.

The selection of genetic marker systems using the statistical concept discrimination probability was discussed qualitatively and is discussed quantitatively below. The probability of occurrence (frequency of occurrence) of a phenotype j for a genetic marker system i in a given blood sample is denoted as p_{ij}. The probability p_i that any two blood samples chosen randomly from the entire population have the same phenotype j for a genetic marker system i is given by:

$$P_i = p_{ij}^2 \qquad (3)$$

Because every genetic marker system contains more than one phenotype, the probability that any two blood samples chosen randomly from the entire population match in any one of the phenotypes from 1 to M, for a genetic marker system having M phenotypes, is given by:

$$P_i = \sum_{j=1}^{M} p_{ij}^2 \qquad (4)$$

Equation (4) gives the discrimination probability for one genetic marker system. The discrimination probability for a set of genetic marker systems (i.e., the probability that any two blood samples chosen randomly from the entire population match in any one of the phenotypes, j(j = 1, ..., M) of each genetic marker system i (i = 1, ..., N)) is given by substituting equation (4) into the right-hand side of equation (1) and obtaining:

$$P = \prod_{i=1}^{N} \left(\sum_{j=1}^{M} p_{ij}^2 \right) \qquad (5)$$

The most efficient set (i.e., the best choice) of genetic marker systems is that which produces the smallest numerical value for the discrimination probability as defined in equation (5). The smaller the probability, the fewer the number of samples that will coincidently match and conversely, the greater the number of samples that do not match. Jones has mathematically proven that the discrimination probability, as defined in equation (5), decreases whenever (a) the phenotypes are more uniformly or equally distributed in each genetic marker system, (b) the number of phenotypes increases for a genetic marker system, and/or (c) the number of genetic marker systems analyzed for increases.[14] Note that by the way that discrimination probability is defined, better discrimination is given by smaller probabilities.

There are similarities and differences between probability of match and discrimination probability. Both probabilities, as expressed in equations (1) and (5), utilize the same frequency-of-occurrence data. The probability of match applies to the situation faced by the criminalistics laboratory in relating a blood clue found at a crime scene to a specific subject. The discrimination probability applies to the situation faced by the criminalistics laboratory in selecting efficient genetic marker systems with which to type a given bloodstain clue. In illustrating the difference between these concepts, Table 4-1, which gives an example of a set of genetic marker systems, contains calculations of the discrimination probability for each system. Table 4-4 shows how the probability of match is used to tie a bloodstain clue found on a suspect's clothing to the victim's blood.

B. PRACTICAL APPLICATIONS

The concept of selecting an efficient set of genetic markers is discussed in relation to applications to discriminate within ethnic groups, between ethnic groups, and between ethnic genders. This section ties together the phenotypic frequency-of-occurrence data collected during this study and the concept of discrimination probability to point out which genetic marker systems are most useful for these applications.

This study, as well as others, notes that a racial bias in phenotypes is manifested by the frequency-of-occurrence data. The distributions of phenotypes with respect to ethnic groups are shown in Figures 4-1 through 4-4. For example, it is clear in Figure 4-1 that there is a strong ethnic bias regarding the gene frequencies in the group specific component (Gc) polymorphic protein system. Although not as pronounced, ethnic bias is clear for all of the genetic marker systems illustrated in these figures.

Table 4-4. Example of Calculation of Probability of Match

Marker System	Phenotype*	p_{ij} = Frequency of Occurrence
ABO	O	0.460
MN	MN	0.498
Hp	2	0.526
PCE_2	C_5^-	0.903
PGM_1	1	0.584
	Product $\Pi\ p_{ij}$ =	0.0635**

*Common to bloodstain clue and victim

**Probability of Match

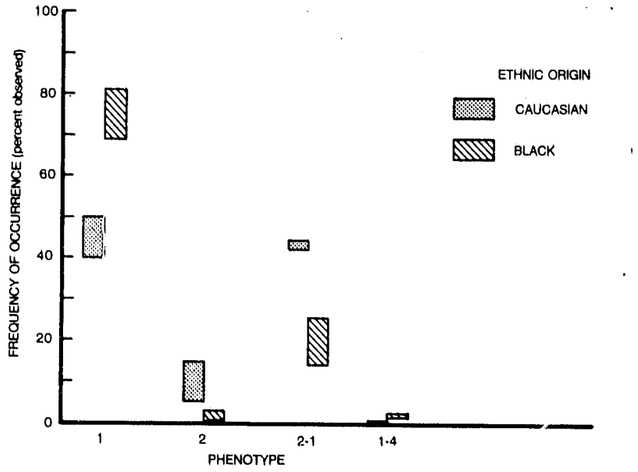


Figure 4-1. Relationship Between Frequency of Occurrence and Ethnic Origin (Group Specific Component (Gc))

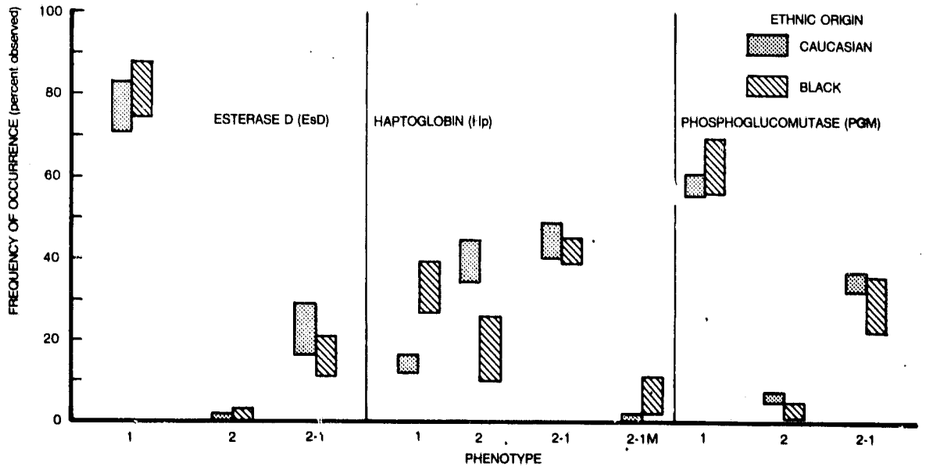


Figure 4-2. Relationship Between Frequency of Occurrence and Ethnic Origin (Esterase D(EsD), Haptoglobin (Hp), Phosphoglucomutase (PGM_1)

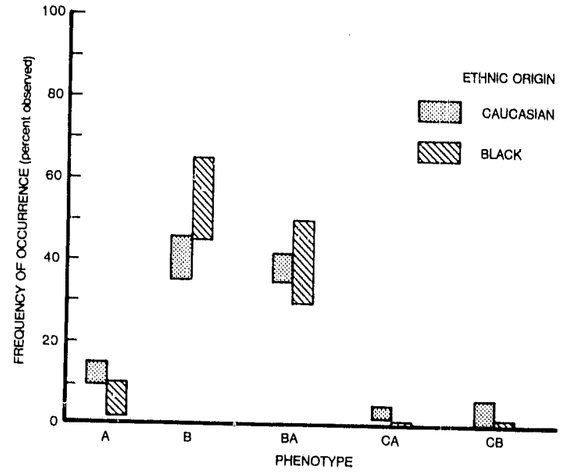


Figure 4-3. Relationship Between Frequency of Occurrence and Ethnic Origin (Erythrocyte Acid Phosphatase (EAP))

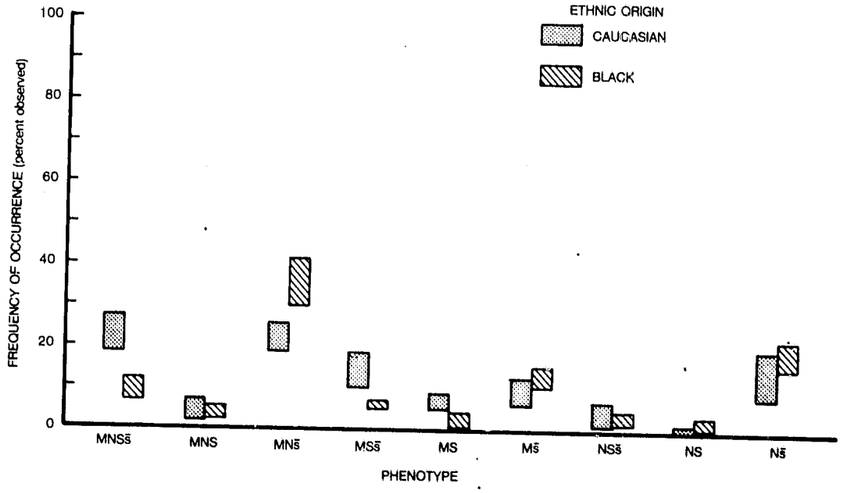


Figure 4-4. Relationship Between Frequency of Occurrence and Ethnic Origin (MNSs)

Visualizing phenotypes of the genetic markers graphically illustrates the practical aspects of using certain systems for differentiating bloods from individuals within ethnic groups or between ethnic groups.

Using the principle that the more uniformly distributed are the phenotypes of a genetic marker system, the better the discrimination probability of that genetic marker system, Figure 4-1 demonstrates Gc is better able to distinguish the blood of Caucasians than of Blacks. The ratios of the average phenotype frequencies for Caucasians are approximately 45:12:43 (for Gc phenotypes 1, 2 and 2-1, respectively) as opposed to 75:2:20:3 (for Gc 1, 2, 2-1, and 1-Y, respectively), for Blacks. The discrimination probability is calculated to be 0.39 for applying Gc to Caucasians and to be 0.56 for applying the system to Blacks. The same approach can be applied to each of the other genetic marker systems. Esterase D (EsD) and phosphoglucomutase-locus one (PGM_1) are better for discriminating blood of Caucasians, while haptoglobin (Hp) is better for discriminating blood of Blacks, as can be seen in Figure 4-2. Figure 4-3 shows erythrocyte acid phosphatase (EAP) can be equally applied to discriminating blood of both Caucasians and Blacks, although the ethnic bias for the different phenotypes is apparent. The MNSs genetic marker system is a good general discrimination system (discrimination probability is 0.37) but appears to be actually better for differentiating the blood of Caucasians than of Blacks.

The discrimination of blood within ethnic groups (e.g., among Caucasians or among Blacks) represents only one practical application of bloodstain analysis, but there is discrimination of blood between ethnic groups (e.g., between Caucasians and Blacks). The ethnic bias in the blood phenotypic frequency-of-occurrence data is not sharp enough to permit clear-cut decisions to be made regarding the ethnic origin of a bloodstain except for a small number of genetic marker systems. Three of these systems,

carbonic anhydrase (CA II), peptidase A (Pep-A), and hemoglobin (Hb) are illustrated in Figure 4-5. The information in Figure 4-5 strongly implies that the presence of phenotypes 2 or 2-1 for either CA II or Pep-A or AS or AC for Hb in a blood sample indicates that its ethnic origin is most probably Black. (Note that inferences about ethnic origin of a bloodstain based on phenotype information must be of a probabilistic nature.) It is noted that because the frequency of occurrence of the aforementioned particular phenotypes is not very great, their potential ability for discrimination of blood among Blacks is small.

Glucose-6-phosphate dehydrogenase (Figure 4-6) is an isoenzyme genetic marker system that provides the forensic serologist with an added dimension of blood discrimination. First, the A or BA phenotype of this system provides a high discrimination of blood between Blacks and Caucasians. This system, in fact, is the best genetic marker of the systems mentioned so far including CAII, Pep-A and Hb. Second, the BA phenotype of this genetic marker system provides a mechanism for singling out Black females.

A particular genetic marker system may have a very good discrimination probability and yet present many difficulties in usage. An example is the Rh system (Figure 4-7) which has a discrimination probability of 0.20 and has been espoused as one of the most efficient systems.[15] However, when the practical aspects of bloodstain analysis are considered, the Rh system loses a lot of its appeal as a viable tool for forensic serology because Rh blood typing is based on identifying both the presence and absence of particular antigens and the analytical procedures require obtaining a meaningful negative result. In wet or whole blood typing, the negative result is more or less conclusive; in forensic serology, typing is performed mostly on bloodstains, inherently complex. For a number of Rh antigens, a negative result due to degradation of an aged bloodstain cannot be distinguished from a

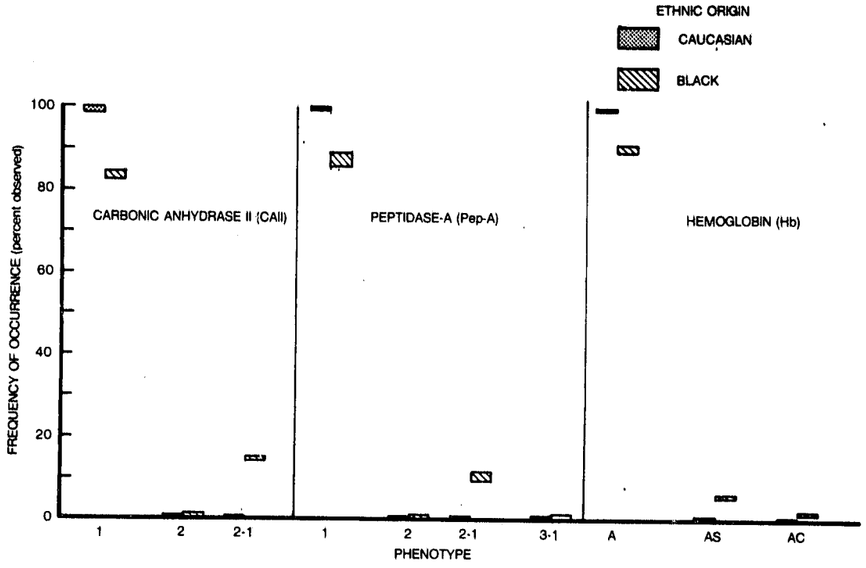


Figure 4-5. Relationship Between Frequency of Occurrence and Ethnic Origin (Carbonic Anhydrase II (CAII), Peptidase A (Pep-A), and Hemoglobin (Hb))

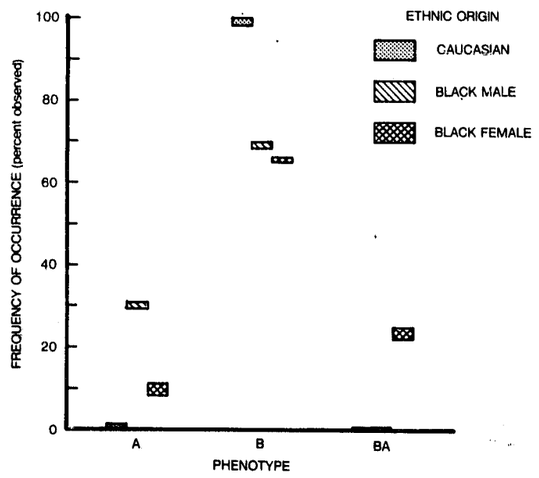


Figure 4-6. Relationship Between Frequency of Occurrence and Ethnic Origin (Glucose-6-Phosphate Dehydrogenase (G6PD))

negative result due to absence of an antigen. Thus, the only meaningful result for the Rh system is the presence of an antigen. Figure 4-8 shows the frequency-of-occurrence distribution for Rh system antigens for which meaningful typing can be performed by forensic serologists. Note that the discrimination probability of this genetic marker system worsens dramatically.

The utility of the Rh systems for discriminating blood within ethnic groups can also be seen in Figure 4-8. An ethnic bias is evident in the figure: The antigens D, c, and e occur in 90 to 100 percent of Blacks, and the antigens C and E occur in less than 30 percent of all Blacks. Thus, in one instance, almost every Black will give a positive result while in the other, the majority will be negative. It can be concluded that the utility of this system for discriminating blood among Blacks is minimal. The utility of the Rh system for discriminating blood among Caucasians is slightly better. The D, C, and c antigens occur with more uniformity (i.e., are more equally distributed--80, 65, and 80 percent, respectively) such that discrimination probability is enhanced. However, considerations of expenditures of time, money, and bloodstained material do not favor utilizing the Rh system in comparison with the other genetic marker systems with commensurate discrimination probability.

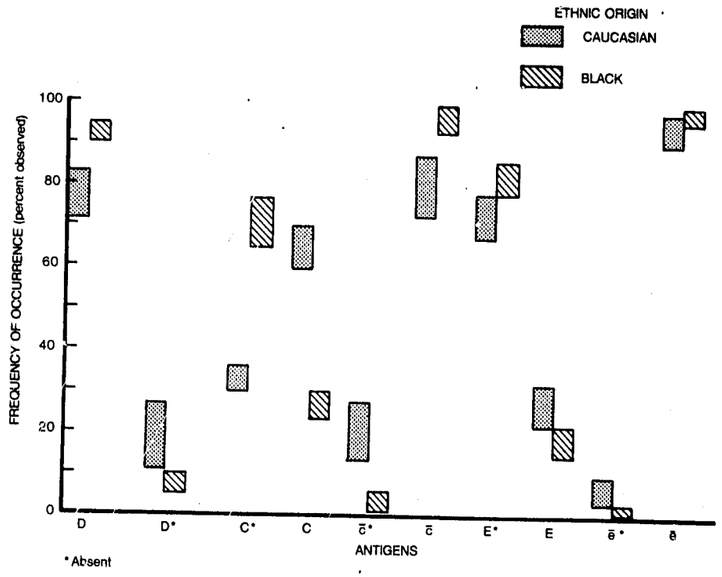


Figure 4-7. Relationship Between Frequency of Occurrence and Ethnic Origin (Rh (Antigens) D, D*, C*, C, $\bar{c}$ *, $\bar{c}$, E*, E, $\bar{e}$* $\bar{e}$)

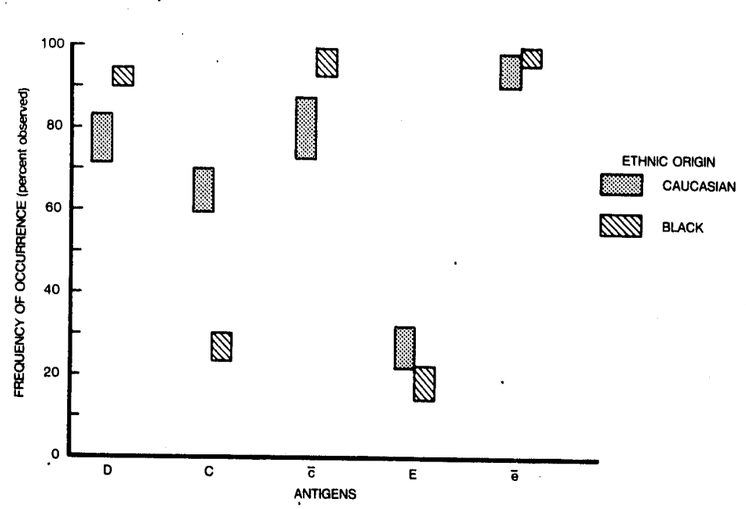


Figure 4-8. Relationship Between Frequency of Occurrence and Ethnic Origin (Rh (Antigens) of Forensic Utility, D, C, $\bar{c}$, E, $\bar{e}$)

V. REFERENCES

1. B.J. Culliford, "The Examination and Typing of Bloodstains in the Crime Laboratory," U.S. Department of Justice, Law Enforcement Assistance Administration, PR 71-7 (December 1971).

2. CU No. 14, 1970 Census, Population Research Section, Los Angeles County, Department of Regional Planning.

3. R.H. Osborne, Amer. J. Phys. Anthro., 16, 187 (1959).

4. W.J. Dixon, ed., "BMDP Biomedical Computer Programs," Berkeley, California, University of California Press (1975).

5. A.E. Mourant, ADA C. Kopeć, Kazimiera Domaniewska - Subczak, The Distribution of the Human Blood Groups and Other Polymorphisms, Oxford University Press, London (1976).

6. D.C. Shreffler, C.F. Sing, J.V. Neel, H. Gershowitz, and J.A. Napier, Amer. J. Hum. Genet., 23, 150-163 (1971)

7. T.E. Reed, Amer. J. Hum. Genet., 20(2), 142-150 (1968).

8. E.R. Giblett, Genetic Markers in Human Blood, Oxford and Edinburgh Blackwell Scientific Publications (1969).

9. R.C. Juberg, Amer. J. Hum. Genet., 22(1), 96-99 (1970).

10. J.C. Detter, G. Stamatoyannopoulos, E.R. Giblett, and A.G. Motulsky, J. Med. Genet., 7, 356-357 (1970).

11. G.J. Brewer, D.R. Bowbeer and R.E. Tashian, Acta Genet., Basel, 17, 97-103 (1967).

12. G.W. Karp and H.E. Sutton, Amer. J. Hum. Genet., 19, 54 (1967).

13. A.M. Hagins, R.C. Shaler, C.E. Mortimer, W.C. Stuver, and P.M. Neilson, J. For. Sci., 23(3), 563 (1978).

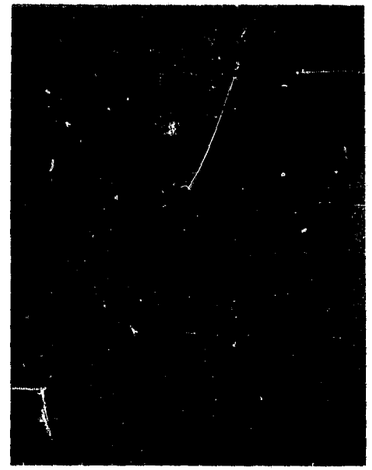
14. D.A. Jones, J. For. Sci. Soc., 12, 355 (1972).

15. G. Denault, H. Takimoto, and E.J. Rattin, "Survey and Assessment Blood and Bloodstain Analysis Program," U.S. Department of Justice, Law Enforcement Assistance Administration, The Aerospace Corporation Report No. ATR-74(7910)-1, Vol. 1, April 1974.

APPENDIX A

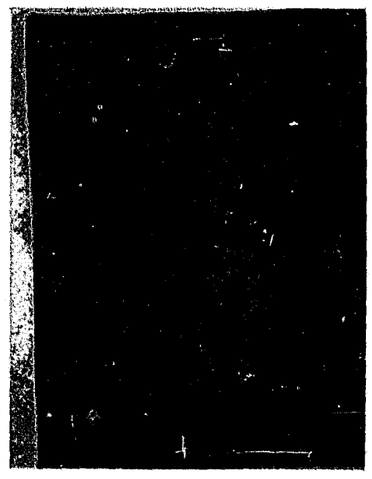
EXAMPLES OF ELECTROPHORETIC RESULTS

Adenosine Deaminase



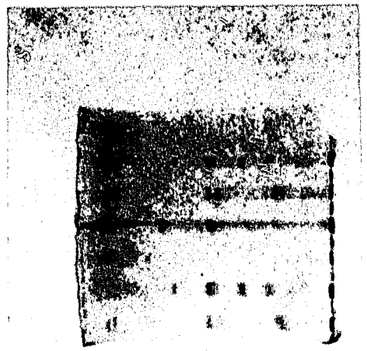
1
1
1
1
1
1
1
1
1
1
1
2-1
1

NEG
1
2
NEG
NEG
1
2
NEG

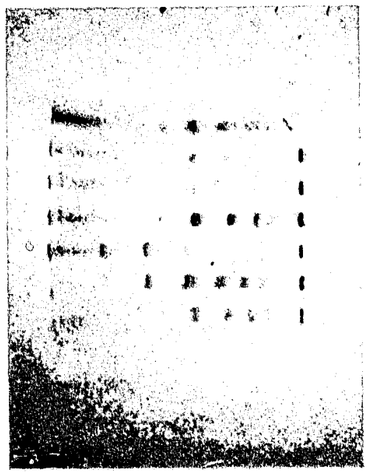


1
1
1
1
1
1
2-1
1
1
1
1
1
1

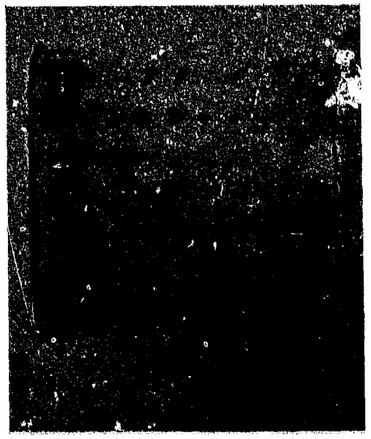
Haptoglobin



1
2
1-J?
2-1
1
1-J?



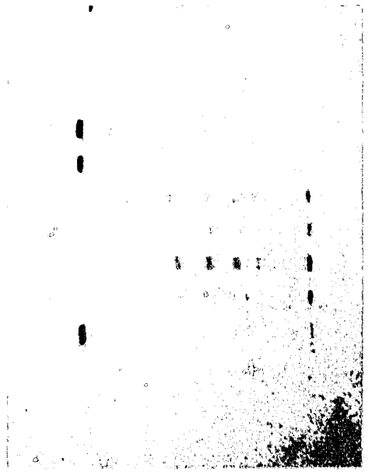
2
2
2
?
2
2-1M
2-1
2



2
2-1
1
2-1M
2-1
2
2

Haptoglobin

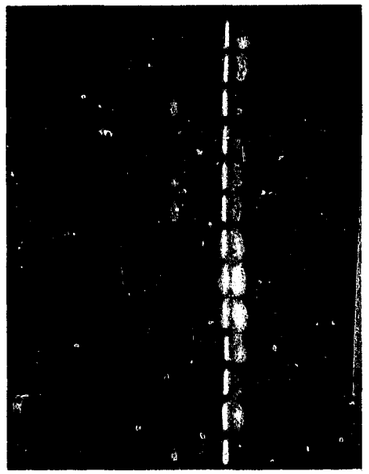
2-1
2-1M
2-1
2-1
2-1
1
1



1
1
2-1
1
2-1
2-1
2
2
1

2
2-1
2-1
1
2-1
2-1
2-1

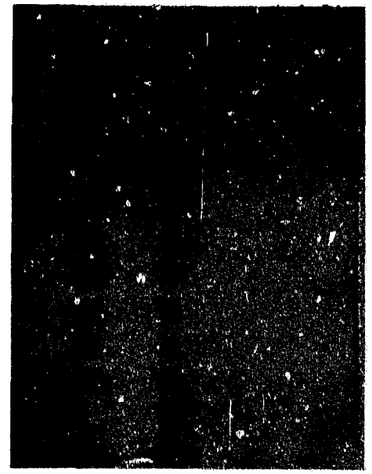
Carbonic Anhydrase II



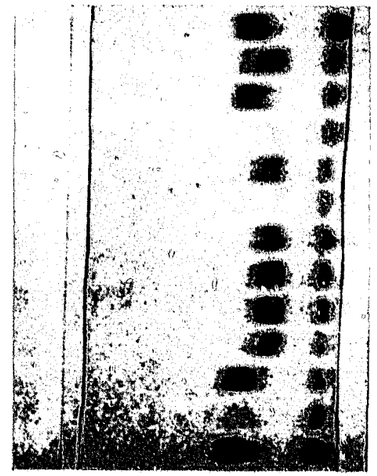
1
2-1
2-1
2-1
2-1
1
1
1
1
2-1
1
2

1
1
1
1
1
2-1
1
1
1
1
1
1
1

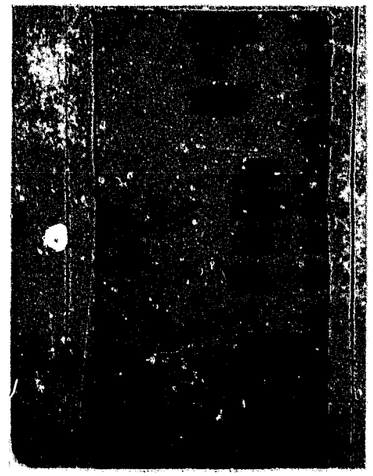
Glucose - 6 - Phosphate Dehydrogenase



B
B
B
A
B
B
B
B
B
B
B
B
A
B

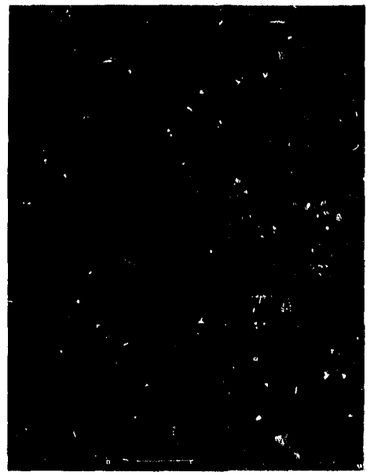


A
BA
A
B
B
B
B
A
A
A



A
A
A
B
B
B
B
B
B
B
B
B
A

6 - Phosphogluconate Dehydrogenase

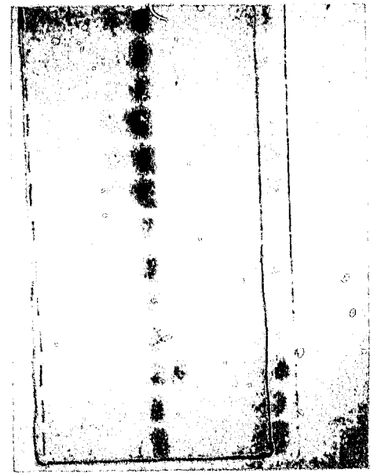


A
A
AC
A
A
A
A
A
A
A
C
A
A
A

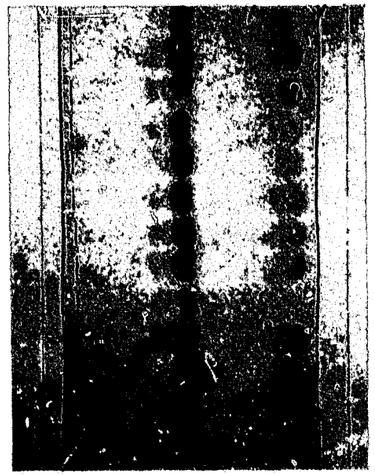
INC
AC
A
A
INC
AC
A
A

A
A
A
A
A
A
C
A
A
A
A
A
A
C
A

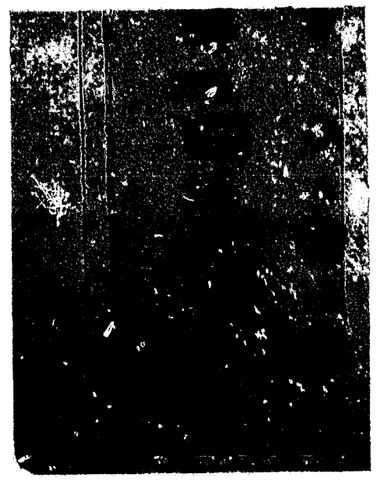
Adenylate Kinase



1
1
1
1
1
1
1
1
1
1
2-1
1
1



1
1
1
1
1
1
1
2-1
1
1

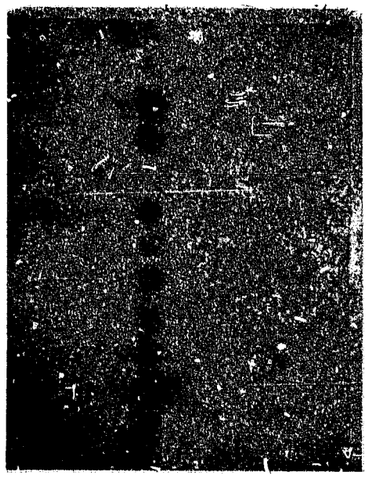


1
1
2-1
1
2-1
2-1
1
1

Peptidase A

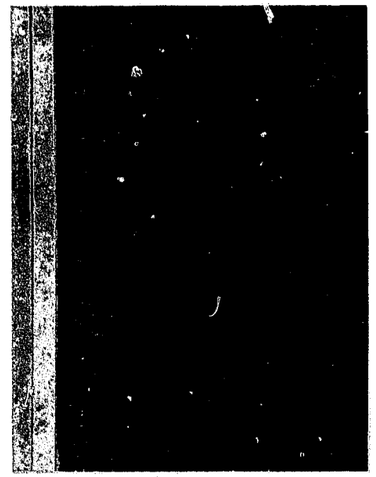
1
1
1
2-1
2-1
1
1
1
1
1
1
1
1

1
1
1
2-1
2-1
1
1
2
1
1
1
2-1
1

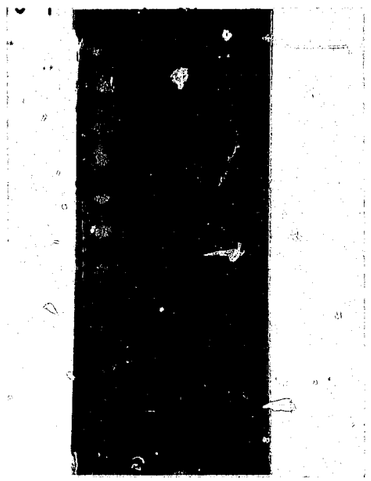


2-1
1
1
1
2
1
1
1
1
1
1
1
1

Glyoxalase - I



2
2
2
1
2
2-1
2
2
2-1
2-1
2-1
2
2-1

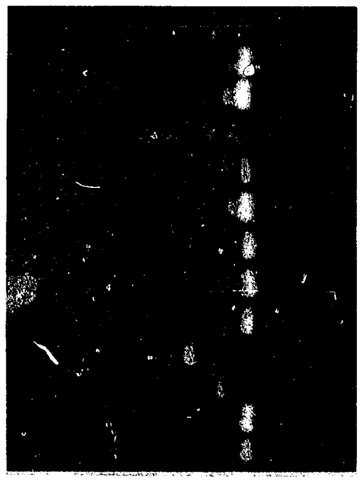


2
2-1
2-1
2-1
2
2-1
2
1
2
2
2
2-1
2-1

2-1
2-1
2
1
2-1
2-1
2-1
2-1
2
2
2
2-1
2

Esterase - D

1
1
1
1
1
2-1
1
1
1
2-1
1
1
2-1



2-1
1
1
2-1
1
1
1
1
1
2
2-1
1
1

1
2-1
1
1
1
1
1
2-1
1
2-1
1
1

A-10

Phosphoglucomutase

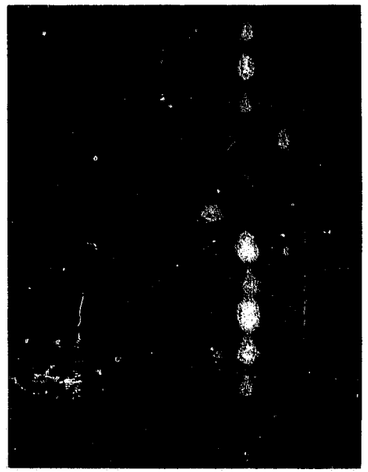
1
2-1
1
2-1
1
1
1
2
1
2-1
1
1
1

1
2-1
1
2-1
1
2-1
1
1
2-1
1
2-1

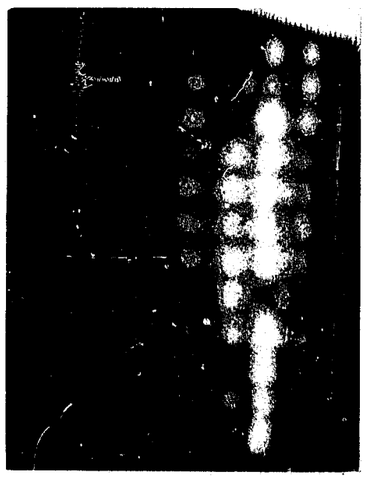
1
2
1
2-1
1
2-1
2-1
1
1
2-1
1
2-1
2-1

A-11

Erythrocyte Acid Phosphatase



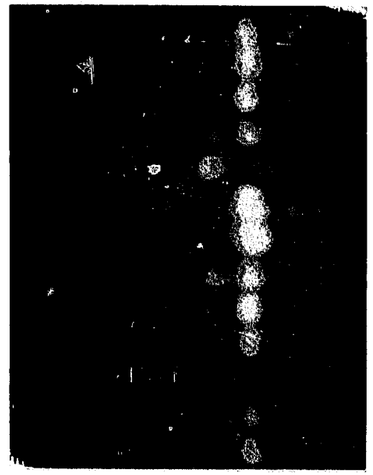
B
B
BA
CA
BA
A
B
BA
B
BA
BA
A
BA



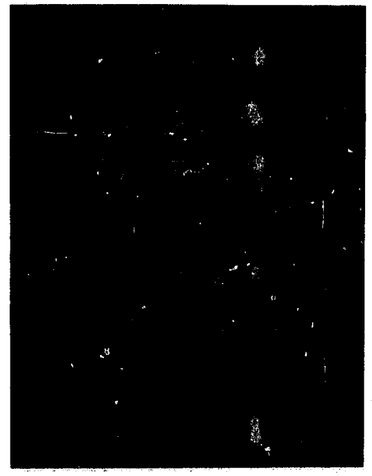
CB
RC
B
BA
BA
BA
BA
A
BA
B
BA
B

DB
RB
BA
BA

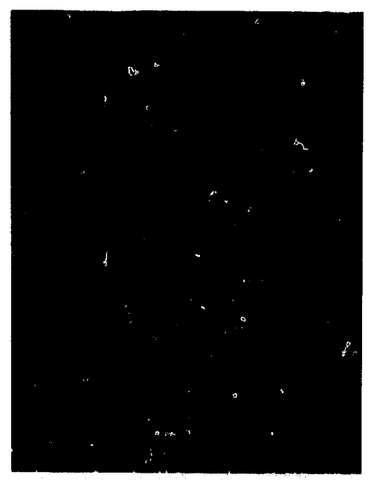
Erythrocyte Acid Phosphatase



CB
B
B
BA
A
B
B
BA
B
B
RA
BA
B

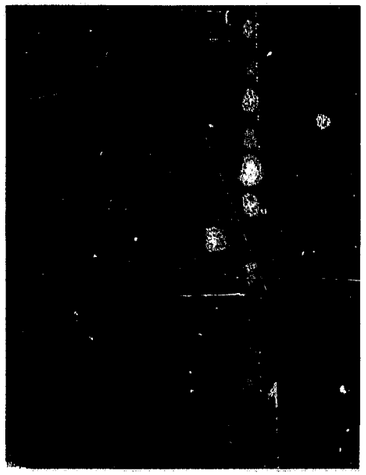


RB
RB
RB
A
BA
CA
A
CB

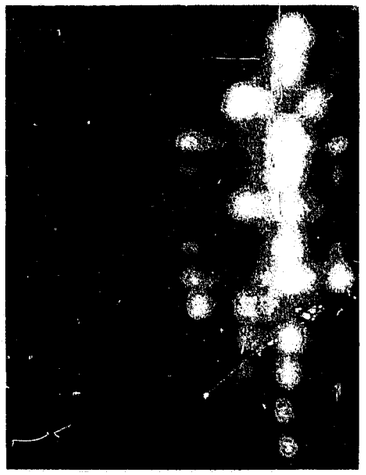


RB
BA
RB
RB
BA
RB

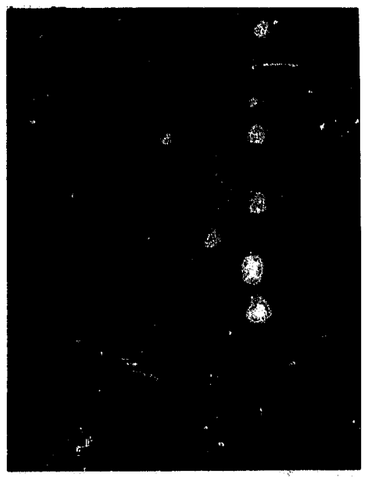
Erythrocyte Acid Phosphatase



B
BA
B
BA
B
B
A
BA
BA
RB
B
BA
B



B
B
A
B
B
BA
B
CB
RA
BA
BA
BA
BA



B
CA
BA
RB
A
BA
A
B
B
A
A
BA
CB

APPENDIX B

BLOOD/BLOODSTAIN ANALYSIS DATA FORM

BLOOD/BLOODSTAIN ANALYSIS DATA FORM

Central Facility Use Only

Central Facility Case Number _____
Entered _____
Verified _____

Key
Pres = Present
Abs = Absent

Data Name	Mnemonic	Data
Laboratory Case No.	CASE (1)	
Lab. No.	LAB (2)	
Date: Year	YR (3)	
Month	MNTH (4)	
Day	DAY (5)	
Gender	GEN (6)	[ ] (A) Male [ ] (B) Female
Ethnic Background	ETH (7)	[ ] (C) Caucasian [ ] (D) Am. Ind. [ ] (E) Oriental [ ] (F) Negro [ ] (G) Other _____ (Specify)
Place of Birth	POB (8)	(city, state) _____
Specimen Source	SRC (9)	[ ] (I) Whole Blood [ ] (J) Dried Blood
Multiplicity	MULT (10)	_____

ANTIGEN DATA

ABO	ABØ (11)	O (L)	A (M)	B (N)	AB (Ø)							
A Subgroups	ASUB (12)	1 (Q)	2 (R)	3 (S)	4 (T)	5 (U)	M (V)	X (W)				
Rh	RH (13)	D (Y)	Ø (Z)	C (A)	c (B)	E (C)	e (D)	C^w (E)	D^u (F)			
Secretor	SEC (14)	Pres (G)	Abs (H)									
Lewis	LEW (15)	Le^a (K)	Le^b (L)	Le^x (M)	Le^c (N)	MAG (Ø)	A_1Le^b (P)					
MN	MN (16)	M (S)	N (T)	S (U)	s (V)							
Dombrock	DØM (17)	Pres (Y)	Abs (Z)									
Duffy	DUFF (18)	Fy^a (A)	Fy^b (B)	Fy^{ab} (C)	Fy (D)	Fy^x (E)						
Kell	KELL (19)	K (H)	k (I)	K_p^a (J)	K_p^b (K)	Js^a (L)	Js^b (M)					
Lutheran	LU (20)	Lu^a (P)	Lu^b (Q)									
Kidd	KIDD (21)	Jk^a (T)	Jk^b (U)									
Diego	DGØ (22)	Di^a (X)	Di^b (Y)									
P	P (23)	P_1 (B)	P_2 (C)									
Vel	VEL (24)	Pres (F)	Abs (G)									
I	I (25)	Pres (I)	Abs (J)									
Yt	YT (26)	Yt^a (L)	Yt^b (M)									
X_g^a	XGA (27)	Pres (Ø)	Abs (P)									
HLA Locus A	HLAA (28)	A1 (R)	A2 (S)	A3 (T)	A9 (U)	A10 (V)	A11 (W)	A28 (X)	A29 (Y)	AW23 (Z)		
		AW24 (A)	AW26 (B)	AW30 (C)	AW31 (D)	AW32 (E)	AW33 (F)	AW34 (G)	AW36 (H)	AW43 (I)		
HLA Locus B	HLAB (29)	B5 (M)	B7 (N)	B8 (Ø)	B12 (P)	B13 (Q)	B14 (R)	B18 (S)	B27 (T)			
		BW15 (U)	BW16 (V)	BW17 (W)	BW21 (X)	BW22 (Y)	BW35 (Z)	BW37 (A)	BW38 (B)			
		BW39 (C)	BW40 (D)	BW41 (E)	BW42 (F)							

ENZYME/PROTEIN DATA

Data Name	Mnemonic	Data
ADA	ADA (30)	1 (L) 2 (P) 2-1 (R)
AK	AK (31)	1 (N) 2 (O) 2-1 (P)
EAP	EAP (32)	A (F) B (H) C (K) HA (S) CA (T) CB (V)
ESD	ESD (33)	1 (D) 2 (E) 2-1 (F)
G6PD	G6PD (34)	A (K) B (L) BA (M)
6PGD	6PGD (35)	A (Q) AC (R) C (S)
PCE 1	PCE1 (36)	E_1^1 (W) E_1^U (X) E_1^{Var} (Y)
PCE 2	PCE2 (37)	C_{6-} (H) C_5 (F)
PCE 3	PCE3 (38)	Pres (H) Abs (I)
GPT	GPT (39)	1 (L) 2 (M) 3 (N) 2-1 (O) 3-2 (P)
GR	GR (40)	1 (S) 2 (T) 2-1 (U)
PEP A	PEPA (41)	1 (X) 2 (Y) 3 (Z) 4 (A) 2-1 (B) 3-1 (C) 3-2 (D) 4-1 (E)
PEP B	PEPB (42)	1 (I) 2 (J) 3 (K) 4 (L) 2-1 (M) 3-1 (N) 3-2 (O) 4-1 (P)
PEP D	PEPD (43)	1 (S) 2 (T) 3 (U) 4 (V) 2-1 (W) 3-1 (X) 3-2 (Y) 4-1 (Z)
UTP	UTP (44)	Gt (C) Gt^D (D) Abs (E)
PGM_1	PGM1 (45)	1 (H) 2 (I) 2-1 (J) 3-1 (K) 3-2 (L) 4-1 (M) 4-2 (N) 5-1 (O) 5-2 (P) 6-1 (Q) 6-2 (R) 7-1 (S) 7-2 (T) 8-1 (U) 8-2 (V) 1AT (W) 2-1AT (X)
PGM_2	PGM2 (46)	1 (A) 2-1 (B) 3-1 (C) 4-1 (D) 5-1 (E)
PGM_3	PGM3 (47)	1 (I) 2 (J) 2-1 (K)
GAII	GA2 (48)	1 (N) 2 (O) 2-1 (P)
Haptoglobin	HP (49)	A (Q) AS (T) AC (U) S (V) SC (W) C (X) AD (Y) AK (Z) AG (A) A|5 (B) A|2 (C) AN10 (D) SG1 (E) AO1 (I)
Transferrin	TF (50)	1 (P) 2 (R) 2-1 (S) D (T) 2-1M (U)
[illegible]	[illegible] (51)	1 (Q) 2 (R) 3 (S)
[illegible]	[illegible] (52)	1 (V) 2 (W) 2-1 (X) X (Y) Y (Z) CHIP1 (A)
[illegible]	[illegible] (53)	Pres (J) Abs (K)
[illegible]	[illegible] (54)	Pres (H) Abs (I)
[illegible]	[illegible] (55)	1 (J) 2 (K) 3 (L) 4 (M) 5 (N) 6 (O) 7 (P) 8 (Q) 9 (R) 10 (S) 11 (T) 12 (U) 13 (V) 14 (W) 15 (X) 16 (Y) 17 (Z) 18 (A) 19 (B) 20 (C) 21 (D) 22 (E) 23 (F)

Comments _____

APPENDIX C

PROCEDURES OF LABORATORIES 4 AND 5

(Michigan State Police Crime Laboratory (4)
and Dade County Crime Laboratory (5))

LABORATORY NO. 4

	Glyoxalase I (GLO I)	Esterase D (ESD)
TANK BUFFER	Tris 0 10 M 42 39 g 12 11 g l Maleic Acid 0 10 M 40 67 g 11 62 g l EDTA (free acid) 0 01 M 10 22 g 2 92 g l $MgCl_2$ $6H_2O$ 0 01 M 7 10 g 2 03 g l Distilled H_2O 3 5 l Adjust to pH 7 4 with 40% NaOH **pH 7.4**	Tris 0 10 M 42 39 g 12 11 g l Maleic Acid 0 10 M 40 67 g 11 62 g l EDTA (free acid) 0 01 M 10 22 g 2 92 g l $MgCl_2$ $6H_2O$ 0 01 M 7 10 g 2 03 g l Distilled H_2O 3 5 l Adjust to pH 7 4 with 40% NaOH **pH 7.4**
GEL BUFFER	1 15 dilution of tank buffer **pH 7.4**	1 15 dilution of tank buffer **pH 7.4**
SUPPORT MEDIUM	1 mm gel	1 mm gel
AMOUNT REQUIRED	2% starch 1% agarose gel	2% starch 1% agarose gel
APPLICATION	LYSATE: 1 cotton thread soaked with lysate diluted 1 1 with 0 10 M Cleland's Reagent (dithiothreitol) STAIN: 1 to 2 threads of stained material soaked in a minimum of Cleland's Reagent (made at 0 10 M with gel buffer) for 5 to 10 minutes	LYSATE: 1 cotton thread soaked with lysate diluted 1 1 with 0 10 M Cleland's Reagent (dithiothreitol) STAIN: 1 to 2 threads of stained material soaked in a minimum of Cleland's Reagent (made at 0 10 M with gel buffer) for 5 to 10 minutes
TEMPERATURE AND CONDITIONS	Cold room at 4°C (or cooling plate) Migration Anodic	Cold room at 4°C (or cooling plate) Migration Anodic
VOLTAGE AND DURATION	Day 250 V for 3-3 4 hours Night	Day: 250 V for 3-3/4 hours Night
REACTION BUFFER	Same as 6-PGD tank buffer **pH 6.8**	Sodium acetate 2 05 g Distilled H_2O 500 ml Adjust to pH 6 0 with 0 1 N Acetic Acid **pH 6.0**
REACTION MIXTURE	(a) Agar (powder) 0 40 g GLO reaction buffer 15 0 ml Heat agar and standby Glutathione reduced 12 0 mg Methyl glyoxal 200 l GLO reaction buffer 5 0 ml Cool agar to 65°C, add glutathione solution, pour over gel Incubate 37°C for 30 minutes (CRITICAL)	(a) Reaction buffer 10 ml (b) 4-methyl umbelliferyl acetate 0 004 g Dissolved in a minimum of acetone Add (a) to (b) and saturate a 5" x 5" inch piece of Whatman 3MM chromatographic paper with the reaction mixture Place paper across plate for 5 minutes at room temperature
	Remove glutathione gel, then (b) Agarose, Type II 0 20 g Distilled H_2O 20 ml Heat to boiling hot Cool to 65°C and add KI I_2 solution 200 l (if too dark, use 150 l) Pour from corner of gel (on cooling plate) KI I_2 solution KI 1 65 g I_2 2 54 g Distilled H_2O 30 0 ml GLO can be run simultaneously with EsD and PGM	Following the 5 to 10 minute reaction period remove the Whatman 3MM overlay and read the results under long-wave UV Acetone is used in the reaction mixture merely to facilitate dissolving the 4-methyl umbelliferyl acetate into solution After the drop of acetone has been added to the 4-methyl umbelliferyl acetate the sodium acetate solution should be added promptly before the acetone can evaporate EsD can be run simultaneously with PGM and GLO

LABORATORY NO. 4

Peptidase A (PEPA)

TANK BUFFER	(a) Tris (b) NaH_2PO_4 (anhydrouse) Add (b) to (a) to reach pH 7.4	0.10 M 0.10M	12.11g/1.0 liter H_2O 18.0g/1.5 liter H_2O **pH 7.4**
GEL BUFFER	(a) Tris (b) Maleic acid Add (b) to (a) to reach pH 7.5	0.01 M 0.01 M	0.61g 500 ml H_2O) 0.58g 500 ml H_2O) **pH 7.5**
SUPPORT MEDIUM	1 mm starch gel		
AMOUNT REQUIRED	10% starch		
APPLICATION	LYSATE 2 cotton threads soaked with lysate	STAIN 2 to 4 stained threads soaked in a minimum of gel buffer for 10 minutes	
TEMPERATURE AND CONDITIONS	Cold room (or cooling plate) Migration: Anodic		
VOLTAGE AND DURATION	Day	Night: 90 V (4.5V/cm) overnight	
REACTION BUFFER	Na_2HPO_4 Adjust to pH 7.5 with 1.0 M HCL	0.2M	14.19g 500 ml H_2O **pH 7.5**
REACTION MIXTURE*	L-valyl-l-leucine $MnCl_2$ 1ml o-dianisidine dihydrochloride (from stock solution of 500mg 20.0ml H_2O) 2ml peroxidase (stock solution = 50mg 2.0ml reaction buffer*) snake venom amino acid oxidase(stock solution = 100mg 2.0ml reaction buffer) .05ml reaction buffer 2.0% agar gel	0.1 M 12.5ml 12.5g	10.0mg 0.25 ml (*for SIGMA P 8000)
NOTES	*The components of the reaction mixture should be mixed in the order which appears above. Upon addition of these components, the formation of a milky precipitate may occur but it is not inimical to reactivity. Rubber gloves must be worn when preparing the reaction mixture and applying it to the plate, due to the amino acid oxidase activity of snake venom. The origin slots should be placed in the gel approximately 1½ inches from the cathodic end of the plate rather than 1/3 the way from the cathode end because the PEP A bands migrate a considerable distance by this method. Because the PEP A bands diffuse quickly, the reaction mixture should be prepared prior to terminating electrophoresis and removal of the plate from the tank.		

Hemoglobin (Hb)

TANK BUFFER **Cathodic pH 8.6**	Tris EDTA (free acid) Boric Acid Distilled H_2O Sodium barbital Barbital Distilled H_2O	2.08×10^{-1} M 3.55×10^{-3} M 3.07×10^{-2} M 3.43×10^{-2} M 7.2×10^{-2} M	25.2 g 2.5 g 1.9 g 1.0 l 5.15 g 0.92 g 1.0 l	Adjust to pH 9.1 with 20% NaOH Adjust to pH 8.6 with 1.0 M NaOH **Anodic pH 9.1**
SOAK BUFFER	*Prepare a solution consisting of:* Anodic tank buffer Cathodic tank buffer		50 ml 50 ml	
SUPPORT MEDIUM	Cellulose acetate			
AMOUNT REQUIRED	12x5 cm strip			
APPLICATION*	LYSATE Apply neat lysate with a single 1-cm long stroke of a capillary pipette		STAIN Extract stained material with 1 drop of strip buffer. Apply extract to cellulose acetate with a single 1 cm long stroke of a capillary pipette	
TEMPERATURE AND CONDITIONS	Room temperature Migration: Anodic			
VOLTAGE AND DURATION	Day: 160 V for 1 hour		Night	
ALTERNATE BUFFER	Beckman "B-3" Glycine Tris Distilled H_2O	 0.29 M 3.72 M	 76.3 g 15.75 g 3.5 l	 21.8 g/l 4.5 g/l
REACTION MIXTURE	Ponseau S 3% Trichloracetic acid Immerse cellulose acetate strip in Ponseau S stain for 4 minutes. Wash strip in 3% acetic acid for 2 to 3 minutes. Rinse off acetic acid under tap water and allow strip to air-dry			0.15 g 100.0 ml
NOTES	*The cellulose acetate strip is floated on top of the solution of soak buffer until soaking is visible, then the strip should be plunged into the soak buffer. When samples are ready to be applied, blot the strip on paper towel and lay the strip across bridges made of several thicknesses of filter paper or Whatman 3MM chromatographic paper. Strips of cellulose acetate, 8x2 cm, soaked in soak buffer and blotted, should be used as edge strips for the main strip, that is, laid across the bridges and running along the top edge and bottom edge of the main strip. The shape of the main strip with edge strips in place is that of the capital letter "I". Washing off the acetic acid with tap water prevents the strip from becoming brittle when dried. The edge strips prevent distortion on the Hb bands during electrophoresis.			

LABORATORY NO. 4

	6-Phosphogluconate Dehydrogenase (6-PGD)	Carbonic Anhydrase II (CAII)
TANK BUFFER	(a) Na_2HPO_4 (anhydrous) 0.20M 56.784g, 28.392g/l Distilled H_2O 2.0 l (b) NaH_2PO_4 (anhydrous) 0.20 M 48.0g, 24.0g/l Distilled H_2O 2.0 l Add (b) to (a) to reach pH 6.8 **pH 6.8**	Tris 8.87×10^{-2} M 37.63 g 10.75 g/l Disodium EDTA 2.5×10^{-3} M 3.26 g 0.93 g/l Boric Acid 8.15×10^{-2} M 17.64 g 5.04 g/l Distilled H_2O 3.5 l (Same as Stock Hp tank buffer for polyacrylamide method) Hp buffer diluted 1:5 **pH 8.2**
GEL BUFFER	1/20 dilution of tank buffer **pH 6.8**	Tris 4.5×10^{-2} M 1.363 g EDTA (free acid) 1.0×10^{-3} M 0.073 g Boric Acid 2.5×10^{-2} M 0.387 g Distilled H_2O 250 ml **pH 8.5**
SUPPORT MEDIUM	1 mm starch gel	1 mm Starch Gel
AMOUNT REQUIRED	10% starch gel	10% Starch Gel (2.0 mg NADP per plate added to gel**)
APPLICATION	LYSATE: 1 cotton thread soaked with neat lysate STAIN: 2 to 4 threads of stained material soaked in a minimum of Cleland's Reagent (dithiothreitol) made at 0.05 M with gel buffer for 5 to 10 minutes	LYSATE: 1 cotton thread soaked with neat lysate STAIN: 1 to 2 threads of stained material soaked in a minimum of gel buffer for 5 to 10 minutes
TEMPERATURE AND CONDITIONS	Cooling plate, 4°C. Migration: Anodic	For day run: Cooling plate, 4°C For night run: Cooling plate or cold room at 4°C Migration: See notes
VOLTAGE AND DURATION	Day: 230V (11.5 V/cm) for 4 hours Night	Day: 330 V (15 V cm) for 4½ hours Night: 100 V (5.4 V cm) overnight
REACTION BUFFER	Tris 0.02M 6.07g Distilled H_2O 250ml Adjust to pH 8.0 with dilute HCL (HCL diluted 1:5 with H_2O), then Add $MgCl_2$ 1.0g **pH 8.0**	(a) NaH_2PO_4 (anhydrous) 0.1 M 18.9 l 5.1 H_2O (b) Na_2HPO_4 (anhydrous) 0.1 M 9.937 g 0.7 1 H_2O Add (b) to (a) to reach pH 6.5 Same as ADA tank buffer **pH 6.5**
REACTION MIXTURE	2.0% agar gel 10g 6-Phosphogluconate 3.0mg NADP 2.0mg PMS 1.0mg MTT 2.0mg Reaction buffer 10.0ml	Fluorescein diacetate 1.0mg Dissolve Fluorescein diacetate in 2 drops of acetone and before acetone can evaporate, add Reaction buffer 5.0 ml Saturate a 3x5½ inch piece of Whatman 3 MM chromatographic paper with the reaction mixture. Place paper across plate and incubate at 37°C for 10 minutes
NOTES	Incubate at 37°C for 30 to 60 minutes If run in conjunction with ADA, use ADA parameters (except use 80V) and include the following ingredients into the ADA reaction mixture 6-Phosphogluconate 6.0mg NADP 4.0 mg(minimum)	*Run CA II simultaneously with G-6-PD **The 2.0 mg NADP in the gel is required for G-6-PD analysis and must be added to the starch gel solution following the boiling stage and prior to the degassing stage Migration of CA II bands in this method is approximately 1 inch anodic to the origin for the CA II - 2 band and approximately 1.2 inch cathodic to the origin for the CA II - 1 band. Therefore the paper containing the reaction mixture must be placed onto the plate directly over the origin such that the origin bisects the paper overlay. Following the 10-minute reaction period, remove the paper overlay and read the results under long-wave UV

Glucose-6-Phosphate Dehydrogenase (G-6-PD)

	LABORATORY 4 *	LABORATORY 5
TANK BUFFER	Tris 8.87x10^{-2} M 37.63 g 10.75 g/l Disodium EDTA 2.5x10^{-3} M 3.26 g 0.93 g/l Boric acid 8.15x10^{-2} M 17.64 g 5.04 g/l Distilled H_2O 3.5 l (Same as stock Hp tank buffer for polyacrylamide method) **pH 8.2**	Trizma base (sigma T-1503) 10.75 g Disodium EDTA (sigma Stock ED2SS) 0.93 g Boric acid (sigma B-0252) 5.04 g Distilled H_2O up to 1 liter **pH 8.2**
GEL BUFFER	Tris 4.5x10^{-2} M 1.363 g EDTA (free acid) 1.0x10^{-3} M 0.073 g Boric acid 2.5x10^{-2} M 0.387 g Distilled H_2O 250 ml **pH 8.5**	Trizma base 5.45 g Free acid EDTA (sigma Stock EDS) 0.29 g Boric acid 1.55 g Distilled H_2O up to 1 liter **pH 8.5**
SUPPORT MEDIUM	1 mm starch gel	2 mm starch gel (starch sigma S-4501)
AMOUNT REQUIRED	10% starch gel (2.0 mg NADP per plate added to gel**)	10% starch gel (10g/100ml) 2 mg NADP must be added*
APPLICATION	LYSATE 1 cotton thread soaked with neat lysate STAIN 1 to 2 threads of stained material soaked in a minimum of gel buffer for 5 to 10 minutes	LYSATE STAIN 2 x 7 mm washed red cell stain soaked 10 minutes in gel buffer containing 7 mg ml NADP.
TEMPERATURE AND CONDITIONS	For day run: Cooling plate, 4°C. For night run: Colling plate or cold room at 4°C Migration: Anodic	Cooling plate at 4°C. Migration: Anodic
VOLTAGE AND DURATION	Day: 330 V (15 V/cm) for 4½ hours Night: 100 V (5V/cm) overnight	Day: 300 V for 5 hours Night: 100 V overnight
REACTION BUFFER	Tris 0.10 M 3.03 g Distilled 250 ml Adjust to pH 8.0 with dil. HCl (HCl diluted 1:5 with H_2O) Then, Add: $MgCl_2$ 1.0g **pH 8.0**	Trizma base 12.2 g Distilled H_2O up to 1 liter Adjust to pH 8.0 with 1:5 HCl. Then add 4 g $MgCl_2$ **pH 8.0**
REACTION MIXTURE	2.0% agar gel 10.0 g Glucose-6-phosphate 10.0 g NADP 4.0 mg (minimum) PMS 1.5 mg MTT 2.0 mg Reaction buffer 10.0 ml	(a) D-Glucose-6-phospate (sigma G-7879) 25 mg Nucleoside adenine diphosphate (sigma N-0505) 10 mg MTT tetrazdium (sigma M-2128) 5 mg Phenazine methosulfate (sigma P-9625) 2 mg Reaction buffer 25 ml (b) Type 1 agarose (sigma A-6013) 0.5 g Distilled H_2O 25 ml
NOTES	*Run G-6-PD simultaneously with CA II. **2.0 mg of NADP must be added to the starch gel solution following the boiling stage and prior to the degassing stage. Incubate in moisture chamber at 37°C for 15 minutes, read the results of the G-6-PD portion of the gel, and return the plate to the incubator for completion of the CA II reaction.	*2 mg NADP added to gel solution following the boiling stage but prior to degassing state. Incubate for 30 minutes at 37°C

LABORATORIES 4 AND 5

TANK BUFFER	Glycine 0.29 M 76.3 g 21.8 g/l Tris 3.72 M 15.75 g 4.5 g/l Distilled H_2O 3.5 l **pH 8.4**
SOAK BUFFER	Same as tank buffer **pH 8.4**
SUPPORT MEDIUM	Cellulose Acetate Membrane
AMOUNT REQUIRED	
APPLICATION	SERUM: Apply neat serum with a single 1 cm long stoke of a capillary pipette STAIN: Extract stained material with 1 or 2 drops of H_2O from a double-drawn pipette, extract this solution with a 4X vol. of $CHCL_3$. Apply as serum.
TEMPERATURE AND CONDITIONS	Room temperature. Migration: Anodic
VOLTAGE AND DURATION	Day: 390 V for 35 minutes. Night:
REACTION BUFFER	Anti-human Gc globulin (Atlantic Antibodies) Immerse membrane in 1:6 dilution* (0.85% isotonic saline) for 5 minutes Wash membrane 30 minutes in agitated saline bath *Dilution depends on batch of antisera
REACTION MIXTURE	0.1% Coomassie Blue R in alcohol:acetic acid solution (w/v) Alcohol 50 ml Acetic acid 10 ml Distilled H_2O 50 ml
NOTES	Wash saline off of membrane with distilled water Stain membrane 3 to 4 minutes Destain with alcohol:acetic acid until contrast is developed

	LABORATORY 4	LABORATORY 5
TANK BUFFER	Citric acid (monohydrate) 137 M 100 52g, 28 72g/l NaOH 267 M 37 4g. 10 7 g l Distilled H_2O 3 5 l Adjust to pH 4 9 with 40% NaOH **pH 4,9**	*Citric acid anhydrous (sigma C-0759)* 26 3 g Sodium hydroxide (Mallinckrodt 7708) 10 7 g Distilled H_2O *up to 1 liter* **pH 4.9**
GEL BUFFER	Succinic acid 0 0164 M 0 945 g Tris 0 0184 M 1 113 g Distilled H_2O 500 ml **pH 5.0**	L-Histidine (sigma H-8000) 0 79 g Distilled H_2O up to 1 liter Adjust to pH 7 0 with concentrated NaOH **pH 7.0**
SUPPORT MEDIUM	*1 mm starch gel*	2 mm starch gel (starch sigma S-4501)
AMOUNT REQUIRED	10% starch gel	10% starch gel (10g 100ml)
APPLICATION	LYSATE: *1 cotton thread soaked in neat lysate* STAIN: 1 thread of stained material soaked in a minimum of gel buffer for 5 to 10 minutes 2 *threads for weak stains*	LYSATE STAIN: 2 x 7 mm washed red cell stain soaked 5 minutes in gel buffer
TEMPERATURE AND CONDITIONS	Cooling plate (or cold room at 4°C*) Migration Cathodic	Cooling plate at 4°C *Migration Cathodic*
VOLTAGE AND DURATION	Day 250 V (12 5 V cm) for 4 hours Night	Day 300 V for 4 hours Night
REACTION BUFFER	Tris 0 10 M 3 03 g Distilled H_2O 250 ml Adjust to pH 8 0 with dilute HCL (HCL diluted 1 5 with H_2O) Then. Add $MgCl_2$ 1 00 g **pH 8.0**	Trizma base (sigma T-1503) 6 06 g Distilled H_2O up to 500 ml Adjust to pH 8 0 with 1 5 HCl Then add 2 g MgCl **pH 8.0**
REACTION MIXTURE	2 0% agar gel 10 g Glucose 18 0 mg Adenosine-5-diphosphate 10 mg (Boehringer) NADP 3 1 mg (minimum) PMS 2 5 mg MTT 4 0 mg Glucose-6-phosphate dehydrogenase (0 5 u plate)20 microliter Hexokinase (1 0 u plate)20 microliter Reaction buffer 10 ml	(a) Glucose (sigma G-5000) 45 mg Adenosine-5'-diphosphate (sigma A-2754) 12 mg Nucleoside adenine diphosphate (sigma N-0505) 10 mg MTT Tetrazolium (sigma M-2128) 10 mg Phenazine methosulfate (sigma P-9625) 2 mg Hexokinase-glucose-6-phospate dehydrogenase (sigma H-8629) 1 mg Reaction buffer 25 ml (b) Type 1 agarose (sigma A-6013) 0 5 g Distilled H_2O 25 ml
NOTES	*If the electrophoresis tank is refrigerated or placed in cold room during electrophoresis, water will condense on the undersides of the top plate and must be wiped away about once per hour Any condensate dropping on the starch gel will result in a localized field disruption and cause a severe distortion of AK bands in the region of the water drop Please remember to disconnect power before drying the top plate and reconnecting power each time* Use a cooling plate in a place at room temperature to alleviate the problem of condensation Remember that AK bands migrate toward the cathode by this method	Spot samples in center of plate to detect variants

PHOSPHOGLUCOMUTASE (PGM)

	LABORATORY 4	LABORATORY 5
TANK BUFFER	Tris 0.10 M — 42.39 g — 12.11 g/l Maleic acid 0.10 M — 40.67 g — 11.62 g/l EDTA (free acid) 0.01 M — 10.22 g — 2.92 g/l $MgCl_2 6H_2O$ 0.01M — 7.10 g — 2.03 g/l Distilled H_2O — 3.5 l Adjust to pH 7.4 with 40% NaOH — **pH 7.4**	Trizma base (sigma T-1503) 12.11 g Maleic acid (sigma M-0375) 11.62 g Free acid EDTA (sigma Stock EDS) 2.92 g $MgCl_2 6H_2O$ (sigma M-0250) 8.12 g Distilled H_2O up to 1 liter Adjust to pH 7.4 with 40% NaOH — **pH 7.4**
GEL BUFFER	1:15 dilution of tank buffer — **pH 7.4**	1:15 dilution of tank buffer — **pH 7.4**
SUPPORT MEDIUM	1 mm gel	2 mm starch gel (starch sigma S-4501)
AMOUNT REQUIRED	2% starch 1% agarose gel	10% starch gel (10g 100ml)
APPLICATION	LYSATE: 1 cotton thread soaked with neat lysate. Use Cleland's Reagent if done in conjunction with EsD or GLO STAINE: 1 to 2 threads of stained material soaked in a minimum of gel buffer for 5 to 10 minutes	2x7 mm washed red cell stain soaked 5 minutes in gel buffer
TEMPERATURE AND CONDITIONS	Cold room at 4°C (or cooling plate) Migration: Anodic	Cooling plate at 4°C Migration: Anodic
VOLTAGE AND DURATION	Day: 250 V for 3-3.4 hours — Night	Day — Night: 120 V overnight
REACTION BUFFER	Tris 0.06M — 1.82 g Distilled H_2O — 250 ml Adjust to pH 8.0 with dil. HCl (HCl diluted 1:5 with H_2O). Then: Add $MgCl_2$ — 0.50 g — **pH 8.0**	Trizma base 7.28 g Distilled H_2O up to 1 liter Adjust to pH 8.0 with 1:5 HCl. Then add 1g $MgCl_2$ **pH 8.0**
REACTION MIXTURE	2.0% agar gel — 10 g Glucose-1-phosphate (containing at least 1% glucose-1,6-diphosphate) — 30.0 mg NADP — 1.5 mg (minimum) PMS — 2.5 mg MTT — 4.0 mg Glucose-6-phosphate dehydrogenase — 20 microliter (1.0 unit/plate) Reaction buffer — 10.0 ml	(a) Glucose-1-phosphate (Wessex Biochemical Limited WB36) 75 mg Nucleoside adenine diphosphate (sigma N-0505) 7 mg MTT Tetrazolium (sigma M-2128) 10 mg Phenazine methosulfate (sigma P-9625) 2 mg Glucose-6-phosphate dehydrogenase (sigma G-7878) 1 mg Reaction buffer 25 ml (b) Type 1 agarose (sigma A-6013) 0.5 g Distilled H_2O 25 ml
NOTES	Incubatae at 37°C for 1 hour Place origin 4 cm from the cathode end If doing PGM, EsD, and GLO together, first develop for EsD. After filter paper for EsD has been removed, pour GLO (I) overlay. Then pour PGM overlay. Place in incubator for 30 minutes. Then remove and pour second GLO overlay	Incubate for one hour at 37°C

	LABORATORY 4	LABORATORY 5
TANK BUFFER	0 1 M Phosphate buffer (a) NaH_2PO_4 (anhydrous) 36 0 g 3 0 l (b) Na_2HPO_4 (anhydrous) 19 87 g 1 4 l Add (b) to (a) to reach pH 6 5 **pH 6,5**	(a) Anhydrous NaH_2PO_4 (sigma S-0751) 23 40 g Distilled H_2O up to 1500 ml (b) Anhydrous Na_2HPO_4 (sigma S-0876) 9 94 g Distilled H_2O up to 700 ml Add (b) to (a) to reach pH 6 4 **pH 6.5**
GEL BUFFER	1 10 dilution of tank buffer **pH 6,5**	1 10 dilution of tank buffer **pH 6.5**
SUPPORT MEDIUM	1 mm starch gel	2 mm starch gel (starch sigma S-4501)
AMOUNT REQUIRED	12% starch gel	10% starch gel (10g 100 ml)
APPLICATION	LYSATE Fresh lysate 1 cotton thread soaked in neat lysate Stored lysate Mix lysate 2 1 with 1 10 dilution of -mercapto-ethanol or 1 1 with 0 05 M Cleland's Reagent (Dithiothreitol) Use 2 cotton threads STAIN Extract stains in a minimum of -mercaptoethanol diluted 1 80 with gel buffer for 30 to 60 minutes and absorb into 2 to 4 clean cotton threads	LYSATE STAIN 2x7 mm washed red cell stain soaked 5 minutes in gel buffer containing 7 mg ml Cleland's Reagent
TEMPERATURE AND CONDITIONS	Cold room at 4°C (or cooling plate) Migration Anodic	Cooling plate at 4°C Migration Anodic
VOLTAGE AND DURATION	Day Night 60 V (3V cm) overnight	Day 300 V for 4 hours Night 60 V overnight
REACTION BUFFER	(a) NaH_2PO_4 (anhydrous) 2.56×10^{-2}M 0 154 g 50 ml H_2O (b) Na_2HPO_4 (anhydrous) 2.4×10^{-2}M 0 710 g 200 ml H_2O Add (a) and (b) to reach pH 7 5 **pH 7.5**	(a) Anhydrous NaH_2PO_4 0 308 g Distilled H_2O up to 100 ml (b) Anhydrous Na_2HPO_4 0 710 g Distilled H_2O up to 200 ml Add (a) to (b) to reach pH 7 5 **pH 7.5**
REACTION MIXTURE	20 g 2 0% agar gel 30 mg Adenosine 4 mg PMS 8 mg MTT 40 microliter Xanthine oxidase (0 10 unit plate) 80 microliter Nucleoside phosphorylase (1 00 unit plate) 20 ml reaction buffer	(a) Adenosine (sigma A-9251) 40 mg Phenazine methosulfate (sigma P-9625) 2 mg MTT Tetrazolium (sigma M-2128) 10 mg Xanthine oxidase (sigma X-1875) 50 l Nucleoside phosphorylase (sigma N-3003) 100 l Reaction buffer 25 ml (b) Type 1 agarose (sigma A-6013) 0 5 g Distilled H_2O 25 ml
NOTES	This concentration of the starch gel is slightly higher in this method than the usual 10% starch gel employed for most other enzymes This method requires extraction of stains from the substrate material and absorption into clean cotton threads because of ADA sensitivity to the adverse effects of substrate contaminants such as detergent or bacteria Place origin 4½ cm from cathode end if run in conjunction with 6-PGD Run at 80 V instead of 60 V	

ERYTHROCYTE ACID PHOSPHATASE (EAP)

	LABORATORY 4	LABORATORY 5(Multisystem with Esterase D(ESD)
TANK BUFFER	NaH_2PO_4 0.245 M 102.9g, 29.4 g l Trisodium citrate 0.10 M 102.9g, 29.4 g l Distilled H_2O 3.5 l **pH 5.9**	Anhydrous NaH_2PO_4 (sigma S-0751) 29.4 g Trisodium citrate (sigma C-7254) 38.6 g Distilled H_2O up to 1 liter **pH 5.9**
GEL BUFFER	1:100 dilution of tank buffer **pH 5.9**	1:100 dilution of tank buffer
SUPPORT MEDIUM	2 mm starch gel	2 mm starch gel (starch sigma S-4501)
AMOUNT REQUIRED	12% starch gel	10% starch gel (10g/100ml)
APPLICATION	LYSATE: 1 Whatman 3MM chromatographic paper "thread" cut to 8x2 mm soaked with lysate diluted 2:1 with 0.10 M Cleland's Reagent (15.4 mg dithiothreitol per 1 ml H_2O, make up fresh) STAIN: Extract stained material with a minimum of gel buffer diluted 1:1 with 0.10 M Cleland's Reagent and soak into 1 Whatman 3MM "thread" measuring 8x3 mm	LYSATE STAIN: 2x7 mm washed red cell stain soaked 5 minutes in gel buffer containing 7 mg/ml Cleland's Reagent
TEMPTERATURE AND CONDITIONS	Cooling plate. 4°C Migration Anodic	Cooling plate at 4°C. Migration: Anodic
VOLTAGE AND DURATION	Day: 400 V (20V cm) for 4½ hours Night	Day: 400 V for 4 hours Night:
REACTION BUFFER	Citric acid 0.05 M 5.25 g Distilled H_2O 500 ml Adjust to pH 5.0 with 0.05 M NaOH **pH 5.0**	EAP REACTION BUFFER pH 5.0: Anhydrous citric acid (sigma C-0759) 5.25 g; Sodium hydroxide (Mallinckrodt 7708) 1.00 g; Distilled H_2O up to 500 ml
		EAP REACTION MIXTURE: 4-Methylumbelliferyl phosphate (sigma M-8883) 4 mg; Reaction buffer 10 ml
		ESD REACTION BUFFER pH 6.0: Sodium acetate, trihydrate (sigma S-8625) 2.05 g; Distilled H_2O up to 500 ml; Adjust to pH 6.0 with 0.1 N Acetic Acid
		ESD REACTION MIXTURE: 4-Methylumbelliferyl acetate (sigma M-0883) 5 mg; Acetone 1 ml; Reaction buffer 9 ml
REACTION MIXTURE	4-Methylumbelliferyl phosphate 0.004 g Reaction buffer 10 ml Saturate a 5½ x 5½ inch piece of 3 MM Whatman chromatographic paper with the reaction mixture Place paper across plate and store in moisture chamber incubated at 37°C for 20 to 30 minutes	
NOTES	Following the reaction period, remove the Whatman 3 MM overlay and read results under long-wave UV The concentration of the starch gel is slightly higher in this method that the usual 10% starch employed for most other enzymes. Avoid use of glycerol due to selective inhibition of the "C" isozyme Place origin 4 cm from cathode edge of plate	Saturate a 3 x 9 inch 3 MM Whatman paper with EsD reaction mixture, place on most anodic portion of plate and incubate 5 to 10 minutes at 37°C. Remove overlay and read results under long-wave UV. Then saturate a 5 x 9 inch paper with EAP reaction mixtured, place on plate from origin towards anode and incubate for 30 to 40 minutes.

	LABORATORY 4	LABORATORY 5
TANK BUFFER	Stock Tris 6.88x10^{-2} M 37.64 g 10.75g/l Disodium EDTA 2.5x10^{-3} M 3.26 g .931g/l Boric acid 8l.1x10^{-2} M 17.64 g 5.04g/l Distilled H_2O 3.5 l Make a 1:5 dilution. **pH 8.2**	Sodium hydroxide (Mallinckrodt No. 7708) 2.0 g Boric acid (sigma B-0252) 18.6 g Distilled H_2O up to 1 liter **pH 7.9**
GEL BUFFER	Same as tank buffer. **pH 8.2**	Trizma base (sigma T-1503) 9.20 g Citric acid anhydrous (sigma C-0759) 1.05 g Distilled H_2O up to 1 liter **pH 8.6**
SUPPORT MEDIUM	Vertical slab polyacrylamide gel	2 mm Acrylamide gel
AMOUNT REQUIRED	4% to 15% continuous density gradient	Cyanogum 41 Gelling Agent (sigma C-588) 5.0 g Tetramethyl-ethylenediamine (sigma T-8133) 0.1 ml Gel buffer 100 ml Ammonium persulfate (sigma A-6761) 100 mg
APPLICATION	SERUM: Serum sample is applied to well at top of gel (5% acrylamide end) formed in homemade gels by acrylic comb or commercial gels by plastic sample-spacer, using a double-drawn pipette. Fill well no more than 1/4 full. STAIN: Extract stained material with aqueous solution of 10%sucrose using 1 to 2 drops from a double-drawn pipette. Apply to sample well at top of gel with a double-drawn pipette to no more than 2/3 full.	2x7 mm Whatman 3MM paper soaked in serum diluted 4:1 with fresh hemolysate for 5 minutes.
TEMPERATURE AND CONDITIONS	Cold room at 4°C; buffer pumped continuously through the system. At R.T., buffer should also be circulated through a cooling bath.	Cooling plate at 4°C. Migration: Anodic
VOLTAGE AND DURATION	Night: Overnight at 250 V or Over-the-Weekend at 110 V.	Day: 300 V for 5 hours Night: 60 V overnight
REACTION MIXTURE*	o-tolidine 0.25 g Ethanol 10 ml Dissolve tolidine in the EtOH first, then add: Acetic acid, glacial 15 ml 1.0% Hydrogen peroxide 10 ml Remove gel from glass brace, slice off tip of gel from lower corner to mark left side from right, immerse gel into reaction mixture contained in a 5-inch petri dish. After 10 minutes, rinse gel with water.	p,p'-Benzylidenebis-(N,N-dimethylaniline) (Eastman 3620) 75 mg Glacial acetic acid (Mallinckrodt 7848) 5 ml Distilled H_2O 2.5 ml Sodium perborate (Mallinckrodt 7848) 240 mg
NOTES	Commercial equipment designed for this method is required. *o-tolidine is carcinogenic and rubber gloves must be worn during the preparation and application of the reaction mixture. Sera samples must contain at least sufficient hemoglobin to saturate the haptoglobin molecules in the samples. This level of Hb is marked by a clearly observable pink coloration of the serum. Whenever a serum sample lacks the pink color, Hb must be added to saturate the Hp prior to electrophoresis. As a source of Hb, use a small volume of lysate from the freshly washed red cells of a "confirmed" Hb A donor. Failure to saturate Hp with Hb results in the production of spurious band patterns formed by unsaturated Hp-Hb complexes. Hb soln: 20 mg Hb in 0.50 ml 10% sucrose. Two drops serum from regular pipette plus 1 Hb soln. from double-drawn pipette. Allow to stand 20 minutes before inserting sample. Lypholyzed Hb can be obtained from Sigma for the Hb soln.	Dissolve p,p'-benzylidenebis-(N,N-dimethylaniline) in glacial acetic acid, then add the water and sodium perborate. Pour over 3MM Whatman paper that extends from the origin to the hemoglobin band and allow to develop.

APPENDIX D

COMMENTS ON LABORATORY PROCEDURES

APPENDIX D. COMMENTS ON LABORATORY PROCEDURES

1. Troubleshooting

A benefit of a bloodstain analysis program is learning to correct problems as they arise--troubleshooting. No matter how good or reliable the methods used, problems are inevitable. They can be attributed in part to human participation and in part to labile biological reagents. In the following paragraphs, problems which were encountered and the attempts to solve them are discussed.

Visualization difficulties were encountered with PGM_1 (due to G-6-PD), ADA (due tc xanthine oxidase and nucleoside phosphorylase), and with AK (due to ADP). Initially these problems involved the lack of visual intensity of isozyme bands and were remedied either by increasing the proportions of reaction components or by changing suppliers.

A lack of intensity of the band patterns in Gc, attributed to reduction of the antisera titer, was addressed by a reduction of the number of cellulose acetate membranes cycled through each batch of Gc antiserum. Severe background staining occurred with the gradient polyacrylamide gels used for Hp analysis, and troubleshooting revealed the excessive staining was caused by unnecessarily high concentrations of ammonium persulfate in the gel solution. In addition, one laboratory found that in about 4 percent of the cases, it was impossible to obtain normal electrophoretic haptoglobin patterns; the cause was not identified.

A disproportionately high frequency of the MN phenotype was observed. This problem was identified as that of insufficient washing of cells. The samples from all but 60 of the blood specimens were recovered and retested. Corrections, where necessary, were made on the data forms.

Two laboratories experienced difficulty resolving the BA and A phenotypes of the glucose-6-phosphate dehydrogenase (G6PD) system. This difficulty was due to the resolving capability of the electrophoretic system being used. Several attempts were made to resolve this situation, but none was satisfactory.

It was noticed that the resolution of GLO I on the starch gel plates began to deteriorate to the point of endangering proper identification of the phenotypes. Troubleshooting revealed the problem to be associated with the use of a new batch of starch. The previous high-quality results had been obtained using a batch of starch from a particular supplier. A new batch of starch from the same supplier consistently yielded poor GLO results. A number of test samples of other available batches were obtained from the supplier but these were also unsuitable. A check with the supplier and distributor revealed that the original good batch was no longer available. The method for the multisystem analysis of GLO, PGM, and EsD using a 1-percent agarose gel containing 2-percent starch* was used for the balance of the project.

Problems were encountered with G-6-PD, CAII, and ADA when diffusion of bands and an apparent loss of activity occurred. Troubleshooting was initiated. However, when the next batch of blood samples were received from the Red Cross, the electrophoretic results were normal. The reason for the difficulties encountered with the former set of blood samples remains undetermined.

2. Electrophoretograms

In addition to providing meaningful statistical information, participation in data gathering provided the opportunity to observe the rare variants in many genetic marker systems. Contact with local blood banks and other serologists helped to form mechanisms for obtaining sources of rare standards and antisera as well as rare variants or their electrophoretograms.

---

* This method, known as "Group I", developed under another project of The Aerospace Corporation on Subcontract No. 67854 with Beckman Instruments, Inc.

END