

STATEMENT OF DR. GEORGE F. SENSABAUGH, JR.

I, George F. Sensabaugh Jr., state the following to be true to the best of my knowledge:

1. I am a Professor of Forensic Science and Biomedical Sciences in the School of Public Health at the University of California, located in Berkeley, California.
2. I received training in forensic science, genetics, and biochemistry at the University of California, Berkeley (1963-69, D. Criminology), the University of California, San Diego (1969-71, Postdoctoral), and the National Institute for Medical Research, London, England (1971-72, Postdoctoral).
3. I have been on the faculty of the University of California, Berkeley, initially in the School of Criminology and subsequently in the School of Public Health, since 1972. I have served as Department Chair and currently serve as Head of the Division of Infectious Diseases in the School of Public Health.
4. A central focus of research in my laboratory over the past 15-plus years has been the application of DNA technology to problems of genetic analysis, specifically in the areas of forensic science, human genetics, and most recently, microbial population genetics. In the course of this work, my laboratory has designed and/or implemented a number of DNA

based genetic typing tests; the interpretation of our findings often entail the use of basic population genetics and statistics.

5. I am author or co-author on more than 170 publications, which include research reports in peer-reviewed scientific journals and review chapters in books. These research papers and other publications include work in the areas of forensic science, human genetics, and microbial population biology.
6. Included among my publications is a chapter on DNA evidence in the Federal Reference Manual on Scientific Evidence, published by the Federal Judicial Center as a guide for judges; this chapter was co-authored with Professor David Kaye (School of Law, Arizona State University). The chapter reviews issues relating to the application of DNA evidence in the legal setting, ranging from technical matters through to the interpretation of findings, which include population genetics and statistics.
7. I currently serve on the editorial boards of the Journal of Forensic Sciences and Forensic Science Review and am an associate editor for the journal Science and Justice. I participate in the professional activities of the International Society for Forensic Genetics and was President of the Society's 18th International Congress in 1999; I served as co-editor of the congress proceedings.

8. I was a member of both the first and second committees of the National Research Council that reported on the application of DNA technology in Forensic Science, to wit: "DNA Technology in Forensic Science" (April, 1992, hereafter referred to as NRC-I) and "Evaluation of Forensic DNA Evidence" (May, 1996, hereafter referred to as NRC-II). The second committee was appointed to update the first report in light of the rapid advances in DNA research since 1992, particularly the research that called into question the validity of some of the primary statistical recommendations made by the first report.
9. I have been qualified as an expert witness in the broad area of forensic biology on approximately 50 occasions and have testified as to forensic aspects of population genetics in approximately half of these occasions.
10. It is an established principle of human biology that, other than identical twins, every individual is genetically unique, i.e., has a unique DNA sequence. A very large number of sites of DNA sequence variation have been identified and it is possible to distinguish between individuals by testing for differences at these sites. If enough sites of variation are tested, it is possible to identify a particular individual (again, identical sibs excepted) to the virtual statistical exclusion of all others.
11. Research workers in human genetics have identified thousands of sites of DNA sequence variation referred to as short tandem repeat (STR) loci. Genetic typing at STR loci is done using a technique known as the polymerase chain reaction (PCR). STR typing by PCR is a

robust technology and is widely used in many different kinds of genetic studies. Genetic types at STR loci are typically reported as a numerical code that allows typing data to be readily exchanged between laboratories and to be stored in data repositories for use in future studies.

12. The forensic community in the United States has selected a set of thirteen (13) highly variable STR loci to be used as the standard typing set by forensic laboratories in the analysis of biological evidence. These same STR loci are also used in forensic laboratories throughout the world. Genetic typing of these STR loci provides a genetic profile (a "DNA profile") that specifies an individual with a high degree of statistical certainty. To date and to the best of my knowledge, no two individuals (identical twins excepted) among the hundreds of thousands of individuals tested have been found to have matching 13 locus DNA profiles. DNA profiles based on fewer loci have a reduced degree of statistical power but nevertheless can be highly discriminating. The statistical discrimination associated with any particular DNA profile can be calculated using well established principles of human population genetics. These principles are described both in the NRC-II report and in the Federal Reference Manual chapter on DNA evidence.
13. The statistical discrimination power associated with any particular DNA profile is typically represented as the random match probability (RMP), the probability that the particular DNA profile would be detected in an individual randomly selected from a population of unrelated individuals. The RMP is calculated as the estimated frequency of the particular

DNA profile in the population using the multiplication rule: the frequencies of the genetic types at each locus are multiplied together to obtain the joint frequency of the single locus types in combination. To illustrate using classical blood group markers, given that the ABO blood group type "O" occurs in approximately 40% of the population, that the MN blood group type "N" occurs in approximately 20% of the population, and that the two blood group systems are genetically independent (i.e., neither one influences the occurrence of the other), then it is possible to multiply the two frequencies together to estimate that approximately 8% of the population ($0.40 \times 0.20 = 0.08$) are ABO type "O" and MN type "N". In other words, we estimate the probability of selecting an individual at random from the population who happened to be both ABO type "O" and MN type "N" to be 8% (1 in 12.5). DNA markers are much more discriminating than the two blood group markers used in the illustration and there are more of them. Accordingly, RMP values for a 13 locus DNA profile can be on the order of 1 in 100 billion or less. Overall, the RMP is an index of the rarity of a DNA profile.

[One may wonder how it can be that a particular DNA profile can be estimated to occur with a frequency of 1 in a 100 billion or less when the number of persons who have ever lived is less than 10 billion. The answer is that the calculation takes into account all hypothetically possible combinations of genetic types in the 13 locus profile, most of which will never have actually occurred. By way of analogy, the number of possible lottery number combinations well exceeds the number of lottery tickets sold; the chance of any

particular ticket being a winning ticket is determined by the number of possible combinations, not by the number of tickets sold.]

14. Statistical assessment of a DNA profile is given to assist the trier of fact in deciding between two alternative hypotheses about the origin of an item of biological evidence, the hypothesis on the one hand that the evidence item originated from a particular specified individual (e.g., the suspect), and the hypothesis on the other hand that the evidence item originated from someone else. If the evidence item in fact originated from the specified individual, the DNA profile obtained from the evidence should be the same as that individual's DNA profile, i.e., the profiles would match. If, on the other hand, the evidence item in fact originated from someone other than the specified individual, the DNA match would be a chance coincidence, the chance being that the specified individual happened to possess a DNA profile indistinguishable from the person who is the true source of the evidence. The probability of selecting an individual who happens by chance to possess the DNA profile in question is indicated by the RMP, the lower the RMP, the lower the chance of a coincidental match.

15. The RMP calculation entails, as described above, a multiplication of genotype frequencies from each locus. Although the loci used in DNA profiling have been demonstrated to be highly variable in a wide variety of racial and ethnic populations, genotype frequencies do vary from population to population. Accordingly, RMP values are typically calculated for

each of the major racial population groups to provide a range of values reflective of the extent of variation in profile frequencies within the total human population.

16. The RMP is predicated on the selection of an individual bearing a particular DNA profile being selected at random from a population of unrelated individuals. The alternative hypothesis to the true match hypothesis may entail involvement of family members, *e.g.*, "It wasn't me, it was my brother". Using basic genetic principles, it is possible to calculate match probabilities for individuals bearing any degree of genetic relationship. Thus, the probability of a coincidental match between a man and his brother can be calculated.

17. When DNA typing was first introduced into forensic application in the late 1980's, genetic typing was done on a different class of genetic marker (restriction fragment length polymorphisms - RFLP) and fewer (4-6) were used, the DNA typing technology was different from current practice, there was a degree of imprecision in the specification of RFLP types, and the use of DNA profile databases for identification purposes was a concept yet to be realized. At the time of the 1992 NRC study (NRC-I), the small number of RFLP loci then in use for DNA profiling coupled with concern about the reliability of comparing RFLP typing data between laboratories due to inherent imprecisions in RFLP typing and the absence of any experience in the use of DNA databases for identification purposes prompted the NRC-I committee to recommend a very cautious approach to database identification. The NRC-I report recommended a two stage process for making identifications from databases, the first stage being to use a DNA profile based on one set

of RFLP loci to identify a candidate individual (or individuals) from the database and then, to protect against possible typing error in that initial identification, to verify the identification testing using a second set of RFLP loci. The RMP for the second typing profile would be reported as the index of the statistical discrimination power of the identification.

18. By the time of the 1996 NRC report (NRC-II), forensic DNA analysis had advanced on many fronts: STR loci were emerging as the profile markers of choice within the forensic community; the technology had changed such that genetic types could be reported categorically (*i.e.*, as a numerical code); the mass of population genetic data had expanded considerably providing empirical evidence of the extent of genetic variation distinguishing individual human beings from each other regardless of racial or ethnic origin; a significant body of experience using DNA profiling for identification purposes had accumulated; and the framework for a workable standardized DNA identification database was in place. Responding to these advances, the NRC-II report rejected the two stage DNA database approach of NRC-I as unnecessary and noted that the reporting of the RMP from only the second stage of testing discarded meaningful genetic information. The NRC-II position reflects by own scientific opinion then and now. I am aware of no forensic laboratory anywhere in the world that actually implemented the NRC-I recommendation for two stage database identification and RMP reporting, nor do I know of any knowledgeable experts in the field who currently adhere to the NRC-I approach.

19. The NRC-II committee took cognizance of the fact that as the size of DNA profile databases expanded, the probability of finding any particular profile in the database would increase. To address the question "what is the probability of finding a profile matching a DNA evidence profile in a database of a known size?", the report recommended a database hit statistic obtained by multiplying the RMP of the evidence profile by the total number of profiles present in the database ("N") (Recommendation 5.1). The usefulness of the database hit statistic changes as the size of the database and the RMP value change. For a DNA profile with a modest RMP value, say 1 in 10,000, the database hit statistic tells us that there would be a 10% expectation of finding a matching profile in a database of size $N=1,000$, an expectation of finding one matching profile in a database of size $N=10,000$, and an expectation of finding ten matching profiles in a database of size $N=100,000$. This sort of information is likely to be helpful in assisting the trier of fact in evaluating the context of a database hit. However, as RMP values become very small (*e.g.*, 1 in 100 billion or less), the database hit statistic remains small even when the database size is large; to illustrate, with the noted RMP and a database of size $N=100,000$, the database hit statistic is 1 in a million. In the limit case - a database containing the profile of every living person - one encounters the contradiction of calculating a database hit statistic that may be greater than one or less than one depending on the RMP used even though we know operationally that the chance of getting a hit on any particular profile is one.
20. Searching databases containing DNA profiles from convicted offenders has proven valuable in identifying crime suspects who might not have been identified otherwise; these

are the "cold hit" cases. The flip side of the coin is that all the other individuals with profiles in the database are excluded as candidate suspects for the crime; this applies as well in cases where no database hit is made. Even in cases with biological evidence where a suspect is identified by conventional investigative practices, a database search showing that no other individual in the database possess a DNA profile matching the evidence lends credence to the investigative process and to the uniqueness of DNA profiles. As the size of the database increases, so also does the size of the excluded population. It may be useful to the trier of fact to know that if a search on a database of 100,000 profiles yielded one prime suspect, that search also eliminated 99,999 individuals as potential suspects.

21. The RMP and the database hit statistic address different questions. Given a particular DNA profile, the RMP informs us as to the rarity of the profile in the population at large, independent of whether the profile is present or not in a database. The RMP is fixed statistic based on the genetic structure of the population. The database hit statistic, in contrast, informs us as to the expectation that the profile would be present in a database containing some number N profiles. This statistic varies with the size of the databases and hence is not fixed but rather is capable of changing as the size of databases change. The RMP thus provides meaningful information in every case involving a DNA profile whereas the database hit statistic is applicable only in cases involving database searches. The RMP is used by forensic laboratories worldwide as the reported statistic for indexing the significance of a DNA profile match, regardless of whether the case involves a database search.

22. I am aware that a few commentators have interpreted the NRC-II report to recommend that in cases involving database searches the database hit statistic should be used in place of rather than in addition to the RMP. Both statistics and their derivation are described in the NRC-II report, the RMP in the general context of assessing the weight of a DNA profile match and the database hit statistic in the specific context of cold hits on a database. As a member of the NRC II committee, I do not believe the database hit statistic was described with the intent of supplanting the RMP in database search cases but rather as a supplement to it.
23. Since the publication of the NRC-II report in 1996, there have been several papers in the peer reviewed scientific/statistical literature discussing the database hit statistic. There is no argument about the derivation of the RMP or of the database hit statistic; the mathematics are not in question. Rather the discussion is between a few statisticians who have differing notions of how questions should be framed by the legal system and of the statistical approaches that should be used to answer these questions. Professor David Balding provides a cogent overview of this discussion in his 2002 paper, "The DNA Database Search Controversy" [Biometrics 52:241-244]. In short, he makes the case that the legal system has evolved a generally functional way of doing things and should take with a grain of salt suggestions for change made by persons who don't know how it works. His paper is the most recent I have seen on the topic and the two years of silence since its publication suggests that whatever interest there may have been in the matter has dissolved.

George Sensabaugh

George Sensabaugh, D.Crim.

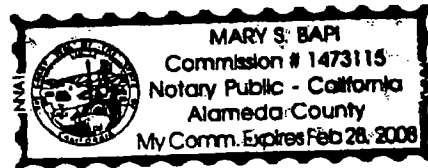
State of California

County of Alameda

Before me the undersigned, a Notary Public for Alameda County, State of California, personally appeared George Sensabaugh, and he being first duly sworn by me upon his oath, says that the statements contained in the above affidavit are true and correct.

Signed and sealed this 10th day of September 2004

Mary S. Bapi



My commission expires *Feb 28, 2008*

GEORGE F. SENSABAUGH, JR.

CURRICULUM VITAE

CURRENT POSITION

Professor of Forensic Science and Biomedical Sciences
Head, Division of Infectious Diseases
School of Public Health
University of California, Berkeley, California 94720

PHONE: (510) 642-1271

FAX: (510) 642-6350

E-Mail: SENSABA@uclink4.berkeley.edu

PERSONAL INFORMATION

Birthdate: 8 June 1941, Palo Alto, California, USA

Marital Status: Married, two children

EDUCATION

B.A., 1963 - Princeton University, Princeton, NJ

Major: Philosophy (Pre-Med)

D. Criminology, 1969 - University of California, Berkeley, CA

Major emphasis: Criminalistics

Minor emphasis: Biochemistry

RESEARCH AND PROFESSIONAL EXPERIENCE

- 1969-1971 Post-doctoral Research Fellow, Department of Chemistry, University of California, San Diego, CA.
- 1971-1972 Post-doctoral Research Fellow, Genetics Division, National Institute for Medical Research, Mill Hill, London, England
- 1972-1975 Assistant Professor of Forensic Science, School of Criminology, University of California, Berkeley, CA
- 1975-1979 Assistant Professor of Forensic Science and Biomedical Sciences, School of Public Health, University of California, Berkeley, CA
- 1979-1986 Associate Professor of Forensic Science and Biomedical Sciences, School of Public Health, University of California, Berkeley, CA
- 1986- Professor of Forensic Science and Biomedical Sciences, School of Public Health, University of California, Berkeley, CA
- 1988-1993 Chairman, Department of Biomedical and Environmental Health Sciences, School of Public Health, University of California, Berkeley, CA
- 1984-1990 Visiting Professor, Forensic Science Unit, Department of Pure and Applied Chemistry, University of Strathclyde, Glasgow, Scotland

HONORS AND AWARDS

Distinguished Service Award, California Association of Criminalists - 1983

Paul L. Kirk Award, American Academy of Forensic Sciences - 1987

Peter Sherry Memorial Lecturer, Georgia Institute of Technology - 1990

Fulbright Research Scholar - 1993

Norman Rosenblatt Memorial Lecturer, Northeastern University - 1995

President, 18th Congress, International Society for Forensic Haemogenetics, 1999

ACADEMIC AFFILIATIONS

Graduate Group in Comparative Biochemistry (Head Graduate Advisor)
Graduate Group in Infectious Diseases and Immunity
Graduate Group in Microbiology
Graduate Group in Nutritional Sciences and Toxicology
Graduate Group in Forensic Science (UC Davis, Executive Comm.)

PROFESSIONAL ASSOCIATIONS

American Association for the Advancement of Science
Sigma Xi
American Society for Human Genetics
American Chemical Society
American Society for Microbiology
California Association of Criminalists
American Academy of Forensic Sciences
California Association of Crime Laboratory Directors
New York Academy of Sciences
International Society for Forensic Genetics
Council on Forensic Science Education

PROFESSIONAL SERVICE

Editorial Secretary, California Association of Criminalists (1977-1982)
Editorial Board, Journal of Forensic Sciences (1980-Present)
Associate Editor for the Americas, Science and Justice (1984-present)
Editorial Board, Forensic Science Reviews (1988-present)
Secretary, Council on Forensic Science Education (1989-1991)
American Society for Human Genetics, ad hoc Committee on Individual Identification by DNA Analysis (1989)
DNA Commission, International Society for Forensic Haemogenetics (1989-1991)
National Research Council, Committee on DNA Technology in Forensic Science (1990-92)
National Research Council, Committee on Forensic DNA - An Update (1994-1996)
Users Advisory Board, California Criminalistics Institute (1994-present)
President, 18th International Congress, International Society for Forensic Haemogenetics, 1999

RESEARCH INTERESTS

Genetic variation in human populations - biological significance and evolutionary origins
Microbial population genetics and evolution
Forensic science - forensic genetics, science-law interactions, concepts of identification