DNA FORENSICS: A POPULATION GENETIC AND BIOLOGICAL ANTHROPOLOGOCAL PERSPECTIVE

Ranajit Chakraborty and **Ranjan Deka**
Center for Genome Information, Department of Environmental Health, University of Cincinnati College of Medicine, Cincinnati, Ohio, USA.

Keywords: convenient sampling, DNA forensics, DNA mixture, exclusion probability, gene diversity, kinship, likelihood ratio, match probability, mitochondrial DNA, population substructure, sample size, short tandem repeat (STR) loci, transfer evidence, Y-chromosome STR.

Contents

Summary

DNA forensics is a discipline in which genetic variations at DNA level is used to aid in forensic investigations to attribute the source of biological samples collected in the context of investigations. Knowledge of human genome diversity as well as operational biological characteristics of DNA markers are essential in selecting genetic markers that are useful in DNA forensics, and in assessing the statistical strength of DNA evidence. For the latter in particular, biological diversity within and between anthropologically defined populations plays a critical role in formulating protocols for evaluating statistical strength. This general thesis is presented in this chapter, with an introduction to the subject of DNA forensics and its brief historical development. For the different generic types of investigative cases handled in DNA forensics, commonly asked

questions of legal relevance are enumerated. It is argued that use of population genetic models of DNA variation, and empirical data on world-wide diversity in anthropologically defined populations, are being used to provide conservative assessment of DNA-based evidence which is helpful for source attribution of forensic specimens. The increasing popularity of DNA forensics demands attention to the future directions in which DNA forensics may be used for human identification, particularly in complex cases arising from mass disasters and natural calamities. The chapter ends with noting the research questions that are to be addressed in enhancing the power of applying the knowledge of human genome diversity in DNA forensics.

1. Introduction

Forensic investigations involving applications of recombinant DNA technology have been described as the most important tool for human identification, since Francis Galton invented the use of fingerprints for such a purpose (NRC 1996). In this chapter, our objective is to present a general overview of the current status of DNA forensics as it is used in criminal as well as in civil proceedings of legal investigations, to exemplify how understanding of anthropological diversity of human populations is used in this area, particularly to assess the statistical strength of forensic DNA evidence. We will begin with a brief history of DNA forensics; outline the generic problems handled in DNA forensics; describe the genetic markers used; and show that anthropological considerations of genetic diversity within and between populations play a critical role in assessing the statistical strength of DNA evidence findings in forensic investigations.

2. A Brief History of DNA Forensics

Jeffreys (1985a,b,c) is credited as the inventor of "DNA Fingerprinting". In these publications he and his colleagues described the technique of DNA profiling using DNA hybridization probes comprised of tandem repeats of core nucleotide sequences detecting multiple variable human DNA fragments by Southern blot hybridization. As the composite profile of genetic variation of such multiple genetic markers resemble bar codes, and creates virtually individualized patterns of each person tested, the term "DNA Fingerprinting" appeared legitimate. Even before that Wyman and White (1980) documented the existence of variable number of tandem repeats (VNTRs) at specific locations of the human genome, suggesting that such repeat variations of core sequences can generate genetic variation several orders larger than that detected by classical serological and biochemical markers.

Though the commercialization of DNA testing by Jeffreys' multilocus probes was started by Cellmark and Lifecodes in 1986, for technical reasons, in 1988 the Federal Bureau of Investigation (FBI) adopted the use of single locus probes to score multiple VNTR loci through repeated re-hybridization by Southern blotting (Butler 2005). In the same year, the UK Home Office and Foreign and Commonwealth Office ratified the use of DNA fingerprinting for resolving family relationships and verification of relationships in immigration cases (Home Office 1988). Used in parentage testing and forensics, this technology soon came under scrutiny during court proceedings for questions regarding procedural and scientific validity (Lander 1989), resulting in a bigger controversy with regard to the statistical strength of DNA evidence for positive

identification of criminals (Lewontin and Hartl 1991; Chakraborty and Kidd 1991).

Invention of the polymerase chain reaction (PCR) technique of DNA amplification (Saiki *et al* 1988) paved the way of addressing the technical limitations of the Southern blot RFLP analysis of VNTR loci, when PCR-based DNA typing methods were introduced in DNA forensics (e.g. HLA-DQA1 typing method, Saiki *et al* 1986). Interestingly, at the same time the National Research Council (NRC 1992) issued the first report, part of which criticized the use of the RFLP method of DNA fingerprinting. Ironically, this report of NRC (called NRC-I) generated even more controversy, as alternatives of RFLP method of DNA typing became popular, and their suggestions for statistical interpretation of DNA evidence were flawed from the viewpoint of population genetic principles. In addition, their recommendations were vague enough to be misused in legal applications. Research and development work in this area, in the meantime, characterized over a dozen of genetic markers, at which genetic variations were ubiquitous (though not as great as those of the VNTR loci, individually), but operationally advantageous in the sense that they each produced discrete alleles, and could be typed with automation and ease of multiplexed PCR methods (Edwards *et al* 1992; Hammond *et al* 1994). Called short tandem repeat (STR), the first commercial kit of PCR-based STR typing was introduced in 1993.

The US Congress DNA Act of 1994 resulted in establishing a governing body (US National DNA Advisory Board) to delineate guidelines, quality control and quality assurance protocols, and adherence of standard operating procedures of DNA typing for forensic use. This body completed its chartered tasks in 2000, resulting in objective sets of criteria to be used in DNA forensics. The so-called "DNA war" lost its fury with the issue of a second National Research Council report, NRC-II (NRC 1996). Even before this, in UK a DNA database was established under the guidance of the Forensic Science Services (FSS) office. In USA, the FBI launched the combined DNA index System (CODIS) in 1998, by which time the PCR-based STR loci (13 of them, along with the sex-typing amelogenin locus) became the major platform of DNA typing for forensic use. With wide acceptance of DNA evidence, nationally as well as internationally, mitochondrial sequencing (of the hypervariable control region) and Y-chromosome linked STR loci were also added to the battery of forensic markers to assist in typing and interpretation of old or degraded evidence samples, and handle DNA mixture with further ease. Two more penta-nucleotide autosomal markers (Penta-D and Penta-E) were added to the battery to autosomal STR kits (Krenke *et al* 2002), and use of DNA forensics became widely popular worldwide, not only to solve criminal and civil cases of human identification on an individual basis, but also for victim identification of mass disaster cases (Beisecker *et al* 2005).

3. Generic Problems Handled in DNA Forensics

The DNA forensic problems related to human identification can be broadly classified in three groups: transfer evidence, DNA mixture analysis, and kinship determination. Transfer evidence relates to scenarios in which DNA profile of an evidence sample (from a crime scene) shows signatures of being DNA from a single source, and the problem is to identify the source of this DNA through comparison of the DNA profile of the evidence sample against those of one or more known persons tested. Three possible

outcomes of such comparisons may arise:

1. Exclusion; i.e. evidence sample profile does not match the profiles of the known persons tested, which results in exclusion of the tested persons as being the source of DNA in the evidence sample;
2. Inconclusive; i.e. due to compromised nature of the evidence samples, their DNA profiles are ambiguous, but neither exclusion nor definitive inclusionary inference can be made; and
3. DNA Match; i.e. at the typed loci, the DNA profile of the evidence sample is indistinguishable from the one found in one of the known persons tested.

The obvious inference drawn under this third scenario is: the tested person cannot be excluded as the contributor of DNA found in the evidence sample. Statistical strength of the evidence is crucial in this third event, since the rarity of the profile would argue against any such coincidental match, should the known person be implicated wrongly in the case.

As most forensic evidence sample are gathered from compromised conditions, by nature, in a great majority of cases, they show signatures of having DNA from multiple individuals. These are termed as DNA mixture, in which, like the transfer evidence scenarios, the results are of three types:

1. Exclusion, occurring when the alleles present in the profiles of known persons are clearly absent in the mixture DNA of the evidence sample;
2. Inconclusive; due to ambiguity of the definitive allele determinations in the mixture DNA; and
3. Inclusion; implying that the known persons cannot be excluded as part contributors of DNA in the mixture sample.

As before, this third alternative observation demands statistical assessment of the evidence. Because of the complexity of the mixture DNA profile, however, the nature of statistical evaluation of DNA mixtures is different (even though the principles are the same) from that of the transfer evidence, which we will discuss later.

Kinship determination, the third type of DNA forensic problem, relates to comparisons of DNA profiles from evidence samples with the ones from one or more individuals biologically or affinally related. The objective is to determine whether or not the evidence sample could belong to a family member of the tested persons. The most popular cases of parentage analyses are indeed special cases of kinship determination, in which through contrasts of alleged father's DNA profile against those of a mother-child pair, one attempts to determine whether or not the questioned person fathered the child. Kinship determination can also attempt to establish stated relationship between evidence sample and that of a family member, or to determine whether or not it represents the DNA of missing offspring of a tested married couple. While all three types of outcomes could result from this type of cases as well, statistics are needed to evaluate the strength of the inclusionary observations. Identification of victims of mass murders, man-made or natural disasters (like the terrorists' attack of 11 September 2001, or the tsunami in South-east Asia in 2004) also fall under this category of DNA forensic issues.

4. Desired Characteristics of DNA Forensic Markers

Though the desirable characteristics of genetic markers that increase their efficiency for application in DNA forensics are related with intra- and inter-population variation, a separate discussion of these features is useful. This is so because some operational features are also relevant as desired characteristics of DNA forensic markers. Historically, genetic markers assayed from coded gene products have been in use for kinship determination and parentage analyses. Blood groups, serum protein and enzyme, and immunological markers have been widely used in parentage analyses since the 1930s. While the efficiencies of such classical genetic markers in ascertaining biological relatedness, individually as well as collectively, have been discussed extensively in the genetic literature, one operational drawback of such markers is the fact that the evidence sample tested should provide uncompromised gene products, suitable enough to type the markers used. Further, barring the immunological markers (e.g. HLA and Gm factors), few of these classical markers have enough segregating alleles to provide high power of discrimination for distinguishing genotypic profiles from different individuals. Relatively, abundance of polymorphic DNA markers in the genome, together with ability to type DNA variation without the gene products, and stability of DNA under compromised condition, offer operational advantages of DNA typing for forensic applications. Further, as DNA can be extracted from a wide variety of biological materials (e.g. blood, saliva, body fluids, tissues, bones, teeth, etc.—in fact any material that contains nucleated cells), applications of genetic markers, which can be assayed with DNA technology, have a much broader scope in comparison to that using the classical genetic markers, such as blood groups, serum proteins and enzymes.

Though the above-mentioned operational advantages are attainable by using several alternative platforms of DNA typing (e.g. Southern blot RFLP method, PCR-based oligo-nucleotide-specific hybridization technique, or PCR-based gel/capillary electrophoresis method), the markers amenable for PCR-based approaches are more advantageous, in the sense that a lower quantity of DNA is enough (e.g. 0.1-1ng versus 50-500ng), degraded DNA is also typable (amplicon size of 500 bp is enough), genotyping reagents are non-isotopic, typing methods are rapid and can be automated, and the allele designations are discrete.

Locus Name	Chromosomal Location	Repeat Motif	Allele Range	Number of Alleles Observed
CSF1PO	5q33.3-34	TAGA	6-16	15
FGA	4q28	CTTT	15-51.2	69
TH01	11p15.5	TCAT	3-14	20
TPOX	2p23-pter	GAAT	6-13	10
VWA	12p12-pter	[TCTG][TCTA]	10-24	28
D13S1358	3p	[TCTG][TCTA]	9-20	20
D5S818	5q21-31	AGAT	7-16	10
D7S820	7q11.21-22	GATA	6-15	22
D8S1179	8	[TCTA][TCTG]	8-19	13
D13S317	13q22-31	TATC	5-15	14
D16S539	16q24-qter	GATA	5-15	10
D18S51	18q21.3	AGAA	7-27	43
D21S11	21q21	[TCTA][TCTG]	24-38	70

Penta-D	21q	AAAGA	2.2-17	>16
Penta-E	15q	AAAGA	5-24	20

Table 1. Characteristics of forensic STR loci

Further, together the 13 or 15 short tandem repeat loci (see Table 1 for their list and other biological characteristics) currently used in the US forensic community, has been studied well enough in worldwide populations to address the initial population genetic criticisms which pertain to the relevance of biological and anthropological diversity in DNA forensics.

-
-
-

Bibliography

Balding D.J. and Donnelly P. (1996). Evaluating DNA profile evidence when the suspect is identified through a database search. *J Forensic Sci*. 41, 603-607. [Likelihood ratios are computed for DNA match when the suspect is identified through database search. Alternative procedures are also given in NRC 1996 and Stockmarr 1999].

Balding D.J. and Nichols R.A. (1994). DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. *Forensic Sci Int*. **64**, 125-140. [This paper derives a formula for the conditional probability of a DNA match given that the profile exists in another individual, under the mutation-genetic drift equilibrium theory of allele frequency distributions in populations of constant finite size. The theory also assumes the infinite allele model of mutations, not necessarily followed by the STR loci currently in use in DNA forensics].

Biesecker L.G., Bailey-Wilson J.E., Ballantyne, J., Baum, H., Bieber, F.R., Brenner, C., Budowle, B., Butler, J.M., Carmody, G., Conneally, P.M., Duceman, B., Eisenberg, A., Foreman, L., Kidd, K.K., Leclair, B., Niezgoda, S., Parsons, T.J., Pugh, E., Shaler, R., Sherry, S.T., Sozer, A. and Walsh, A. (2005). Epidemiology. DNA identifications after the 9/11 World Trade Center attack. *Science* **310**, 1122-1123. [This report provides a synopsis of use of established a improvised technologies to handle human identification in complex mass disaster cases, from the experiences learned to identify the victims of the World Trade Center attack on September 11, 2001].

Buckleton J., Triggs C.M. and Walsh S.J. (2005). *Forensic DNA Evidence Interpretation*. CRC Press, Washington D.C. [This edited volume, through its various chapters, reviewed the methods of DNA typing as well as statistical interpretation of DNA data obtained in the context of DNA forensic case work analyses].

Budowle B., Smith J., Moretti T. and DiZinno J. (2000). DNA *Typing Protocols: Molecular Biology and Forensic Analysis.* Eaton Publishing, Natick, MA. [This book describes the protocols for collection and storage of forensic biological samples, and methods of extraction and typing DNA for forensic investigations].

Budowle B. (2001). Statistics and Mixture Interpretation Workshop. In: *Twelfth International Symposium of Human Identification*, 2001; Biloxi, Mississiippi, USA. [This presentation provides an easy to use DNA mixture interpretation, emphasizing the ease and conservativeness of exclusion probability approach in mixture interpretation].

Budowle B., Monson K.L., and Chakraborty R. (1996). Estimating minimum allele frequencies for DNA profile frequency estimates for PCR- based loci. *Int. J. Leg. Med.* **108**, 173-176. [This note provides an algorithm to compute a minimum threshold allele frequency for STR loci, based on the degree of polymorphism at the loci and the sample size of the study to arrive at conservative estimates of DNA profile frequency in populations taking into account the limited sample size of DNA typing databases].

Budowle B., Carmody G., Chakraborty R., and Monson K.L. (2000). Source attribution of a forensic DNA profile. *Forensic Sci. Communications* **2(3).** http://www.fbi.gov/programs/lab/fsc/backissu/july2000/source.htm. [This paper describes the logic and circumstances under which a source attribution statement of a forensic DNA profile can be made].

Budowle B., Shea B., Niezgoda, S. and Chakraborty, R. (2001a). CODIS STR loci data for 41 sample populations. *J. Forensic. Sci.* **46**, 453-489. [This paper presents allele frequency, estimates of minimum threshold allele frequency, and other DNA forensics related summary statistics for the CODIS STR loci in samples of regional populations of the continental United States, empirically demonstrating the levels of allele frequency differences in the surveyed populations].

Budowle B. and Chakraborty R. (2001b). Population variation at the CODIS core short tandem repeat loci in Europeans. *Legal Med.* **3**, 29-33. [Analyses of allele frequency data on the STR loci used in DNA forensics from several national populations of Europe are presented in this publication, demonstrating that the coefficient of gene differentiation between different European populations is well within the bound prescribed in the NRC 1996 recommendation].

Budowle B., Allard M.W., Wilson M.R., and Chakraborty R. (2003a). Forensics and mitochondrial DNA: Applications, debates, and foundations. *Ann. Rev. Genomics and Human Genet.***4**:119-141. [This paper reviews the utility of mitochondrial DNA typing in DNA forensics, pointing out issues that are to be addressed in such applications of mitochondrial DNA diversity].

Budowle B., Sinka S.K., Lee H.S., and Chakraborty R. (2003b). Utility of Y chromosome STR haplotypes in forensic applications. *Forensic Science Review* **15:**153-164. [This paper discusses procedural as well as statistical protocols for using Y-chromosome STR haplotypes in DNA forensics, showing how this type of lineage markers can supplement other DNA typing methods for enhancing the power of human identification from forensic samples].

Butler J.M. (2005). *Forensic DNA Typing: Biology, Technology, and Genetics of STR Markers.* Elsevier Academic Press, New York. [This is an excellent monograph describing the molecular genetic techniques as well as forensic utility of tandem repeat markers of the human genome].

Cerda-Flores R.M. (2001). *Estructura y Mezcla Genetica de las Poblaciones Mestizas del Noreste de Mexico Mediante el uso de Marcadores Moleculares Autosomicos, Mitochondriales y del Chromosoma Y.* PhD Dissertation, Faculty of Biological Sceinces, Universidad Autonoma de Nuevo Leon, Mexico. [This dissertation showed that while the autosomal genome of the Mestizo population of Nueovo Leon, Mexico consists of genes from Europeans, Africans, and Native Americans, their mitochondrial gene pool is predominantly from Native Americans, and Y-chromsomal genes mostly from the Europeans, suggesting the evidence of gender-biased genetic admixture in this population].

Cerda-Flores R., Budowle B., Jin L., Barton S.A., Deka R., and Chakraborty R. (2002). Maximum likelihood estimates of admixture in Northeastern Mexico using 13 short tandem repeat loci. *Amer. J. Hum. Biol.* **14**, 429-439. [This paper estimated the proportion of genes contributed by Europeans, Africans, and Native Americans in the present day population of Nuevo Leon, Mexico, based on 13 short tandem repeat markers used in DNA forensics].

Chakraborty R. (1986). Gene admixture in human populations: Models and predictions. *Yearbook of Physical Anthropol.* **29**:1-43. [This is a review article on the subject of gene migration between populations and its impact on the genetic structure of populations inferred from autosomal, mitochondrial, and Y-chromosomal markers].

Chakraborty R. (1992). Sample size requirements for addressing the population genetic issues of forensic use of DNA typing. *Hum Biol*. **64**,141-160. [This paper develops an analytical formulation through which database sample sizes can be computed with the condition that frequencies below certain threshold values cannot be reliable. This lead to the development of the concept of minimum threshold allele frequency, used currently in DNA statistics computations].

Chakraborty R. and Jin L. (1992). Heterozygote deficiency, population substructure and their implications in DNA fingerprinting. *Hum Genet.* **88**, 267-272. [This paper shows that the observed heterozygote deficiency at VNTR loci in population data cannot be totally explained by population substructure effects alone].

Chakraborty R. and Kidd K.K. (1991). The utility of DNA typing in forensic work. *Science* **254**, 1735-1739. [This is a rebuttal of the criticism against using statistical genetic principles in DNA forensics, supporting the conservativeness of the used protocols].

Chakraborty R., Shaw M. and Schull, W.J. (1974). Exclusion probability: the current state of the art. *Amer. J. Hum. Genet*. **26**, 477-488. [This paper discusses the cumulative efficiency of using blood group and protein polymorphism markers in parentage analysis].

Chakraborty R., Lee H.S. and Budowle B. (2004). Authors' response. Jour. Forens. Sci. 49, 1-4. [This letter to the editor exemplifies the population genetic flaws of general criticism against DNA typing databases, showing that even if the assumptions of the strict product rule for computing coincidental DNAS match probability are violated, they have no consequence on the conservative methods currently used in such computations].

Chakraborty R., Srinivasan M.R., and Daiger S.P. (1993). Evaluations of standard error and confidence interval of estimated multilocus genotype probabilities, and their implications in DNA forensics. *Am. J. Hum. Genet.* **52**, 60-70. [This paper derives the confidence limits of DNA profile frequencies based on the sampling theory of several independent multinomial distributions].

Chakraborty R., Stivers D.N., Su B., Zhong Y. and Budowle B. (1999). The utility of STR loci beyond human identification: Implications for development of new DNA typing systems. *Electrophoresis* **20**, 1682-1696. [This paper shows that the allele frequency data on STR loci used in DNA forensics, even when they are estimated from convenient sampling of populations, accurately reflect phylogenetic relationships between populations. Further, this paper provides a comparative study of STR and SNP loci to suggest the number of SNP loci which would be needed to equal or exceed the power of the current panel of STR loci used in DNA forensics].

Chakraborty, R. and Zhong, Y. (1994) Statistical power of an exact test of Hardy Weinberg proportions of genotypic data at a multiallelic locus. *Hum. Hered.* **44**, 1-9. [This paper provides power computation of tests of independence of alleles within a multiallelic locus, showing that the low power is associated with small levels of departure from the assumption of independence].

Clayton T. and Buckleton J. (1995). *Mixtures.* In: *Forensic DNA Evidence Interpretation*. (Buckleton J., Triggs C.M., Walsh S.J., eds.), 217-274. CRC Press, Washington DC. [This chapter describes the biological features as well as methods of interpretation related to DNA mixtures in forensic samples].

Curran J.M., Buckleton J.S., Triggs C.M. and Weir B.S. (2002). Assessing uncertainty in DNA evidence caused by sampling effects. Sci. Justice 42, 29-37. [This paper showed that analytically derived confidence limits of multilocus DNA profile frequencies are justified for single donor profiles, and the 10-factor variation is also empirically justifiable].

Deka R., Shriver M.D., Yu L.M., Ferrell R.E. and Chakraborty R. (1995). Intra- and inter-population diversity at short tandem repeat loci in diverse populations of the world. *Electrophoresis* **16**, 1659-1664. [This paper presents gene diversity data within and between anthropologically-defined global human populations at the short tandem repeat loci].

Deka R., Shriver M.D., Yu L.M., Heidreich E.M., Jin L., Zhong Y., McGarvey S.T., Agarwal S.S., Bunker C.H., Miki T., Hundrieser J.,Yin S-J., Raskin S., Barrantes R., Ferrell R.E. and Chakraborty, R. (1999). Genetic variation at 23 microsatellite loci in 16 human populations. *J. Genet.* **78**:99-121. [This paper presents allele frequency data on microsatellite loci in anthropologically defined global human populations, showing that the genetic diversity between populations at these loci conforms to the notion of larger differences between continental populations in comparison with the local populations within the major continents of the world].

Edwards A., Hammond H.A., Jin L., Caskey C.T. and Chakraborty R. (1992). Genetic variation at five trimeric and tetrameric tandem repeat loci in four human populations. *Genomics* **12**, 241-253. [This report consists of characterization of several short tandem repeat loci, showing that multiplex PCR methods are suitable for generating multilocus genotype data on such loci].

Evett I.W. and Weir B.S. (1998). *Interpreting DNA Evidence – Statistical Genetics for Forensic Scientists*. Sinauer Associates, Inc., Sunderland. [This monograph describes the likelihood-based methods for evaluating strength of DNA evidence in forensic casework and parentage analysis involving DNA typing].

Gaensslen R.E., Bell S.C. and Lee, H.C. (1987a). Distributions of genetic markers in United States populations. I. Blood group and secretor systems. *Jour. Forens. Sci.* **32**, 1016-1058. [Compilation of allele frequency data from regional populations of the continental United States are presented for the blood group and secretor systems suggesting that the combined data provide a better information on the frequencies in the US at larger than the data from individual regional samples].

Gaensslen R.E., Bell S.C. and Lee, H.C. (1987b). Distributions of genetic markers in United States populations. II. Isoenzyme systems. *Jour. Forens. Sci.* **32**, 1348-1381. [Isoenzyme system allele frequencies are compiled in this presentation to provide information on genetic variation at these loci in the US at large, in comparison to that in regional populations within the US].

Gaensslen R.E., Bell S.C., Lee H.C. (1987c). Distributions of genetic markers in United States populations. III. Serum group systems and hemoglobin variants. *Jour. Forens. Sci.* **32**, 1754-1774. [Frequencies of variants at the Group-specific component and hemoglobin systems are compiled in this publication for the US populations to show that results obtained from the combined data give better information on frequencies for the US populations at large than is obtainable from data from restricted geographic areas].

Gill P. and Buckleton J. (2005). Biological basis for DNSA evidence. In: Forensic DNA Evidence Interpretation. (Buckleton J., Triggs C.M., Walsh S.J., eds.), 1-25. CRC Press, Washington DC. [This chapter describes the biological features related to genetic profiles obtained by DNA typing of forensic samples].

Gill P., Sparkes, R., Pinchin R., Clayton T., Whitaker, J. and Buckleton, J. (1998). Interpreting simple STR mixtures using allele peak areas. *Forensic Sci Int.* **91,** 41-53. [This paper presents an algorithm to decompose the STR profiles of a DNA mixture into individual components using quantitative data on peak areas of alleles present in the mixture profile].

Hammond H.A., Jin L., Zhong Y., Caskey C.T. and Chakraborty R. (1994). Evaluation of 13 short tandem repeat loci for use in personal identification applications. *Am J Hum Genet.* **55**, 175-189. [This paper forms the first characterization of population genetic properties of short tandem repeat loci that started the formation of the CODIS panel of PCR-based genetic markers used world-wide in DNA forensics].

Harding H.W.J. (1998) DNA database size. *Jour. Forens. Sci*. **43**, 248-249. [This note provides an empirical validation that for estimating allele frequencies above certain threshold values, forensic population databases of size 200 individuals are generally adequate].

Hedrick P.W. (2000). Genetics of Populations, 2nd ed., Jones and Bartlett Publishers, Sudbury, MA. [A text-book on population genetics and its applications in various fields, including parentage testing and human identification using genetic markers. On pp. 384-390, statistical methods related to parentage testing and human identification are reviewed].

Holland M.M. and Parsons T.J. (1999). Mitochondrial DNA sequence analysis – validation and use for forensic casework. *Forensic Sci Rev.* **11**, 21-50. [This paper reviews the use of mitochondrial DNA sequence typing in DNA forensics, and discusses the laboratory protocols, database issues, and statistical computations].

Home Office (1988). DNA profiling in immigration casework. Report of a pilot trial by the Home Office and Foreign and Commonwealth Office (Home Office, London). [This report describes the details of first use of DNA fingerprinting in resolving disputed relationships in immigration cases in UK].

Jeffreys A.J., Brookfield J.F.Y. and Semeonoff R. (1985a). Positive identification of an immigration test-case using human DNA fingerprints. *Nature* **317**, 818-819. [This is the first report describing the successful use of DNA fingerprinting to resolve a dispute of stated relatedness in an immigration case in UK].

Jeffreys A.J., Wilson V. and Thein S.L. (1985b). Individual specific "fingerprints" of human DNA. *Nature* **316**, 76-79. [This paper coined the term "DNA fingerprinting", and describes DNA sequence level

homologies of many minisatellite loci, defining a "core"sequence].

Jeffreys A.J., Wilson V. and Thein S.L. (1985c). Hypervariable minisatellite regions in human DNA. *Nature* **314:**67-73. [The term "minisatellite" is used for the first time in this paper to describe length polymorphism of tandem repeat sequences].

Jin L. and Chakraborty, R. (1995). Population substructure, stepwise mutations, heterozygote deficiency and their implications in DNA forensics. *Heredity* **74**, 274-285. [Under the assumption of contraction-expansion stepwise mutation model, this paper shows that the effect of population substructure on genotype frequencies is smaller for loci with higher mutation rate. Implication of this theory is that the level of population substructure adjustments needed for the STR loci used in DNA forensics cannot be predicted from the population substructure statistics inferred from bi-allelic markers, which generally have a much lower mutation rate].

Krane D.E., Doom T.E., Mueller L.D., Raymer M.L., Shields W.M. and Thompson W.C. (2004). Commentary on: Budowle B, Shea B, Niezgoda S, Chakraborty R. CODIS STR loci data from 41 populations. J Forensic Sciu 2001; 46:453-89. Jour. Forens. Sci. 49, 5-6. [This letter to the editor criticizes DNA typing population data, suggesting that ignorance of sporadic departures from assumptions of allelic independence invalidates DNA match probability computation using the strict product rule. The flaws of such criticisms are pointed out in the response by Chakraborty *et al* 2004].

Krenke, B.E., Tereba A., Anderson S.J., Buel E., Culhane S., Finis C.J., Tomsey C.S., Zachetti J.M., Masibay A., Rabbach D.R., Amiott E.A. and Sprecher C.J. (2002). Validation of a 16-locus fluorescent multiplex system. *Jour. Forens. Sci*. **47**, 773-785. [This paper describes the validation work for the Promega multiplex PowerPlex 16 system kit used in DNA forensics, which include the two pentanucleotide repeat loci Penta D and Penta E].

Lander E. S. (1989). DNA fingerprinting on trial. *Nature* **339**, 501-505. [First major criticism of procedural and scientific validity of DNA typing in forensics].

Lewontin, R.C., Hartl, D.L. (1991). Population genetics in foresic DNA typing. *Science* **254**, 1745-1751. [This report documents the first major attack in US on the statistical aspects of DNA typing for positive identification of criminals].

Mueller L.D. (1999). The DNA typing controversy and NRC II. In: *Statistical Methods in Health Sciences: Genetics* (Halloran M.E. and Geisser S., eds.). Springer-Verlag, New York. [This chapter documents sporadic deviation from assumptions of allelic independence within and between loci in DNA typing databases, without taking into consideration of effects of multiple testing and rarity of allele frequencies contributing to such deviations].

National Health Survey (1980). *Selected genetic markers of blood and secretions*. U.S. Department of Health, Education, and Welfare, Publ. No. (PHS) 80-1664, Washington D.C. [This document provides empirical data showing that the allele frequencies at genetic loci estimated from convenient sampling corresponds well with those obtained from statistically designed stratified random sampling methods in the US populations].

Neel, J.V. and Schull W.J. (1946). *Human Heredity*. The University of Chicago Press, Chicago. [An introductory text of human genetic principles and their applications].

NRC (1992). National Research Council Committee on DNA Technology in Forensic Science, Board on Biology, Commission of Life Sciences, DNA Technology in Forensic Science, 1992; National Academy Press, Washington DC. [This report addressed procedural as well as statistical genetic issues with regard to the use of DNA typing in forensic investigations. The recommendations of this report were subsequently found to be scientifically unjustifiable, prompting a further report from NRC].

NRC (1996). National Research Council Committee on DNA Forensic Science, The Evaluation of Forensic DNA Evidence. 1996; National Academy Press, Washington DC. [This report provides guidelines for statistical genetic computations for evaluating the strength of DNA forensic evidence, using principles and empirical data on population genetics and human genome diversity].

Perlin M.W. and Szabady, B. (2001). Linear mixture analysis: a mathematical approach to resolving mixed DNA samples. *Jour. Forens. Sci.* **46**, 1372-1378. [This paper provides a mathematical model to translate the DNA profile of a mixture of DNA into individual component genotypes from which the mixture could have been generated].

Rivas F., Zhong Y., Olivares N., Cerda-Flores R.M. and Chakraborty R. (1997) World-wide gene diversity at the HLA-DQA1 locus. *Am. J. Hum Biol***.** 9,735-749. [This paper presents empirical data on the levels of departure from independence of alleles within a locus due to population substructure and genetic admixture effects at the HLA-DQA1 locus. Empirical data on the extent of population substructure effects at this locus is also presented for a world-wide compilation of genotype data at this locus].

Robinson W.P., Asmussen M.A. and Thomson, G. (1991). Three-locus systems impose additional constraints on pairwise disequilibria. *Genetics* **129**, 925-930. [This paper shows mutilocus systems impose additional constraints on linkage disequrilibria, and hence, without such constraints invoked tests of multilocus linkage disequilibria may not be valid].

Robin E.D. and Wong R. (1988). Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. *Jour. Cell Physiol*. **136**, 507-513. [This paper developed a biochemical method for estimating the virtual number of mitochondria per cell, and showed that the amount of mtDNA per cell is closely regulated with given cell types but differ widely by cell type].

Saiki R.K., Bugawan T.L., Horn, G.T., Mullis, K.B. and Erlich H.A. (1986). Analysis of enzymatically amplified beta-globin and HLA-DQ alpha DNA with allele-specific oligonucleotide probes. *Nature* **324**, 163-166. [This paper describes a PCR-based method to detect polymorphisms due to single base substitution detected by an allele-specific oligonucletide hybridization assay].

Saiki R.K., Gelfand D.H., Stoffel S., Scharf S.J., Higuchi R., Horn G.T., Mullis K.B., Erlich and H.A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487-491. [This paper reports the technique of polymerase chain reaction, PCR, to amplify DNA segments, which eased and revolutionized genotyping and sequencing polymorphic regions of any genome].

Sans M., Weimer T.A., Franco M.H., Salzano F.M., Bentancor N., Alvarez I., Bianchi N.O., and Chakraborty R. (2002). Unequal contributions of male and female gene pools from parental populations in the African descendants of the city of Melo, Uruguay. *Am. J. Phys. Anthrop.* **118**, 33-44. [This paper describes the impact of gender-biased gene migration on the population structure of a genetically admixed population].

Sinha S.K., Budowle B., Chakraborty R., Pauovic A., Guidry R.D.V., Larsen C., Lal A., Schaffer M., Pineda G., Sinha S.K. Jr., Schneida E., Nasir H., and Shewale J.G. (2004). Utility of the Y-STR Typing Systems, Y-PLEX^{TM} 6 and Y-PLEX^{TM} 5 in Forensic Casework and 11 Y-STR Haplotype Database for Three Major Population Groups in the United States. *Jour. Forens. Sci.* **49**, 1-10. [This paper describes Y-linked haplotype population data using 11 STR markers for three major population groups of continental United States].

Stockmarr A. (1999). Likelihood ratios for evaluating DNA evidence when the suspect is found through a database search. *Biometrics* **55**, 671-677. [This paper provides the statistical logic of computing the likelihood ratios of DNA match in cases where the suspect is found through database search. Alternative similar procedure has also been suggested by Balding and Donnelly, 1996; and NRC 1996].

Sun G. McGarvey S.T., Bayoumi R., Mulligan C. J., Barrantes R., Raskin S., Akey J., Chakraborty R., and Deka R. (2003). Global genetic variation at nine short tandem repeat loci and their implications on forensic genetics. *Eur. Jour. Hum. Genet.* **11**, 39-49. [Gene diversity analyses of 9 STR loci used in DNA forensics with new data on 20 globally distributed human populations are presented in this paper to support that the extent of population effects recommended in NRC 1996 is empirically well supported by real data on the humen genome diversity at these forensic loci].

Thompson W.C., Taroni F. and Aitkin, C.G.(2003). How the probability of a false positive affects the value of DNA evidence. *Jour. Forens. Sci.* **48**, 47-54. [This paper uses Bayesian inferential procedures for modifying the likelihood of DNA match using strong assumptions regarding the probability of false positives in case work analysis].

Walker R.H. (ed). (1983). *Inclusion Probabilities in Parentage Testing.* American Association of Blood Banks, Arlington, VA. [A symposium proceedings describing the standard serological markers used in parentage testing and a review of statistical methods for for evaluating strength of evidence for paternity].

Weir B.S. (1996). *Genetic Data Analysis II*. Sinauer, Sunderland, MA. [A textbook on statistical genetics

describing methods of analyzing different types of genetic data].

Wyman A.R. and White R. (1980). A highly polymorphic locus in human DNA. *Proc. Natl. Acad. Sci. USA* **77**, 6754-6758. [This is the first characterization of a single-locus tandem repeat locus on the human genome].

Biographical Sketches

Ranajit Chakraborty is the Director of the Center for Genome Information and Robert A. Kehoe Professor at the Department of Environmental Health of the University of Cincinnati College of Medicine in Cincinnati, Ohio. Professor Chakraborty's research and teaching activities are related in the general area of complex disease genomics and development of design of such studies and data analysis methods. Another focus of his research is the use of recombinant DNA technology in human DNA forensics and microbial forensics of pathogen detection and identification. Inventors of several new design strategies of complex disease gene mapping (such as Mapping by Admixture Linkage Disequilibrium, MALD), Chakraborty has also worked extensively on formulating population genetic methods for assessing strength of DNA forensic evidence for human and pathogen identification. Raised in a family of religious background, Professor Chakraborty is also active in local Indian community serving in spiritual activities.

Ranjan Deka is a Professor at the Center for Genome Information in the Department of Environmental Health of the University of Cincinnati College of Medicine, Cincinnati, Ohio. He is also the Director of the Core Genotyping Laboratory at this Center. Trained as a physical anthropologist, Dr. Deka moved into the field of complex disease research. His primary interest is identification of genetic variants associated with type 2 diabetes, obesity and metabolic syndrome. His research involves utilization of isolated populations in understanding the genetic basis of these complex diseases.