Available online at www.sciencedirect.com
ScienceDirect
Forensic Science International 165 (2007) 238–243
Forensic Science International
www.elsevier.com/locate/forsciint

Scientific standards for studies in forensic genetics

Peter M. Schneider *

Institute of Legal Medicine, University of Cologne, Melatenguertel 60-62, D-50858 Cologne, Germany

Received 13 June 2006; received in revised form 13 June 2006; accepted 19 June 2006
Available online 27 July 2006

Abstract

Forensic molecular genetics has evolved from a rapidly developing field with changing technologies into a highly recognized and generally accepted forensic science, leading to the establishment of national DNA databases with DNA profiles from suspects and convicted offenders. DNA evidence has taken a central role by carrying a significant weight for convictions, as well as by excluding innocent suspects early on in a criminal investigation. Due to this impact on the criminal justice system, guidelines for research in forensic genetics have been introduced already since many years. The most important issues regarding the selection and definition of typing systems both for paternity testing and for forensic identification, the criteria for technical and biostatistical validation, as well as the use of mitochondrial DNA analysis are summarized and discussed.

Keywords: PCR; Forensic DNA typing; Short tandem repeat (STR); mtDNA; Population genetics; Quality assurance

1. Introduction

Modern DNA-based forensic genetics comprise a number of important applications. Examples are the investigation of biological stains to obtain evidence for the presence of an alleged perpetrator at the crime scene by comparing the genetic profiles from crime scene samples of human origin to those of potential stain donors, the identification of unknown corpses in the context both of natural death and of crime or mass disaster, as well as the investigation of family relationships.

In the last 20 years, forensic molecular genetics has evolved from a rapidly developing field with changing technologies into a highly recognized and generally accepted forensic science, leading to the establishment of national DNA databases with DNA profiles from suspects and convicted offenders [1–4]. The methodology has become quite reliable and the analytical equipment has reached a high level of automation. The genetic typing systems are standardized based on generally accepted recommendations from scientific bodies such as the National Research Council (NRC) of the Academy of Sciences of the United States, and the International Society for Forensic Genetics (ISFG). At a very early stage, the ISFG has recognized the potential significance of DNA-based typing methods for the entire field of criminal investigation and paternity testing, and has addressed relevant topics initially regarding the introduction of hybridization-based single and multi-locus minisatellite probes [5,6], and very soon as well on PCR-based typing systems [7,8].

DNA evidence has taken a central role by carrying a significant weight for convictions in the court of justice, as well as by excluding innocent suspects early on in a criminal investigation. A typical and highly publicized crime case involving the rape and murder of two girls clearly demonstrated this capacity, and was the first example for carrying out a mass screening among a defined subgroup of the male population to identify the true perpetrator [9,10]. In contrast to other fields of applied research where quality problems leading to genotyping errors are being discussed only recently [11], the consequences of error in the forensic field carry a heavy burden on all scientific studies, as these may end up as a basis for a ruling on the admissibility, or for analysing and interpreting evidence in a crime case. In this regard, the guidelines developed for forensic genetics may serve as a model system for establishing and assessing the quality of research also for other fields of molecular genotyping studies [12]. To provide an overview about the standards for applied research in forensic genetics, typical areas relevant for scientific studies will be subsequently addressed.

* Tel.: +49 221 4784251; fax: +49 221 4783294.
E-mail address: peter.schneider@uk-koeln.de.

0379-0738/$ – see front matter © 2006 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.forsciint.2006.06.067

2. Paternity testing

Genetic studies on family relationships are pivotal to any application of DNA markers as they serve to establish and validate new typing technologies and genetic systems. Typically, family trios with confirmed genetic relationship using a reference method and from a homogeneous population are selected for study. The results of such validation studies are relevant both for paternity testing and for forensic identification.

2.1. Definition of systems and alleles

Following the initial ISFG guidelines [5,6], a new genetic typing system is defined by ''a segment of unique DNA sequence'' occupying a specific chromosomal position, and by the ''unique identification of complementary primer sequences''. Furthermore, the alleles of a system are defined by ''DNA fragments of variable length agreeing with a formal genetic model''. This model should be Mendelian inheritance which is applicable for all DNA typing systems, notwithstanding the possible occurrence of ''silent'' alleles [13] which can be formally handled following the example of classical blood group markers such as the ABO system. Furthermore, chromosomal linkage data to other systems should be known to consider the possible relevance of haplotype information.

2.2. Standardization requirements

Appropriate size markers with discrete sizes flanking and spanning the entire range to be analyzed are necessary for determining the length of allelic DNA fragments. For the currently used PCR-based short tandem repeat (STR) systems, allelic ladders comprising a complete range of the commonly observed allelic fragments have to be included in each analysis [7,8]. Specific guidelines have been established regarding the nomenclature of STR alleles [14,15]. They are based on the agreement that only denaturing gel systems shall be used for electrophoretic separation of DNA fragments. Thus, only the molecular weight of a fragment (i.e. the size in basepairs) is relevant but not a variation within the DNA sequence, which would become visible using a non-denaturing gel system due to sequence-based secondary structures within the variant alleles [16]. Allele designations are based on the number of variable repeats. These are derived from the coding DNA sequence information in the human genome database, e.g. the GenBank, or, if missing, based on the first original description in the literature. In some cases, this has led to confusion in particular for X- und Y-chromosomal STR systems which had to be resolved by exchange of samples [17,18]. Specific issues regarding Y-STRs have also been addressed by the DNA Commission [19,20]. A useful tool in this context are reference DNA samples such as generally available cell lines with known genotypes, as these can be used to standardize typing systems and allelelic ladders for population studies [21]. This is particularly relevant since it is recommended to include a known human control DNA sample in each experiment. An additional approach is the provision of reference DNA samples available from public institutions such as the National Institute of Standards and Technology (NIST) offering traceable standard reference materials for PCR typing [22].

2.3. System validation studies

To establish the relevant genetic parameters such as Mendelian inheritance, mutation rates and allele frequencies, at least 500 meioses should be studied using a standardized approach [7,8]. It can be expected that mutation rates for autosomal STRs are in the range of 0.1–0.5%, and that paternal mutations exceed maternal mutations [23]. For the Y-chromosomal systems, similar mutation rates have been obtained [24]. By counting the parental genotypes from family trios, allele frequencies can be established provided that the families have been randomly selected from a representative and homogeneous population (see also below). Under these conditions, observed and expected genotype frequencies based on the assumption of Hardy–Weinberg equilibrium (HWE) have to be compared using adequate statistical tests [25]. In any case, it is recommended to apply an exact test [26] rather than the less accurate chi-square test which may be misleading when rare alleles are found in a population sample. It has to be kept in mind, however, that a deviation from HWE may be more indicative for a technical problem in the typing procedure rather than for a non-representative population sample. If a significant excess of homozygous genotypes is observed, a critical assessment of allele detection procedures and electrophoretic conditions should be made, as these may have been a cause for a loss of DNA fragments outside the detection range, or due to poor electrophoretic resolution.

If multiple systems are validated such as multiplex STR kits, the results from the population studies have to be corrected for multiple testing, and the independence of loci has to be demonstrated, e.g. using a modification of the exact test [27,28]. Although slight deviations will occasionally be observed for individual loci, this is normally not critical, as the chromosomal locations of the tested systems are known and not closely linked. Another typical source for such deviations are small population sample sizes of less than 500 individuals. Testing at least 500 meioses implies that 250 family trios have to be tested comprising 250 unrelated male and female individuals with a total number of 1000 alleles. From the scientific literature it becomes evident that not all published studies fulfill this requirement, mostly for practical reasons of obtaining a sufficient number of validated family trios. Therefore, data from such studies may have to be treated with caution. It should also be noted that it is highly desirable to have access to the original genotype data (or haplotype data, respectively, in the case of Y-chromosomal markers) in addition to the normally published allele frequencies. Nowadays, using online publication resources, such data can easily be made available as supplementary data for downloading from the journal's website.