The future of forensic DNA analysis

John M. Butler

National Institute of Standards and Technology, Gaithersburg, MD, USA

The author’s thoughts and opinions on where the field of forensic DNA testing is headed for the next decade are provided in the context of where the field has come over the past 30 years. Similar to the Olympic motto of ‘faster, higher, stronger’, forensic DNA protocols can be expected to become more rapid and sensitive and provide stronger investigative potential. New short tandem repeat (STR) loci have expanded the core set of genetic markers used for human identification in Europe and the USA. Rapid DNA testing is on the verge of enabling new applications. Next-generation sequencing has the potential to provide greater depth of coverage for information on STR alleles. Familial DNA searching has expanded capabilities of DNA databases in parts of the world where it is allowed. Challenges and opportunities that will impact the future of forensic DNA are explored including the need for education and training to improve interpretation of complex DNA profiles.

1. Introduction

Since its introduction in the mid-1980s [1], forensic DNA testing has played an important role in the criminal justice community through aiding conviction of the guilty and exoneration of the innocent. Remains from missing persons and victims of mass disasters have been re-associated and identified through linking reference samples to recovered remains [2]. New technologies are regularly introduced and validated to expand the capabilities of laboratories working to recover DNA results with improved sensitivity and informative-ness. Forensic laboratories have embraced automation for sample preparation and data interpretation in order to meet increasing throughput demands. Short tandem repeat (STR) typing continues to be the primary workhorse in forensic DNA analysis although other genetic markers are used for specific applications [3,4].

(a) DNA capabilities

DNA analysis provides capabilities not found in most of the other forensic disciplines. When biological material is transferred between perpetrator and victim in violent crimes such as murder and rape, DNA recovered from the crime scene has power to potentially identify the perpetrator. Theoretically when testing sufficient genetic markers, probabilistic ‘individualization’ of a DNA profile is statistically achievable except with identical twins—and even twins can sometimes be separated with additional genetic information [5]. Use of genetic markers that are inherited independent of one another enables application of the product rule (and subpopulation adjustments [6]), where statistical rarity of results at each marker can be combined across multiple genetic markers. Hence the weight of evidence can be expressed in quantitative terms that can be quite rare (1 in trillions or rarer) thanks to the multiplicative nature of independent multi-locus results. These calculations involve a genetic principle known as Hardy–Weinberg equilibrium [7]. DNA statistical interpretation rests on a solid scientific foundation, much of which originally came from academic efforts outside of the forensics arena. However, as will be discussed later, interpretations of complex DNA mixture profiles can be subjective and inconsistent between analysts and across laboratory protocols.

Interpretation of all forensic evidence involves comparing question (Q) results with known (K) references. This Q-to-K comparison depends on the quality of results obtained from crime scene evidence (Q) and the availability
of suitable reference samples (K). When samples from a suspect or multiple suspects are available during forensic DNA analysis, the Q-to-K comparison is fairly straightforward through examining the Q and K samples at the same genetic markers. When no suspects are available, DNA databases, which have been developed over the past 20 years to provide potential reference profiles from known previous offenders, can be searched to try to find a match to the unknown Q profile. DNA database profiles essentially enable throwing a wide net at an unknown Q profile to try to identify the perpetrator.

Another important capability of DNA is that by the nature of its inheritance pattern, where half of an individual’s genetic code comes from his or her mother and half comes from his or her father, close biological relatives can be used for reference points. In other words, the possibility exists to reach beyond information available in a sample itself due to the capabilities of established genetic trait transmission principles. Missing persons and disaster victims can be identified from kinship associations if no direct reference sample is available for comparison purposes [8]. For example, if a child is missing, then the child’s mother and father can supply DNA for comparison purposes that may be used to link their DNA profiles to putative sample(s) from their child.

Superb sensitivity is available with forensic DNA testing due to amplification of target regions with the polymerase chain reaction (PCR). DNA results can be obtained from as little as a single cell depending on methods used [9]. However, this exquisite sensitivity is both a blessing and a curse. When high-sensitivity techniques are used, a very real possibility exists for contamination from DNA coming from someone not associated with the crime sample under investigation. If consumables such as swabs or tubes are not DNA-free, then the manufacturer’s DNA may be detected and even lead investigators down the wrong path [10]. A number of measures, such as examining negative controls, are typically used to prevent drawing incorrect conclusions if DNA contamination occurs.

Quality assurance measures have been developed over the years for DNA testing to promote confidence in results obtained. The DNA testing quality infrastructure is probably more advanced than that of many other forensic disciplines, because of experience from organizations like the European DNA Profiling Group (EDNAP) and the European Network of Forensic Science Institutes (ENFSI) in Europe and the Federal Bureau of Investigation’s DNA Advisory Board (DAB) and Scientific Working Group on DNA Analysis Methods (SWGDAM) in the USA.

Forensic DNA technology development has directly benefitted from the Human Genome Project (HGP) and subsequent scientific efforts and discoveries in genetics and genomics. Genomic information from HGP provided the STR genetic markers for testing and more recent studies have supplied population data for assessing genetic variation. Tools, talent and technology have spilled over into the forensic DNA field from the much larger DNA sequencing market. Current DNA typing with STR markers involves use of fluorescent dyes to label PCR products and capillary electrophoresis (CE) to rapidly separate and detect these dye-labelled PCR products. The dyes and the CE methods used in forensic DNA testing were initially developed for DNA sequencing purposes. Future advances in forensic DNA analysis will probably mirror genomic technology development.

(b) Author’s experience

I have sought throughout my career to understand past and present activities in the field through writing textbooks and providing training materials to practitioners. I have also tried to help shape the future of forensic DNA analysis through performing relevant research. My first contributions to forensic DNA analysis involved pioneering quantitation methods for mitochondrial DNA [11] and CE separation methods for genotyping STR markers [12]. Twenty years of research, 150 publications, more than 300 presentations and workshops, and five textbooks have contributed to an informed perspective on forensic DNA analysis. Over the past decade, I have given more than a dozen presentations on my thoughts regarding the future of forensic DNA (http://www.csl.nist.gov/srhtmlbase/NISTpub.htm). The Applied Genetics Group in the US National Institute of Standards and Technology, which I had the privilege of leading from 2000 to 2013, has actively worked to improve the future of forensic DNA.

In 2001, two colleagues and I shared our thoughts on the types of genetic markers that would play a role in the future and why [13]. Those thoughts regarding continued use of STR markers with single nucleotide polymorphism (SNP) markers remaining supplemental solely for ancestry and phenotype application have proved prophetic in many ways. In my Fundamentals textbook, I included an entire chapter on ‘Future Trends’ [14]. I hope that the thoughts presented in this article can provide some useful input into where we are as a field and where we need to go in the foreseeable future.

2. The past reviewed

Accurately predicting the future is always challenging due to unforeseen innovation. However, by examining the past and understanding present challenges, it is often possible to extrapolate to reasonable predictions for the future. Appreciating the trajectory of activities in any field of endeavour enables those acting in it to know where they are going and to work to improve the future. The Research and Development Working Group of the National Commission on the Future of DNA Evidence in 2000 [15], Peter Gill in 2002 [16] and Lutz Roewer in 2013 [17] provided their perspectives at the time on the past, present and future of forensic DNA.

(a) Major themes and time periods for forensic DNA analysis

Table 1 summarizes my views regarding primary activities in each phase of forensic DNA analysis development over the past three decades (see also [18]). The first decade of forensic DNA analysis involved what I would term the ‘exploration’ phase, where different methodologies were attempted involving restriction fragment length polymorphism (RFLP) followed by more rapid and sensitive PCR assays. However, the early PCR assays did not have the specificity (i.e. power of discrimination) that original RFLP methods provided with its many possible alleles and genotypes.

By the beginning of the second decade of DNA typing, efforts solidified around STRs, as these genetic markers provided both a high degree of sensitivity through PCR amplification and specificity when results from multiple STR loci were combined in multiplex PCR amplifications. This decade is what I refer to as the ‘stabilization and standardization’
come due to the success of DNA to solve crimes and the new capabilities for DNA analysis and use. Laboratories as a ‘sophistication’ phase, where expanding tool sets provide and adopted by the European Union [25] and the USA [26].

Database sizes, an expanded set of core loci have been proposed and national DNA databases around the world. With larger DNA funding. Similar growth patterns have occurred in other laboratory capacity thanks to significant federal government due to expansion of DNA database collection state laws and NDIS is 12 million profiles larger (14 million total profiles) by the end of 2004. Ten years later, the National DNA Index System (NDIS) in October 1998 that involved the SGM loci [22]. The USA launched its the first one being the UK National DNA Database in April 1995 that involved the SGM loci [22]. The UK Forensic Science Service (FSS) led the application of STR markers to forensic casework with a first-generation quadruplex consisting of TH01, vWA, FES/FPS and F13A1 [19]. Collaborative studies conducted by EDNAP helped to ensure consistency of early methods between laboratories [20]. A second-generation multiplex (SGM) followed a few years later and examined six STR loci (TH01, vWA, FGA, D8S1179, D18S51 and D21S11) and the sex-typing marker amelogenin [21]. National DNA databases began to be developed, with the first one being the UK National DNA Database in April 1995 that involved the SGM loci [22]. The USA launched its National DNA Index System (NDIS) in October 1998 with 13 core STR loci (TH01, vWA, FGA, D8S1179, D18S51, D21S11, CSF1PO, TPOX, D3S1358, D5S818, D7S820, D13S317 and D16S539) [23]. Once core loci were decided upon, commercial kits began to be used for autosomal STR and Y-chromosome STR markers. Analysis platforms, which originally involved slab gel electrophoresis with silver-stain or fluorescence detection, eventually stabilized on CE with laser-induced fluorescence detection that permitted sensitive, automated and high-throughput analysis. Quality assurance measures were put into place during this decade by the FBI’s DNA Advisory Board in the USA and ENFSI in Europe [24].

The third decade of DNA analysis involved a ‘growth’ phase. The US NDIS began this decade with around two million DNA profiles at the end of 2004. Ten years later, NDIS is 12 million profiles larger (≈14 million total profiles) due to expansion of DNA database collection state laws and laboratory capacity thanks to significant federal government funding. Similar growth patterns have occurred in other national DNA databases around the world. With larger DNA database sizes, an expanded set of core loci have been proposed and adopted by the European Union [25] and the USA [26].

The next decade and the future beyond it can be labelled as a ‘sophistication’ phase, where expanding tool sets provide new capabilities for DNA analysis and use. Laboratories must navigate the flood of sample submissions that often come due to the success of DNA to solve crimes and the ability to conduct more tests from a single item than previously feasible. DNA testing will become more rapid, more informative and more sensitive. Probabilistic methods of expressing complex results will require software and deeper thinking about the meaning of data obtained. An ever-increasing understanding of human genomic information will require confronting genetic privacy concerns now and in the future.

(b) Research leadership
Research and development is crucial to the future of any field. During its existence, the UK Forensic Science Service played an important role in the development and application of forensic DNA techniques [27]. The pioneering research efforts of Peter Gill, Ian Evett, Gillian Tully and others led the way for development of new DNA markers, methodologies and interpretation frameworks. Other important centres of research in forensic DNA over the past several decades include University of Innsbruck (Austria), University of Copenhagen (Denmark), University of Santiago de Compostella (Spain), NIST Applied Genetics Group (USA), University of North Texas Health Sciences Center (USA) and the Institute of Environmental Science and Research (ESR, New Zealand). These groups regularly publish their research in Forensic Science International: Genetics, Forensic Science International, International Journal of Legal Medicine, Legal Medicine and Journal of Forensic Sciences.

3. The present considered
Understanding the present situation and challenges can aid looking to the future of forensic DNA analysis.

(a) Genetic marker systems
The primary genetic tests in use today involve autosomal STR markers—especially those core STRs that are used in national
DNA databases. Current commercial kits typically amplify in the range of 15–22 autosomal STR loci in order to cover European and/or USA core loci. Table 2 shows the current practice and future potential for various genetic marker systems used in forensic DNA (see [3] for more information on these markers). A primary benefit for STRs is that their multiple alleles enable detection and interpretation of evidentiary DNA mixtures more effectively than bi-allelic markers.

(b) DNA database growth and use in the USA
The number of samples (both known references and crime scene specimens) involved in DNA databases means that genetic markers used to generate the DNA profiles in those databases will drive the future of DNA testing. Over the past 15 years, DNA collection laws have expanded to provide a greater number of samples for testing purposes. Table 3 reviews the growth in the number of states over the years that have required DNA collection for specific categories of offences. Note that while only sex crimes were required by all 50 states back in 1999, the range of crimes requiring mandatory DNA collection has steadily increased. In 1999, only five states required that everyone convicted of a felony offence would need to provide a DNA sample. All 50 states now collect DNA from all convicted felons. Likewise, the number of states permitting collection of DNA from

<table>
<thead>
<tr>
<th>marker</th>
<th>current practice (as of 2014)</th>
<th>future potential</th>
</tr>
</thead>
<tbody>
<tr>
<td>autosomal STRs</td>
<td>core loci used to create DNA profile databases and to perform casework; data generated in laboratories with CE systems</td>
<td>expanded core set of loci enabling more international comparisons; data generated by NGS</td>
</tr>
<tr>
<td>Y-chromosome STRs</td>
<td>casework examination of 12−27 Y-STR loci with haplotype frequencies searched in population databases (e.g. YHRD.org); familial searching candidate pool restricted with Y-STR screening</td>
<td>larger population databases to improve haplotype frequency estimates; genetic genealogy database information combined with Y-STR casework data to help provide potential surname of perpetrator in some cases; rapidly mutating Y-STRs used to separate close male relatives</td>
</tr>
<tr>
<td>X-chromosome STRs</td>
<td>population data collected for 12± loci but only used occasionally in kinship cases</td>
<td>X-STRs and X-SNP markers routinely used to help address challenging kinship questions with testing performed on NGS platform in parallel with autosomal STRs</td>
</tr>
<tr>
<td>mitochondrial DNA</td>
<td>control region Sanger sequencing with haplotype frequencies estimated through population database searches (e.g. EMPOP.org)</td>
<td>full mtGenome by NGS to produce the highest resolution possible; larger population databases to improve haplotype frequency estimates</td>
</tr>
<tr>
<td>bi-allelic markers</td>
<td>a few dozen SNPs examined with multiple SNPePshot assays on CE platforms for simple phenotype or biogeographic ancestry prediction; some population data collected with insertion/deletion (InDel) assays</td>
<td>hundreds of SNPs or InDels for biogeographic ancestry and phenotype predictions tested on NGS platform in parallel with STRs</td>
</tr>
</tbody>
</table>

Table 2. Current practice and future potential for genetic markers used in forensic DNA analysis.

<table>
<thead>
<tr>
<th>offences</th>
<th>no. of states</th>
</tr>
</thead>
<tbody>
<tr>
<td>sex crimes</td>
<td>50</td>
</tr>
<tr>
<td>all violent crimes</td>
<td>36</td>
</tr>
<tr>
<td>burglary</td>
<td>14</td>
</tr>
<tr>
<td>all felons</td>
<td>5</td>
</tr>
<tr>
<td>juveniles</td>
<td>24</td>
</tr>
<tr>
<td>arrestees</td>
<td>1</td>
</tr>
<tr>
<td>familial searching performed</td>
<td>—</td>
</tr>
</tbody>
</table>

those arrested of a crime has increased from one in 1999 to 32 in 2014.

This growth has come because developers of DNA databases and software, suppliers of DNA testing kits and victim advocates have been effective in lobbying federal and state legislatures for new laws in the USA. With significant federal funding since 2004, DNA testing in the USA has grown to a level of around 1.5 million samples processed each year. Private laboratories, such as Bode Technology Group, Cellmark and Myriad Genetics, supplied the bulk of the initial DNA profiles generated because public laboratories did not have the needed capacity. More recently (through federal funding support), public forensic laboratory capabilities have grown in many regions to match the needs of convicted offender and arrestee DNA testing. FBI records indicate that over 250,000 criminal investigations have been aided through DNA testing as of late 2014 (http://www.fbi.gov/about-us/lab/biometric-analysis/codis/ndis-statistics).

(c) Critical challenges faced today
The success of DNA testing has brought significant growth, which in turn has brought new challenges particularly in the areas of sample backlogs and data interpretation. Laboratory automation and expert system data review can ease sample backlogs as can restrictive case acceptance policies. In some cases, law enforcement investigators may conduct ‘swab-a-thons’ at crime scenes and submit numerous items to try to solve a case through a DNA database hit rather than thinking carefully about which items may be most probative.

Data interpretation uncertainties are highest and errors are most likely to be made in situations with DNA mixtures from three or more individuals, especially with low-template DNA ‘touch’ samples. Work with these types of samples has been enabled in recent years with an increase in DNA detection sensitivity. With greater sensitivity comes the need for greater responsibility in data interpretation. Unfortunately, inconsistencies with handling DNA interpretation of complex mixtures adds to the challenge of obtaining reproducible results from multiple analysts and/or forensic laboratories.

4. The future predicted
The Olympic motto of Citius, Altius, Fortius—which is Latin for faster, higher, stronger (http://www.olympic.org/)—provides a framework for our discussion of where I believe the future of forensic DNA is headed. DNA testing protocols can be expected to get faster with rapid DNA instrumentation. Improved sensitivity and technology in recent years has enabled higher amounts of data to be recovered from biological evidence. Conclusions that are stronger can be drawn in many cases with probabilistic approaches under development. Each of these areas is discussed in greater detail later.

(a) Faster results
In the past few years, instrumentation providing a fully automated DNA profiling system has been introduced [28–31]. These rapid DNA instruments integrate the steps of DNA extraction, rapid PCR amplification of 15 or more STR loci, DNA separation, detection, sizing and genotyping. The initial versions of these rapid DNA systems can provide swab-in to profile-out results in less than 90 min for five buccal swab reference samples, but at a reagent expense that is currently about 10 times that of conventional laboratory testing. These systems are being marketed for generating DNA data in police booking station environments where a suspect may be held in custody for about 4 h while mug shots and fingerprints are taken. In order for these rapid DNA devices to reach their full potential in deployed environments outside of forensic laboratories, real-time DNA database searches will be needed and communication improved across the various elements of the criminal justice system. The legal framework to permit arrestee testing must also be maintained in order to collect DNA profiles in a police booking station environment. As these rapid methods are implemented in situations outside of a typical laboratory environment, it is important to not sacrifice quality for speed.

How fast will DNA profiles be able to be generated in the future? There are technological limits with any technique, which in the case of PCR amplification is a combination of the polymerase biochemistry, primer binding kinetics and the thermal cycling device. The fastest full 15-locus STR profiles generated to-date require 14 min of multiplex PCR amplification with 28 cycles of heating and cooling, which can lead to DNA profiles being produced in a laboratory setting in less than an hour [32]. While instruments may exist to heat and cool small volumes of liquid more quickly, sufficient time for polymerase extension of multiple DNA templates is required to obtain useful DNA profiles. As PCR amplification becomes faster, DNA separation and detection may become the rate-limiting step.

(b) Higher sensitivity and information content
The future of higher amounts of information in forensic DNA involves improved detection sensitivity, higher information content from expanded sets of core STR loci and possibly supplemental genetic markers, and deeper information from sequence analysis of alleles.

More data are available from biological samples due to improved sensitivity in PCR assays and information content of profiles generated. A consequence of improving DNA test sensitivity in recent years is the generation of more complicated DNA profiles for interpretation. It is important to keep in mind that just because a DNA profile can be obtained from as little as a single cell does not mean that the source of the profile is relevant to the crime event being investigated [10].

In terms of sensitivity, fundamental limits exist with PCR amplification due to stochastic (random) variation in sampling each allele at a locus [33]. These stochastic effects lead to variation in peak heights and peak height ratios for heterozygous samples during replicate PCR amplification. Even though two alleles are present at an STR locus in equal amounts in the DNA template, stochastic variation in the early rounds of PCR can lead to selective amplification of one allele over the other, which in extreme situations leads to allele drop-out (i.e. failure to detect the allele). Studies with single-source samples have shown that stochastic effects such as elevated stutter and allele drop-out occur at around 15–20 cells or 100–125 pg [33,34]. Allele drop-in may also occur when the number of PCR amplification cycles is increased to improve sensitivity [16,35]. When stochastic variation occurs, it becomes increasingly difficult to confidently pair alleles into genotypes and to correctly separate individual contributors in DNA mixtures. Thus, with stochastic
effects ever-present in low-level DNA PCR amplifications, allele drop-out and potential allele sharing from multiple contributors lead to greater uncertainty in the specific genotype combinations that can be reliably assumed. Combining results from replicate amplifications can help identify the true alleles in a profile [36,37].

Uncertainty in an evidentiary Q profile due to potential allele or locus drop-out can make it difficult to confidently exclude a specific suspect when a Q-to-K comparison is conducted. Probabilistic genotyping approaches (see next section) are being developed to try to explain observed data in some cases using computer simulations that estimate relative contributions of potential contributors. Some laboratories in the future may decide to establish a complexity threshold in order to halt efforts on poor-quality data (see ch. 7 in [4]). Because sensitive DNA detection technology has the potential to outpace reliable interpretation, the forensic DNA community needs to be vigilant in efforts to appropriately interpret challenging evidence without pushing too far. Clearly communicating the limitations of interpretation approaches used by investigators and officers of the court is crucial to avoid improper use of DNA.

Higher information content will be present in future DNA profiles that incorporate an expanded number of required core loci for inclusion in national DNA databases [25,26]. International data exchange will benefit from having more STR loci in common across various databases [26,38]. However, more information collected means more analysis and interpretation time and effort. While mixture detectability will improve with use of more STR loci, mixture interpretation will take longer with the larger PCR multiplexes since there will be more data to review. Although the capability may exist in future assays to simultaneously collect information from autosomal STRs, Y-STRs, X-STRs and numerous SNPs as has been proposed [39], it is unclear what the benefits would be of routinely attempting to collect far more information from evidentiary samples than would be available to compare to with typical reference profiles that only contain core STR loci. Routinely adding extensive information to reference samples in the hopes that this information may be of benefit in the future also means investments in interpretation time and effort—and may not yield a return on this investment unless evidence samples are examined at all or a substantial number of the tested STR and SNP loci. Because forensic DNA analysis involves a Q-to-K comparison, an optimal investment of effort is achievable when the same loci are examined in the Q and the K profiles to provide as many points of comparison as possible.

Next-generation sequencing (NGS), also known as massively parallel sequencing, provides the opportunity to collect information from numerous STRs and SNPs simultaneously [40]. In addition, sequence analysis of STR alleles provides a deeper depth of information by characterizing internal sequence variation for same size alleles that cannot be distinguished with CE analysis. For example, eight different sequence versions of allele 21 were identified in D12S391 among 197 samples examined [41]. The single detected peak in apparent homozygotes using CE may be resolvable into heterozygous alleles with NGS at some hypervariable STR loci [41,42]. In addition, mtDNA genome sequences can be generated with NGS [43]. However, the orders-of-magnitude difference between mtDNA, which may contain thousands of copies per cell, and nuclear DNA, with its two copies of each allele per cell, will probably preclude the routine combination of autosomal STRs and mtDNA into a single, robust NGS assay.

Only time will tell whether future routine DNA testing proceeds down the path of collecting and analysing a deeper depth of allele information available from NGS methods. The additional collected information with NGS requires specialized bioinformatics tools to process the data and a more complicated allele nomenclature scheme, and comes at a cost of more analysis time and more data storage space. What is interesting for research studies may not always be practical or necessary for routine application to forensic cases.

(c) Stronger conclusions with challenging samples

Improving the ability to decipher and interpret DNA results from challenging samples provides probably the largest opportunity for future advances in forensic DNA analysis. Challenges can come through poor-quality samples containing degraded DNA or PCR inhibitors that fail to produce an informative Q profile. Smaller PCR amplicons in the form of miniSTRs [44,45] or SNPs [46] have been shown to help recover information from damaged DNA templates. The DNA target size can be reduced with some STR markers in NGS assays versus what can be done when CE is the detection platform where limitations exist in terms of dye colour space.

DNA database searches that fail to produce a matching K reference profile can be essentially expanded through loosening the stringency of the search and using genetic inheritance principles to produce a ‘familial’ search in an attempt to identify close biological relatives of the source of the unknown Q profile [47]. Familial searching typically experiences low success rates due to a lack of close relatives of the true perpetrator in the database. In addition, true relatives in the database may be missed due to the search strategy used or specific inheritance patterns that leave a true sibling (in the database) with different alleles compared with the Q search profile (see appendix 2 in [3]). False positives when the Q profile has common alleles that are shared by unrelated people can result in long candidate lists. Y-STR testing on male samples from a ranked candidate list can help filter false positives [48]. While familial searching is unlikely to be conducted on a national level in the USA [49], it has produced some successes in the UK through aiding identification of 41 perpetrators in 188 police investigations [50]. Genetic privacy concerns are often raised in opposition to familial searching [51], and so the future of this technique is probably limited due to practical and privacy issues.

When no K reference sample is available for comparison following a database search, it may be possible that future DNA capabilities will include prediction of external visible characteristics [52], such as eye colour or hair colour [53]. Biogeographic ancestry estimation may be helpful in some case scenarios [54]. Microbial DNA may also play a role in future investigations to lend support to circumstances involving microbial transfer when touching a computer keyboard [55] or having sexual intercourse [36]. This analysis will likely be performed with NGS or other high-throughput methods in the future.

Complex DNA mixtures, containing genetic data from more than two individuals, especially if any of the individuals are related, offer one of the largest challenges for the future of forensic DNA analysis. Probabilistic genotyping offers a way to strengthen conclusions with challenging DNA mixture results...
that might otherwise be declared inconclusive under a binary approach to interpretation (see appendix 4 in [4] for a detailed example). Software programs have been developed to perform probabilistic genotyping using what has been termed ‘semi-continuous’ (where only allele information is used) or ‘fully continuous’ (where alleles and peak height information are used) [57–63].

Supporting stronger conclusions with probabilistic approaches on complex evidence will require validation of software used to understand its limitations [64]. Scientists using these software tools must not treat them as ‘black boxes’ generating reports that they themselves do not understand. Equally important will be the need for effective communication of the meaning of results obtained to those who will receive and use the laboratory reports to make decisions. These reports must appropriately communicate limitations of methods, models, assumptions made and interpretation applied with the case results.

(d) Other factors

I believe in many ways we are at an interesting juncture in forensic DNA. We have over the past few years developed great technological capabilities that in some ways have outpaced our ability to appropriately interpret evidence results obtained. We have emphasized methods for enhanced sensitivity while not working equally hard to improve our understanding of possible genotypes that may compose the evidence results (i.e. the specificity of results). Thus, for forensic DNA typing to move forward responsibly we must improve the framework and consistency of DNA evidence interpretation [4]. To aid efforts in moving forward in DNA interpretation, I support the 13 recommendations that Peter Gill makes in his recent book ‘Misleading DNA evidence: reasons for miscarriages of justice’ [10].

Cost will play a role in the future use of DNA in forensic science. Important factors influencing cost include competition, centralization and communication. More competition is needed in many areas of the forensic DNA supply chain in order to drive prices down. Likewise, automation within laboratories and supply manufacturing lines is crucial to reduce potential quality challenges either from potential sample swaps or consumable contamination. To accomplish this type of automation cost-effectively, economies of scale are needed. The future will probably be better served by central, focused laboratories than by many small operations with redundant capabilities. While these thoughts would seem to suggest that rapid DNA testing at police stations will be unsuccessful, it is important to realize that the current political climate supports local funding over sending money to central, specialized facilities. Crime is largely committed by local perpetrators [65]. In a large country like the USA, we have many local laboratories with efforts that may be uncoordinated and in some cases redundant to neighbouring areas. Smaller countries, such as the UK, may be able to support more centralized laboratory systems. Thus, there is not a one-size-fits-all solution for the future of forensic DNA.

A second reason that technological innovation may be slow or non-existent relates to validation requirements for methods used in forensic applications. Slow technology adoption by the forensic community can result in few companies wanting to invest in long-term efforts to bring new products to market where a return-on-investment may be slower than it is in other fields. The scrutiny that products and manufacturing processes can undergo during adversarial legal battles may also keep potential companies out of the forensic DNA marketplace.

Another important reason that forensic DNA may not involve cutting-edge technology is the need to maintain connection to legacy information from results already collected. The fact that the USA has 14 million STR profiles (most with the 13 core STR loci selected in 1997) in its national DNA database means that switching to completely different genetic markers, such as SNPs or InDels, is not feasible (see also [15,66]). Switching to another set of genetic markers for established DNA databases does not make sense when doing so would, in the case of the USA, render obsolete a billion dollar investment. Thus, genetic markers used in future DNA databases will most probably include current core STR loci plus additional ones to avoid adventitious matches [25,26].

Knowledge and understanding of our DNA sequence information are constantly evolving. Human genomes are more complex than originally imagined and research studies are continuing to uncover additional information (e.g. [67]). While currently used STR markers show no significant value for predicting genetic diseases [68], the information landscape is continually changing with additional data being gathered.

Early uses of DNA databases involved using a large net to gather as many reference profiles as possible from previous offenders and more recently from those arrested for specific crimes. In the USA, laws have steadily increased the number of offences requiring DNA collection for testing purposes (table 3). However, I would predict that technology advances will not be the rate-limiting step in the future of DNA database growth. Rather the appetite of the general population to accept DNA use and continue to make significant resource investments in this effort will be. Public forensic laboratories will probably need to market the value of their work more in a fiscally challenging future. UK database growth has slowed in recent years due to privacy concerns raised. Other national DNA databases are likely to face similar challenges. I would not be surprised to find continued support for the position argued by Robert Gaensslen on this topic in a 2006 article that ‘absent specific legislative guidance, . . . the ‘default’ position [should] be more protective of individual privacy rights’ [69, p. 375].

5. Concluding thoughts

Successful research is an important key to the future in any scientific endeavour and is dependent on sufficient, sustainable funding. Beyond just publishing scientific articles and presenting work to colleagues at conferences, training of practitioners and stakeholders is essential for effective implementation of new techniques and ideas. Keys to success include resources and recognition of the value of training support in terms of funding for the presenters and setting aside time for the participants. Continuing education needs to mean more to participants than just checking a box that helps a laboratory
pass an audit. Quality work results when scientists have a desire to learn and try to think critically about casework results rather than mindlessly following a protocol.

The future of forensic DNA will have an impact on other areas of forensic science. As DNA has become successfully used in the past few decades to solve crimes, funding and personnel resources have been reduced in other areas such as trace evidence—and sometimes completely eliminated from a forensic laboratory’s capabilities, which may bring long-term consequences. In my new role at NIST as Special Assistant to the Director for Forensic Science (which I have had since April 2013), I have an opportunity to look beyond DNA. Many challenges exist as we seek to improve forensic science efforts across various disciplines. I am involved in setting up and running the US National Commission on Forensic Science and the Organization of Scientific Area Committees [70]. There are many lessons from DNA that can be applied to other areas in forensic science. We should seek to understand issues, challenges and potential solutions outside of our focused, discipline-specific view and bring potential solutions back to our own areas of expertise. Improving communication across various forensic disciplines is just as important as improving communication between laboratories, law enforcement and the legal community. International cooperation and standards setting will enable forensic DNA testing of the future to become even more robust than it currently is.

Better communication across stakeholders in the criminal justice system is a primary need in order to maximize the value of DNA testing efforts. More than a billion dollars has been invested over the past decade to grow the USA national DNA database to its current state of almost 14 million profiles. While many crimes are solved with DNA database matches, I think it is safe to say that many more probably could be solved if the communication between investigators and forensic DNA scientists were improved. With fiscal restrictions for the foreseeable future, significant investments in DNA databases may be a thing of the past. Forensic scientists will have to work smarter with the resources available to them. I believe it is also important to appreciate that a DNA test, even with the best technology available, is only as good as the sample collected. Thus, working closely with crime scene investigators is essential in gathering quality evidence. As we learn from the past and perform well in the present, we will create a future where forensic DNA analysis can play an important role in criminal justice systems around the world.

Competing interests. I declare I have no competing interests.

Funding. The author is supported by internal funds from the National Institute of Standards and Technology as part of its forensic science efforts.

Acknowledgments. I have been fortunate to work with many excellent scientists, colleagues and friends over my career. I am grateful for the encouragement and mentorship of Ralph Allen, Bruce McCord and Dennis Reeder early in my career. I have enjoyed working with members of the NIST Applied Genetics Group including Pete Vallone, Mike Coble, Becky Hill, Margaret Kline and Dave Duewer. We have conducted a lot of interesting research together over the years and I miss regular association with them as I have moved into a new role at NIST. I am thankful for the opportunity to serve as an Associate Editor for Forensic Science International: Genetics over the past 8 years. Ideas expressed in this article have come from reading the literature, writing my Forensic DNA Typing textbooks and learning from others at numerous conferences over the past two decades. Points of view in this document are those of the author and do not necessarily reflect the official position or policies of the US Department of Justice or the National Institute of Standards and Technology. Certain commercial equipment, instruments and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that any of the materials, instruments or equipment identified are necessarily the best available for the purpose.

References


