300 Matches Per Week – The Effectiveness and Future Development of DNA Intelligence Databases – Parts 1 and 2

D.J. Werrett and R. Sparkes, Forensic Science Service

PART 1

The first part of this paper reviews the performance of the National DNA Database to date, places it into context with criminality in the United Kingdom and demonstrates the success of the Database and how its role is developing as an important part of the Criminal Justice System in the UK. The opportunity is also taken to review the operating processes and quality framework as well as outlining the future strategy for the developing situation in the UK and Europe.

This paper will give the reasons for the setting up of the Database and its Terms of Reference, the headline statistics achieved since April 1995 through October 1998. It will consider the operational framework with a view to getting the basics right and achieving high throughput in an environment where multiple suppliers are providing profiles to the National DNA Database through a quality framework.

It will introduce the concepts of the strategic use of DNA databasing by the police and consider the strategy for the future in response to the requirement.

WHY A DATABASE?

The concept of a National DNA Database was developed in early of 1994 and passed through legislation in October of 1994 before enactment in April 1995 (1). The Database was predicated through a tripartite arrangement between the police, the Home Office and the Forensic Science Service. Careful examination of crime patterns in the UK revealed that 60% of court appearances deal with approximately 21% of offenders and that these offenders commit a variety of crime. Additionally volume crime, such as burglary and car theft, is often associated with drug abuse and this in itself is a major multi million pound problem for society. Further, the trend of recidivism has been rising over the years, such that individuals with six or more court appearances have an 80% chance of re-offending and there is a rising recidivism rate after the first offence; several cohorts of individuals have been followed with regard to criminality: In 1953, a 14 year old, convicted for the first time, had a 51% chance of re-offending, by 1973 the chance of re-offending had risen to 77%.

The criminal population is dynamic, individuals, mainly men (85%), gaining a first conviction between the age of 14 and 19 and then from the age of 25 onwards an increasing number of individuals desist, reaching 60% by the age of 34 (2). Submissions to the National DNA Database reflect this age pattern, 30% of buccal scrapes submitted to the National DNA Database are from 14 -19 year olds. Further, anecdotal evidence from fingerprint databases has suggested that there are individuals who commit a variety of crime, ranging from petty to very serious crime, often starting their careers off with petty crime. The tripartite group, therefore, proposed that the National DNA Database should contain DNA profiles from suspects of crime as well as those convicted of crime, with the proviso that in those cases where the prosecution was not proceeded with, or an individual was acquitted, the record and the DNA profile should be removed from the Database. The proposals were supported by the Home Secretary and subsequently by Parliament and this resulted in the launch of the Database on 10 April 1995.

The law was re-drafted so that buccal scrapes or hair roots could be taken by police officers from suspects of recordable crime (A recordable crime being defined as any crime that might carry a prison sentence). In either case there is a compulsion to provide the sample with no right of refusal. Each year in the UK there are approximately two million arrests and about 450,000 of these are new individuals, 85% male, 15% under 17 and 54% unemployed, who will receive new criminal record office numbers and therefore potentially could be sampled for the National DNA Database. Forty per cent of those arrested are subsequently convicted.

CUSTODIAN, SUPPLIER AND USER REQUIREMENT

The arrangements for the Database are such that the Forensic Science Service acts as custodian and supplier with the data being owned by the police forces. Under the custodian arrangement the Forensic Science Service manages the data and supplies information to the forces with regard to samples being analysed and matches being found between suspects and crime scenes, for example.

At the outset the police, through the auspices of the Association of Chief Police Officers (ACPO), set out their user requirement which comprised:

- A 7 - 14 day turnaround time;
- Individualisation of DNA profiles;
- An agreed economical price, currently approximately $55;
- Storage in perpetuity at minus 80° (should the technology change the samples would all be re-analysed);
- Zero error rate and the issuing of matches from person to person, person to scene and scene to scene.

Over the years the requirements have changed and these will be discussed later.

THE NATIONAL DNA DATABASE PERFORMANCE FIGURES

The total number of samples from suspects and convicted loaded to the Database is now 391,494. And 62,857 profiles have been removed because the individual has either been acquitted or the case has not been proceeded with. The backlog now stands at 99,273, but at the time of writing has been reduced to 84,000. The backlog has fallen by over 35,000 samples in the last two and a half months. The total number of profiles held on the undetected crime scene Database is 39,142. The total number of matches reported between suspects and crime scenes is 28,128 and the total number of crime scene to crime scene matches is 5,936. The number of matches reported per week varies from 300-500. However, in the current week 576 matches have been reported. These include person to crime matches of: three murders/ manslaughters, five rapes, three other sexual offences, one arson, nine serious robberies, two grievous bodily harms, two abductions and two aggravated burglaries. The success of the Database has been such that clearly the customers want more and it is expected that approximately 250,000 samples from suspects of crime will be analysed this year rising to 350,000 next year. In addition, 36,000 stains from undetected scenes of crime are likely to be submitted as well as 30,500 stains being generated from casework. The National DNA Database the facility is also used to analyse samples from intelligence screens. These are screens where the investigating officer chooses to investigate a case through asking for volunteer samples from potential male suspects who live in the area. These are entirely voluntarily given and approximately 15,000 samples are analysed each year (see below).

OPERATIONAL FRAMEWORK

The Database is used for intelligence purposes only. If a police officer wishes to use the results of the analysis in court then a further examination is undertaken within a caseworking laboratory.

At the outset, with regard to analysis of samples for the Database it was decided that processes should be duplicated when appropriate and if necessary to reinforce quality cover at the expense of throughput to reinforce quality. In a Database situation there is no “safe” error. For example, in casework it is often argued that an error of omission, that is failure to produce evidence, is much less important than an error of false inclusion of an innocent individual. In the Database situation failure to identify an individual who may go on to commit more and serious crimes could be just as important as the wrongful matching of an individual. At the outset it was therefore decided that the following regime would be followed:

From each individual two buccal scrapes are submitted. A portion of one of these is stored at minus 80° and will be re-analysed should the technology change, a half of one swab is then extracted and taken through the analysis process.

In 1995 when we first assessed the technique the interpretation stage was identified as the most demanding. It was therefore decided that this would be carried out by two operators independently, designated genotyper A and genotyper B, and that the results of their independent interpretations would be brought together and examined by a third individual who would check for differences of interpretation. Thus, the first half of the swab is taken through the process, the interpretation is carried out and the profile is added to the Database. Overnight the profiles from suspects are used to challenge the Database of profiles from undetected crime scenes. If a match is obtained a portion of the second swab from the suspect is retrieved from storage and the process, including genotyper A and genotyper B, is repeated. After this second complete analysis from extraction to DNA profile has been used to confirm the original result a report is then issued to the police forces for intelligence purposes only.

In the first two years of operation assessment was made of the error rate using this regime. About 5,000 duplicate results were examined and the error rate found to be less than 0.05%. This was brought to the attention of the customer, the police, who agreed that for volume crime cases, such as burglary, it would no longer be necessary to completely repeat the analysis. Thus, for volume crime matches they are reported after the first analysis stage, but for serious crime matches the samples are completely re-analysed for a match report to be issued.

PEOPLE AND EQUIPMENT

When setting up the Database careful consideration was given to the contribution made by individuals to the process. A consultant process engineer devised a system that assigned teams of individuals to be given tasks: extraction, amplification, gel running etc. The output for each team would therefore be passed along the line to the next team. This did not work, staff failed to gain ownership of the final output, the DNA profiles, individuals were reorganised into teams to perform all of the techniques of the process from beginning to end. We estimate that a team of about 16 scientists will analyse approximately 35,000 samples a year using the regime described above and that it takes about one month to train a technician in the front end of the processes and a good graduate three months for the back end. A team made up of technicians and graduates is sustainable for six months to one year before training throughout the process becomes a necessary part of team building. We have learnt through the training programme that great emphasis should be placed, not only on grades and qualifications but also on ability and competence. Competency records, details of proficiency testing programmes and outputs are maintained on all members of staff within the Database. Re-training programmes have been very successful at running the quality of output. Another aspect of the Database is that where possible, teams are given dedicated equipment, again this encourages ownership of the process and outputs. We have benchmarked outputs on the 377 ABD sequencer, they are run for up to 24 hours a day and as we convert them to 96 lane machines we would expect to achieve outputs of between 50,000 and 100,000 profiles per sequencer.

INDIVIDUALISATION

The National DNA Database of England and Wales uses six STRs: HUMTH01, D21S11, D18S57, D851179, HUMVWFA/A, HUMFIBRA (FGA) which gives a match probability of about 1 in 50 million (3,4) We also use the XY homologous locus amelogenin (5) to determine the sex of the donor. This match probability, 1 in 50 million means that for the UK for any given profile there is a 50% probability that there are one or more individuals with the same profile within the country. This has proved to be a useful statistic in educating both the judiciary and the police and there is now a tacit agreement between the judiciary, the Crown Prosecution Service, the police, that when a case is brought to court it is expected that there will be supporting evidence alongside the DNA evidence. It would be highly unusual for any case to be brought to court in the UK with the sole evidence being DNA. Thus, the 1 in 50 million statistic forces the issue of corroborating evidence and over the years individualisation has no longer become a requirement of the customer, the police.

ADVENTITIOUS MATCHES

In the next year we expect to examine approximately 200,000 samples from suspects and 40,000 stains from scenes of crime. If we consider the number of possible pairways comparisons that that will allow us to make (200,000 x 40,000) approximately, and the relationship with the average probability match figure of 1 in 50 million, we would expect to get between 100-200 adventitious matches a year, i.e. 'innocent' by chance matches between an individual on the database and a crime scene stain. Although this is less than 0.1% of matches, and of course we will get adventitious matches through partial profiles placed on the database, we feel it operationally necessary to move to a more powerful multiplex to eliminate adventitious matches. We are doing this within European framework, in the next six months we expect to move to a multiplex system which will provide probability match figures of approximately 1 in a million million. Of course this will in itself raise serious issues about individualisation which, up until now, have not had to be addressed because of the 1 in 50 million probability match figures of the current system.

MATCHES BY OFFENCE TYPE

One of the statistics that the Database can generate is matches by offence type. We tend to group these into matches from persons to serious crime offences and matches from persons to volume crime offences. Taking a period from April 1997 to August 1998 we find that for serious categories of crime: such as murder/manslaughter and rape, we have reported 34 and 136 matches respectively. For volume crime cases, such as burglary of dwellings and other burglary (which includes burglary of industrial premises), we have reported 8,787 and 8,193 matches respectively. Interestingly for auto crime, which is not within the categories targeted by the police, we have reported 3,912 matches. For the sake of comparison between countries, such as the UK and United States, it is worth placing these matches in the context of criminality within the country itself. In 1997 in the UK there were 711 murders, 655 manslaughters, and 6,337 rapes. These represent approximately 0.1%, 0.1%, and 0.6% of the overall crime statistics. In comparison there were 520,000 burglaries of dwelling houses and 495,000 burglaries of other premises, equivalent to ~50.0% and 48% of the crime statistics. If we compare these percentages with the match percentages then we find that for murders the match percentage is ~0.15% on the Database as compared to ~0.2% for murders and manslaughters within the crime statistics, rapes ~0.6% compared with ~0.6% and for

burglary of dwelling and other burglary ~39% compared to ~50.3% and ~37% compared to approximately ~48%. Thus, Database matches approximately reflect the criminality in the country in similar proportions. Countries with, for example, higher murder rates may expect to get higher matches for unsolved murders.

Match numbers and the effectiveness of matches may not be equitable, for example, in the UK approximately 90% of murders are solved and 50% of those tend to be domestic related offences. This leaves a relatively small number of murders, 10% of the 711, approximately 70, which fall into the 'difficult to investigate' category. If the matches from the DNA Database for murder/ manslaughter is approximately 30 in a given year, and the information is directed against those cases that are 'difficult to investigate' then the impact on policing, saving of police time could be dramatic. Another aspect of the performance of the Database are the links that are achieved for unsolved cases between volume crime and serious crime. Whilst reflecting on criminality in general it is also means that sometimes detection of, for example, fingerprints or DNA profiles at volume crime cases and putting a name to those cases could in themselves solve some very serious crime. In the last year or so we have shown links between murders and burglaries, links between rapes, traffic offences and burglaries, links between serious robberies and auto crime, links between woundings and auto crime. Some of these links undoubtedly will lead to successful prosecutions.

SUBMISSION RATES AND EFFECTIVENESS OF MATCH REPORTS

We have examined the correlation between the number of stains that a Force were submitting and the success rate in terms of matches reported to that Force. We were concerned that there may be a tendency for those Forces that submit lots of stains to get less information back per stain. It appears there is an apparent straight line correlation between the number of stains from scenes of crime submitted and the number of matches reported back.

A most telling factor is the speed with which the stains are analysed. For example in one particular Force, Sussex, we found that a return of 225 matches yielded 74 primary detections, but there was a high rise in detections when analysis of the stain took less than 28 days. The Force now believes that the detection rate could be as high as 78% of all matches, if the matches are reported back promptly to the Force. Another Force, Kent, examined the effectiveness of matches and found that 74% made a positive contribution to detection and that some led to multiple primary detections. The best case scenario being 50 primary detections accrued from a single Database match. A large metropolitan force, known as the West Midlands Police Force that covers conurbations within Central England, ran an evaluation coupled with an initiative to collect samples from suspects and stains from scenes of crime. In the last nine months of 1997 they collected ~12,500 samples from suspects and ~2,500 stains from scenes of crime. They received ~1,700 matches, a 73% success rate for scene stains and if one takes the suspect samples on their own for that period a 10% chance match for suspect samples.

EVIDENCE SUBMITTED

The type of evidence submitted has changed considerably since the inception of the National DNA Database. Traditionally blood stain analysis and indeed semen stain analysis has formed a major part of evidence analysis within the Forensic Science Service in the UK. The growing emphasis on taking material from scenes of crime, and there are now more stains analysed from undetected crime scenes than there are stains analysed within normal casework procedure, has switched the emphasis away from semen to saliva. Blood stains and blood swabs still comprise 12% and 46% of submissions respectively but now 3% and 27% of submissions are saliva stains and cigarette butts respectively. Saliva stains, particularly from face mask discarded at the scene of armed robberies and also head hair from similar sources, make up 4% of submissions. An interesting statistic is that more than 50% of cigarette butts analysed produce DNA profiles that can be sent to the National DNA Database. Approximately 10% give mixed profiles but a significant proportion of these give major and minor profiles and the major profile is sent to the National DNA Database.

Through the Database the changing work pattern goes further than changes in evidence type submissions. Traditionally forensic science has been part of a supply chain: samples would be obtained from scenes and suspects, some information with regard to the offence would be supplied to the laboratory. The laboratory would carry out examinations, a report would be produced that would go the Prosecution Service, or indeed the Defence, and finally where appropriate a court case would ensue. Through information being provided by the DNA Database the laboratory is now instigating investigations. Inceptive intelligence information produced by the National DNA Database is leading to new forms of crime investigation that are now becoming integrated within police procedures in the UK.

QUALITY FRAMEWORK

When the National DNA Database was set up in April 1995 it was decided that a framework for Quality should be developed that would allow other organisations capable of DNA analysis to carry out DNA profiling for police forces and supply those DNA profiles to the National DNA Database. All of the DNA profiles on the National DNA Database are owned by the police forces through the Association of Chief Police Officers (ACPO). The Database is administered and the software owned by the Forensic Science Service. Two roles were developed, the Forensic Science Service as Custodian, and the Forensic Science Service as Supplier. As a Supplier to the Database we carry out DNA profiling according to the standards set by the FSS as Custodian. There is, therefore, a Chinese wall within the Forensic Science Service. The Chief Scientist acts as Custodian to the Database and the Director of DNA Services the operational Supplier. All Suppliers must comply with the quality framework, it comprises: accreditation to international standards, including NIS46 and NIS96, ISO25, ISO9000, Series ISO9001. External accreditation was considered to be important. The accreditation procedure involves scientists, not in themselves forensic scientists, auditing the work to the international standards. Following accreditation, laboratories are expected to carry out two proficiency tests to demonstrate, not only their ability to obtain the right results but also to obtain results of a quality that is well above the base line. The assessors of the proficiency test programme are looking for the laboratory to obtain a benchmark not just the right result. When a laboratory has successfully completed the proficiency tests it is allowed to supply profiles to the Database but must also subscribe to an ongoing Quality Assurance Programme. The Quality Assurance Programme comprises both 'Declared samples' and 'Undeclared samples'. The former being buccal scrapes submitted to the profiling laboratories as a known Quality Assurance Trial, the latter being buccal scrapes submitted as profiles from suspects through the police forces without the knowledge of the profiling laboratory. Finally, within the UK a DNA Suppliers Group has been set up where all the leading laboratories meet on regular occasions, agree changes to protocols and provide evidence to one another of the validation of changes to the protocol. This has proved to be a very fertile group in terms of ensuring that common high quality standards are maintained for the supply of DNA profiles to the Database.

INTELLIGENCE SCREENS

Finally to conclude Part 1 of this presentation I would like to mention the use of Intelligence Screens within the UK. To date there have been 81 screens. Investigating Officers use these screens to eliminate individuals within a given area where an offence has been committed. Samples are given entirely on a volunteer basis, it is not within the auspices of the National DNA Database. We have 43 active screens and 10 have been discontinued. They have involved the analysis of 24,000 samples and 28 screens have resulted in matches between individuals and serious crimes: 9 murders and 17 rapes. Prosecutions have taken place and/or pending for each of these investigations. On average about 500 samples are submitted before a screen reveals a match, but there is a large range between 30 and 4,500, however for more than 90% of matches less than 1500 samples are submitted. Investigating Officers within the UK now recognise intelligence screens as potentially a very powerful tool for investigating murders and rapes.

Part 2

INTRODUCTION

Since its inception in April 1995 the National DNA Database has become well established and demonstrated the efficacy of DNA profiling to crime detection. We are now in a position to move on technologically and intend to take as far as possible the application of STR analysis using refinements to existing technology and the supporting infrastructure. This will be achieved by considering and improving four aspects of the current technique. Firstly we will augment the current multiplex to deal with the adventitious hits and robustness of performance issues; in conjunction with this we will introduce automation to develop high throughput reliably; thirdly we will improve the output of current machinery, including changing the 377 automated sequencer from a 36 lane to a 96 lane machine; and finally we hope to improve routine interpretation through expert systems.

ENHANCEMENT OF MULTIPLEX

Over the next few years it is anticipated that the Forensic Science Service will analyse between 300,000 - 500,000 samples per annum, this would obviously require a large scale manufacturing commitment if we were to manufacture the multimix in-house. Dealing with the number of adventitious hits that occur with a database of this proportion is also a priority; as is increasing our ability to analyse highly degraded forensic material and generally improving the overall robustness of a high throughput technique.

In addition to our own requirements there is currently a need to exchange data across Europe, the FSS, together with laboratories from the European Network of Forensic Sciences Institute (ENFSI) DNA Working Group is undertaking this work. It is envisaged that this will be achieved by the development and use of a common multiplex which will provide the information necessary to track criminals, particularly sexual offenders, across borders. It is also important that the multiplex is sufficiently robust to work in a variety of laboratories in a reliable way. The Strategy has been therefore to combine two user requirements.

Through the ENFSI DNA WG Interpol has designated loci for international use. HUMTH01, HUMVWFA, HUMFIBRA and D21S11. For the FSS to avoid loss of existing data, or re-analysis of samples, D851179 and D18S51 should also be included. The multiplex should be optimised for use with the various machines currently available throughout Europe, such as the 377 and 310 automated sequencers.

The performance criteria for new multiplex acceptance includes: probability of match of 10^{-10} or less, one tube reaction; enhanced performance with degraded samples; balanced profile morphology; single base pair resolution and low artefact activity.

Two companies; Promega Corporation and Perkin Elmer, have made proposals for a new multiplex that could meet the above criteria and they were invited to take part in an exercise involving 36 laboratories from 25 countries. The proposals submitted were PowerPlex 2.2™ (6) and AmpF*l*STR™ SGMPlus™ (7), respectively, (both have now been commercially advertised as potential new products). The purpose of the exercise was to allow each laboratory the opportunity to evaluate the products. A common set of DNA samples were produced by the laboratory in an attempt to eliminate sample variation and protocols were supplied to facilitate a standard approach. As far as possible the DNA samples were designed to initiate casework material. The results of this exercise should be available in December of this year.

INTRODUCTION OF AUTOMATION

High throughput production of DNA profiles has become routine with the development of robust multiplexes and automated fluorescent based detection. However, to maximise the efficiency of production, it is necessary to consider the streamlining and automation of sample and data handling at all stages of the DNA profiling process. The current processing practices were evaluated to determine the steps that would most benefit from the intervention of automation. Currently Automation within the National DNA database has focused on the development of a sample logging machine and two robotics systems (8). The latter are controlled by Overlord, a supervisory software package which allows the execution of both DOS and Windows™ - based packages plus integration with a Laboratory Information Management System (LIMS). The LIMS affords the database complete sample tracking, which can be both laborious and time consuming, and handling continuity for individual samples and/or batches of samples.

This integrated system forms the basis of an automated high throughput DNA process. As with the original manual approach, samples are extracted, quantified, amplified, separated by electrophoresis and analysed. However, processing is streamlined by the use of 96-well microtitre plates in conjunction with sample tracking based on uniquely barcoded microtitre plates and the position of the samples within these plates. Buccal scrapes are submitted to the unit for analysis and entered

onto the logging system. To date we have been unable to successfully develop a suitable automated extraction technique using Chelex™; the current extraction medium used by the FSS, however, we are currently looking at several options including magnetic bead technology and Qiagen™. Therefore, manually extracted DNA samples are loaded into microtitre plates and automated quantification, dilution and PCR set up are undertaken by a robotic system comprising a 9 channel Hamilton 2,200 pipetting stations, an articulated robotic arm on a linear 3 metre track, plate hotels and a fluorimeter for DNA quantification using a Picogreen assay. Following amplification, the PCR products are aliquotted by a second pipetting station and then loaded into an automated DNA sequencer by multichannel pipette. Once analysed completed STR profiles are added to the database whilst samples requiring further work are flagged on the LIMS and exported to the appropriate robotic system for re-processing.

At present the system described above is in validation with a pilot study pending. It is our intention to roll this out to full production by April 1999. Early results indicate a first pass success rate of over 85%. A re-PCR then has a 70% chance of success and a re-run 90% this yields an overall success rate, including repeats, of 97% which is comparable to the manual system.

96 LANE UPGRADES

Gels are currently run in 48 lane format but it is envisaged that this will change to 96 lanes using ABD 377 sequencers. Two routes are being explored. Upgrades of the machines with proprietary 96 lane software from the manufacturer. Alternatively modifying existing 377 hardware and replacing proprietary software, to include bespoke software which feeds directly off a modified data collection system (9). The latter is an attempt to improve both quality and quantity of the data collected. It would also give access to raw data for expert system processing.

In theory, current sequencers are run 6 times per day, over a total of 302 working days per year, with an 85% first pass success rate. The 36 lane format incorporates 6 control lanes and 20 for the 96 lane format. In practice machines will undergo downtime due to any number unforeseen circumstances. Once implemented the output from any one sequencer in a year could potentially increase output from 46,000 to 117,000, more realistically this will increase from 15,000 to 50,000.

EXPERT SYSTEMS

The rapid development of the NDNADB made us acutely aware of the need for detailed examination of the process flows required to ensure that samples are analysed correctly, results interpreted and when necessary samples re-analysed (as mentioned in part 1). At the genotyping stage, after analysis, interpretation is required prior to allele designation and therefore there is room for subjectivity. This step is therefore carried out independently by two operators and their resulting files compared and examined by a third individual, who will make a final decision. Each part of this process flow was examined in detail, documented and timed. This initiated the STRess (STR expert systems suite) project (10). The goal of this project was to address the considerable time spent analysing and genotyping gels. STRess is a Windows™ - and Macintosh based program that accepts raw data, generates a file of allele designations and then compares this file to one generated by a human operator. The primary objective was to provide a system that would carry out the same analytical processes as a human, to at least the same, if not higher standard. An additional benefit would be a decrease in processing time, but this would have to be achieved without sacrificing quality. Quality monitoring and the integrity of results was an important part of the specification; this required a fully documented audit trail.

The basic function of STRess includes 4 processes:

- Process 1 simply accepts the data produced, this can either be obtained from Genescan™ software by exporting a data table, or from Genotyper™ software by running a raw data macro. This raw data file defines each peak in terms of position, height and area.

- Process 2 is the heart of the system and is responsible for cleaning the raw data ready for allelic designation. Sample lanes are cleaned using rules generated from vast operator experience and data collections. The underlying philosophy of the system is to move data from one file to another (rather than remove the data altogether), this allows a clear audit trail that can show the fate of every peak from the input file.

- Process 3. Once the sample data is cleaned the remaining peaks have to be designated. In theory this needs to be done by reference to allelic ladders run in pre set lanes on the gel, however, this presents a number of problems; the gel ladder may shift from the ideal; the gel ladder may be incomplete with not all possible allele represented or there may be missing

ladder peaks. To circumvent these problems, STRess constructs a 'virtual' ladder. This is done by comparing the gel ladder with a known pattern of peaks determined when the acrylamide mix is validated. The shift between ideal ladder and the gel ladder is determined. This shift is then used to compensate for missing peaks. Thus the virtual ladder is built up from true gel peaks and peaks calculated for the observed shift. Following the creation of the virtual ladder, the remaining peaks can be designated. Once the designation phase is complete customised comments can be added depending on a range of post designation rules–this is known as allele qualification and will draw the attention of a human operator to any anomalous results which will require further qualification or interpretation.

- Process 4. The final stage is to compare the human output with that from STRess. The STRess compare function produces a table of differences that allows the human operator to investigate and arbitrate before sending the profile to the NDNADB.

In summary, by studying thousands of sample operations of the system and comparing them to the interpretation by human operator, rules can be continually refined and the program tuned for maximum efficiency. To date a saving in time of more than 40% has been achieved by use of the STRess program. It is currently in use in all NDNADB teams and is now being introduced into some caseworking teams. It has become an invaluable troubleshooting tool and has given us a better understanding of the rules required for concise interpretation.

Thus, the goal of the FSS is to produce a benchmark system using a 377 in 96 lane format supported by automated extraction, quantification, PCR, followed by automated analysis of results.

REFERENCES

1. Werrett D. J. (1997) The National DNA Database. *Forensic Science International*, **88:**33-42.
2. Graham J., & Bowling B. Young People and Crime. *Home Office Research Study 145.*
3. Sparkes R., Kimpton C., Watson S., Oldroyd N., Clayton T., Barnett L., Arnold J., Thompson C., Hale R., Chapman J., Urquhart A., and Gill P. (1996) The Validation of a 7-locus Multiplex STR Test for Use in Forensic Casework (I) *Int. J. Leg. Med.*, **109:**186-194.
4. Sparkes R., Kimpton C., Gilbard S., Carne P., Andersen J., Oldroyd N., Thomas D., Urquhart A., Gill P. (1996) Validation of a 7-locus Multiplex STR Test for Use in Forensic Casework (II). *Int. J. Leg. Med.*, **109:**195-204.
5. Sullivan K.M., Manucci A., Kimpton C.P. and Gill P. (1993) A Rapid and Quantitative DNA Sex Test: Fluorescence - Based PCR Analysis of X-Y Homologous Gene Amelogenin. *BioTechniques,* **Vol 15** No 4. 636-640.
6. PowerPlex™: a product of Promega: Promega Corporation, 2800 Woods Hollow Road, Madison, WI 53711-5399, USA.
7. Profiler: AmpF*l*STR™ Profiler Plus™: a Product of PE Applied Biosystems, A Division of Perkin Elmer, 850 Lincoln Centre Drive, Foster City, CA 94404 USA
8. Hopwood A., Brookes J., Shariff A., Cage P., Tatum E., Mirza R., Crook M., Brews K., Sullivan K. (1997) A Fully integrated Robotic System for High Sample Throughput Within a DNA Databasing Unit. *Proceedings: LabAutomation'98*, January 17-21 1998. San Diego, CA, USA
9. Tibbetts C. Presentation to Workshop at 1998 AAFS Meeting San Francisco; February 9th -14th.
10. Werrett D. J., Pinchin R., Hale R. (1998) Problem Solving: DNA Data Acquisition and Analysis. *Profiles In DNA.* **Volume 2** No. 1 May 1998