Can the Validation Process in Forensic DNA Typing Be Standardized?

John M. Butler[1], Christine S. Tomsey[2], Margaret C. Kline[1]

[1]National Institute of Standards and Technology, Biotechnology Division, Gaithersburg, MD 20899-8311
[2]Pennsylvania State Police DNA Laboratory, Greensburg, PA 15601

Abstract

Validation of procedures used in forensic DNA typing is essential to ensure that reliable results can be obtained with a particular method and associated materials. Laboratories spend a significant amount of time performing validation of new DNA typing kits, software, and analytical instrumentation. As technologies are constantly evolving and new kits becoming available, the need to validate and implement new procedures is on-going. The wide variety of approaches and opinions that exist on the topic of validation make it challenging for a laboratory to deduce a minimum set of criteria for evaluation to ensure that a method can be relied upon to produce quality information. Through conducting a questionnaire of DNA caseworking and databasing laboratories, carefully surveying the literature, and interviewing representatives from a small lab, a large lab, and a contract forensic lab, we are attempting to establish a standardized model for validation that meets the needs of the forensic DNA typing community. The survey summaries found that a wide range of responses exist throughout the community making it difficult to define specific recommendations for minimum sample numbers given particular scenarios that would likely be adopted. A new website has been established as part of STRBase to collate the community's past, present, and future validation studies in order to provide a dynamic resource with helpful information (see http://www.cstl.nist.gov/biotech/strbase/validation.htm).

Introduction

Validation has been and always will be a tremendously important aspect of adopting and implementing new technologies in forensic DNA typing. With a growing availability and constant evolution of new methodologies in DNA testing, validation will continue to play an important role in forensic laboratories. Demonstration that an adopted technique is robust, reproducible, and reliable across a defined range of conditions is essential in an environment where results help put the guilty behind bars and set the innocent free. Error prone techniques cannot be tolerated when so much is at stake.

There are generally thought to be two types of validation. The manufacturer of a DNA test or a group of laboratories typically performs developmental validation due to its more extensive nature while an internal validation is more specific to the needs of a particular forensic laboratory (Butler 2001, FBI 2000). Both developmental and internal validations help to establish that the performance characteristics for an analytical procedure are adequate for the intended use.

While validation studies are often viewed as being tedious and frankly downright boring, the consequences of not performing adequate validation of a technique can be lost information and wasted time, resources, and money as a technique is implemented and utilized. However, laboratories can also spend excessive time performing validation studies and not get to the business of solving cases or running offender database samples in a timely fashion. A delicate balance exists between thorough investigation of a technique and rapid implementation. Of course, the risk of adopting a particular technology in a laboratory goes down as more and more forensic laboratories implement it and the community at large embraces the technique. We believe that a widely accepted set of criteria for validation including an established minimum number of samples to be analyzed as part of a validation study would be helpful to forensic laboratory analysts, their supervisors, the court system, and laboratory accreditation inspectors. Such a set of standardized validation studies involving specific numbers of recommended samples for various testing scenarios have not yet been defined, much less

accepted by the community, primarily out of a desire to avoid any kind of perceived rigidity that may be associated with inflexible minimum sample numbers.

A number of organizations and documents exist to aid in framing the validation standardization information discussed here. Within the United States, the DNA Advisory Board (DAB) Standards govern forensic DNA analysis (FBI 2000). Section 8 of the DAB standards speak specifically to the topic of validation but only in broad terms. Rather than providing specific recommended numbers for various studies, the focus of the DAB standards is on "appropriately documenting" developmental and internal validation along with material modifications.

Validation guidelines for quality assurance programs in DNA analysis were previously published in 1989, 1991, and 1995 by the FBI-sponsored Technical Working Group on DNA Analysis Methods (TWGDAM). In July 2004, the FBI's Scientific Working Group on DNA Analysis Methods (SWGDAM) published revised validation guidelines (SWGDAM 2004). These most recent SWGDAM validation guidelines supercede the 1995 TWGDAM guidelines and attempt to provide a little more detail to the general validation information described in the DAB standards that were issued in July 1998 and April 1999 (FBI 2000).

The SWGDAM Revised Validation Guidelines emphasize that a total of at least 50 samples should be run as part of internal validation (SWGDAM 2004). These 50 samples can come as part of studies examining some or all of the following: known and non-probative evidence samples (section 3.1), reproducibility and precision (section 3.2), match criteria (section 3.3), sensitivity and stochastic studies (section 3.4), mixture studies (section 3.5), contamination (section 3.6), and a qualifying test (section 3.7). Furthermore, these guidelines state that assessment of the limits of new assays or typing technologies should be examined with authentic case samples when possible.

It is probably worth noting that one of the authors of this report (JMB) was a member of the SWGDAM validation committee and has wrestled with these analytical validation issues for over a decade in developing a number of new methods for application in human identity testing laboratories including the pioneering of capillary electrophoresis for STR typing (Butler 1994). Another author (CST) is the DNA technical leader for a state forensic laboratory that has been directly involved in validating six different amplification kits across the ABI 310, ABI 377, ABI 3100, and FMBIO platforms (Tomsey 2001, Krenke 2002, Greenspoon 2004). She also has 10 years of experience as a trained DNA auditor against the TWGDAM and DAB standards and has organized in conjunction with the FBI the first formalized DNA training for auditors. The last author (MCK) has conducted numerous interlaboratory studies among forensic labs for the past decade (Kline 1997, Kline 2002, Kline 2003) and personally examined well over 20,000 DNA samples involving every forensic DNA typing technology developed since the early-1990s.

The purpose of this study is to better define the philosophy of validation as currently practiced in forensic DNA laboratories with a goal to outline recommended minimum sample numbers for testing in various validation scenarios. It is our hope that we can learn from the past as we move forward into the future.

Materials and Methods[1]

Information on current practices and attitudes towards validation was gathered from members of the forensic DNA typing community through multiple mechanisms.

[1] 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.

First, a questionnaire was prepared in June 2004 that contained 14 primary questions (the entire survey is reproduced at the end of this article). The distribution of this validation survey was through a general request at the National Institute of Justice DNA Grantees meeting held June 28-30, 2004 followed up by direct solicitation through email. The primary audience targeted included those who attended the NIJ Grantees meeting as well as participants in recent NIST interlaboratory studies or contacts through the STRBase website (see http://www.cstl.nist.gov/biotech/strbase). Thus, hundreds of DNA analysts and supervisors from across the United States and a few dozen from other parts of the world were given an opportunity to contribute to this voluntary survey.

By the end of August 2004, we received 52 responses to our validation questionnaire, which are summarized in the Results and Discussion section. These responses were from forensic scientists located in 27 different states and Puerto Rico along with 2 from outside the U.S. In addition, representatives from four different companies comprising some of the largest contract DNA testing laboratories in human identity testing expressed their opinions on validation. Each response was entered into a Microsoft Access database and Excel worksheets to enable comparison statistics to be performed across the survey participants.

Second, interviews were conducted with several forensic scientists to gain their perspective on the validation process and current practices. Dr. Robin Cotton, laboratory director of Orchid Cellmark (Germantown, MD), Dr. Timothy McMahon, validation coordinator for the Armed Forces DNA Identification Laboratory (Rockville, MD), and Karolyn Tontarski, DNA analyst at the Montgomery County Crime Laboratory (Rockville, MD) provided perspective on validation from a private lab, a large government lab, and a small local lab, respectively. Dr. Cotton's perspective from having presented numerous court testimonies regarding DNA typing and being a member of American Society of Crime Laboratory Directors/Laboratory Accreditation Board (ASCLD/LAB) was especially valuable.

Finally, a careful literature review was conducted after identifying articles published in the last 12 years regarding DNA typing with STRs using a variety of technologies. A total of 64 articles pertaining to validation of STR kits, in-house assays, software, and instruments were reviewed as noted in **Table 1**. Complete citations for these articles are found in the reference list. Validation summary sheets for many of these studies are now or will shortly become available on the newly created STRBase Validation Homepage: http://www.cstl.nist.gov/biotech/strbase/validation.htm.

Results and Discussion

Validation Survey Responses

The 52 survey responses received were all quite positive towards our efforts of gathering information in an attempt to standardize the validation process. Of course, this number does not reflect all of those provided with the opportunity who failed to respond in a timely manner either because of busy schedules, lack of interest in the topic, or no desire to support our efforts to standardize validation.

Overall 85% of the respondents felt that the process of validation could be standardized. Those who responded "no" to this question included comments such as "to some degree it can be, however, validation is specific to the platform, kits, …", "a start-up lab should do much more than an experienced lab…", "validation builds on previous work by lab or published data", "parts of it can be standardized; I don't think the non-probative cases could be", and "only in a general way, as with the SWGDAM guidelines. The uniqueness of each new procedure would make standardization difficult."

To the question, "*if a standard protocol or set of guidelines existed for validation, would you use it?*" there were 47 "yes" (90%), 1 "probably", 1 "maybe-if adopted by the community", and 3 "no" responses. Comments in conjunction with this question included "No-I would reference them. I may not completely abide by them but I would certainly review them", "No-but it would be taken into consideration", "Yes-we would have to or there would be problems in court", "Yes-as long as they

remain updated, relevant and feasible guidelines and do not become dogma", "Yes-if it would pass an audit for validation", and "Yes-unless they were far less stringent than current practice."

When asked "*if a standard set of samples existed for performing validation testing, would you use them?*" the responses included "Yes-would love to have something like that available; we are always eager to have benchmarks for assessment", "Yes-these types of samples would cut down on time for validation. It would be efficient if they were ready for the particular type of validation…", "Yes-as long as they are readily available at a reasonable price", "Yes-now THAT I like", "Yes-under the condition that it holds at least our laboratory's internal guidelines for minimum number of samples, etc.", "Yes-for some studies", and "No-this approach is not recommended. It is most important that systems work with the materials available in individual laboratories. Laboratories should be allowed, even encouraged, to select their own preferred materials. Choices for such selection of standard materials for within laboratory analyses and cross-laboratory comparison already exist from a variety of government and commercial entities." Overall, 90% of survey respondents indicated an interest in a standard set of samples for performing validation testing and would use them if available. However, no specific recommendations were given as to the type of samples that should be made available.

Different sources cited for current guidance in terms of performing validation studies included the DAB standards, SWGDAM, ASCLD/LAB, ISO 17025, Promega's validation guide (see Promega 2001), publications in the literature, information shared via meeting presentations, NIST studies and publications, FBI studies and publications, previous scientific training, previous validations in the lab, in-house statisticians, consultation with the manufacturer and kit user's manual, common sense, and one-on-one conversations with other forensic scientists and DNA technical leaders.

Responses to the request for the *number of total samples recommended for internally validating a new forensic kit* ranged from 5 to 500. The most frequent number seen in the responses received was 50 samples, which matches the newly released SWGDAM Revised Validation Guidelines recommendation of at least 50 samples for internal validation (SWGDAM 2004).

We requested as part of our validation questionnaire *specific numbers or ranges of numbers for commonly performed validation studies*. These studies include precision, sensitivity, mixture ratios, non-human DNA, non-probative cases, heterozygote peak height ratios, and stutter percentage studies. As can be imagined, there was a wide variety of responses to requests for specific numbers thought to be appropriate or sufficient to demonstrate validity of a particular procedure.

Precision numbers ranged from as few as 5 injections of a single sample to more than 100 separate DNA samples. Typically allelic ladders or positive controls, such as the cell line 9947A, were recommended for use in precision studies. The recommended numbers for sensitivity studies also varied in the suggested numbers of samples to be tested as well as the maximum and minimum amounts that should be evaluated. Most responses, however, involved less than 10 samples with dilution series in the range of 10 ng down to 30 pg. Suggested mixture studies contained a few responses with extreme ratios, such as 1:1000, 1:200, or 1:100, but most were in the reasonable range for detection of 1:20 to 20:1. Some recommended numbers of samples included 5 different 2-person mixtures, 50 amplifications from at least 10 different mixtures, and 1 set of samples ranging from 1:10 to 10:1. For the non-human samples, responses included numbers such as "10-20 food animals, companion animals, local wildlife, and ferrets." However, many respondents said that they did not intend to perform animal studies as part of internal validation because the manufacturer had already examined these types of samples during developmental validation. Most of the non-probative case responses were between 5 and 10 cases with the full range spanning 3 to 25. One responder commented, "complete cases are not required to test a system." This individual went on to recommend that a lab "run at least 8 mock non-probative samples" and then noted "non-probative cases are not guaranteed to provide complete profiles. They are needed only to show that false results are not generated. Lack of results or incomplete results does not affect the validity of a validation." Survey results for the recommended numbers of samples to determine heterozygote peak height ratio imbalance ranged from 0 to 400 samples with a median of 50. The number of samples recommended for stutter percentage values ranged from 5 to 400 with a median of 63.

As noted by Vincent and Bessetti (2003) instruments are not static over time and STR kit reagents can vary slightly from one lot to the next. Thus testing a range of input DNA quantity is helpful to demonstrate that the technique is robust over the range of interest. Many labs perform quality control of a new kit by running an optimal amount of DNA template (e.g., 1 ng) whereas a dilution series of 5 concentrations from a single sample across the expected range of results can be more informative.

In response to the question, "*how do you know when you are finished validating a kit, instrument, software, or procedure?*" answers included comments such as "when you have demonstrated that it works as expected over a range of samples that is representative of what is seen in casework", "when repeat performance gave the same result", "when you pull the toothpick out and it is dry?... Meet at least minimum expectations and DAB guidelines", "you are very comfortable that you know how it works and your documentation will convince a reviewer you have put the kit thru a rigorous review/test", "Once a reasonable body of data has been assembled and analyzed, quirks have been revealed, and the upper and lower limits of the system have been challenged using a range of samples that one could expect to encounter in the everyday operation of the system", "when you achieve accuracy and precision to the desired statistical level of certainty", and "validation is never complete".

Respondents stated that they had *plans to validate in the next year the following kits, software, or instruments*: DNA IQ and Qiagen extraction methods on robotic platforms, the Quantifiler kit for DNA quantitation on the ABI 7000, PowerPlex 16, Identifiler, PowerPlex Y, Yfiler, Profiler Plus/COfiler reduced volume amplifications, GeneMapper*ID*, GeneScan/Genotyper for Windows, TrueAllele as an expert system, SQL*LIMS and Forensic Solution for sample tracking, the ABI 3100-*Avant*, the ABI 3100, the MegaBACE, and the FMBIO III+. We used this information in guiding the examples and scenarios provided in the recommendations section below.

Interviews

The various challenges faced within different environments including small forensic laboratories (e.g., <5 analysts), large forensic laboratories (e.g., >50 analysts), and private contract laboratories were explored through personal interviews with key personnel to gain their perspectives on validation studies. The laboratories interviewed all have extensive experience with a range of forensic DNA testing methodologies and are well-respected in the human identity testing community. All three laboratory representatives interviewed agreed that their names could be acknowledged as long as specific comments were not attributed to them. Thus, the comments listed here are done so in an anonymous fashion.

One of the primary challenges from the perspective of each of the laboratories, and particularly small laboratories, is having time to validate new equipment or technologies. In some cases, equipment that could increase laboratory capacity is left sitting in boxes due to the perceived potential energy barrier of performing extensive validation studies. During the initial effort to bring STR testing on-line, one lab interviewed was shutdown to casework acceptance and all samples were outsourced to a contract laboratory during the time it took to get validated. Laboratories do not always have this luxury and thus must carefully plan and execute their validation around other on-going projects and casework loads.

In one laboratory, validation of PowerPlex 16 took about 8 months overall, but with probably only 4 months of actual lab time. For this lab, approximately 2 months were taken just to purchase the necessary supplies and the remainder of the time involved writing up the validation results and standard operating procedures. This laboratory found the Validation of STR Systems Reference Manual (Promega Corporation 2001) prepared by Promega Corporation in 2001 to be a helpful resource in deciding on and designing experiments for the validation studies performed.

In another one of the interviewed labs, validation of PowerPlex 16 took around 4 months with most of the time in the initial startup and results write-up. A full month was taken to create the tables of data

and analyze them. One interviewee made a comment that the time required to perform validation of a new amplification system, such as Identifiler, is generally around 3-4 months but that it could be compressed to a few weeks with careful planning.

All of those interviewed felt that providing actual numbers or a spread of possible numbers for minimum sample testing would be helpful in any efforts to standardize the validation process. More specifics were felt to be helpful along with the possibility of resources to aid future validations including summaries of previous validation work. Teaching the philosophy around validation was also stressed with particular emphasis on the fact that not every study needs to be performed for every validation of an analytical method. Before a set of validation experiments is performed, the question should be asked "Do we already know the answer to this question from the literature or a previous study performed in-house?" If the answer is "yes" and we document how we know this answer, then there is no need to perform that set of validation experiments. A good example of this scenario is non-human DNA studies. If a manufacturer or another lab has demonstrated and documented performance of a particular STR kit with a variety of animal and bacterial samples, then those same studies do not need to be performed in your own laboratory.

In two of the interviewees' opinions, the comfort level reached when a process or piece of equipment works properly in their hands is the most important part of validation and helps the scientist performing the studies to know that enough samples have been examined. One interviewee admitted that there is no logic behind the number of samples tested in currently performed validation studies and that very few samples are actually needed when there is no reason to expect that two people's DNA behave differently (other than the natural variation in size of the actual observed alleles).

When asked how you know when you have enough samples for validation, one reply was "you can never know…but it is always nice to have more samples." Another interviewee stated that they had seen a range of studies from gross overkill to what is barely enough because everyone has a different opinion on what is sufficient. The issue of the time it takes to validate methods and instruments and to train new personnel is very important. This is both an expense issue and a philosophy issue. The philosophy of one laboratory interviewed is to perform only a minimum amount of forensic sample testing during validation because DNA will likely behave about the same in a case sample even though there are contaminants. Since every case is different and one cannot artificially mimic casework with exactness, there is no point in trying to test every possible scenario.

To the inquiry of what would be done differently if your validation studies could be done again, the focus was on spending more time with relative fluorescence unit (RFU) cut-off values and carefully defining the assay limit of detection. The number of samples run in precision studies would have been drastically reduced was another response.

More than one of those interviewed stressed the importance of learning how to think through the validation process before going into the laboratory to conduct experiments. Validation efforts should avoid addressing a question that you already know the answer to. Unfortunately, many decision making people do not know how to design effective experiments. It is also important to remove variables, such as pipetting, as much as possible from experiments. For example, a large sample batch could be prepared and then aliquoted to multiple tubes rather than preparing each tube individually. We often do not assess very well where the variation is coming from in a set of experiments.

Regarding mixture studies, it was noted that a minor component is rarely detected at 1:10 and never below 1:20 so there is no reason to test beyond these ranges (e.g., 1:100). Mixture dilutions are typically performed with only a single mix of two different individuals rather than many possibilities. It was also stressed that it is essential to know how much stuff you have in setting up a mixture experiment. Mixing blood samples in particular ratios is not reliable and mixture studies should always begin with quantified amounts of DNA. A laboratory may do lots of mixtures in the beginning as part of the learning process but with experience only a few samples are needed to see how a technology

responds to mixture analysis. Mixture ratio studies should be done at the end of validation after defining the target input DNA concentration.

The interviewees all felt that sensitivity is among the most important studies to perform in validation with forensic DNA testing. Parameters that can be varied such as ABI 310 injection time, PCR reaction volume, and amplicon amount in the ABI 310 analysis tube should all be examined. Typically with sensitivity experiments only 1 or 2 different individuals are examined over the desired concentration range (e.g., 10 ng down to 50 pg). One lab ran 8 samples in their dilution series while another ran only 5. Precision studies are usually not necessary once an instrument has been validated and the physical environment has not changed.

Those interviewed felt that putting together detailed examples and scenarios for validation with an effort to gain consensus from the community would be helpful. However, it was noted that forensic laboratories are often slow to change and can be resistant to efforts of standardization. Several of the interviewees emphasized that validation results should be well organized to aid inspections and commented that usually validation efforts are not questioned if they are well laid out and clearly documented.

Regarding the possibility of having standard samples to use for validation studies, one interviewee felt that initial tubes with pre-made mixtures and sensitivity titrations would be helpful but that non-probative samples must come from the lab itself.

Literature Summary

The peer-reviewed literature contains a number of publications on the topic of validation as it pertains to specific work performed with STR kits or other procedures in human identity testing. We have tried to create a fairly comprehensive compilation of these studies in the reference list at the end of this document. The 64 references examined as part of this study are summarized in Table 1 according to the kit, assay, instrument, or software validated. More details regarding these studies may be found on the new NIST STRBase Validation Homepage: http://www.cstl.nist.gov/biotech/strbase/validation.htm.

To illustrate the significance of validation over the years in terms of what is discussed at forensic DNA meetings, we reviewed the titles for the more than 1,200 talks and posters presented at the various International Symposiums on Human Identification sponsored by the Promega Corporation over the past 15 years. We found that greater than 10% of the talks and posters contained the term “validation” in the title (**Table 2**). A more detailed examination of abstracts or presentation content would likely increase the percentage of validation presentations particularly if discussions of allele frequency information as part of “validating” population databases were added to this breakdown. The number of validation presentations seems to ebb and flow by year probably based largely on timing for adoption of new technologies and kits by various forensic laboratories. It is interesting to note that the highest percentage of validation presentations was last year in 2003.

In carefully considering the needs of the human identity testing community in terms of standardizing the validation effort, it is helpful to look to other fields that may have faced similar challenges. The pharmaceutical industry is concerned with validation of analytical methods used during drug development and product testing. The Food and Drug Administration (FDA) has adopted several documents regarding validation that were produced a few years ago by the International Council on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). These ICH documents are intended as guidelines for industry validation of analytical procedures:

> Q2A Guideline for Industry: Text on Validation of Analytical Procedures
> http://www.fda.gov/cder/guidance/ichq2a.pdf
>
> Q2B Guideline for Industry: Validation of Analytical Procedures: Methodology
> http://www.fda.gov/cder/guidance/1320fnl.pdf

There have also been some insightful publications on the topics of analytical method validation and quality assurance of chemical measurements published in the journal *Analytical Chemistry* over the past several decades (Taylor 1981, Taylor 1983, Green 1996).

It is important to note that interlaboratory studies play an important role in verifying that a methodology works well across multiple laboratories. The European DNA Profiling Group (EDNAP) has run a number of constructive interlaboratory studies over the years (see Table 14.1 in Butler 2001). The use of positive control DNA samples, standard cell lines, and certified or standard reference materials also brings a degree of confidence to one's results when performing validation studies (Szibor 2003).

Examining the Steps Involved in Validation

Since the terms "validate" and "validation" often mean different things to various people it is worth exploring briefly the steps involved in bringing a process "on-line" in a forensic DNA laboratory and to point out that pre-validation learning and post-validation training are not validation of the technique but rather involve a human variable and "comfort level" that is more challenging to standardize.

The steps involved in bringing a procedure (assay, instrument, or software) "on-line" in a forensic lab setting typically include (1) installation of the instrumentation or software and purchase of assay reagents, (2) learning about the technique and how to perform it properly, (3) validation of the analytical procedure to define its range and reliability, (4) creation of the standard operating procedures with interpretation guidelines based on the validation studies, (5) training of other personnel on the technique, and (6) each trained analysts passing a qualification test for initial use in forensic casework. After a procedure has been successfully been implemented into use with forensic casework, proficiency tests are performed on a regular basis (usually twice a year) to demonstrate successful application of the technique over time by qualified analysts. In addition, new materials and instruments need to be evaluated over time through a quality control process involving a performance check on the validated procedure.

Defining and Recommending a Minimum Sample Number

As noted previously, SWGDAM recently released a set of revised validation guidelines through publication in the FBI's on-line journal *Forensic Science Communications* (see http://www.fbi.gov/hq/lab/fsc/backissu/july2004/standards/2004_03_standards02.htm). While these guidelines provide further information supplemental to the DAB standards (FBI 2000), they still lack specific recommendations in terms of the minimum number of samples to be tested that would be helpful to the community. The only concrete figure provided in the SWGDAM guidelines is that a minimum of at least 50 samples should be run for internal validation purposes by a forensic laboratory.

While among our original goals for this project was the definition of a minimum sample number for various studies performed in forensic laboratories, we have come to the conclusion that everyone will always have a different comfort level and inflexible, absolute numbers for defined studies will not likely be widely accepted. Instead, we would like to note that each study does not require an excessive number of samples to be run and that not all studies are essential for every procedure under consideration for validation. **Figure 1** illustrates that for replicate experiments there is little gained in terms of the confidence around a set of results after 5-10 measurements. This figure illustrates why the selection of 5 replicate experiments is often selected as a minimum sample number since the Student *t* value is 2.78 and can only go down to 1.96 with an infinite number of replicates to reflect the true variation that exists within a 95% confidence interval.

It is also worth noting that there is no time requirement for how long a validation study should take and thus the "at least 50 samples" recommended by SWGDAM could be run in a matter of days rather than weeks or months. As noted by the SWGDAM Revised Validation Guidelines (SWGDAM 2004)

not all studies may be necessary due to the method involved. Experimental design can also be implemented in many studies to aid in examining the variables under investigation. Some suggested approaches involving experimental design are available at http://www.haag.com/Assorted File Folder/How Many Data Handout.pdf.

New Validation Homepage on STRBase

The NIST STRBase website has been widely used by the forensic DNA typing community since it was introduced in 1997 (Ruiberg 2001). We have established a new validation homepage at http://www.cstl.nist.gov/biotech/strbase/validation.htm that can serve as a repository of helpful information on kit, assay, software, and instrument validation studies performed in forensic DNA laboratories. This site also contains links to the DAB standards and SWGDAM revised validation guidelines. Publications in the literature involving validation of forensic DNA tests are also listed and summarized. Validation Summary Sheets make up the core of this new validation section of STRBase and contain a simple summary of the studies conducted along with the numbers and types of samples examined as part of the analytical validation for a particular assay or procedure. **Table 3** includes an example Validation Summary Sheet that details 17 studies conducted as part of the developmental validation of the new Y-STR kit PowerPlex Y (Krenke 2004).

Besides the literature review to capture past published information, we would like to encourage the posting of information from individual forensic laboratory internal validation studies so that the community can quickly gain a sense of the numbers of samples and types of samples run in other labs. Forensic science journals rarely accept internal validation studies any more due to the fact that they are not novel. We feel that this validation summary format will easily enable labs to document their work as well as others and thus establish new techniques and technologies on firmer ground. **Table 4** contains an example of this type of internal validation information from the Pennsylvania State Police efforts in validating the PowerPlex 16 kit on the ABI 310 Genetic Analyzer a few years ago.

Conclusions

An important way in which validation can be standardized is through the documentation process. The "Validation of STR Systems Reference Manual" published by Promega Corporation in February 2001 is a good example of how validation studies can be summarized with a statement of purpose and introduction to each study along with what experiments were performed, what results were observed, and what conclusions could be drawn. This format makes it easy for an inspector or another analyst to quickly get a sense of the work performed to understand aspects of the DNA test under investigation.

Minimum sample recommendations can be made for some processes but not easily for all. Nor do we believe that proposed minimum sample numbers would likely be widely accepted due to current differences of opinion among forensic DNA scientists. Perhaps as more validation summaries and surveys are completed, there will come to be greater unity in approaches for conducting validation studies throughout the human identity testing community.

A new validation homepage has been setup on the NIST STRBase website to aid current and future work with validation studies (see http://www.cstl.nist.gov/biotech/strbase/validation.htm). We plan to include some standard formats for writing up validation results in the future. Helpful information, such as summaries of species specificity studies, will be added by ourselves or contributors from the community as this Validation Homepage grows in usefulness.

Our hope is that by approaching the validation process in a more unified and standardized manner validation studies may be performed more quickly and with greater confidence so that forensic laboratories may get on to the important business of solving crimes with technologies and procedures that have been verified to be robust and reliable.

Acknowledgments

This project was made possible by funding from the National Institute of Justice through an interagency agreement with the NIST Office of Law Enforcement Standards. The support of many members of the forensic DNA typing community who took time out of their busy schedules to respond to the questions in our validation survey in a meaning manner is greatly appreciated. The helpful perspectives of Robin Cotton (Orchid Cellmark, Germantown, MD), Kari Tontarski (Montgomery Country Crime Laboratory, Rockville, MD), and Tim McMahon (AFDIL, Rockville, MD) are gratefully acknowledged as well. Discussions with David Duewer (Analytical Chemistry Division, NIST) were invaluable as always.

References

Andersen JF, Greenhalgh MJ, Butler HR, Kilpatrick SR, Piercy RC, Way KA, Myhill HS, Wright JC, Hallett R, Parkin BH. Further validation of a multiplex STR system for use in routine forensic identity testing. *Forensic Sci Int* 1996; 78:47-64.

Applied Biosystems. *Quantifiler Human DNA Quantification Kit User's Manual*. Applied Biosystems, Foster City, California, 2003.

Budowle B, Moretti TR, Keys KM, Koons BW, Smerick JB. Validation studies of the CTT STR multiplex system. *J Forensic Sci* 1997; 42(4):701-707.

Buse EL, Putinier JC, Hong MM, Yap AE, Hartmann JM. Performance evaluation of two multiplexes used in fluorescent short tandem repeat DNA analysis. *J Forensic Sci* 2003; 48(2):348-357.

Butler JM, McCord BR, Jung JM, Allen RO. Rapid separation of the short tandem repeat HUMTH01 by capillary electrophoresis. Biotechniques 1994; 17:1062-1070.

Butler JM. *Forensic DNA Typing: Biology and Technology behind STR Markers*. London: Academic Press, 2001. Chapter 14 "Laboratory Validation" pp. 205-222.

Christian GD. *Analytical Chemistry* (6th Edition). New York: John Wiley & Sons, 2004.

Clayton TM, Whitaker JP, Fisher DL, Lee DA, Holland MM, Weedn VW, Maguire CN, DiZinno JA, Kimpton CP, Gill P. Further validation of a quadruplex STR DNA typing system: a collaborative effort to identify victims of a mass disaster. *Forensic Sci Int* 1995; 76:17-25

Collins PJ, Hennessy LK, Leibelt CS, Roby RK, Reeder DJ, Foxall PA (2004) Developmental validation of a single-tube amplification of the 13 CODIS STR loci, D2S1338, D19S433, and amelogenin: the AmpFISTR Identifiler PCR amplification kit. *J. Forensic Sci.* 49(6), *in press*. (on-line at JFS website Sept 2004)

Coticone SR, Oldroyd N, Philips H, Foxall P. Development of the AmpFISTR SEfiler PCR amplification kit: a new multiplex containing the highly discriminating ACTBP2 (SE33) locus. *Int J Legal Med* 2004; 118(4):224-234.

Cotton EA, Allsop RF, Guest JL, Frazier RR, Koumi P, Callow IP, Seager A, Sparkes RL. Validation of the AMPFISTR((R)) SGM plus system for use in forensic casework. Forensic Sci Int 2000; 112(2-3):151-161.

Crouse C, Schumm JW. Investigation of species specificity using nine PCR-based human STR systems. *J Forensic Sci* 1995; 40(6):952-956.

Crouse CA. Implementation of forensic DNA analysis on casework evidence at the Palm Beach County Sheriff's Office Crime Laboratory: historical perspective. Croat Med J 2001;42(3):247-251.

Daniels DL, Hall AM, Ballantyne J. SWGDAM developmental validation of a 19-locus Y-STR system for forensic casework. *J Forensic Sci* 2004; 49(4):668-683.

Federal Bureau of Investigation. Quality assurance standards for forensic DNA testing laboratories, *Forensic Science Communications* [Online]. (July 2000A). Available : http://www.fbi.gov/hq/lab/fsc/backissu/july2000/codispre.htm.

Federal Bureau of Investigation. Quality assurance standards for convicted offender DNA databasing laboratories, *Forensic Science Communications* [Online]. (July 2000B). Available: http://www.fbi.gov/hq/lab/fsc/backissu/july2000/codispre.htm.

Food and Drug Administration. General principles of software validation; final guidance for industry and FDA staff. Document issued January 11, 2002. Available: http://www.fda.gov/cdrh/comp/guidance/938.pdf.

Fox JC, Cave CA, Schumm JW. Development, characterization, and validation of a sensitive primate-specific quantification assay for forensic analysis. *Biotechniques* 2003; 34(2):314-8, 320, 322.

Frank WE, Llewellyn BE, Fish PA, Riech AK, Marcacci TL, Gandor DW, Parker D, Carter RR, Thibault SM. Validation of the AmpFlSTR Profiler Plus PCR amplification kit for use in forensic casework. *J Forensic Sci* 2001; 46(3):642-646.

Frazier RR, Millican ES, Watson SK, Oldroyd NJ, Sparkes RL, Taylor KM, Panchal S, Bark L, Kimpton CP, Gill PD. Validation of the Applied Biosystems Prism™ 377 automated sequencer for forensic short tandem repeat analysis. Electrophoresis 1996; 17(10):1550-1552.

Fregeau CJ, Bowen KL, Fourney RM. Validation of highly polymorphic fluorescent multiplex short tandem repeat systems using two generations of DNA sequencers. *J Forensic Sci* 1999; 44(1):133-166.

Fregeau CJ, Bowen KL, Leclair B, Trudel I, Bishop L, Fourney RM. AmpFlSTR Profiler Plus short tandem repeat DNA analysis of casework samples, mixture samples, and nonhuman DNA samples amplified under reduced PCR volume conditions (25 microL). *J Forensic Sci* 2003; 48(5):1014-1034.

Gaines ML, Wojtkiewicz PW, Valentine JA, Brown CL. Reduced volume PCR amplification reactions using the AmpFlSTR Profiler Plus kit. *J Forensic Sci* 2002; 47(6):1224-1237.

Gill P, Koumi P, Allen H. Sizing short tandem repeat alleles in capillary array gel electrophoresis instruments. *Electrophoresis* 2001;22:2670-2678.

Green JM. A practical guide to analytical method validation. *Anal Chem* 1996; 68:305A-309A.

Greenspoon SA, Lytle PJ, Turek SA, Rolands JM, Scarpetta MA, Carr CD. Validation of the PowerPlex 1.1 loci for use in human identification. *J Forensic Sci* 2000; 45(3):677-683.

Greenspoon SA, Ban JD, Pablo L, Crouse CA, Kist FG, Tomsey CS, Glessner AL, Mihalacki LR, Long TM, Heidebrecht BJ, Braunstein CA, Freeman DA, Soberalski C, Nathan B, Amin AS, Douglas EK, Schumm JW. Validation and implementation of the PowerPlex 16 BIO System STR multiplex for forensic casework. *J Forensic Sci* 2004; 49(1):71-80.

Greenspoon SA, Ban JD, Sykes K, Ballard EJ, Edler SS, Baisden M, Covington BL. Application of the BioMek 2000 Laboratory Automation Workstation and the DNA IQ System to the extraction of forensic casework samples. *J Forensic Sci* 2004; 49(1):29-39.

Holland MM, Parsons TJ. Mitochondrial DNA sequence analysis - validation and use for forensic casework. *Forensic Sci Rev* 1999; 11(1):22-50.

Holt CL, Buoncristiani M, Wallin JM, Nguyen T, Lazaruk KD, Walsh PS. TWGDAM validation of AmpFlSTR PCR amplification kits for forensic DNA casework. *J Forensic Sci* 2002; 47(1):66-96.

Isenberg AR, Allen RO, Keys KM, Smerick JB, Budowle B, McCord BR. Analysis of two multiplexed short tandem repeat systems using capillary electrophoresis with multiwavelength florescence detection. *Electrophoresis* 1998;19:94-100.

Johnson CL, Warren JH, Giles RC, Staub RW. Validation and uses of a Y-chromosome STR 10-plex for forensic and paternity laboratories. *J Forensic Sci* 2003; 48(6):1260-1268.

Junge A, Madea B. Validation studies and characterization of variant alleles at the short tandem repeat locus D12S391. *Int J Legal Med* 1999; 112(1):67-69.

Junge A, Lederer T, Braunschweiger G, Madea B. Validation of the multiplex kit genRESMPX-2 for forensic casework analysis. *Int J Legal Med* 2003; 117(6):317-325.

Kadash K, Kozlowski BE, Biega LA, Duceman BW. Validation study of the TrueAllele automated data review system. *J Forensic Sci* 2004; 49(4):660-667.

Kimpton CP, Fisher D, Watson S, Adams M, Urquhart A, Lygo J, Gill P. Evaluation of an automated DNA profiling system employing multiplex amplification of four tetrameric STR loci. *Int J Leg Med* 1994; 106:302-311.

Kimpton CP, Oldroyd NJ, Watson SK, Frazier RR, Johnson PE, Millican ES, Urquhart A, Sparkes BL, Gill P. Validation of highly discriminating multiplex short tandem repeat amplification systems for individual identification. *Electrophoresis* 1996; 17(8):1283-1293.

Kline MC, Duewer DL, Newall P, Redman JW, Reeder DJ, Richard M. Interlaboratory evaluation of short tandem repeat triplex CTT*. *J Forensic Sci* 1997; 42(5):897-906.

Kline MC, Duewer DL, Redman JW, Butler JM, Boyer DA. Polymerase chain reaction amplification of DNA from aged blood stains: quantitative evaluation of the "suitability for purpose" of four filter papers as archival media. *Anal Chem* 2002; 74(8):1863-1869.

Kline MC, Duewer DL, Redman JW, Butler JM. NIST Mixed Stain Study 3: DNA quantitation accuracy and its influence on short tandem repeat multiplex signal intensity. *Anal Chem* 2003; 75(10):2463-2469.

Koumi P, Green HE, Hartley S, Jordan D, Lahec S, Livett RJ, Tsang KW, Ward DM. Evaluation and validation of the ABI 3700, ABI 3100, and the MegaBACE 1000 capillary array electrophoresis instruments for use with short tandem repeat microsatellite typing in a forensic environment. *Electrophoresis* 2004; 25(14):2227-2241.

Krenke BE, Tereba A, Anderson SJ, Buel E, Culhane S, Finis CJ, Tomsey CS, Zachetti JM, Masibay A, Rabbach DR, Amiott EA, Sprecher CJ. Validation of a 16-locus fluorescent multiplex system. *J Forensic Sci* 2002; 47(4):773-785.

Krenke BE, Viculis L, Richard ML, Prinz M, Milne SC, Ladd C, Gross AM, Gornall T, Frappier JRH, Eisenberg AJ, Barna C, Aranda XG, Adamowicz, Budowle B. Validation of a male-specific, 12-locus fluorescent short tandem repeat (STR) multiplex. *Forensic Sci Int* 2004; *in press* (on-line at FSI website, Sept 2004)

LaFountain MJ, Schwartz MB, Svete PA, Walkinshaw MA, Buel E. TWGDAM validation of the AmpFlSTR Profiler Plus and AmpFlSTR COfiler STR multiplex systems using capillary electrophoresis. *J Forensic Sci* 2001; 46(5):1191-1198.

LaFountain M, Schwartz M, Cormier J, Buel E. Validation of capillary electrophoresis for analysis of the X-Y homologous amelogenin gene. *J Forensic Sci* 1998; 43(6):1188-1194.

Lazaruk K, Walsh PS, Oaks F, Gilbert D, Rosenblum BB, Menchen S, Scheibler D, Wenz HM, Holt C, Wallin J. Genotyping of forensic short tandem repeat (STR) systems based on sizing precision in a capillary electrophoresis instrument. *Electrophoresis* 1998;19(1):86-93.

Leibelt C, Budowle B, Collins P, Daoudi Y, Moretti T, Nunn G, Reeder D, Roby R. Identification of a D8S1179 primer binding site mutation and the validation of a primer designed to recover null alleles. *Forensic Sci Int* 2003; 133(3):220-227.

Levedakou EN, Freeman DA, Budzynski MJ, Early BE, Damaso RC, Pollard AM, Townley AJ, Gombos JL, Lewis JL, Kist FG, Hockensmith ME, Terwilliger ML, Amiott E, McElfresh KC, Schumm JW, Ulery SR, Konotop F, Sessa TL, Sailus JS, Crouse CA, Tomsey CS, Ban JD, Nelson MS. Characterization and validation studies of powerPlex 2.1, a nine-locus short tandem repeat (STR) multiplex system and penta D monoplex. *J Forensic Sci* 2002; 47(4):757-772.

Lygo JE, Johnson PE, Holdaway DJ, Woodroffe S, Whitaker JP, Clayton TM, Kimpton CP, Gill P. The validation of short tandem repeat (STR) loci for use in forensic casework. *Int J Legal Med* 1994; 107(2):77-89.

Mandrekar MN, Erickson AM, Kopp K, Krenke BE, Mandrekar PV, Nelson R, Peterson K, Shultz J, Tereba A, Westphal N. Development of a human DNA quantitation system. *Croat Med J* 2001; 42(3):336-339.

Micka KA, Sprecher CJ, Lins AM, Comey CT, Koons BW, Crouse C, Endean D, Pirelli K, Lee SB, Duda N, Ma M, Schumm JW. Validation of multiplex polymorphic STR amplification sets developed for personal identification applications. J Forensic Sci 1996; 41(4):582-590.

Micka KA, Amiott EA, Hockenberry TL, Sprecher CJ, Lins AM, Rabbach DR, Taylor JA, Bacher JW, Glidewell DE, Gibson SD, Crouse CA, Schumm JW. TWGDAM validation of a nine-locus and a four-locus fluorescent STR multiplex system. J Forensic Sci 1999; 44(6):1243-1257.

Moretti TR, Baumstark AL, Defenbaugh DA, Keys KM, Brown AL, Budowle B. Validation of STR typing by capillary electrophoresis. *J Forensic Sci* 2001; 46(3):661-676.

Moretti TR, Baumstark AL, Defenbaugh DA, Keys KM, Smerick JB, Budowle B. Validation of short tandem repeats (STRs) for forensic usage: performance testing of fluorescent multiplex STR systems and analysis of authentic and simulated forensic samples. *J Forensic Sci* 2001; 46(3):647-660.

Morley JM, Bark JE, Evans CE, Perry JG, Hewitt CA, Tully G. Validation of mitochondrial DNA minisequencing for forensic casework. *Int J Legal Med* 1999; 112(4):241-248.

Mozer TJ. Requirements for complete validation of an STR product. *Profiles in DNA* 2001; 4(3):14-15.

Nelson MS, Levedakou EN, Matthews JR, Early BE, Freeman DA, Kuhn CA, Sprecher CJ, Amin AS, McElfresh KC, Schumm JW. Detection of a primer-binding site polymorphism for the STR locus D16S539 using the Powerplex 1.1 system and validation of a degenerate primer to correct for the polymorphism. *J Forensic Sci* 2002; 47(2):345-349.

Nicklas JA, Buel E. Development of an *Alu*-based, real-time PCR method for quantitation of human DNA in forensic samples. *J Forensic Sci* 2003;48(5):936-944.

Pawlowski R, Maciejewska A. Forensic validation of a multiplex containing nine STRs--population genetics in northern Poland. *Int J Legal Med* 2000; 114(1-2):45-49.

Pestoni C, Lareu MV, Rodriguez MS, Munoz I, Barros F, Carracedo A. The use of the STRs HUMTH01, HUMVWA31/A, HUMF13A1, HUMFES/FPS, HUMLPL in forensic application: validation studies and population data for Galicia (NW Spain). *Int J Legal Med* 1995; 107(6):283-290.

Potter T. Co-amplification of ENFSI-loci D3S1358, D8S1179 and D18S51: validation of new primer sequences and allelic distribution among 2874 individuals. *Forensic Sci Int* 2003; 138(1-3):104-110.

Prinz M, Ishii A, Coleman A, Baum HJ, Shaler RC. Validation and casework application of a Y chromosome specific STR multiplex. *Forensic Sci Int* 2001; 120(3):177-188.

Promega Corporation (2001) Validation of STR Systems Reference Manual.
http://www.promega.com/techserv/apps/hmnid/referenceinformation/powerplex/ValidationManual.pdf

Richard ML, Frappier RH, Newman JC. Developmental validation of a real-time quantitative PCR assay for automated quantification of human DNA. *J Forensic Sci* 2003; 48(5):1041-1046.

Ruitberg CM, Reeder DJ, Butler JM. STRBase: a short tandem repeat DNA database for the human identity testing community. *Nucleic Acids Res* 2001; 29(1):320-322.

Ryan JH, Barrus JK, Budowle B, Shannon CM, Thompson VW, Ward BE. The application of an automated allele concordance analysis system (CompareCalls) to ensure the accuracy of single-source STR DNA profiles. *J Forensic Sci* 2004; 49(3):492-499.

Sgueglia JB, Geiger S, Davis J. Precision studies using the ABI prism 3100 genetic analyzer for forensic DNA analysis. *Anal Bioanal Chem* 2003; 376(8):1247-1254.

Shewale JG, Nasir H, Schneida E, Gross AM, Budowle B, Sinha SK. Y-chromosome STR system, Y-PLEX 12, for forensic casework: development and validation. J Forensic Sci 2004;49(6), *in press* (on-line at JFS website, Sept 2004)

Sinha SK, Budowle B, Arcot SS, Richey SL, Chakrabor R, Jones MD, Wojtkiewicz PW, Schoenbauer DA, Gross AM, Sinha SK, Shewale JG. Development and validation of a multiplexed Y-chromosome STR genotyping system, Y-PLEX 6, for forensic casework. *J Forensic Sci* 2003; 48(1):93-103.

Sinha SK, Nasir H, Gross AM, Budowle B, Shewale JG. Development and validation of the Y-PLEX 5, a Y-chromosome STR genotyping system, for forensic casework. *J Forensic Sci* 2003; 48(5):985-1000.

Sparkes R, Kimpton CP, Watson S, Oldroyd NJ, Clayton TM, Barnett L, Arnold J, Thompson C, Hale R, Chapman J, Urquhart A, Gill P. The validation of a 7-locus multiplex STR test for use in forensic casework. (I). Mixtures, ageing, degradation and species studies. *Int J Legal Med* 1996; 109(4):186-194.

Sparkes R, Kimpton C, Gilbard S, Carne P, Andersen J, Oldroyd N, Thomas D, Urquhart A, Gill P. The validation of a 7-locus multiplex STR test for use in forensic casework. (II), Artefacts, casework studies and success rates. *Int J Legal Med* 1996; 109(4):195-204.

Szibor R, Edelmann J, Hering S, Plate I, Wittig H, Roewer L, Wiegand P, Cali F, Romano V, Michael M. Cell line DNA typing in forensic genetics--the necessity of reliable standards. *Forensic Sci Int* 2003; 138(1-3):37-43.

Taylor JK. Quality assurance of chemical measurements. Anal Chem 1981; 53(14):1588A-1596A.

Taylor JK. Validation of analytical methods. Anal Chem 1983; 55(6):600A-608A.

Tomsey CS, Kurtz M, Kist F, Hockensmith M, Call P. Comparison of PowerPlex 16, PowerPlex1.1/2.1, and ABI AmpFISTR Profiler Plus/COfiler for forensic use. *Croat Med J* 2001; 42(3):239-243.

van Oorschot RA, Gutowski SJ, Robinson SL, Hedley JA, Andrew IR. HUMTH01 validation studies: effect of substrate, environment, and mixtures. *J Forensic Sci* 1996; 41(1):142-145.

Vincent EB, Bessetti J. Validation questions and answers. *Profiles in DNA* 2003; 6(2):13-14.

Wallin JM, Buoncristiani MR, Lazaruk KD, Fildes N, Holt CL, Walsh PS. TWGDAM validation of the AmpFISTR Blue PCR amplification kit for forensic casework analysis. *J Forensic Sci* 1998; 43(4):854-870.

Wallin JM, Holt CL, Lazaruk KD, Nguyen TH, Walsh PS. Constructing universal multiplex PCR systems for comparative genotyping. *J Forensic Sci* 2002; 47(1):52-65.

Wiegand P, Budowle B, Rand S, Brinkmann B. Forensic validation of the STR systems SE 33 and TC 11. *Int J Legal Med* 1993; 105(6):315-320.

Wilson MR, DiZinno JA, Polanskey D, Replogle J, Budowle B. Validation of mitochondrial DNA sequencing for forensic casework analysis. *Int J Legal Med* 1995; 108(2):68-74.

Validation Standardization Questionnaire

Please return to John Butler (NIST): john.butler@nist.gov or 301-975-8505 (fax)

Purpose of questionnaire: We are embarking on an effort to define the minimum number of samples needed to reliably validate DNA typing procedures. As part of this effort, we are conducting a survey of standard practices currently used by practitioners in forensic DNA laboratories. Your honest responses to the following questions will help the entire community as we compile this information. Results will be summarized at the Promega meeting in October 2004 and made available on the NIST STRBase web site.

General Questions

What does the term validation mean to you? *(define in a single sentence if possible)*

How do you know when you are finished validating a kit, instrument, software, or procedure?

What steps are needed in internal validation and how many samples should be run at a minimum?

- Precision studies ___(indicate types of samples –i.e., ladders), # samples/run ____; # runs ___
- Sensitivity studies_____, what range? _____
- Mixture studies _____ what mixture ratios are needed? _____
- Non-human DNA studies _____
- Non-probative cases _____

How many total samples do you think it takes to internally "validate" a new forensic kit?

- 10
- 50
- 500
- Other: _____

How many samples are necessary to determine heterozygote ratios?_____ Stutter values?_____
How many different sets of samples are needed?____ Do they have to be run over a particular time period?____
How are validation, training, and proficiency testing related to one another?
Do you think that the process of validation can be standardized? **Y or N**

Specific Needs

Where do you look for guidance currently in terms of validation?_____
What are some kits, software, instruments that you are considering for validation in the next year?

Commercial Kit	**Software**	**Analysis Instrument**
Extraction: _____	Genotyper for Windows	ABI 310
DNA quant: _____	GeneMapper	ABI 3100 *Avant*
STR amp kit: _____	TrueAllele	ABI 3100
	Other: _____	ABI 7000
		Other: _____

Outcome of this Study

If a standard protocol or set of guidelines existed for validation, would you use it? **Y or N**
If a standard set of samples existed for performing validation testing, would you use them? **Y or N**
Other Comments:_____

Contact Information

Please provide name and email address for follow-up questions if needed (will not be included in final summary so as to keep all results anonymous). Name: _____ Email: _____
(Information for presentation collected from June-Sept 2004)

Table 1. References to various validation studies conducted using commercial STR kits, in-house assays, instrumentation, and software that were examined as part of this study.

Kit, Assay, or Instrument	Reference
Profiler Plus	Frank *et al.* (2001), LaFountain *et al.* (2001), Tomsey *et al.* (2001), Holt *et al.* (2002), Fregeau *et al.* (2003), Buse *et al.* (2003), Wallin *et al.* (2002), Pawlowski *et al.* (2000), Moretti *et al.* (2001)
COfiler	LaFountain *et al.* (2001), Tomsey *et al.* (2001), Moretti *et al.* (2001), Holt *et al.* (2002), Buse *et al.* (2003), Wallin *et al.* (2002)
SGM Plus	Cotton *et al.* (2000)
AmpFlSTR Blue	Wallin *et al.* (1998)
AmpFlSTR Green I	Holt *et al.* (2002)
Profiler	Holt *et al.* (2002)
Profiler Plus *ID*	Leibelt *et al.* (2003)
Identifiler	Collins *et al.* (2004)
SEfiler	Coticone *et al.* (2004)
PowerPlex 1.1	Micka *et al.* (1999), Tomsey *et al.* (2001), Greenspoon *et al.* (2001)
PowerPlex 1.1 + D16 primer	Nelson *et al.* (2002)
PowerPlex 2.1	Tomsey *et al.* (2001), Levedakou *et al.* (2002)
PowerPlex 16	Krenke *et al.* (2002), Tomsey *et al.* (2001)
PowerPlex 16 BIO	Greenspoon *et al.* (2004)
Y-PLEX 6	Sinha *et al.* (2003a)
Y-PLEX 5	Sinha *et al.* (2003b)
Y-PLEX 12	Shewale *et al.* (2004)
PowerPlex Y	Krenke *et al.* (2004)
genRES MPX-2	Junge *et al.* (2003)
Reduced volume PCR for Profiler Plus STR kit	Gaines *et al.* (2002), Fregeau *et al.* (2003)
SGM	Sparkes *et al.* (1996a), Sparkes *et al.* (1996b), Kimpton *et al.* (1996)
TH01, VWA, F13A1, FES	Lygo *et al.* (1994), Clayton *et al.* (1995), Andersen *et al.* (1996)
CTT	Budowle *et al.* (1997)
D3S1358, D8S1179, D18S51	Potter (2003)
TH01, VWA, F13A1, FES, LPL	Pestoni *et al.* (1995)
STR sets	Crouse and Schumm (1995), Micka *et al.* (1996)
TH01	Van Oorschot *et al.* (1996), Wiegand *et al.* (1993)
D12S391	Junge *et al.* (1999)
Amelogenin	LaFountain *et al.* (1998)
Y-STR 4plex	Prinz *et al.* (2001)
Y-STR 10plex	Johnson *et al.* (2003)
19-locus Y-STR system	Daniels *et al.* (2004)
ABI 377	Frazier *et al.* (1996), Fregeau *et al.* (1999)
ABI 310	Lazaruk et al. (1998), Isenberg et al. (1998), Moretti et al. (2001)
ABI 3100	Sgueglia et al. (2003), Koumi et al. (2004)
ABI 3700	Gill et al. (2001), Koumi *et al.* (2004)
MegaBACE	Koumi *et al.* (2004)
DNA quant (RT-PCR *Alu*)	Nicklas and Buel (2003)
DNA quant (RT-PCR CSF)	Richard *et al.* (2003)
DNA quant (AluQuant)	Mandrekar *et al.* (2001)
BodeQuant	Fox *et al.* (2003)
Quantifiler	Applied Biosystems (2003)
Biomek 2000 with DNA IQ	Greenspoon *et al.* (2004)
mtDNA sequencing	Wilson *et al.* (1995), Holland and Parsons (1999)
mtDNA minisequencing	Morley *et al.* (1999)
TrueAllele software	Kadash *et al.* (2004)
CompareCalls software	Ryan *et al.* (2004)

Table 2. Summary of the number of times validation is mentioned in the title of talks and posters presented at the International Symposium on Human Identification meetings sponsored by the Promega Corporation (Madison, WI). More than 10% of the 1,220 presentations made prior to 2004 clearly involved validation.

Meeting #	Year	Validation in Title	Total Talks	Validation in Title	Total Posters	%
1st	1989	1	10	--	--	10.0
2nd	1991	0	21	0	14	0
3rd	1992	0	24	0	26	0
4th	1993	0	28	5	38	7.6
5th	1994	5	26	5	46	13.9
6th	1995	2	26	5	57	8.4
7th	1996	2	30	1	77	2.8
8th	1997	3	34	11	81	12.1
9th	1998	3	25	14	80	16.2
10th	1999	0	44	7	70	6.1
11th	2000	8	33	11	107	13.6
12th	2001	4	30	7	76	10.4
13th	2002	2	27	8	78	9.5
14th	2003	4	26	17	86	**18.8**
15th	2004					
	TOTAL	34	**384**	91	**836**	10.2

Table 3. Summary of PowerPlex Y validation studies and numbers of samples examined (Krenke et al. 2004). Heterozygote peak height ratios are not applicable (N/A) for single-copy Y-STR loci.

Study Conducted (17 studies done)	**Description of Samples Tested (performed in 7 labs and Promega)**	**# Run**
Single Source (Concordance)	5 samples x 8 labs	40
Mixture Ratio (male:female)	6 labs x 2 M/F mixture series x 11 ratios (1:0,1:1,1:10,1:100,1:300,1:1000,0.5:300, 0.25:300,0.13:300, 0.06:300, 0.03:300 ng M:F)	132
Mixture Ratio (male:male)	6 labs x 2 M/M mixtures series x 11 ratios (1:0,19:1,9:1,5:1,2:1,1:1,1:2,1:5,1:9,1:19,0:1)	132
Sensitivity	7 labs x 2 series x 6 amounts (1/0.5/0.25/0.125/0.06/0.03)	84
Non-Human	24 animals	24
NIST SRM	6 components of SRM 2395	6
Precision (ABI 3100 and ABI 377)	[10 ladder replicates + 10 sample replicated for 3100] + [8 ladders + 8 samples for 377]	36
Non-Probative Cases	65 cases with 102 samples	102
Stutter	412 males used	412
Peak Height Ratio	**N/A** (except for DYS385 but no studies were noted)	
Cycling Parameters	5 cycles (28/27/26/25/24) x 8 punch sizes x 2 samples	80
Annealing Temperature	5 labs x 5 temperatures (54/58/60/62/64) x 1 sample	25
Reaction volume	5 volumes (50/25/15/12.5/6.25) x [5 amounts + 5 concentrations]	50
Thermal cycler test	4 models (480/2400/9600/9700) x 1 sample + [3 models x 3 sets x 12 samples]	76
Male-specificity	2 females x 1 titration series (0-500 ng female DNA) x 5 amounts	10
TaqGold polymerase titration	5 amounts (1.38/2.06/2.75/3.44/4.13 U) x 4 quantities (1/0.5/0.25/0.13 ng DNA)	20
Primer pair titration	5 amounts (0.5x/0.75x/1x/1.5x/2x) x 4 quantities (1/0.5/0.25/0.13 ng DNA)	20
Magnesium titration	5 amounts (1/1.25/1.5/1.75/2 mM Mg) x 4 quantities (1/0.5/0.25/0.13 ng DNA)	20
	TOTAL SAMPLES EXAMINED	**1,269**

Table 4. Example of Pennsylvania State Police validation for the PowerPlex 16 kit (see http://www.cstl.nist.gov/biotech/strbase/validation/VSS_PASP_PP16.htm). This work represents both the internal validation performed by their lab plus assistance with Promega's developmental validation efforts (see Krenke *et al.* 2002). We encourage submission of similar information from internal validations completed by other laboratories to the new STRBase Validation Homepage: http://www.cstl.nist.gov/biotech/strbase/validation.htm.

Study Conducted	**Description of Samples Tested for PP16 Validation**	**# Run**
Single Source (Concordance)	8 samples (Promega concordance) + 200 samples (part of population concordance study)	208
Mixtures	45	45
Mixture Ratio	1 sample x 11 ratios (1:0, 19:1, 9:1, 4:1, 2:1, 1:1, 1:2, 1:4, 1:9, 1:19, 0:1) x 2 injections (5/10 seconds)	22
Sensitivity	5 samples x 8 amounts (5/2/1/0.5/0.25/0.125/0.06/0.03 ng) + [5 samples x 3 points (at/above/below dropout)]	55
Non-Human	11 animals	11
NIST SRM 2391b	All 12 components tested	12
Precision (ABI 310)	(5 samples x 10 injections each) + 10 injections of allelic ladders	60
Non-Probative Cases	5 cases x 4 samples each (evidence EF & SP/victim/suspect)	20
Stutter	200 samples (data used from population samples)	-
Peak Height Ratio	200 samples (data used from population samples)	-
Cycling Parameters	14 samples x 2 different cycle numbers (30/32) x 2 injection times (3/5 seconds)	56
Annealing Temperature	3 samples x 4 concentrations (2.0/1.0/0.5/0.25 ng) x 5 temperatures (56°C/58°C/60°C/62°C/64°C)	60
Proficiency	9 sets x 4 samples per set	36
Substrate	9 common substrates x 1 sample each	9
Environment	5 conditions (outside/80°C/50°C/4°C/RT) x 6 time points (3/6/12/25/48/85 days)	30
Various tissues	Bone, hair, teeth, semen, perspiration, urine, blood, semen, vaginal swab (minimum of one sample each)	9
	TOTAL SAMPLES RUN	**633**

Figure 1. Impact of number of experiments conducted on confidence associated with data collected. The curve is associated with the student's t test and reflects the number that is multiplied by the standard deviation of measurements made after various numbers of experiments. With a single experiment, repeatability is meaningless and therefore not defined with a data point. With two experiments, the Student *t* value is 12.7 and with three repeated experiments it drops to 4.30. Note that there is a diminishing return in terms of improvement with increasing the number of experiments. Thus, the number 5 is often selected as an optimal number of replicates to reflect the true population without having to run too many samples (e.g., consider that the minimum allele frequency is typically defined as 5/2N, or at least 5 observations are needed).