Meeting the Growing Demand for DNA Typing Results in the Justice System

By Kevin C. McElfresh, Young Kim, and Debbie DiPierro,
The Bode Technology Group, Inc.
Springfield, Virginia, USA
www.bodetech.com

By keeping the focus clearly on catching the largest number of felons, it is much easier to understand how to address the issue of processing as many samples as possible over the shortest period of time.

Easily one of the most important sociological impacts of biotechnology has been the use of DNA in the prosecution of criminals—forensic DNA typing. Recently there has been renewed interest in DNA typing. This is due to the impact that DNA databanks are having on unsolved criminal cases and as a result of highly publicized calls by Mayor Guiliani in New York City and others for point-of-arrest DNA typing. The justice community's need for large quantities of DNA typing data in a short period of time has grown dramatically in the last 5 years, and all indicators show that the need will continue to increase. Clearly, short tandem repeat (STR) typing, now the standard in DNA identification, is capable of meeting the scientific challenge of identification at a pace that will satisfy the justice system. However, the science is not the issue; the issue is the technological and mechanical support of STR typing. Therefore, the question is not whether to use DNA to identify someone, but rather how to process all of the samples.

While all the media attention generated by the success of DNA databanking has been gratifying to those of us laboring in the trenches to bring about these results, it is most important to remember why databanks exist. The recidivism rate for felons is stunning. Felon offender databanks exist to catch the felons who are repeat offenders and prevent further horrific violence. By keeping the focus clearly on catching the largest number of felons, it is much easier to understand how to address the issue of processing as many samples as possible over the shortest period of time.

FROM 8 WEEKS TO 8 HOURS

In 1986, the very first DNA identification cases took approximately 8 weeks to complete. Even with the long timeframe, DNA clearly had made its mark. However, the long and labor-intensive processing time, as well as the sample size and quality limitations, clearly precluded DNA's use as an investigative tool in criminal cases. Due to the long processing time for a restriction fragment length polymorphism (RFLP) case and backlog, a typical DNA identification case could take as long as eight months from sample receipt to final report. With the advent of the polymerase chain reaction[a] (PCR) and STRs, DNA identification became a much more capable tool for something more than a "cleanup" technology (PCR/STR technology can realistically be used in the investigative phase of a case). It is now possible to accomplish a DNA test in just eight hours or less. Now, the questions asked are how many samples can be processed per unit time, and how many loci are really necessary to accomplish the task of finding a single felon in a database? Once again, focusing on the reason that DNA databanks exist is the key to answering these questions realistically. At the highest level, the answers to how many samples per unit time, and how many loci, are easy: As many samples as possible, and just enough loci to do the job. On closer inspection the implementation of as many samples as possible and enough loci is a lot tougher.

The truly significant advantage of STRs is that there is platform independence. Unlike RFLP, this means that there is no need to be a slave to a single format, but rather a laboratory can choose one format for casework and a different format for databanking.

IT'S THROUGHPUT…

There is but one way to calculate throughput: number of samples entered into the databank per unit time. While there is but one calculation, there would appear to be as many ways to enter numbers into the categories of sample number and unit time as STR loci in the human genome. The truly significant advantage of STRs is that there is platform independence. Unlike RFLP, this means that there is no need to be a slave to a single format, but rather a laboratory can choose one format for casework and a different format for databanking so that throughput needs can be met. Also significant is that, while the Federal Bureau of Investigation has sanctioned 13 STR loci, until there is a single multiplex system for all 13 loci, there is no reason that

Table 1. Probability of a Coincidental Match.

Locus	TPOX	D13S317	D5S818	CSF1PO	TH01	D16S539	D7S820	vWA
N (individuals)	103,566	103,480	104,130	100,649	105,418	105,783	106,623	107,493
f[1]	0.113	0.109	0.107	0.105	0.102	0.099	0.096	0.092
P(1 match)[2]	1	1	1	1	1	0.278	0.004	3.8×10^{-5}
Genotype[3]	11,8	12,11	12,11	12,11	8,7	12,11	11,10	17,16
Number of Matches	23,856	4761	1040	170	14	5	0	0

[1] f is the geometric mean of the homozygosities of the loci. Therefore, loci are listed by ascending heterozygosity (1 – homozygosity) starting at the first column.

[2] Probability of at least 1 match (upper bound) using Equation 5.5 from the National Research Council II Report, 1996 (1). A '1' in the cell indicates statistical certainty.

[3] Most frequent genotype based on calculated frequencies. A missing genotype at a locus was not scored in the search.

all 13 should be analyzed prior to the entry of the data into the Combined DNA Index System (CODIS). In effect, 100,000 samples analyzed at 8 loci will solve more crimes than 50,000 samples analyzed at 13 loci. The data in Table 1 show why this is true.

The number of matches is based on a search of the database for the most common genotype. For example, a TPOX genotype of '11,8' yielded 23,856 matches on the database of 103,566 individuals. Adding the D13 locus genotype of '12,11' dropped the number of matches to 4,761 individuals. Clearly, after seven loci, no additional loci generate data that will contribute to additional discrimination within the database. Most importantly, this is the worst-case scenario of a single felon having the most common alleles at all eight loci. Further, while the number of matches, or "hits," per database search is ideally only one, the AFIS (automated fingerprint identification system) fingerprint experience demonstrates that there is substantial information gained even if there is more than one hit. Practically, just a few hits in the database will narrow the search enough for law enforcement to follow the leads and eliminate the other suspects in addition to running additional STR loci. It is also worth pointing out that adding an additional five loci (i.e., searching on all CODIS 13 loci) does not narrow the number of hits. If the search in Table 1 is reversed so that the most heterozygous loci are searched on first, then the number of loci needed to find a single person, using the most common genotype, is only six.

The other issue that requires discussion is the platform for processing the samples. Currently, there are slab-gel and capillary formats for analyzing STRs. Colloquially these methods are referred to as the FMBIO® and PE 310 (ABI® 310) formats. At The Bode Technology Group (TBTG) we are moving toward processing more than 3,000 samples per week, utilizing the Hitachi FMBIO® platform, SA43 gel boxes and analysts working standard hours. Based on our review of systems capabilities, single-channel capillary machines cannot meet this throughput capacity. (The obvious assumption is that the capital investment and space requirement for 30 PE310s is unreasonable and clearly unwarranted.) Multichannel capillary machines may well change the current situation. However, the reality is that the allele calls are NOT platform dependent as they were in RFLP testing, so it makes sense to use the highest throughput systems to get the samples typed and in the databank, no matter what system is used in the casework laboratory.

Once a felon is caught via a databank search, then every available locus should be used to make sure that there is no chance of this person ever getting away with another crime. The key however, is to catch him. To catch him, the samples have to be typed and in the databank. That is the easy part. The truly hard part is explaining to a grieving family why a felon, who should have been in prison, killed their loved one.

REFERENCES

1. National Research Council (1996) *The Evaluation of Forensic DNA Evidence*, Washington, DC, National Academic Press.

[a] The PCR process is covered by patents issued and applicable in certain countries. Promega does not encourage or support the unauthorized or unlicensed use of the PCR process.