IMPLEMENTING A LARGE STR DATABASE: CONSEQUENCES AND SURPRISES

Kevin C. McElfresh, Deborah DiPierro, Robert Bever, and Jim Schumm
The Bode Technology Group, Inc.

◇◇◇

At the time of the first Promega conference, ten years ago, a forensic case using DNA took 6 to 8 weeks on average to complete. Gels were generally run once a week, Friday to Monday. In any given week 12 to 15 cases containing 50 to 100 samples would reach completion. In spite of the time constraints, arguably, the first databank case of sorts was run. Timothy Spencer was subsequently convicted of rape and murder and finally executed. Last week, our laboratory typed over 2000 DNA samples, processing approximately 22 gels per day, each gel running about 2 hours. Last year our laboratory typed over 70,000 DNA samples. Now we have processed more than 100,000 samples and have discovered some rather interesting ways in which DNA has arranged itself with respect to tandem repeats.

One serious consequence of large scale DNA data banking is that μ variants are very common. While early attempts to characterize short tandem repeat (STR) systems sought to eliminate variation that was not "on-ladder" it has been our experience that the μ variants are the heart and soul of STR systems and produce additional heterozygosity. Our lab has been systematically working to substantiate and validate μ variants so that they can be made part of the standard group of alleles within their respective loci. During the course of this project we have also uncovered variation within variation. For example, we have sequenced several vWA 22 alleles and discovered that while all alleles are in fact 22 repeat blocks long, the arrangement of repeat blocks can differ.

One very positive consequence and a surprise, is that DNA is playing a quality assurance role for the samples that are being submitted from the various systems submitting samples. From the population of people within one of the databank systems we are examining, one person may be sampled several times in several different jurisdictions. Not surprisingly, the person has given a different name than previously, but more importantly, the race and even the sex has changed. These samples all show up as duplicates and are flagged as a result of the different demographic characteristics.

Perhaps most remarkably, the staff of both our databank and casework groups have established working teamwork relationships with the state and federal groups that we serve. This approach has lead to the solution of many issues long before any of them became problems. We have been able to mesh our systems together in such a way that data can move as fast as possible and as transparently as possible to the host organization.