MACRO

Customizing Genotyper® Allele-Call Macro—PowerTyper™ 16 Macro, Version 2

By John R. Ertl
DNA Analysis Unit, Wisconsin Department of
Justice Crime Laboratory, Madison Wisconsin

…modifications were made to the allele-call macros to provide concordance with manual analysis of data. One of the resulting allele-call macros, designed for use with the PowerPlex® 16 System, became the basis for version 2 of the PowerTyper™ 16 Macro.

INTRODUCTION

The Federal Bureau of Investigation describes a number of requirements designed to ensure a high level of quality, integrity, and competency within forensic DNA testing laboratories (1). Among these is the requirement for the internal validation of DNA typing systems prior to implementation. Part of this validation process involves the generation of empirical data upon which match criteria (allele calls) are to be based.

The Wisconsin State Crime Laboratory System is utilizing commercially available amplification kits to analyze the thirteen (13) short tandem repeat (STR) nuclear DNA loci required for entry into the Combined DNA Index System (CODIS). Internal validation studies have been done with the AmpFlSTR® Profiler Plus™ and COfiler™ Amplification Kits (Applied Biosystems) and more recently with the PowerPlex® 16 System (Promega Corporation). In each case, the ABI PRISM® 310 Genetic Analyzer and the associated software suite was used for automated fluorescent microsatellite analysis.

Occasional discrepancies were noted between the results of the automated Genotyper® Software (version 2.5) analysis and the results of a manual analysis of the GeneScan® Analysis Software (version 3.1) data. The causes of these discrepancies were investigated, and where possible, modifications were made to the allele-call macros to provide concordance with the manual analysis of the data. One of the resulting allele-call macros, designed for use with the PowerPlex® 16 System, became the basis for version 2 of the PowerTyper™ 16 Macro.

THE AUTOMATED ALLELE-CALLING PROCESS

The GeneScan® Analysis Software generates three pieces of information about each peak in the electropherogram that are ultimately used by the Genotyper® Software for assigning allele designations: the color of the peak (according to the fluorescent primer tag), the position of the peak (the DNA fragment size in nucleotide bases) and the intensity of the peak (peak height in relative fluorescence units, RFUs).

The locus/allele designations are predefined for peaks of each particular color and at specific positions. These are found under the appropriate headings in the categories window of the Genotyper® Software. A similar set of offset categories (category.os) use these defined positions and the actual allelic ladder peak positions to fine-tune the definitions of the category positions. The difference between the expected position and the actual position of a given peak determines its offset. The offsets are then incorporated into the appropriate category definitions. This fine-tuning compensates for general electrophoretic mobility shifts due to subtle changes in the instrumentation or the environmental conditions with time.

MACRO

Locus/allele designations are assigned to each peak of a sample profile based on these adjusted category definitions. Locus/allele designations thus assigned to probable stutter peaks are then filtered out based on their peak intensity relative to that of the associated allele peak.

In the discussion that follows, it is assumed that there were no obvious problems with the data (i.e., that peaks are on scale and above the detection threshold and that there are no interfering artifacts, spikes, dye blobs or baseline shifts present that might otherwise lead to a problem. Discussions regarding stutter filters based on an associated allele peak apply to the casework (POWER) macro only. PowerTyper™ 16 Macro, version 2, also contains an allele calling macro designed for databank samples (POWER 20%), which uses a non-specific peak filter based on the tallest peak in the category.

DISCREPANCIES—CAUSES AND SOLUTIONS

Rarely, an allelic ladder could not be analyzed properly in order to define the category offsets. The macro would terminate prematurely when an expected peak could not be found. The initial definitions of the expected peak positions were compared to the observed positions and the variability in the positions from our internal validation ladder precision study. At least two of the default locus/allele definitions differed significantly from our empirical data. Redefining the category default positions and adjusting the allowed variation (+/– three standard deviations) has eliminated this problem in our laboratories. The default locus/allele definitions in version 2 of the PowerTyper™ 16 Macro, however, remain the same as in version 1.

Occasionally, a leading stutter peak at a given locus in the allelic ladder would be incorrectly assigned as the first allele at that locus. It then appears as if the allele designations are shifted one position to the left in the affected locus. In part, redefining the default position and variation, especially of the first peak at each locus, as described above served to reduce the occurrence of this problem. The real solution came in redefining the first offset category of each locus to allow for the possibility of the presence of more than one peak. A subsequent filter then removes the smaller stutter peak assignment, leaving the proper allele peak assignment. This modification has been implemented in version 2 of the PowerTyper™ 16 Macro.

Several discrepancies between the results obtained using the Genotyper® Software automatic allele-call macros and those obtained via manual analysis of the GeneScan® Software data have been resolved using the combination of the PowerTyper™ 16, version 2, and the customization of the macro utilizing data from our laboratory's internal validation studies.

In forensic mixtures, there may be major and minor contributors to the total amount of DNA that is recovered from a sample. The resulting profile may then contain peaks with a wide range of intensities. In these situations it was noted that many of the smaller peaks were not given locus/allele assignments regardless of their position. This problem was traced to the presence of generalized size filters being present in the original macro at each locus. They were removing any and all locus/allele designations from peaks with intensities less than a given percentage of the most intense peak at that locus. This general filter may have acted as a crude stutter filter in earlier versions of the Genotyper® Software, however more specific stutter filters are now possible. These general filters have been removed from version 2 of the PowerTyper™ 16 Macro. It should be noted that if a peak from a minor contributor lies in the stutter position of a peak from a major contributor, and has an intensity less than might be expected from stutter alone, it will not be given a locus/allele assignment.

In a similar situation, true allele peaks of lower intensity falling in the stutter position of a more intense peak may not be given a locus/allele designation even though they are more intense than is defined for a stutter at that locus. The cause has been described previously (2) and is attributable to the fact that the Genotyper® Software sequentially applies each stutter filter to all categories within that same color. Accordingly, once the appropriate stutter cutoffs are determined from the internal validation data (e.g., three standard deviations above the average at each locus), it is imperative that the order in which the loci are processed within each color proceed from the highest to the lowest level of stutter filter. This is accomplished by reordering the macro script locus

command blocks within the PowerTyper™ 16 Macro step window.

There are also instances where peaks in a stutter position are given a locus/allele designation even though they are less intense than is defined for a stutter at that locus. This problem comes from the arithmetic used in determining if a peak falls above or below the defined stutter cutoff. Typically, stutter cutoffs are determined from the internal validation data analyzing single-source samples with stutter peaks, and are expressed as a simple percentage of the two peak intensities,

$$\left(\frac{L}{H}\right) \times 100,$$

(where L is the intensity of the stutter peak and H is the intensity of the allele peak). In comparing the stutter cutoff values used in version 1 with the generally accepted values for these loci, it appeared that the Genotyper® Software used the inverse expression,

$$\left(\frac{H}{L}\right) \times 100$$

to evaluate stutters (i.e., the filter for a locus with a 15% stutter cutoff had been set to 666%). Upon closer examination, it was determined that the Genotyper® Software uses a "percent higher" expression,

$$\left(\frac{(H-L)}{L}\right) \times 100$$

in which a filter set to 567% corresponds to a 15% cutoff while one set to 666% corresponds to a 13% cutoff. Thus peaks in a stutter position with an intensity greater than 13% and less than 15% would not be given a locus/allele designation. The stutter filters in PowerTyper™ 16 Macro, version 2, are based on the manufacturer's validation (3) but may be adjusted according to a laboratory's own internal validation.

Even given the modifications described here, stutter filter will fail any time there is a peak of the same color between the stutter peak and the allele peak. In version 1, the allele peak window was set from 0 to 5 bases (or 6 bases for the Penta loci) to the right of the putative stutter peak. Depending on the relative intensities of the three peaks involved, true stutter peaks may or may not be designated as alleles, and true allele peaks may or may not be designated. In version 2, the window to the right of the putative stutter peak in which the allele peak might be found is set from 3.5 to 4.5 bases for the tetranucleotide repeats and from 4.5 to 5.5 bases for the pentanucleotide repeats but these values may also be adjusted according to the internal validation data. Narrowing this window has two effects. First, only peaks in a stutter position will have the locus/allele designation removed (i.e., be deemed as a stutter), and second, minus-A peaks will be acknowledged as off-ladder peaks. A true stutter peak separated from its allele peak by another peak of the same color, of any height, will be given a locus/allele designation and will need to be dealt with manually.

SUMMARY

Several discrepancies between the results obtained using the Genotyper® Software automatic allele-call macros and those obtained via manual analysis of the GeneScan® Software data have been resolved using the combination of an improved macro template (available as PowerTyper™ 16 Macro, version 2) and the customization of the macro utilizing data from our laboratory's internal validation studies.

The macro is now designed to leave the original locus/allele designation on any and all peaks where the software is known to fail. This will hopefully alert the analyst to the presence of the peak and allow him/her to make the appropriate determination.

REFERENCES

1. Federal Bureau of Investigation (2000) Quality assurance standards for forensic testing laboratories, *Forensic Science Communications* **2**(3).
2. Kinsey, P.T. and Horman, S. (2000) Modification of the stutter position label-filtering macro in the PE Biosystems Genotyper® version 2.5 software package: Resolution of stutter-filter back talk, *Forensic Science Communications* **2**(3).
3. Krenke, B. *et al.* (2002) Validation of a 16-locus fluorescent multiplex system, *J. Forensic Sci.* **47**, 773–85.