David L. Duewer[1]
John M. Butler[2]

[1]Analytical Chemistry Division, National Institute of Standards and Technology, Gaithersburg, MD, USA
[2]Biochemical Science Division, National Institute of Standards and Technology, Gaithersburg, MD, USA

Received March 3, 2006
Revised April 21, 2006
Accepted May 11, 2006

Research Article

Multiplex_QA: An exploratory quality assessment tool for multiplexed electrophoretic assays

Multiplex_QA is a data analysis tool for visualizing short- and long-term changes in the performance of multiplexed electrophoretic assays, particularly the commercial short tandem repeat (STR) kits used by the human forensic identity community. A number of quality metrics are calculated from the signal collected for the internal size standard included in nearly all multiplex assays. These quality metrics are related to the signal intensity, symmetry, retention, resolution, and noise of data collected by capillary electrophoresis systems. Interlocking graphical displays enable the identification of changes in the quality metrics with time, evaluation of relationships among the metrics, and detailed examination of electropherographic features of particularly interesting analyses. While primarily intended for exploring which metrics are most useful for documenting data quality, the current version of the tool is sufficiently robust for use by forensic scientists with an interest in data analysis and access to a fast desktop computer.

Keywords: Electropherograms / Exploratory data analysis / Quality assessment / Resolution
DOI 10.1002/elps.200600116

1 Introduction

Short tandem repeat (STR) markers are used in a variety of applications including genetic mapping, disease diagnostics, and human identity testing to evaluate variation among individuals [1]. Commercial kits enable robust co-amplification of 15 or more STR loci using multiplex PCR assays [2–4]. The fluorescently labeled PCR products generated with the STR kits are typically separated by PAGE or CE with entangled polymer solutions and detected with laser-induced fluorescence [1, 5, 6].

A number of genetic analyzer instrument platforms have been introduced over the years and validated through examination of collected data [7–10]. Currently, the forensic human identity community mostly uses CE systems that evaluate one to a few dozen samples at a time. These commercial genetic analyzer systems do not readily provide the user with sufficient information for evaluating sensitivity, resolution, or other analytical parameters for quantitatively documenting electrophoretic quality of an individual sample or for many samples over time. Electrophoretic quality assessment is thus typically monitored indirectly through genotype results for negative and positive controls and through visual evaluation of electropherograms, although software for quantitatively evaluating genotype quality is now available [11–13]; OSIRIS, http://www.nlm.nih.gov/pubs/plan/lrp06/briefing/panel4/osiris.html).

With any data collection device, it is valuable to be able to assess possible changes in instrument performance over time. Key parameters with electrophoretic separations and fluorescent detection include signal intensity and stability as well as electrophoretic resolution. Evaluation of these and other quality metrics can help ensure data quality over time in a laboratory. Since there may be user-selectable signal processing options in software used to extract the desired information from the "raw" data, quantitative study of quality metrics may help optimize analysis protocols. A spreadsheet system named Multiplex_QA has been developed at the National Institute of Standards and Technology to assess and display these

Correspondence: Dr. David Duewer; NIST, 100 Bureau Dr., Mail Stop 8390, Gaithersburg, MD 20899–8390, USA
E-mail: david.duewer@nist.gov
Fax: +1-301-926-8671

Abbreviations: ABI, Applied Biosystems; **bp**, base pair; **D**^{3}, display, document, and discover; **IQRe**, expanded interquartile range; **ISS**, internal size standard; **NIST**, National Institute of Standards and Technology

key electrophoretic parameters for multiplexed STR systems. The Multiplex_QA system primarily exploits information provided by the internal size standards (ISSs) used to convert electrophoretic time to fragment size [6]. Since an ISS is added to virtually every sample analyzed, including negative controls and other blanks, electrophoretic performance can be monitored even when no allelic peaks are present.

The Multiplex_QA system operates within the Excel (Microsoft Corp., Redman, WA USA) environment. The entire open-source system and its User's Manual are freely available at http://www.cstl.nist.gov/biotech/strbase/software.htm. The User's Manual (i) details all of the system's capabilities, how they are intended to interact, and how to invoke them all, (ii) defines all of the available quality metrics, and (iii) specifies the nature and format of the data needed to make the system work. This article introduces the key features and capabilities of Multiplex_QA along with potential ways this program might be used to monitor electrophoretic data quality.

2 Source data and definitions

2.1 .fsa Binary files and BatchExtract text files

Currently, most forensic STR analysis is performed using instrumentation and proprietary software purchased from Applied Biosystems (ABI, Foster City, CA USA), particularly the single-capillary ABI 310 and the 16-capillary ABI 3100. These and other ABI systems store all "raw" and interpreted data for a particular analysis in an .fsa binary file. Multiplex_QA does not itself interpret .fsa files but rather reads data that have been extracted into text files by the BatchExtract system (ftp://ftp.ncbi.nlm.nih.gov/pub/forensics/BATCHEXTRACT/, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD USA). The November 2005 version of BatchExtract was accessed for this work. Information contained in these BatchExtract text files is used to create quality metrics. In principle, data from non-ABI STR analysis systems can similarly be extracted into text files; the information and format requirements are detailed in the Multiplex_QA User's Manual. Figure 1 displays the flow of information from sample workup to allelic typing, genotype assessment, or Multiplex_QA.

2.2 STR multiplex kits

A number of commercial STR multiplex kits are routinely used within the human identity community. Current kits typically use a combination of product size and dye labeling to evaluate allelic types at 7–17 or more simultaneously amplified genetic loci (Table 1). These STR kits label amplification products with three or four distinguishable dyes. One other dye is used to label the components of some ISS added to the PCR amplification products (amplicons) before electrophoretic analysis.

Although the identity of the multiplex kit used is immaterial to calculation of the quality metrics, an attempt is made to classify samples by multiplex kit to facilitate compar-

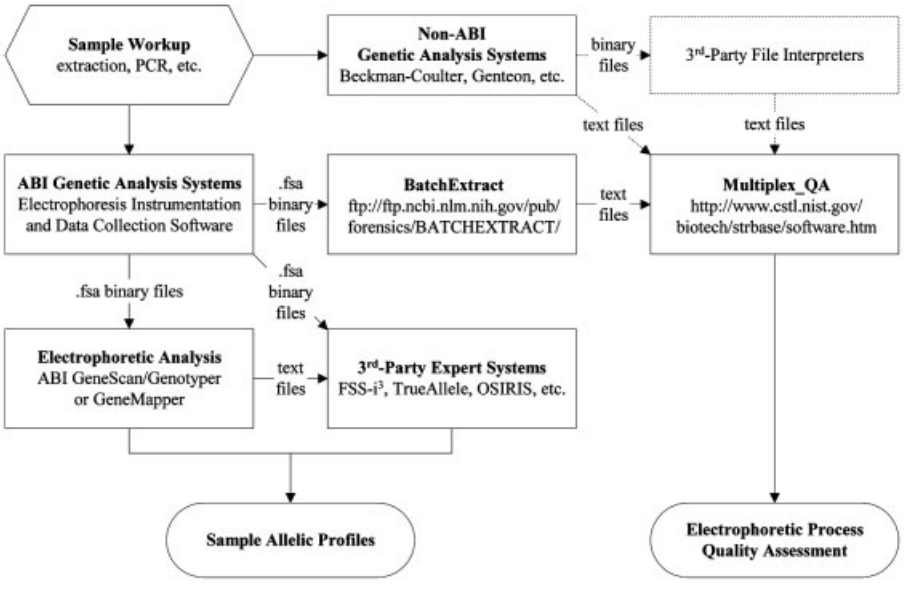

Figure 1. Information flow from sample workup to assessment of genotype quality or electrophoretic performance. Dotted lines indicate paths or processes that are not currently implemented.

Table 1. Commercial multiplex kits currently recognized by Multiplex_QA

Code	Name	#Loci	#Dyes
COF	ABI AmpFℓSTR COfiler	7	3
IDF	ABI AmpFℓSTR Identifiler	16	4
PP16	Promega PowerPlex 16	16	3
Pro	ABI AmpFℓSTR Profiler	10	3
ProP	ABI AmpFℓSTR Profiler Plus	10	3
SGMP	ABI AmpFℓSTR SGM Plus	11	3
Yfiler	ABI AmpFℓSTR Yfiler	17	4

ison and interpretation of the metrics. Table 1 lists the commercial STR multiplex kits that Multiplex_QA currently recognizes; however, the system is designed to evaluate data from any multiplex kit that uses four or fewer dyes.

2.3 Internal size standards

An ISS consists of fragments of known size labeled with a different color fluorescent dye from those used to label STR PCR products. Data for the ISS components enable association of electrophoretic retention time with DNA fragment size, expressed as the number of constituent nucleotide base pairs (bp's). This association is, in turn, used to align electropherograms of unknown samples with those for one or more "allelic ladder" external standards. Once aligned, each allelic signal in the unknown can be confidently evaluated against the known alleles present in the external standard. For successful allelic typing, the ISS components must respond to all electrophoretic influence factors, as the PCR products do, from the samples of interest. ISS components having electrophoretic behavior very similar to that of amplicon products are therefore present in nearly any multiplex assay.

Many of the quality metrics currently in the Multiplex_QA system are calculated from ISS information, either from the background-corrected ISS channel of the multi-dye electropherogram or from processed ISS retention time *versus* component bp size information. All such uses require that the identity of the ISS used in a given assay be recognized so that the proper time *versus* size relationship can be established. Four commercially available ISSs are currently recognized: ABI GS350, ABI GS400, ABI GS500, and Promega (Promega Corp, Madison, WI, USA) ILS600.

2.4 Peaks

The signal for any given DNA fragment of interest, whether an ISS component or an amplicon fragment, is evaluated as a "peak" in the background-corrected, dye-resolved electropherogram. The ABI software characterizes each peak with the time the signal is first recognized above baseline (t_{begin}), the time at which the signal reaches maximum intensity (t_{max}), the time the signal returns to baseline (t_{end}), the maximum height (H), and the total area under the signal envelope from t_{start} to t_{end} (A).

2.5 Samples

Although the number and nature of non-ISS marker peaks are irrelevant to the calculation of most quality metrics, three sample classes are recognized to facilitate comparison and interpretation of the metrics: "blanks", "allelic ladders", and "unknowns". Samples are considered to be blank if the only recognized peaks in the electropherogram are for ISS components. Samples are considered to be allelic ladders if the number of peaks in the non-ISS dyes is large and the variability in their signal intensities is low. Allelic ladder samples that cannot be associated with one of the known multiplex kits (Table 1) are labeled "unknown ladder". All non-blank, non-allelic ladder samples are considered to be unknowns.

2.6 Plates

Samples that are analyzed closely together in time have typically been amplified using the same reagents and conditions. To enable calculation of "average" quality metrics, the Multiplex_QA system attempts to group samples into "plates" of related samples on the basis of injection time, storage location, and/or capillary index.

3 Quality metrics

Multiplex_QA calculates a number of quality metrics for every sample. Table 2 lists the metrics, loosely classified as relating to signal intensity, noise, peak symmetry, retention, height/area, and resolution.

3.1 Intensity

Signal intensity metrics are estimated from the peak parameters stored by the ABI analysis software: the number of recognized peaks of each dye, robust estimates of the average value and variability of the $\log_{10}$(peak maximum signal) for all peaks of each color, and robust estimates of the average value and variability of the $\log_{10}$(peak areas) for all peaks of each color. The median is used to estimate the robust average; the expanded interquartile range (IQRe) is used to estimate a robust standard deviation [14, 15].

Table 2. Quality metrics

Type	No.	Description
Intensity	≤5	Number of peaks for each dye
	≤5	Median of $\log_{10}$(peak areas) for each dye
	≤5	IQRe of $\log_{10}$(peak areas) for each dye
	≤5	Median of $\log_{10}$(peak heights) for each dye
	≤5	IQRe of $\log_{10}$(peak heights) for each dye
Noise	1	Average height of ISS components used in regression models
	1	Number of data used to calculate noise
	1	Average noise
	1	SD of noise
	1	Log_{10}(% S/N)
Symmetry	≤5	Median of peak left/right symmetry for each dye
	≤5	IQRe of peak left/right symmetry for each dye
Retention	3	Summary statistics for regression model (t_{max} *versus bp*)
	7	Model parameters and parameter errors
	1	Predicted retention time for a peak of size 250 bp
	3	Summary statistics for inverse-model (*bp versus* t_{max})
	≤2	Predicted bp for the observed t_{max} of the ISS 250 bp component
Height/Area	3	Summary statistics for regression model (*H/A versus bp*)
	6	Model parameters and parameter errors
	2	Min and max observed/predicted ratio
	1	Predicted *H/A* for a peak of size 250 bp
Resolution	1	Expected resolution between peaks of size 186 and 187 bp

Since the values for the peak maximum signal and peak areas of different samples can vary over several orders of magnitude, these values are logarithmically transformed, $Y = \log_{10}(X)$, before they are summarized. The median of $\log_{10}$-transformed values is easily back-transformed, $X = 10^Y$, to a value that is very similar if not exactly the same as the median of the original values. However, the IQRe of the $\log_{10}$-transformed values back-transforms to a multiplicative factor standard deviation and not the usual additive quantity [16]. That is, about 68% of the original values summarized by the median and the IQRe of the $\log_{10}$-transformed data are expected to be in the interval $10^{\text{median}}/10^{\text{IQRe}}$ to $10^{\text{median}} * 10^{\text{IQRe}}$, not within the interval $10^{\text{median}} - 10^{\text{IQRe}}$ to $10^{\text{median}} + 10^{\text{IQRe}}$.

3.2 Noise

Signal noise metrics are estimated from the ISS electropherographic data in the between-component regions between ISS components from 200 to 300 bp. The average and SD of all data within these regions provide estimates of the absolute baseline noise. The ratio of SD to the average height of the ISS components from 200 to 300 bp estimates the relative noise for the region where the retention and peak height/peak area models are most descriptive (see below).

3.3 Peak symmetry

Peak symmetry metrics for each dye are estimated as the robust median and IQRe summaries of a symmetry parameter, *PS*, calculated from the t_{begin}, t_{max}, and t_{end} times estimated in the ABI analysis software

$$PS = \frac{t_{max} - t_{begin}}{t_{end} - t_{max}}$$

3.4 Retention

Peak retention metrics are calculated using an empirical regression model to relate the t_{max} times of the ISS peaks from 150 to 350 bp to their nominal bp size, *bp*. Since the 250-bp component used in some ISSs has anomalous electrophoretic properties [17], it is not used. To facilitate interpretation of the regression coefficients, times are expressed relative to the observed t_{max} of the nominal 200 bp ISS component, $t_{max,200}$. In a further attempt to standardize the interpretation of the coefficients, the regression is performed on orthogonalized variables; however, the conceptual model used to predict the time of the peak maximum for a fragment of nominal size *bp*, $t_{max,bp}$, is

$$t_{max,bp} = \alpha + \beta * bp + \gamma * bp^2$$

where α, β, and γ are the regression coefficients for this model.

The quality metrics evaluated include the regression parameters and their estimated SDs, the correlation between the observed and predicted t_{max} of the ISS peaks used to establish the model, and the SD of the differences between the observed and predicted time values. The residual SD of the inverse model (*bp* as a function of t_{max}) is also calculated, as is the average bias between the ABI-assigned peak sizes and those predicted by the inverse

regression model. When the ISS includes a nominal 250-bp component, the bias between the predicted size and the nominal size is calculated.

3.5 Peak height/peak area

The peak height/peak area ratio metrics are calculated from an empirical regression model relating the ratio of the ABI-estimated *H* and *A*, *H*/*A*, to the nominal size of the ISS peaks from 150 to 350 bp. The ISS component of nominal 250 bp is again not used in the regression. To help standardize the interpretation of the coefficients, the regression is performed on orthogonalized variables. The conceptual model used to predict the ratio for a fragment of nominal size *bp*, H/A_{bp}, is

$$H/A_{bp} = a + b * \log_{10}(bp) + c/\log_{10}(bp)$$

where *a*, *b*, and *c* are the regression coefficients for this model.

The quality metrics evaluated include the regression parameters and their estimated SDs, the correlation between the observed and predicted H/A_{bp} of the ISS peaks used to establish the model, the SD of the differences between the observed and predicted ratios, and the predicted H/A_{250}.

To provide a rough guide to samples with abnormally thin peaks (*e.g.*, spikes) and wide peaks (*e.g.*, dye blobs), the ratio between the observed and the calculated *H*/*A* for all peaks of all colors is calculated. The minimum and the maximum of these ratio-of-ratio values are retained as potential quality metrics.

3.6 Resolution

A combination of the retention and *H*/*A* models enables calculation of expected electrographic resolution. Assuming that the peaks can be validly approximated as triangles, peak area = ½(peak height)(peak width). Therefore, the expected peak width for a fragment of size *bp*, W_{bp}, is

$$W_{bp} = 2\frac{A_{bp}}{H_{bp}} = \frac{2}{H/A_{bp}} = \frac{2}{a + b * \log_{10}(bp) + c/\log_{10}(bp)}.$$

The usual resolution metric for two given peaks is the ratio of the separation between the peak maxima to the average width of the peaks. Since resolution of peaks at 186 bp and 187 bp (the size of the HUMTH01 9.3 and 10 alleles in the ABI COfiler multiplex) is of practical interest, a representative resolution metric, R_{187}, is then

$$R_{187} = \frac{(t_{max,187} - t_{max,186})}{(W_{186} + W_{187})/2}$$

$$= \frac{(\alpha + \beta(187) + \gamma(187^2)) - (\alpha + \beta(186) + \gamma(186^2))}{\left(\frac{2}{a + b * \log_{10}(186) + c/\log_{10}(186)} + \frac{2}{a + b * \log_{10}(187) + c/\log_{10}(187)}\right)/2}$$

$$\cong (\beta + \gamma(373))(a + b * \log_{10}(186.5) + c/\log_{10}(186.5))/2$$

4 Graphical displays

Multiplex_QA contains utilities for importing BatchExtract-ed information, calculating the various quality metrics, combining the metrics for newly processed samples with those previously processed, selecting subsets of the available information, and otherwise managing the mass of available numbers. However, the heart of the system is graphical display of the calculated metrics against the date the .fsa files were created. Other displays enable visual investigation of relationships between specified pairs of quality metrics, the electropherograms for particular samples, and the details of the retention, *H*/*A*, and resolution models for such samples.

4.1 D^3 Plots

The Display, Document, and Discover (D^3) plot is intended to help identify patterns in electrophoretic performance over time. Figure 2 displays data for 2912 analyses by one NIST researcher on a particular ABI 310 from 12 January 2000 to 29 December 2000. More than 53% of these analyses were performed using the ABI COfiler multiplex, about 40% with ABI ProfilerPlus, about 5% Promega PowerPlex 16, and about 2% ABI Profiler. These analyses were performed for quite a variety of research projects, including the evaluation of CE column washing procedures; they do not represent the performance characteristics expected for an instrument in a production environment.

Because of the limited graphical space on the typical computer screen, no more than five metrics can be simultaneously displayed on a single D^3 plot. The metrics displayed in Fig. 2 include the resolution-related "Resolution 187" (Rc_{187}), intensity-related "Area ISS", retention-related "T for ISS250", and *H*/*A*-related "H/A for ISS250" (H/A_{250}). To give all of the metrics about the same graphical scale, they are shown in a robust standardized format: $Y = (X\text{-median})/\text{IQRe}$. All values for each metric are displayed, but those greater- or less-than ± 6 IQRe units of the median are displayed at the upper or lower edge of the panel. For metrics that are normally distributed, this should include virtually all of the observed values. Horizontal ± 3 IQRe lines help differentiate representative values from extreme outliers. A histogram of the values for

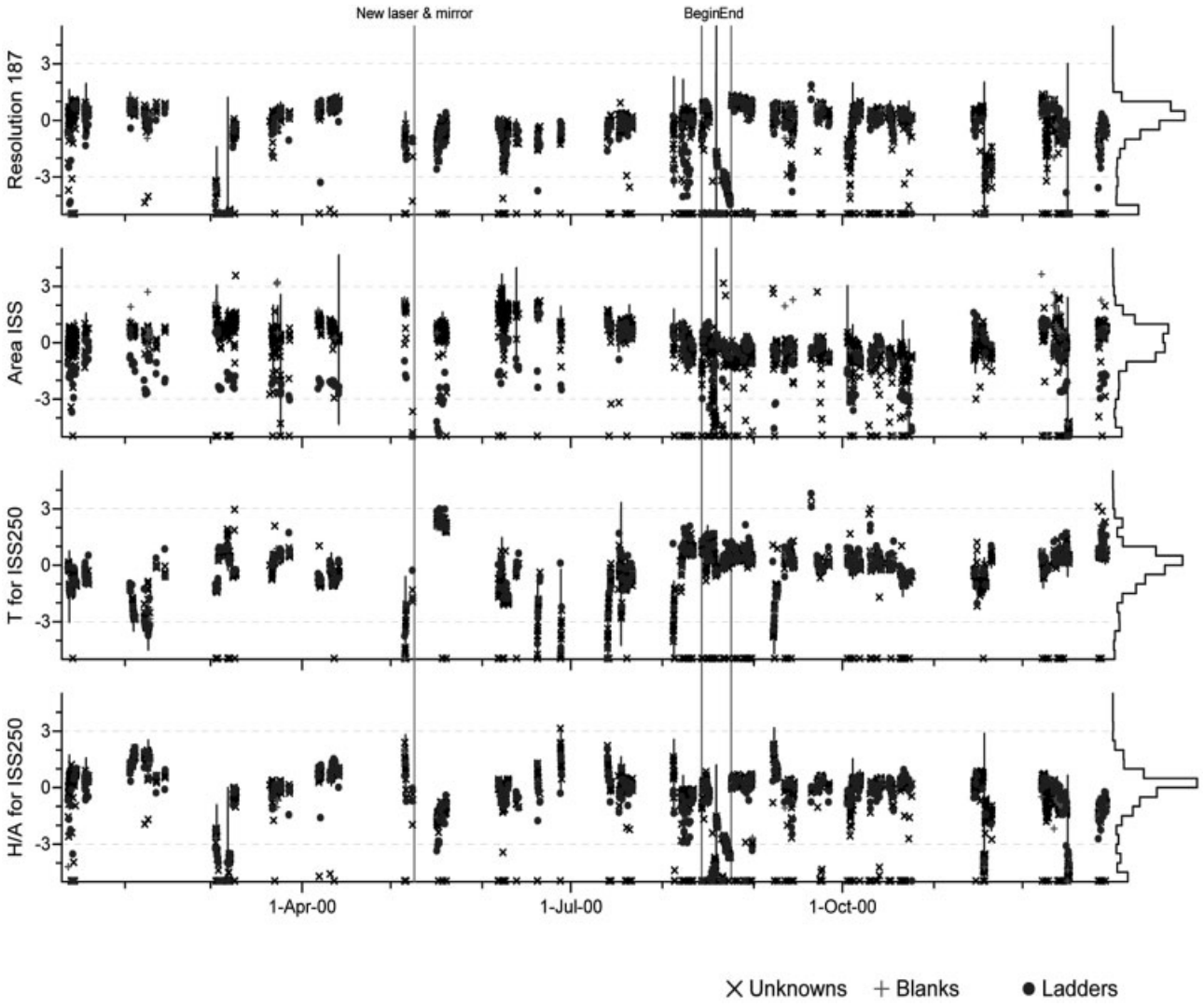

Figure 2. "Display, Document, and Discover" (D^3) plot of a year in the life of an ABI 310. This composite time-series displays standardized values for four quality metrics (vertical axis) for 2912 analyses performed from 2 January 2000 to 29 December 2000 on one ABI 310 as a function of calendar date (horizontal axis). D^3 plots can display one to five metrics on one page, at a time scale from years to hours. Unknown, blank, and allelic ladder samples are plotted separately. "Error crosses" provide approximate 95% confidence summaries of each metric for each plate of samples. The horizontal lines within each subplot denote approximate 99% confidence intervals, assuming that the metrics are normally distributed. The histogram to the extreme right of each subplot visualizes the data's empirical distribution. Vertical lines indicate when particular events occurred, such as maintenance or change in analysis protocol. Each of these graphical features can be independently activated.

each metric helps to visualize whether they can be usefully summarized with "normal" distribution estimates or whether they should first be transformed in some manner. A number of the metrics have underlying distributions that are significantly skewed from normal or have multimodal distributions. In addition to having a majority population that appears to be somewhat skewed from normal, all but the "T for ISS250" metrics have a minority population of very low values.

For each of the chosen metrics, Fig. 2 displays the values for all blank, ladder, and unknown samples analyzed. Plate summary values are also displayed; these "error bars" cover about 95% of the values of a given metric for all samples analyzed from the same plate. Vertical lines mark documented events in the life of the particular instrument. Each of these graphical features can be independently activated.

Long-term D^3 plots compactly document when and how heavily particular instruments are used and can reveal the influence of environmental factors: some of the structure in Fig. 2 is believed to be associated with the changes in room temperature and humidity associated with a Maryland summer. D^3 plots covering shorter periods, thus allowing better graphical resolution of metrics for individual analyses, are also of interest. Figure 3 displays metrics for the subset of data shown in Fig. 2 that were injected from 10 to 25 August 2000, including a period (from the "Begin" to the "End" events) of a column rejuvenation experiment. Both the *H*/*A*- and resolution-related metrics rather steadily degrade over the course of the experiment, with a sudden decrease for analyses performed on 18 August 2000 and a partial recovery during 19 August 2000.

4.2 Correlation

While the relationship between H/A_{250} and R_{187} is both obvious in the D^3 plot and expected from their definitions, more subtle relationships among the metrics can easily be masked by the presence of non-representative data. Correlation scatterplots, such as shown in Fig. 4, can help visualize any such relationships. Figure 4 displays all of the data shown in the D^3 plot of Fig. 3, with each unique pair of metrics plotted in its own panel. The ellipse in each panel should enclose about 95% of the pairs of representative values, assuming both metrics are approximately normally distributed [18]. In addition to the very strong positive correlation between H/A_{250} and R_{187}, both may be positively correlated with the signal intensity.

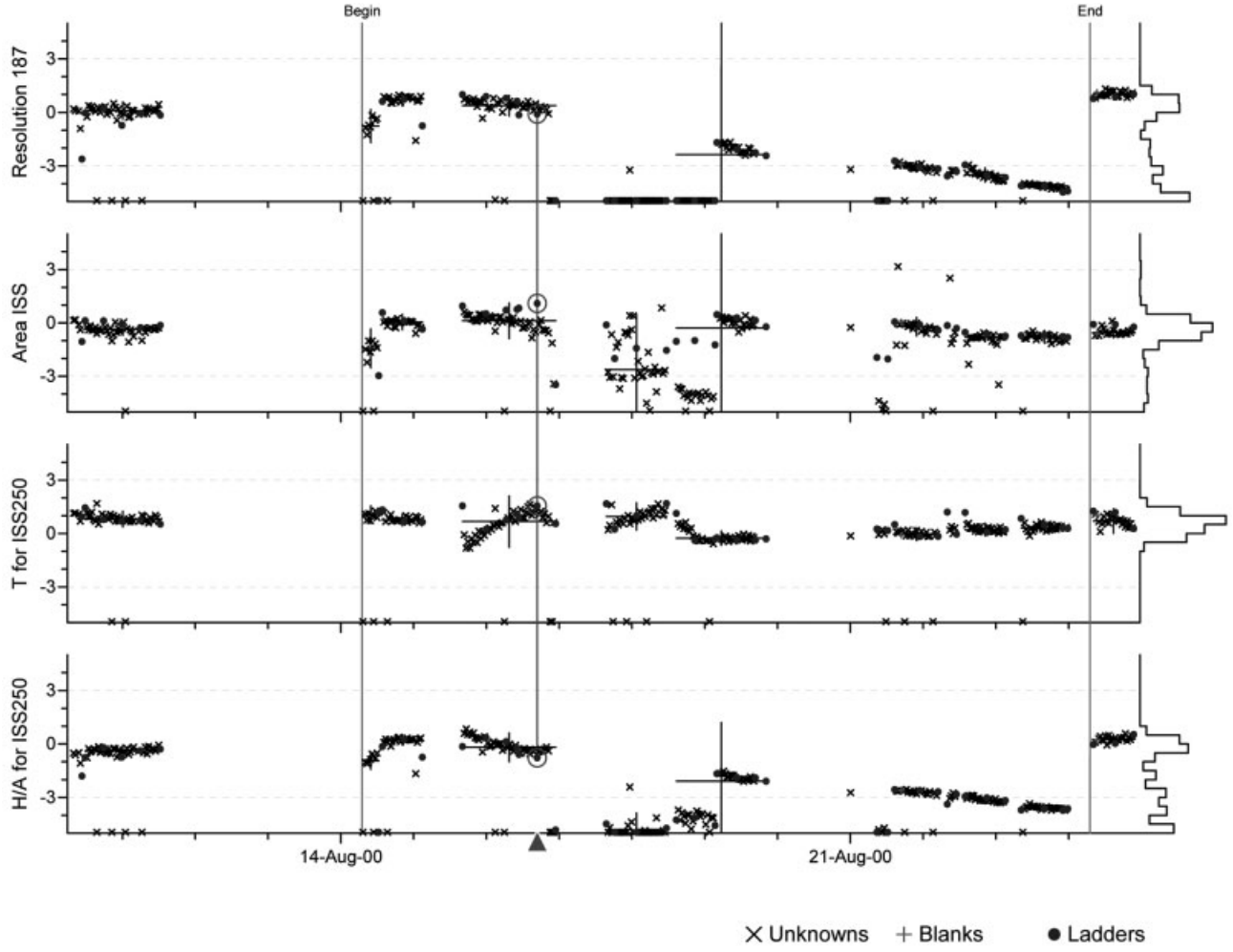

Figure 3. D^3 plot for the 10 August 2000 to 25 August 2000 segment of the data shown in Fig. 2. The higher temporal resolution unmasks a period of steadily declining electrophoretic resolution. The solid triangle at the bottom edge and the open circles connected by a solid line above it indicate the selection of a particular sample (injected 16 August 2000 16:46) for further analysis.

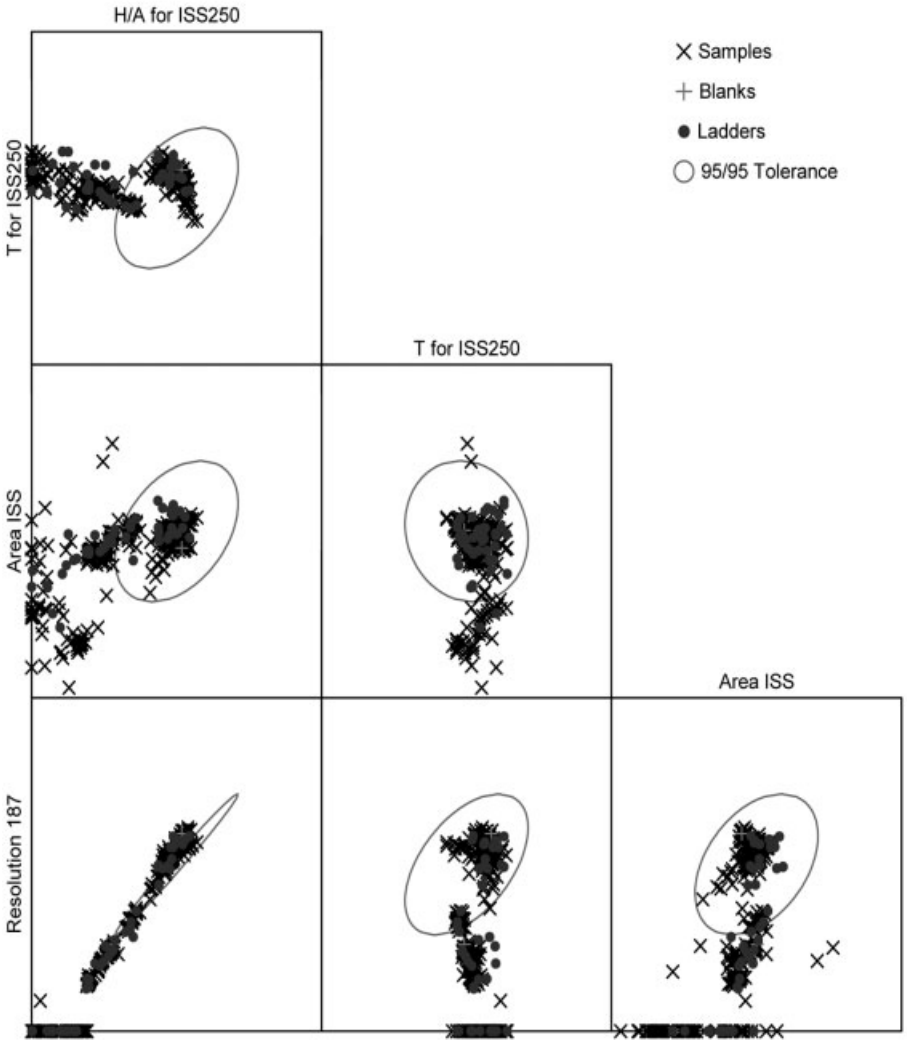

Figure 4. Correlations among the data of Fig. 3. This composite scattergram plots all values for each pair of metrics displayed in the active D^3 plot. As in the D^3 plot, unknowns, blanks, are plotted separately, as are 95% confidence error crosses for same-plate samples. The ellipse in each panel encloses about 95% of the paired values, assuming both metrics are normally distributed. Each graphical element can be independently activated. The display of samples within the ellipse can be suppressed.

4.3 Electropherograms

Individual electropherograms can be selected for close visual evaluation from a D^3 plot. The solid triangle at the bottom edge of Fig. 3 and the line-connected open circles above it indicate the selected analysis (injected 16 August 2000 16:46, a Promega PowerPlex 16 allelic ladder). Once selected, either a complete electropherogram or some designated portion of it can be displayed. Figure 5 displays a section of the GS500 ISS and D8S1179 and TPOX sections of the "yellow ladder" signals for this sample. To better visualize the low-level signals, the electropherogram is displayed using both a linear (upper segment) and logarithmic intensity axis (lower segment). While not readily apparent when viewed in the typical linear format, the logarithmic display reveals that the yellow and the ISS dyes are not completely separated ("pull-up"). However, the signals and the resolution are otherwise fairly typical.

4.4 Model plots

It is also sometimes useful to visualize the data and regression models used in the estimation of the retention, *H*/*A*, and resolution metrics. Like the electropherogram, model plots can be viewed for analyses selected from the D^3 plot. Figure 6 displays the *H*/*A* model and data in the panels to the lower left and the retention and resolution models in the panels to the lower right. Values are displayed for all of the observed ISS component peaks, with those used to parameterize the models shown as open circles and the rest shown as solid circles. The upper

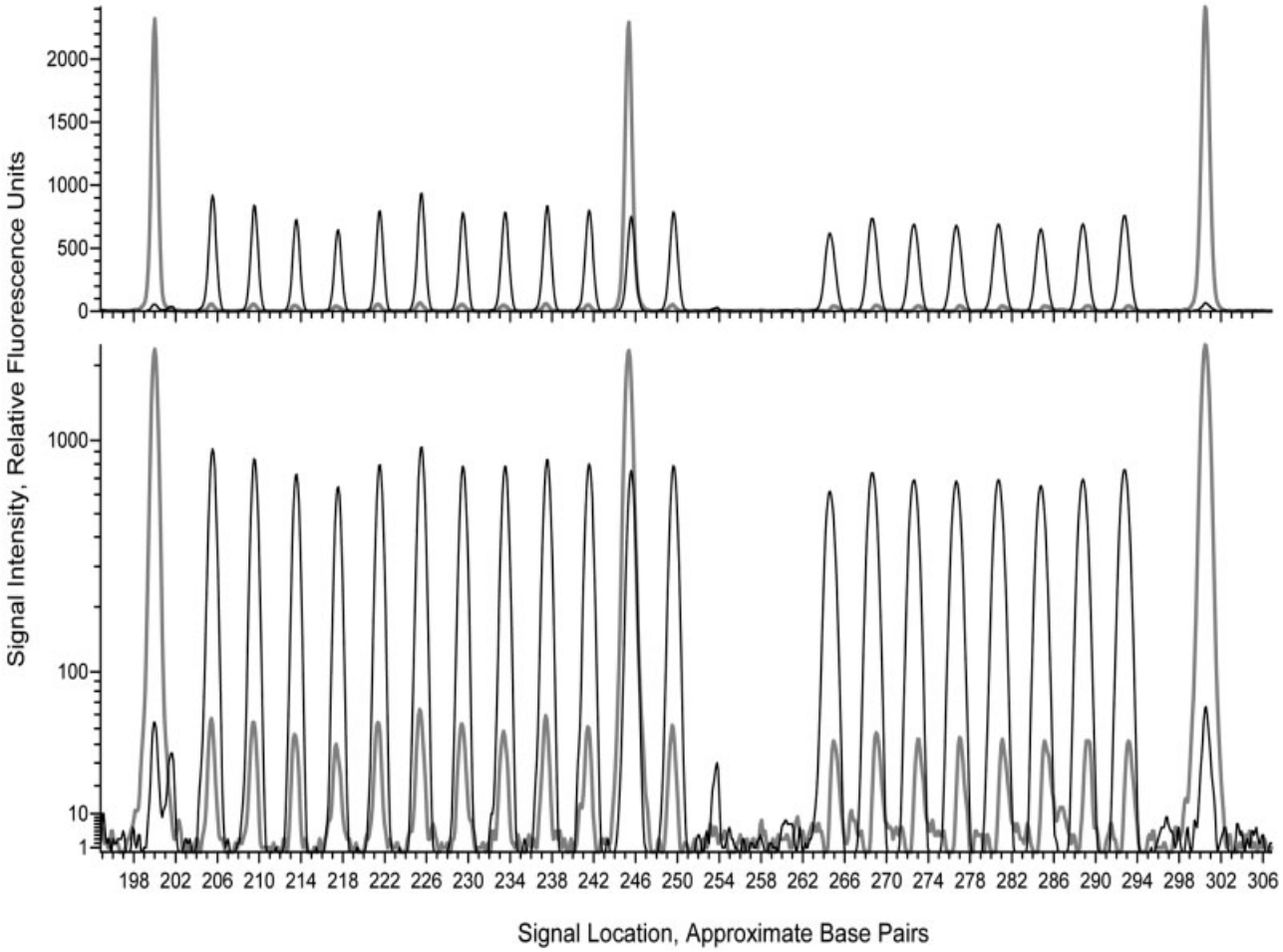

Figure 5. High-resolution electropherogram for the sample selected in Fig. 3. The fluorescence intensity (vertical axis) *versus* fragment size (horizontal axis) electropherogram for any contiguous interval of a selected analysis can be displayed. Here, the ISS and "yellow" dye signals for an allelic ladder sample are displayed over the fragment size range from 196 to 306 bp. From one to all of the dye signals available for the sample can be selected for display. To better visualize the low-level signals, the electropherogram is displayed using both a linear (upper segment) and logarithmic intensity axis (lower segment).

panels display the relative residuals, predicted minus observed, for the *H/A* and retention models. For this analysis, both models describe the behavior of all ISS components quite well.

5 Example exploratory analysis

Figure 7 displays a composite of D^3, electropherogram, and model plots used to visualize a section of an ABI COfiler allelic ladder sample injected 18 August 2000 01:28 during the period that resolution was significantly reduced. While the intensity of both the GS500 ISS and D7S820 "yellow ladder" signals are lower by about half and pull-up is much reduced, the ISS peaks are now quite right-tailed. While visible in the linear-intensity display, this asymmetry is most readily apparent when viewed in the logarithmic-intensity panel. Even though the retention model appears unchanged, the *H/A* model is badly degraded by the tailing. While the cause of the degradation cannot now be assigned to a specific cause, resolution was partially restored through capillary washing.

6 Conclusions

Multiplex_QA is an exploratory tool. It does not provide magic answers. Rather, it is intended to provide the tools you need to ask questions of your data such as: has electrophoretic performance changed?, when did it

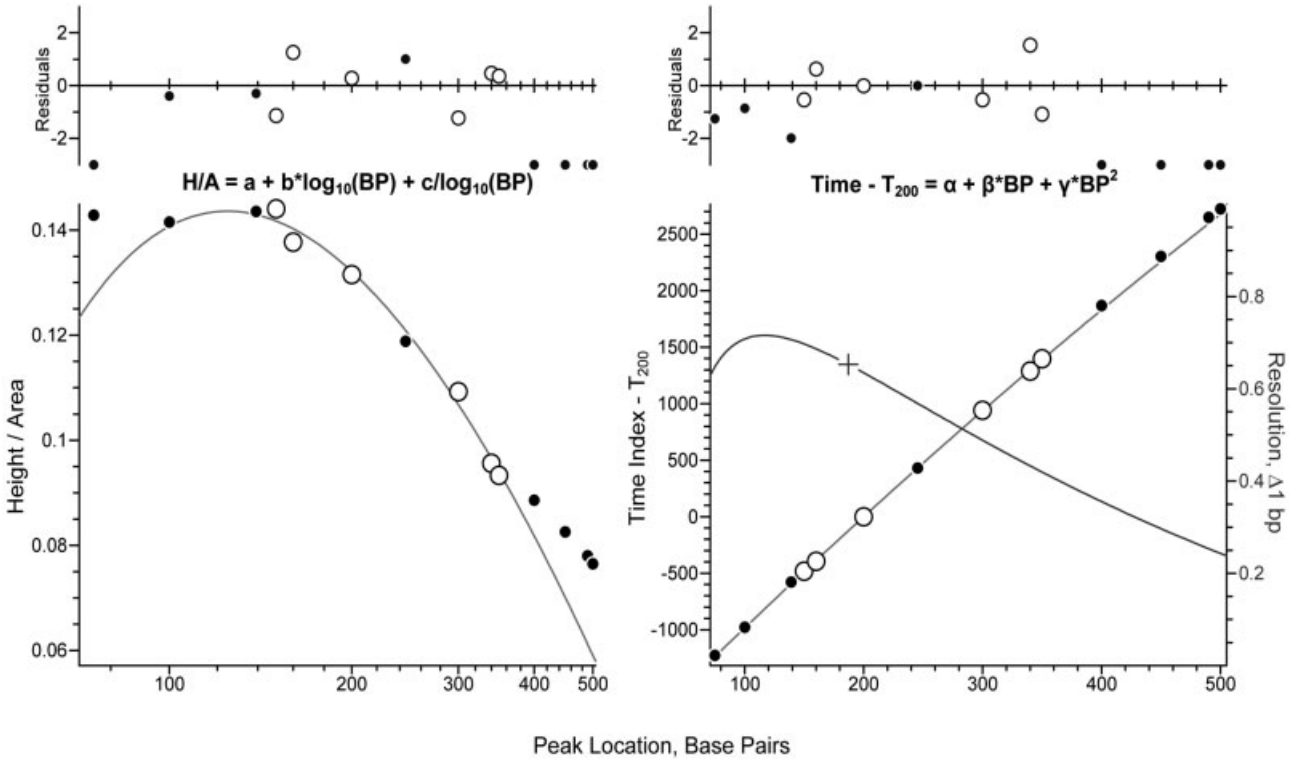

Figure 6. Height/area, retention, and resolution model plots for the sample selected in Fig. 3. The panel to the lower left presents the ISS height/area data (circles) and regression model (line) as a function of $\log_{10}$(fragment size). The panel to the lower right likewise presents the ISS retention time data and regression model as a function of fragment size. The data used to parameterize the models are shown as open circles. The upper panels display the relative residuals, predicted minus observed, for these models. The expected resolution for allelic peaks 1 bp different in size is also displayed in the lower right panel (light line); the "+" on this line denotes the expected resolution for the HUMTH01 9.3 and 10 alleles in the ABI COfiler multiplex.

change?, how much did it change?, which analyses are affected?, and what is going on in the electropherograms when a particular change is noted?

The D^3 plot is the heart of the system, visualizing different proposed quality metrics over time. Once "something funny" is recognized, a user of Multiplex_QA can identify the individual samples affected, confirm the data used in the regression models, look for relationships among the metrics, and view the complete or selected regions of the electropherogram. This information may help the user understand why things changed and whether corrective action is required.

A number of summary estimates are proposed as quality metrics, primarily addressing signal intensity, noise, peak symmetry, retention, height/area, and resolution. Although most of the electrophoretic data evaluated thus far reflect research usage, some of the observed patterns of change are associated with environmental and instrumental events. We anticipate that similar patterns exist in production data and that routine evaluation of electrophoretic performance of such data will prove useful in demonstrating and maintaining data quality over time.

We thank Margaret C. Kline, Janette W. Redman, Peter M. Vallone, Michael D. Coble, and Mary Satterfield of the Biochemical Science Division, Chemical Science and Technology Laboratory, National Institute of Standards and Technology, for their support, advice, criticism, data, and – above all! – time and patience during the development and debugging of Multiplex_QA. We thank Janos Murvai, Jon Baker, and Stephen Sherry of the National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, for developing BatchExtract and making it publicly accessible. We thank Helle Smidt Mogensen of the Department of Forensic Genetics, Institute of Forensic Medicine at the University of Copenhagen, for critically evaluating the Multiplex_QA system with fresh eyes and enthusiasm. This project was funded in part by the National Institute of Justice through interagency agreement 2003-IJ-R-029 with the NIST

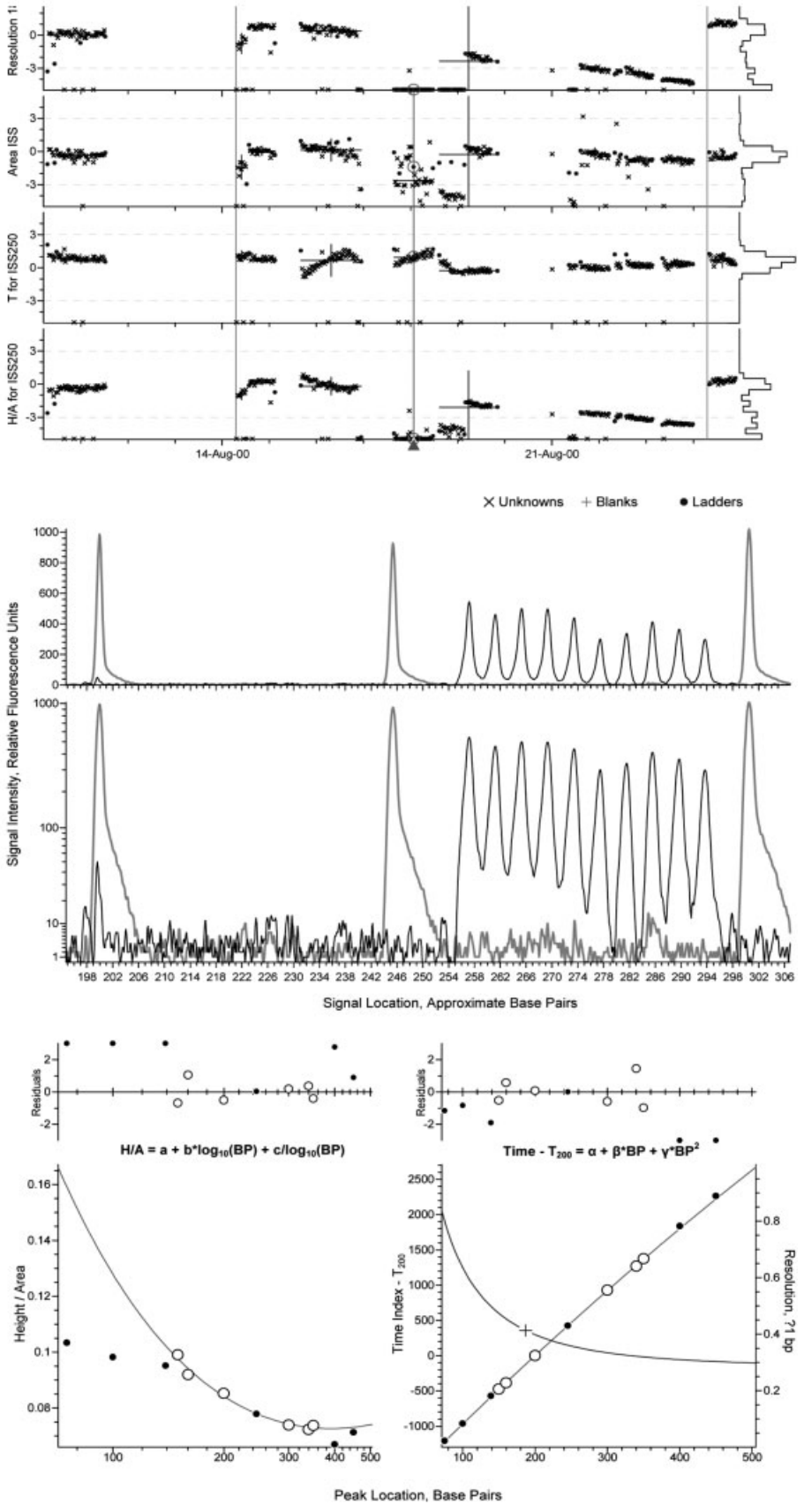

Figure 7. D^3, electropherogram, and model plots for an allelic ladder sample showing peak tailing and thus lower resolution.

Office of Law Enforcement Standards, awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. Contribution of the U.S. National Institute of Standards and Technology. Not subject to copyright. 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. Points of view are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Justice.

7 References

[1] Butler, J. M., *Forensic DNA Typing: Biology, Technology, and Genetics of STR Markers*, 2nd edn., Elsevier New York, 2005.

[2] Krenke, B. E., Tereba, A., Anderson, S. J., Buel, E. *et al.*, *J. Forensic Sci.* 2002, *47*, 773–785.

[3] Collins, P. J., Hennessy, L. K., Leibelt, C. S., Roby, R. K. *et al.*, *J. Forensic Sci.* 2004, *49*, 1265–1277.

[4] Mulero, J. J., Chang, C. W., Calandro, L. M., Green, R. L. *et al.*, *J. Forensic Sci.* 2006, *51*, 64–75.

[5] Wallin, J. M., Buoncristiani, M. R., Lazaruk, K., Fildes, N. *et al.*, *J. Forensic Sci.* 1998, *43*, 854–870.

[6] Butler, J. M., Buel, E., Crivellente, F., McCord, B. R., *Electrophoresis* 2004, *25*, 1397–1412.

[7] Moretti, T. R., Baumstark, A. L., Defenbaugh, D. A., Keys, K. M. *et al.*, *J. Forensic Sci.* 2001, *46*, 647–660.

[8] Moretti, T. R., Baumstark, A. L., Defenbaugh, D. A., Keys, K. M. *et al.*, *J. Forensic Sci.* 2001, *46*, 661–676.

[9] Holt, C. L., Buoncristiani, M., Wallin, J. M., Nguyen, T. *et al.*, *J. Forensic Sci.* 2002, *47*, 66–96.

[10] Wallin, J. M., Holt, C. L., Lazaruk, K. D., Nguyen, T. H., Walsh, P. S., *J. Forensic Sci.* 2002, *47*, 52–65.

[11] Kadash, K., Kozlowski, B. E., Biega, L. A., Duceman, B. W., *J. Forensic Sci.* 2004, *49*, 660–667.

[12] Perlin, M. W., in: *Proceedings of the Fourteenth International Symposium on Human Identification 2003*, Phoenix, AZ, September 29–October 2, 2003. http://www. promega.com/geneticidproc/ussymp14proc/oralpresentations/perlin.pdf.

[13] Bill, M., Knox, C., *Profiles DNA*, 2005, *8*, 8–10. http://www. promega. com/profiles/.

[14] Analytical Methods Committee, *Analyst* 1989, *114*, 1693–1698.

[15] Stuart, A., Ord, J. K. (Eds.), *Kendall's Advanced Theory of Statistics*, 5th edn., Oxford University Press, New York 1987, Vol. 1, Chapter 10, Section 11.

[16] Sharpless, K. S., Duewer, D. L., *Anal. Chem.* 1995, *67*, 4416–4422.

[17] Rosenblum, B. B., Oaks, F., Menchen, S., Johnson, B., *Nucleic Acids Res.* 1997, *25*, 3925–3929.

[18] Duewer, D. L., Liu, H.-K., Reeder, D. J., *J. Forensic Sci.* 1999, *44*, 969–977.

