applied
biosystems
life technologies

GeneMapper®*ID-X* Software

Version 1.5

**For Research, Forensic Paternity and Cell Line Authentication.
Not for use in diagnostic and therapeutic applications.**

life
technologies

The information in this guide is subject to change without notice.

DISCLAIMER

LIFE TECHNOLOGIES CORPORATION AND/OR ITS AFFILIATE(S) DISCLAIM ALL WARRANTIES WITH RESPECT TO THIS DOCUMENT, EXPRESSED OR IMPLIED, INCLUDING BUT NOT LIMITED TO THOSE OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, OR NON-INFRINGEMENT. TO THE EXTENT ALLOWED BY LAW, IN NO EVENT SHALL LIFE TECHNOLOGIES AND/OR ITS AFFILIATE(S) BE LIABLE, WHETHER IN CONTRACT, TORT, WARRANTY, OR UNDER ANY STATUTE OR ON ANY OTHER BASIS FOR SPECIAL, INCIDENTAL, INDIRECT, PUNITIVE, MULTIPLE OR CONSEQUENTIAL DAMAGES IN CONNECTION WITH OR ARISING FROM THIS DOCUMENT, INCLUDING BUT NOT LIMITED TO THE USE THEREOF.

Limited Use Label License No. 540: Sequencing or Fragment Analysis Intellectual Property for Human Identity Testing Software

Notice to Purchaser: No right to resell this software product, any upgrades, modified versions, updates, additions, copies, or any of its components is conveyed expressly, by implication, or by estoppel. This software product may be used to perform internal research and development for the sole benefit of the purchaser; provided, however, that Life Technologies Corporation ("LTC") does not convey any right or license, whether expressly, by implication, by estoppels, or otherwise, under any LTC patents that claim DNA sequencing or fragment analysis methods, to the purchaser by the purchase of this software product to use such DNA sequencing or fragment analysis methods. This software product is not for use in commercial applications of any kind, including, without limitation, quality control and commercial services such as reporting the results of purchaser's activities for a fee or other form of consideration, except as follows. A limited license to use the DNA sequencing or fragment analysis methods covered by such patents can be obtained for certain research and development activities (a) through the purchase of certain LTC reagents that convey rights for commercial use (e.g., LTC kits for performing human identity testing and LTC kits for performing in vitro diagnostics testing) when such reagents are used on an authorized LTC instrument, or (b) directly from LTC. In addition, LTC grants the purchaser a limited right to use this software product in conjunction with: (1) a second product purchased from LTC or that conveys rights for commercial use (e.g., LTC kits for performing human identity testing and LTC kits for performing in vitro diagnostics testing) and (2) an authorized LTC instrument that conveys rights for commercial use to perform services on a fee per test or contract basis.   For information on obtaining additional rights, please contact outlicensing@lifetech.com or Out Licensing, Life Technologies, 5791 Van Allen Way, Carlsbad, California 92008.

GeneMapper® *ID-X* Software v1.5 has undergone a verification process pursuant to Life Technologies Corporation and/or its affiliate(s) internal quality process. However, we recommend that customers installing or using GeneMapper® *ID-X* Software v1.5 perform any internal validation testing they deem necessary according to relevant guidelines. Each laboratory is solely responsible to ensure that its GeneMapper® *ID-X* Software v1.5 satisfies or will satisfy SWGDAM Guidelines or other applicable guidelines and is fit for each laboratory's applications.

This product includes software developed by the Apache Software Foundation.

This product includes software developed by the ExoLab Project.

JNIRegistry is copyrighted © by ICE Engineering, Inc.

TRADEMARKS

All trademarks are the property of Thermo Fisher Scientific and its subsidiaries unless otherwise specified.

Microsoft and Windows are trademarks of Microsoft Corporation.

Oracle is a trademark of Oracle Corporation.

Contents

About this guide

Revision history

Revision	Date	Description
A	February 2015	New document for v1.5 software.
		Replaces Pub. no. 4426481.
		Added information for ACC. Updated information for OS, PHR, SPK, AN, SSPK, MIX, OMR, CGQ. Updated allelic ladder requirements and Baseline Window size. Added Bin Overlap appendix.

Peak Detection, Sizing, Allelic Ladder, and Sample Analysis Algorithms

This chapter covers:

Allelic Ladder Requirements

Allelic Ladder Run Requirements

To accurately genotype samples, you must run an allelic ladder sample along with the unknown samples.

Instrument	Number of allelic ladders to run	One injection equals	Number of samples per allelic ladder(s)
3500	1 per 3 injections	8 samples	23 samples + 1 allelic ladder
3500xL	1 per injection	24 samples	23 samples + 1 allelic ladder
3130	1 per 4 injections	4 samples	15 samples + 1 allelic ladder
3130*xl*	1 per injection	16 samples	15 samples + 1 allelic ladder
3730	3 per injection	48 samples	15 samples + 1 allelic ladder

IMPORTANT! Variation in laboratory temperature can cause changes in fragment migration speed and sizing variation between both single- and multiple-capillary runs (with larger size variations seen between samples injected in multiple-capillary runs). We recommend the above frequency of allelic ladder injections, which should account for normal variation in run speed. However, during internal validation studies, verify the required allelic ladder injection frequency to ensure accurate genotyping of all samples in your laboratory environment.

It is critical to genotype using an allelic ladder run under the same conditions as the samples, because size values obtained for the same sample can differ between instrument platforms because of different polymer matrices and electrophoretic conditions.
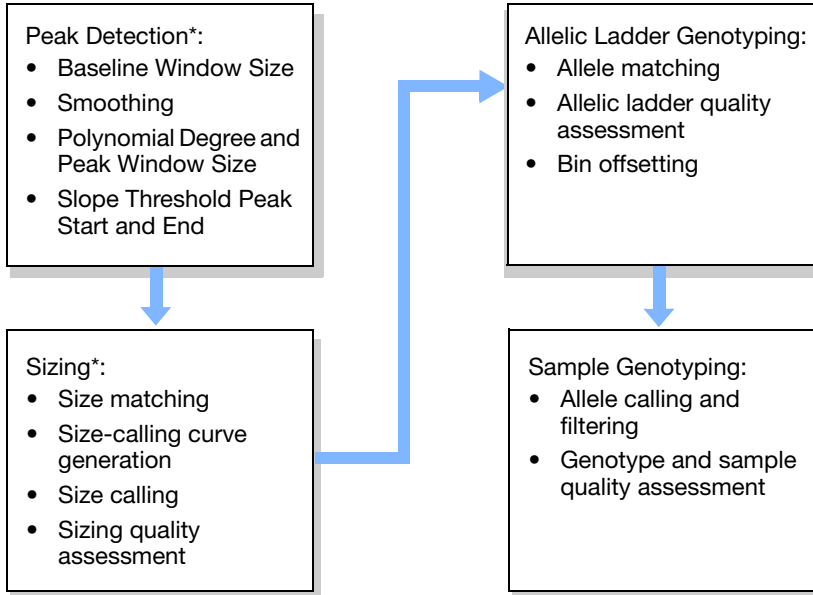
Allelic Ladder Analysis Requirements

- To accurately genotype samples, the GeneMapper® *ID-X* Software requires at least one allelic ladder sample per run folder.

  IMPORTANT! If a run folder contains multiple allelic ladder samples, the GeneMapper *ID-X* Software calculates bin offsets using an average of all ladders that use the same panel within a run folder (for more information, see "Bin Offsetting" on page 27).

- Allelic ladder samples must be identified as "Allelic Ladder" in the Sample Type column in a project.
- Analyze the allelic ladder sample(s) using the same analysis method and parameters used for samples.

Chapter 1 Peak Detection, Sizing, Allelic Ladder, and Sample Analysis Algorithms
Overview of the Peak Detection, Sizing, and Genotyping Process

1

Overview of the Peak Detection, Sizing, and Genotyping Process

Peak Detection*:
- Baseline Window Size
- Smoothing
- Polynomial Degree and Peak Window Size
- Slope Threshold Peak Start and End

Allelic Ladder Genotyping:
- Allele matching
- Allelic ladder quality assessment
- Bin offsetting

Sizing*:
- Size matching
- Size-calling curve generation
- Size calling
- Sizing quality assessment

Sample Genotyping:
- Allele calling and filtering
- Genotype and sample quality assessment

*Sample peak detection and sizing is performed after allelic ladder genotyping.

Peak Detection

Peak Detection:
- Baseline Window Size
- Smoothing
- Polynomial Degree and Peak Window Size
- Slope Threshold Peak Start and Peak End

Parameters in the Analysis Method Peak Detector tab determine how the raw data are baselined and smoothed, and how peaks are detected.
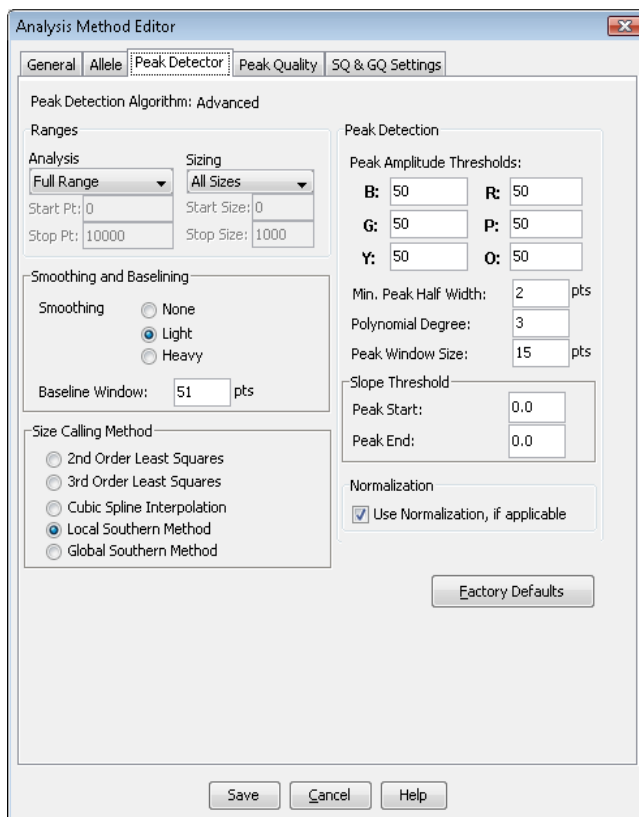
This section describes how these parameters work. For more information on these parameters, see the *GeneMapper® ID-X Software Help*.

IMPORTANT! This section provides examples of the effect of changing certain peak detection parameters. We provide default settings in the software, but your laboratory should verify these settings, or changes to these settings, as part of your laboratory's internal validation studies.

Examining Peak Definitions

To examine how GeneMapper® *ID-X* Software has defined a peak, select **Plot ▸ Peak Positions** in the Samples plot window. The peak positions (start, apex, end) are tick-marked in the electropherogram.
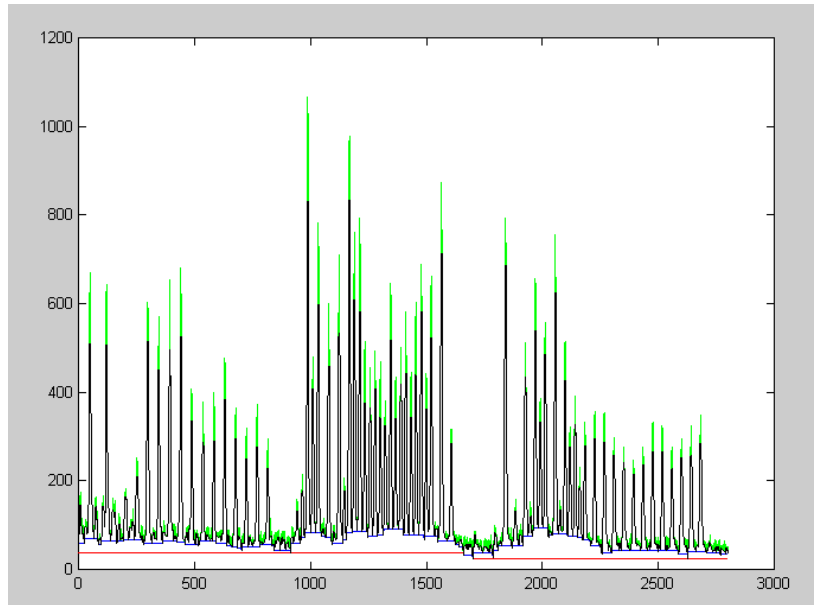
Baseline Window Size

Peak Detection:

- **Baseline Window Size**
- Smoothing
- Polynomial Degree and Peak Window Size
- Slope Threshold Peak Start and Peak End

The Baseline Window size parameter controls baselining for a group of peaks.

The software determines a reference baseline value for each data point. In general, the software sets the reference baseline to be the lowest value that it detects in a specified window size (in data points) centered on each data point.

A small baseline window relative to the width of a cluster, or grouping of peaks spatially close to each other, can result in shorter peak heights.

Larger baseline windows relative to the peaks being detected can create an elevated baseline, resulting in peaks that are elevated or not resolved to the baseline.

Guideline for Using This Parameter

The following settings are optimal for analysis:

- GlobalFiler®, GlobalFiler® Express, and Yfiler® Plus kit data: Change setting to 33
- All other AmpF*l*STR® kit data: Use default setting of 51

Choose a value that encompasses the width in data points of the peaks being detected while preserving a qualitatively smooth baseline.

The trade-off for a smoother baseline that touches all peaks is a reduction in peak height.

Effects of Varying Baseline Window Size

Figure 1 on page 7 shows a sample with different reference baselines (zero in the analyzed electropherogram) that result from different baseline window size settings:

- The red trace shows a baseline derived from an extreme baseline window size value of 2801. At this setting, the reference baseline does not touch all peaks and elevates peak heights.
- The blue trace shows a baseline derived from the default value of 51 data points.

• The black trace shows a baseline that is derived from an extreme baseline window size value of 5 data points. At this setting, the reference baseline tracks the peaks, significantly reducing peak height.

Figure 1   Baseline Window example

Smoothing

Peak Detection:
- Baseline Window Size
- **Smoothing**
- Polynomial Degree and Peak Window Size
- Slope Threshold Peak Start and Peak End

This parameter smooths the outline of peaks, and reduces the number of false peaks that are detected.

Smoothing is performed before peak detection and can be set to:

Option	Description
None	Applies no smoothing. Select for slower runs with very broad peaks, or to avoid the detection of sharp edges.
Light	Provides the best results for typical data. Light smoothing slightly reduces peak height.
Heavy	Select for data with very sharp, narrow peaks of interest. Heavy smoothing can significantly reduce peak height.

Figure 2 shows the effect of smoothing settings on a peak.
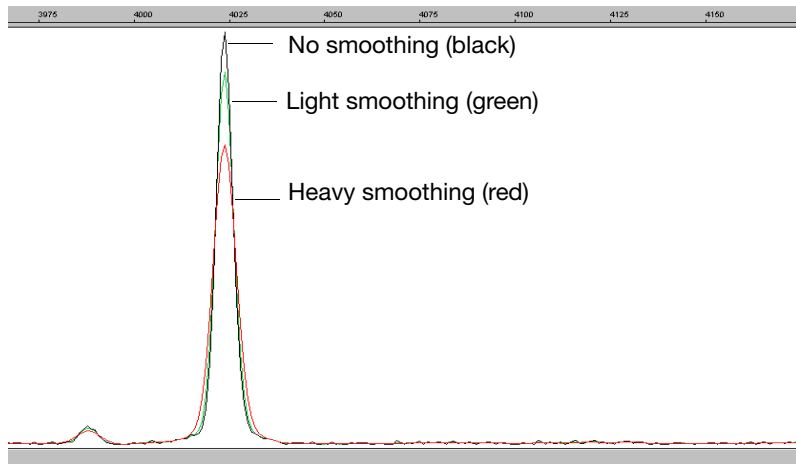
Figure 2   Smoothing example

Polynomial Degree and Peak Window Size

Peak Detection:
- Baseline Window Size
- Smoothing
- **Polynomial Degree and Peak Window Size**
- Slope Threshold Peak Start and Peak End

Polynomial Degree and Peak Window Size settings affect the peak detection sensitivity. You can adjust these parameters to detect a single base pair difference while minimizing the detection of shoulder effects and/or noise.

The peak window size functions with the polynomial degree to set the sensitivity of peak detection. The peak detector calculates the first derivative of a polynomial curve fitted to the data within a window that is centered on each data point in the analysis range.

Using curves with larger polynomial degree values allows the curve to more closely approximate the signal and, therefore, the peak detector captures more of the peak structure in the electropherogram.

The peak window size sets the width (in data points) of the window to which the polynomial curve is fitted to data:

- Higher peak window size values smooth out the polynomial curve, which limits the structure being detected.
- Smaller window size values capture more of the peak structure.

How to Use the Peak Detection Parameters

Use the table below to adjust the sensitivity of detection.

Function	Polynomial Degree Value	Window Size Value
Increase sensitivity	Higher	Lower
Decrease sensitivity	Lower	Higher

Effects of Varying the Polynomial Degree

Figure 3 on page 10 shows peaks detected with a window size of 15 data points and a polynomial curve of degree 2 (green), 3 (red), and 4 (black). The diamonds represent a detected peak using the respective polynomial curves.

Note that the smaller trailing peak is not detected using a degree of 2 (green). As the peak detection window is applied to each data point across the displayed region, a polynomial curve of degree 2 could not be fitted to the underlying data to detect its structure.
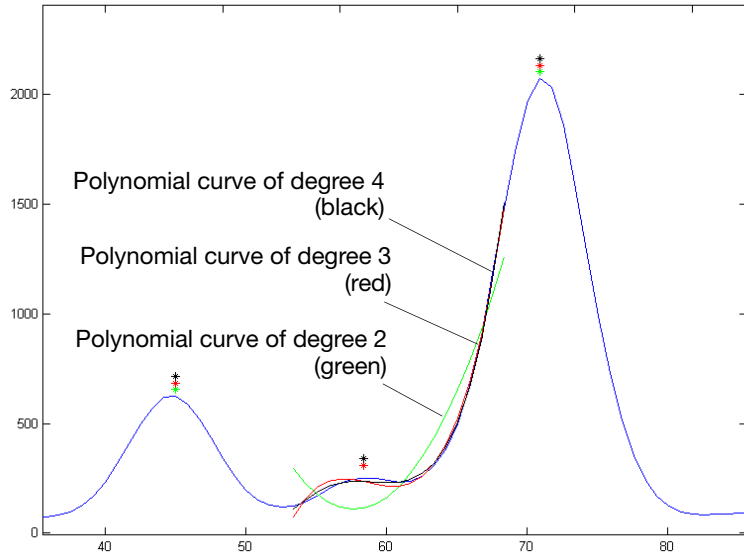
Figure 3   Electropherogram showing peaks detected with three different polynomial degrees

Effects of Increasing the Window Size Value

In Figure 4 on page 11 both polynomial curves have a degree of 3 and the window size value was increased from 15 (red) to 31(black) data points.

As the cubic polynomial is stretched to fit the data in the larger window size, the polynomial curve becomes smoother. Note that the structure of the smaller trailing peak is no longer detected as a distinct peak from the adjacent larger peak to the right.

Window size value of 31 (black)

Window size value of 15 (red)
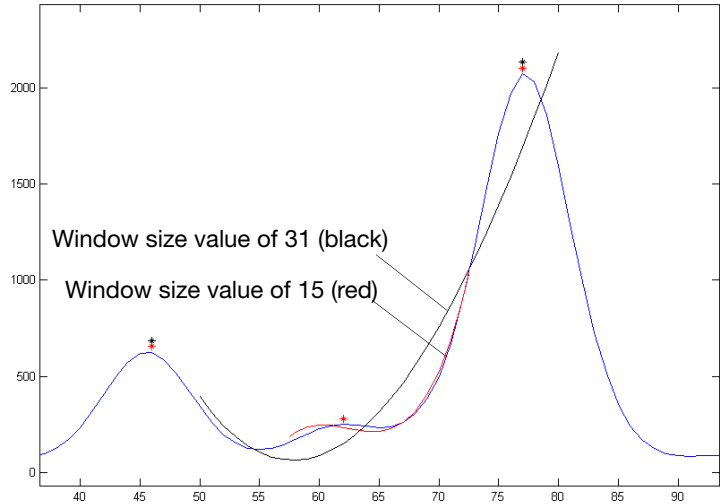
Figure 4   Electropherogram showing the same peaks as in Figure 3 after increasing the window size value but keeping the polynomial degree the same

Optimizing Peak Detection Sensitivity – Example 1: Reducing Peak Window Size

Peak Detection:
- Baseline Window Size
- Smoothing
- **Polynomial Degree and Peak Window Size**
- Slope Threshold Peak Start and Peak End

Figure 5 on page 12 shows the initial electropherogram with two resolved alleles of known fragment lengths (they differ by one nucleotide) that are detected as a single peak. The analysis was performed using a polynomial degree of 3 and a peak window size of 19 data points.
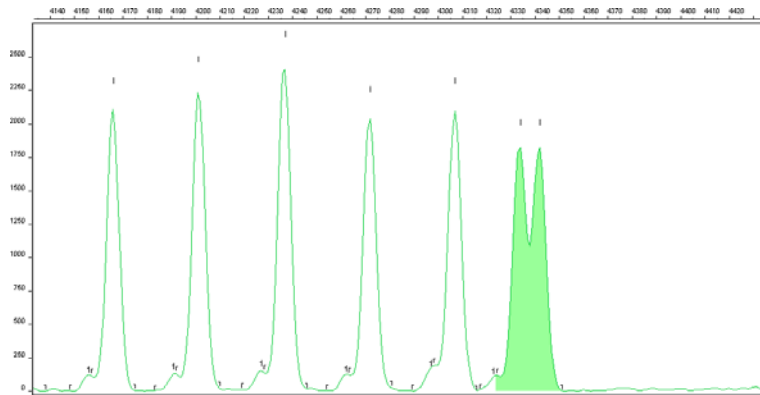
Figure 5   Electropherogram showing two resolved alleles detected as a single peak

Effects of Reducing Window Size

Figure 6 shows that both alleles are detected after reanalyzing with the polynomial degree set to 3 and the window size value decreased to 15 (from 19) data points.
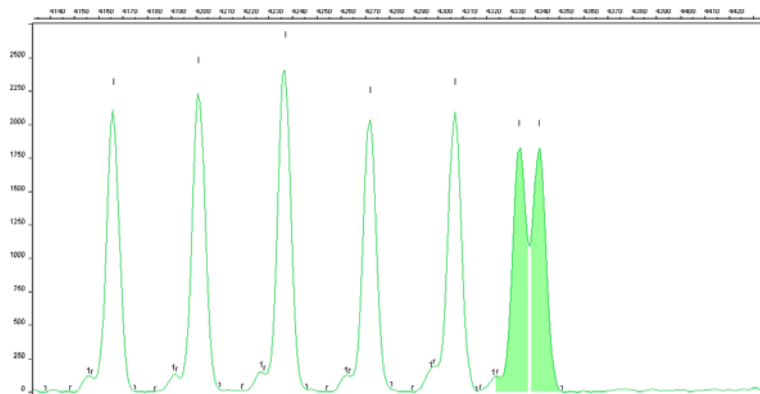
Figure 6   Electropherogram showing the alleles detected as two peaks after decreasing the window size value

1

Optimizing Peak Detection Sensitivity – Example 2: Extreme Settings

Peak Detection:
- Baseline Window Size
- Smoothing
- **Polynomial Degree and Peak Window Size**
- Slope Threshold Peak Start and Peak End

Figure 7 shows the result of an analysis using a peak window size value set to 10 and a polynomial degree set to 9. These extreme settings for peak detection caused several peaks to be split and detected as two separate peaks.
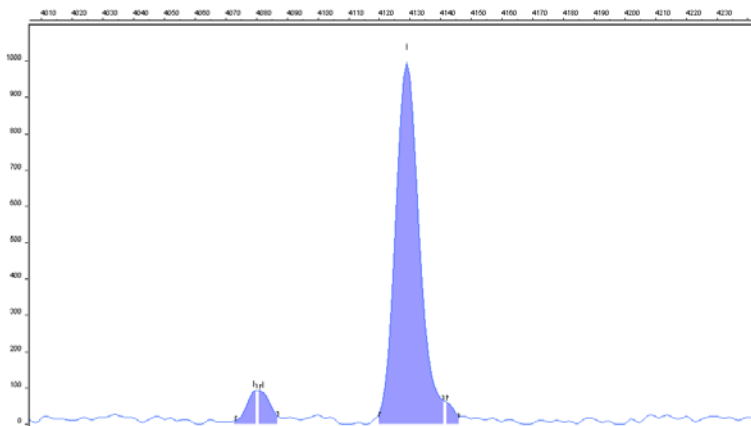
Figure 7   Electropherogram showing the result of an analysis using extreme setting for peak detection

Slope Thresholds for Peak Start/End Parameters

Peak Detection:
- Baseline Window Size
- Smoothing
- Polynomial Degree and Peak Window Size
- **Slope Threshold Peak Start and Peak End**

The Slope Threshold for Peak Start and Slope Threshold for Peak End parameters adjust the start and end points of a peak.

The values assigned to these parameters can be used to better position the start and end points of an asymmetrical peak, or a poorly resolved shouldering peak to more accurately reflect the peak position and area.

In general, from left to right, the slope of a peak increases from the baseline up to the apex. From the apex down to the baseline, the slope decreases negatively until it returns to zero at the baseline (Figure 8 on page 14).
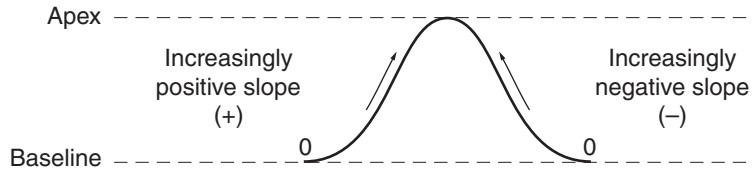
Figure 8   Peak slope

If either of the slope values you enter exceeds the slope of the peak being detected, the software overrides your value and reverts to zero.

Guidelines for Using These Parameters

- For typical or symmetrical peaks, use a value of zero.
- For asymmetrical peaks, select values other than zero to better reflect the beginning and end points.
- A value of zero does not affect the sizing accuracy or precision of an asymmetrical peak.

Using These Parameters

Note: The size of a detected peak is the calculated apex between the start and end points of a peak. Peak size does not change based on start and end settings.

To move the…	Then…	Example
Start point of a peak closer to its apex	Change the Slope Threshold for Peak Start value from zero to a positive number.	
End point of a peak closer to its apex	Change the Slope Threshold for Peak End value to a more negative number.	

Slope Threshold Example – Asymmetrical Peak

Peak Detection:
- Baseline Window
- Smoothing
- Polynomial Degree and Peak Window Size
- **Slope Threshold Peak Start and Peak End**

Initial Electropherogram

Figure 9 shows the initial electropherogram analyzed with value of 0 for Peak Start and Peak End. Note the asymmetrical peak with a noticeable tail on the right side.
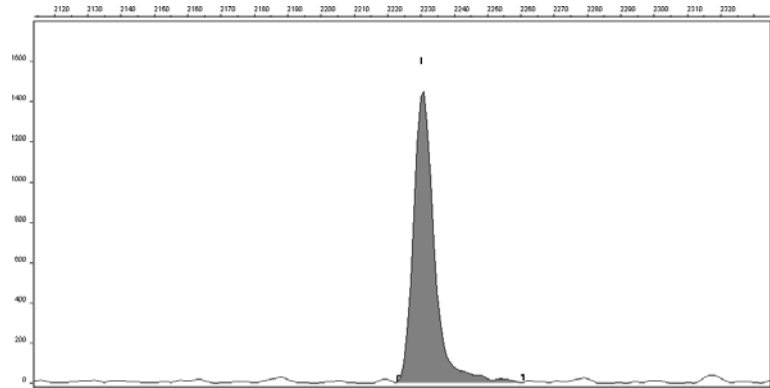
Figure 9   Electropherogram showing an asymmetrical peak

Adjusting Slope Threshold for Peak End

After reanalysis with a value of –35.0 for the Slope Threshold for Peak End, the end point that defines the peak moves closer to its apex, thereby removing the tail (Figure 10). Note that the only change to tabular data is the area (peak size and height are unchanged).
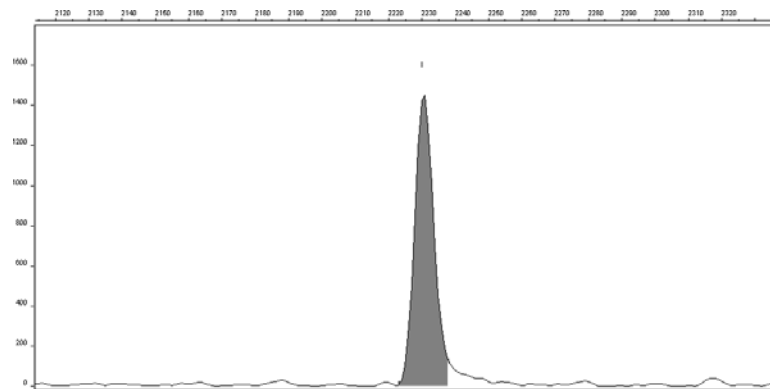
Figure 10   Electropherogram showing the effect of changing the slope threshold for peak end

Sizing

Sizing:
- Size matching
- Size-calling curve generation
- Size calling
- Sizing quality assessment

During sizing, the software:

- Performs size matching of the internal size standard in all samples against the size standard definition selected in the software.
- Generates a size-calling curve.
- Sizes DNA sample peaks.
- Assesses the sizing quality.

Size Matching

Sizing:
- **Size matching**
- Size-calling curve generation
- Size calling
- Sizing quality assessment

During size matching, the software matches the size standard fragments from the electropherogram to the list of fragment sizes in the size standard definition specified in the software.

Size matching uses ratio matching, based on relative height and distance of neighboring peaks. It then derives quality values statistically by examining the similarity between the theoretical (from the size standard definition) and actual (observed) fragment patterns.

The software ignores anomalous peaks that do not match the expected patterns. The software constructs a best-fit curve using the data points of each size standard fragment detected. A comparison between the sizes calculated from the best-fit curve and the matched peaks from the size standard definition file using the array of numbers is performed. Size-matching (and subsequent size calling) fails if significant differences in peak patterns are found, if no match can be made based on the expected patterns, or if all peaks are not found.

Because the software uses ratio-matching (looks for the expected number of alleles and expected peak patterns instead of specific data points), it is not necessary to define new size-standard definitions due to migration shifts.

Figure 11 on page 17 illustrates the size matching and size-calling curve generation algorithm.

Determines the expected peak spacing and height ratios
Uses the list of sizes from the Size Standard definition file.
Note: The values that are used are for example only and do not indicate typical size-standard values.

Evaluates peaks in the size standard data
Ignores peaks that do not meet expected pattern (dotted peak).

Plots the Size-calling curve
Uses peaks that meet expected pattern

Figure 11   Size matching and size-calling curve generation

Size-Calling Curve Generation and Size Calling

Sizing:
- Size matching
- **Size-calling curve generation**
- **Size calling**
- Sizing quality

To generate the size-calling curve, the software plots the actual data points of the size standard against the expected size of each size standard peak. The size-calling method determines how the size-calling curve is generated and used to size each sample.

During size matching and size calling:

- Two size-calling curves are generated for each sample:
  - Black: A best-fit second order curve, regardless of the size-calling method that is selected.
  - Red: A curve based on the size-calling method that is selected in the analysis method.
- The data points of non-size-standard peaks are plotted against the size-calling curve
- Peaks are sized according to the size-calling method that is selected in the analysis method

Size-calling methods are:

- Local Southern
- Global Southern
- Least Squares (2nd Order or 3rd Order)
- Cubic Spline Interpolation

1
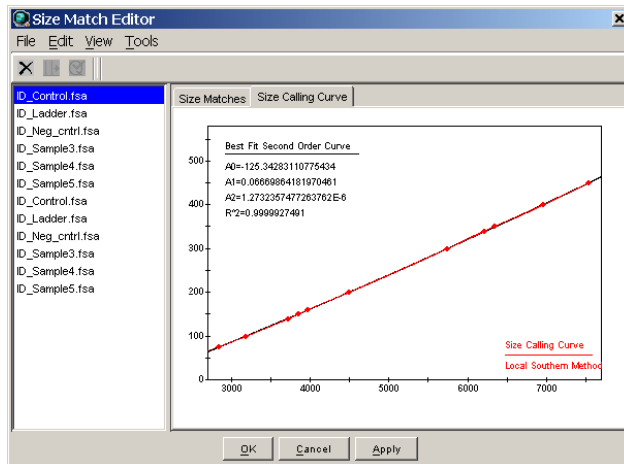
Size Calling – Local Southern Method

Sizing:
- Size matching
- Size-calling curve generation
- **Size calling**
- Sizing quality assessment

The Local Southern method (Figure 12) determines the sizes of fragments by using the reciprocal relationship between fragment length and mobility, as described by E. M. Southern (1979).

Best Fit Second Order Curve

A0=-125.53167570444131
A1=0.06876961532950268
A2=1.268935001923412E-6
R^2=0.9999928907588569

Size Calling Curve

Local Southern Method
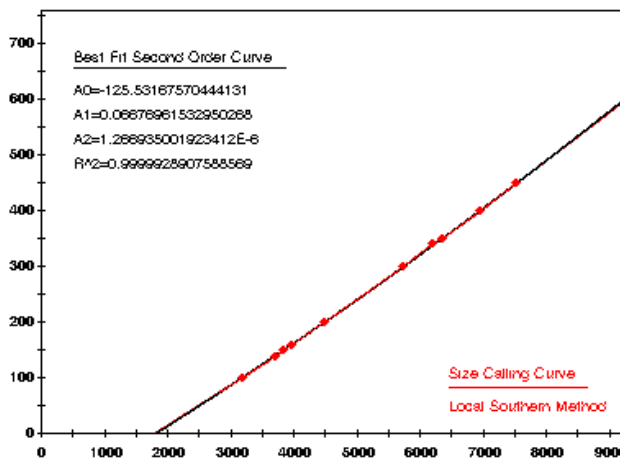
Figure 12   Local Southern method

IMPORTANT! Do not use the Local Southern size-calling method for AmpFℓSTR® MiniFiler™ PCR Amplification Kit data. The Local Southern method requires two data points below the smallest fragment for accurate sizing. Because of the small size of amplicons generated using the MiniFiler™ kit, the GeneScan™500 LIZ® Size Standard 50 bp peak may be hidden by the primer peak, and allele 5 in the D16S539 marker can size at <75 bp. Use the 3rd Order Least Squares size-calling method described on page 22.

Local Southern Method Equation

The equation attempts to describe the reciprocal relationship between the mobility, m, and the length, L0, of the standard fragments.

$$L = [c/(m - m0)] + L0$$

How The Local Southern Method Works

This method, which is similar to the Cubic Spline method (page 24), uses the four fragments closest in size to the unknown fragment to determine a best-fit line value. Only the region of the size standard near the fragment of unknown length is analyzed.

Note: Size estimates may be inaccurate if any of the standard fragments run anomalously.

In the Local Southern method (Figure 13):

1. The fitting constants of the curve are calculated for each group of three neighboring points on the standard.

2. A curve is then created using three standard points (two points below and one point above the fragment), then a fragment size is determined.

3. Another curve is created using an additional set of three points (one point below and two points above the fragment), to assign another value.

4. The two curves are averaged to determine the unknown fragment length.
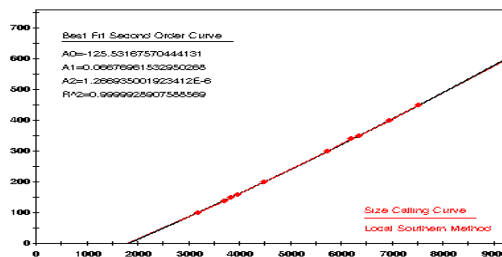
To size call fragments between b and c:
- Fit curve through a, b, and c
- Fit curve through b, c, and d
- Average curves between b and c

Figure 13   Local Southern method

Size Calling – Global Southern Method

Sizing:
- Size matching
- Size-calling curve generation
- **Size calling**
- Sizing quality assessment

The Global Southern method (Figure 14) is similar to the Least Squares method (page 22) in that it compensates for standard fragments that may run anomalously. The method creates a best-fit line through all the available points, and then uses values on that line to calculate the fragment values.
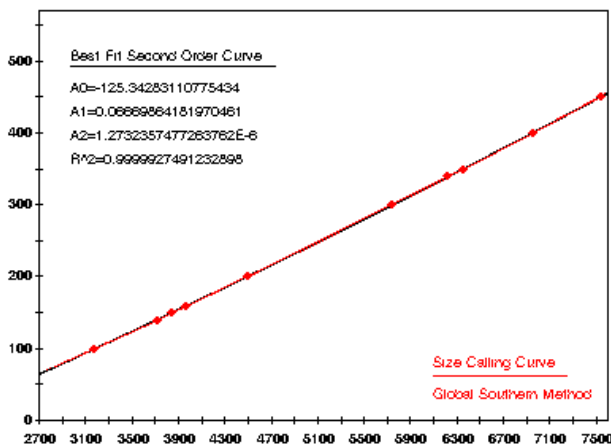
Figure 14   Global Southern method

Global Southern Method Equations

Equation	Description
$L = [c/(m - m0)] + L0$	Attempts to describe the reciprocal relationship between the mobility, m, and the length, L0, of the standard fragments.
$\sum i\{Li - [c/((mi - m0) + L0)]\}^2$	The fitting constants L0, m0, and c are calculated by a least-squares fit to minimize the left side quantity.

How the Global Southern Method Works

All points in the standard are weighted equally, and the curve is not constrained to go through any specific point. The software can analyze a large range of fragment sizes with this method. For best results, use a standard that brackets all the fragments of interest.

Size Calling – Least Squares Method

Sizing:
- Size matching
- Size-calling curve generation
- **Size calling**
- Sizing quality assessment

Both Least Squares methods (2nd-Order and 3rd-Order) use regression analysis to build a best-fit size-calling curve. This curve:

- Produces the minimum additive distance from the curve to the plotted data points.
- Compensates for any fragments that may run anomalously.

Consequently, this method typically results in the least amount of deviation for all the fragments, including the size standards and the samples.

Depending on whether you choose the 2nd- or 3rd-Order Least Squares Method in the Analysis Parameters dialog box, the resulting size curve is either a quadratic or a cubic function. The software uses the known standard fragments and the associated data points to produce a sizing curve based on Multiple Linear Regression.

Using 3rd Order for MiniFiler™ Kit Data

A 3rd-order curve (which uses a higher polynomial degree and captures more of the peak structure) provides the most flexibility when generating best-fit curves for sizing samples with anomalously migrating fragments.

Examples

Figure 15 on page 23 and Figure 16 on page 24 show that in nearly all instances the mobility of an individual DNA fragment is coincident with the best curve fit of the entire data set. Stated differently, the mobility of most DNA fragments is strictly length dependent. This method automatically compensates for fragments that run anomalously.

To generate the Least Squares curve, the software:

- Plots the known fragment sizes (bp) versus data points.
- Generates a best-fit curve using regression analysis.

- Applies the following calculation to determine the size in data points of the unknown fragments:

$$Y = Ax^3 + Bx^2 + Cx + D$$

where:

  Y = size (bp)

  x = datapoint

  A = First order coefficient

  B = Second order coefficient

  C = Third order coefficient

  D = Zeroth Coefficient or constant
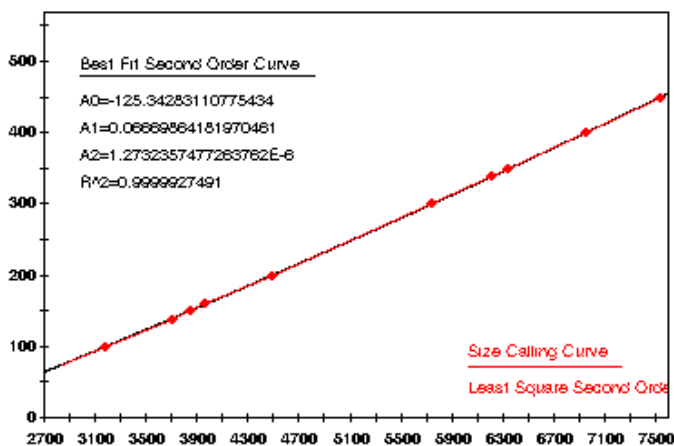
Figure 15   2nd-Order Least Squares size-calling curve (quadratic)
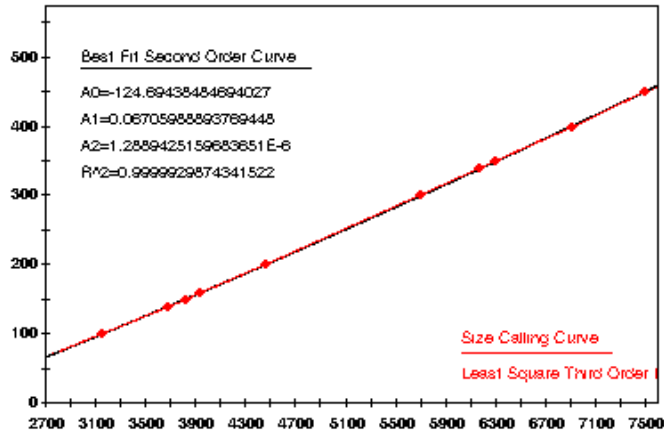
Figure 16   3rd-Order Least Squares size-calling curve (cubic)

Size Calling – Cubic Spline Interpolation Method

Sizing:
- Size matching
- Size-calling curve generation
- **Size calling**
- Sizing quality assessment

The Cubic Spline method (Figure 17), which is similar to the Local Southern method (page 19), forces the sizing curve through all the known points of the selected size standard. Although this enforcement produces exact results for the values of the standards themselves, it does not compensate for standard fragments that may run anomalously.
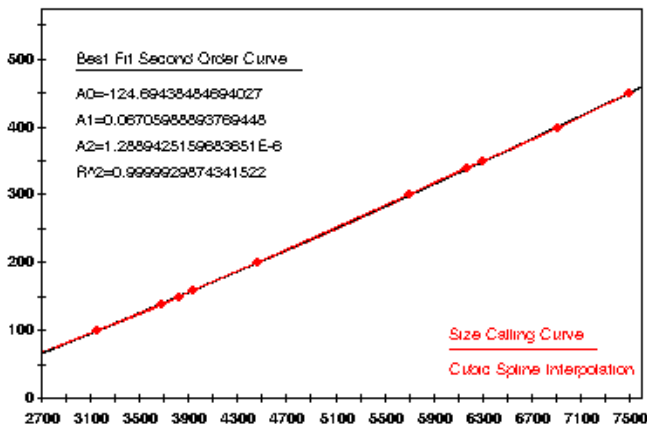
Figure 17   Cubic Spline Interpolation method

Possible Local Sizing Inaccuracy

Mobility of any DNA fragment can be affected by its sequence, and by secondary and tertiary structure formation. If any internal size standard fragment has anomalous mobility, the Cubic Spline method may exhibit local sizing inaccuracy.

For example, assume that a standard fragment is close in molecular length to an unknown sample fragment. Assume further that the standard fragment runs anomalously. The Cubic Spline method assigns the official value to this standard fragment, even though it may be slightly incorrect. The size of the unknown fragment is then likely to be calculated incorrectly as well.

Note: This method does not determine the degree of sizing accuracy error.

Sizing Quality Assessment

Sizing:
- Size matching
- Size-calling curve generation
- Size calling
- **Sizing quality assessment**

The last step in sizing is the sizing quality assessment. Sizing quality is discussed in "Sizing Quality Assessment" on page 43.

The software does not genotype samples with ⬤ sizing quality.

Allelic Ladder Genotyping

Allelic ladder genotyping:
- Allele matching
- Allelic ladder quality assessment
- Bin offsetting

After sizing quality is determined, the following occur:

- Allele matching
- Allelic ladder quality assessment
- Bin offsetting

Allele Matching

Allelic ladder genotyping:
- **Allele matching**
- Allelic ladder quality assessment
- Bin offsetting

During allele matching, the software assigns allele labels to allelic ladder peaks by matching peaks with the bins that are specified in the panel. The matching algorithm uses ratio-matching, based on relative distance and height of neighboring peaks.

Figure 18 is an example of the blue dye plot in the AmpF*l*STR® Identifiler® PCR Amplification Kit allelic ladder.
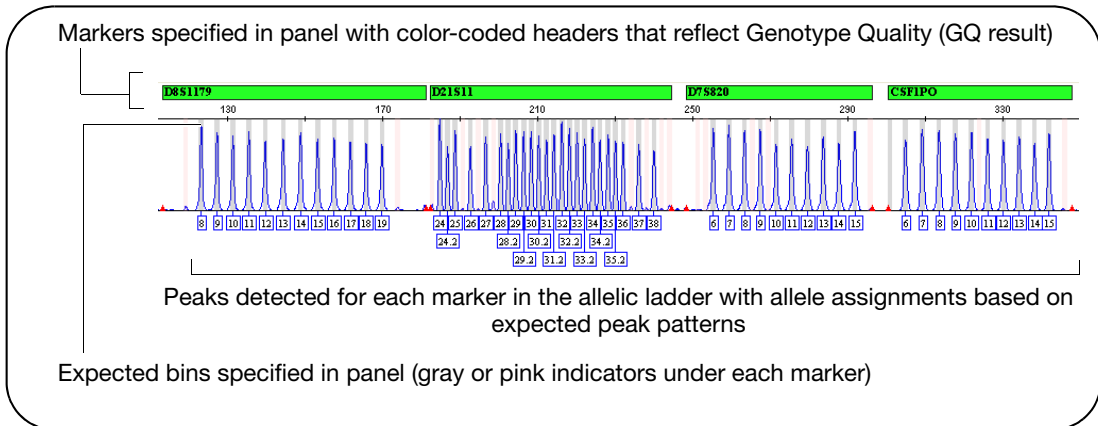
Figure 18   Identifiler® kit allelic ladder blue dye

Allelic Ladder Quality Assessment

Allelic ladder
genotyping:
- Allele matching
- **Allelic ladder quality assessment**
- Bin offsetting

After allele matching, allelic ladders are evaluated to determine whether they are used to create bin offsets, described on page 27.

Allelic ladder quality assessment is discussed in "Allelic Ladder Quality Assessment" on page 45.

Bin Offsetting

Allelic ladder
genotyping:
- Allele matching
- Allelic ladder quality assessment
- **Bin offsetting**

A bin is a region that defines an allele within a marker. Bin offsets are the size difference (in base pairs) between the reference bin locations in the panel and the actual sizes of the allelic ladder fragments. Bin offsets are automatically calculated during analysis of the allelic ladder samples. The bin offsets determine how much and in which direction (higher or lower) the reference bins shift. The offset bins are then used to genotype the other sample types.

Bin offsetting evaluates the alleles found in allelic ladder samples and adjusts bins before allele-calling. Marker ranges are also adjusted based on the bin offsets.

AmpFℓSTR® Allelic Ladder and Bin Sets

The allelic ladder for each AmpFℓSTR® kit contains the most common alleles for each locus and is used to generate the reference sizes in the bin sets provided with the software.

Reference sizes consider:

- The actual size (bp) of DNA fragments.
- 3´ A nucleotide addition.
- Mobility shifts related to the non-nucleotide linkers in the allelic ladder. Non-nucleotide linkers enable reproducible positioning of the alleles to facilitate inter-locus spacing. (Grossman *et al*., 1994 and Baron *et al*., 1996)

Physical and Virtual Bins

The bin set provides reference allele sizes for:

• Alleles physically present in the allelic ladder (physical bins).

• Alleles that are not present in the allelic ladder (virtual bins), but have either been reported in STRbase (www.cstl.nist.gov/div831/strbase) or discovered during developmental validation of a particular chemistry, and can be genotyped by reference to the alleles present in the ladder.

Figure 19 illustrates a bin set with eight physical bins and six virtual bins.
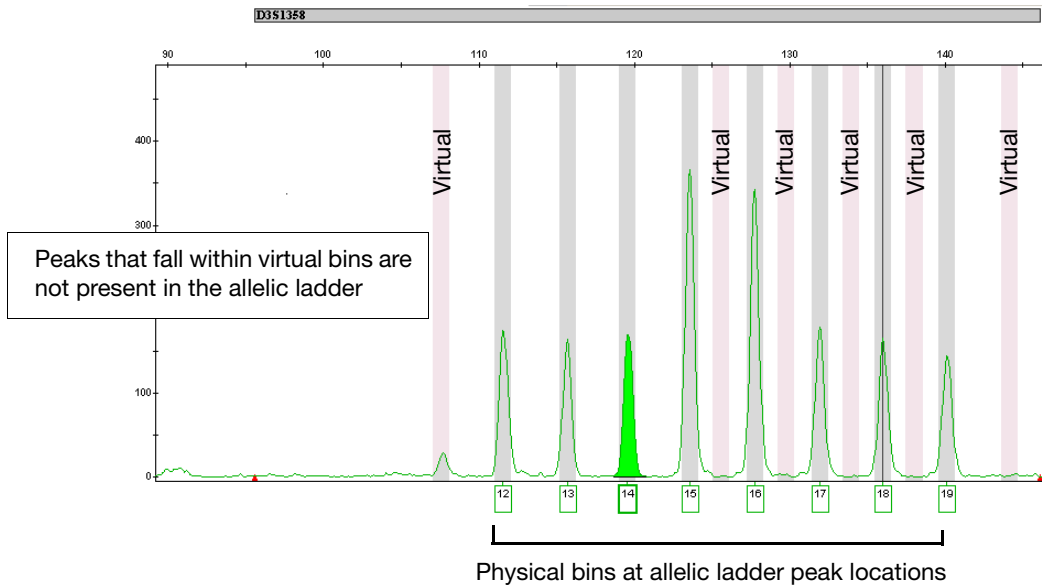
Figure 19   Bin set with physical and virtual bins

Note: The GeneMapper® *ID-X* Software allows you to display virtual bins in pink. For information, see the GeneMapper® *ID-X* Software Help.

GeneMapper® *ID-X* Software Version 1.5 Reference Guide

How the Software Calculates and Applies Bin Offsets

When analyzing allelic ladder samples, the GeneMapper® *ID-X* Software offsets the reference sizes defined in the bin set relative to the actual sizes obtained for the alleles in the allelic ladder.

To determine offsets, the software:

1. Performs an allelic ladder quality assessment (described on page 45).

2. Uses ▮ CGQ allelic ladders with ▮ SQ only.

   CGQ is described on page 58. SQ is described on page 43.

3. Identifies the size of a physical allele in the allelic ladder. Virtual alleles are not used to calculate offsets.

   If more than one allelic ladder sample is present in the run folder, determines the offset for each allele in each of the allelic ladder samples, then averages the offsets for the allele.

4. Offsets the corresponding bins for the physical alleles by the difference between the observed size for the allele in the run data and the reference size for the allele.

5. Offsets the virtual bins. To offset the virtual bins, the software uses the offset from the closest physical bin or virtual bin to the left of the bin in question. If the first bin is a virtual bin, the software uses the offset from the first physical bin (counting from left to right).

6. Offsets the:

- Marker Range Min Size value by the calculated bin offset of the first physical bin of the allelic ladder in the marker.
- Marker Range Max Size value by the calculated bin offset of the last physical bin of the allelic ladder in the marker.

For more information, see "Displaying Offset Marker Margins" on page 33.

IMPORTANT! The software displays the *reference* marker range in two places: in the Samples plot status bar when you click the marker header above the plot and in the Panel Manager. To display the *offset* marker ranges, examine the value that is displayed in the Samples plot status bar when you click the red marker indicators on the x-axis of the Samples plot.

---

7. Repeats the process for each physical allele in all samples with a Sample Type of Allelic Ladder.

Example: One Allelic Ladder Is Present In a Run Folder

A reference size in the bin set for an allele is 122 bp. The observed size is 119.56 bp. The offset value is calculated as $119.56 - 122 = -2.44$. The bin for this allele is shifted down 2.44 base pairs to 119.56 ($\pm$ 0.5 bp).

Figure 20 on page 31 shows the reference size (bin) for allele 14 in the D3S1358 marker before analysis.
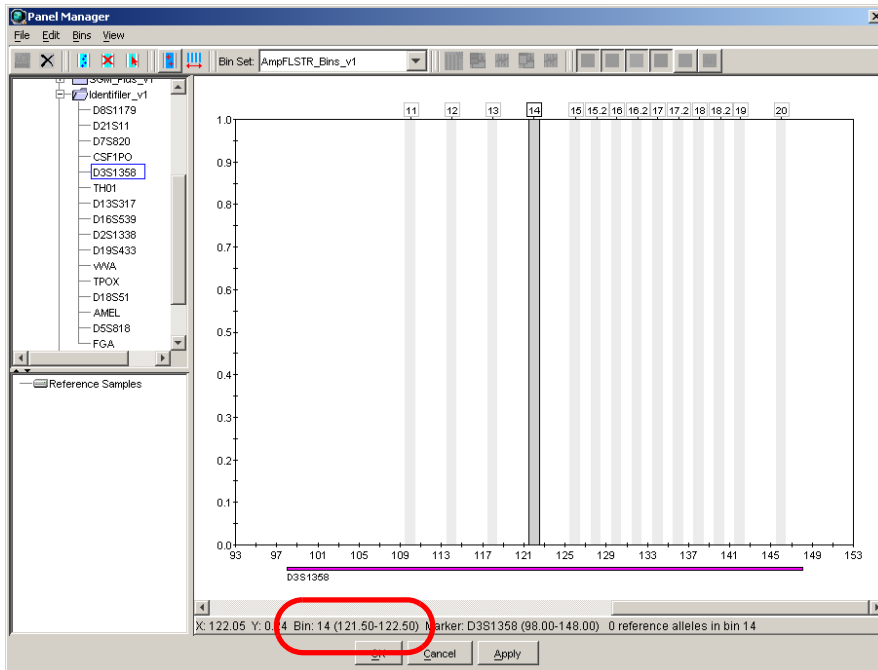
Reference allele size (position of the bin) 122 ± 0.5 bp

Figure 20   Bin assignment before offsetting

Figure 21 on page 32 shows that after analysis of the allelic ladder sample, the bin is offset based on the observed size of the allele in the allelic ladder.
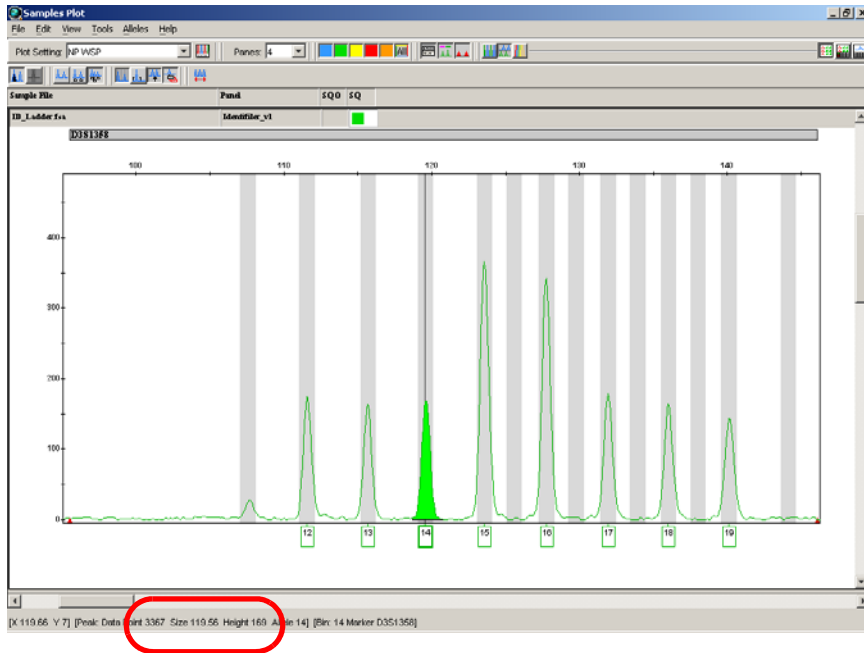
Offset bin 119 ± 0.5 bp

Figure 21   Bin assignment after offsetting

Example: More Than One Allelic Ladder Sample Is Present In a Run Folder

A reference size in the bin set for an allele is 122 bp. The observed size is 119.56 bp for one allelic ladder sample and 119.50 bp for another allelic ladder sample. The offset for the each allelic ladder sample is calculated (119.56 – 122 = –2.44) and (119.50 – 122 = –2.50). The software averages the two offsets [(–2.44 + –2.50)/2 = –2.47]. The bin for this allele is shifted down 2.47 base pairs to 119.53 (± 0.5 bp).

Example: How Offsets Are Applied To Virtual Alleles

Virtual allele 11 in Figure 21 above uses the same offset value as allele 12. The sizes between these two alleles differ by 4 bp. Specifying a size for allele 11 that is 4 bp shorter than allele 12 is a reasonable estimate because alleles 11 and 12 differ by a single repeat unit (4 bp).

Displaying Offset Marker Margins

When bins are offset, marker margins are also offset. To display the shifted marker margins, click the red marker indicators in the Samples plot (Figure 22).

Click the red tick mark at marker start to display the offset marker start in the status bar
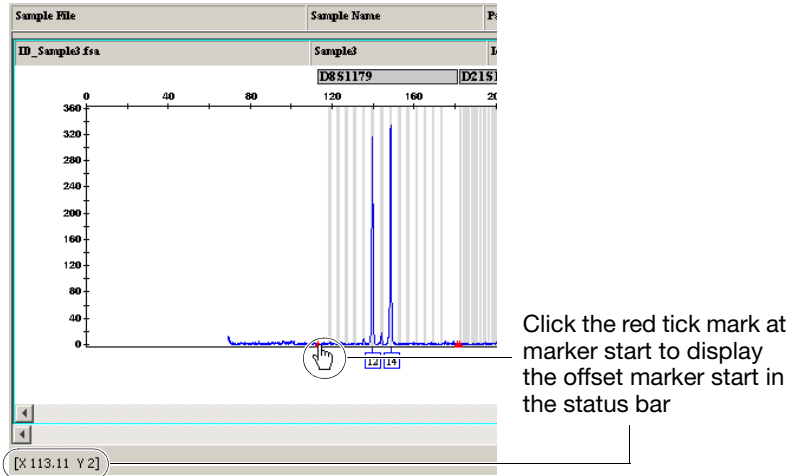
Figure 22   Displaying offset marker margins

Note: Reference marker margins that are defined in the panel are displayed when you click the marker header in the Samples plot and in the Panel Manager (Figure 23).

Original marker margins (not shifted)
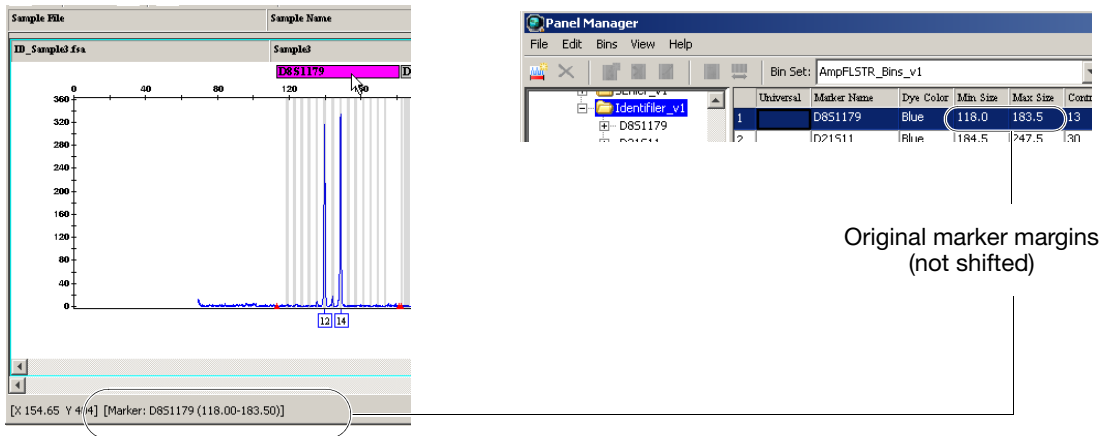
Figure 23   Original marker margins are displayed when you click the marker header in the Samples plot and in the Panel Manager

Sample Genotyping

Sample genotyping:
- Allele calling and filtering
- Genotype and sample and quality assessment

After allelic ladders are detected and sized, sizing quality is determined, then allelic ladders are genotyped.

After allelic ladders are genotyped, samples are detected and sized, sizing quality is determined, then samples are genotyped.

During genotyping of samples, the GeneMapper® *ID-X* Software:

- Performs allele-calling and filtering.
- Assesses genotype quality.
- Assesses sample quality.

Allele Calling and Filtering of Samples

Sample genotyping:
- **Allele calling and filtering**
- Genotype and sample and quality assessment

During allele calling of samples, the software:

- Labels allele peaks in unknown samples:
  - Compares sizes of the unknown peaks to the offset bins.
  - Assigns the relevant allele label if a peak falls within one of the defined bins. Figure 24 on page 35 is an example of a 14 allele in D3S1358.
  - Assigns an OL (Off-Ladder) label if a peak falls outside of the defined bins or within a bin overlap region (see "Bin Overlap" on page 61).
- Applies filters to eliminate peaks from consideration as allele peaks:
  - Allele Tab analysis method settings (Global Cut-off value, Minus A, Global Stutter, Amelogenin Cutoff, and Range Filter)
  - Marker-specific stutter ratios if specified in the Analysis Method (Allele tab) and defined in the panel
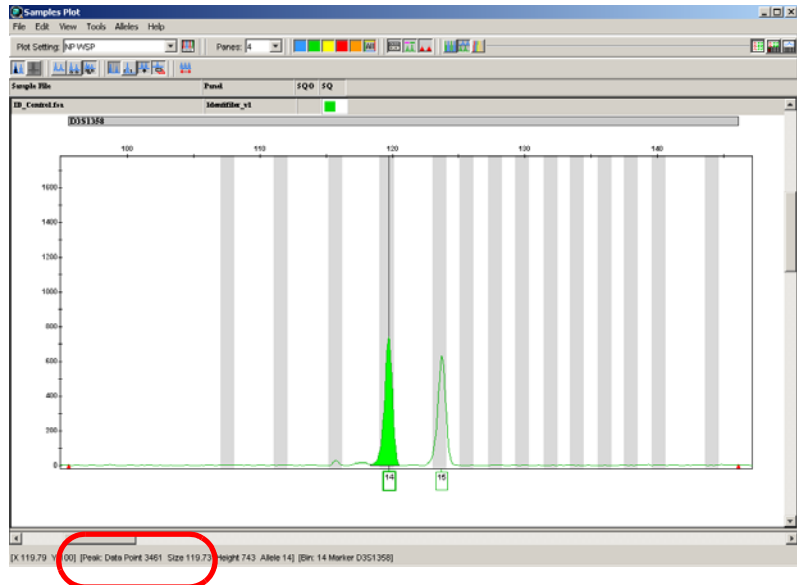
Reference allele size (position of the bin) 119 ± 0.5 bp

Figure 24   Allele calling

Notice the size of this allele is 119.73, which is within the shifted bin calculated previously (119.53 ± 0.5 bp, see Figure 21 on page 32). Therefore, this peak is labeled as a 14.

Off Ladder (OL) Allele Calls

If a peak falls outside one of the defined bins or within a bin overlap region, the software labels it OL (Off Ladder).

OL calls may be caused by:

- The presence of a microvariant allele at a particular locus
- Sample migration anomalies and artifacts
- A high density of allelic ladder bins that are less than 0.1nt apart after bin offsetting, which causes bin overlap (see "Bin Overlap" on page 61)

Allele Naming Convention

Allele names are assigned based on the number of complete four-base-pair repeat-units and partial repeat-units:

- Complete four base pair repeat units are designated by an integer.

- Variant alleles that contain a partial repeat are designated by a decimal followed by the number of bases in the partial repeat. For example:
  – An FGA 26.2 allele contains 26 complete repeat units and a partial repeat unit of two base pairs.
  – A D21S11 allele 33.1 contains 33 complete repeat units and one nonconsensus base pair.
  – A D21S11 29.3 allele contains 29 complete repeat units and a partial 3-bp unit (Moller et al., 1995; Gill et al., 1997).

Genotype and Sample Quality Assessment

Sample genotyping:
- Allele calling and filtering
- **Genotype and sample and quality assessment**

To determine sample quality, the software:

- Applies user-defined and software-defined thresholds to generate marker-level quality value flags (described in "Marker-Level Quality Assessment" on page 47) and sample-level quality value flags (described in "Sample-Level Quality Assessments" on page 55).
- Applies GQ weighting from ⚠ marker-level quality value flags to determine the Genotype quality (GQ) (described in "Sample-Level Quality Assessments" on page 55).
- Uses GQ to determine the sample-level CGQ composite genotyping quality assessment (described in "CGQ (samples)" on page 58).

2

GeneMapper® *ID-X* Software Quality Value System

This chapter covers:

2

Overview of the Quality Value System

The GeneMapper® *ID-X* Software provides a quality value system and a set of streamlined data review tools and features that can be optimized and validated for use in both expert systems and traditional data analysis workflows.

The GeneMapper® *ID-X* Software quality value system:

- Assesses the quality of allelic ladders before analysis and does not consider low-quality allelic ladders for genotyping.
- Assesses the quality of data at the sample and marker level using process quality values (PQVs).
- Can be used in an optimized and validated expert system or traditional data analysis workflow to quickly identify data quality issues and aid in interpretation of samples that do not meet all thresholds.
- Can be used in an optimized and validated expert system workflow to quickly segregate samples that require manual review from those that do not.

The PQV results of the quality assessment are displayed as color-coded flags: . The color of the flag depends on software-specified thresholds or user-defined thresholds set in the analysis method.

Optimizing and Validating an Expert System

Before using any software as an expert system, optimize and validate the thresholds for each AmpFℓSTR kit and instrument platform combination by processing a variety of samples that challenge each of the quality flags.

IMPORTANT! Different kit/instrument combinations may require different thresholds.

**Displaying
Numeric SQ and
GQ Values**

By default, the GeneMapper *ID-X* Software displays Pass , Check- , or Low Quality for Sizing Quality (SQ) and Genotyping Quality (GQ) to represent the numeric score of the assessment.

For example, the Sizing Quality (SQ) yields a value between 1 and 0. Based on the SQ threshold settings of the SQ & GQ Settings tab, the software displays Pass , Check- , or Low Quality to indicate the SQ result.

You can alternatively display numeric values for SQ and GQ.

Sizing Quality Representation	Samples Table Example
Symbols (default)	
Numbers	

To display numeric values for SQ and GQ:

1. Select **File ▸ Project Options**, then select the **Analysis** tab.

2. In the Quality Metrics Display settings, select **Numbers**.
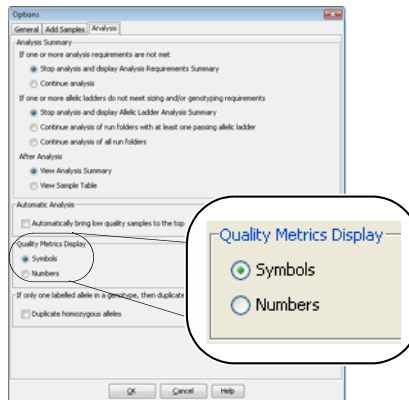
3. Click **OK** to apply the settings.

2

Chapter 2 GeneMapper® ID-X Software Quality Value System
Overview of the Quality Value System Checks and Assessments

Overview of the Quality Value System Checks and Assessments

The GeneMapper *ID-X* Software quality value system performs the following checks and assessments:

- **Analysis requirements checks:**

   Before analysis starts, identifies any conditions that may prevent analysis or cause unexpected results. If one or more analysis requirements are not met, the ARNM (Analysis Requirements Not Met) PQV is set to ⚠. If Project Options are set accordingly, the Analysis Requirements Summary is also displayed (described in GeneMapper® *ID-X* Software Help).

- **Allelic ladder genotyping**:
   - **Sizing quality assessment** – Evaluates the quality of the size standard profile in each allelic ladder.
   - **Allelic ladder quality assessment** – Evaluates allelic ladder quality. Also determines if an allelic ladder is used for creating bin offsets. If the software finds one or more allelic ladders that do not meet sizing and genotyping requirements and Project Options are set accordingly, the Allelic Ladder Analysis Summary is displayed (described in GeneMapper® *ID-X* Software Help).
   - **Marker-level quality assessment** – Evaluates labeled peaks within each marker. The SPK and OS PQVs contribute to the allelic ladder GQ assessment if spikes and/or off-scale fluorescence signals are present in the allelic ladder (and Spike Detection is Enabled for the Allelic Ladder and the Allelic Ladder GQ Weighting for Spikes and Off-scale is set to 1[on]. If Spike Detection is Disabled for the Allelic Ladder, spikes will not be detected or labeled.) Other PQVs do not affect allelic ladder GQ.
   - **Sample-level quality assessment** – Evaluates the quality of the entire sample.

- **Sample genotyping**:
   - **Sizing quality assessment** – Evaluates the quality of the size standard profile in each sample.

– **Marker-level quality assessment** – Evaluates labeled peaks within each marker. Contributes to the genotype quality assessment.

– **Sample-level quality assessment** – Evaluates the quality of the entire sample.

• **Genotype quality assessment** – Evaluates the quality of each marker in a sample. Contributes to the overall composite genotype quality assessment.

Analysis Requirement Checks

The analysis requirements checks are performed before data analysis starts. If one or more analysis requirement checks is ⚠, analysis may not occur or may not give the expected results (for example, analyzing without a size standard or allelic ladder affects sizing quality or genotyping quality).

Analysis requirements checks that are displayed in the Samples table before the analysis starts

Acronym	Full Name	Function/Rule(s)	Flag	
SFNF	Sample File Not Found	Indicates if the software finds the sample (.fsa or .hid) files in the original location from which they were added to the project. Select **Edit ▸ Define New Sample File Path** to re-associate the samples file with the project.	🟩	Sample file found
			⚠	Sample file not found
MNF	Matrix Not Found	Indicates if the software finds in the database the matrix file that is associated with the sample.	🟩	Matrix file found
			⚠	Matrix file not found
SNF	Size Standard Not Found	Indicates if the software finds in the database the size standard definition associated with the sample.	🟩	Size standard definition found
			⚠	Size standard definition not found

> **Note:** The SFNF, MNF, and SNF flags are included in the ARNM flag described on page 42. To conserve space, display only the ARNM column in the Samples table.

Analysis requirements checks that are displayed in the Analysis Requirements Summary after the analysis starts

Acronym	Full Name	Function/Rule(s)		Flag Indicator
ARNM	Analysis Requirement Not Met	Indicates if all analysis requirements are met. These requirement checks are performed when analysis is started, before actual analysis begins: • Sample File Not Found • Analysis Method Not Selected • Analysis Method Not Found In The Database • Panel Not Selected • Panel Not Found in the Database • Binset Not Selected • Binset In Analysis Method does not match Binset selected in the Panel Manager • Size Standard Not Selected • Size Standard Not Found In Database • Size Standard Dye Color Is Not Present In The Sample Dye Set • Matrix Not Selected • Matrix Not Found or Contains Invalid Data • No Allelic Ladder Selected In Run Folder • GMID V3.*x* Analysis Method Selected • Basic Or Classic Size Standard Selected • SNP Panel Selected **Note:** The ARNM PQV includes the SFNF, MNF, and SNF PQVs described on the previous page. To conserve space, display only the ARNM column in the Samples table.		All analysis requirements are met.
				One or more analysis requirements are not met.

Sizing Quality Assessment

The quality value system evaluates the quality of the size standard profile within each sample (SQ) and allows you to flag size standards with poor peak resolution. You can adjust the ranges that correspond to the Pass ■, Check ▲, and Low Quality ● ranges in the analysis method (SQ & GQ Settings).

Sizing Quality that is displayed in the Samples Table after the analysis is complete

Acronym	Full Name	Function/Rule(s)		Flag Indicator
SQ	Sizing Quality	Evaluates the similarity between the fragment pattern for the size standard dye that is specified in the size standard definition and the actual distribution of size standard peaks in the sample, calculates an interim SQ (a value between 0 and 1) as described in "Sizing" on page 16, then applies the broad peak weighting that is specified in the analysis method. **Weighting** The Broad Peak (BD) threshold and SQ Broad Peak weighting that is specified in the analysis method affect the SQ. To determine the final SQ value, the software: • Evaluates peak widths in the sample in the dye color specified in the size standard definition. • If the width of any peak in the sizing range exceeds the broad peak threshold, applies the SQ weighting: Interim SQ × (1 − SQ weighting)	■	SQ is within the user-defined Passing range (Default = 0.75 to 1.0)
			▲	SQ is between the user-defined Passing range and Low Quality range (Default = 0.26 to 0.74)
			●	SQ is within the user-defined Low Quality range (Default = 0.0 to 0.25)

Sizing Quality that is displayed in the Samples Table after the analysis is complete

Acronym	Full Name	Function/Rule(s)	Flag Indicator	
SQ *(continued)*	Sizing Quality	**Example** A size standard peak exceeds BD threshold, has an Interim SQ = 0.8, and SQ weighting = 0.5. The final SQ is: SQ = $0.8 \times (1 - 0.5) = 0.4$ Based on the default SQ range, this SQ is 🔺. **Note:** Broad Peak (BD) threshold and SQ Broad Peak weighting are user-defined. **Note:** The GeneMapper® *ID-X* Software does not genotype samples with 🛑 SQ.		

Allelic Ladder Quality Assessment

The quality value system performs an allelic ladder quality assessment to determine if a ladder is used in genotyping (to create bin offsets, described on page 27).

Allelic ladder quality is reported per marker by the GQ (Genotyping Quality is described on page 53) and per sample by the CGQ (Genotyping Quality is described on page 58).

Quality Rules Allelic ladder samples are analyzed before all other samples. An allelic ladder sample must have a ■ SQ and a ■ CGQ to be used for creating bin offsets. For an allelic ladder to have a ■ CGQ, all the markers within the allelic ladder must pass the following rules:

Rule	Description
1	All ladder alleles specified in the panel used to analyze are detected.
2	In each marker, the peak height ratio of the first and second peak is greater than 50%. This rule eliminates allelic ladders if the stutter peak before the first true allele peak is labeled as an allele.
3	No spikes are detected above 20% (default) of the highest allele peak in the same dye color within the extended marker range. **Note:** When spike detection is enabled for Allelic Ladders, spike detection is performed within each extended marker range (no gaps are present between markers; the end point of each marker is extended past the marker definition in the panel to the beginning of the next marker). **Note:** When spike detection for allelic ladders is set to Disable, spikes will not be detected or labeled. **Note:** The Allelic Ladder Spike Cut-off value is user-definable in the Peak Quality tab of the analysis method.
4	The peak height ratio between the lowest and highest peak is equal to or greater than 15%.
5	In each marker, the base pair spacing between any two ladder alleles specified in the panel used to analyze is within the expected range.
6	No off-scale (OS) fluorescent signal is detected within each extended marker range. **Note:** The Allelic Ladder GQ Weighting for Off-scale is user-definable.

Effect on GQ   Allelic ladder quality rules are evaluated at the marker level and affect GQ results.

If any of the rules do not pass, the GQ (described on page 53) for the allelic ladder marker is set to ⬤.

Note: If rule 3 does not pass and the Allelic Ladder GQ Weighting for Spikes is set to 0 (off in the SQ & GQ Settings tab in the analysis method), the GQ for the allelic ladder marker is set to ▪, even if spikes are present.

Other than SPK, no other individual marker-level PQV flags (described on page 47) affect the GQ of allelic ladder samples. Therefore, they do not affect the allelic ladders used for genotyping.

Marker-Level Quality Assessment

Marker-level quality assessment indicates the quality of each marker in a sample.

Marker-level quality assessment evaluates only labeled peaks within a defined marker size range defined in the panel and can be weighted to contribute to the calculation of the GQ (Genotype Quality).

Note: Gray PQVs ( ) indicate that at least one label in the marker was edited or the GQ was overridden.

Marker-level quality flags that are displayed in the Genotypes table after the analysis is complete

Acronym	Full Name	Function/Rule(s)		Flag Indicator
OS	Off-scale	• **Allelic Ladders** – Indicates if any off-scale fluorescence signal is detected within each extended marker range (no gaps are present between markers and the end point of each marker is extended past the marker definition in the panel to the beginning of the next marker). • **Samples** – Indicates if any fluorescence signal within a marker size range exceeds the detection threshold of the instrument. The software displays a pink bar above off-scale peaks. **Note:** The detection limit varies on different instruments. The OS quality flag indicates if the CCD camera is saturated at any data point within the analysis range. It is not based on an absolute RFU value (peak height).	🟩	No fluorescence signal exceeds the detection limit of the instrument.
			⚠️	One or more fluorescence signals exceed the detection limit of the instrument.
BIN	Out of Bin Allele	Indicates if labeled peaks do not fall inside bins. These peaks are labeled with OL (off ladder). For information of bin offsetting, see page 27.	🟩	No labeled peaks are detected outside an offset bin.
			⚠️	One or more labeled peaks are detected outside an offset bin.

Marker-level quality flags that are displayed in the Genotypes table after the analysis is complete *(continued)*

Acronym	Full Name	Function/Rule(s)		Flag Indicator
PHR	Peak Height Ratio	Indicates if the peak height ratio between the lowest and highest peak is less than the Min Peak Height Ratio defined in the analysis method. **Note:** The PHR PQV flag used and displayed in the Genotypes table of a GeneMapper *ID-X* project differs from the PHR Status PQV flag used and displayed in the Genotype Combinations tables of the Mixture Analysis tool.	🟩	The peak height ratio calculated between the lowest and highest peaks is equal to or greater than the Min Peak Height Ratio threshold.
			🔺	The peak height ratio calculated between the lowest and highest peaks is less than the Min Peak Height Ratio threshold.
			N/A	Only one peak is labeled.
LPH	Low Peak Height	Indicates if any peak heights (RFU) within a marker size range are below the following thresholds set in the analysis method: • Homozygous Min Peak Height • Heterozygous Min Peak Height	🟩	All detected peak heights are equal to or greater than the thresholds.
			🔺	One or more peak heights are below the thresholds.
MPH	Max Peak Height	Indicates if any peak heights (RFU) exceed the Max Peak Height (RFU) value set in the analysis method.	🟩	No peaks exceed the Max Peak Height (RFU) value (all peak heights are less than or equal to the threshold).
			🔺	One or more peaks exceed the Max Peak Height (RFU) value.
			N/A	No peaks are detected or all peaks are off-scale (OS).

Marker-level quality flags that are displayed in the Genotypes table after the analysis is complete *(continued)*

Acronym	Full Name	Function/Rule(s)	Flag Indicator	
SPK	Marker Spike	• **Allelic Ladders** – Indicates if spikes are detected within each extended marker range (no gaps are present between markers; the end point of each marker is extended past the marker definition in the panel to the beginning of the next marker). • **Samples** – Indicates if spikes are detected within a marker size range. The software uses a proprietary algorithm that detects spikes based on the peak morphology.	🟩	No spikes are detected within the range *or* One or more spikes, but no alleles | (⚠ AN), are detected within the range for Sample Type = Sample.
			⚠	One or more spikes are detected within the range.
			N/A	PQV does not apply to the marker, *or* One or more spikes are detected within the range for Sample Type = Negative Control.
AN	Allele Number	Indicates if the software detects: • More than the Max Expected Alleles for Autosomal/Amelogenin markers set in the analysis method (Peak Quality tab) *or* • More than the Max Expected Alleles are detected for any markers specified as Y markers in the Panel Manager *or* • No alleles are detected for autosomal or amelogenin markers *or* • No X allele in Amelogenin	🟩	At least one allele and not more than the Max Expected Alleles are detected for the marker, or the X allele is detected for the Amelogenin marker.
			⚠	No alleles, more than the Max Expected Alleles, or no X allele for Amelogenin is detected.

Marker-level quality flags that are displayed in the Genotypes table after the analysis is complete *(continued)*

Acronym	Full Name	Function/Rule(s)		Flag Indicator
ACC	Amelogenin Cross Check	For analysis of samples that contain both Y and autosomal STRs. Indicates if Amelogenin and Y marker results are concordant. The same flag is displayed across Amelogenin and all markers with the Y Marker checkbox selected in the Panel Manager.		Sample or positive control contains the amelogenin marker and one or more Y markers, and amelogenin and all Y marker results are concordant.
				Amelogenin and one or more Y marker results are not concordant: • Amelogenin genotype is not X,Y (for example it is X,OL or OL,OL) • Amelogenin genotype is X or X,X but Y markers contain called or OL alleles • Amelogenin genotype is X,Y and one or more Y markers do not contain called or OL alleles

Marker-level quality flags that are displayed in the Genotypes table after the analysis is complete *(continued)*

Acronym	Full Name	Function/Rule(s)	Flag Indicator	
ACC *(continued)*	Amelogenin Cross Check		N/A	• Negative control • Allelic Ladder • Sample or positive control that does not contain the amelogenin marker and at least one Y marker selected in the Panel Manager (for example, samples generated with the AmpFISTR® Yfiler® PCR Amplification Kit, which does not contain amelogenin)
BD	Broad Peak	Indicates if the width of any peak exceeds the Max Peak Width (half height in basepairs) defined in the analysis method (Peak Quality tab). Broad peaks may be a result of dye artifacts or poor resolution. Peaks with poor resolution: • May be caused by electrophoresis issues • May not be true DNA peaks		No peaks exceed the Max Peak Width threshold.
				One or more peaks exceed the Max Peak Width threshold.

Marker-level quality flags that are displayed in the Genotypes table after the analysis is complete *(continued)*

Acronym	Full Name	Function/Rule(s)	Flag Indicator	
CC	Control Concordance	Indicates if a positive, custom, or negative control produces the expected profile. The expected profile is obtained from: • Positive Control – The Control Alleles in the panel used for analysis. • Custom Control – The profile associated with the Profile ID in the Profile Manager. • Negative Control – No peaks expected. Applies only to Sample Type = Positive Control or Negative Control. Weighting of the CC PQV is non-editable and set to 1.0.	🟩	Positive or Custom Control – Expected profile is detected.
				Negative Control – No peaks are detected.
			🔺	Positive or Custom Control – Expected profile is not detected.
				Negative Control – Peaks are detected.
			N/A	Sample Type = Sample or Allelic Ladder
OVL	Overlapping Alleles	Indicates if a labeled peak (allele or artifact) is within the size ranges of two neighboring markers. Applies only if you modify the default marker size ranges to overlap one another.	🟩	No labeled peaks are detected within overlapping marker size ranges.
			🔺	One or more labeled peaks are detected within overlapping marker size ranges.

Genotype Quality Assessment

For samples, the quality value system assigns the genotype quality (GQ) for each marker based on the individual marker quality flags.

For allelic ladders, the quality value system assigns a GQ for each marker based on the allelic ladder quality requirements, as described in "Allelic Ladder Quality Assessment" on page 45.

The GQ is used to determine the CGQ.

For samples, you can adjust the ranges that correspond to the Pass ▮, Check ⚠, and Low Quality ⬟ ranges in the analysis method (SQ & GQ Settings).

Genotype Quality that is displayed in the Genotypes Table after the analysis is complete

Acronym	Full Name	Function/Rule(s)	Flag Indicator	
GQ (samples)	Genotype Quality	Indicates the genotype quality of the marker in the sample. The genotype quality for a sample marker is determined based on the presence of labeled peaks detected (after filtering) and the GQ weighting specified in the analysis method (SQ & GQ Settings tab): • If no labeled peaks are detected (and the sample is not a negative control), the GQ is set to 0. • If one or more labeled peaks are detected, the GQ is initially set to 1 with a final value determined by the GQ weighting of individual marker-level quality values as specified in the analysis method using the formula: $GQ = 1 \times [(1 - \textit{weight of flag 1}) \times (1 - \textit{weight of flag 2}) \times (1 - \textit{weight of flag n})]$ where: *weight of flag 1* through *weight of flag n* represent the GQ weight for the following PQVs (applied only if the PQV is ⚠): OS, BIN, PHR, LPH, AN, BD, CC, SPK, and OVL.	▮	GQ is within the Passing range (Default = 0.75 to 1.0)
			⚠	GQ is between the Passing range and Low Quality range (Default = 0.26 to 0.74)
			⬟	GQ is within the Low Quality range (Default = 0.0 to 0.25) or the SQ for the sample is ⬟.

Genotype Quality that is displayed in the Genotypes Table after the analysis is complete *(continued)*

Acronym	Full Name	Function/Rule(s)		Flag Indicator
GQ *(continued)* (samples)	Genotype Quality	**Weighting** The weighting of individual flags affects the GQ. The CC weight is defined by the software at a value of 1.0. If CC is ⚠, GQ is set to 0 (🛑). All other PQVs apply user-defined weights: • 1: If a PQV is triggered, it sets the GQ value to 🛑 . • > 0 <1:If a PQV is triggered, it reduces the GQ value according to the formula on the previous page. The higher the value, the greater the effect on GQ. • 0: No effect on the GQ calculation. **Example** A marker contains a ⚠ LPH and ⚠ OS flag, all other flags are 🟩 . The GQ weight for LPH and OS are set to 0.3. GQ = 1 × [(1 − 0.3) × (1 − 0.3)] GQ = 1 × 0.49 GQ = 0.49		
GQ (allelic ladders)	Genotype Quality	Indicates the genotype quality of the marker in the allelic ladder. The genotype quality for an allelic ladder marker is determined using system-defined quality rules (described on page 45) to ensure: • All expected peaks are present. • Peak height ratio of the first and second peak is greater than 50%. • No spikes are present in the extended marker range (within or between markers). • The peak height ratio between the lowest and highest peak is equal to or greater than 15%. **IMPORTANT!** If the Allelic Ladder GQ Weighting for Spikes is set to 0 (off) (SQ & GQ Settings tab in the analysis method), the GQ may be 🟩 , even if spikes are present in the allelic ladder.	🟩	All allelic ladder quality requirements are met.
			🛑	One or more allelic ladder quality requirements are not met or the SQ for the sample is 🛑 .

Sample-Level Quality Assessments

Sample-level quality assessment indicates the quality of the entire sample. Samples are grouped in the Analysis Summary based on the status of sample-level PQVs.

Note: Gray PQVs (▢ ▲ ⬡) indicate that a marker in the sample was edited.

Sample-Level Quality Flags that are displayed in the Samples Table after the analysis is complete

| Acronym	Full Name	Function/Rule(s)	Flag	
SOS	Sample Off-scale	Indicates if any fluorescence signal within the analysis range exceeds the detection threshold of the instrument. **Note:** The detection limit varies on different instruments. The OS quality flag indicates if the CCD camera is saturated at any data point within the analysis range. It is not based on an absolute RFU value (peak height).	▢	No fluorescence signal within the analysis range exceeds the detection limit of the instrument.
			⚠	One or more fluorescence signals within the analysis range exceed the detection limit of the instrument.

2

Sample-Level Quality Flags that are displayed in the Samples Table after the analysis is complete *(continued)*

Acronym	Full Name	Function/Rule(s)	Flag	
SSPK	Sample Spike	• **Allelic Ladders** – Indicates if spikes are detected within the sizing range. • **Samples** – If enabled in the Analysis method>Peak Quality tab, indicates if spikes are detected within or between two defined marker size ranges. Does not indicate if spikes are detected before the first marker or after the last marker. The software uses a proprietary algorithm that detects spikes based on the peak morphology. **Note:** If an allele label is changed to a Spike label by the user in the Samples or Genotypes plot, the peak is considered an artifact and is not added to the Genotypes table. **Note:** If an allele label is changed to a Spike label by the user in the Samples or Genotypes plot, the peak is considered an artifact and is not added to the Genotypes table. **Note:** Peaks with magenta artifact labels such as Spike peaks are not interpreted by the Mixture Analysis tool.	🟩	No spikes are detected within the range.
			🔺	One or more spikes are detected within the range.

Sample-Level Quality Flags that are displayed in the Samples Table after the analysis is complete *(continued)*

Acronym	Full Name	Function/Rule(s)	Flag	
MIX	Mixed Source	Indicates a potential mixed source sample. A sample is considered a potential mixture if it meets either of the following conditions: • **Condition 1**: Two or more markers contain three or more called alleles *or* • **Condition 2:** – One or more markers contain three or more called alleles *and* – Another two or more markers have ⚠ PHR (peak height ratio, described on page 48) **Note:** The MIX PQV flag identifies samples in a GeneMapper *ID-X* project that may contain DNA from a mixed source. This flag provides information to aid in data examination; it is not used by the GeneMapper® ID-X Software to identify samples for mixture analysis or in the filtering process by which samples enter the Mixture Analysis tool.	🟩	Neither condition is met.
			⚠	Either or both conditions are met, potential mixture.
			N/A	Sample Type = Allelic Ladder, or a sample contains no labeled peaks.
OMR	Outside Marker Range	Indicates if labeled peaks are detected between two marker size ranges defined in the panel. Does not indicate if labeled peaks are detected before the first marker or after the last marker. If an OMR peak is determined to be a true DNA peak, you can edit the label and assign the allele to a marker. **Note:** The default marker size ranges for neighboring markers do not overlap, but microvariants can fall between two marker size ranges.	🟩	No labeled peaks are detected between marker size ranges.
			⚠	One or more labeled peaks are detected between marker size ranges. **Note:** If you set the Cut-off value to 1, OMR peaks are not filtered.
			N/A	Sample Type = Allelic Ladder

Sample-Level Quality Flags that are displayed in the Samples Table after the analysis is complete *(continued)*

Acronym	Full Name	Function/Rule(s)	Flag	
CGQ (samples)	Composite Genotype Quality	Overall sample genotype quality indicator. Considers the individual marker GQ (Genotype Quality) values.	🟩	All individual marker GQ values are 🟩.
			🔺	One or more individual marker GQ values are 🔺 and no marker GQ values are 🔴.
			🔴	One or more individual marker GQ values are 🔴.
			🟢	Manually overridden Status was originally 🟩, 🔺, 🔴, ⬛, ▲, or ⬤ but was set to 🟢 by the user.
CGQ (allelic ladders)	Composite Genotype Quality	Allelic ladder quality indicator. Considers the allelic ladder quality assessment (see page 45). Other than SPK and OS, does not consider individual marker GQ (Genotype Quality) values. **Note:** Allelic ladder samples with 🔴 CGQ are not used to create bin offsets. For more information, see page 27. **IMPORTANT!** If the Allelic Ladder GQ Weighting for Spikes is set to 0 (off), or spike detection for the allelic ladder is disabled (SQ & GQ Settings tab in the analysis method), the CGQ may be 🟩, even if spikes are present in the allelic ladder.	🟩	All allelic ladder quality requirements are met for every marker. Used for bin offsetting.
			🔴	One or more allelic ladder quality requirements are not met for one or more markers. Not used for bin offsetting.
			🟢	Manually overridden Status was originally 🟩, 🔺, 🔴, ⬛, ▲, or ⬤ but was set to 🟢 by the user. Used for bin offsetting.

Troubleshooting

Complete the following form before contacting Technical Support.

Software Issues

Does the computer and operating system meet the required computer specifications? List any potential discrepancies.
Problem description:
Did you receive any error messages or alerts? If so, what did they say (if possible, take screen captures)? At what point did these errors/alerts occur?
Is the problem reproducible?

Steps taken prior to encountering or to reproduce the problem:

1.

2.

3.

4.

5.

Bin Overlap

B

Definition of bins and overlapping bins

Physical and virtual bins

A bin is a region that defines an allele within a marker. There are two types of bins, physical and virtual. Physical bins are alleles physically present in the allelic ladder. Virtual bins are alleles that are not present in the allelic ladder, but have been previously reported[‡] or discovered during developmental validation of a particular chemistry. In the GeneMapper® *ID-X* software, physical bins are shaded grey and virtual bins are shaded pink.

When creating the bin definitions, the goal is to always leave a gap of at least 0.1 nt or more between bins. However, once migration due to electrophoresis transpires and subsequent bin offsetting is applied, there are opportunities for overlapping bins to occur between bins that are only 0.1 nt apart.

Figure 1   Example of two physical bins located 1 nt apart with a gap of 0.1 nt.

‡ Previously reported in the STRbase (**www.cstl.nist.gov/div831/strbase**).

Table 1   Describes alleles 11.3 and 12 in Figure 1

Bin	Bin type	Size (bp)	Bin width (nt) left of peak	Bin width (nt) right of peak	Total bin width (nt)	Gap size between bins (nt)
Allele 11.3	Physical	91.5	0.50	0.45	0.95	0.1
Allele 12	Physical	92.5	0.45	0.50	0.95	

Bin offsets   Bin offsets are the size difference (in base pairs) between the reference bin locations in the panel and the actual sizes of the allelic ladder fragments. Bin offsets are automatically calculated during analysis of the allelic ladder samples. The bin offsets determine how much and in which direction (higher or lower) the reference bins shift. The offset bins are then used to genotype the other sample types.

Overlap between bins   After bin offsets are applied, bin overlap may occur between two bins, physical or virtual, that have a pre-defined separation gap of only 0.1 nt. Migrational influences may shift the bins such that the pre-defined 0.1 nt gap is eliminated, resulting in an overlap of the two bins.

Figure 2   Example of overlap between two physical bins
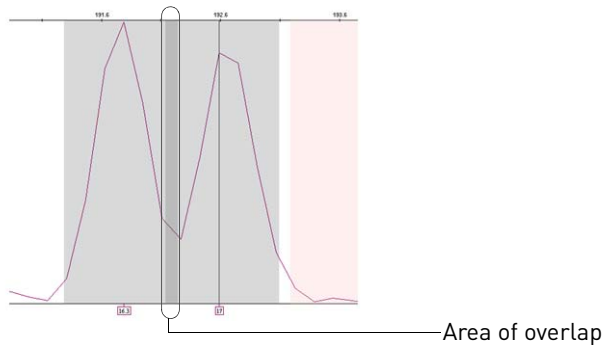
Area of overlap

Figure 1 Example of overlap between one virtual bin and one physical bin
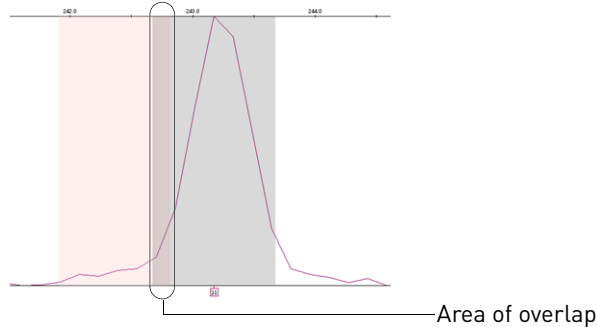
Area of overlap

Possible causes of overlapping bins

STR Kits with a high density of physical and virtual bins may be more prone to the occurrence of overlapping bins.

Due to migration conditions, bin-to-bin overlap can occur when adjacent ladder alleles within a single ladder are less the 0.9 nt apart or when the 0.1 nt gap is eliminated due to the applied bin offsets.

IMPORTANT! Reducing the bin width to correct bin overlap is not recommended. Alleles falling within the overlap region are already identified as Off Ladder (OL). Reducing the bin width causes correct allele calls to fall in the gap region. This results in more OL calls and a loss of information.

IMPORTANT! Bin overlap has been observed in data generated on both the Applied Biosystems® 3730 DNA Analyzer using POP-7™ polymer and the Applied Biosystems® 3500 Genetic Analyzer using POP-4® polymer. The frequency of bin overlap is higher in data generated on the 3730 instrument using POP-7™ polymer.

Figure 3 Example of bin width that is overlapping by 0.1 nt
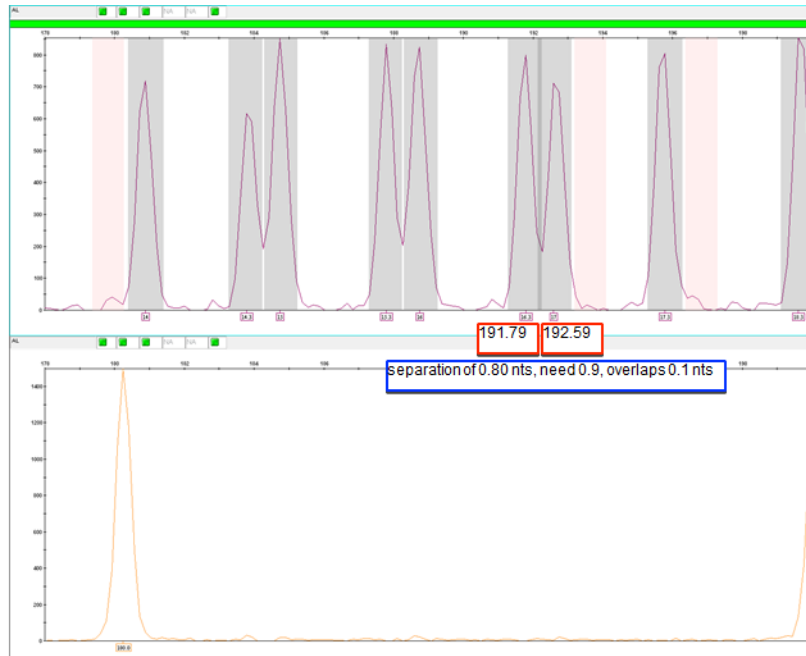
Size calling method

The Size Calling Method of the GeneMapper® *ID-X* software may affect bin overlap. The size calling algorithm impacts the peak size and changes the distance between two adjacent peaks.

IMPORTANT! The Local Southern Method sizing algorithm appears to prevent overlap more than the 3rd Order Least Squares sizing algorithm when using GeneScan™ 600 LIZ® Size Standard.

Multiple ladders
The use of multiple allelic ladders may result in different bin offsets upon comparison to analysis with a single allelic ladder. The migration rates of the allelic ladder peaks among different ladder injections may vary during capillary electrophoresis. This causes variation in the bin offset values among the multiple ladders. Therefore, when multiple allelic ladders are defined within a given GeneMapper® *ID-X* project, the bin offset values are averaged among all the defined ladders. If the averaged bin offsets eliminate the pre-defined 0.1 nt gap between bins that are 1 nt apart, then bin overlap occurs.
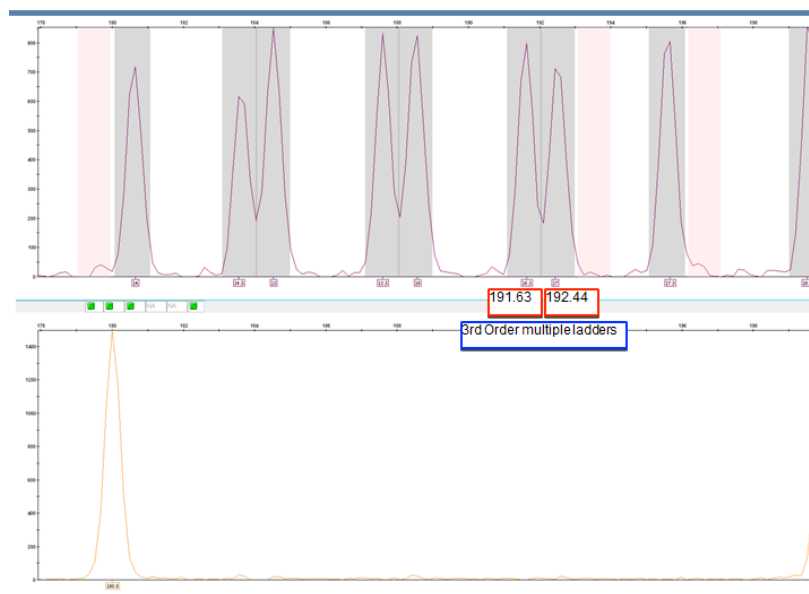
Minimizing overlapping bins

Here is a list of potential ways to minimize overlapping bins.

Category	Actions
Use a different ladder file	Remove the offending ladder from the data set, or replace it with another ladder that ran without overlap.
Size Calling Method	In the Peak Detector tab of the Analysis Method Editor, an option would be to choose the Local Southern Method sizing algorithm, which may reduce the affect of overlapping bins.
	Using the Local Southern Method on GlobalFiler® Express data from the 3730 platform may generate sizing failure because the 60 bp size peak migrates into the primer region and may not be recognized by the software.
	Conduct the appropriate internal verification studies before implementing modified analysis methods.

Recommended action

GeneMapper® *ID-X* software assigns alleles in overlapping bins as "Off Ladder" or OL. Therefore, there is no risk of incorrect genotyping. You can leave the allele call as OL, or manually assign an allele call after performing the necessary data interpretation.

IMPORTANT! Because alleles in overlapping bins are assigned OL, we recommend that you take no action to prevent overlapping bins.

Figure 4   Allele peak in a bin (top panel) and allele peak in overlapping bins called as OL (middle and bottom panels)
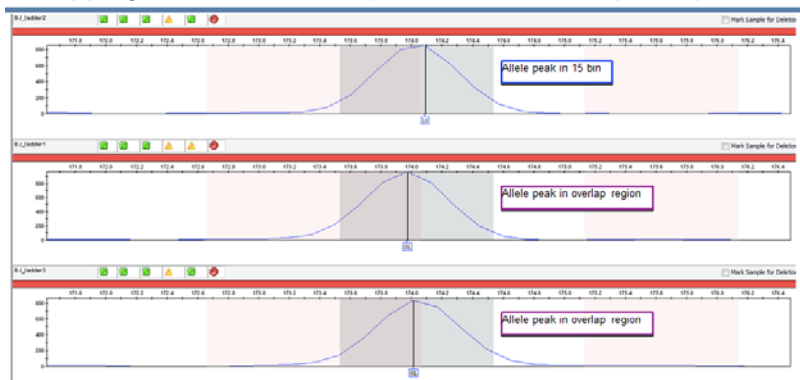
Bibliography

Baron, H., Fung, S., Aydin, A., Bahring, S., Luft, F.C., Schuster,H.1996. Oligonucleotide ligation assay (OLA) for the diagnosis of familial hypercholesterolemia. *Nat Biotechnol.*14(10):1279–82.

Brinkmann, B., Moller, A., and Wiegand, P. 1995. Structure of new mutations in 2 STR systems. *Int. J. Legal Med.* 107:201–203.

Gill, P., d'Aloja, E., Andersen, J., Dupuy, B., Jangblad, M., Johnsson, V., Kloosterman, A.D., Kratzer, A., Lareu, M.V., Meldegaard, M., Phillips, C., Pfitzinger, H., Rand, S, Sabatier, M., Scheithauer, R., Schmitter, H., Schneider, P., and Vide, M.C. 1997. Report of the European DNA Profiling Group (EDNAP): an investigation of the complex STR loci D21S11 and HUMFIBRA (FGA).

Grossman, P.D., Bloch, W., Brinson, E., Chang, C.C., Eggerding, F.A., Fung, S., Iovannisci, D.M., Woo, S., Win-Deen, E.S.1994. High-density multiplex detection of nucleic acid sequences: oligonucleotide ligation assay and sequence-coded separation. *Nucleic Acids Res*. 22(21):4527–34.

Southern, E.M. 1979. Measurement of DNA length by gel electrophoresis. *Anal Biochem*. 100(2):319–23.

Documentation and support

Related documentation

Document	Part number	Description
GeneMapper® ID-X Software Version 1.5 Administrator's Guide	100031703	Describes how to configure the administrative features of the software, maintain the GeneMapper *ID-X* and database dashboard software, and use the command-line interface.
GeneMapper® ID-X Software Version 1.5 Installation Guide	100031706	Explains how to install the GeneMapper *ID-X* software.
GeneMapper® ID-X Software Version 1.5 Getting Started Guide — Basic Features	100031701	Explains how to set up the software, set up a project, and analyze the example data provided with the GeneMapper *ID-X* Software.
GeneMapper® ID-X Software Version 1.5 Quick Reference Card — Basic Features	100031702	Provides an easy-to-follow workflow for using basic system features.
GeneMapper® ID-X Software Version 1.5 Getting Started Guide — Mixture Analysis Tool	100031704	Explains how to set up the software for mixture analysis and analyze the example data provided with the GeneMapper *ID-X* Software.
GeneMapper® ID-X Software Version 1.5 Quick Reference Card — Mixture Analysis Tool	100031705	Provides an easy-to-follow workflow for using the mixture analysis tool.

How to obtain support

For HID support:

- In the United States and Canada – send an email to HIDTechSupport@lifetech.com, or call 888-821-4443 option 1.
- Outside the United States and Canada – contact your local support office.

For the latest services and support information for all locations, go to:

www.lifetechnologies.com/support

At the web site, you can:

- Access worldwide telephone and fax numbers to contact Technical Support and Sales facilities.
- Search through frequently asked questions (FAQs).
- Submit a question directly to Technical Support.
- Search for user documents, Material Safety Data Sheets (MSDSs), vector maps and sequences, application notes, formulations, handbooks, certificates of analysis, citations, and other product support documents.
- Download .pdf documents.
- Obtain information about customer training.

Limited product warranty

Life Technologies Corporation and/or its affiliate(s) warrant their products as set forth in the Life Technologies' General Terms and Conditions of Sale found on Life Technologies' website at **www.lifetechnologies.com/termsandconditions**. If you have any questions, please contact Life Technologies at **www.lifetechnologies.com/support**.

Index