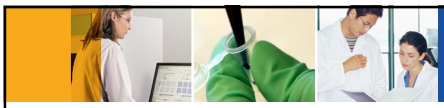
DNA Sequencing Setup and Troubleshooting

Lara Cullen, PhD
Scientific Applications Specialist
Australia and New Zealand

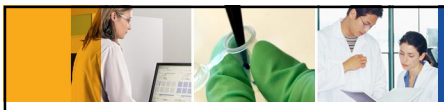
Applied Biosystems



Applied Biosystems

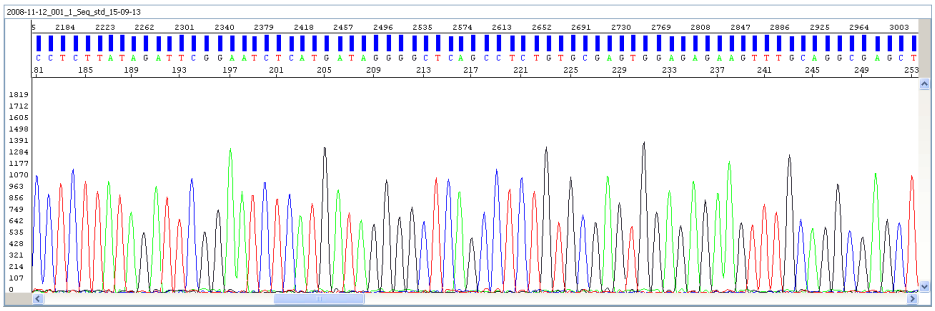
Reviewing Sequencing Data

- *Review the Electropherogram*
- *Review the Raw Data (Signal intensity, data start points, baseline, length of read, artefacts)*
- *Review the EPT Plot*
- *Review Data Analysis settings (correct basecaller, mobility file?)*
- *Is there any pattern to the problem? (eg: specific capillary?, specific primer?)*



Good Quality Sequencing Data

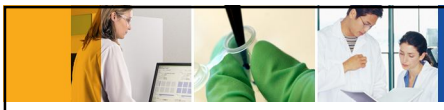
Electropherogram



Peaks should be evenly spaced, not too broad and no tailing

Read length (QV =20) should be ~600bp with POP6 on 50cm Array

Baseline "noise" under the main peaks should be minimal



Quality Values Chart

Quality Value = -10log$_{10}$ (Pe)

Pe is probability of error

KB basecaller generates QVs from 1 to 99

Typical high quality pure bases will have QV 20- 50

Typical high quality mixed bases will have QV 10-50

Size and color of QVs bars are identical for QVs 50-99

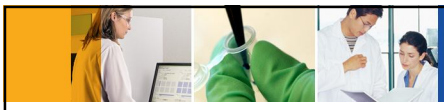
QV	Pe	QV	Pe	QV	Pe
1	79%	21	0.790%	41	0.0079%
2	63%	22	0.630%	42	0.0063%
3	50%	23	0.500%	43	0.0050%
4	39%	24	0.390%	44	0.0039%
5	31%	25	0.310%	45	0.0031%
6	25%	26	0.250%	46	0.0025%
7	20%	27	0.200%	47	0.0020%
8	15%	28	0.150%	48	0.0015%
9	12%	29	0.120%	49	0.0012%
10	10%	30	0.100%	50	0.0010%
11	7.9%	31	0.079%	60	0.0001%
12	6.3%	32	0.063%	70	0.00001%
13	5.0%	33	0.050%	80	0.000001%
14	4.0%	34	0.040%	90	0.0000001%
15	3.2%	35	0.032%	99	0.00000012%
16	2.5%	36	0.025%		
17	2.0%	37	0.020%		
18	1.6%	38	0.016%		
19	1.3%	39	0.013%		
20	1.0%	40	0.010%		

More information available in Chapter 6 of Sequencing Analysis User Guide and Chapter 10 of SeqScape User Guide

Quality Values are an indicator of the chance of an incorrect basecall in your sequencing data

QV 10 = 1 in 10 chance
QV 20 = 1 in 100 chance
QV 30 = 1 in 1000 chance
QV 40 = 1 in 10000 chance

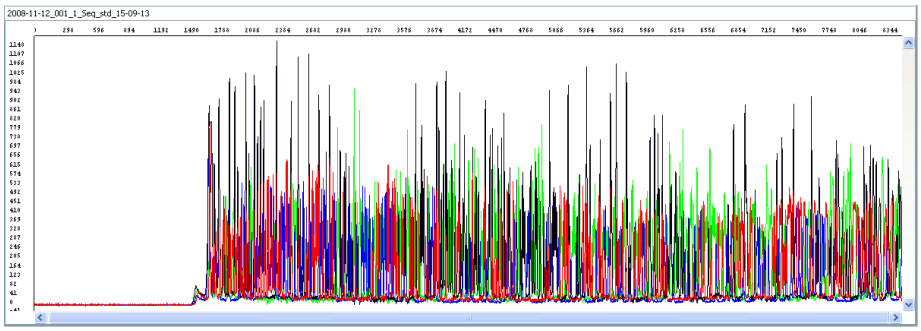
Quality Values of 20 or higher will give blue bars when the default settings are used in Sequencing Analysis



Applied Biosystems

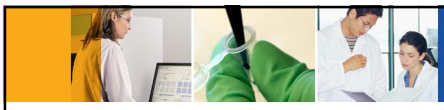
Good Quality Sequencing Data

Raw Data



Aim for signal strength in the range of a few hundred to a couple of thousand.

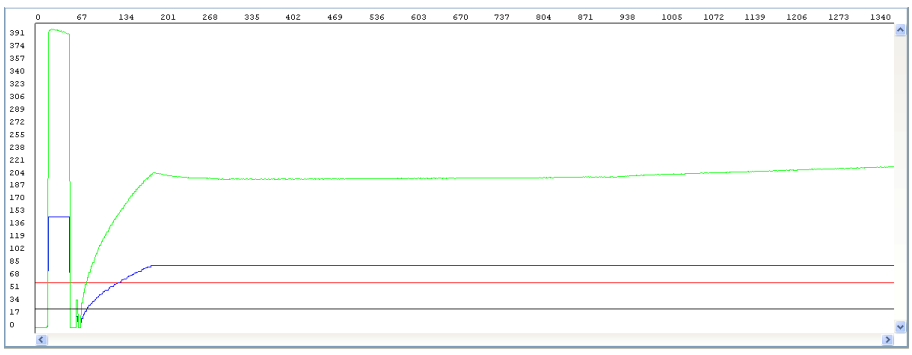
There is no official spec for Signal to Noise Ratio, but samples that have ratios >100 are generally considered to be good.



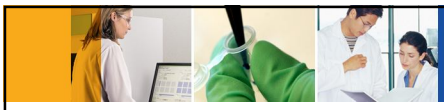
Applied Biosystems

Good Quality Sequencing Data

EPT Data



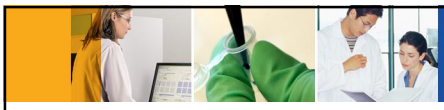
Check the current and voltage both have the normal profile, anything unusual could indicate a reagent or instrument problem



Sequencing Troubleshooting:
Defining the Problem

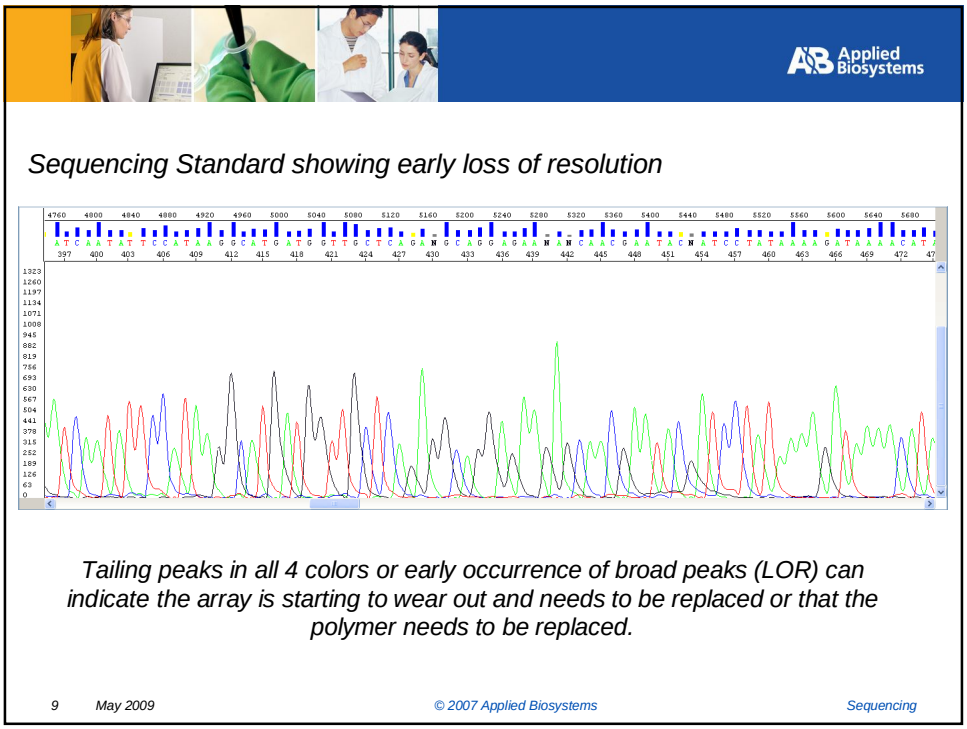
- *Poor data can occur for many reasons*
  - *Instrument Problem?*
  - *Array Problem?*
  - *Polymer Problem?*
  - *Sequencing Primer Problem?*
  - *Template DNA Quality Problem?*
  - *Sequencing Primer Problem?*
  - *Data Analysis Problem?*

Controls are an essential part of defining the problem



BigDye® Terminator v3.1 Sequencing Standard

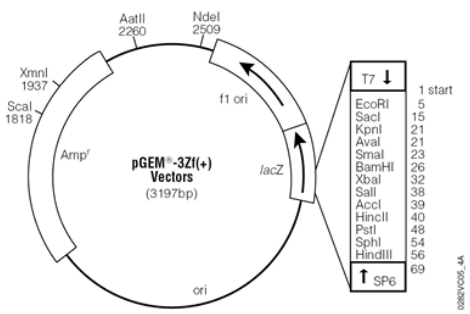
- *Lyophilized sequencing reactions that require only resuspension and denaturation before use*
- *Validate the instrument performance and rule out problems with common reagents and consumables such as polymer, array, buffer, plasticware and septa*



Sequencing Standard showing early loss of resolution

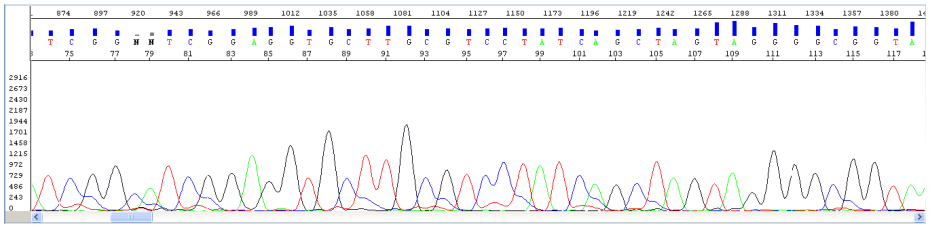
Tailing peaks in all 4 colors or early occurrence of broad peaks (LOR) can indicate the array is starting to wear out and needs to be replaced or that the polymer needs to be replaced.

pGEM®-3Zf(+) control DNA + M13 forward primer



- *High quality plasmid DNA and primer that can be used in a control reaction with BigDye Terminator ready reaction mix*
- *Run and clean up the pGEM reaction in the same way as you do your samples*
- *Validates the Chemistry (BDT kit), thermal cycling conditions, thermal cycler and the sequencing reaction clean-up*

pGEM control reaction showing trailing peaks on C's only



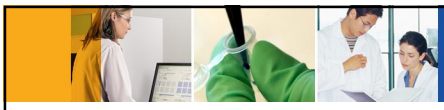
Tailing in the C peaks alone may indicate a problem with the BDT ready reaction mix. BDT exposed to light during storage or excessive freeze thaw cycles may show this problem. Samples loaded in water instead of Hi-Di formamide may also be more prone to this problem

Sequencing Reaction Setup Example

Reaction Component	**Volume/reaction**
BDT Ready Reaction Premix (2.5x)	1.0 µl
BigDye Sequencing Buffer (5x)	3.5 µl
Primer — 3.2 pmol/ul	1.0 µl
Template DNA (10ng/ul)	1.0 µl
Water	13.5 µl
Final Volume (1X)	**20 µl**

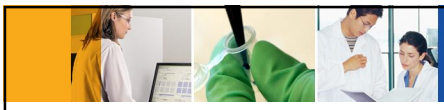
Sequencing Reactions contain only BDT Reaction premix, buffer and your template DNA and primer.

If sequencing standard and pGEM controls give good quality data the problem may lie with the template DNA preparation method, template DNA quantity or the primer quality.



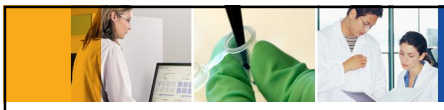
Additional controls for DNA and Primer quality

- *Are some samples giving good quality data and others bad with the same primer?*
- *Are you sequencing the same template DNA with more than one primer (eg: forward and reverse) and finding one works much better than the other?*
- *Have you recently switched DNA preparation methods?*
- *Do you quantitate your DNA before sequencing?*
- *Do you have a sample that you know works well that you can use as a lab specific control?*



Primer Design

- *Use Generic Sequencing primers if possible (eg: M13 Forward and M13 Reverse or T3 and T7) when sequencing plasmids*
- *When designing primers for Sequencing keep in mind the following:*
  - *Primers should be at least 18bp and avoid runs of identical nucleotides to ensure they are specific to the intended target and hybridise well*
  - *Avoid primers that have the potential to form dimers or have secondary structure*
  - *GC content should be between 30-80% (50% optimal)*



Primer Design

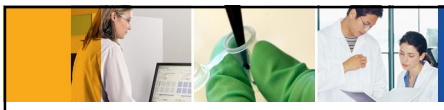
- *In most cases PCR primers with Tm of about 60°C will work as sequencing primers*
- *If you are having problems, primers with Tm of about 55°C may work better than higher or lower Tm since reaction conditions are as follows:*

96°C for 1 min] **1 cycle**

96°C for 10 sec
50°C for 5 sec] **25 cycles**
60°C for 4 min

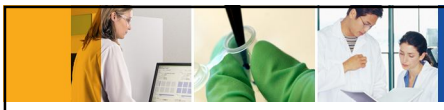
4°C Hold

Dilute fresh primers from stocks regularly as they aren't as stable when stored at low concentrations



Template DNA Quality

- *When Sequencing PCR products ensure the PCR is specific*
  - *High Tm primers (60 degrees or more)*
  - *One specific band when run on a gel*
  - *PCR clean up kits or ExoSAP-IT® should work well if the PCR product is specific*
- *Commercial plasmid miniprep kits generally give DNA of sufficiently high quality for sequencing*
  - *Try not to overload the columns*
  - *Some spin column based kits may leave residual resin that can interfere with injection and cause failed samples. Centrifuge then take from the top of the sample for sequencing*



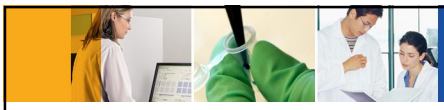
Template DNA Quantity

Template	Quantity
PCR product:	
100–200 bp	1–3 ng
200–500 bp	3–10 ng
500–1000 bp	5–20 ng
1000–2000 bp	10–40 ng
>2000 bp	20–50 ng
Single-stranded	25–50 ng
Double-stranded	150–300 ng
Cosmid, BAC	0.5–1.0 µg
Bacterial genomic DNA	2–3 µg

Too much DNA or too little will reduce the length of read and the quality of base calls

The suggested template DNA quantities should be used as a guide however you may need to optimise your own quantities in some cases

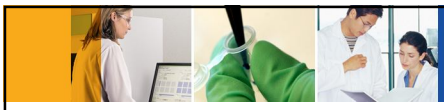
Quantitate your DNA by gel electrophoresis or UV absorbance



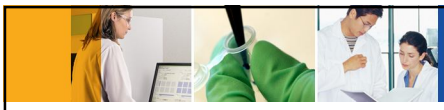
DNA prep, PCR, PCR purification → Cycle Sequencing & post purification → Instrument

BDT Sequencing Standard	✓ Validated
pGEM from the sequencing kit	✓ Validated
Laboratory internal DNA quality Control	✓ Validated

Running controls helps in focusing troubleshooting efforts and reduces time taken to determine the cause of the problem



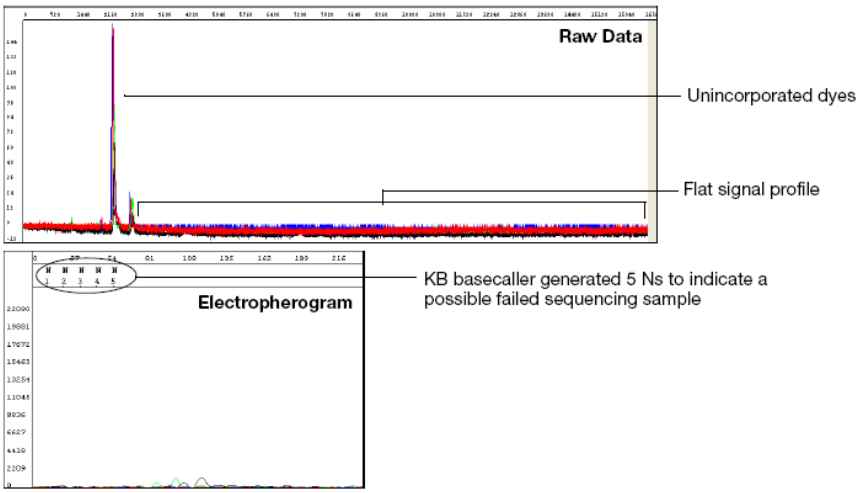
Common Sequencing Problems



Sample Prep Issues

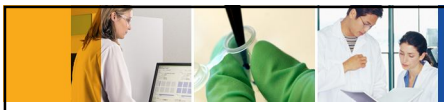
- *No Injection/sample*
- *Overloading*
  - *Too much template*
  - *Spectral pull up*
- *Multiple Sequencing Products*
- *Poor Sequencing Primer quality*
- *Dye Blobs*

No Injection/sample

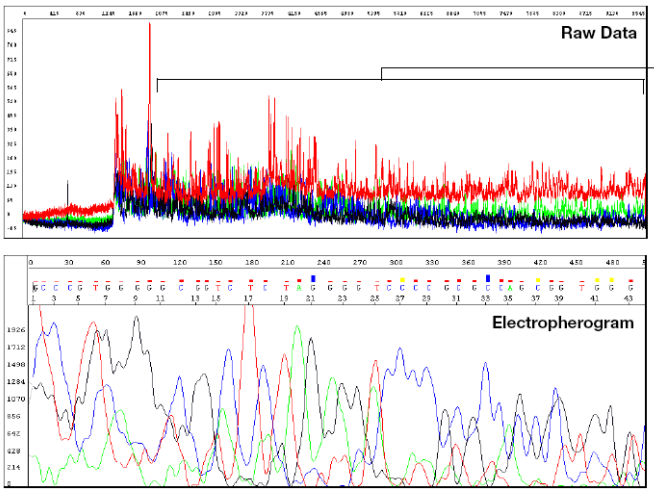


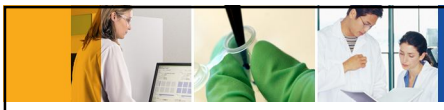
No Injection/sample

- *Failed Reaction*
  - *Component left out of reaction*
  - *Wrong primer used*
  - *Enzyme not working*
  - *Thermal cycler problem*
  - *Not enough DNA*
  - *Inhibitor in DNA*
- *Labelled product lost during cleanup*



Low Signal





Low Signal

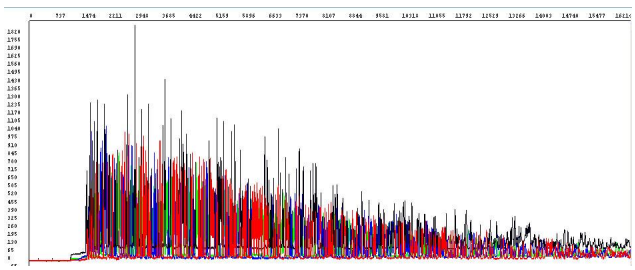
- *Sequencing reaction failed*
- *Not enough primer/tempate/Big Dye Terminator*
- *Partial loss of product during cleanup*
- *Difficult template sequence (Adding 5% DMSO or 1M Betaine to the reaction may help in some cases)*
- *Salts in sample interfering with electrokinetic injection*

Sample Overloading

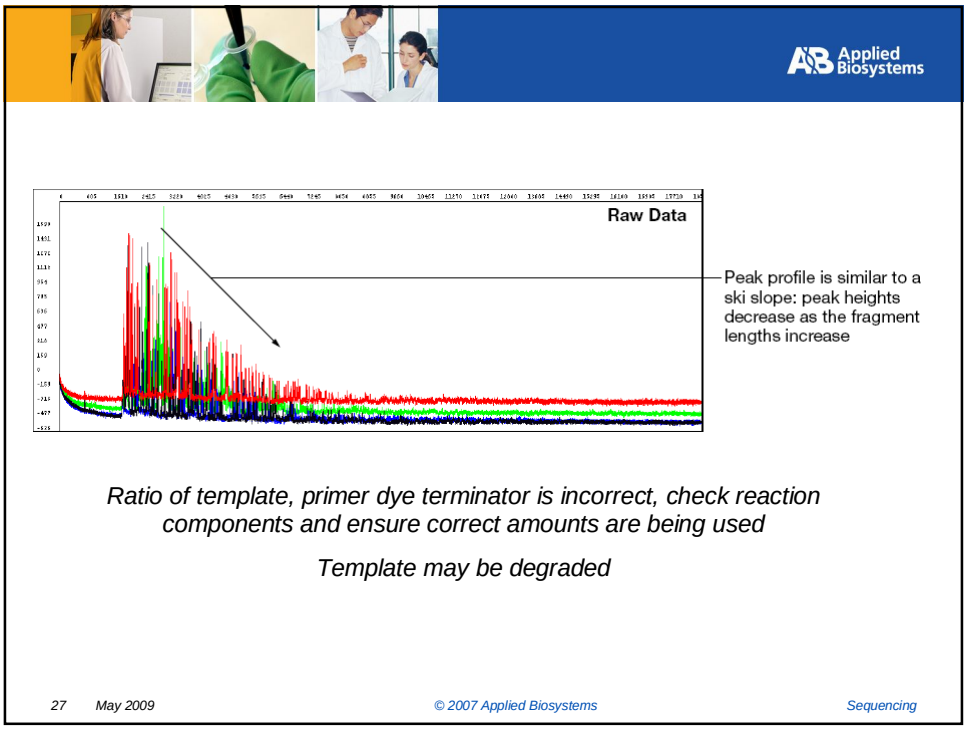
- *Can refer to too much template being present and/or a very robust reaction resulting in a very strong signal.*
- *In the Raw Data, if the signal exceeds the values below, the Analysis Software may not analyze the data properly:*
  - *3100 and 3130/3130xl Genetic Analyzers: >8000 rfu*
  - *3730/3730xl Genetic Analyzers: >32,000 rfu*

Note: RFU values for overloading are based on software analysis values. Sample overloading and miscalling may occur with rfu values much lower than listed. Pull-up (peaks under peaks) in the Sequencing data may occur with very strong/overloaded signal.

Overloading – Too Much Template



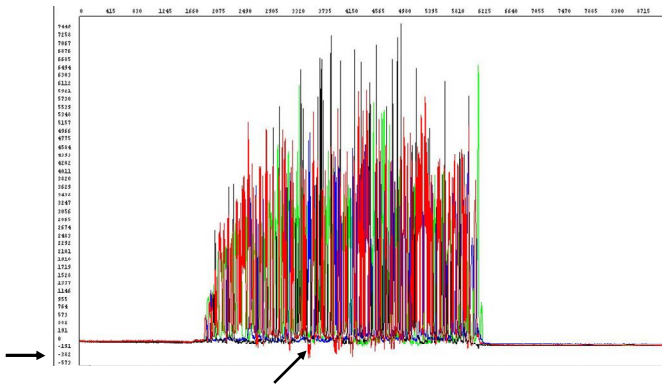
If samples contain too much DNA, the labelled ddNTPs can be incorporated early on, creating a "top heavy" reaction where the data looks strong in the front and then gets weaker as the run progresses, resulting in shorter reads. Careful quantitation of the DNA template can usually avoid this.



Ratio of template, primer dye terminator is incorrect, check reaction components and ensure correct amounts are being used

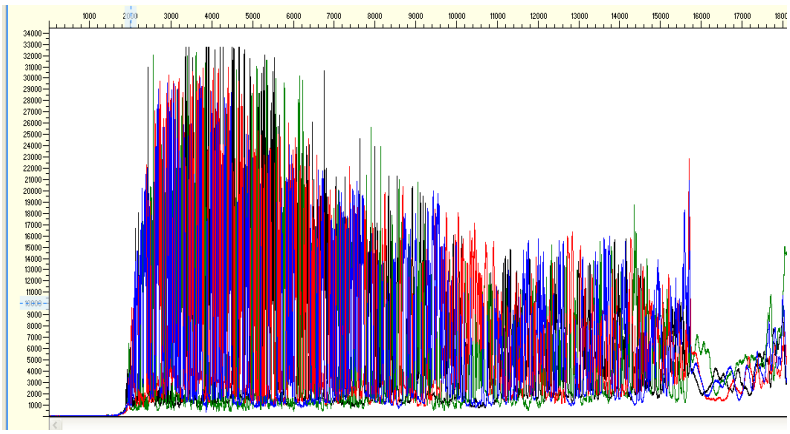
Template may be degraded

Overloading – Spectral Pull up



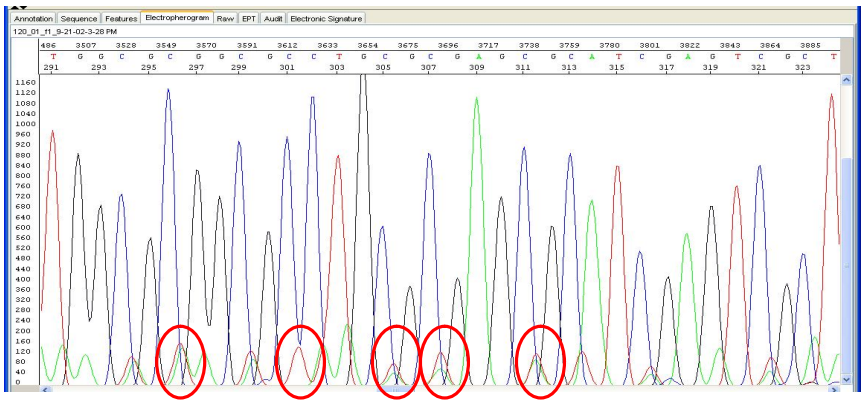
Raw data showing overloaded data on an Applied Biosystems 3100 Genetic Analyzer. In the areas where the signal is very strong, you can start to see pull down (negative peaks) in the baseline on multi-capillary instruments.

Overloading – Spectral Pull up



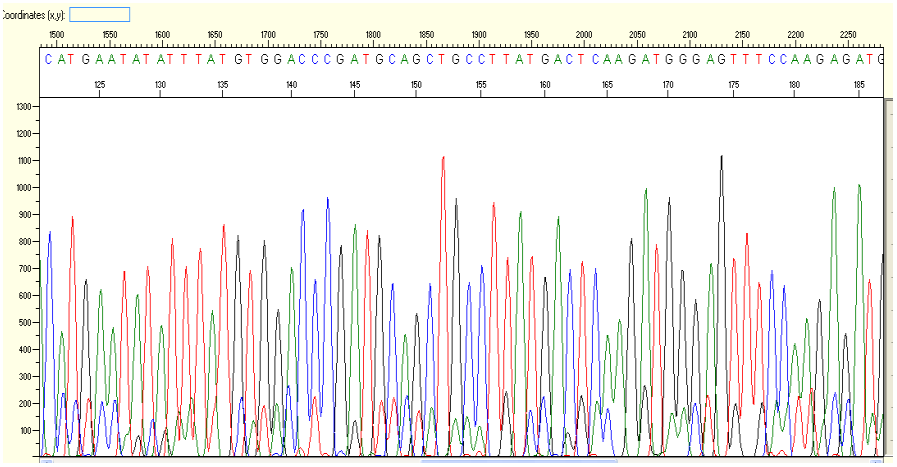
Raw Data – Peaks are off scale on the Applied Biosystems 3730 Instrument.

Overloading – Pull Up Peaks

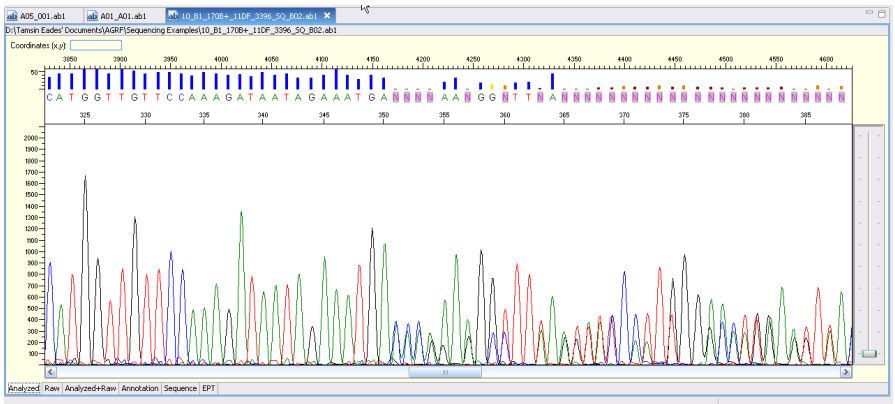


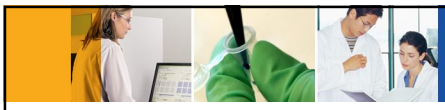
Peaks appearing under peaks in a discernable pattern. Similar pattern can also occur if the wrong spectral/matrix is used or if the spectral needs to be re-done due to changes in the optics.

Sample Prep Issues
Multiple Products



Multiple Products: Heterozygous insertion or Deletion

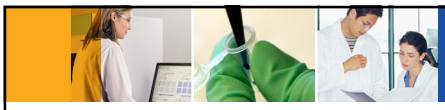




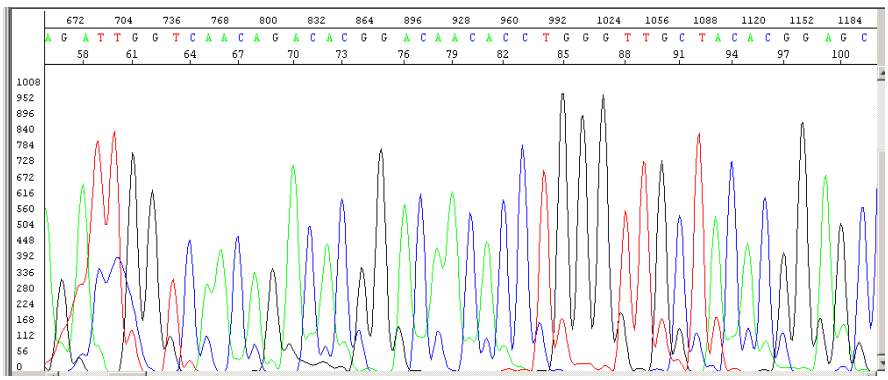
Sample Prep Issues - Multiple Products

Multiple products can be caused by:

- *Non-specific binding of the primer to the template during PCR or Cycle Sequencing*
- *Multiple clones or colonies present during sample prep or PCR products*
- *Heterozygous insertions or deletions (HIM)*
- *Contamination from water/environment*
- *Re-using Septa*



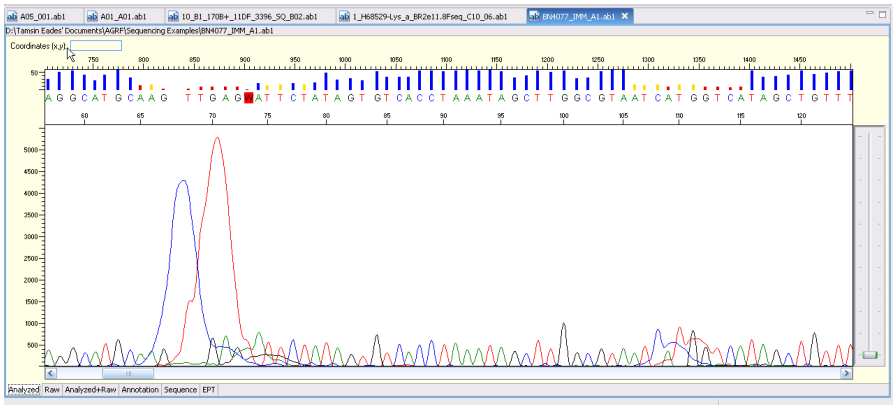
Poor Sequencing Primer quality



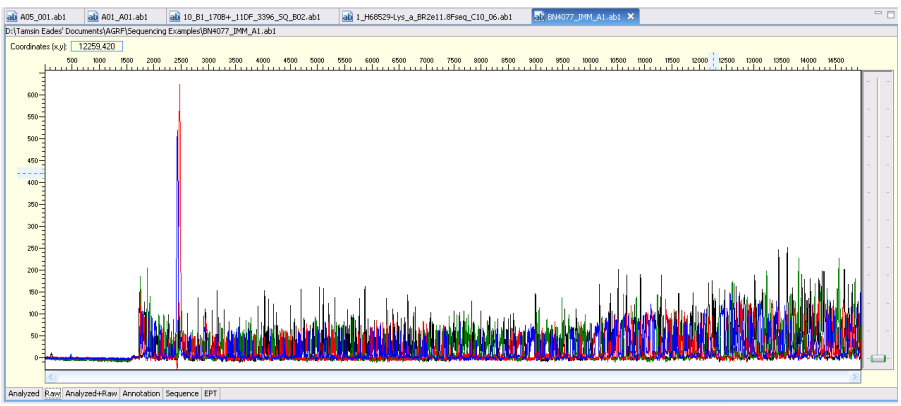
N+1 effect due to poor primer manufacturing

Dye Blobs – Problem with sequencing clean-up

Electropherogram showing Dye Blobs



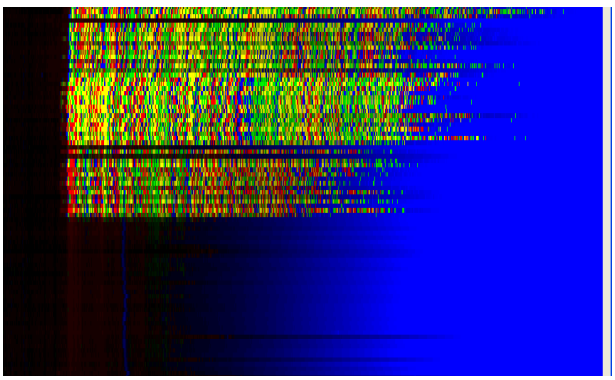
Raw Data showing Dye Blobs



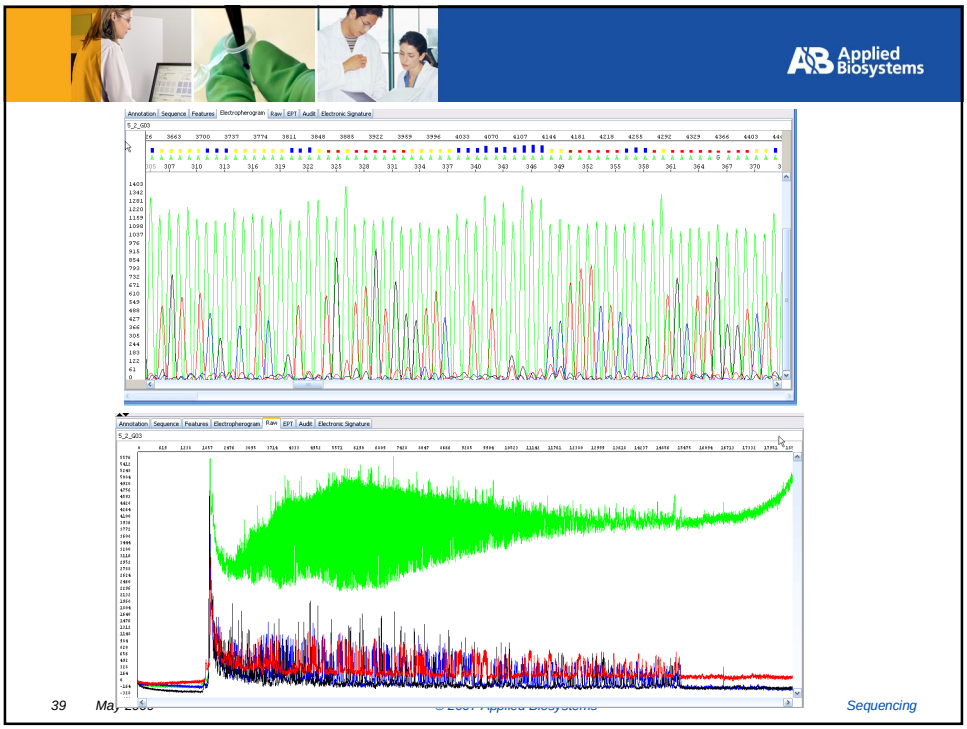
General Issues - Hardware

- *Hazes – Red, Blue, Green*
- *Dust*
- *Bubbles*
- *Electrophoresis Problem*

Hazes – Red, Blue, Green



Hazes can appear in almost any color and are usually the result of some contaminant getting into the system.

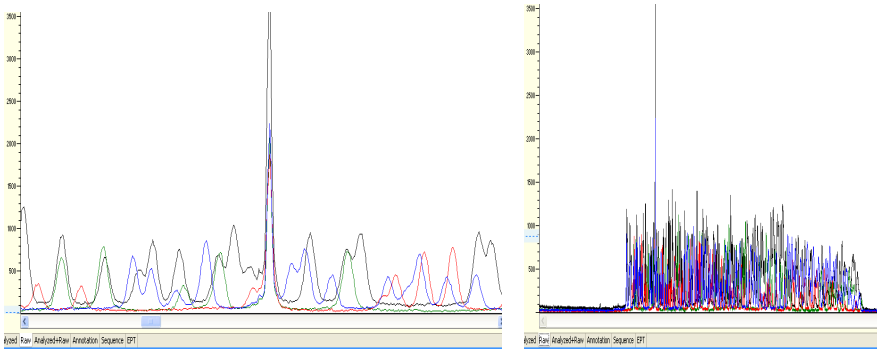


Hazes – Red, Blue, Green

Potential Reasons:

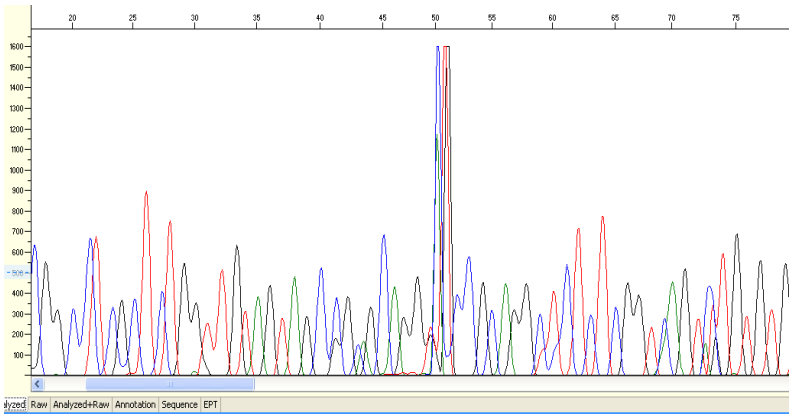
- *Improper/little maintenance*
- *Contaminant in the water used to clean the system*
- *Use of solvents or cleaners to clean instrument components.*
- *Residual Carbon or Ozone from an Arcing event.*

Dust causing Spikes in the Data

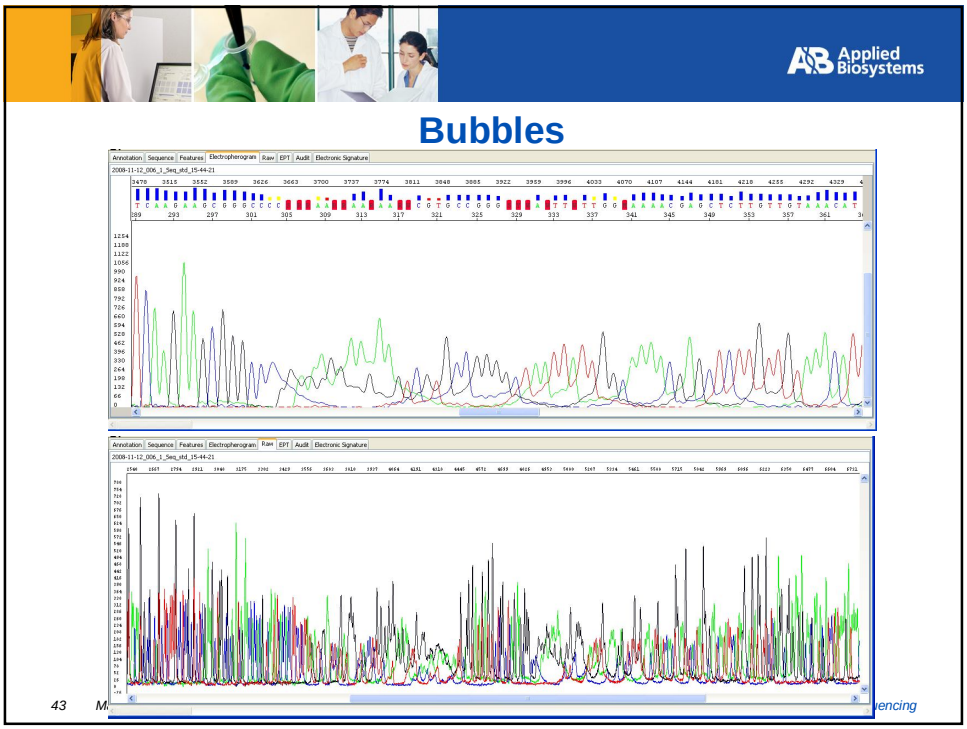


Raw Data view. The figure on the left is a zoomed in view of the spike in the data seen in the zoomed out view on the right. Note that all 4 colors are present in a thin spike.

Dust causing Spikes in the Data



Analyzed data view of a dust spike crossing the read region.



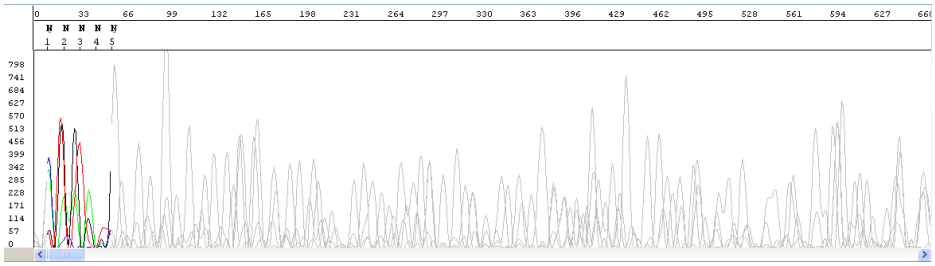
Bubbles

Bubbles

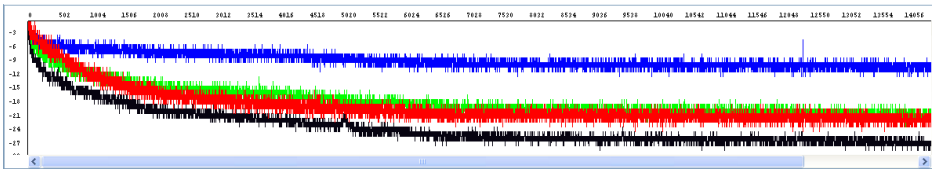
- *Check for leaks on the system.*
- *Polymer should be allowed to equilibrate and degas for 30 – 60 minutes prior to placing on the instrument.*
- *Run the Bubble Remove Wizard on the Applied Biosystems 3100/3130 and 3730 series to remove bubbles.*
- *Make sure all fittings are tight.*
- *Make sure the Ferrule Tip area is clear of bubbles or microbubbles.*
- *If bubbles persist in spite of all fittings being tight please call AB Technical Support for assistance*

Electrophoresis Problem

Electropherogram

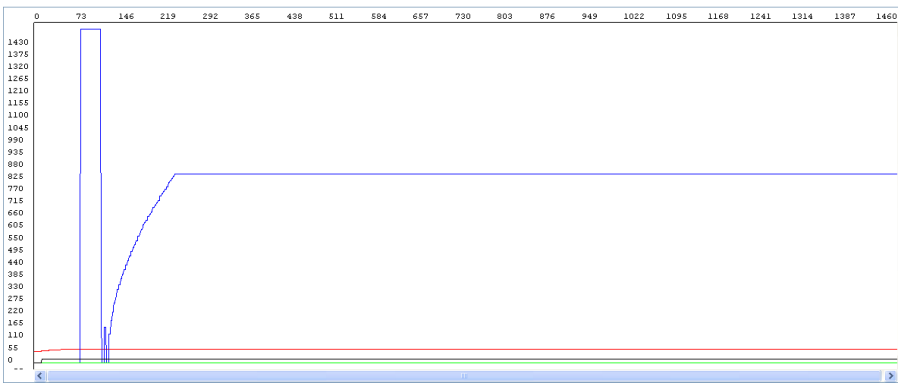


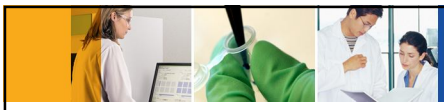
Raw Data



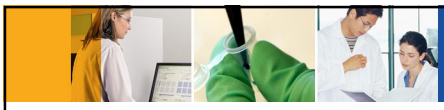
Electrophoresis Problem

EPT Data





Analysis Troubleshooting



Dye Sets/Virtual Filter Sets

Dye set/Filter set needs to be chosen before the run, normally samples need to be re-run if the wrong dye set is chosen

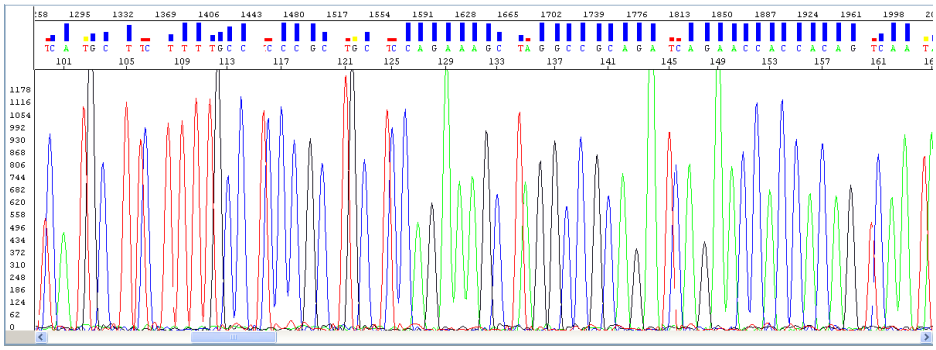
Sequencing Dye Sets

- *E: BigDye® Terminator v.1.1*
- *Z: BigDye® Terminator v.3.1*

Fragment Analysis Dye Sets

- *D: 6-FAM™, NED™, HEX™, ROX™*
- *D: 6-FAM™, NED™, VIC®, ROX™*
- *F: 5-FAM™, JOE™, NED™, ROX™*
- *G5: 6-FAM™, VIC®, NED™, PET®, LIZ®*
- *E5: dR110, dR6G, dTAMRA™, dROX™, LIZ®*

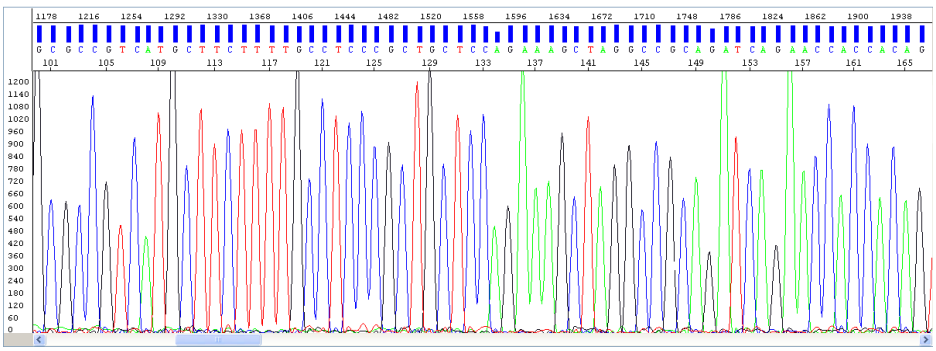
Analysis Problem – Wrong Dye Set/Primer File



The sample above was Analysed with the Dye Set/Primer file for BDTv1.1 but the chemistry used to set up the reactions was BDTv3.1

Overlapping bases and poor quality values result

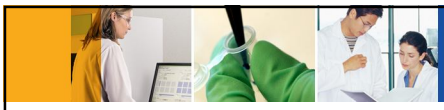
Analysis Problem – Wrong Dye Set/Primer File



This is the same sample re-analysed with the Dye Set/Primer file for BDTv3.1

Note the improvement in the spacing and quality values

This is the only circumstance in which you can change the dye set after the run



For any questions you have or assistance you need contact AB Technical Support:

anztechsupport@appliedbiosystems.com

1800 636 327 (Aus) / 0800 636 327 (NZ)