



MASSEY UNIVERSITY
COLLEGE OF SCIENCES
TE WĀHANGA PŪTAIAO

MASSEY GENOME SERVICE

**Sanger Sequencing and Genotyping
using Life Technologies 3730
capillary instrumentation
Operational since 2003**

SEQUENCING TECHNICAL BULLETIN
February 2019



BULLETIN INCLUDES

**Quality Values, Mixed Bases and IUB codes
Information included in the Analysis Report
Primer Design Recommendations
Sequencing Reaction Recommendations**

**Enquiries regarding sequencing technical information contact:
Xiaoxiao Lin – Laboratory Manager**



MASSEY GENOME SERVICE (MGS)

Sanger Sequencing using Life Technologies 3730 capillary instrumentation

TECHNICAL BULLETIN

Quality Values and Mixed Bases

Quality Values

The Sequencing Analysis Software used by the Massey Genome Service to analyse customers sequencing data uses an algorithm that calculates quality values (QVs) for pure and mixed bases. The QV is a per-base estimate of the basecaller accuracy. The per-base QVs are calibrated on a scale corresponding to:

$QV = -10 \log_{10}(P_e)$, where P_e is the probability of error.

The range of a QV is 1-50, with 1 being low confidence and 50 being high confidence.

| Quality Value (QV) | Probability the Basecall is incorrect |
|--------------------|---------------------------------------|
| 10 | 10% |
| 20 | 1% |
| 30 | 0.1% |
| 40 | 0.01% |
| 50 | 0.001% |

QVs are displayed as bars above each base in the sample electropherogram. You will see QVs on your electropherogram printouts only. EditView, Chromas and Sequencher will not show QVs. DNA Sequencing Analysis Software® version 5.0, Sequence Scanner® v1.0 and SeqScape® Software will show QVs.

The colour of a QV bar indicates its value.

| QV Bar | Default Colour and Range | The range identifies data that is: |
|--------|--------------------------|------------------------------------|
| Low | Red 0 to 14 | Not acceptable |
| Medium | Yellow 15 to 19 | Needs manual review |
| High | Blue 20 to 50 | Acceptable |

Mixed Bases

The Sequencing Analysis Software uses an algorithm called “KB basecaller” for calling bases. Mixed bases are one-base positions that contain two bases. The KB basecaller assigns A, C, G, or T to pure bases and an IUB code to mixed bases. The IUB codes are in the table below.

IUB codes – International Universal Base coding

| Code | Translation |
|------|---------------------------|
| A | Adenosine |
| C | Cytidine |
| G | Guanine |
| T | Thymidine |
| B | C, G, or T |
| D | A, G, or T |
| H | A, C, or T |
| R | A or G (puRine) |
| Y | C or T (pYrimidine) |
| K | G or T (Keto) |
| M | A or C (amino) |
| S | G or C (Strong -3H bonds) |
| W | A or T (Weak -2H bonds) |
| N | ANY base |
| V | A, C, or G |

Information included in the Analysis Report

You will be sent an analysis report of the sequencing data, if you tick this on your online submission. The report will show the success and/or failure of your sequencing data. The report displays a summary of QVs and length of reads (LORs), as well as individual sample information and errors. The report is used to help troubleshoot and provide easy assessment of data quality.

The report contains four tables of information:

- **Summary:**
 - Displays the number of samples in the report
 - Displays the number of samples that contain QVs (sequences analysed with the KB basecaller).
 - Displays the definition of the QV ranges.
- **Length of Read (LOR):**
 - Displays the definition of the LOR ranges and colours, and the number of samples in each range.
- **Sample Details:**
 - This part of the report contains a list of each sample and its associated analysis information. The following information is given:
 - Well - The well position in the plate that the sample was in.
 - Cap# - The capillary number that the sample was run on.
 - Peak1- The scan number that represents the first data point in the file that is from the sample, not including primer peaks.
 - Base Spacing -The values represents the calculated base spacing for the sample. Base spacing is defined by the number of scan points from the crest of one peak to the crest of the next peak.
 - Bases with Low, Medium, or High QVs - The number of bases where the QVs are in the low, medium, or high range.
 - Sample Score - The average quality value of the bases in the clear range sequence for that sample.
 - LOR - The usable range of high-quality or high-accuracy bases, as determined by quality values.
 - A, G, C, or T S/N - The value represents the average signal/average noise of a bases in a sample.
 - Avg S/N -The value represents the average signal/noise value of all the bases in a sample.
- **Errors Table:**
 - This part of the report displays the errors that occurred during analysis and post processing.
 - This part of the report gives the name of the sample that the error occurred in, where in Basecalling or post processing the error occurred, and a description of the error.

Primer Design

The following recommendations are provided to help optimize primer selection:

- Primers should be at least 18 bases long to ensure good hybridization.
- Avoid runs of an identical nucleotide, especially runs of four or more Gs.
- Keep the G-C content in the range 30-80%, preferably 50-55%.
- For cycle sequencing, primers with $T_M > 45^\circ\text{C}$ produce better results than primers with lower T_M using Applied Biosystems Inc. thermal cycling parameters.
- For primers with G-C content less than 50%, it may be necessary to extend the primer sequence beyond 18 bases to keep the $T_M > 45^\circ\text{C}$.
- Use of primers longer than 18 bases also minimizes the chances of having a secondary hybridization site on the target DNA.
- Avoid primers that hybridize to form dimers.
- Avoid palindromes because they can form secondary structures.
- The primer should be as pure as possible, preferably purified by HPLC.

Sequencing Reaction Recommendations

The following is useful information if you are using the **Capillary Separation Service, Capillary Separation Service with reaction clean-up, Plate Service, or Plate service with reaction clean-up**, and are doing your own sequencing reactions using BigDye™ Terminator chemistries.

The Massey Genome Service recommends doing a 20µl total volume sequencing reaction.

If you are diluting the sequencing reaction mix and are using the 5x sequencing reaction buffer supplied with the BigDye™ Terminator chemistries, please follow the dilution guidelines below for doing the dilutions:

| Reaction | 1X | ½ X | ¼ X | 1/8 X |
|-----------------------|----------|----------|----------|----------|
| BDT | 8 | 4 | 2 | 1 |
| 5X Seq. Buffer | 0 | 2 | 3 | 3.5 |
| Primer | 3.2 pmol | 3.2 pmol | 3.2 pmol | 3.2 pmol |
| Template | - | - | - | - |
| Water | To 20µl | To 20µl | To 20µl | To 20µl |

For cleanup of your fluorescent labelled sequencing products, and remove the unincorporated fluorescent ddNTPs, the Massey Genome Service recommends using the following commercial kits:

- X-Terminator™: Supplied by Applied Biosystems Inc.
- Agencourt Cleanseq™: Supplied by Beckman Coulter

With capillary sequencing where the ions in the sample are electrokinetically injected into the capillaries, it is critical to remove all unincorporated labelled ddNTPs when purifying your extension products, because they can be preferentially injected.

You may also choose to use ethanol precipitation for cleaning up your sequencing reactions. This method is much less costly than using a commercial kit, but is less effective in the removal of the unincorporated fluorescent ddNTPs from sequencing reactions.

Ethanol/EDTA/Sodium Acetate Precipitation method

NOTE: This method is for 20µl total volume sequencing reactions.

- If you did the sequencing reaction in 0.2ml strip tubes or 0.2ml individual PCR tubes, pipette the full contents of the sequencing reaction into a 0.5ml or 1.5ml tube. If you did your sequencing reactions in a 96 well plate, carry out the ethanol precipitation in this plate.
- Make up a 1:1 premix solution of 3M sodium acetate: 125mM EDTA.
- Add 4µl of the premix solution to each of the sequencing reactions.
- Add 50µl of 100% ethanol to each sequencing reaction. Vortex and spin down briefly so that the contents is at the bottom of the tube.
- Incubate at room temperature for 15 minutes.
- Spin at 13000rpm for 15 minutes at room temperature.
- Aspirate off the supernatant using a pipette. Make sure that when aspirating that all of the liquid is removed, before proceeding to the next step. If there is any supernatant left after aspiration, then give the tube a quick spin and pipette the rest of the supernatant off.
- Add 70µl of 70% ethanol to each tube. Vortex and spin at 13000rpm for 10 minutes. Aspirate off the supernatant using a pipette. Make sure that when aspirating that all of the liquid is removed, before proceeding to the next step.
- Perform a second wash with 70µl of 70% ethanol and spin for another 10 minutes at 13000rpm.
- Aspirate off the supernatant using a pipette. Make sure that when aspirating that all of the liquid is removed, before proceeding to the next step.
- Make sure the tubes are completely dry before submitting to the Massey Genome Service for capillary separation.

NOTE: When using ethanol precipitation make sure that all liquid is removed after each spin, before proceeding to the next step. If precipitating in tubes, aspirate off the supernatant with a pipette. If there is any supernatant left after aspiration, then give the tube a quick spin and pipette the rest of the supernatant off. If precipitating in a plate, tip the plate upside down and remove the supernatant onto a lint-free tissue. Then spin the plate upside down at 185 X g for 1 minute, to remove the rest of the supernatant.

Make sure you make up regular stocks of 3M sodium acetate pH5.2, 125mM EDTA, and 70% ethanol, for the ethanol/EDTA precipitation and the ethanol/EDTA/sodium acetate precipitation methods.