Sanger Sequencing Best and Worst Practices

A resource for troubleshooting your Sanger Sequencing data

Sanger sequencing was first described by Fred Sanger *et al*. in 1977. In 1986, the Applied Biosystem's ABI 370A became the first automatic fluorescence-based Sanger sequencing instrument. That first instrument used a polyacrylamide gel for electrophoresis, could run 16 samples at a time, and produced about 450 bases of sequence per sample. In 1996, Applied Biosystems introduced the ABI 310, which replaced gel electrophoresis with electrophoresis in a single capillary. Improvements in chemistry increased read lengths to 600 bases. The ABI 3730 was released in 2002, and it offered the ability to sequence 48 samples at a time. The MSU Genomics Core has an ABI 3730XL, which sequences 96 samples at a time and can process many 96-well plates per day. Read lengths of 800 to 900 bases are possible with the 3730XL.

More than 45 years after the first description of Sanger sequencing and over 35 years since the first automated Sanger sequencing instrument was first introduced, Sanger sequencing is still an important tool for molecular genetics. While automated Sanger sequencing is extremely convenient, it has been around for so long that many researchers take it for granted. To keep costs low, most Sanger sequencing facilities prepare 1/16th reactions, which use very small amounts of template, primer and fluorescent reaction mix. The nature of these small volume reactions means that when templates and primers are not present at prescribed levels, the quality of the sequence data may suffer.

In the RTSF Genomics Core, we have seen many examples of how improper template or primer preparation can affect Sanger sequencing quality. This Sanger Sequencing guide has been developed to show best and worst practices when performing Sanger Sequencing. This guide shows examples of good- and poor-quality sequence. Explanations are provided about how results were obtained. In particular, examples of poor-quality sequences are provided to help researchers to troubleshoot their own Sanger sequencing results.

The pGEM-3Zf(+) control (supplied with BigDye Terminator v3.1 Cycle Sequencing Kit, Applied Biosystems, part number 4337456 or Promega, part number P2271) was used as the template for the examples in this guide (except where noted). Sequence Scanner 2 software (free from ThermoFisher) was used to view the results (.ab1 files) and capture images for this guide. There are many options available to view Sanger Sequencing (.ab1) files. The Genomics Core does not recommend a particular software. You will have to determine which software works best for your situation. A non-exhaustive list of available software is available on page 27.

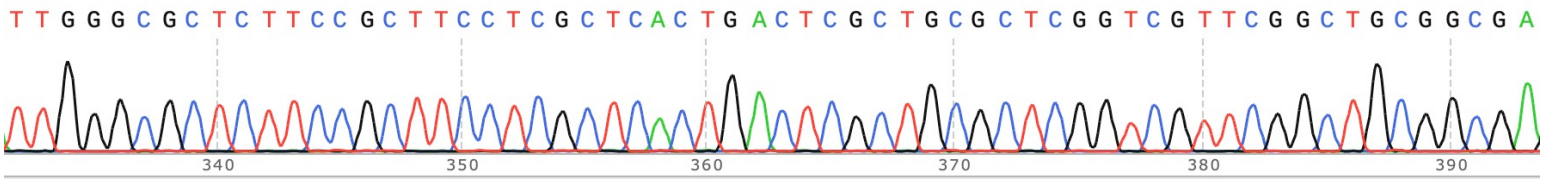


Table of Contents

Best Practices

The RTSF Genomics Core's Sanger Sequencing sample requirements are briefly described below. For detailed requirements, please see the following webpage: https://rtsf.natsci.msu.edu/genomics/sample-requirements/sanger-sequencing-sample-requirements.aspx

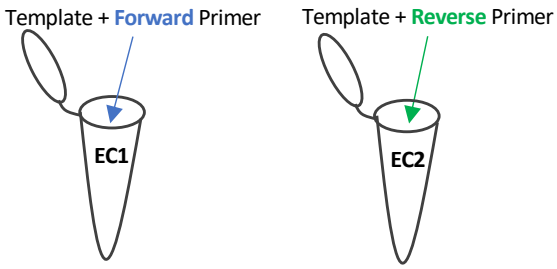
The RTSF Genomics Core accepts samples in which the researcher has already added their own primer, or the researcher can submit the sample and request that the Core add one of six available in-house primers (M13 forward, M13 reverse, SP6, T7, T7 terminator, or T3).

Fluorometric quantification, such as Qubit, is recommended. NanoDrop concentrations are inaccurate and often overestimate sample concentration by 10-fold or more. The most common culprit of poor sequencing results is insufficient template. The recommended DNA and primer concentrations are given in the table below.

		When primer is added by RESEARCHER		When primer is added by GENOMICS CORE	
Sample Type & Size	**DNA Mass (ng)***	**Volume of 10µM primer to add (µl)**	**Total Volume (µl)**	**Volume of 10µM primer to add (µl)**	**Total Volume (µl)**
Plasmid					
Single-stranded DNA	200	3	12	0	8
Double-stranded DNA (up to 10 kb)	1000	3	12	0	8
Purified PCR Product					
<100 – 200 bp	2 - 6	3	12	0	8
200 – 500 bp	6 - 15	3	12	0	8
500 – 1000 bp	10 - 40	3	12	0	8
1000 – 2000 bp	20 - 80	3	12	0	8
>2000 bp	80 - 200	3	12	0	8

*Please note that these are general recommendations and optimization may be required by the researcher.

To sequence in both the forward and reverse direction, two reactions are required. One primer is added per reaction.

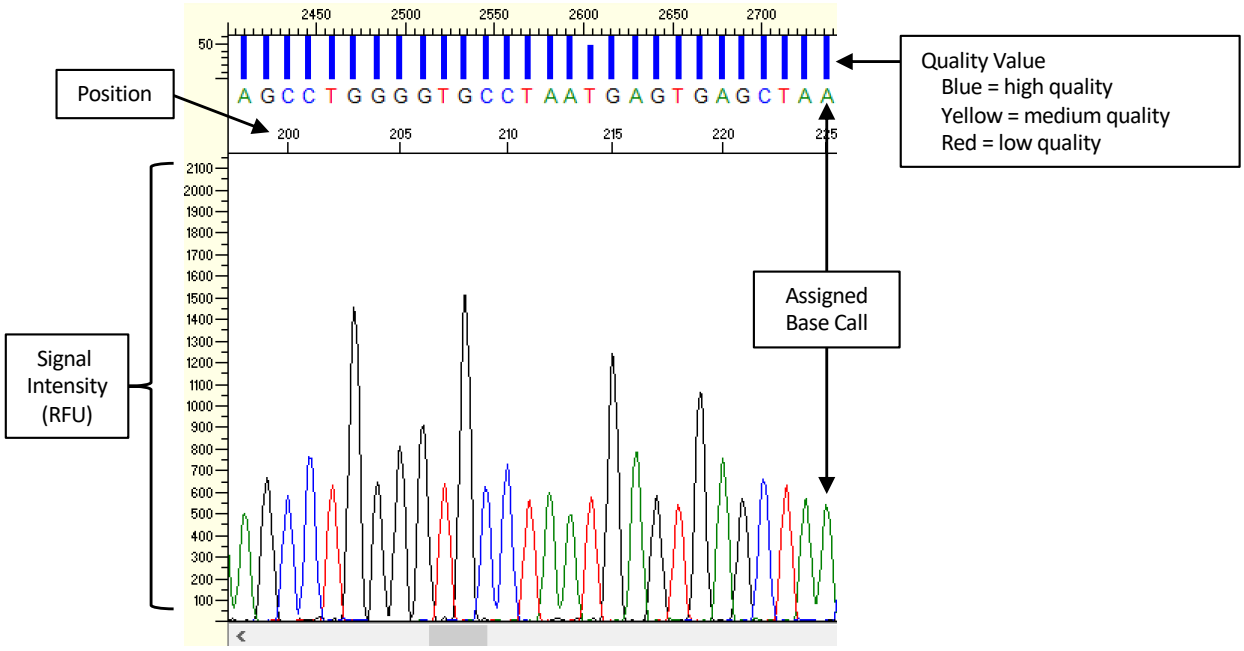


For PCR products it is crucial to purify the products to remove remaining primers, unincorporated nucleotides, enzyme, buffer, and salts. Recommended purification methods include bead or column-based purification, enzymatic cleanup (i.e. ExoSAP-IT), or gel extraction. Additionally, it is important to visualize PCR products on an agarose gel to confirm that only one product has been generated. Reactions containing multiple PCR products (i.e. desired product and an off-target product) will result in unusable data.

The Genomics Core will return the Sanger sequencing result in the form of an .ab1 file. One .ab1 file is returned per reaction. Results are uploaded to Genomics Depot for researchers to download. It is the responsibility of the researcher to download and save all data files. The Genomics Core will retain data for only 6 months from time of collection

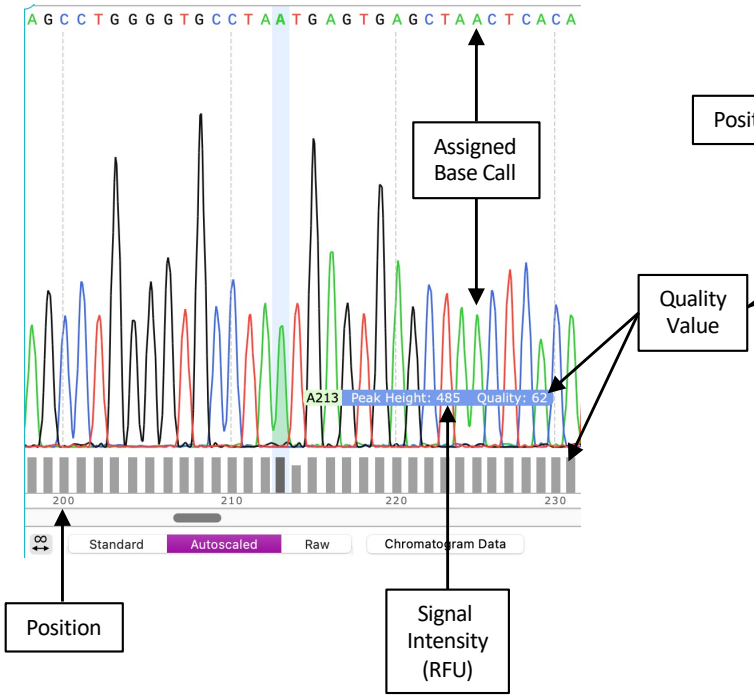
Anatomy of the Electropherogram

The electropherogram below is labeled to show the various aspects that should be reviewed when assessing the quality of the sequence. The sequence immediately below was viewed with Sequence Scanner 2 software. Sequence Scanner 2 was used to collect the images for this guide, except where noted.

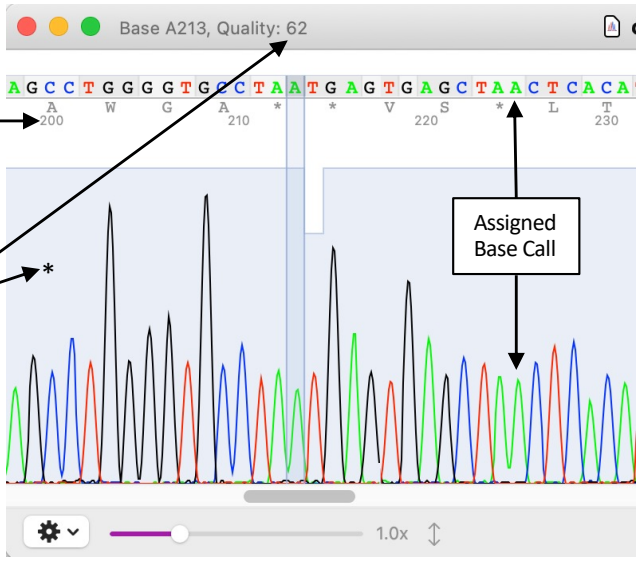


Other software (free or paid) is available for viewing Sanger Sequencing data. The same sequence from the image above was viewed with two other (free) software programs, SnapGene Viewer (left image) and 4Peaks (right image), to show how they compare to Sequence Scanner 2.

SnapGene Viewer



4Peaks



*Light blue background shows quality values.

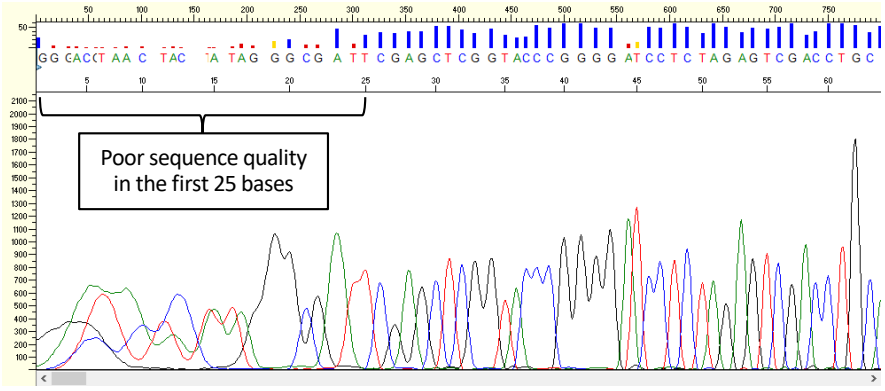
Assessing your Sanger Sequencing results

When reviewing your Sanger sequencing results, it is important to **visually** review the electropherogram, sequence quality, and signal intensity to determine if your results are high quality. Start by viewing the electropherogram.

Basic Expectations of the Electropherogram

The beginning of the sequence

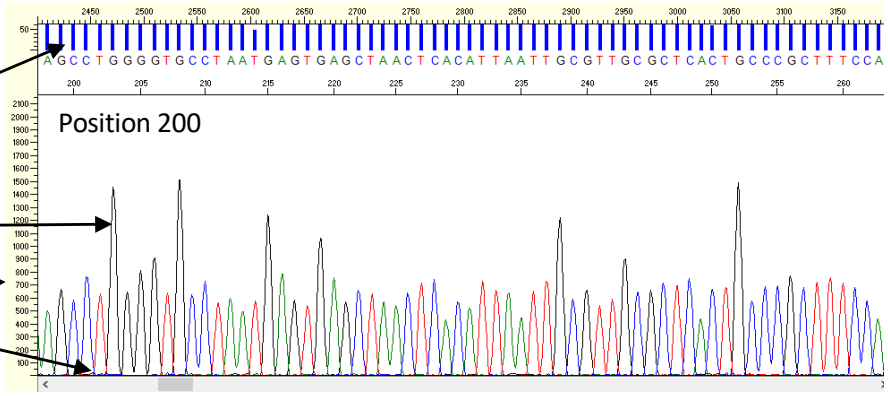
The quality of the first 25 to 35 bases of the sequence is expected to be poor. In this electropherogram, the quality is poor through the first 25 bases. After the initial poor-quality bases, the quality improves, the peaks are sharp and there is no background noise.



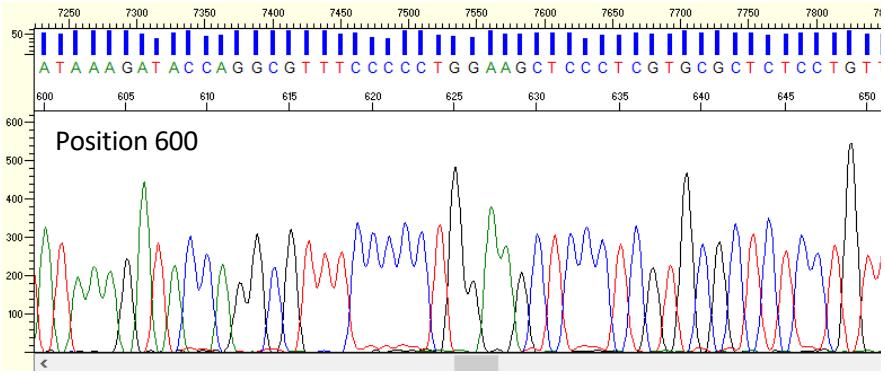
The middle of the sequence

After the first 25-35 bases of a high-quality sample, you should expect:

- high quality values
- sharp peaks
- good signal intensity
- very little to no background noise

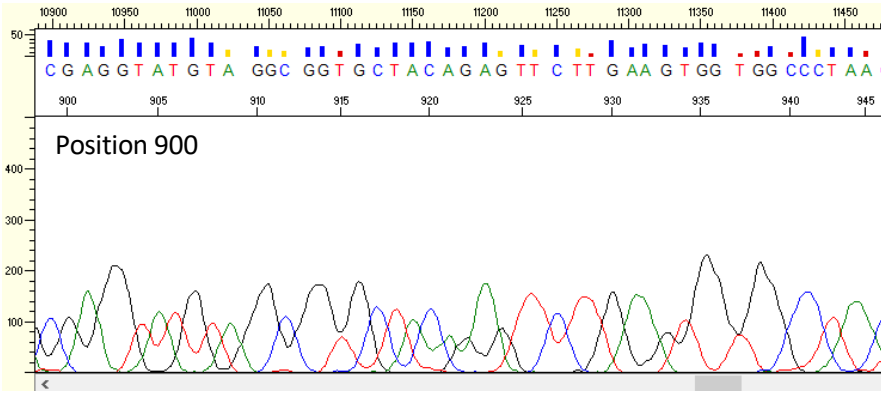


As the sequence continues, the signal intensity decreases. In the figure to the right, the signal intensity has dropped noticeably by position 600. The quality is still high – the peaks are well defined, the signal intensity is >175 RFU, and there is little to no background noise.



The end of the sequence

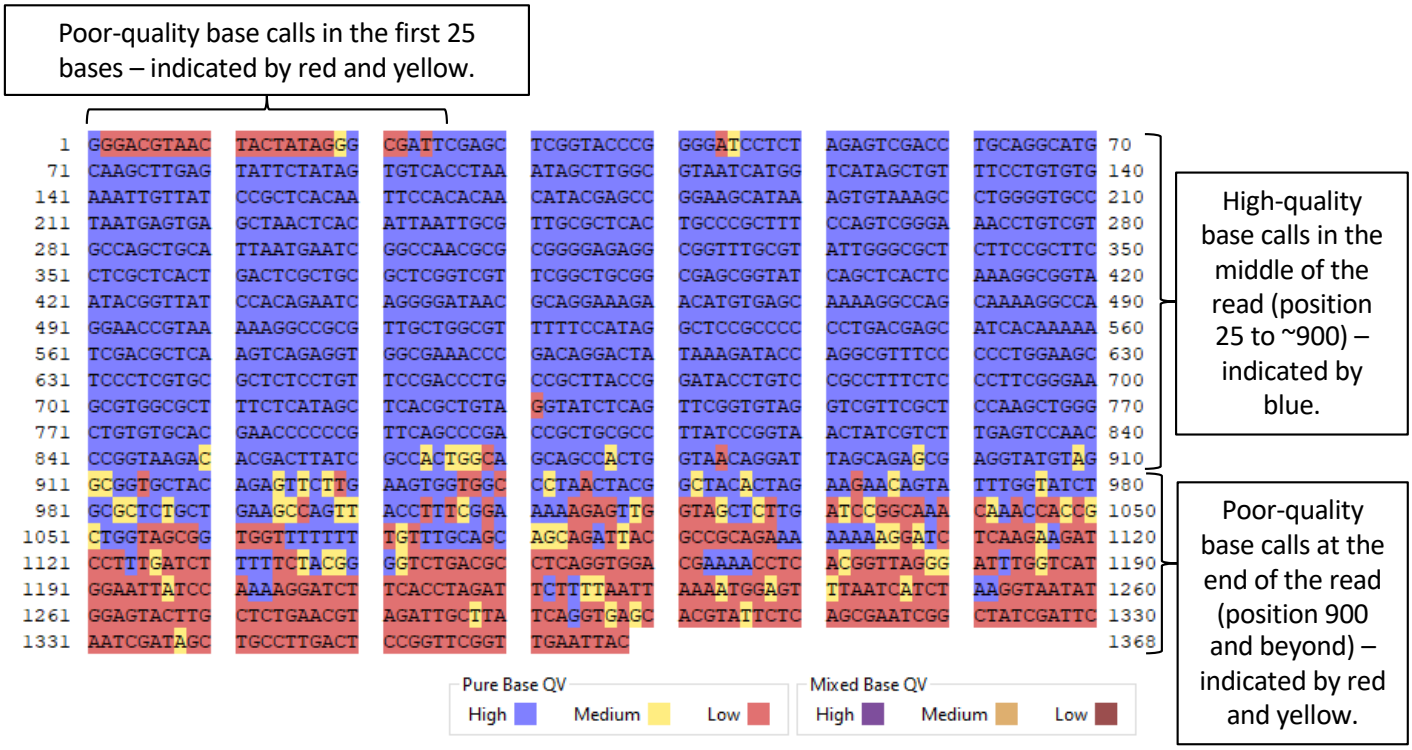
Sequence lengths of 800 to 900 bases are possible for high-quality templates. It is common for the the signal intensity and quality to decrease as the read gets longer. In the electropherogram to the right, the signal intensity is lower than it was at the beginning and middle of the read, the quality is beginning to drop, and the peaks are broader. While sequence beyond 900 bases may be useful, those bases should be interpreted with caution.



Sequence and Base Call Quality

The base call quality can be assessed by looking at the quality bars on electropherogram itself, or if your program allows, you can review the base call quality in conjunction with the sequence as seen below. The sequence immediately below is the same sequence that was viewed in the electropherograms on the previous page.

Example of a high-quality sequence



Example of a poor-quality sequence

The base calls are low to medium quality throughout the entire sequence – indicated by red and yellow. There are very few consecutive high-quality base calls (blue). **This sequence is unusable.**

Signal Intensity

The average raw signal intensities should be between 175 to 10,000 RFU (relative fluorescence units). Data with average raw signal intensities below 175 RFU or above 10,000 RFU often result in poor-quality data and should not be used. The example on the left has signal intensities in the acceptable range. The example on the right has very low signal intensities – this sequence data was unusable.

Good Signal Intensities

Data Analysis

Basecaller	KB.bcp
Basecaller Version	KB 1.4.1.8
Mobility File	KB_3730_POP7_BDTv3.mob
Basecall Date/Time	2023-09-05 07:28:20 -04:00
Total # of Scans Collected	29961
Basecall Start Scan#	2183
Basecall Stop Scan#	25953
Peak 1 Scan#	2181
Base Spacing	20.1
Average Raw Signal Intensity	A(890),C(1324),G(664),T(1411)
Average Noise	A(7) , C(7) , G(6) , T(8)
Average Raw Signal to Noise Ratio	A(121) , C(193) , G(117) , T(167)
Trace Score	39
Contiguous Read Length	1001
QV20+ (# Bases w QV >=20)	1002

Low/Poor Signal Intensities

Data Analysis

Basecaller	KB.bcp
Basecaller Version	KB 1.4.2.5
Mobility File	KB_3730_POP7_BDTv3.mob
Basecall Date/Time	2023-10-09 07:25:24 -05:00
Total # of Scans Collected	29961
Basecall Start Scan#	2260
Basecall Stop Scan#	25789
Peak 1 Scan#	2258
Base Spacing	20.18
Average Raw Signal Intensity	A(56),C(68),G(39),T(83)
Average Noise	A(9) , C(10) , G(7) , T(13)
Average Raw Signal to Noise Ratio	A(5) , C(6) , G(5) , T(6)
Trace Score	13
Contiguous Read Length	5
QV20+ (# Bases w QV >=20)	106

Quality Values

A quality value (QV) is given for each base. The QV is the probability of a basecall error. For example, a QV of 20 means that the error rate is 1% for that basecall.

Quality Value = $-10\log_{10}(Pe)$

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

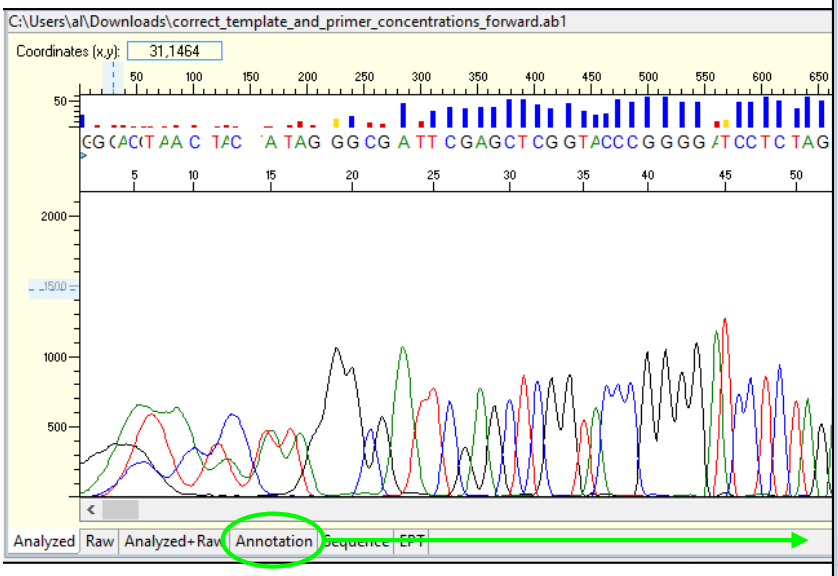
Size and color of QVs bars are identical for QVs 50-99

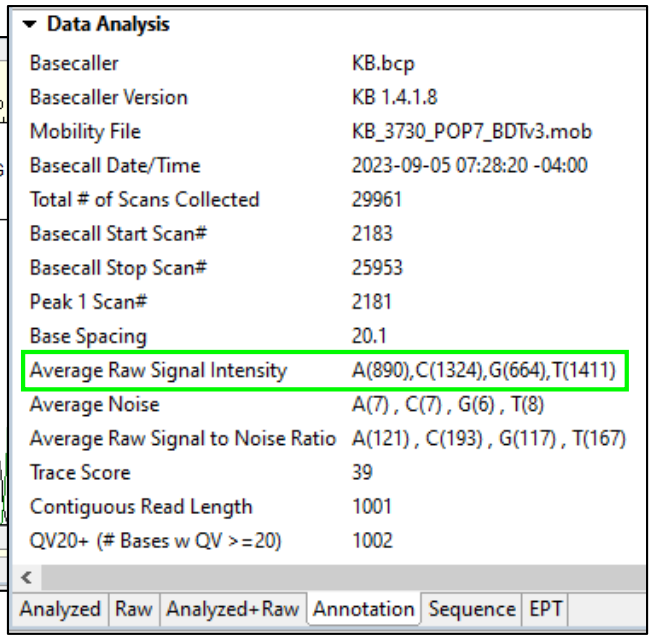
QV	*Pe*	QV	*Pe*	QV	*Pe*
1	79%	21	0.790%	41	0.0079%
2	63%	22	0.630%	42	0.0063%
3	50%	23	0.500%	43	0.0050%
4	39%	24	0.390%	44	0.0039%
5	31%	25	0.310%	45	0.0031%
6	25%	26	0.250%	46	0.0025%
7	20%	27	0.200%	47	0.0020%
8	15%	28	0.150%	48	0.0015%
9	12%	29	0.120%	49	0.0012%
10	10%	30	0.100%	50	0.0010%
11	7.9%	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%		

Where do I find the signal intensity?

Sequence Scanner 2

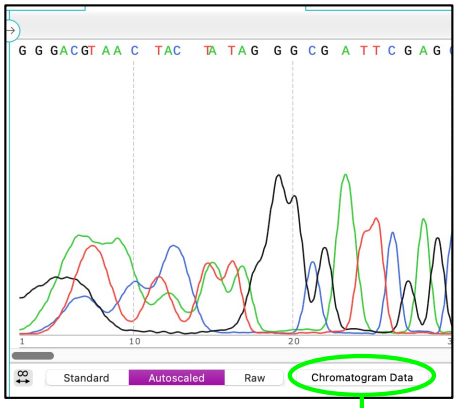
Open the file in Sequence Scanner 2; click “Annotation” at the bottom of the screen; and the window will switch to the annotation view.

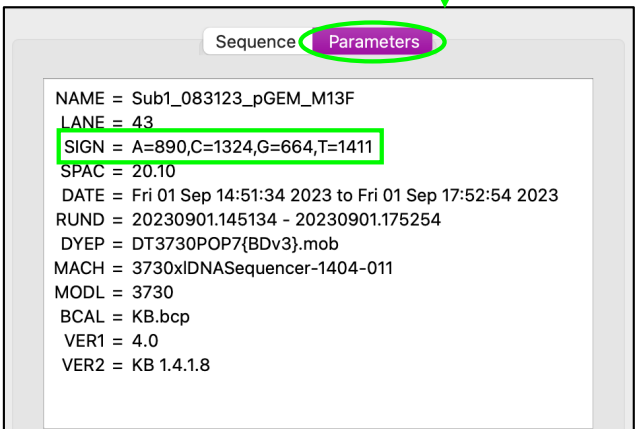




SnapGene Viewer

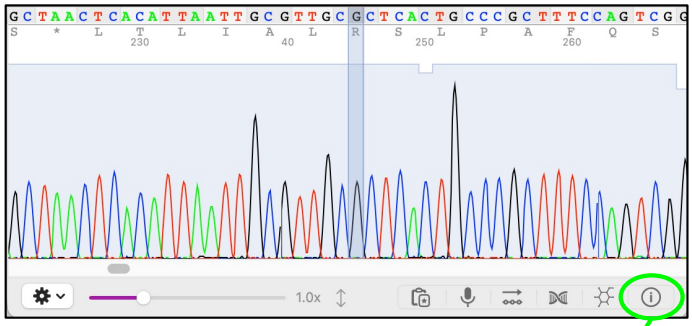
Open the file in SnapGene Viewer; click “Chromatogram Data” at the bottom of the window. Click “Parameters” when the new window opens.

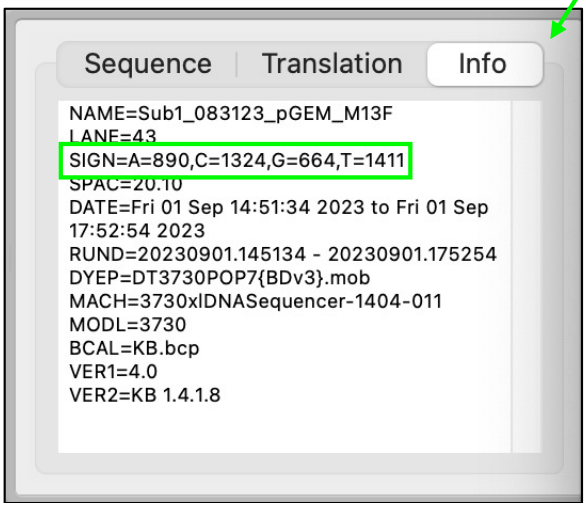




4Peaks

Open the file in 4Peaks; click the “i” icon in the bottom right corner and a new window will open.





Example 1: Correct template and primer concentration (plasmid)

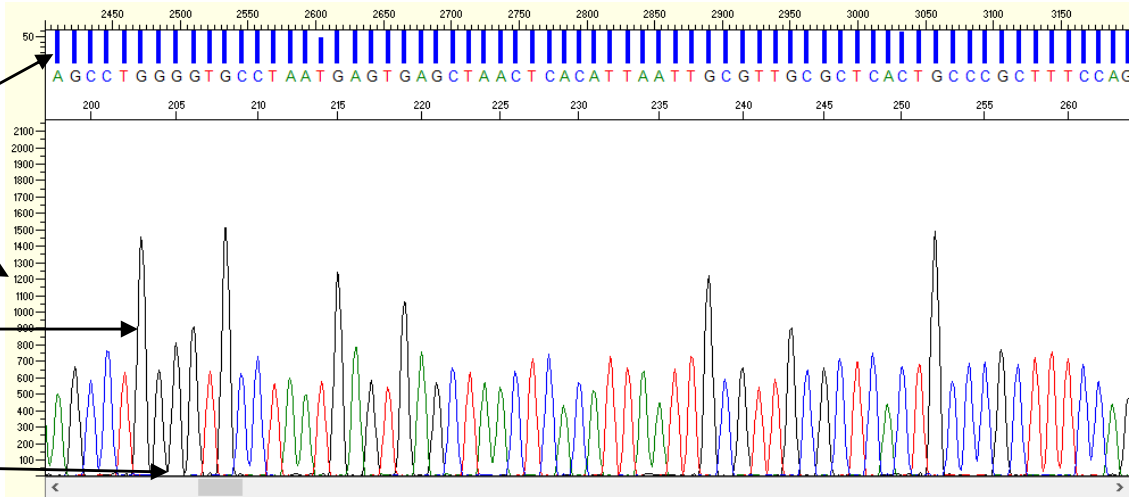
This is an example of a good sequencing reaction that resulted in high quality data.
The template source is the pGEM-3Zf(+) control; a double stranded DNA plasmid that is 3,197 bp.

Template Mass	Number of Primers Added	Volume of 10µM Primer	Total Volume	File Name
1000 ng	1	3 µl	12 µl	1_correct_plasmid.ab1

Position 200

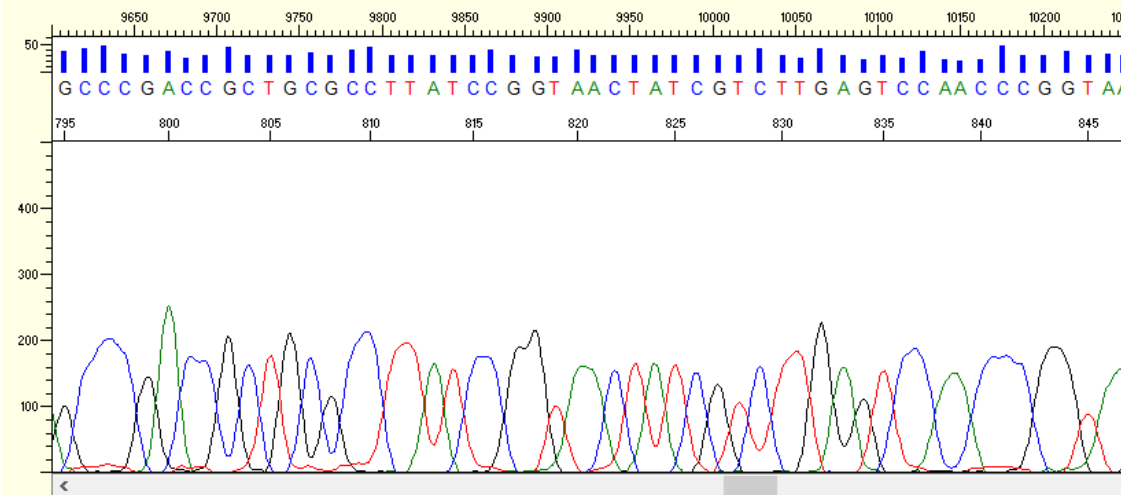
The criteria that make this a high-quality sequence are:

- Quality values are high throughout the read
- Signal intensity is > 175 RFU
- Each position only has one peak (i.e. one assigned base call)
- There is no background noise (i.e. peaks along the bottom of the electropherogram)



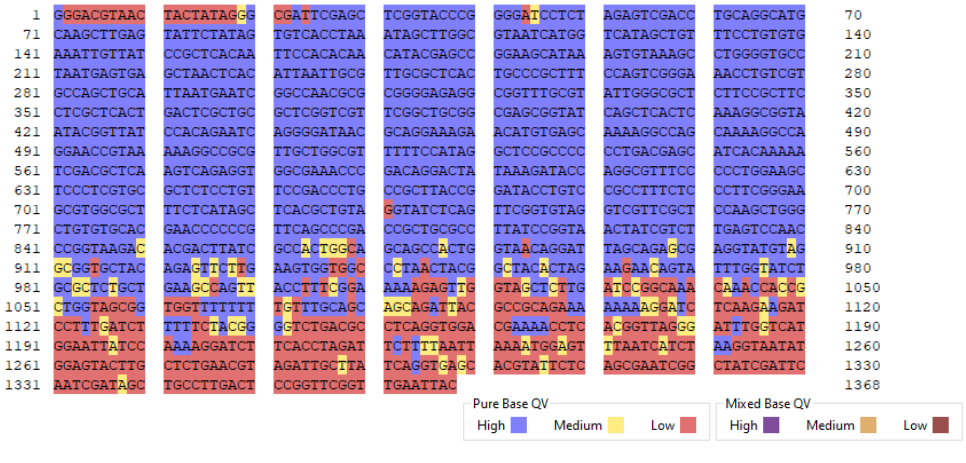
Position 795

At position 795, the signal intensity has decreased, but continues to be ~200 RFU. The quality values continue to be high, each position has only one peak (i.e. assigned base call) and there is no background noise.



Sequence Annotation and Quality

The base calls in blue are high-quality bases with only one base being called.
The sequence quality begins to drop at base 909 (yellow and red bases).



Signal Intensity

Acceptable range: 664 to 1411 RFU.

Data Analysis

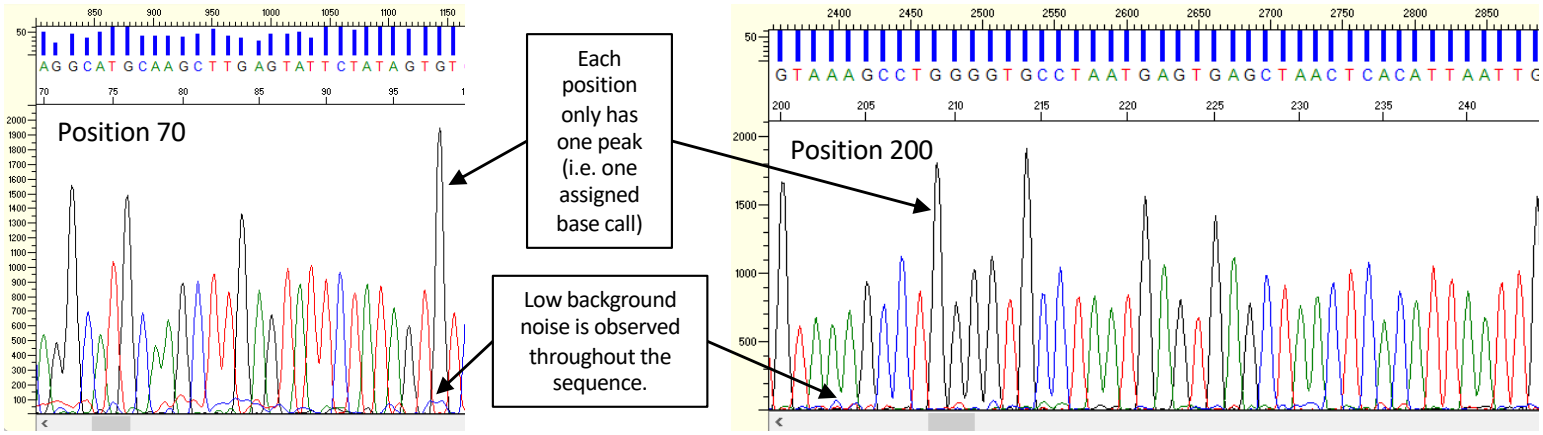
Basecaller	KB.bcp
Basecaller Version	KB 1.4.1.8
Mobility File	KB_3730_POP7_BDTv3.mob
Basecall Date/Time	2023-09-05 07:28:20 -04:00
Total # of Scans Collected	29961
Basecall Start Scan#	2183
Basecall Stop Scan#	25953
Peak 1 Scan#	2181
Base Spacing	20.1
Average Raw Signal Intensity	A(890),C(1324),G(664),T(1411)
Average Noise	A(7) , C(7) , G(6) , T(8)
Average Raw Signal to Noise Ratio	A(121) , C(193) , G(117) , T(167)
Trace Score	39
Contiguous Read Length	1001
QV20+ (# Bases w QV >=20)	1002

Example 2a: 1/10th recommended template (plasmid)

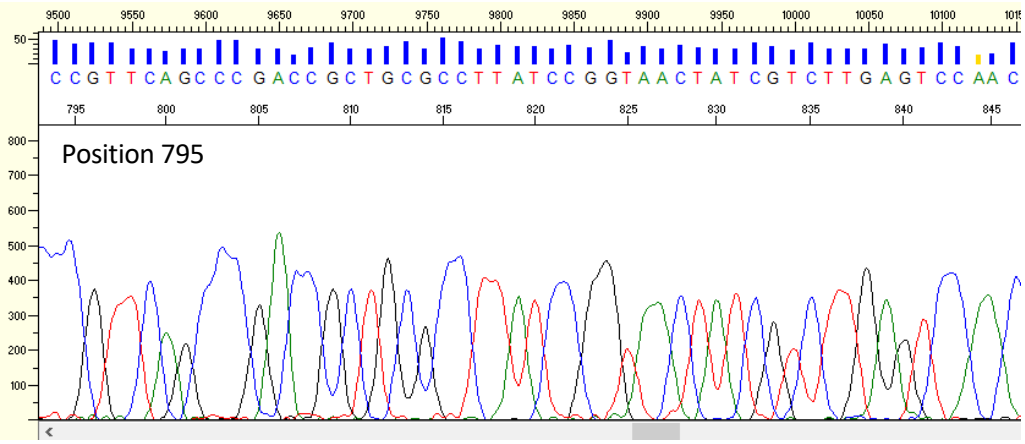
The template was diluted to 1/10th the recommendation. The primer was added at the correct concentration.
The template source is the pGEM-3Zf(+) control; a double stranded DNA plasmid that is 3,197 bp.

Template Mass	Number of Primers Added	Volume of 10µM Primer	Total Volume	File Name
100 ng	1	3 µl	12 µl	2a_0.1_template_plasmid.ab1

The quality of this sequence is still good despite having a lower than recommended template mass. There are two characteristics that indicate that the template mass was lower than recommended. 1.) There is noise; these are the peaks along the bottom of the electropherogram. This is most pronounced around position 70 of this example, but it can be seen throughout the read. 2.) The signal intensities are lower than was observed in Example 1 (correct template), however, the signal intensities are still above the threshold of 175 RFU.

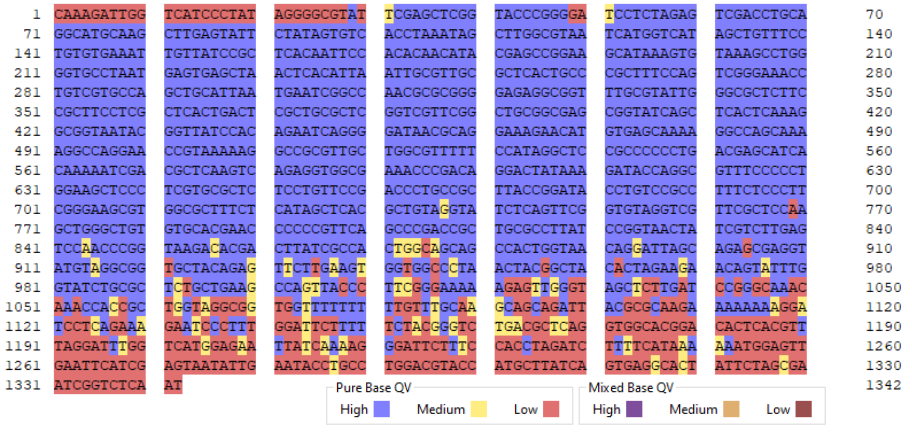


At position 795, the signal intensity has decreased, but continues to be >175 RFU. The quality continues to be high, and each position has only one peak (i.e. one assigned base call).



Sequence Annotation and Quality

The base calls in blue are high quality with only one base being called. The sequence quality begins to drop around base 904 (yellow and red bases). The quality of this sequence is high, like Example 1 (correct template).



Signal Intensity

The Average Raw Signal Intensities are in the acceptable range.

A(194), C(232), G(230), T(241)

Sequence is good.

No changes needed.

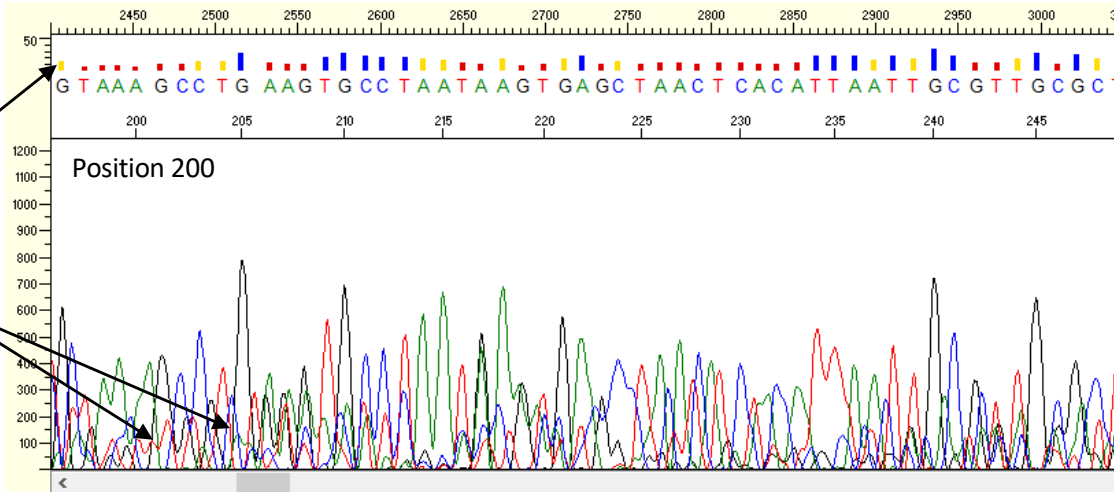
Example 2b: 1/100th recommended template (plasmid)

The template was diluted to 1/100th the recommendation. The primer was added at the correct concentration.
The template source is the pGEM-3Zf(+) control; a double stranded DNA plasmid that is 3,197 bp.

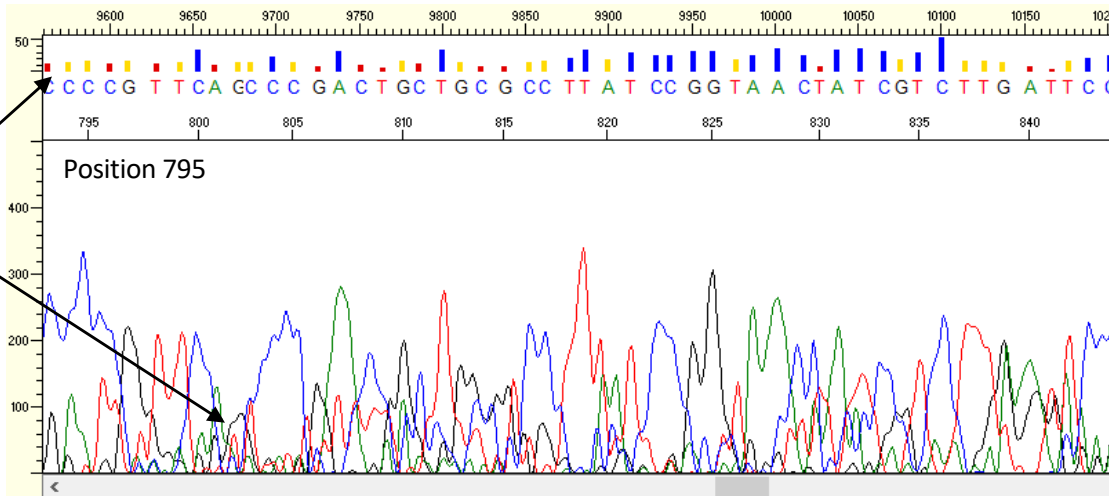
Template Mass	Number of Primers Added	Volume of 10µM Primer	Total Volume	File Name
10 ng	1	3 µl	12 µl	2b_0.01_template_plasmid.ab1

This is a poor-quality sequence that did not generate usable data.

- Quality values are low throughout the sequence.
- The background signal is high relative to the primary signal, which contributes to low quality base calls.

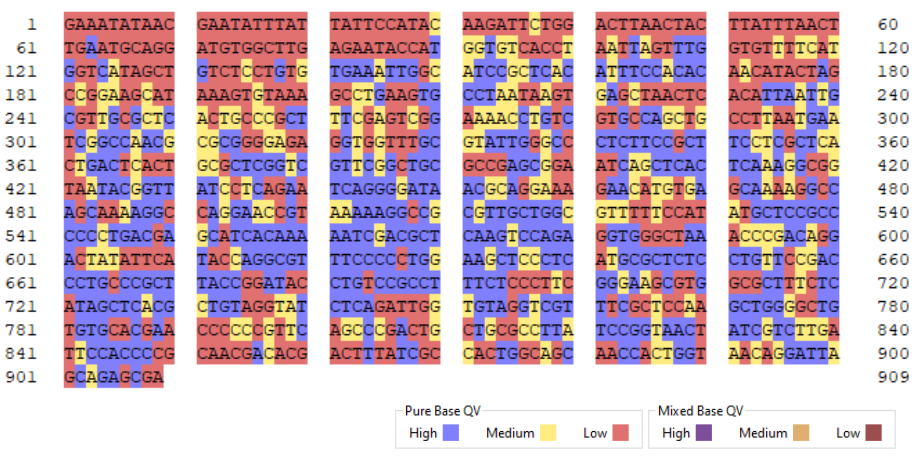


At position 795, the quality continues to be poor with multiple peaks per position as was seen at position 200 above.



Sequence Annotation and Quality

The base calls in yellow and red are medium and low quality, respectively. Most of the sequence has medium to low quality base calls with a few bases that are high quality, but noncontinuous. **This sequence is poor quality and should not be used.**



Signal Intensity

The Average Raw Signal Intensities are below the cut off (175 RFU).

A(44), C(47), G(39), T(54)

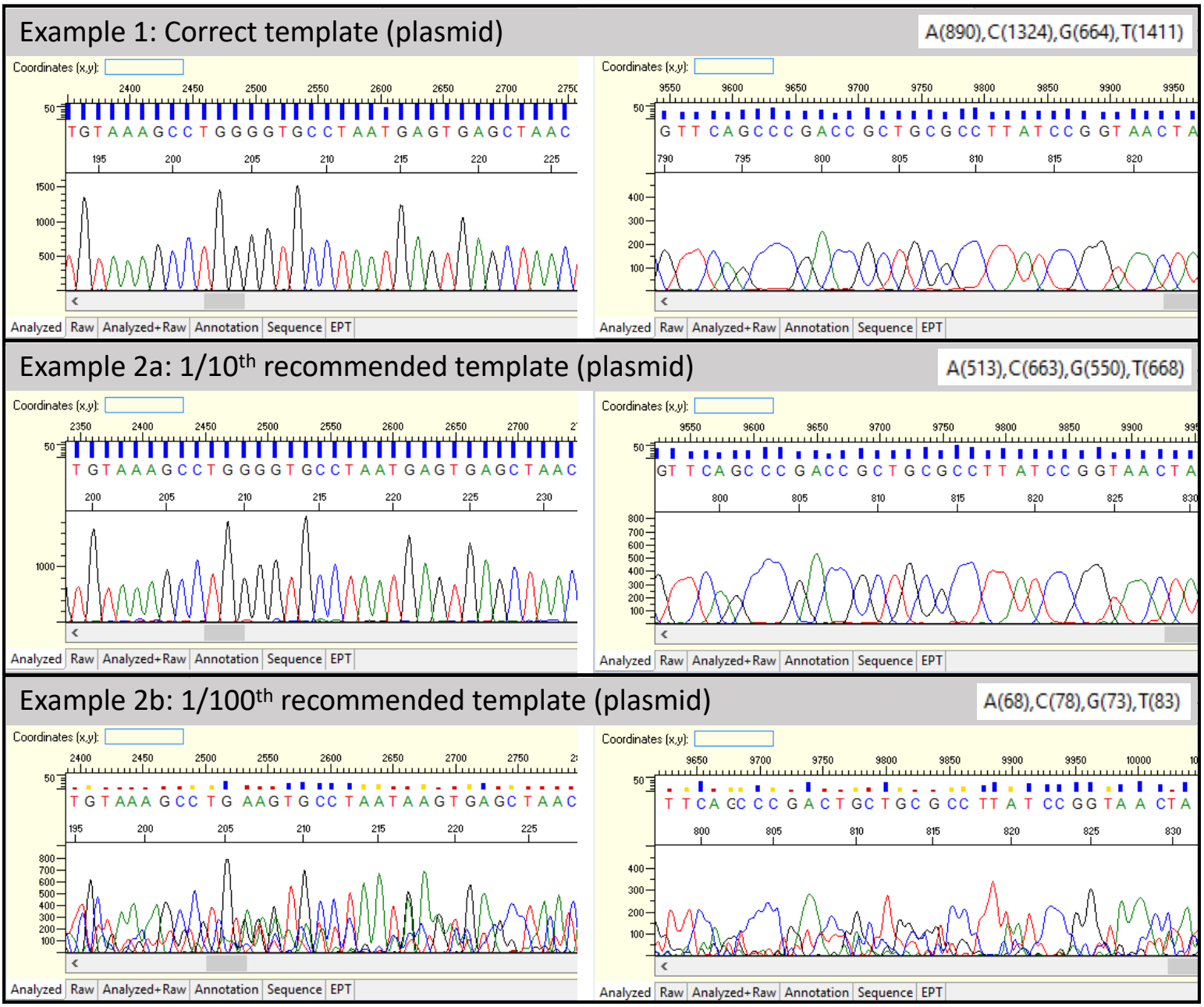
Summary

Sequence is poor quality and should not be used.

To improve results, quantify sample with fluorometric method, such as Qubit. Submit fresh sample using recommended mass. Note that the NanoDrop is not an accurate method for quantifying nucleic acids.

Comparison & Summary: Samples with too little template (plasmid)

The panel below shows how the quality of the sequence decreases as the amount of template decreases. The correct template mass (example 1) is on top, 1/10th the recommended template mass (example 2a) is in the middle, and 1/100th the recommended template mass (example 2b) is on the bottom. Two regions of the electropherogram are displayed: position 200 on the left and position 795 on the right. The average raw signal intensities are also displayed for each example.



Examples 1 and 2a are good quality sequences. While example 2a used 1/10th the recommended template and still achieved a good quality result, it should be noted that this is not necessarily expected or recommended. However, it does demonstrate that high quality templates may generate acceptable quality data even when deviating from recommended inputs. Example 2b is a poor-quality sequence and should not be used. Refer to pages 9 – 11 for a more detailed summary of each example.

Example 3: Correct template and primer concentration (PCR product)

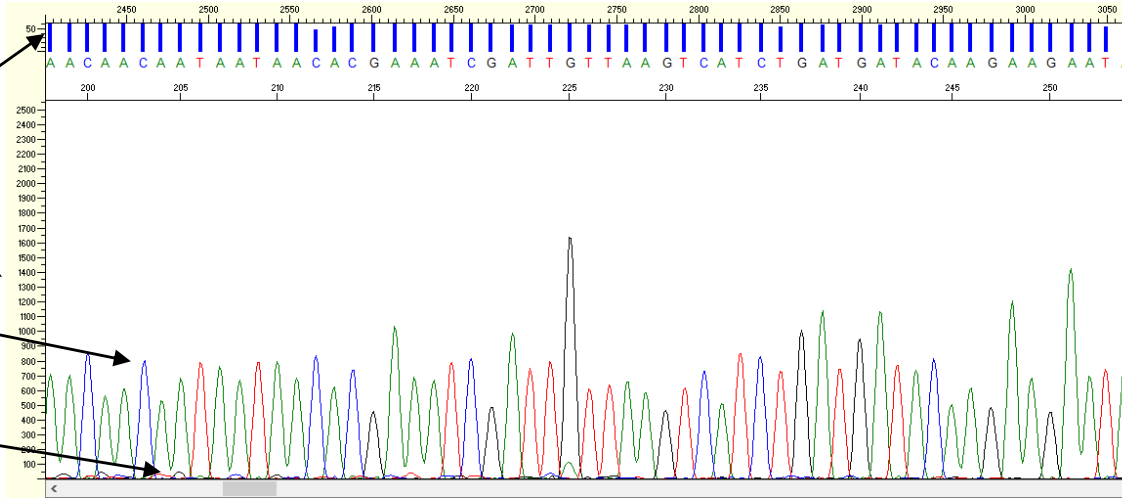
This is an example of a good sequencing reaction that resulted in high quality data.
The template source is a 408 bp double stranded PCR product that was purified with ExoSAP-it.

Template Mass	Number of Primers Added	Volume of 10µM Primer	Total Volume	File Name
15 ng	1	3 µl	12 µl	3_correct_PCR_product.ab1

Position 200

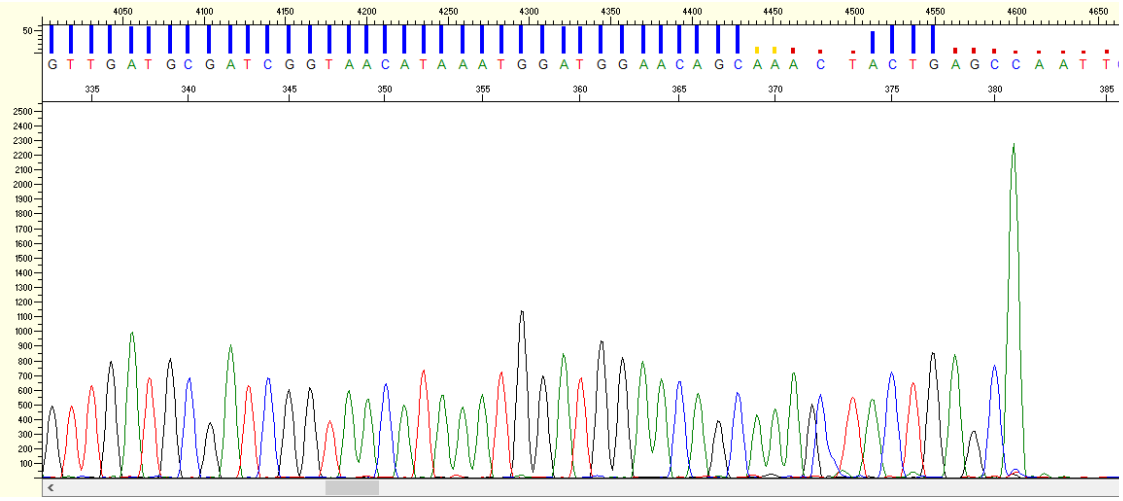
The criteria that make this a high-quality sequence are:

- Quality values are high throughout the read
- Signal intensity is > 175 RFU
- Each position only has one peak (i.e. one assigned base call)
- There is very little to no background noise (i.e. peaks along the bottom of the electropherogram)

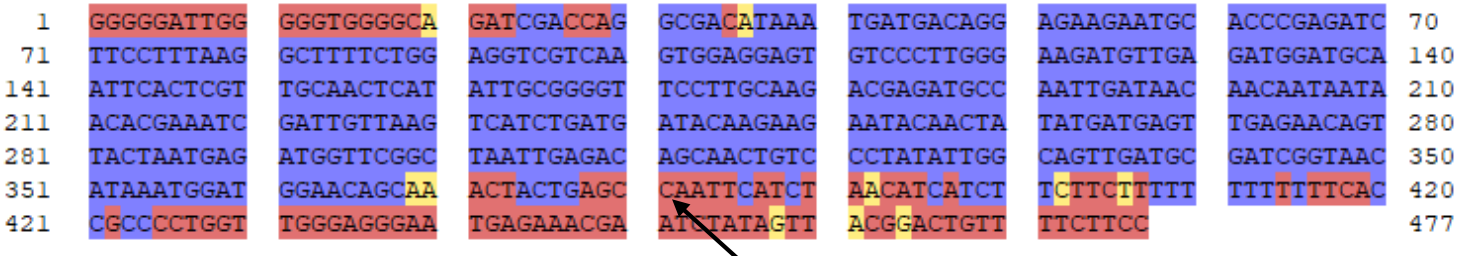


End of sequence

At position 381 the sequence abruptly ends. This is expected because the template is a PCR product.



Sequence Annotation and Quality



The basecalls end at 381 in the electropherogram. Basecalls beyond this point should be ignored.

Signal Intensity is in the acceptable range

A(1522), C(863), G(950), T(1458)

Example 4a: 1/10th recommended template (PCR product)

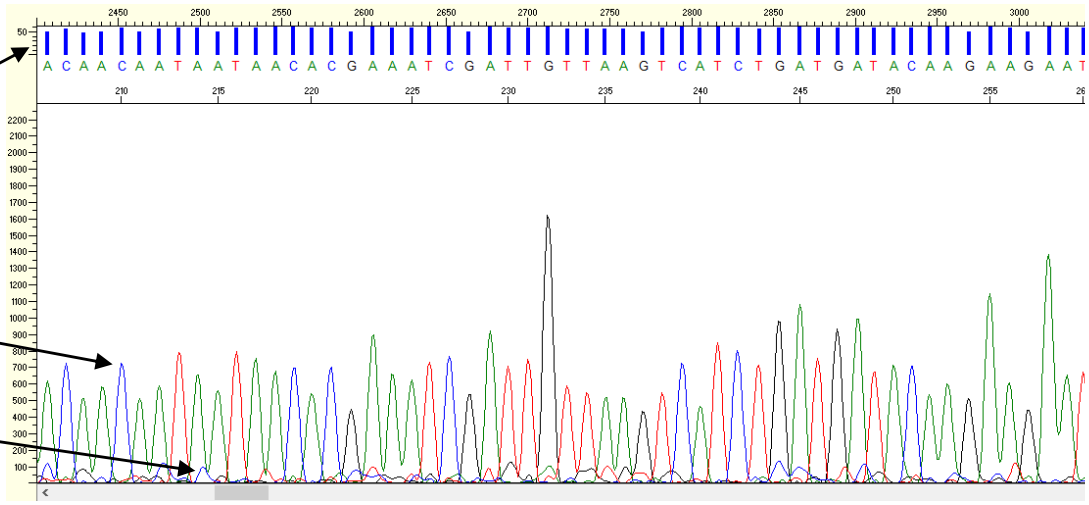
The template was diluted to 1/10th the recommendation. The primer was added at the correct concentration. The template source is a 408 bp double stranded PCR product that was purified with ExoSAP-it.

Template Mass	Number of Primers Added	Volume of 10µM Primer	Total Volume	File Name
1.5 ng	1	3 µl	12 µl	4a_0.1_template_PCR_product.ab1

Position 210

The criteria that make this an acceptable quality sequence are:

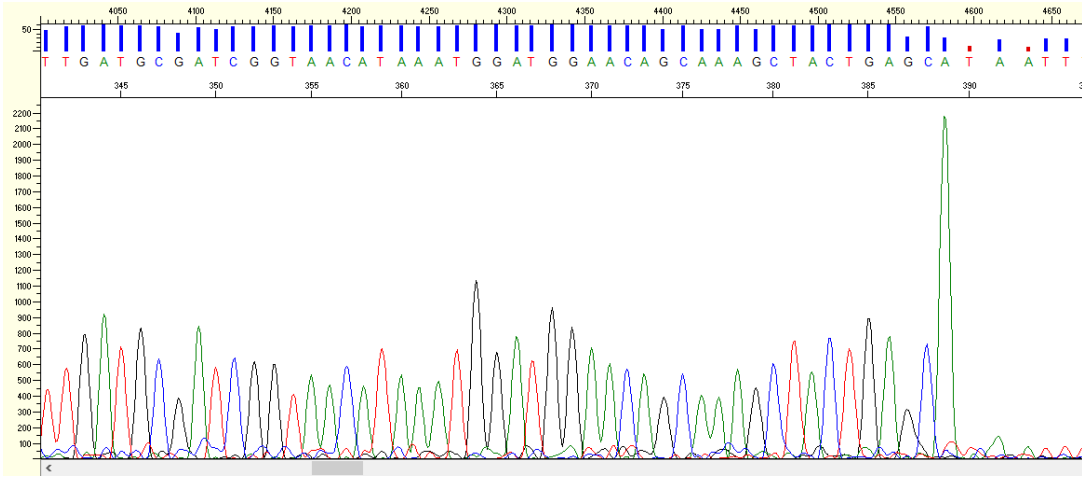
- Quality values are high throughout the read
- Signal intensity is > 175 RFU
- Each position only has one peak (i.e. one assigned base call)
- There is some background noise (i.e. peaks along the bottom of the electropherogram), but the true base call can be determined.



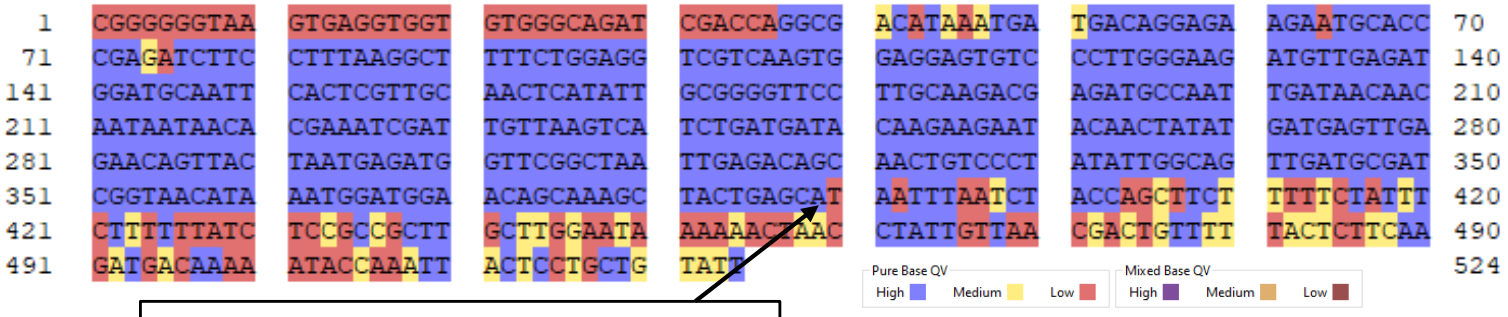
End of sequence

Quality and signal intensity continue to be high until the end of the read. The background noise continues as well, but the base calls can still be determined.

At position 389 the sequence abruptly ends. This is expected because the template is a PCR product.



Sequence Annotation and Quality



The basecalls end at 389 in the electropherogram. Basecalls beyond this point should be ignored.

Signal Intensity is in the acceptable range

A(204),C(129),G(171),T(185)

Sequence is good.

No changes needed.

Example 4b: 1/100th recommended template (PCR product)

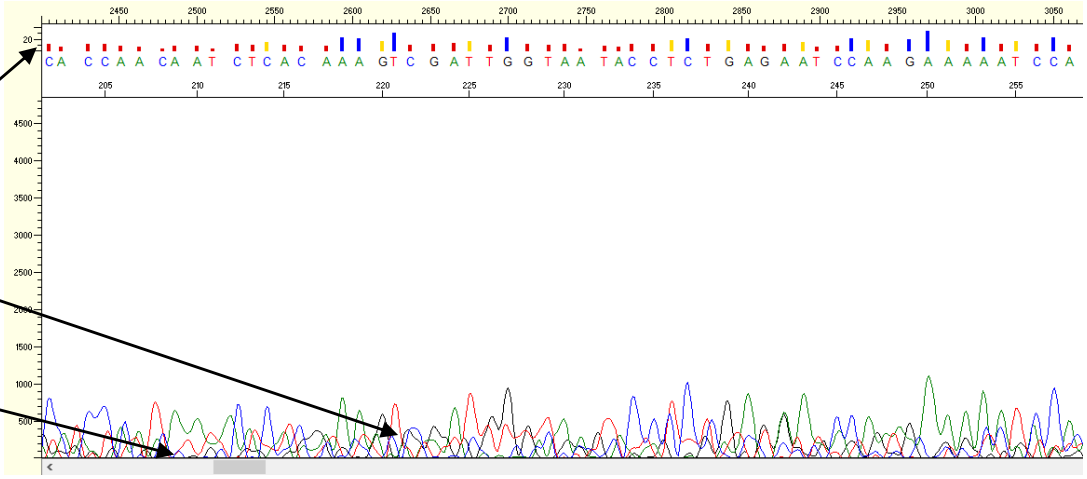
The template was diluted to 1/100th the recommendation. The primer was added at the correct concentration. The template source is a 408 bp double stranded PCR product that was purified with ExoSAP-it.

Template Mass	Number of Primers Added	Volume of 10µM Primer	Total Volume	File Name
0.15 ng	1	3 µl	12 µl	4b_0.01_template_PCR_product.ab1

Position 205

The criteria that make this a poor-quality sequence are:

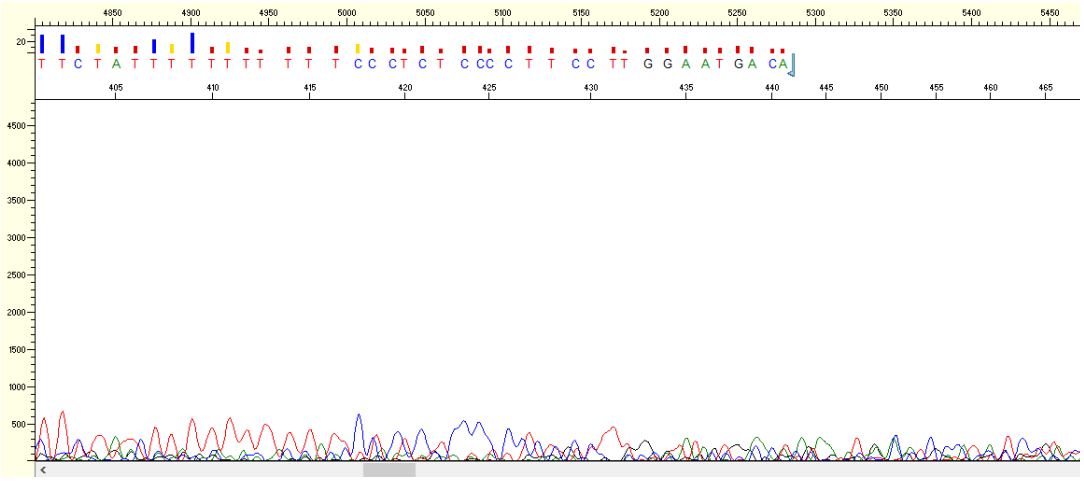
- Quality values are low throughout the read.
- Each position has multiple peaks and/or there is a lot of background noise.



End of sequence

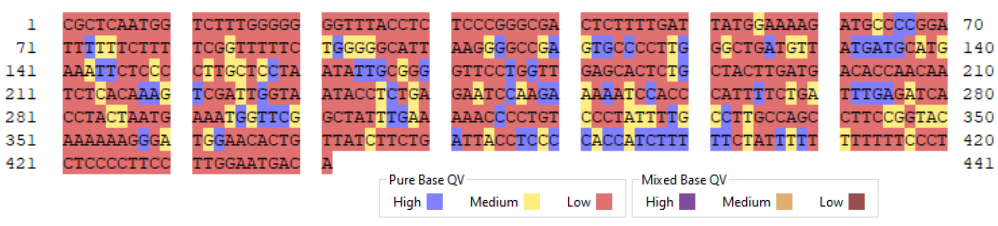
Bases are no longer called after position 440. This is beyond the expected length of the PCR product.

Quality values continue to be poor through the entire read and multiple peaks are called at each position.



Sequence Annotation and Quality

This sequence is made up almost exclusively of low-quality base calls (red). This is a failed sequence and cannot be used.



Signal Intensity is below the acceptable range.

A(61), C(59), G(49), T(74)

Summary

Sequence is poor quality and should not be used.

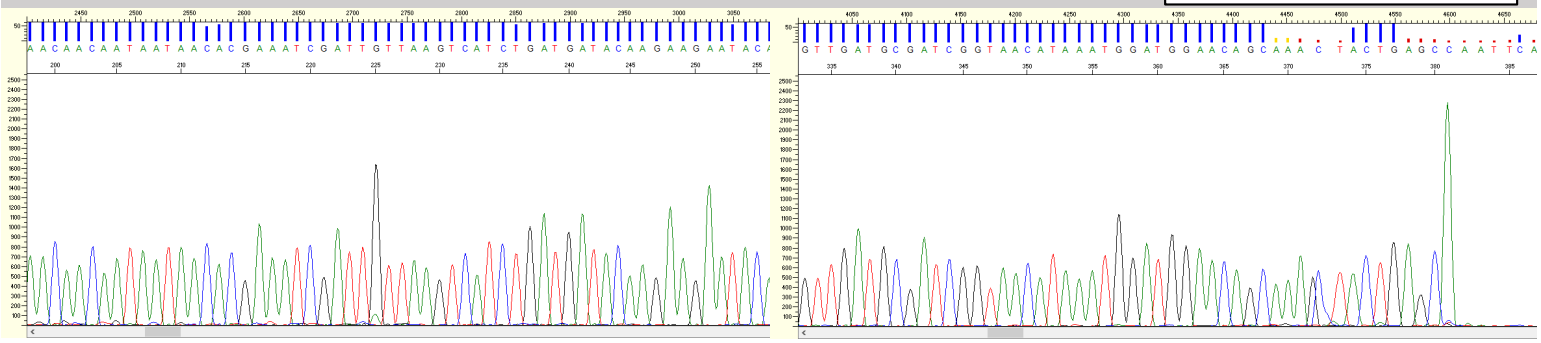
To improve results, quantify sample with fluorometric method, such as Qubit. Submit fresh sample using recommended mass. Note that the NanoDrop is not an accurate method for quantifying nucleic acids.

Comparison & Summary: Too little template (PCR product)

The panel below shows how the quality of the sequence decreases as the amount of template decreases. The correct template mass is on top (example 3), 1/10th the recommended template mass (example 4a) is in the middle, and 1/100th the recommended template mass (example 4b) is on the bottom. Two regions of the electropherogram are displayed: position ~210 on the left and the end of the sequence on the right. The average raw signal intensities are also displayed for each example.

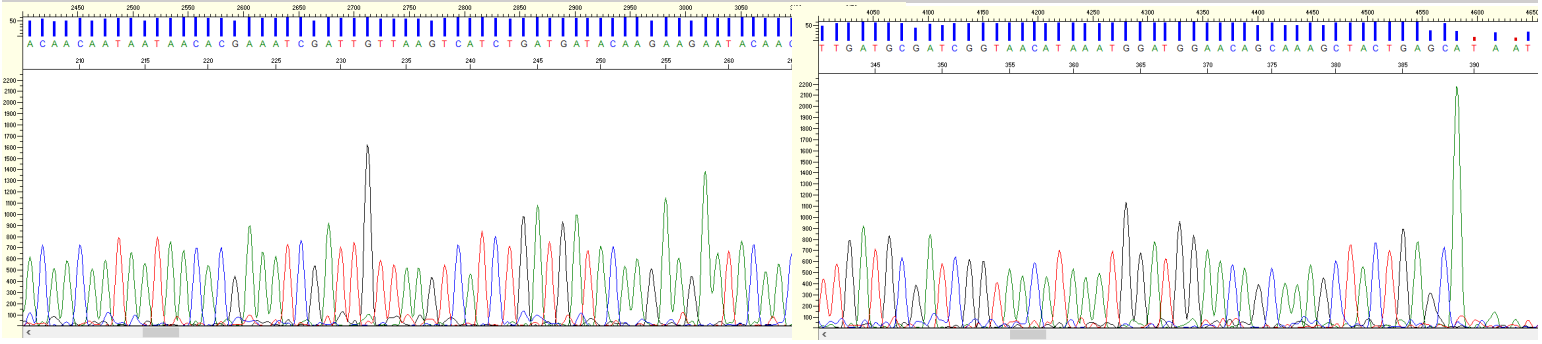
Example 3: Correct template (PCR product)

A(1522), C(863), G(950), T(1458)



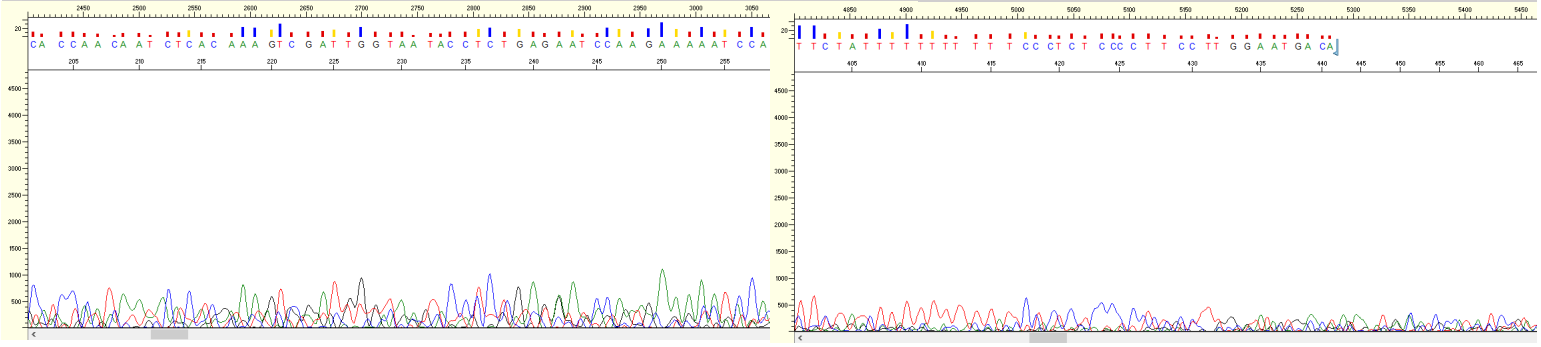
Example 4a: 1.5 ng template (PCR product)

A(204), C(129), G(171), T(185)



Example 4b: 0.15 ng template (PCR product)

A(61), C(59), G(49), T(74)



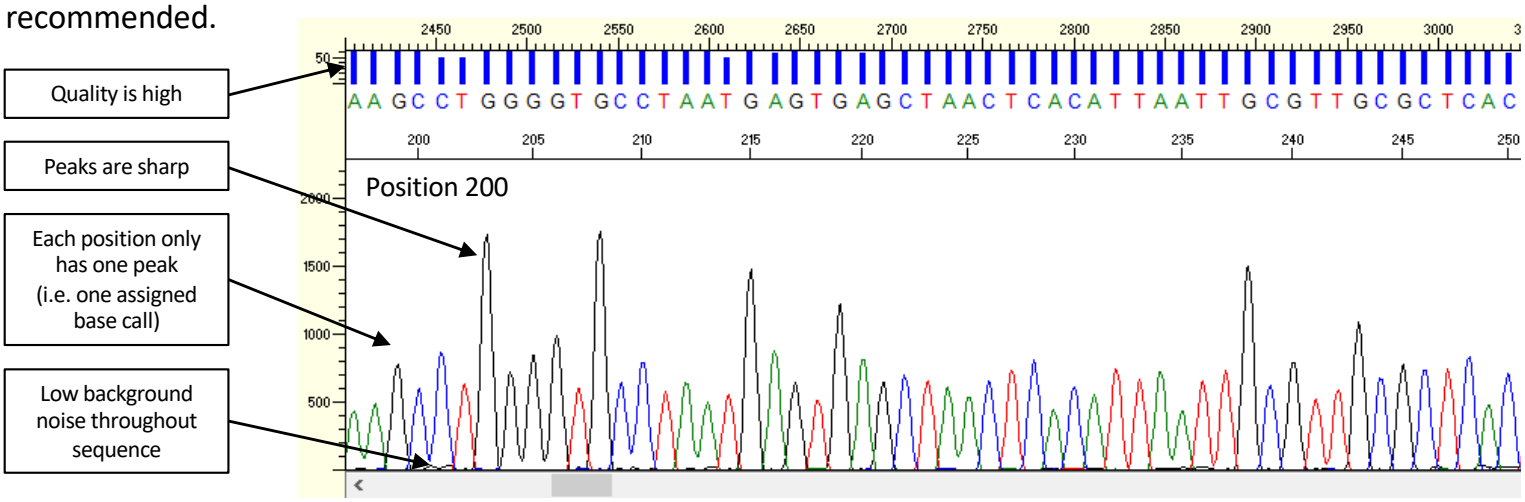
In the examples above you can see that a moderate deviation from the recommended template mass for PCR products generates acceptable results (example 4a). However, providing a very small amount of PCR product results in sequencing failure (example 4b). See pages 13-15 for more details on each of these examples.

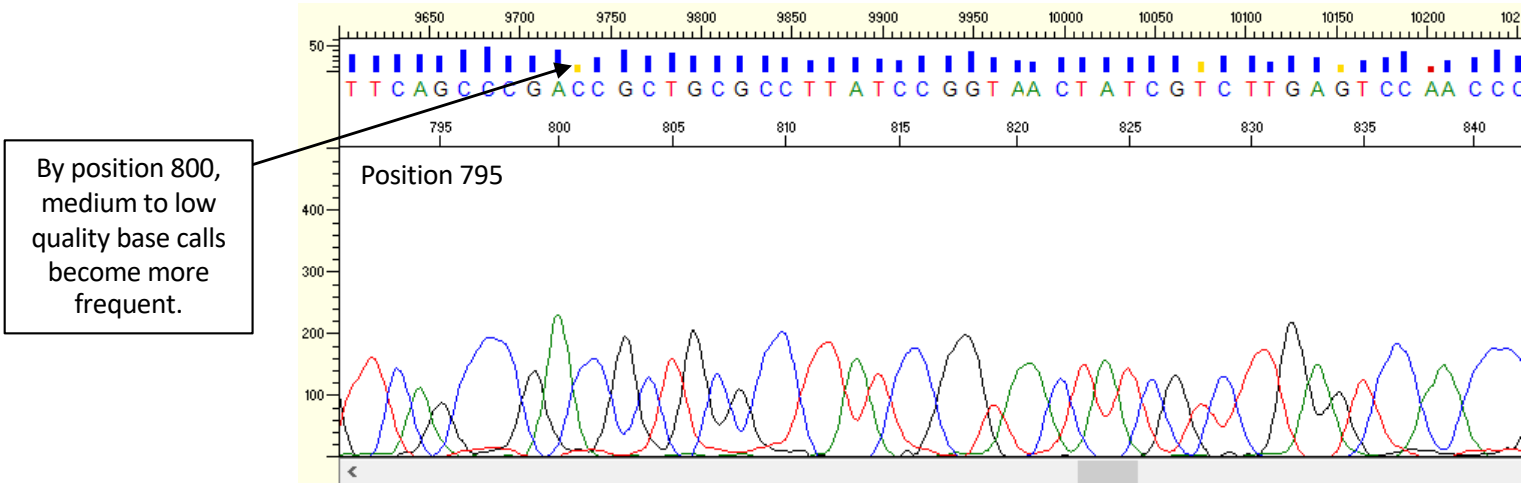
Example 5: 5X recommended template (plasmid)

The template was added at 5X the recommendation. The primer was added at the correct concentration.
The template source is the pGEM-3Zf(+) control; a double stranded DNA plasmid that is 3,197 bp.

Template Mass	Number of Primers Added	Volume of 10µM Primer	Total Volume	File Name
5000 ng	1	3 µl	12 µl	5_5X_template_plasmid.ab1

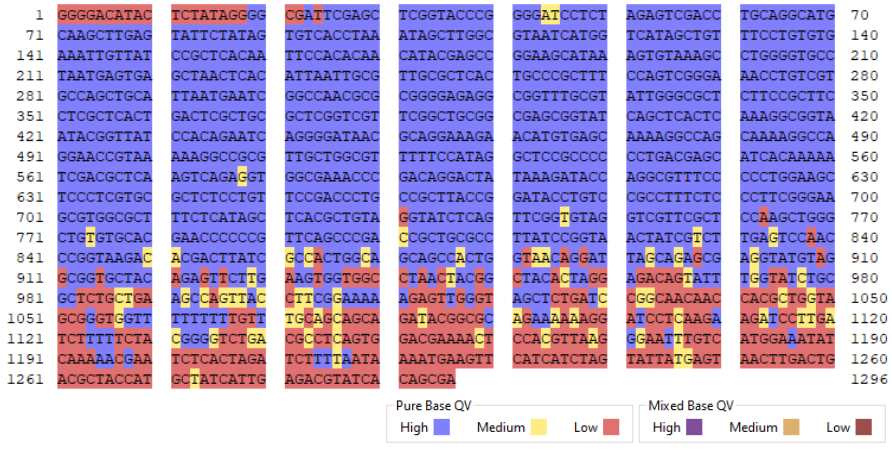
The quality of this sequence is still good despite the template being at a greater mass than is recommended.





Sequence Annotation and Quality

The base calls in blue are high quality with only one base being called. Medium and low-quality values become more frequent around base 850 (yellow and red bases). The quality of this sequence is high, like example 1 (correct template). But, the usable sequence may be slightly shorter.



Signal Intensity

The Average Raw Signal Intensities are in the acceptable range.

A(1818), C(2384), G(890), T(2914)

Sequence is good.

No changes needed.

Example 6: 5X recommended template (PCR product)

The template was added at 5X the recommendation. The primer was added at the correct concentration.
The template source is a 408 bp double stranded PCR product that was purified with ExoSAP-it.

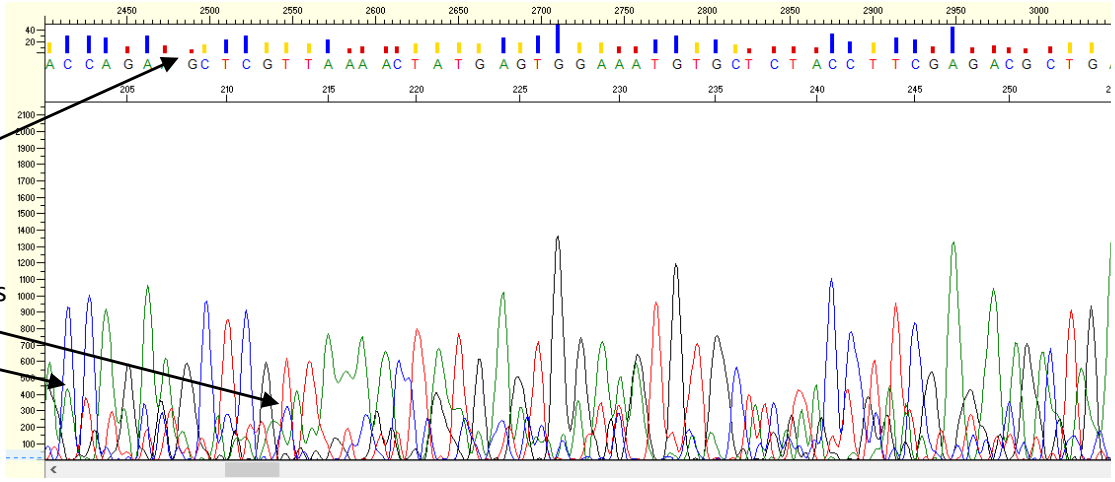
Template Mass	Number of Primers Added	Volume of 10µM Primer	Total Volume	File Name
75 ng	1	3 µl	12 µl	6_5X_template_PCR_product.ab1

The quality of this sequence is poor.

Position 205

The criteria that make this a poor-quality sequence are:

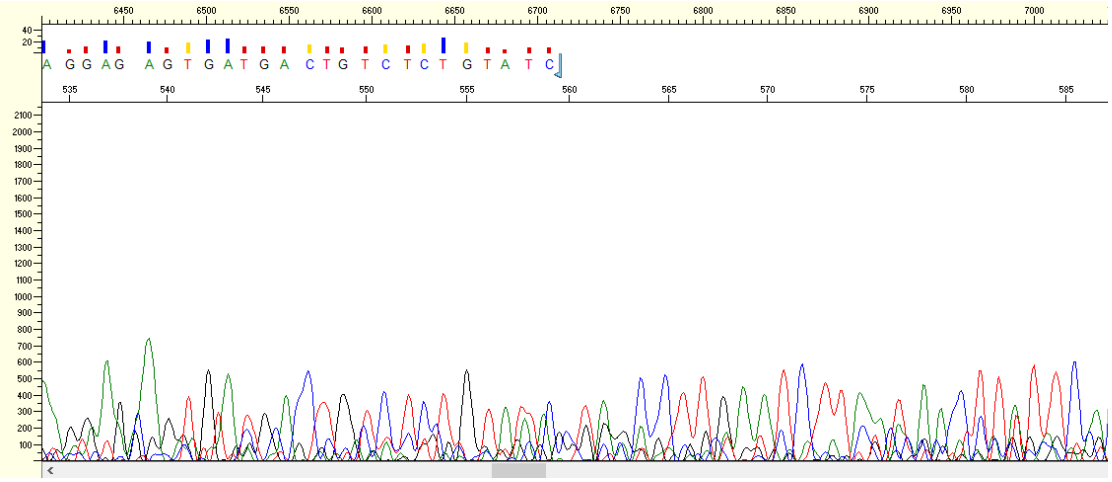
- Quality values are low throughout the read.
- Each position has multiple peaks and/or there is a lot of background noise.



End of sequence

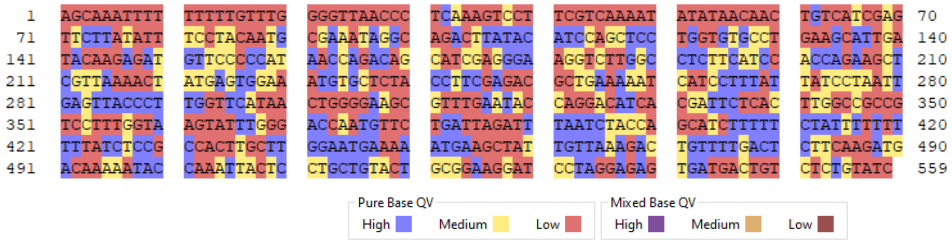
Bases are no longer called after position 559, however, this is beyond the expected length of the PCR product.

Quality values continue to be poor through the entire read and multiple peaks/noise are present at most, if not all, positions.



Sequence Annotation and Quality

This sequence has mostly medium- to low-quality base calls.
This is a failed sequence and cannot be used.



Signal Intensity is below the acceptable range.

A(57), C(60), G(42), T(77)

Summary

Sequence is poor quality and should not be used.

To improve results, quantify sample with fluorometric method, such as Qubit. Submit fresh sample using recommended mass. Note that the NanoDrop is not an accurate method for quantifying nucleic acids.

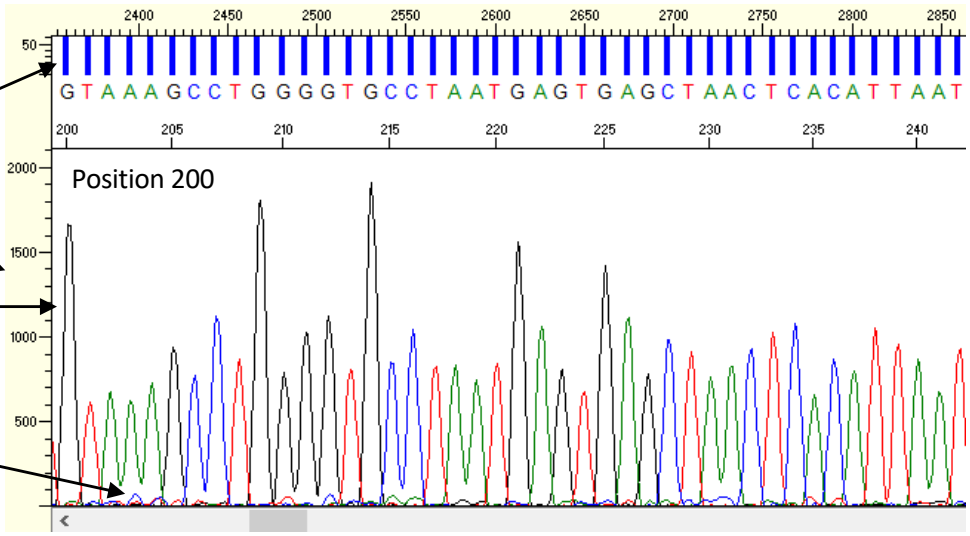
Example 7a: 1/10th primer concentration but correct template

The primer was diluted to 1/10th the recommended concentration. The template was added at the correct mass. The template source is the pGEM-3Zf(+) control; a double stranded DNA plasmid that is 3,197 bp.

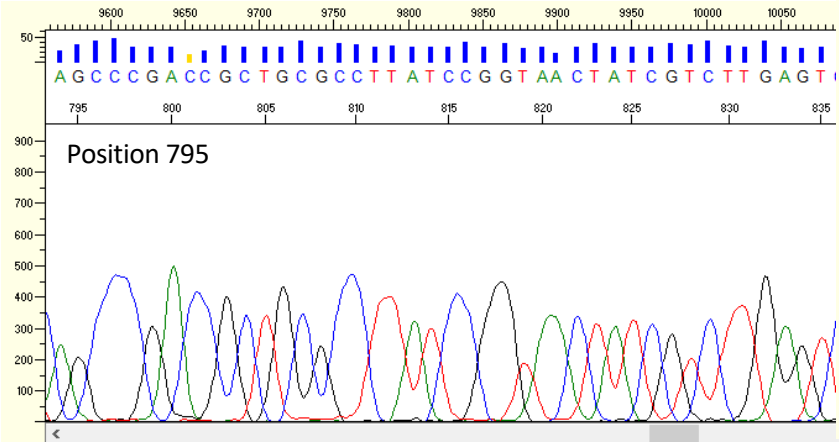
Template Mass	Number of Primers Added	Volume of 1µM Primer	Total Volume	File Name
1000 ng	1	3 µl	12 µl	7a_0.1_primer_plasmid.ab1

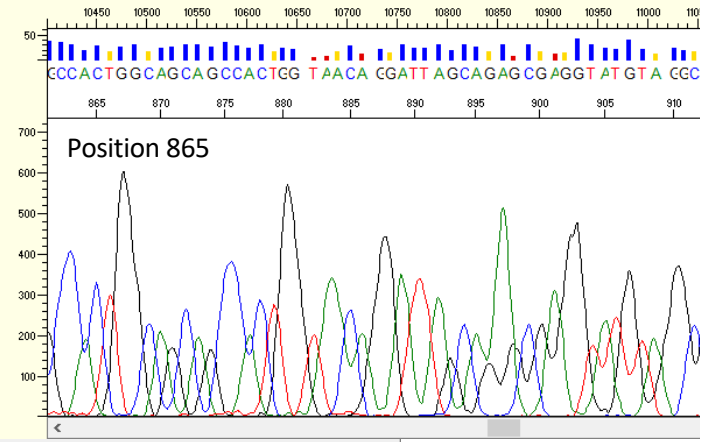
The quality of this sequence is still acceptable at the beginning and middle of the read:

- Quality is high throughout the read
- Signal intensity is > 175 RFU
- Each position only has one peak (i.e. base call) and peaks are sharp
- There is a very small amount of background noise (i.e. peaks along the bottom of the electropherogram)



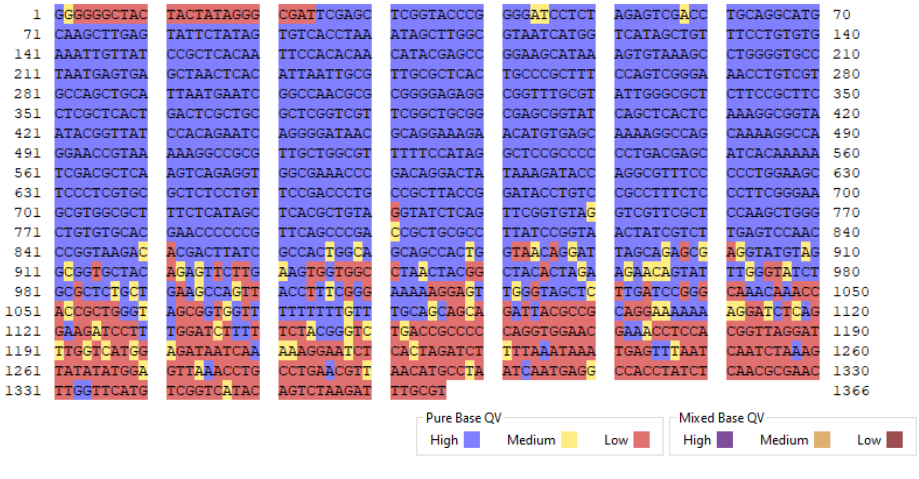
The quality of the sequence decreases earlier in the read than in example 1. At position 795, the signal intensity has decreased, but the quality is OK. The quality really begins to drop off by position 850.





Sequence Annotation and Quality

The base calls in blue are high quality bases with only one base being called. The sequence quality begins to drop slightly earlier than in example 1.



Signal Intensity

The Average Raw Signal Intensities are in the acceptable range.

A(513),C(663),G(550),T(668)

Summary

Sequence is acceptable, but it may be necessary to trim lower quality bases from the end of the read.

To improve results, double check primer concentration and amount that was added to sample. Then submit a fresh sample with the appropriate amount of primer.

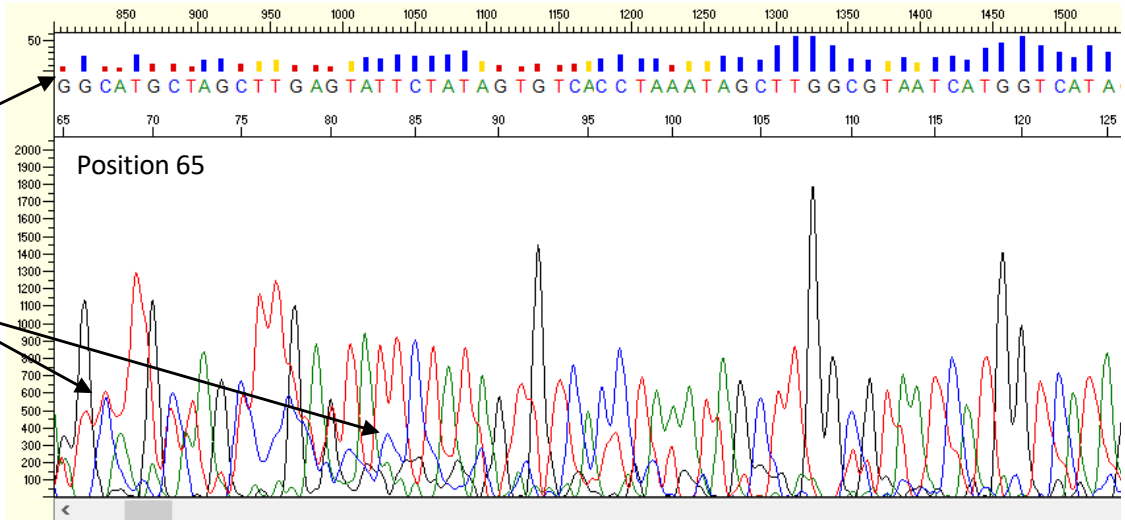
Example 7b: 1/100th primer concentration but correct template

The primer was diluted to 1/100th the recommended concentration. The template was added at the correct mass. The template source is the pGEM-3Zf(+) control; a double stranded DNA plasmid that is 3,197 bp.

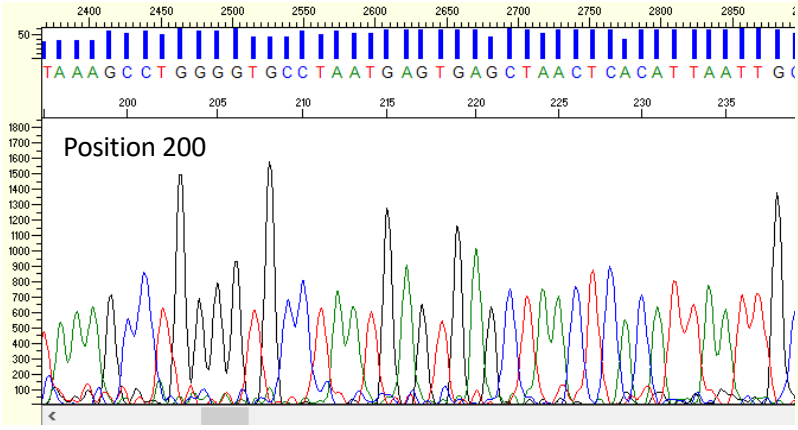
Template Mass	Number of Primers Added	Volume of 0.1µM Primer	Total Volume	File Name
1000 ng	1	3 µl	12 µl	7b_0.01_primer_plasmid.ab1

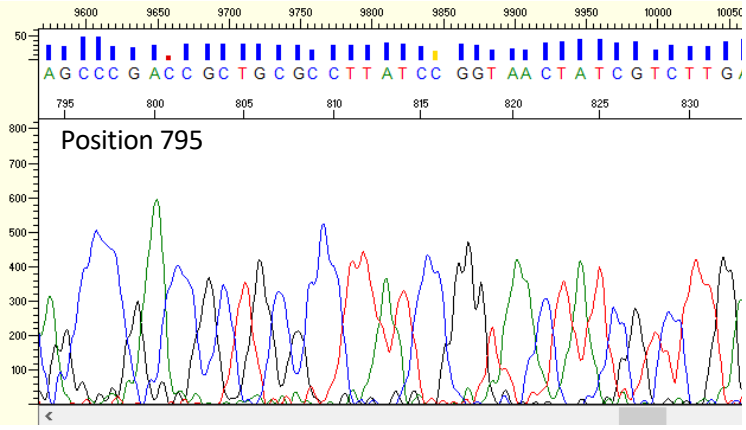
The base call quality is very poor in the first hundred bases. The quality improves after that.

There is background noise and multiple peaks per position.



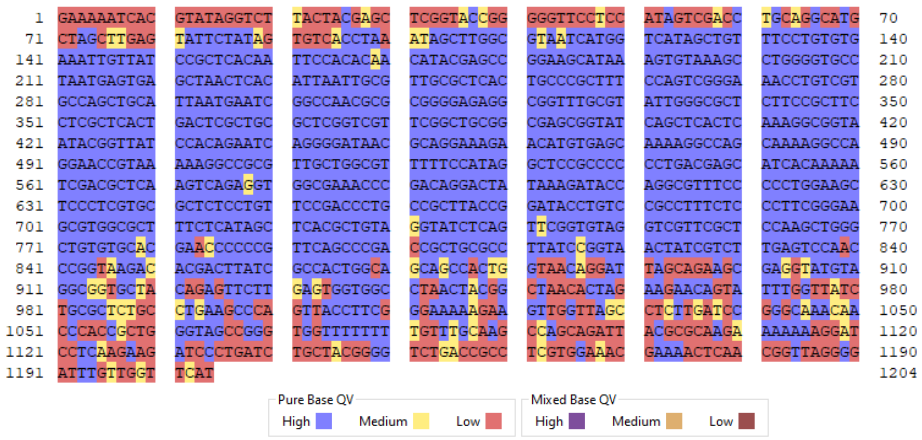
Some background noise continues throughout the sequence and the quality drops off earlier than in example 4a (1/10th primer concentration).





Sequence Annotation and Quality

Sequence quality is poor in the first 100 bases and drops off around base 845.



Signal Intensity

The average raw signal intensities are very low. Data quality may be compromised.

A(68),C(78),G(73),T(83)

Summary

Sequence can be used, but manual review of electropherogram and quality scores must be done to determine what is usable.

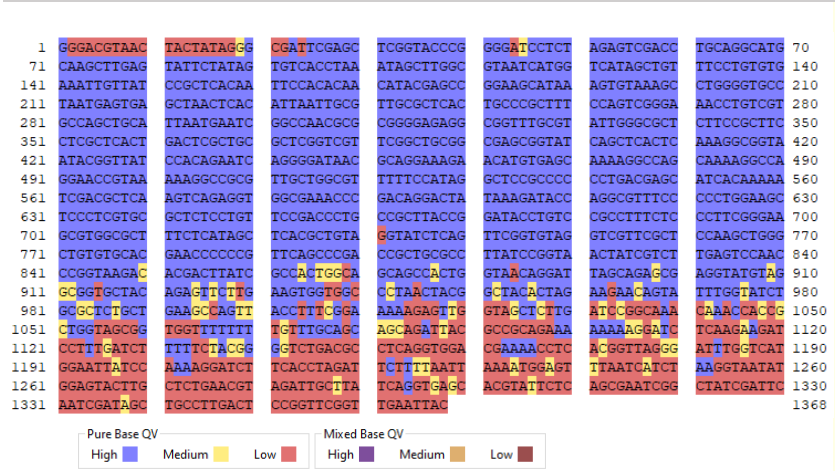
To improve results, double check primer concentration and amount that was added to sample. Then submit a fresh sample with the appropriate amount of primer.

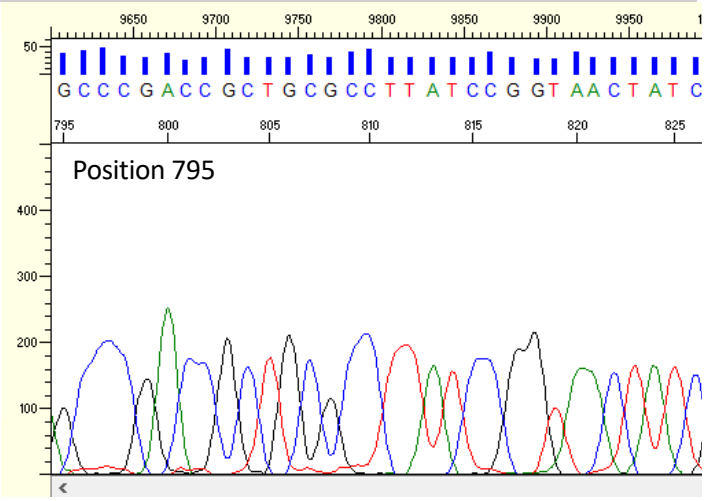
Comparison & Summary: Samples with too little primer

The images below show the sequence and base call quality, the electropherogram at position 795, and the average signal intensities for comparison of a sample (plasmid) with the correct amount of primer and template (example 1), 1/10th recommended primer concentration (example 7a), and 1/100th recommended primer concentration (example 7b). The examples with too little primer still yielded acceptable results. However, manual review of the electropherogram and trimming to remove poor-quality bases at the beginning and end of the read is necessary with 1/100th the recommended primer concentration. Refer to pages 9 and 19-20 for a more detailed summary of each example.

Example 1: Correct template (plasmid)

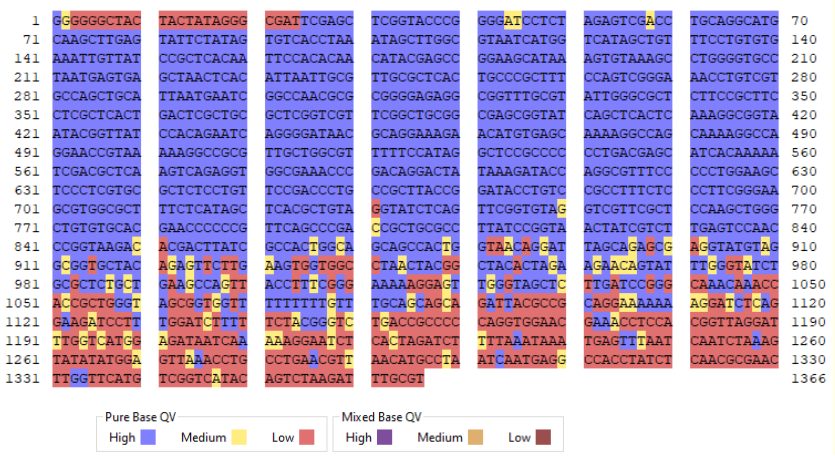
A(890), C(1324), G(664), T(1411)

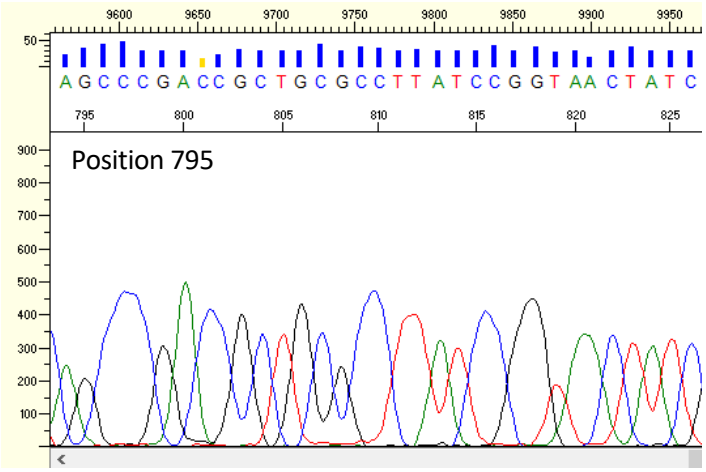




Example 7a: 1/10th recommended primer concentration (plasmid)

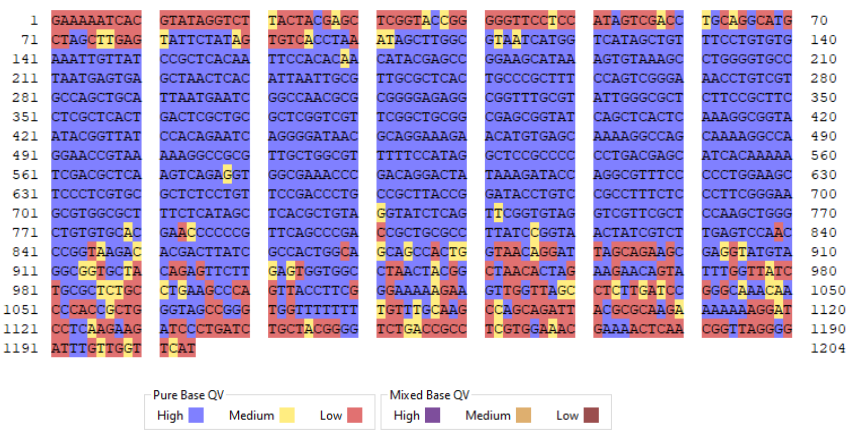
A(513), C(663), G(550), T(668)

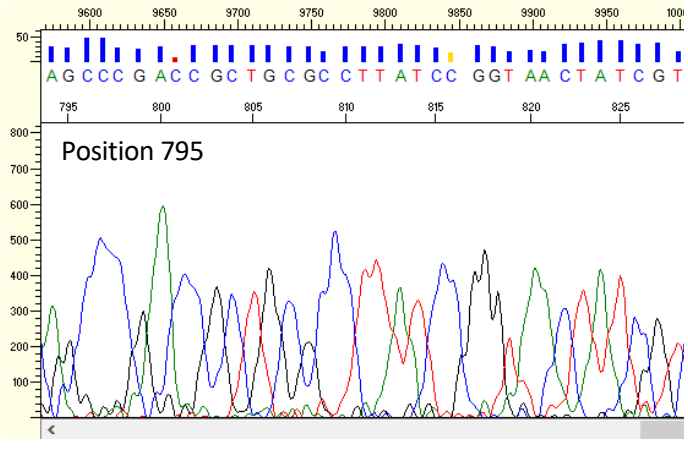




Example 7b: 1/100th recommended primer concentration (plasmid)

A(68), C(78), G(73), T(83)



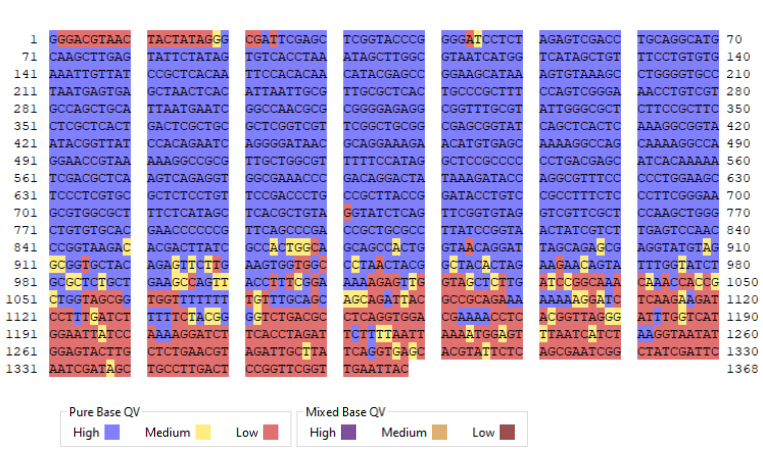


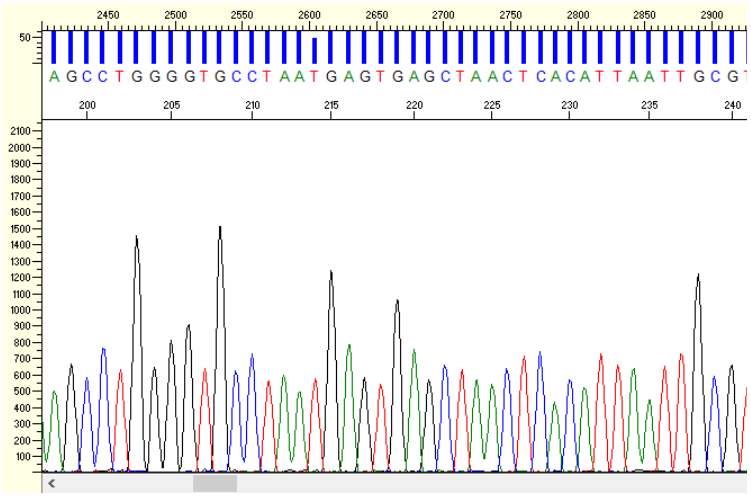
Comparison & Summary: Samples with too much primer

We tested the effects of too much primer (10X and 40X recommended primer concentrations) with the correct template mass (plasmid). In both cases the quality of the sequence data was relatively unaffected, and the data was acceptable. The sequence and base call quality and the average signal intensities are presented below for comparison. The correct primer concentration is on top (example 1), 10X the recommended primer concentration is in the middle (example 8a), and 40X the recommended primer concentration is on the bottom (example 8b).

Example 1: Correct template (plasmid)

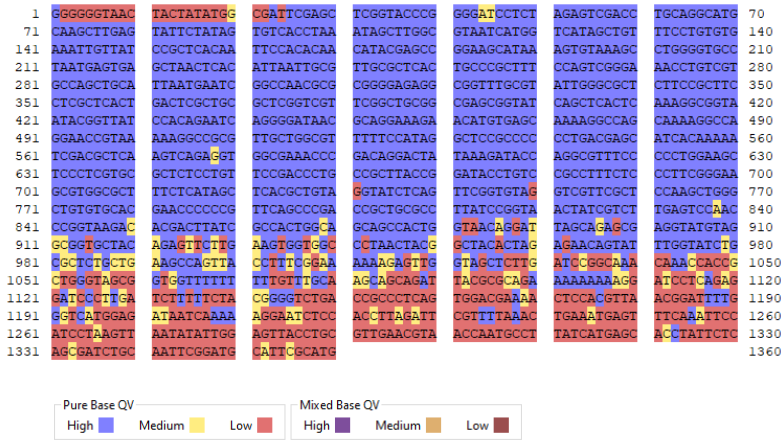
A(890), C(1324), G(664), T(1411)

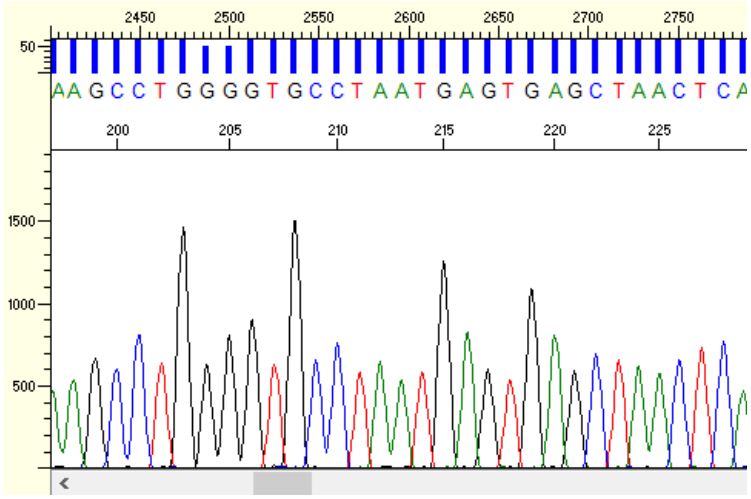




Example 8a: 10X recommended primer concentration (plasmid)

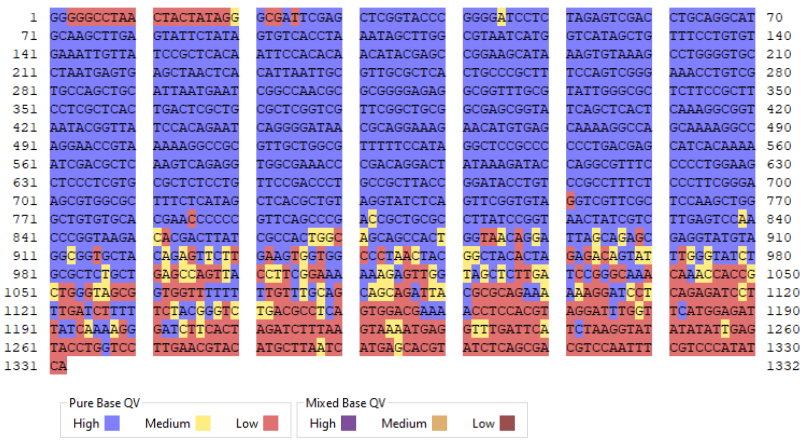
A(815), C(1210), G(597), T(1288)

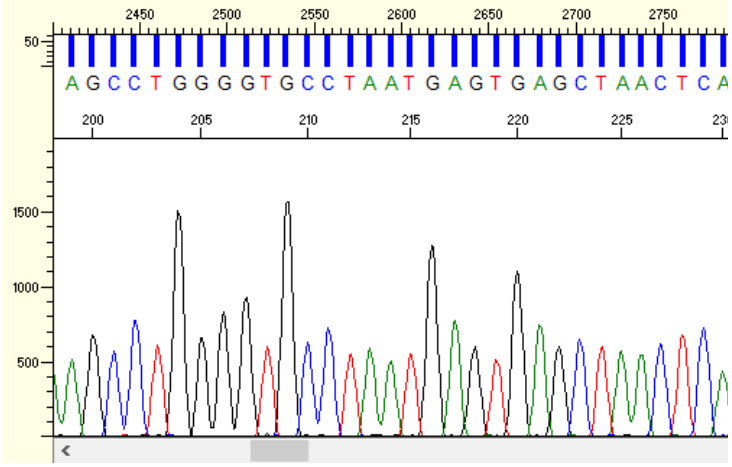




Example 8b: 40X recommended primer concentration (plasmid)

A(851), C(1207), G(625), T(1323)



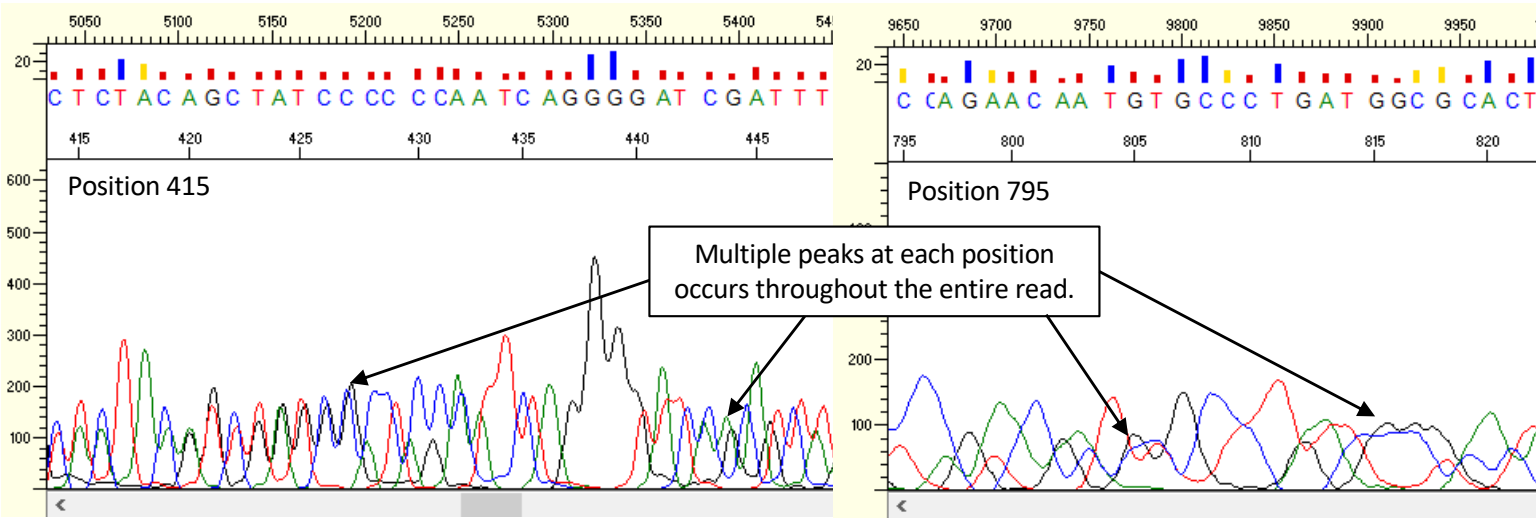


Example 9a: Two primers and correct template

Two primers were added, both at the recommended concentration. The template was added at the correct mass. The template source is the pGEM-3Zf(+) control; a double stranded DNA plasmid that is 3,197 bp.

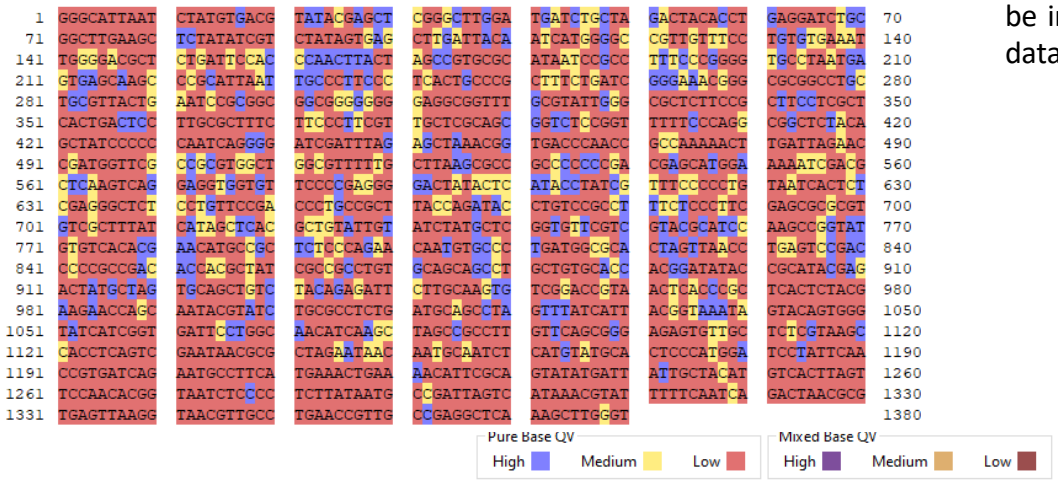
Template Mass	Number of Primers Added	Concentration and volume of primers	Total Volume	File Name
1000 ng	2	3 µl forward primer (10µM) 3 µl reverse primer (10µM)	12 µl	9a_two_primers_1X_plasmid.ab1

The quality of this sequence is poor throughout the entire read. This sequence cannot be used. The phenotype shown here is a mixed template, which means there are multiple peaks at each position.



Sequence Annotation and Quality

The base calls are almost exclusively red, which are low-quality calls. **This sequence is cannot not be used.**



Signal Intensity

Average raw signal intensities can still be in the acceptable range even if the data is poor quality.

A(830), C(1449), G(464), T(1620)

Summary

This example demonstrates the results if two primers are added to a Sanger reaction. Similar results can be seen when primers are not cleaned from a PCR product. The quality of this sequence is very poor and cannot be used.

To achieve usable results, prepare a fresh sample and add only one primer. If sequencing is needed in both the forward and reverse directions, two reactions are necessary.

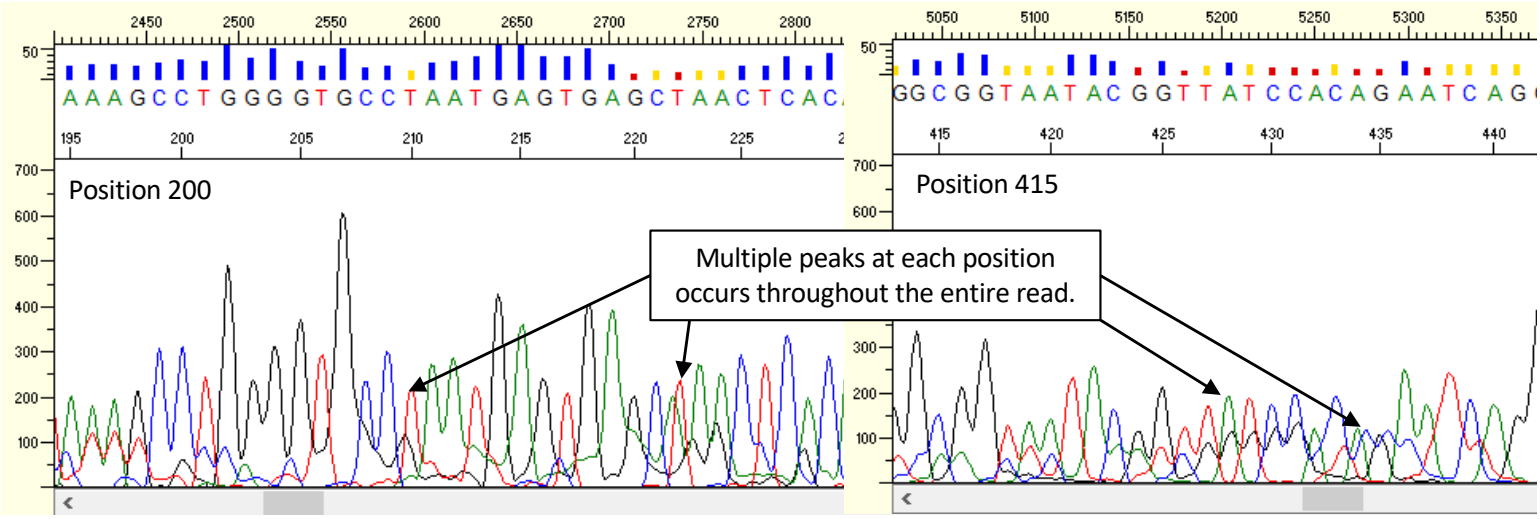
If mixed template is observed in PCR products, visualize the products on an agarose gel to confirm that only one product is present. If there are multiple products, the sample will need to be purified so that only one product remains before Sanger sequencing.

Example 9b: Two primers and correct template

Two primers were added. One at the recommended concentration and the other at 1/10th the recommended concentration. The template was added at the correct mass.
The template source is the pGEM-3Zf(+) control; a double stranded DNA plasmid that is 3,197 bp.

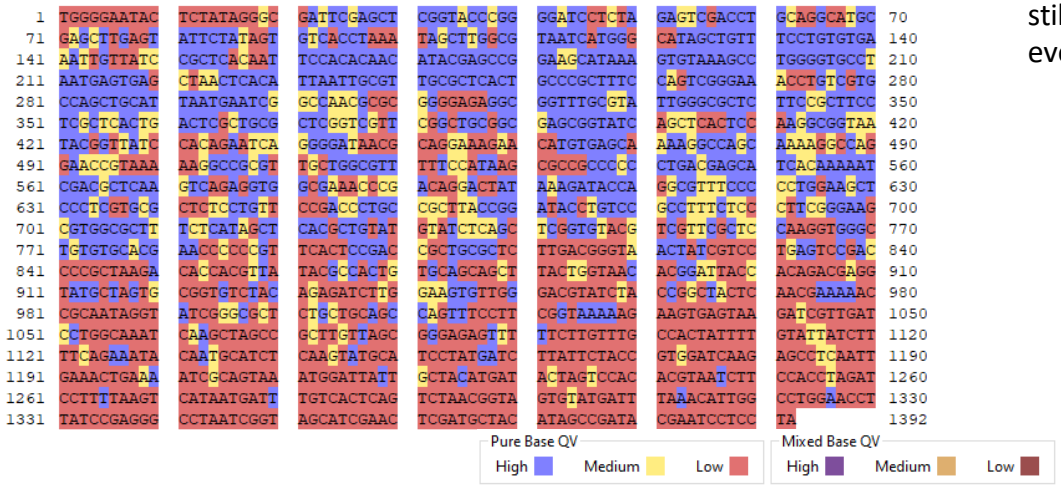
Template Mass	Number of Primers Added	Concentration and volume of primers	Total Volume	File Name
1000 ng	2	3 µl forward primer (10µM) 3 µl reverse primer (1µM)	12 µl	9b_two_primers_1X_0.1_plasmid.ab1

The quality of this sequence is poor, showing the phenotype of mixed template at most positions throughout the read. Mixed template means that there are multiple peaks (base calls) at a single position.



Sequence Annotation and Quality

Most of the base calls are red and yellow, which are low- and medium quality calls. **This sequence is poor quality and should not be used.**



Signal Intensity

Average raw signal intensities can still be in the acceptable range even if the data is poor quality.

A(826), C(1333), G(465), T(1528)

Summary

This example demonstrates the results if two primers are added to a Sanger reaction. Similar results can be seen when primers are not cleaned from a PCR product. The quality of this sequence is very poor and cannot be used.

To achieve usable results, prepare a fresh sample and add only one primer. If sequencing is needed in both the forward and reverse directions, two reactions are necessary.

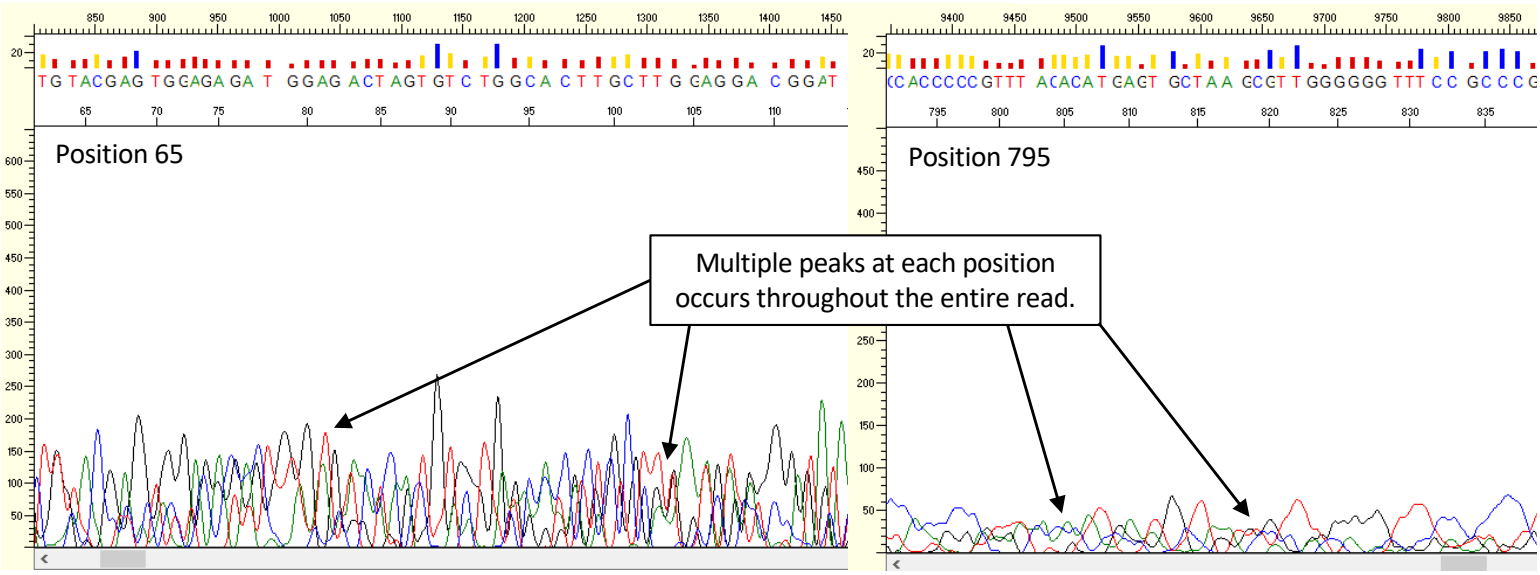
If mixed template is observed in PCR products, visualize the products on an agarose gel to confirm that only one product is present. If there are multiple products, the sample will need to be purified so that only one product remains before Sanger sequencing.

Example 10: Mixed template but correct primer

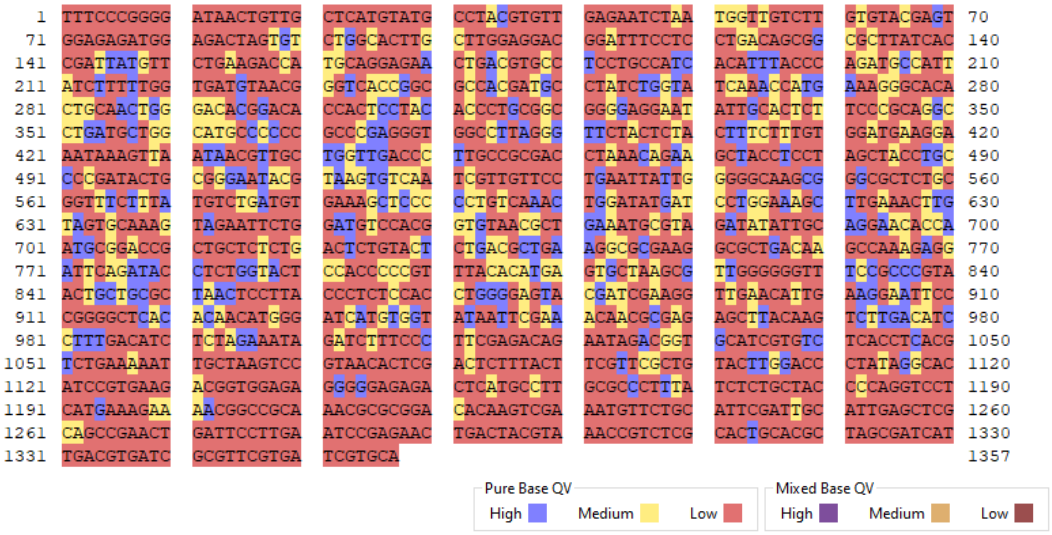
The template source is the ZymoBIOMICS Microbial Community DNA Standard (Zymo Cat# D6306), a mock community composed of genomic DNA from eight bacteria plus two fungi. Full length 16S primers (27F and 1492R) were used to amplify the 16S rRNA gene. The product was purified with ExoSAP-IT. The product size is 1465 bp.

Template Mass	Number of Primers Added	Concentration and volume of primer	Total Volume	File Name
20 ng	1	3 µl forward primer (10µM)	12 µl	10_mixed_template_PCR_product.ab1

This sequence has multiple peaks (base calls) at each position throughout the read indicating a mixed template. This sequence is poor and should not be used.



Sequence Annotation and Quality



Signal Intensity

Average raw signal intensities are in the acceptable range for this sample even though the quality is very poor.

A(497), C(449), G(370), T(455)

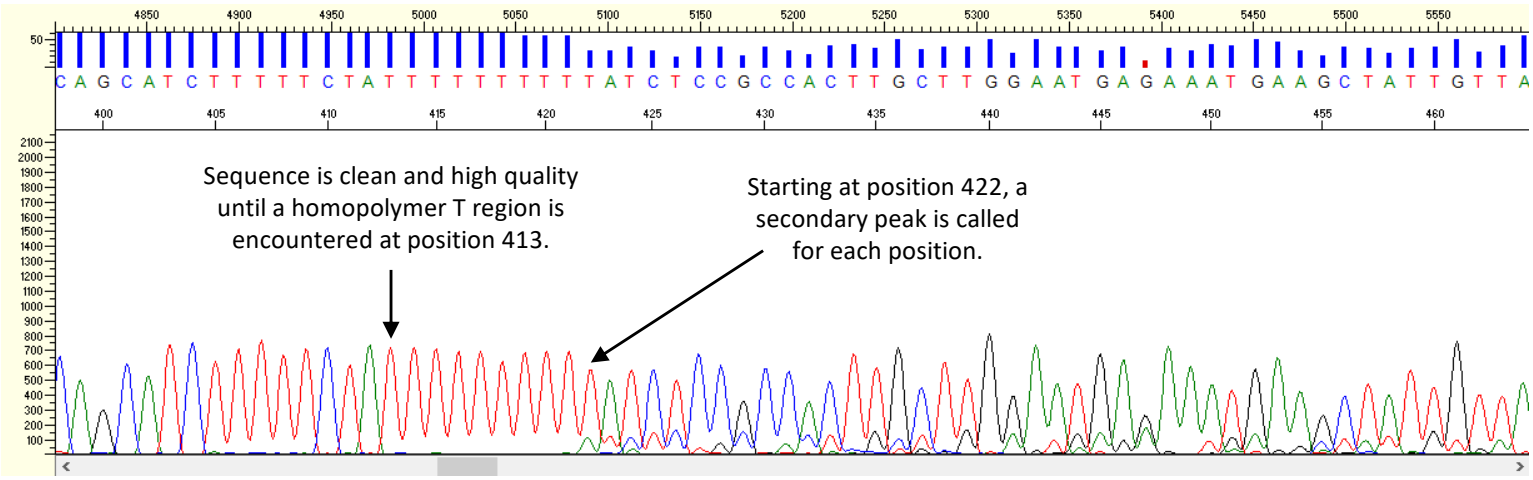
Summary

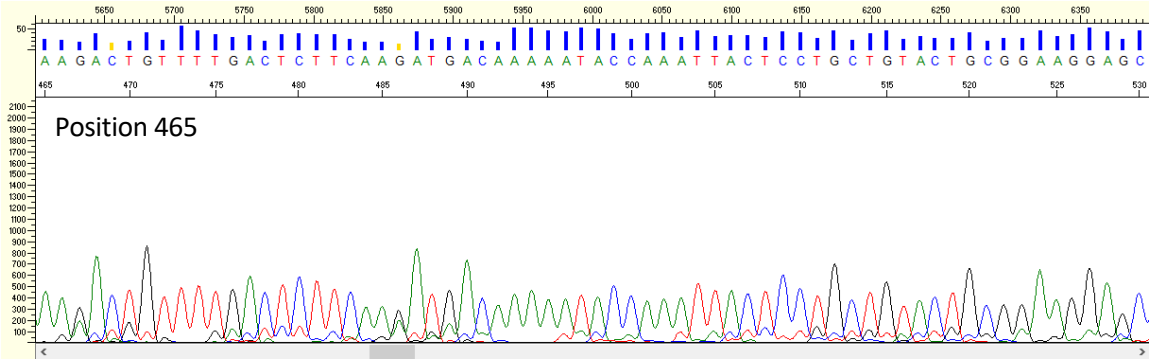
This example demonstrates the results if one primer is added to a Sanger reaction that contains multiple targets. In this case, the target is the 16S rRNA gene from multiple bacterial species. The quality of this sequence is very poor and cannot be used. For this example, it is not possible to use Sanger sequencing. Either short read or long read sequencing is required.

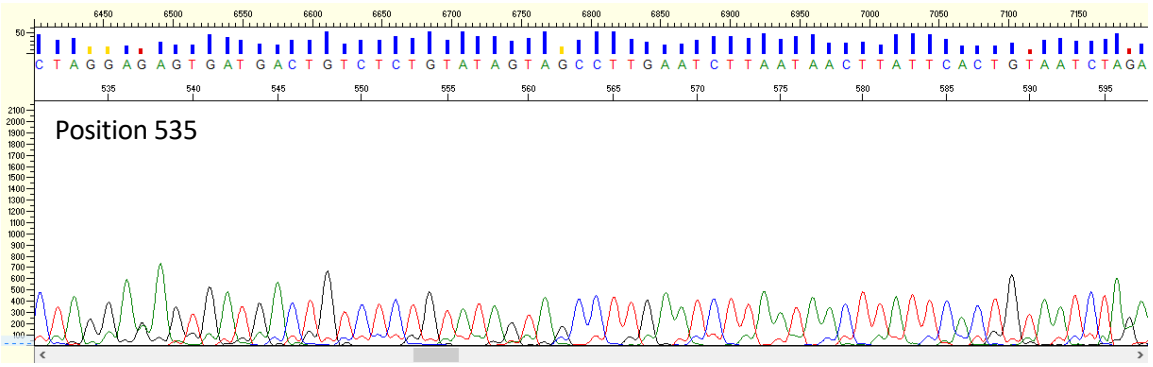
Example 11: Slippage (PCR product)

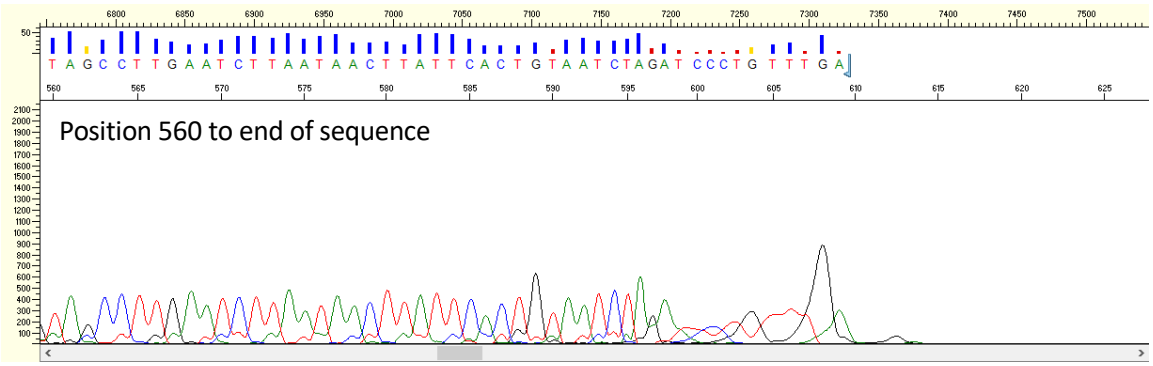
The template was added at the correct concentration. The primer was added at the correct concentration.
The template source is a 510 bp double stranded PCR product that was purified with ExoSAP-it.

Template Mass	Number of Primers Added	Concentration and volume of primer	Total Volume	File Name
20 ng	1	3 µl forward primer (10µM)	12 µl	11_slippage_PCR_product.ab1









Summary

Secondary peaks continue through the entire read.

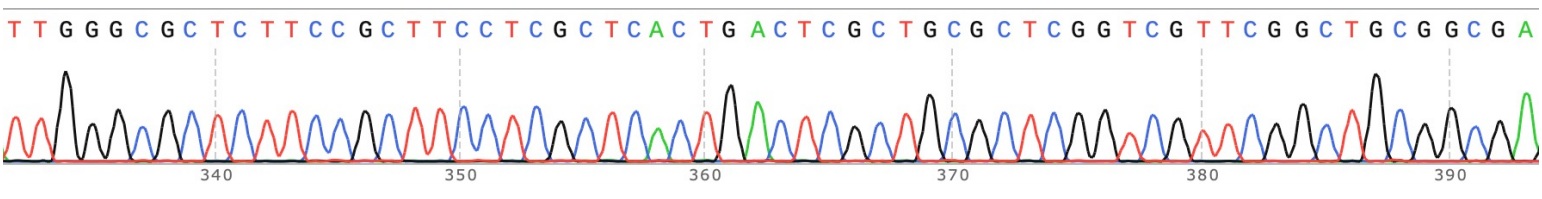
If the secondary peaks are too extreme, it may be necessary to sequence the opposite strand up to the homopolymer that is causing the slippage. Or, a new primer can be used to continue sequencing immediately after the homopolymer that caused the slippage.

In some cases, it may not be possible to satisfactorily sequence around homopolymers.

Software for viewing Sanger Sequencing Data

Below is a non-exhaustive list of software available for viewing Sanger Sequencing data. Some of the software is free and some is paid. The images used in this guide were captured while viewing data in Sequence Scanner 2.

- Sequence Scanner 2
- SnapGene Viewer
- 4Peaks
- Chromas
- GeneStudio
- UGENE
- Chromaseq
- Geneious
- DNASTAR Lasergene
- Sequencher
- CodonCode
- DNA Baser



File Descriptions & Where to download

File names and descriptions are in the table below. All .ab1 files presented in this guide are available to download from the Genomics Core's website at:

https://rtsf.natsci.msu.edu/genomics/technical-documents/sanger-sequencing-best-and-worst-practices.aspx

Page	File Name	Description
9	1_correct_plasmid.ab1	Correct template and primer concentration (plasmid)
10	2a_0.1_template_plasmid.ab1	1/10th template concentration and correct primer concentration (plasmid)
11	2b_0.01_template_plasmid.ab1	1/100th template concentration and correct primer concentration (plasmid)
13	3_correct_PCR_product.ab1	Correct template concentration and primer concentration (PCR product)
14	4a_0.1_template_PCR_product.ab1	1/10th template concentration and correct primer concentration (PCR product)
15	4b_0.01_template_PCR_product.ab1	1/100th template concentration and correct primer concentration (PCR product)
17	5_5X_template_plasmid.ab1	5X template concentration and correct primer concentration (plasmid)
18	6_5X_template_PCR_product.ab1	5X template concentration and correct primer concentration (PCR product)
19	7a_0.1_primer_plasmid.ab1	1/10th primer concentration and correct template concentration (plasmid)
20	7b_0.01_primer_plasmid.ab1	1/100th primer concentration and correct template concentration (plasmid)
22	8a_10X_primer_plasmid.ab1	10X primer concentration and correct template concentration (plasmid)
22	8b_40X_primer_plasmid.ab1	40X primer concentration and correct template concentration (plasmid)
23	9a_two_primers_1X_plasmid.ab1	Two primers added at the recommended concentration. The correct template concentration was used (plasmid)
24	9b_two_primers_1X_0.1_plasmid.ab1	Two primers added. One at the recommended concentration and the other at 1/10th the recommended concentration. The correct template concentration was used (plasmid)
25	10_mixed_template_PCR_product.ab1	A microbial community standard was amplified. The correct template mass and primer concentration were used.
26	11_slippage_PCR_product.ab1	This example demonstrates slippage in a PCR product. The correct template mass and primer concentration was used.

Questions?

Contact the RTSF Genomics Core at gtsf@msu.edu