

A Brief Guide to Interpreting the DNA Sequencing Electropherogram Version 3.0



Plant-Microbe Genomics Facility
The Ohio State University
484 W.12th Ave., Columbus, OH 43210
Ph: 614/247-6204
FAX: 614/247-8696
pmgf@osu.edu
www.biosci.ohio-state.edu/~pmgf/

This guide includes an example of high quality sequence as well as many different problems that occur with DNA sequences attained from the 3700 DNA Analyzer in the Plant-Microbe Genomics Facility. For those figures that demonstrate a problem there is a (1) description of the problem, (2) the most likely cause(s), and (3) one or more solutions.

If you have additional questions or comments about the figures below, then please do not hesitate to contact the facility.

<i>Figure 1 Electropherogram with high quality sequence.</i>	2
<i>Figure 2 Electropherogram that demonstrates the limit of the resolution of the 3700 DNA Analyzer</i>	3
<i>Figure 3 Electropherogram with unincorporated nucleotide peaks.</i>	4
<i>Figure 4 Electropherogram with mobility errors.</i>	5
<i>Figure 5 Electropherogram that demonstrates a lack of extension products, i. e. no bands.</i>	6
<i>Figure 6 Electropherogram that has multiple sequences.</i>	7
<i>Figure 7 Electropherograms that have homoN slippage.</i>	8
<i>Figure 8 Electropherogram that has a strong stop.</i>	9
<i>Figure 9 Electropherogram that has a micro-air bubble or debris.</i>	10
<i>Figure 10 Electropherogram that has "the spread".</i>	11
<i>Figure 11 Electropherogram that has Primer N-1.</i>	12

mrz; 9-03

Figure 1 Electropherogram with high quality sequence.

DNA sequence of high quality is characterized by sharp peaks and little to no background as seen below. DNA sequences of high quality typically result in a read length of 650 to 750 bases with an accuracy of 99%, but in some cases the read length can exceed 900 bases. At this facility the positive control reaction has read length of 720 bases with an accuracy of 99%.

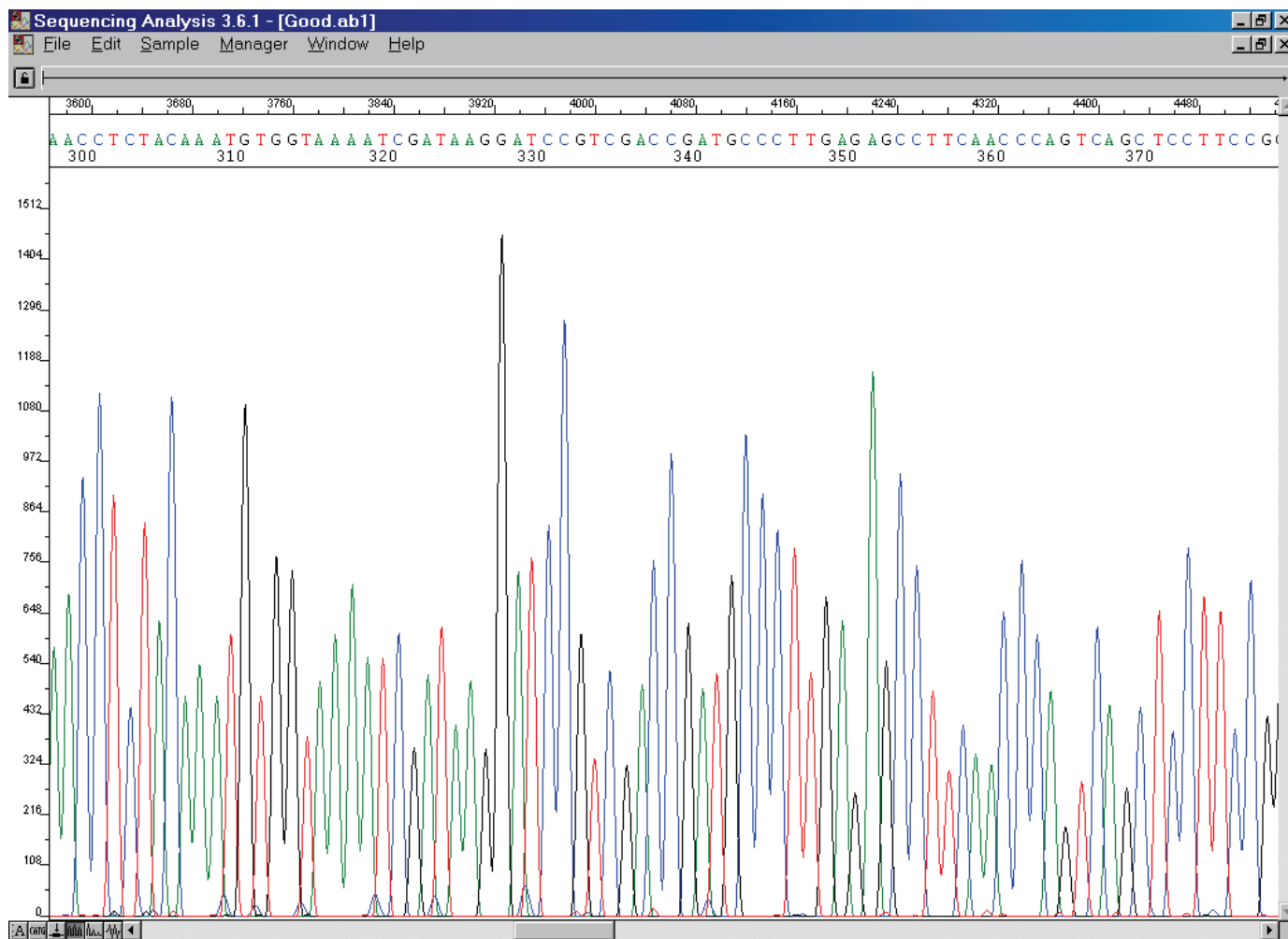


Figure 2 Electropherogram that demonstrates the limit of the resolution of the 3700 DNA Analyzer

The electropherogram below demonstrates how the bands for the extension products eventually become too wide for proper interpretation by the sequencing analysis software. When the width of the base of the peak begins to approach 1/4 of the peak height, then resolution may be lost. Even though the absolute sequence is not always correct the sequence is indicative of the bases present. For example the "T"s (underline and lower case) at 782 and 830 were called as "N"s by the sequencing program. The program often times inserts bases to accommodate a broad peak, for example the sequence from 801 to 810 should be CTTCTGAG and the sequence from 815 to 820 should be AGTGG. The best solution for this limitation is to manually edit the sequence.

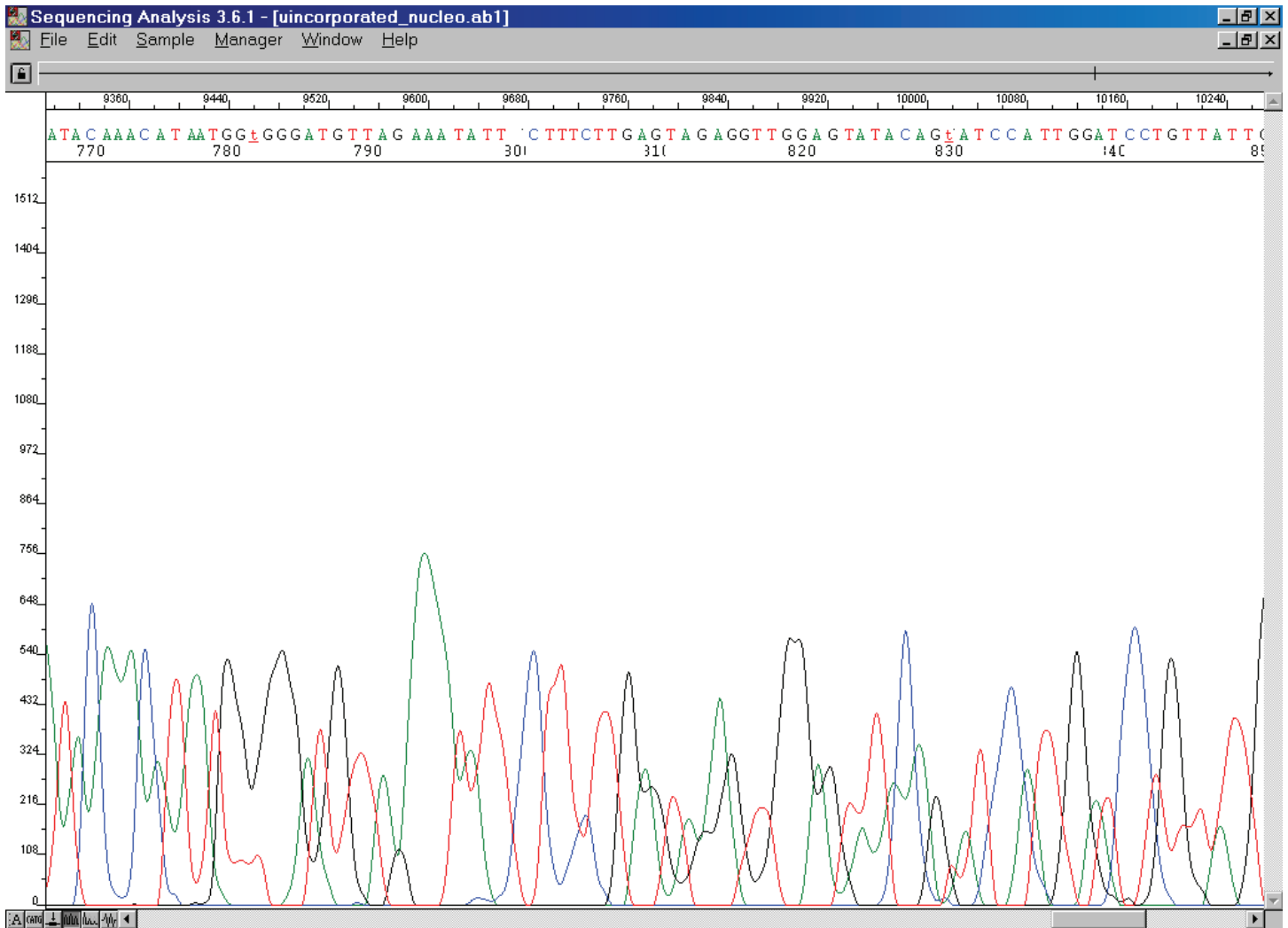


Figure 3 Electropherogram with unincorporated nucleotide peaks.

The electropherogram below has very large peaks at the beginning of the sequence (scan lines 0 to 320), and these peaks are unincorporated nucleotides that can mask the true sequence of the template. The true sequence for bases 1 to 25 is "TAGATTCGGGTACCTTAGTGA". The intensity of the unincorporated peaks varies depending upon the efficiency of the sequencing reaction and the efficiency of the sequencing reaction cleanup procedure (gel filtration or solid phase extraction) after the reaction. The cleanup procedure is performed to remove salts and unincorporated nucleotides prior to electrophoresis. The best solution to this limitation is to edit the sequence manually by looking at the peaks underneath the unincorporated nucleotide peaks.

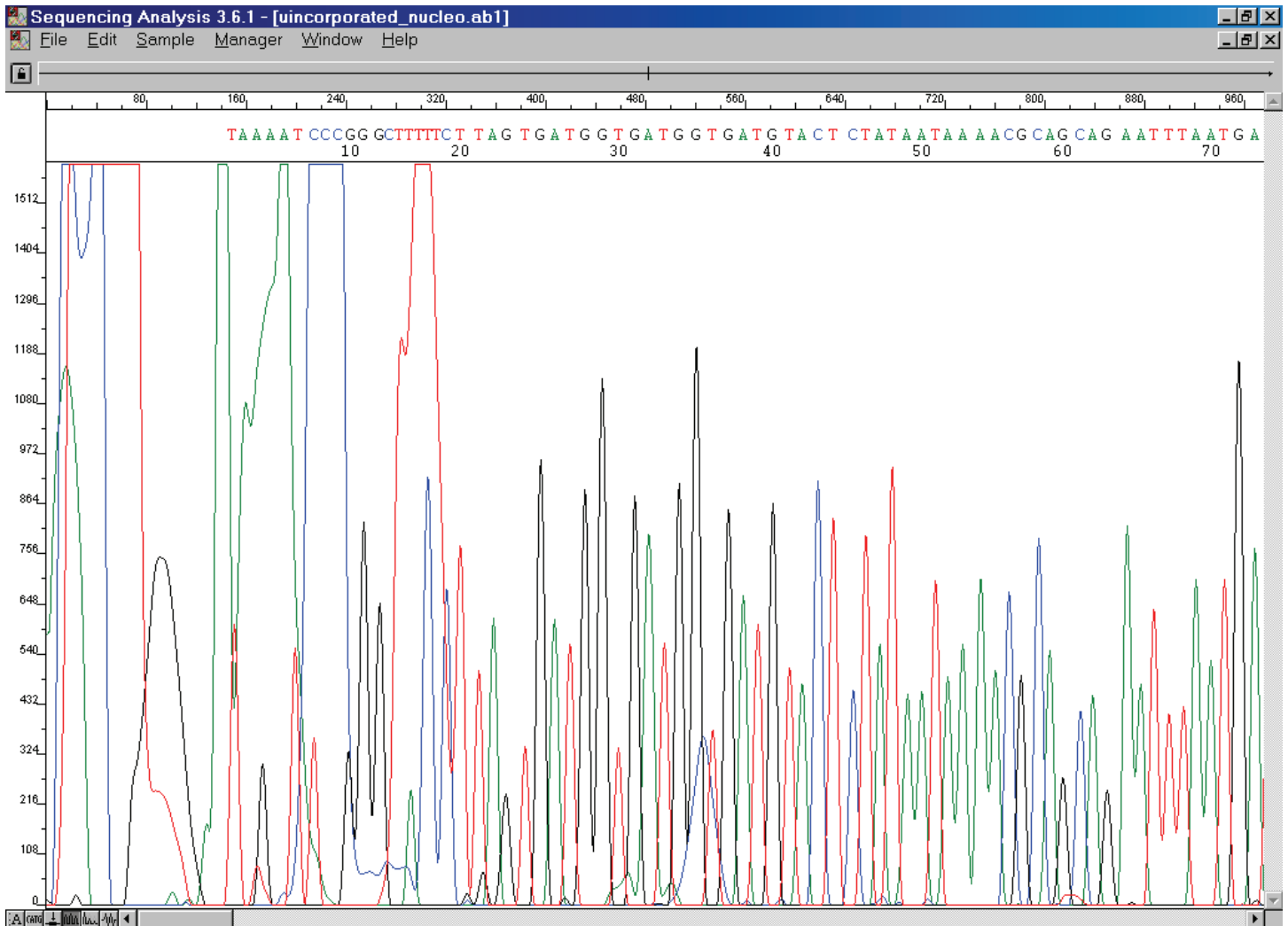


Figure 4 Electropherogram with mobility errors.

Mobility errors only occur in the first 100 bases, and are characterized by peaks that are too close together or overlapping. This problem is sequence dependent and caused by limitations in the software's inability to accurately judge the spacing between the peaks in the electropherogram. For example, bases 10 through 16 are "CTNNCT" due to overlap of the peaks, but the sequence is actually "CTCACT". Also, the "N"s at base 47 and 56 should not be there, but the software is expecting a peak since the previous peaks are shifted to the left. The best solution is to edit the sequence manually.

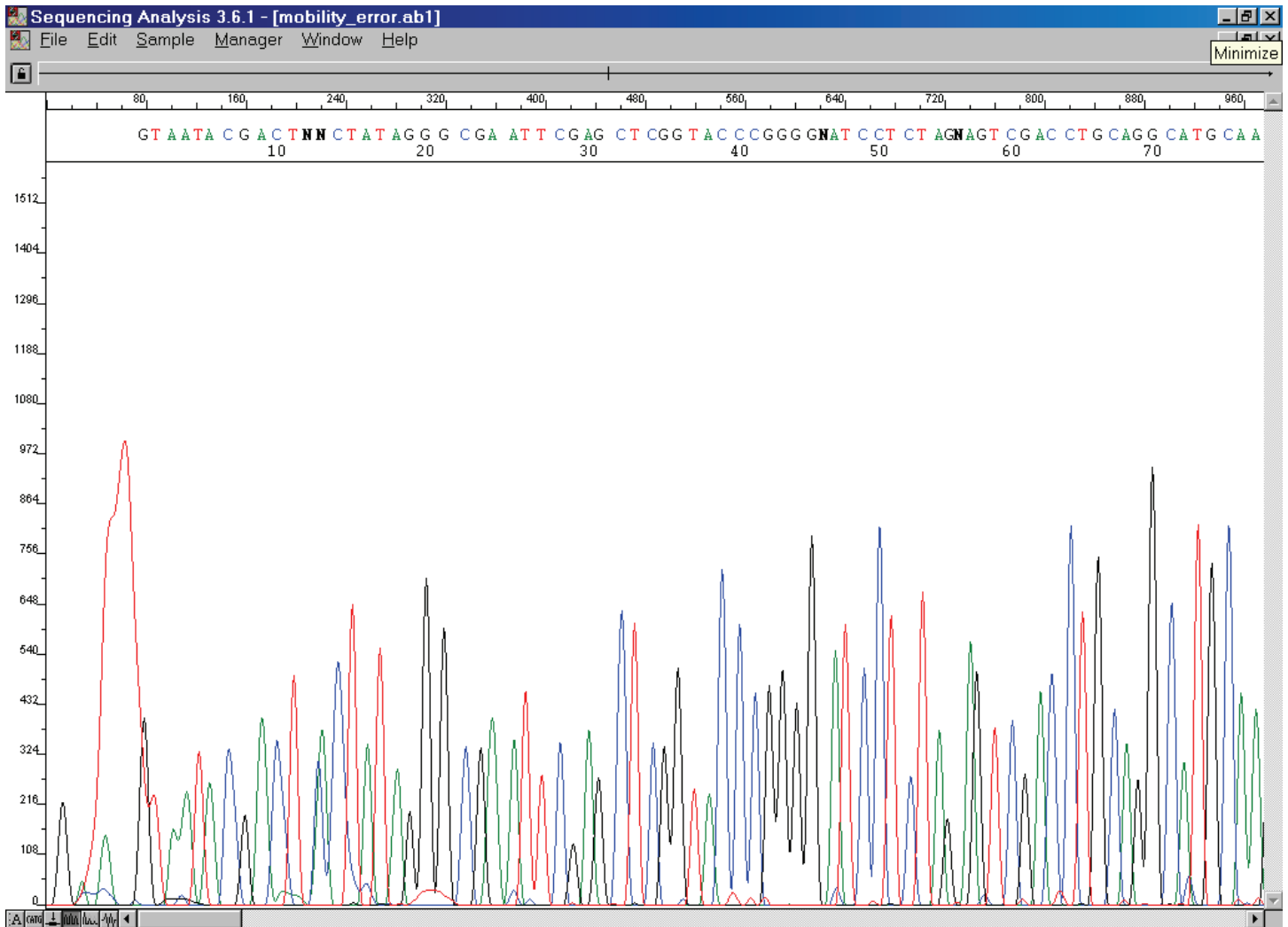


Figure 5 Electropherogram that demonstrates a lack of extension products, *i. e.* no bands.

Large initial peaks from the unincorporated nucleotides followed by a nearly flat line with all 4 colors mixed together characterize the electropherogram below. The causes of this result are numerous including template or primer that is: low quality, low concentration, or degraded. Another cause could be the lack the primer binding site. Also, this may be the result of a simple error on the part of the operator, such as not adding a reagent to the reaction, or a malfunction by the 3700 DNA Analyzer, such as an air bubble in the transfer syringe. The solutions are as varied as the problems described above, *e.g.* repeating the reaction, purifying the template again or redesigning the primer.

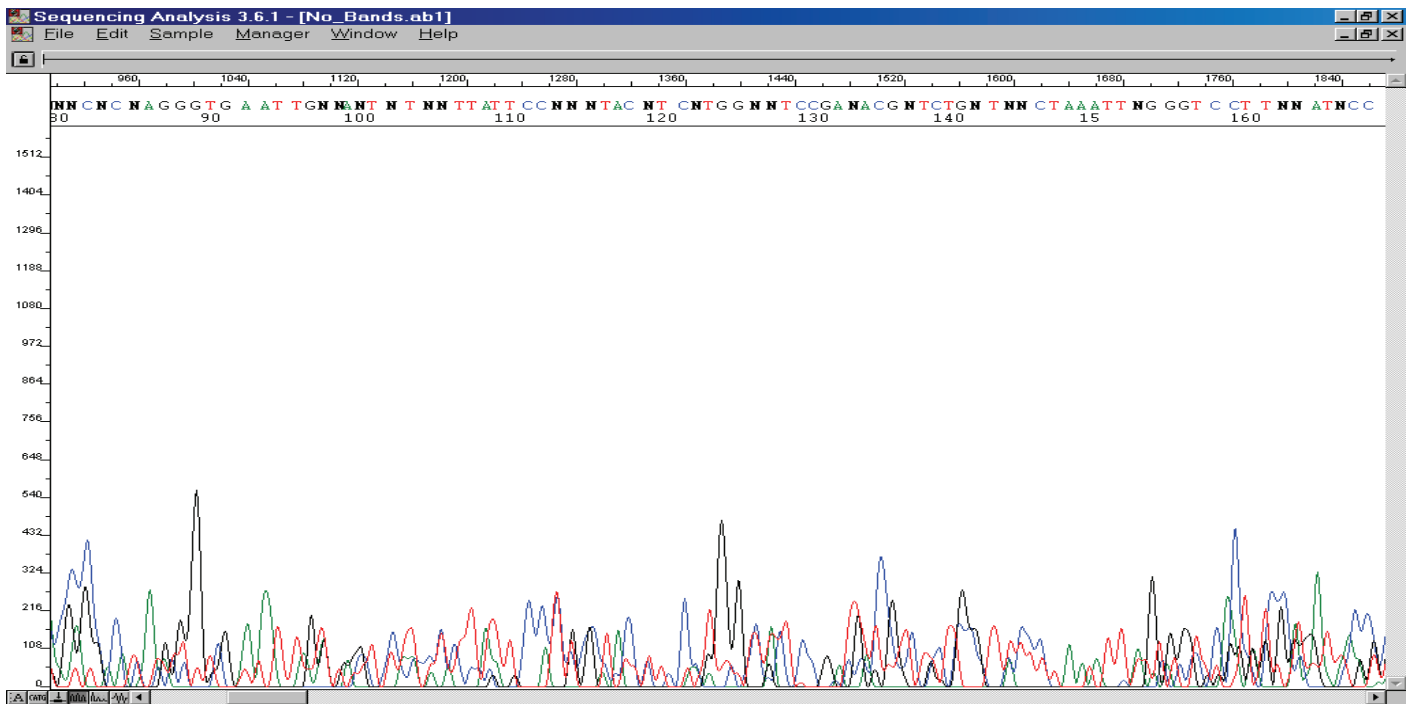
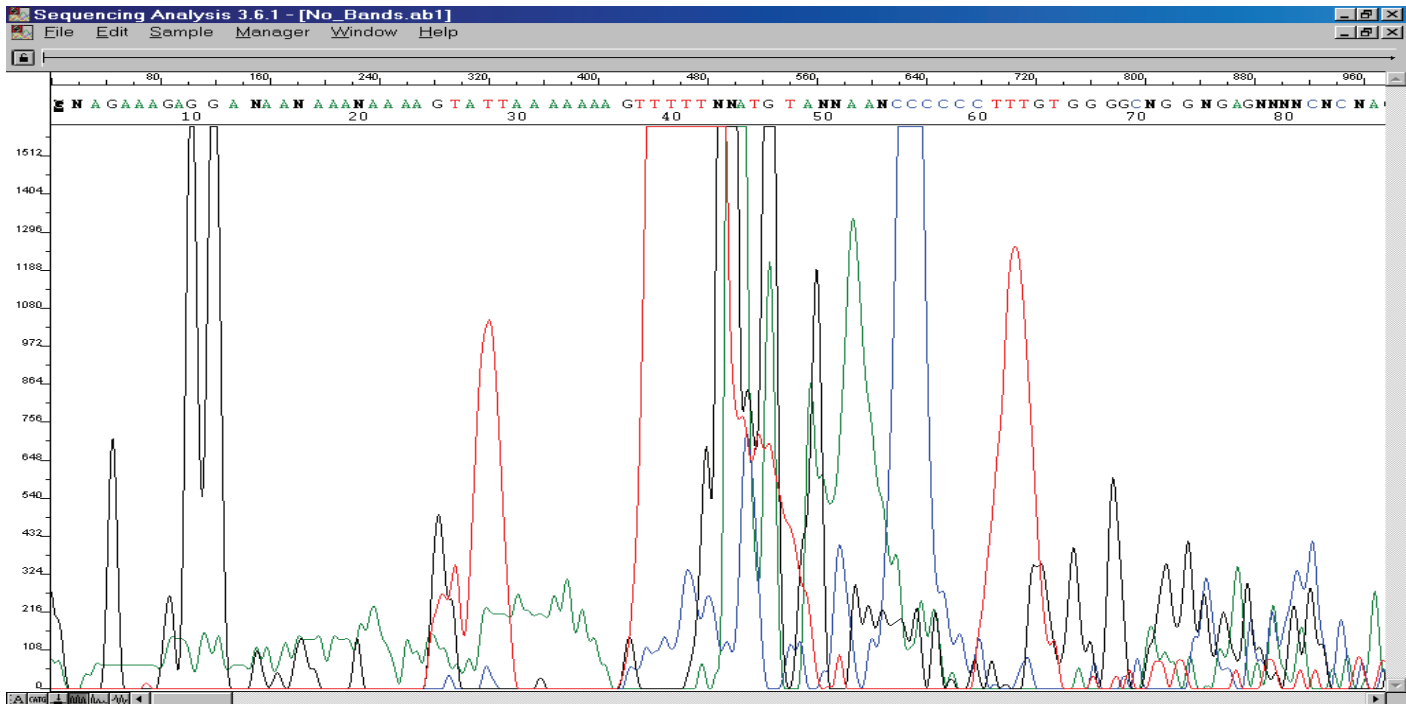


Figure 6 Electropherogram that has multiple sequences.

The electropherogram demonstrates multiple peaks for each position, and the peaks are in phase with each other. For example, base 34 has a “T” peak under the “G” peak. This problem can be due to multiple templates, multiple priming sites or multiple primers. Below is a special case, multiple inserts, in which there are two different plasmids present that share the same vector (bases 1 – 33), but they have different inserts (bases 34 - 160). The solution is to separate the plasmids and resequence, e.g. restreaking the bacterial culture that contains the plasmids and then purify the plasmid again.

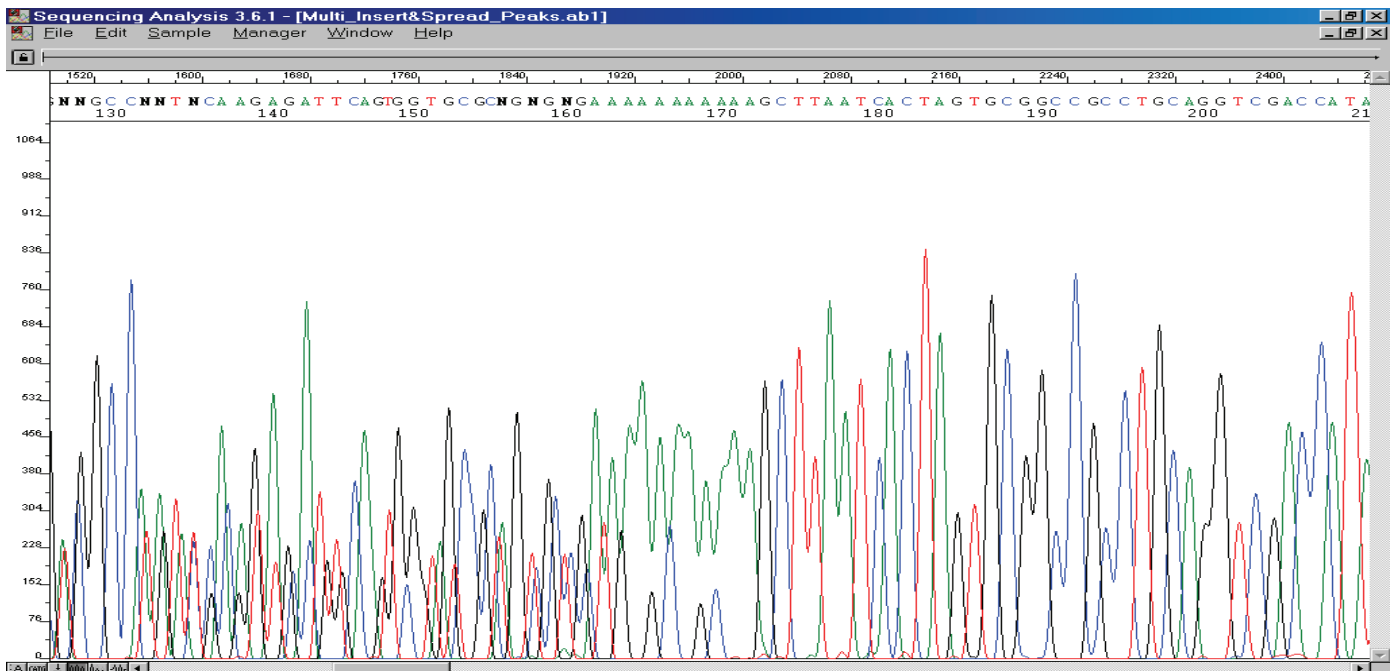
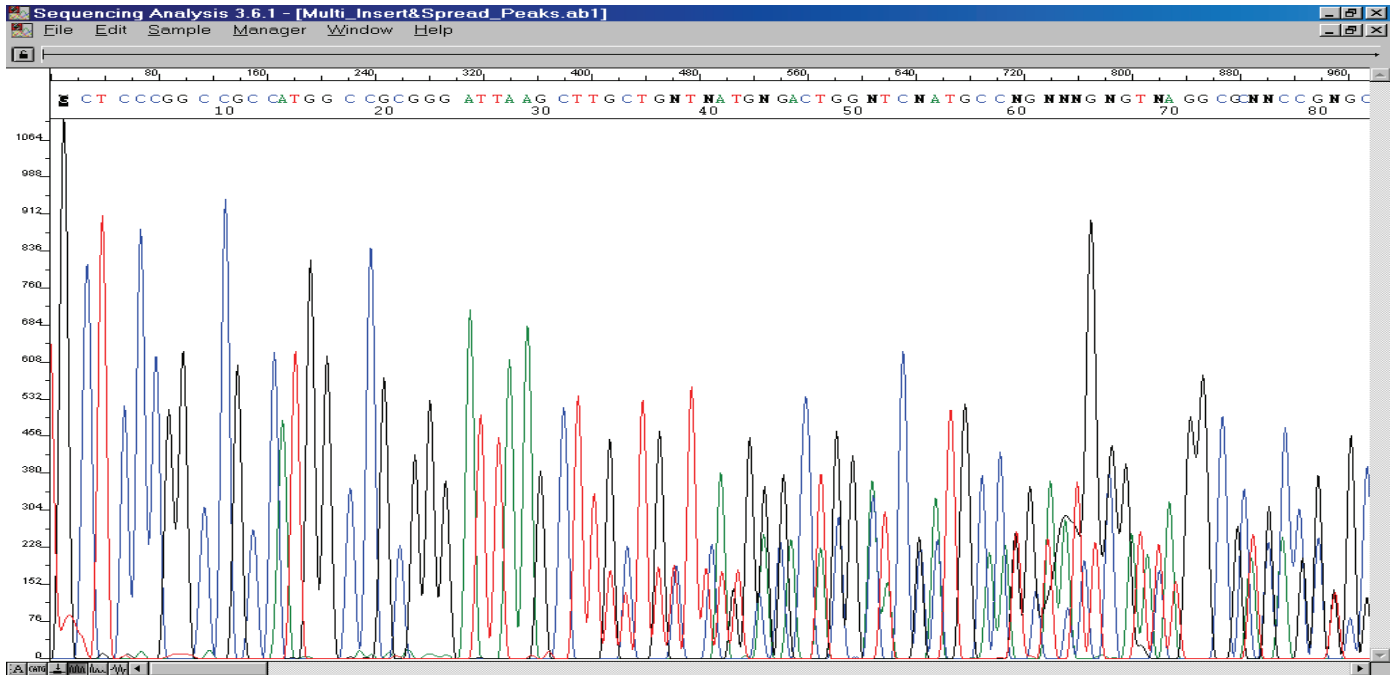


Figure 9 Electropherogram that has a micro-air bubble or debris.

The sequence is characterized by a large spike of all four colors and is caused by either a micro-air bubble or small debris migrating out of the capillary into the laser path and therefore scattering the light. For example, the spike is at base 450 and is masking the true base at this location. The best solution is to run the sample again on the DNA Analyzer.

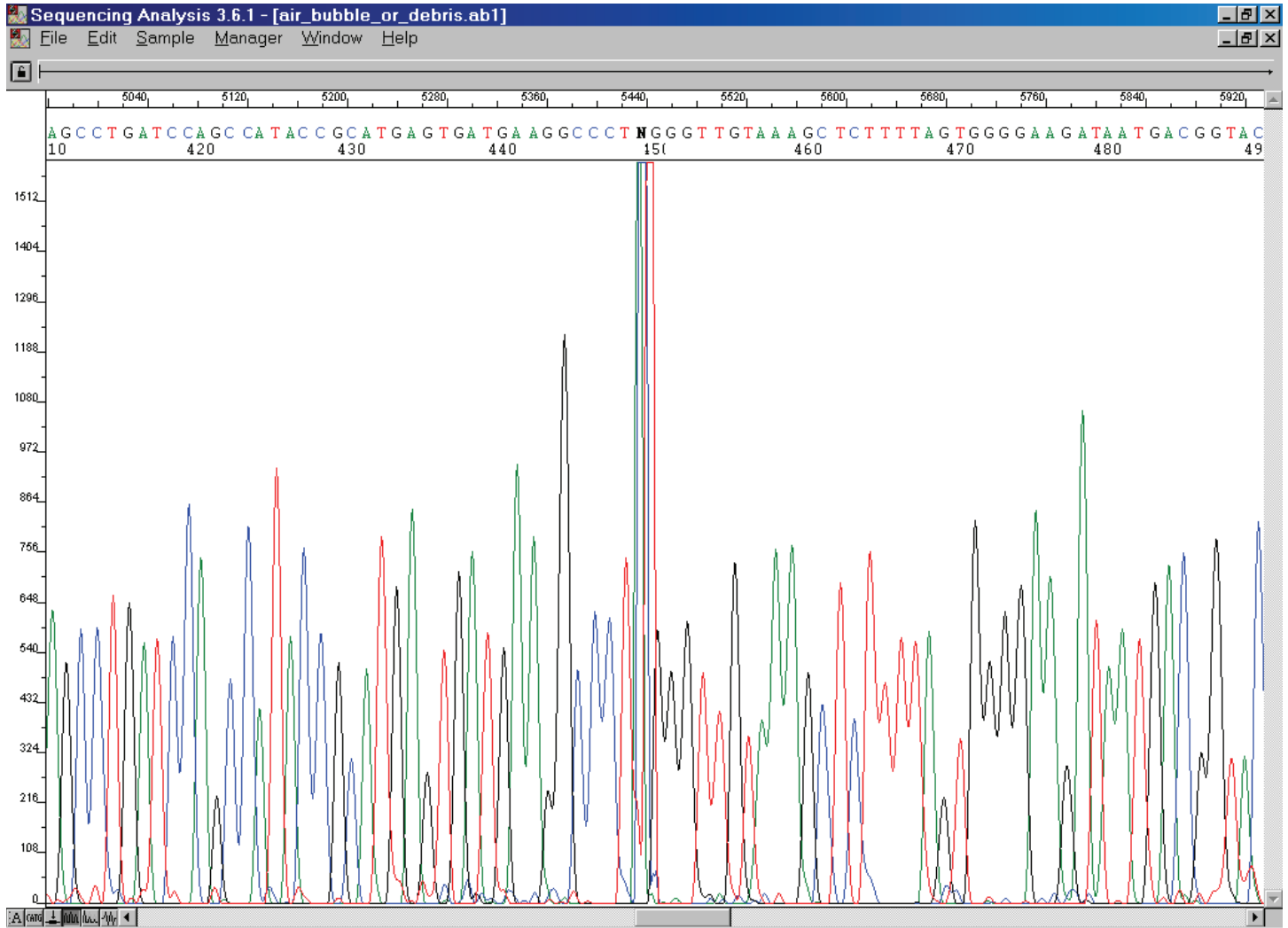


Figure 10 Electropherogram that has "the spread".

Bands that gradually spread out and therefore become irresolvable characterize the sequence that has "the spread". This problem is characterized by delayed migration of the extension products, *i.e.* the wider the first peak the longer the delay. The first peak that is too broad (irresolvable) can be from base 1 to as late as base 400. The problem is caused by a small, ionic contaminant in the DNA preparation. Most commonly the contaminant is associated with plasmids that were purified with a solid phase extraction kit, *e.g.* Qiaprep or Wizard. *This is a common problem that is not well understood.* To solve or minimize the problem the facility will add EDTA to the sequencing reaction, repeat the cleanup procedure, and then analyze the reaction again. When this does not work we repeat the reaction with different reaction conditions to minimize the amount of contaminant and maximize the amount of extension products. At the moment we know of no confirmed methods the customer can do to minimize the contaminant in their plasmid preparation. Since there is little the customer can do to address this problem, then the reactions with this problem are automatically rerun without the need for customers to request that the reaction be repeated.

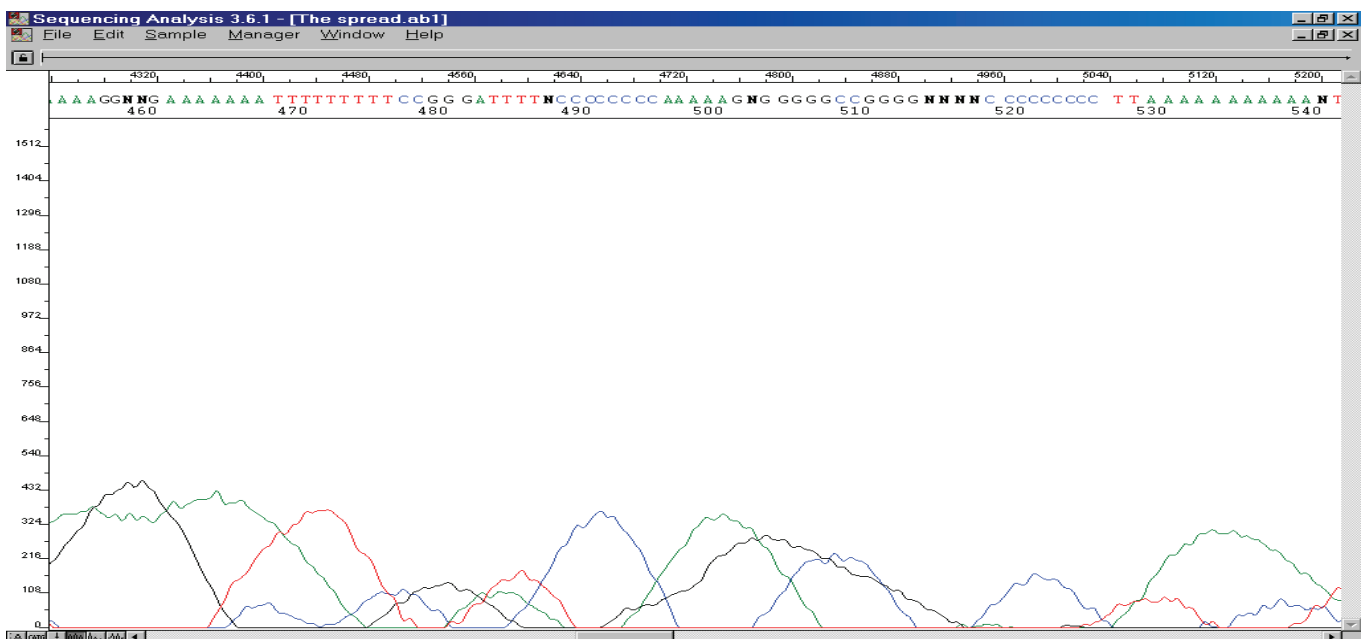
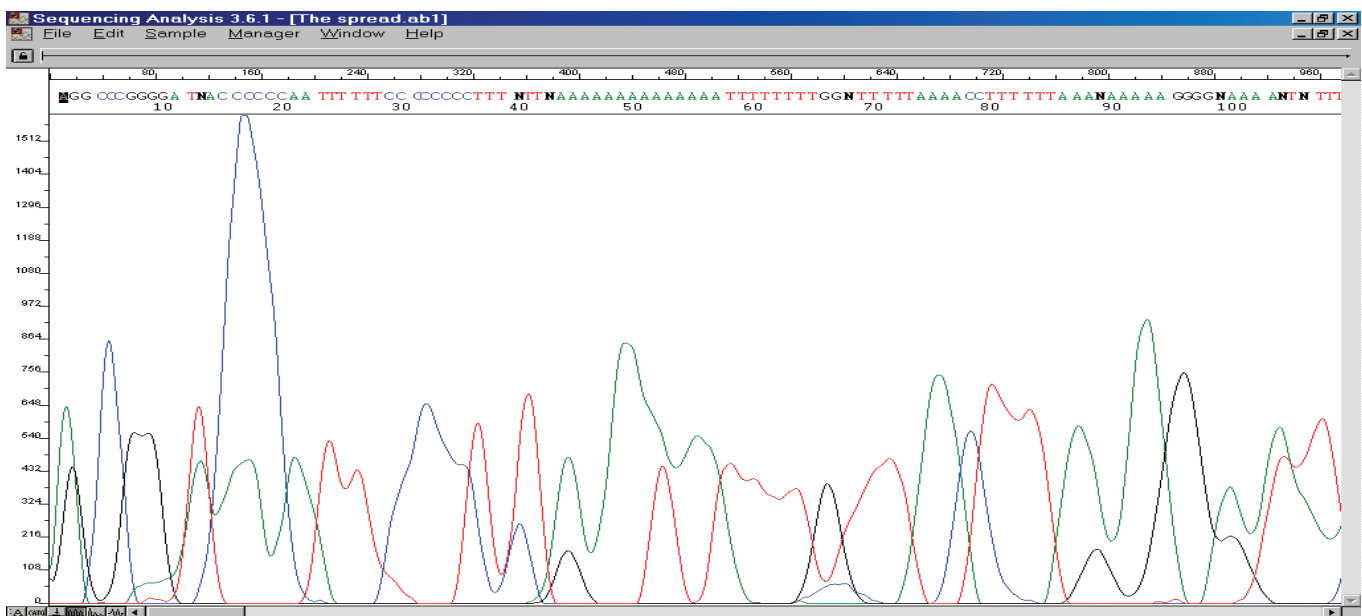


Figure 11 Electropherogram that has Primer N-1.

The sequence is characterized by having multiple peaks at all of the locations and these peaks are arranged such that the same sequence is present but shifted by one, two or more positions. For example, the “C” at position 289 is represented by a smaller peak at base 288 and an even smaller peak at base 287. This problem is caused by a significant percentage of the primers being short by 1, 2 or more bases. For example, the primer for the reactions below was 20 bases in length, but many molecules are 19 bases long and a few are 18 bases long. The best solution to the problem is to use another primer and do another sequencing reaction. Another solution is to purify the primer, or order a primer that is purified followed by repeating the sequencing reaction.

