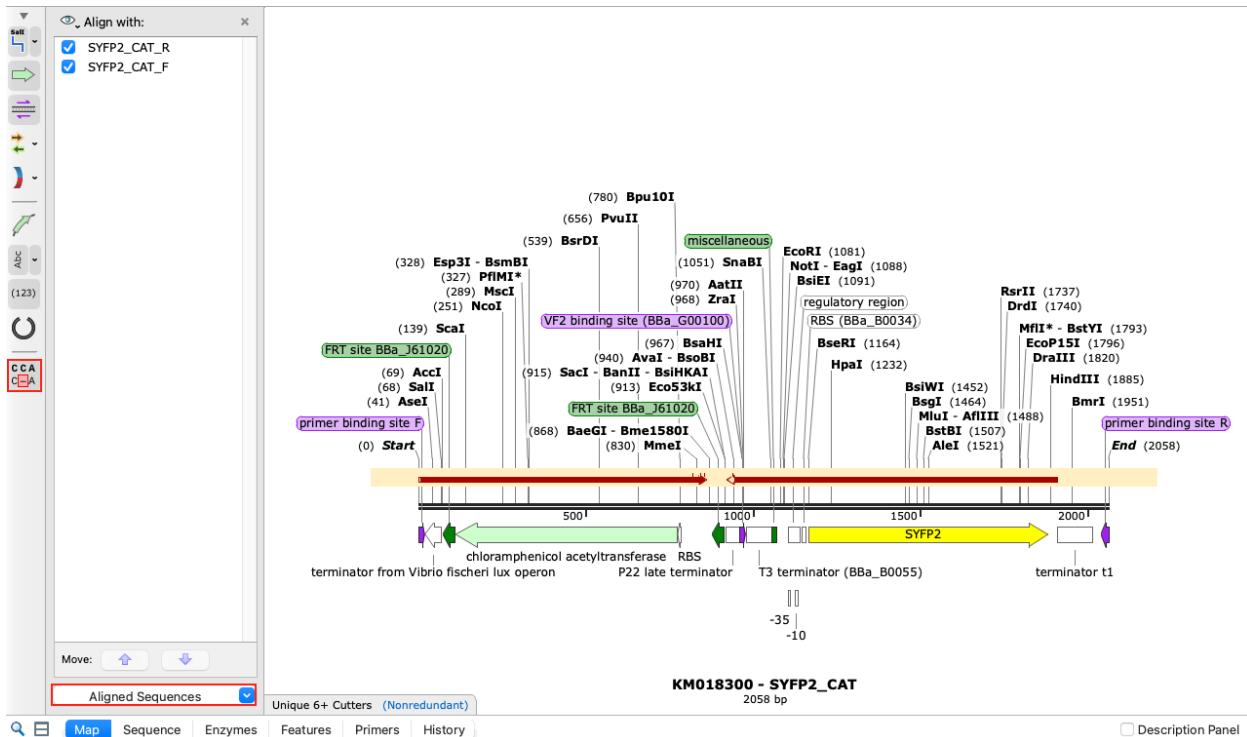


3. Validating Results

Align to DNA Reference

This folder contains sample data demonstrating the **Align to DNA Reference** tool in SnapGene. This tool is commonly used to align sequencing data (such as Sanger trace files) to a reference DNA sequence. This tool can be accessed in one of two ways:

1. From the **Tools>Align to Reference DNA Sequence** menu option
2. From the **Show Alignments** button in the sidebar (shown in the screenshot below in the middle-left)



From there, the **Tools>Align to Reference DNA Sequence** will give you three options:

1. **Align Imported Sequences...** this will take you to the file explorer on your operating system so that you can select files on your computer to align.
2. **Aligned Copy Sequence**, allowing you to align a sequence previously copied to your clipboard.
3. **Align Open Sequences...** so that you can align sequences that you already have open in SnapGene.

To access the same options when using the **Show Alignments** button, select the **Aligned Sequences** option outlined in the screenshot above. This window will also allow you to drag files from your computer file system directly into the Align with box shown in the screenshot above.

The sample reference contained in this folder already has the two Sanger reads provided in this folder aligned to it. The reference sequence is a Super Yellow Fluorescent Protein 2 (SYFP2) expression cassette, containing the fluorescent protein SYFP2 ([Bba_K864100](#)) from the Registry of Standard Biological Parts. This cassette also expresses chloramphenicol acetyltransferase, conferring resistance to the antibiotic chloramphenicol. The cassette has been cloned into the linearized [pGEM®-T Easy](#) (Promega) by TA cloning. Sanger sequencing was then performed using the M13F and M13R primers. For the purposes of this sample data, the vector

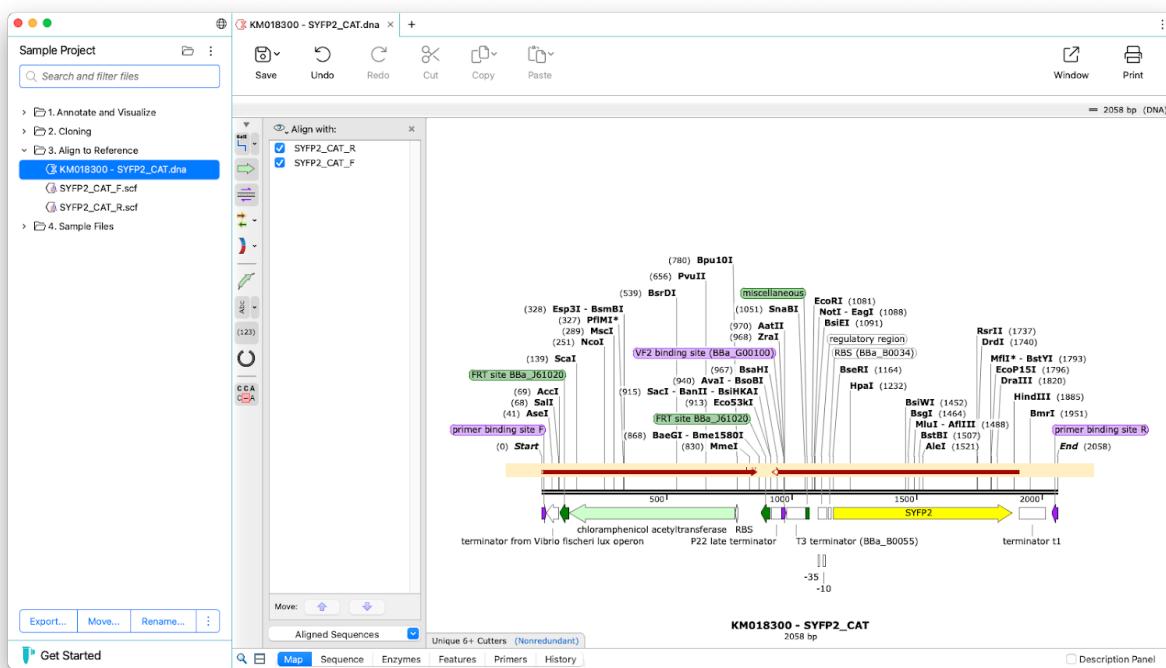
portion of the sequencing data has been trimmed, leaving only the regions that map to the expression cassette, then the two reads have been aligned to the reference sequence as described above.

The Align to DNA reference folder contains three files:

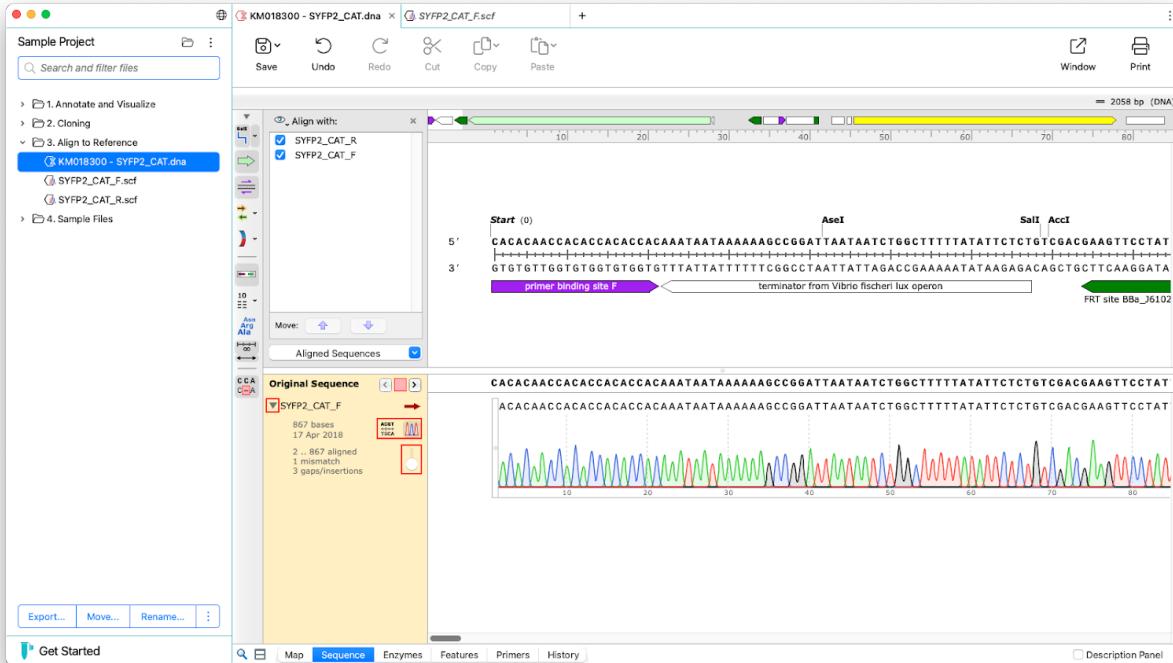
1. KM018300 - SYFP2_CAT.dna. This is the reference sequence.
 2. SYFP2_CAT_F.scf
 3. SYFP2_CAT_R.scf

2 and 3 are the chromatogram files that have already been mapped to the reference sequence. These chromatograms are the format that many sequencing providers will send data in.

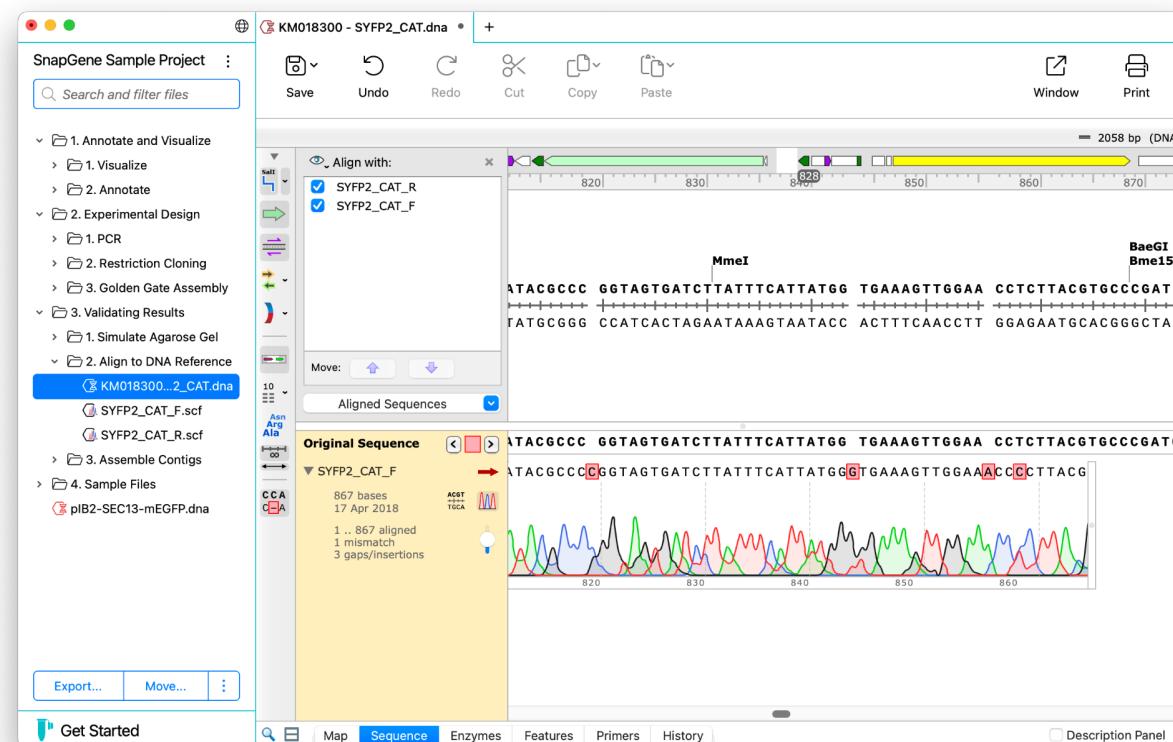
Looking closer at the reference sequence you will see that the two reads map in opposite directions along a good portion of the sequence.



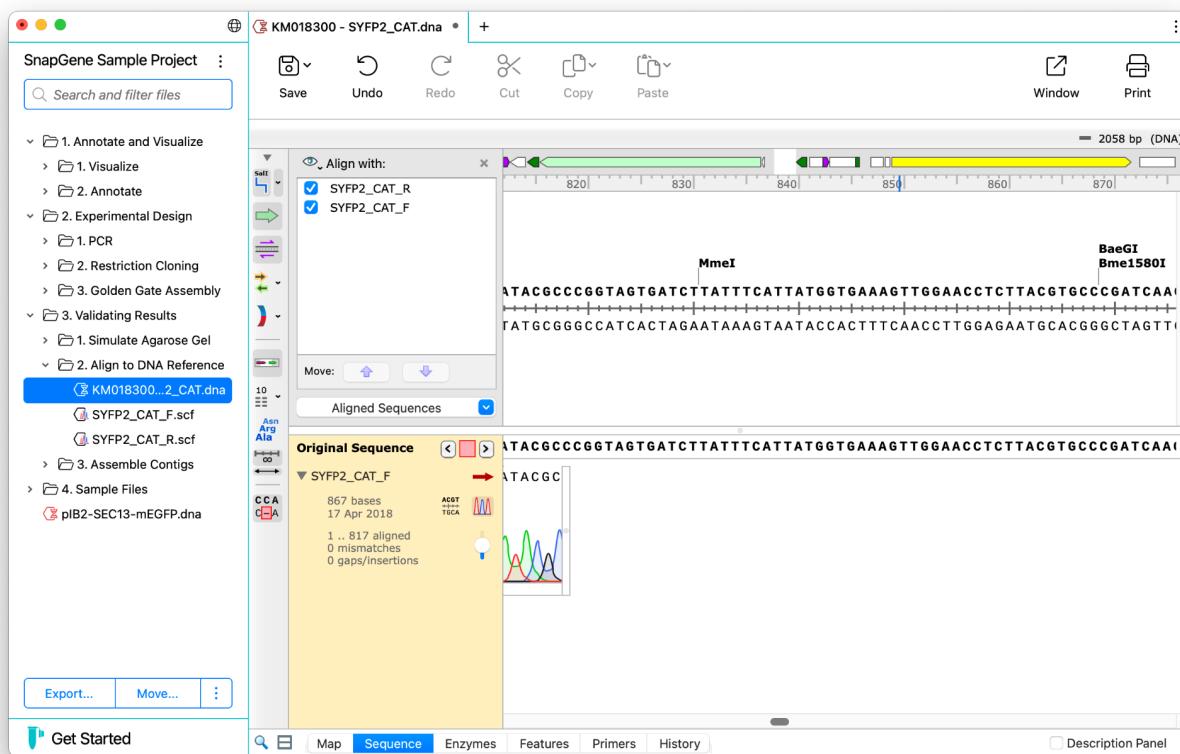
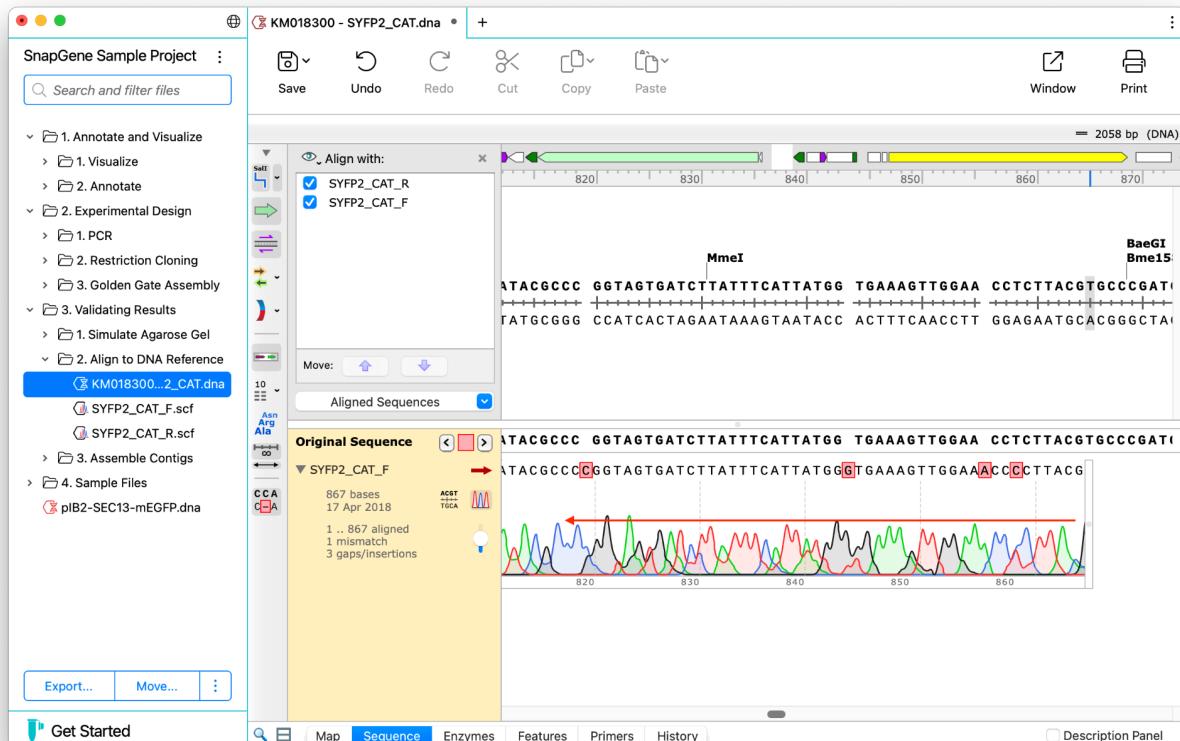
Switching to the sequence view by selecting the Sequence tab at the bottom of the window will show this alignment in more detail. Selecting the dropdown arrow highlighted in red in the screenshot below will toggle the chromatogram view on or off, allowing you to assess the quality of the sequencing data. The other settings highlighted in red will allow you to adjust other view settings for the chromatogram, such as showing features instead of the chromatogram peaks, or adjusting the size of the peaks.



The majority of this sequencing data seems to align to the reference sequence well. However, as we scroll towards the end of the forward read, we can start to see a few mismatches reported in the sequencing data.



The quality of this region of the data looks to be far lower than the earlier regions of the chromatogram. This can be excluded from the alignment by dragging the slider at the end of the sequence towards the left.



Scrolling to the middle of the reference sequence, where the reverse read begins to align shows a similar pattern with the quality of sequencing dropping off at the ends. This can be excluded from the alignment in the same way, this time dragging the slider to the right,



Other than these two regions of lower quality, the data seems to align well to the reference sequence, showing that the reference sequence is correct in these regions. More information about how to use the align to DNA reference tool in SnapGene is provided in our [User Guide](#).