

# SEQUENCHER®

## Tutorial for Windows and Macintosh

---

### Trimming Sequence

© 2016 Gene Codes Corporation

Gene Codes Corporation



Gene Codes Corporation  
775 Technology Drive, Ann Arbor, MI 48108 USA  
1.800.497.4939 (USA) +1.734.769.7249 (elsewhere)  
+1.734.769.7074 (fax)  
[www.genecodes.com](http://www.genecodes.com) [gcinfo@genecodes.com](mailto:gcinfo@genecodes.com)

# Trimming Sequence

There are two reasons why you may want to trim a sequence that you have imported into **Sequencher**. The first has to do with quality. **Sequencher's** quality trimming functions are described in detail in the **Quality Scores** tutorial and in the **Tour Guide**. The second reason for removing data from the ends is not quality specific, but is rather sequence specific. **Sequencher** has a command for removing vector sequence, **Sequence > Trim Vector...**, but this function is much more versatile than its name implies.

In this tutorial, you will become familiar with **Sequencher's** sequence trimming functions, starting with vector trimming and then focusing on trimming intron sequence from amplified exons. You can also use **Sequencher** to trim primer sequence and there are special options built into the vector trimming function that facilitate the trimming of transposons.

## TRIMMING VECTOR

Vector contamination is the inevitable result of sequencing DNA libraries. Whether you have shotgun sequencing libraries or EST libraries, your sequencing primer originates in your vector, so your sequencing data will always contain some vector sequence at one or both ends of your insert. The vector trimming tool in **Sequencher** allows you to trim one sequence at a time, or thousands of sequences at a time. It is an invaluable tool in large sequencing projects.

The parameters for the vector trimming function are defined in the **Vector Insertion Sites** dialog, accessible from the **Window > Specify Vector Insertion Sites...** command. The trim is executed from the **Sequence** menu, **Trim Vector...** In this example, we will trim BlueKSM vector from 10 sequences that were cloned into the vector's BamHI site.

- Launch **Sequencher**.
- Choose **New Project** from the **File** menu.
- From the **File** menu, select **Import > Sequencher Project...** and navigate to the **Sequencher > Sample Data** folder.
- Choose **Trimming Vector.SPF** and then click **Open**.

You will have imported 5 forward and 5 reverse sequences into the **Project Window**.

- Under the **Window** menu, select **Specify Vector Insertion Sites...**

The top of the window describes the default parameters for the recognition of the vector sequence. For most uses, the default parameters work well. You may, however, adjust these parameters to suit your project.

Minimum overlap to consider as vector contamination:	3	▲▼
Approx. match percentage to consider as contamination:	80	▲▼
Minimum overlap allowed without exact matches:	10	▲▼
Additional bases to remove from a contaminated 5' end:	0	▲▼
Additional bases to remove from a contaminated 3' end:	0	▲▼

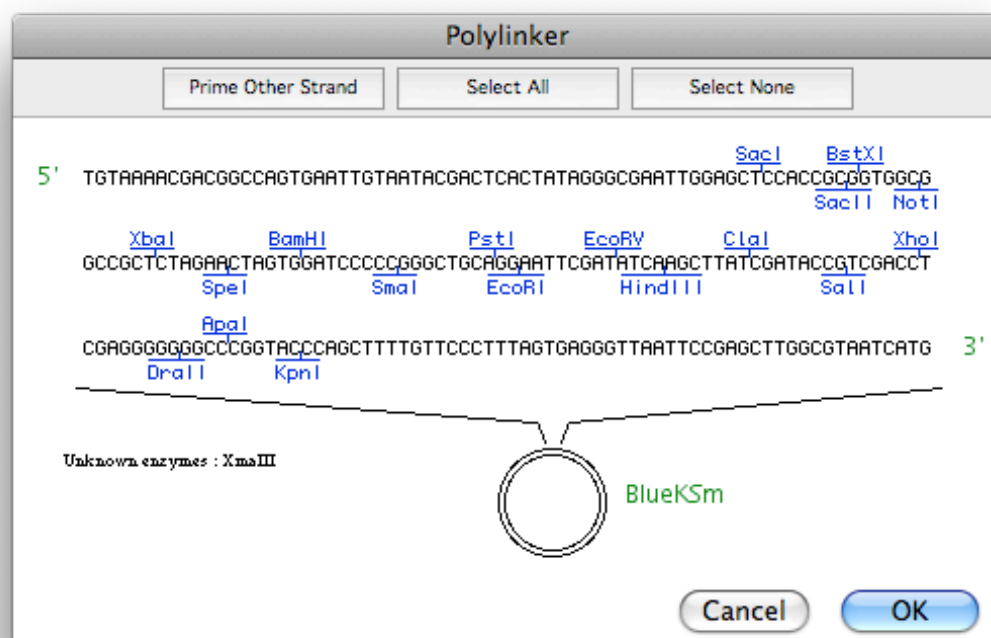
If a sequence has a lot of ambiguity in the region of the vector junction, you can decrease the *Approx. match percentage to consider as contamination* and thereby increase the likelihood that **Sequencher** will recognize the vector sequence. If it is essential to remove even a single base of vector contamination, you may want to decrease the *Minimum overlap to consider as vector contamination*. Gene Codes' technical staff is also happy to work with you to optimize your trimming parameters.

- Click on the button in the middle of the window, **Use Vecbase File...**

Use Vecbase File...

If the VecBase file is already loaded, then **Sequencher** will display the list of vectors in the VecBase. Otherwise:

- Open **3 Vectors from VecBase.TXT** in the **Sequencher Sample Data** folder.
- Select **BlueKSm** and click on **Select**.



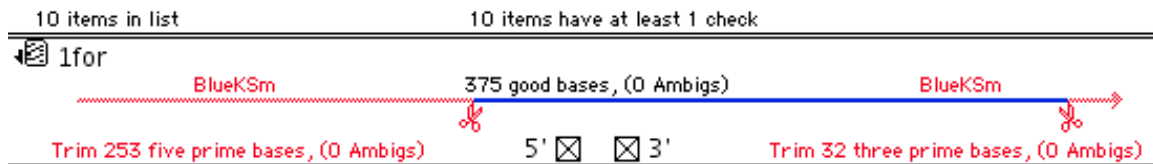
- Select the cut site **BamHI**. The BamHI site will now display in a contrasting color.
- Click on the **OK** button.

The sequence flanking the BamHI insertion site is now in the first sequence window. The location of the insertion is marked with a "•".

- Close on the **Vector Insertion Sites** dialog.

If you were interested in repeating this vector specification in a different session with **Sequencher**, you could save your settings by clicking on the **Save Sites** button before closing the dialog.

- To execute the trim, select all of the sequences in the **Project Window** by selecting **Select > Select All**.
- From the **Sequence** menu, choose the **Trim Vector...** command.



Immediately, **Sequencher** opens the **Vector Contamination** dialog, graphically displaying all of the vector sequences as red lines flanking a blue line, which represents the insert. If you click on the **Show Bases** button, **Sequencher** will display the specific bases surrounding the insertion site so you can judge the validity of the proposed trim.

**Sequencher** also allows you to sort the sequences in this dialog according to a number of parameters including the severity of the trim. If you find that your parameters are too stringent or too loose, then you may adjust them before executing the trim.

- If you are satisfied, click on the **Trim Checked Items** button and confirm the action with the **Trim** button.
- Close the **Vector Contamination** dialog.

The sequences are now trimmed and will assemble into one contig with no mismatches and a complete open reading frame.

## BATCH REVERTING TRIMMED ENDS

After you have made the initial trim of your sequences in the **Project Window**, you may feel that the trim criteria were too stringent. You can revert all of the trim or a portion of the trim for the current sequence selection by going to the **Sequence** menu, selecting the **Batch Revert Trim Ends...** menu item, and entering the number of 5' bases to revert and/or the number of 3' bases to revert.

Batch Revert Trim Ends Criteria

☒ Number of 5' bases to revert

☒ Number of 3' bases to revert

## TRIMMING INTRONS

Just as vector contamination is the natural result of sequencing libraries, intron contamination is often the undesired consequence of sequencing PCR amplified exons. In order to amplify and sequence a complete exon, the sequencing reaction must be primed in the intron. Very often the intron data is not relevant to the study, so it is preferable that it be removed. The same functions that are used to recognize vector sequence can be used to remove any sequence.

In this example, we will remove the flanking intron sequences from exons 2, 3, and 4 from HLA:A, and then assemble a virtual cDNA using a Reference Sequence.

- Under the **File** menu, select **New Project**.
- Under the **File** menu again, select **Import > Sequencer Project...** and open **Trimming Introns.SPF**.

You will now have six autoseq sequences and one text sequence selected in your **Project Window**.

- Under the **Window** menu, select **Specify Vector Insertion Sites...** and click on the **Load Sites** button.
- Select the file **Introns234.VCT** in the **Sequencer > Sample Data** folder and then select the **Open** button.

The sequences that are loaded into this window were identified from the assembly of the genomic sequence with the cDNA sequence. For more information on this type of assembly, refer to the **Assembly Strategies** tutorial.

#1	Name:	FlankingExon2
	Polylinker:	CGCAGGACCGGGGAGCCGCGCCGGGACGAGGGTCGGGCAAGTCTCAGCCACTGCTC CCCCAG●GGTGAAGTACCCCGGCCGGGGCGAGGTCAGGACCCCTCATCCCCAC GGACGGGCCAGGTCGCCACAGTCTCCGGTCCAGATCCACCCGAGCCGCGGG
	<input style="width: 100px;" type="button" value="Use Vecbase File..."/>	
#2	Name:	FlankingExon3
	Polylinker:	AGGCCAAATATCCCCCGGGTTGGTCGGGGCTGGGCGGGCTCGGGGACTGGGCTG ACCGCGGGGTCGGGGCCA●GTACAGGGGCCACGGGCGCTCCCTGATCGCTGTAG ATCTCCCGGGCTGGCCTCCACAGGGGGAGACCATTTGGGACCCACACTAGGATA TCACCTTCTTTGGTTCTGAGGGAGGGGATTCTTCTTGTT
	<input style="width: 100px;" type="button" value="Use Vecbase File..."/>	
#3	Name:	FlankingExon4
	Polylinker:	CTCAGATGGCCACATGCGTCTGGTGGAGTGTCCCATGACAGATGCARAATGCCTG ATTTTCTGACTCTTCCGTCAG●GTAGGAGGAGATGGGGGTGTCATGTCTCTTA GGGAAGCAGGAGCCTCTCTGGAGCCTTTAGCAGGGTCAGG
	<input style="width: 100px;" type="button" value="Use Vecbase File..."/>	

The asterisk or bullet in the middle of the sequence represents the exon. **Sequencher** identifies the trimming site from the sequences 5' and 3' of the •.

- Close the **Vector Insertion Sites** dialog.
- Check to be sure that your sequences are still selected.

Before you trim sequence for vector or introns, it is often wise to trim first for quality by executing the **Trim Ends...** command. For more information on trimming ends, see the **Quality Scores** tutorial.

- Execute the trim from the **Sequence** menu by selecting **Trim Ends...**

**Sequencher** displays a graphic representation of the proposed trim. With this set of data, there are no quality scores, so **Sequencher** uses a concentration of ambiguous base calls to define poor quality data.

- Click on the **Change Trim Criteria** button.
- Enter the following criteria in the second line of 5' trim pane and check the corresponding box.

<input checked="" type="checkbox"/>	Trimming no more than 25 %, trim until the first	25	bases contain less than	3	ambiguities.
-------------------------------------	--	----	-------------------------	---	--------------

- Enter the following criteria in the second line of 3' trim pane and check the corresponding boxes.

<input checked="" type="checkbox"/>	Starting	100	bases after 5' trim, trim the first	25	bases containing more than	3	ambiguities.
<input checked="" type="checkbox"/>	Trim from the 3' end until the last	25	bases contain less than	3	ambiguities.		

- Click **OK**.

**Sequencher** flags the 3' ends of Exon4R and Exon4F for trimming.

- Click on the **Trim Checked Items** button.
- Click on the **Trim** button.
- Close the **Ends Trimming Criteria** dialog.

You are now ready to trim the intron sequences that you defined previously.

- Insure that all of your sequences are still selected.
- Execute the trim from the **Sequence** menu by selecting **Trim Vector...**

Note: it is not necessary to deselect the Reference Sequence. Sequences that do not match the trimming criteria will not be affected.

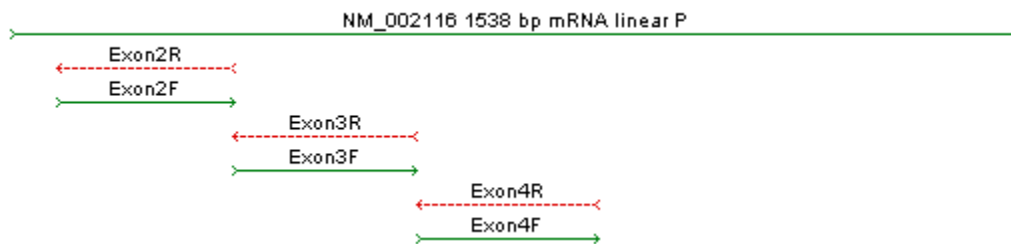
The names of each of the trimming sequences are listed at the positions where the trim will take place. For instance, Exon2F will be trimmed on both the 5' and 3' ends with the Flanking exon2 sequence entered into the **Vector Insertion Sites** dialog.

- Select the **Trim Checked Items** button, then the **Trim** button, then close the **Vector Contamination** dialog.
- Make sure all of the items in the **Project Window** are selected by either dragging your mouse around the items or using the menu item **Select > Select All**.

The Reference Sequence has an “R” in the icon to indicate that the sequence has special properties. For more information on the Reference Sequence, see the **Reference Sequence** tutorial.

- Click on the **Assemble to Reference** button and then close the **Assembly Completed** dialog.
- Double-click on the **Contig [0001]** icon to open the **Contig Editor Overview**.

The new contig is a virtual cDNA created from the virtual blunt-end ligation of the trimmed exon sequences.



The Vector Trimming function allows you to target any known sequence for trimming. Just as it is used in these examples to remove vector and intron, it has been used to remove primer sequence, transposons, and polyA tails. It is a flexible and powerful tool in your bioinformatics arsenal.

- Close out of the **Contig Editor** dialog.
- Select **Close Project** from the **File** menu to close the project and no need to save changes when prompted.
- Exit out of **Sequencher** if desired.