MacVector 13.5

for Mac OS X

Getting Started
with
MacVector 13.5

Support:
1 866 338 0222
+44 (0) 1223 410552
support@macvector.com

MacVector, Inc.
Software for Scientists

MacVector Resources

Tutorials

A number of tutorials are available for download;

http://www.macvector.com/downloads.html

Videos

http://www.macvector.com/Screencasts/screencasts2.html

Manual

There is a downloadable PDF version of the manual (12.0) at

http://www.macvector.com/downloads.html#MacVector12UserGuide

Discussion Forums

To post questions or follow ongoing discussions, check out the user forums at;

http://www.macvector.com/phpbb/index.php

Copyright Statement

This version of the workshop guide was published in November 2014.

Contents

Introduction

MacVector 13.5 is the culmination of a series of significant updates to MacVector over the last seven years that feature a redesigned tabbed interface, OS X toolbars, increased interactivity between views, improved sequence and feature editing, an Auto Annotation function, a new Primer3 based primer design module, a revised *Entrez* browser, support for Next Generation Sequencing data analysis and numerous other enhancements. This document provides a basic overview of MacVector functionality for new users, in-depth discussions of common functionality and practical examples of how to use new functionality introduced in recent versions. All of the sequence files used in the examples can be found in `/Applications/MacVector/Sample Files/` or `/Applications/MacVector/Tutorial Files/`.

Getting Sequences Into MacVector

You can open saved sequence files in many different formats, copy sequence data from other applications and paste it into a new Sequence window, or open sequences directly from *Entrez*.

Importing sequence files

The **Starting Point** dialog is displayed every time you start MacVector and provides quick and easy access to sample files and the other folders and files you use most.

To open a MacVector-supplied sample sequence to get you going using the **Starting Point** dialog, choose Sample File as the starting point from the list on the left, then select a sequence from the list of files on the right and click Choose.

You can also open sequence files from the main MacVector menu. To do this, select **File | Open** from the menu, browse to the required folder, select the sequence file from the list, and click Open.

MacVector can automatically identify and open sequence files in all of the major Text formats (GenBank, EMBL, FastA, GCG), as well as chromatogram ("ABI") files and sequence files from other programs such as VectorNTI, LaserGene, GeneWorks and DNAStrider.

Creating new sequences

You can also create an entirely new sequence file and paste content into it. To do this, choose **File | New** from the menu and select the required sequence type from the submenu. Then paste valid DNA or protein sequence characters into the new sequence file.

Alternatively, copy and paste the sequence data to the clipboard first, then select **File | New from Clipboard** from the menu.

Note: As of MacVector 12.6, you can copy text data containing annotated sequences (e.g. in GenBank or EMBL format) and MacVector will parse the data and insert a fully annotated sequence into the target sequence.

Opening sequences from Entrez

You can locate and open complete sequences from the online NCBI databases directly within MacVector.

Choose **Database | Internet Entrez Search...** from the menu and then pick the database you want to search from the Database drop-down menu.

Note: The Protein Sequence Record database for proteins and the Core Nucleotide db for DNA are recommended and should appear at the top of the list of databases.

Perform a search by choosing a category in the first All Fields drop-down menu, typing appropriate text in the adjacent text box and clicking Search.

Any matches are listed in the top results panel. Double-click on a result to display the sequence in a Sequence window on the MacVector desktop.

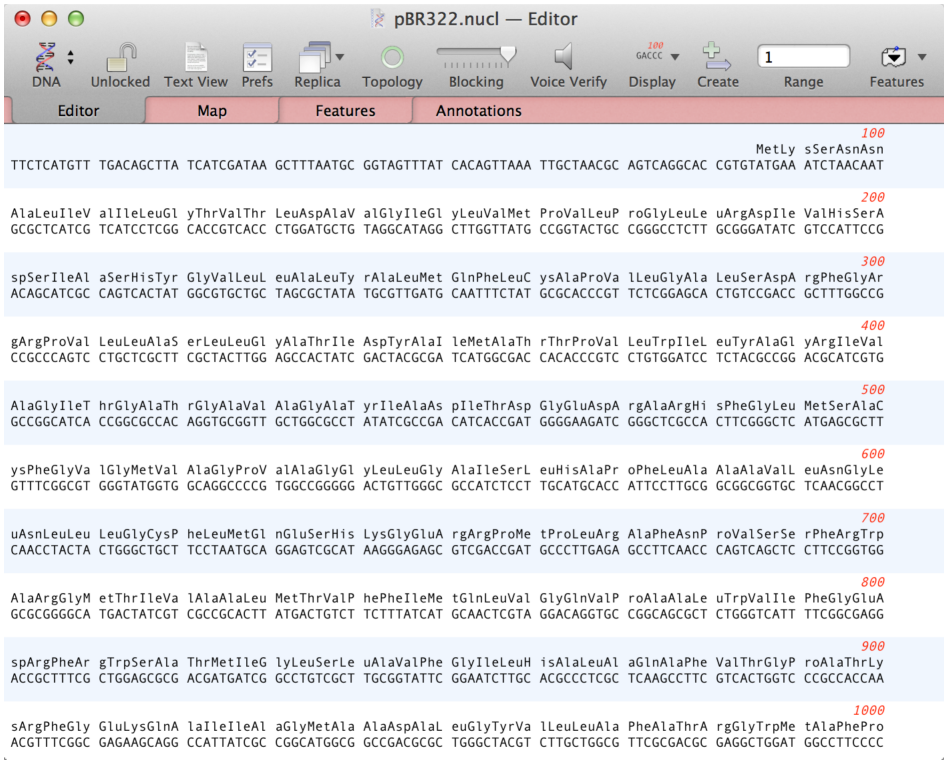
Note: You can select multiple results and open them all in separate Sequence windows by clicking To Desktop.

Try this: To find all the ribosomal protein sequences from the organism "Canis" select the Protein sequence record database, click on the **+** button to add another query field and Search using "Search the Field" Organism for `canis*` And Text Word for `ribosomal`.

Working With Sequences in MacVector

Tabbed Interface

When you open a sequence using MacVector you will se a single window with a set of tabs offering different views of the sequence.

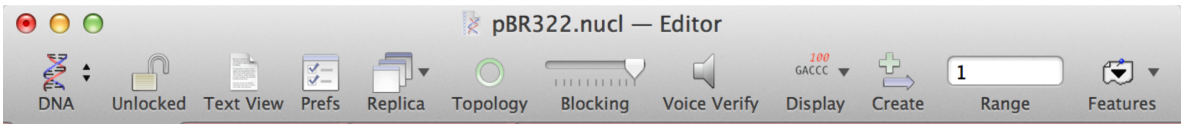


The color-coding MacVector has always used for distinguishing nucleic acid and protein sequences is retained in the background of the tab bar, where light red indicates a nucleic acid sequence and blue indicates protein.

- **Editor** - this displays the standard editor view of the sequence. You can display the sequence in single stranded and double stranded forms and also show translations above and below the sequence.
- **Map** - this displays the graphical view of the sequence, where features are displayed graphically according to the settings you have applied for the sequence. This view also displays a default set of restriction enzyme sites.
- **Features** - this displays a list of features associated with the sequence using a modern OS X style list display. The list can be sorted and features can be edited by simply double-clicking on them.
- **Annotations** - this displays "annotations" which in MacVector parlance is all the data associated with a sequence that does not have a defined start and stop location on the sequence. Like the Features tab, it uses a modern OS X style list display and you can sort and edit annotations with simple mouse clicks.

Note that you can switch between tabs by holding down the command key pressing "1", "2", "3" or "4" on the keyboard in addition to clicking with the mouse.

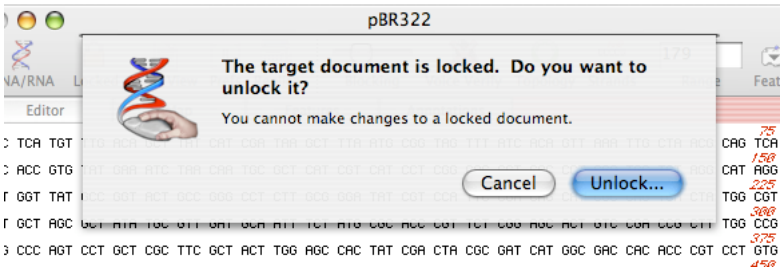
The Editor Tab



Common Toolbar Items

The first six toolbar buttons are common to all of the tabs.

- **DNA/RNA** - use this button to toggle the molecule between DNA and RNA. (Protein windows show a non-functional Protein icon).
- **Locked** - the padlock helps prevent you from inadvertently modifying a sequence - you are prompted to unlock the sequence if you try to perform a destructive edit.

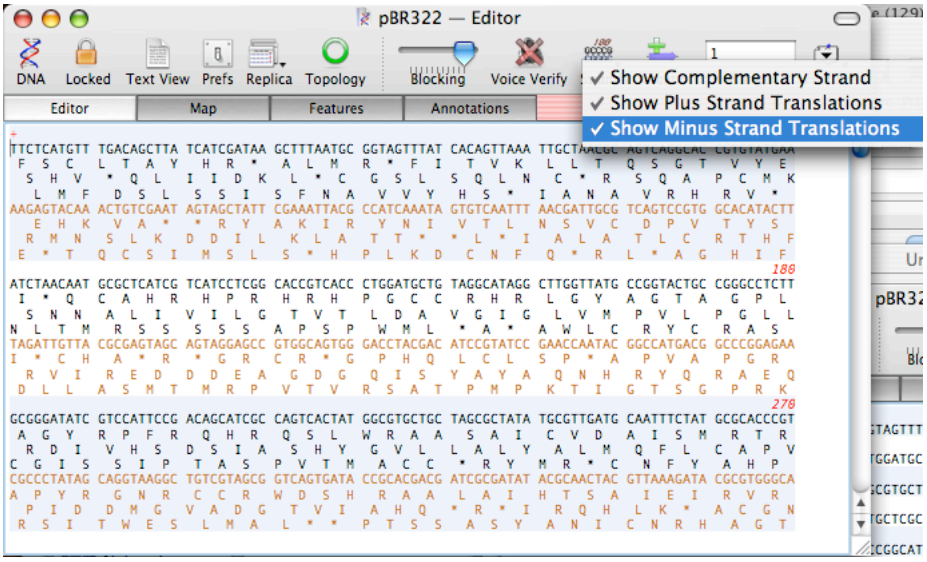


- **Text View** - this opens a separate window with the "Annotated Sequence" text view.
- **Prefs** - this opens a Preferences Pane dialog, usually set to the Fonts pane. Clicking on this is equivalent to selecting the **MacVector | Preferences**... menu item and switching to the Fonts tab.
- **Replica** - this opens a second identical window. This is useful if you want to view e.g. the Editor tab and the Map tab at the same time. You can open as many sequence windows as you like - changes in one window will be reflected in the other open windows. The Replica button displays a dropdown menu that lets you select which tab you want the replica window to open displaying.
- **Topology** - toggles between linear and circular. This affects certain functions such as restriction enzyme searching. For example, pBR322 has an *Eco RI* site at its origin - in circular mode, the site will be displayed, but in linear mode the site is considered to be "split" and with half appearing at each end and thus will not be shown. Sequences *must* have a circular topology before you can view them as circular molecules in the Map tab.

Editor Specific Toolbar Items

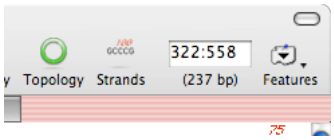
- **Blocking** - this adjusts the "blocking" of the sequence residues to make them easier to visualize.
- **Voice Verify** - turns on/off the voice playback. Turn this on to hear each residue spoken out loud as you type it in.

- **Display** - with earlier versions of MacVector, this simply toggled the complementary strand on and off. With MacVector 13, you can use this to display a 3 or 6 frame translation underneath the sequence, or to display translations of any CDS features present on the sequence. The translation uses the currently selected genetic code and honors the one vs. three-letter amino acid code set in the Text Display preference pane. The color of the complementary strand can be set using the Colors preference pane.



- **Create** - this button opens the Feature editor with the start and stop locations of the new feature set to the current selection.

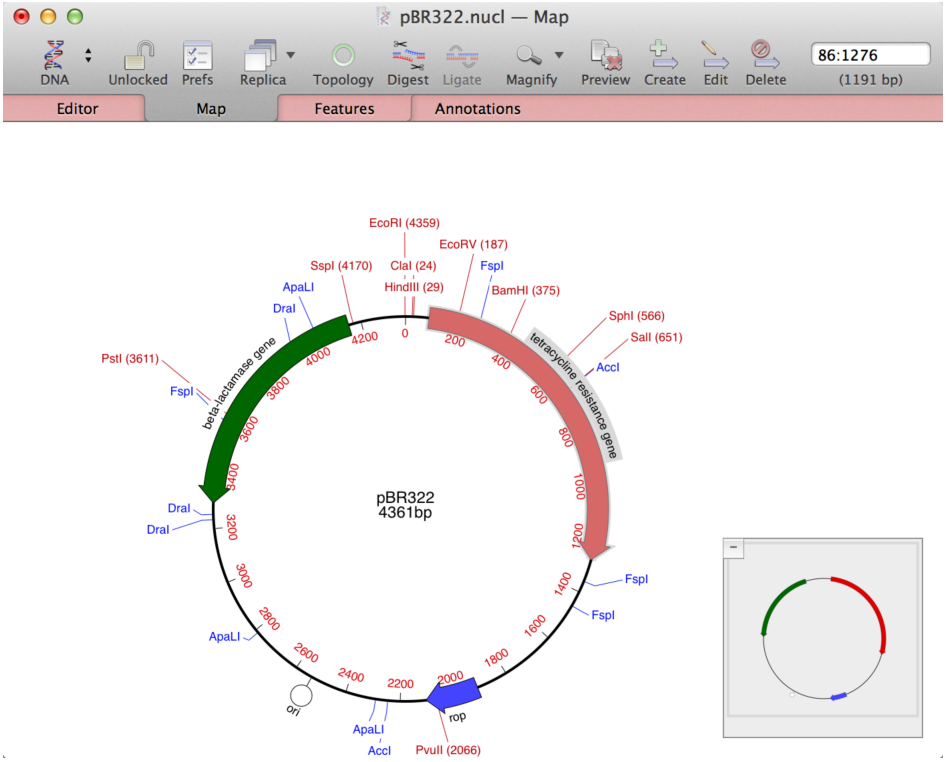
- **Range** - shows the current selection as a range. You can also type into this box to either jump to a specific location or to select a particular range by separating numbers with the ":" character. The "Range" text is replaced by the number of residues selected when a selection is in place.



Note: When you type or paste in the range box, it acquires the “focus”, indicated by a blue focus ring. If you want to then copy the selected sequence, you must switch the focus to the actual sequence pane. The easiest way to do this is to press the `<esc>`, `<enter>` or `<tab>` keys. Any of these will move the focus to the main sequence editor pane without losing any selections.

- **Features** - you can click on this to display a popup menu with all of the features associated with the sequence and choose one to select that region.

The Map Tab



The Map tab displays a graphical representation of the sequence, including the location of common restriction enzyme sites. The enzymes used for this analysis are controlled by the Map View Options dialog. This can be opened by clicking on the Prefs button, or by selecting the **Options | Map View Options...** menu item. The use of this is discussed in more detail later.

- If you double-click on an enzyme name, all of the enzymes of that name become selected so you can quickly identify multiple cut sites.

- Unique sites are displayed in a different color to others (the default color is red).

- If you have copied a DNA fragment to the clipboard, any sites that are compatible with the ends of that fragment are displayed with a pastel red (left end) or green (right end) background.

- An overview (bottom right in the screenshot) quickly enables you to visualize your sequence and your position. The size of the overview is controlled by a setting in the **MacVector | Preferences | Map View** pane. You can always hide the overview by clicking in the "-" button in the top left corner of the overview. Once collapsed, click on the "+" button in the bottom right corner of the window to show the overview.

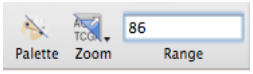
Map Specific Toolbar Items

- **Digest** - this button is enabled whenever two restriction enzyme sites have been selected. Clicking this will copy the sequence between

the two sites to the clipboard along with the structure of the ends produced by the enzymes. This is discussed in more detail later.

- **Ligate** – this is enabled whenever a sequence is present on the clipboard and one or more restriction enzymes sites have been selected. Clicking on the button will bring up a ligation dialog that lets you manipulate the ends of the fragments and/or flip the source fragment before inserting into the target molecule. Again, this is discussed in more detail later.
- **Magnify** – this lets you choose a magnification value for the view, using a dropdown menu.
- **Preview** – this turns on the display of page outlines so that you can see exactly how the graphic will be split between pages when sent to the currently selected printer.
- **Create** - this invokes the Feature Editor, creating a new feature.
- **Edit** - this is only active if you have a feature selected. It brings up the new Feature Editor, loaded with the existing details of the feature you have selected. Note that this lets you edit the GenBank-style representation of the feature. To edit the feature appearance, double-click on the feature to bring up the Symbol Editor.
- **Delete** - this deletes the selected feature. This physically deletes it - if you just want to hide a feature from display use the floating graphics palette window to turn it off.

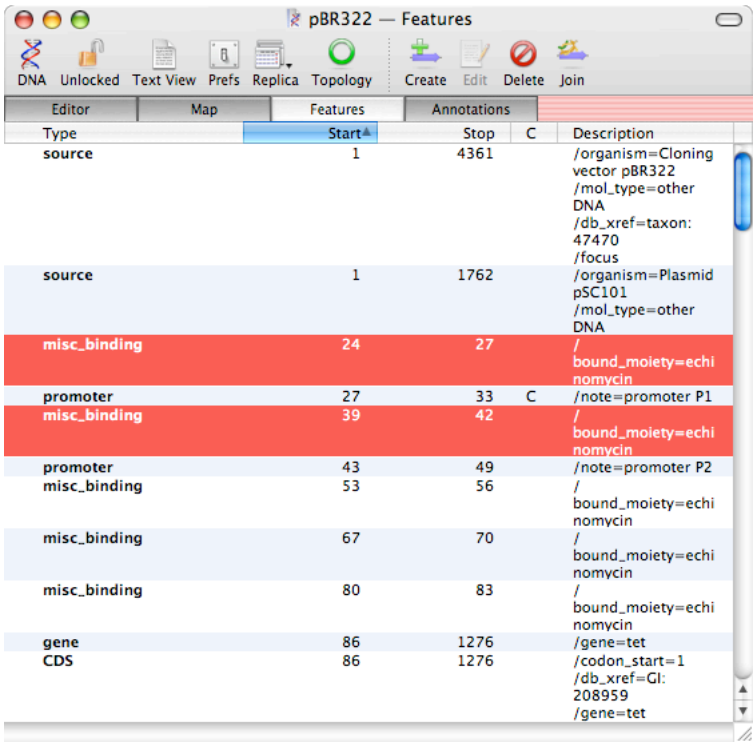
In addition to the default buttons, there are a number of other useful buttons that can be added using the Customize Toolbar menu.



- **Palette** - this toggles the visibility of the floating graphics palette window.
- **Zoom** - this toggles between Fit to Window and Zoom to Sequence – essentially shortcuts to those functions in the floating graphics palette.
- **Range** - this is identical to the range menu in the Editor tab.

Note: MacVector 11 made a change to the way circular sequences are handled in the Map tab and in the floating graphics palette. **Sequences can now only be shown as circular if the Topology of the sequence is set to be circular**. This is to prevent users thinking their sequence is being treated as circular in algorithmic calculations simply because it is being displayed as a circle in the Map tab. If the circular tab in the floating palette is disabled and you want your sequence to appear as a circle, press the Topology button so that it forms a circle.

The Features Tab



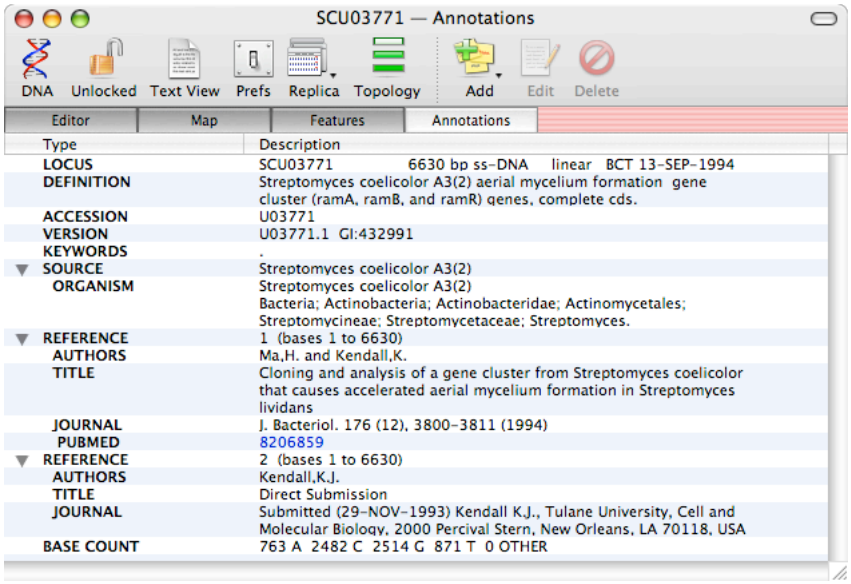
Type	Start	Stop	C	Description
source	1	4361		/organism=Cloning vector pBR322 /mol_type=other DNA /db_xref=taxon:47470 /focus
source	1	1762		/organism=Plasmid pSC101 /mol_type=other DNA
misc_binding	24	27		/bound_moiety=echinomycin
promoter	27	33	C	/note=promoter P1
misc_binding	39	42		/bound_moiety=echinomycin
promoter	43	49		/note=promoter P2
misc_binding	53	56		/bound_moiety=echinomycin
misc_binding	67	70		/bound_moiety=echinomycin
misc_binding	80	83		/bound_moiety=echinomycin
gene	86	1276		/gene=tet
CDS	86	1276		/codon_start=1 /db_xref=GI:208959 /gene=tet

Features Specific Toolbar Items

- **Create** - creates a new feature using the Feature Editor.
- **Edit** - only active if a feature is selected.
- **Delete** - deletes the selected feature(s).
- **Join** – joins two features together. You can use it to join two or more features to create a single segmented feature. This is particularly useful for annotating CDS features on a genomic sequence – you can annotate each exon region individually using other tools in MacVector, then join them all together at the end to create a standard GenBank segmented CDS feature.

The Annotations Tab

The Annotations tab displays all of the annotations associated with a sequence that do not have a specific location on the sequence. This includes such things as keywords, journal references and general comments.

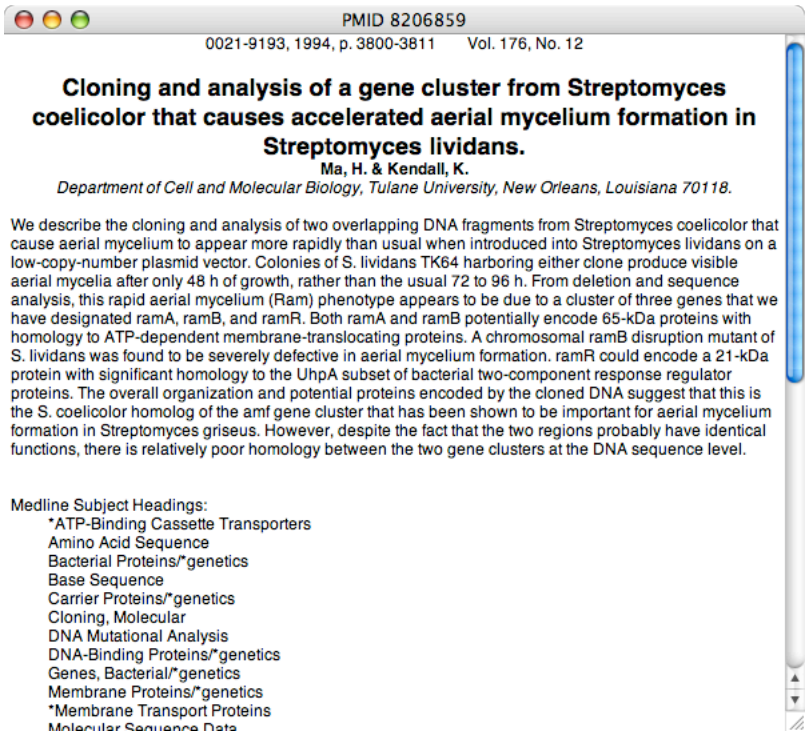


Annotations Specific Toolbar Items

- **Add** - creates a new annotation using the new Annotation Editor.
- **Edit** - only active if an annotation is selected.
- **Delete** - deletes the selected annotation(s).

PubMed Links

Functional PubMed (formerly MedLine) links are now shown in blue underlined text to note that they can be clicked on. If you click on one of these, MacVector will retrieve the corresponding abstract from the NCBI PubMed service and display it in a separate window. Note that prior to version 13 this required a ***double-click***, not just a single click. MacVector formats these using a typical journal abstract appearance.

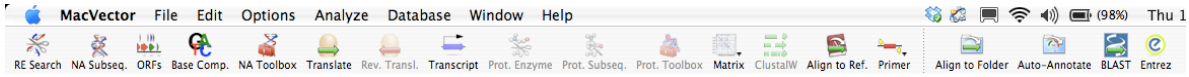


You can use **File | Save** to save abstracts in Rich Text Format (RTF).

Analyzing Sequences

Analysis Tools Toolbar

MacVector provides a large number of functions for analyzing DNA and protein sequences. Most of the analysis functions are accessed through the **Analyze** menu items, though there are also a few under the **Database** menu. All of these can also be accessed from the floating toolbar that runs across the top of the screen.



Users who are comfortable with the traditional menu-based analysis functions may wish to hide the toolbar – this can be done by choosing **Window | Hide Analyses Tools**.

The default toolbar contains both DNA and Protein analysis functions. Most of the buttons directly correspond to items in the **Analyze** menu. The buttons to the right correspond to functions in the **Database** menu. The buttons become enabled and disabled exactly as the menu items do, so functions like Restriction Enzyme searching (RE Search) and Translation (Translate) are only available if you have a nucleic acid sequence window currently active. Clicking a button is functionally identical to selecting the corresponding item from the appropriate menu.

Analysis Workflow Overview

Most MacVector analysis functions use a standard workflow;

(a) Make sure the sequence you want to analyze is the front-most window

(b) Select the analysis function from the Analysis Tools toolbar or from the **Analyze** menu.

(c) A setup dialog appears letting you assign the parameters to be used in the analysis

(d) The algorithm runs and then a "filter" dialog appears, letting you choose which outputs you would like to see and optionally apply filters that that only a subset of the results is displayed.

(e) Result windows are then opened, which you can typically interact with. When a result window is active, you can choose the analysis menu item (or toolbar button) again and the filter dialog will reappear, letting you change the data you want to be displayed.

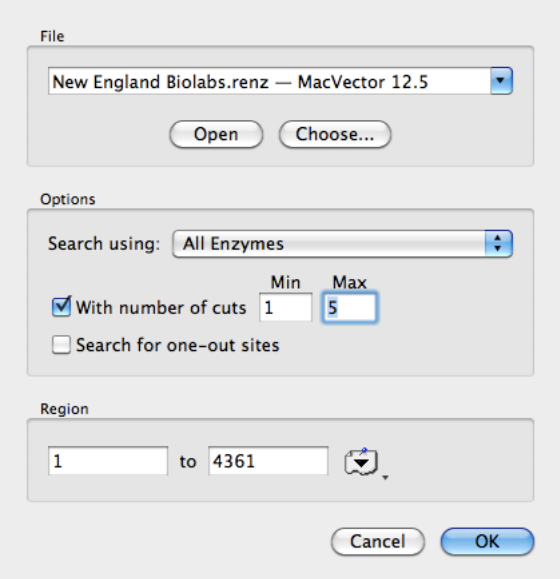
Performing a Restriction Enzyme Analysis

To illustrate a typical analysis workflow, we will perform a **Restriction Enzyme Analysis** on one of the MacVector sample sequences.

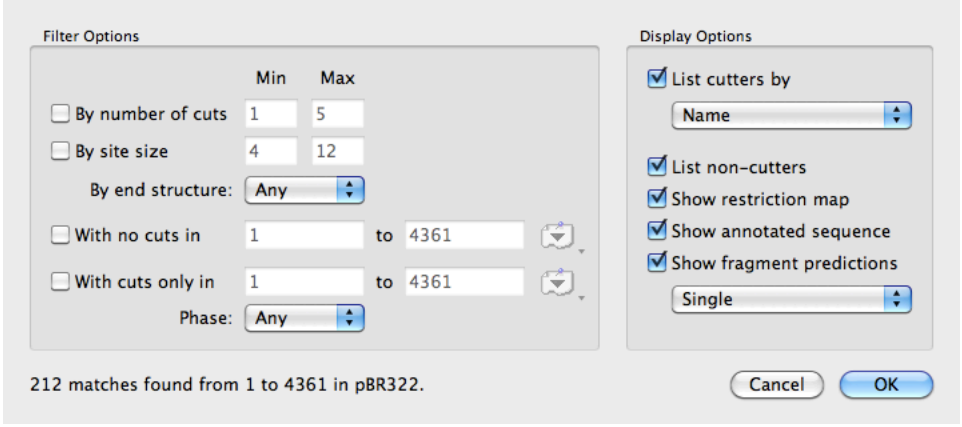
Open `/Applications/MacVector/Sample Files/pBR322.nucl`.

Choose **Analyze | Restriction Enzyme...**

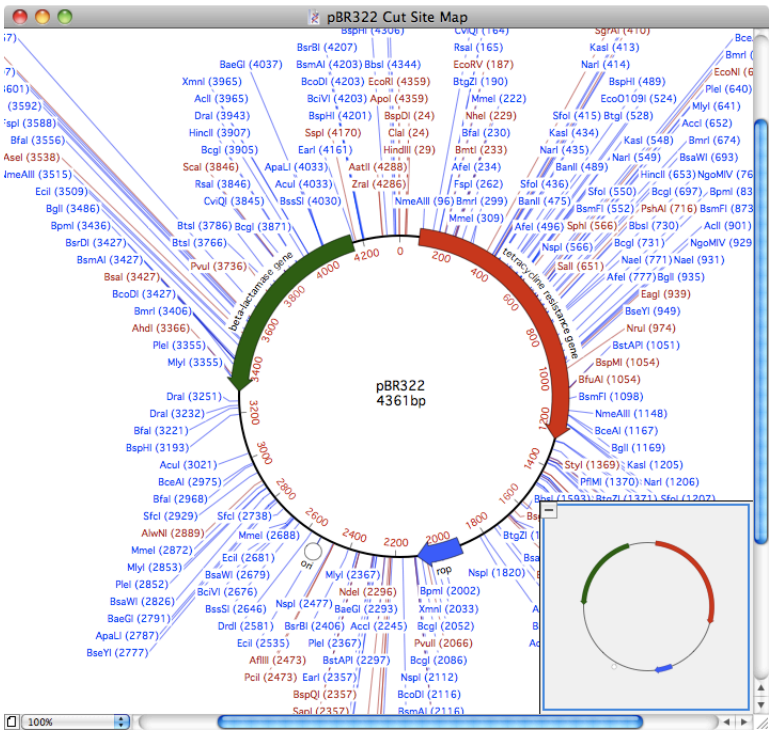
The setup dialog appears;



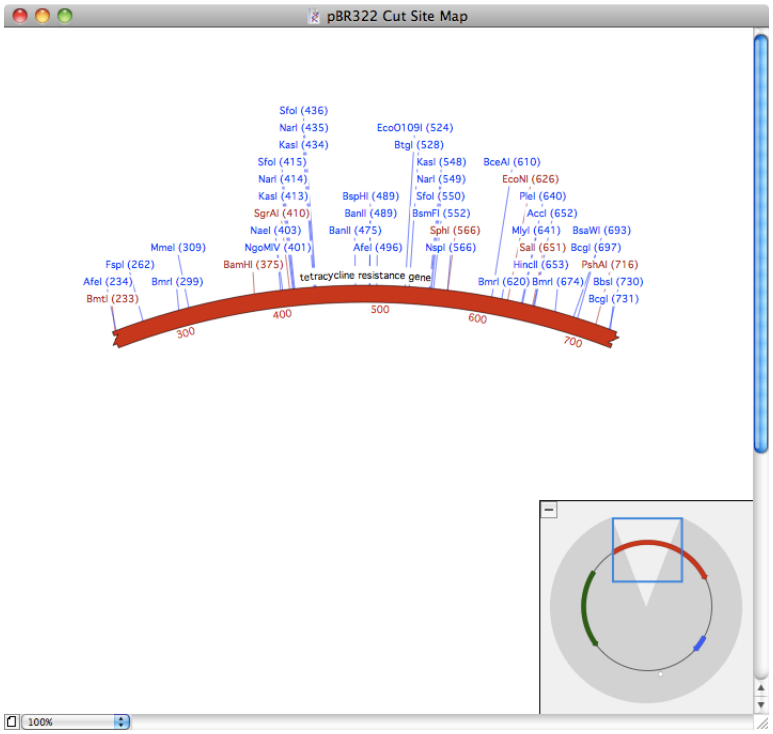
For this analysis, we'll Search using **All Enzymes**, but select the checkbox to restrict the number of cuts to between 1 and 5. After clicking OK the analysis runs and very quickly the filter dialog appears;



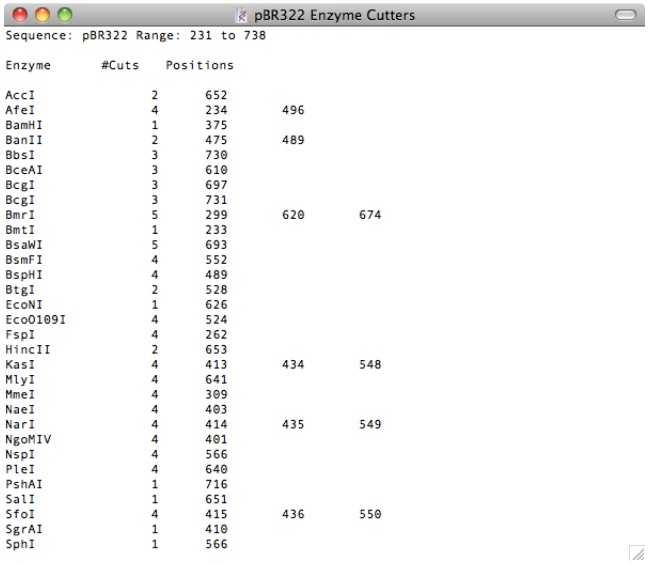
The Display Options section controls the result windows that will appear. For this example we will display all, but take particular notice of the Restriction Map window;



Click just inside the red tetracycline resistance gene, hold down the mouse and drag around the circle. An arc appears showing you the selected region. When you let go, the Map refreshes and is “zoomed in” to the region you selected;



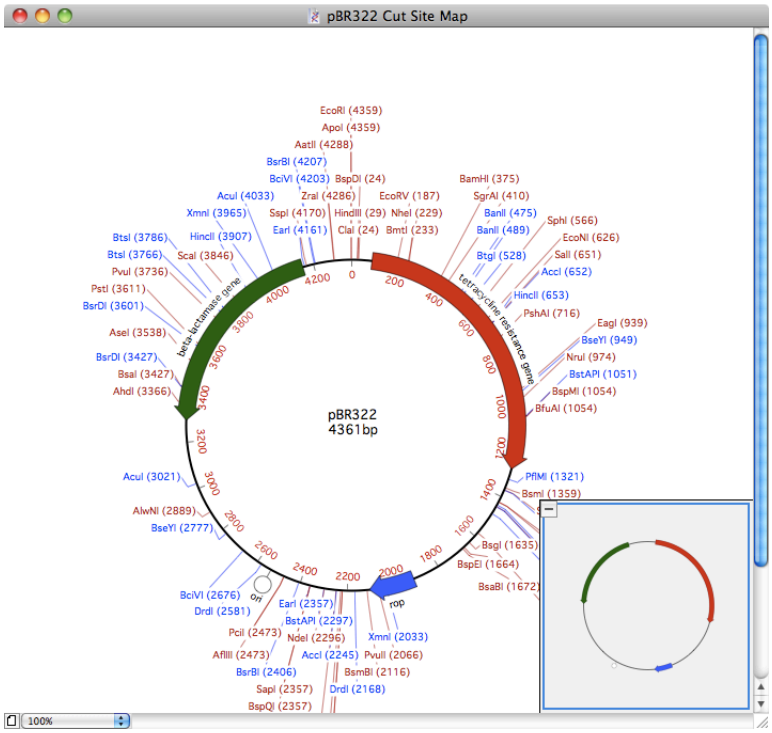
Not only does the Map view now display the zoomed region, but all of the other result windows refresh to show just the restriction sites present in the selected region;



Double-click in the Restriction Map window and the results will reset to the full range. You can also use the Up and Down arrow keys to increase or decrease the zoom level by two-fold increments. When zoomed in to a region, you can also use the Left and Right arrow keys to nudge the zoomed region to the left or right on the sequence.

Now choose **Analyze | Restriction Enzyme** once again with the Restriction Map window front most. The filter dialog reappears. Select the By number

of cuts checkbox and change the range to Min = 1 and Max = 2 and click OK. All the result windows refresh to display the new subset of result data;



If you want to start the entire Restriction Enzyme analysis again from scratch, bring the original `pBR322` window to the front and choose **Analyze | Restriction Enzyme...**

DNA Analysis Functions

Restriction Enzyme – this provides more in-depth analysis of sequences for restriction sites than the simple overview seen in the sequence Map tab. You can filter results based on number of cuts, type of overhang produced, size of site or even search for one-out sites. MacVector comes supplied with all the known restriction enzymes broken out by supplier, so if you primarily use just New England Biolabs enzymes for example, you can just search with enzymes in their catalog.

Nucleic Acid Subsequence – this lets you search sequences for short subsequences, such as transcription factor binding sites, promoter consensus sequences or even a database of your own primers. The search is extremely flexible – subsequences can have up to 3 parts with varying sized regions between gaps and you can define the number of mismatches permitted in the search for each individual part of each subsequence.

Open Reading Frames – you can scan a sequence for open reading frames, specifying the minimum length you want reported and even the genetic code to use.

Nucleic Acid Analysis Toolbox – this is a collection of different algorithms primarily aimed at helping you identify protein coding open

reading frames. As well as displaying possible open reading frames on each strand, there are algorithms that plot skewed base composition and codon usage analysis. You can zoom in to the plots to evaluate which open reading frames correspond to likely protein coding regions.

Base Composition – as well as a basic ACGT count, these algorithms list molecular weight and plot melting temperature, A/C/G/T content across the sequence and occurrence of di- and tri-nucleotides.

Primer3 – Primer3 is a popular 3rd party algorithm that scans target DNA to find pairs of primers suitable for PCR. You can lock one or both primers in position and find a matching internal primer for use in real-time PCR. The results a presented in both graphical and tabular form, letting you interact with them to easily copy any predicted products to save and/or insert into other vectors.

Quicktest Primer – this displays a floating dialog where you can paste/type primer sequences and get instant feedback on potential hairpin loops, melting temperature etc. The interactive display lets you nudge primers along a sequence, add mismatches and tails and even see the effects of mutations on open reading frames and restriction enzyme recognition sequences.

Sequencing Primers/Probes – similar to the PCR Primer Pairs, but finds just one primer using conditions more suited to sequencing or hybridization experiments.

Test Sequencing Primer/Probe – tests a single primer, showing potential binding sites and potential structural problems.

Translation – lets you translate one or more segments of a DNA molecule, creating a new protein sequence and/or displaying the codon usage for the translation.

Generate Transcript – you can use this to create a new RNA sequence from one or more segments of a DNA sequence. In particular, the interface lets you take advantage of existing features on the DNA, such as CDS features, RNA features, introns and exons to create a correctly spliced mRNA.

Create Dotplot->Pustell DNA Matrix – this performs a pair-wise comparison between two DNA sequences (or even for a single sequence against itself) and displays the alignment graphically as a dot-plot and also as text. You can zoom in the graphical window to focus in on individual aligned segments.

Align Multiple Sequences Using – lets you align multiple DNA sequences using the ClustalW, T-Coffee or Muscle algorithms. Note that none of these algorithms will "flip" DNA sequences to get a better alignment. If you think you need that functionality, you should probably be looking at the Align to Reference or Assembler features.

Align to Reference – this lets you align one or more DNA sequences against a reference sequence. The sample "Read" sequences can be "ABI" trace files, which will be displayed graphically in the window. The alignment WILL "flip" sequences to maintain better alignments and has

full editing and display functionality aimed at identifying mismatches between the Reads and the reference. You can also use Align to Reference to align cDNA clones against a genomic sequence, where the algorithm will also take canonical splice site donor and acceptor sequence into account when determining the intron/exon boundaries.

Align to Folder – scans a hierarchy of folders on your hard drive for sequences (in any format that MacVector can read) that match the sample sequence. With MacVector 13.5, you can scan NGS data files (e.g. in fastq format) and retrieve matching hits into a separate fastq file.

Auto Annotation – if you have a sequence that you have been sent from a collaborator or central lab, or that you downloaded from the Internet, you can use this function to automatically annotate known features on the sequence. You simply point the algorithm at a folder on your hard drive and it scans every sequence looking for features that match regions on the sample sequence. For features that match, it not only annotates the sequence, but also copies the graphical appearance, so you can use this to make sure that all your common genes have a consistent appearance.

Internet BLAST Search – performs an online BLAST search against the databases at the NCBI.

Aligning Sequences Overview

MacVector offers a number of different approaches for aligning sequences, each tailored to a different molecular biological problem. Most of the functions are applicable to both DNA and protein and some can even compare DNA with protein using the currently selected genetic code.

Internet BLAST – use this to scan your sequence against online databases at the NCBI to find known sequences that are most closely related and show each of those sequences aligned to yours.

Align to Folder – similar to BLAST, but this aligns your test sequence to sequences already saved on your hard drive (or accessible over a local network).

Pustell Matrix (Dotplot) – this compares two sequences using a graphical "dot-plot" approach along with a text alignment of the most significant matching segments. It can also be used to compare a sequence to itself to identify direct and inverted repeat regions. This is the best approach for getting an overall view of the relationship between two sequences as it can show duplications, rearrangements and also very weak similarity between two sequences that can be hard to identify by looking just at text alignments.

Align To Reference – this is an important alignment function in MacVector for comparing one or more sample DNA sequences against a known reference sequence. As it can directly handle and display chromatogram data ("ABI" files) and automatically "flip" source sequences to match the reference it is ideal for re-sequencing applications such as sequence confirmation, mutagenesis analysis or simply comparing

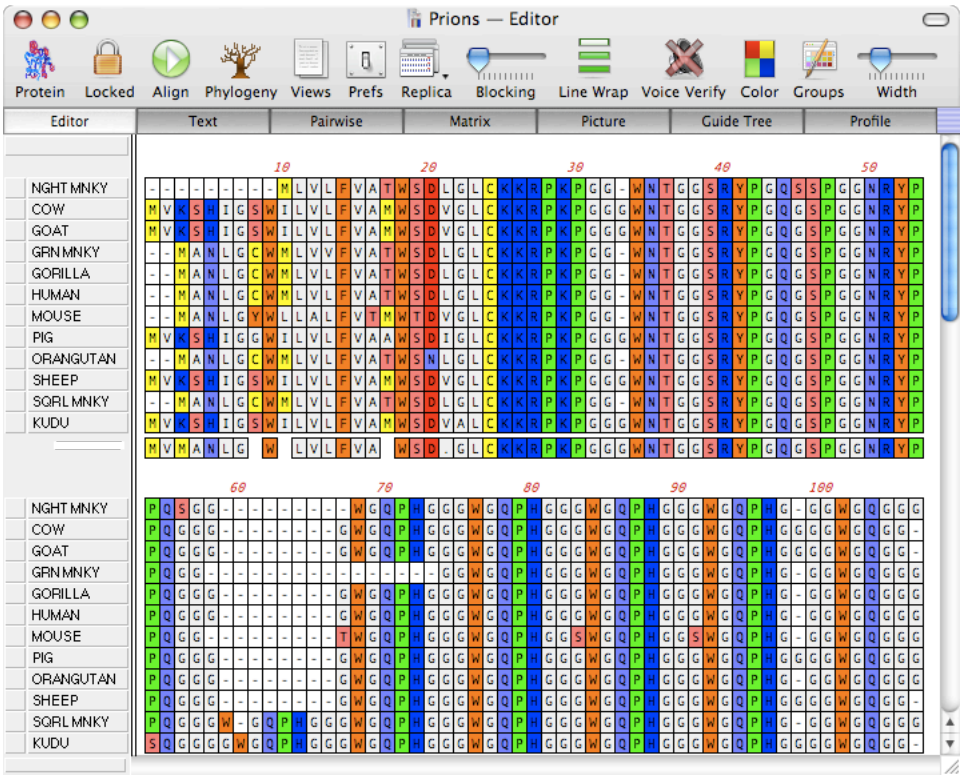
clones or related sequences to a reference. There is also a mode that can align cDNA clones to a genome, taking splice sites into account.

Multiple Sequence Alignment – this feature lets you align two or more sequences without requiring a reference. Tuned for protein sequences, but also applicable to DNA, alignments can be generated automatically using the popular ClustalW, T-Coffee and Muscle algorithms. The generated alignments can be edited and displayed in many ways and submitted to additional phylogenetic reconstruction algorithms to help determine the relationship between sequences.

Sequence Assembly – finally, with the use of the optional Assembler module, you can assemble multiple "Read" sequences into a single consensus sequence using the popular phred, phrap and cross_match algorithms. Assembler can also align many millions of "next generation sequencing" reads against one or more genome sequences using the fast Bowtie algorithm.

The Multiple Sequence Alignment Window

Like the single sequence windows described above, the multiple sequence alignment window also uses a tabbed interface. Open the file `/MacVector/Sample Files/Prions`.



The Editor tab provides the only editable view of multiple sequence alignments. You can cut/copy/paste and perform a variety of editing functions in this tab. All of the other tabs provide read-only views of the

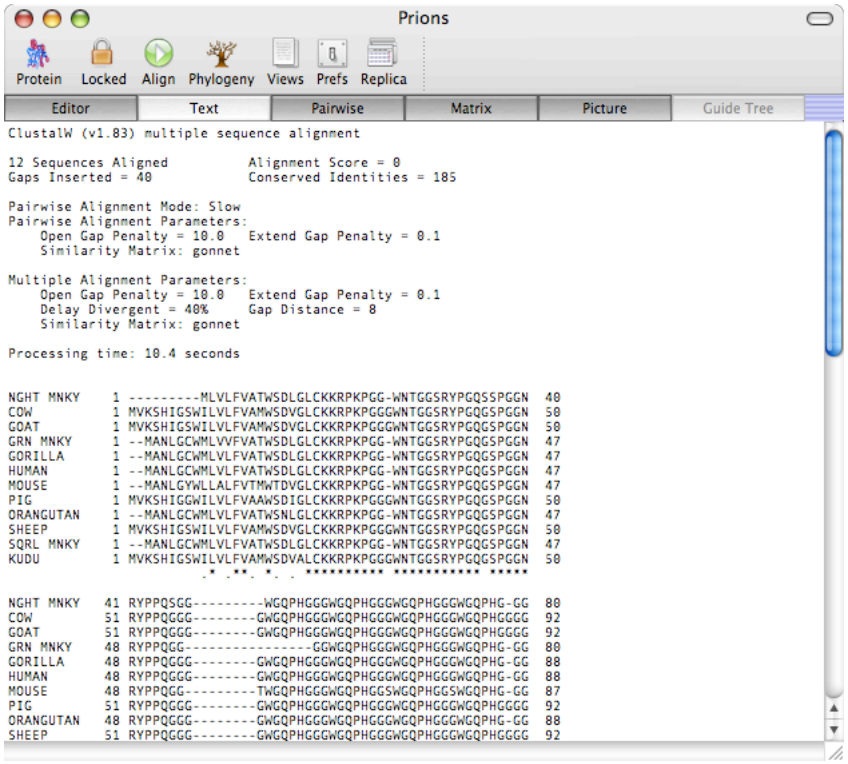
alignment, but they do update in response to edits in the Editor tab. In the same way that the different tabs in the single sequence window are just different views of the same underlying data object, the views in the multiple sequence alignment window are just different formatted views of the same underlying multiple alignment data.

Editor Tab

- **DNA/RNA** - use this button to toggle the molecule between DNA and RNA. (Protein windows show a non-functional Protein icon).
- **Locked** - the padlock helps prevent you from inadvertently modifying a sequence - you are prompted to unlock the sequence if you try to perform a destructive edit.
- **Align** – use this to automatically align the sequences using the ClustalW, Muscle or T-Coffee algorithms.
- **Phylogeny** – invokes the built-in MacVector phylogenetic reconstruction functionality.
- **Views** - this replaces the old Results button that previously defined which result windows would appear, and also allowed you to create a new sequence from the consensus. The dialog displayed is identical to before, except that now the dialog controls which tabs are active in the alignment window. Normally, you can leave all of the tabs active, but if you are analyzing large numbers of sequences or particularly long sequences you may want to turn off some of the tabs as the text in them (particularly the pair-wise alignment tab) can take some time to be generated.
- **Prefs** – opens the multiple sequence alignment document preferences dialog, giving you control over many aspects of the multiple sequence alignment display, including how the consensus is calculated and the fonts and layout of many of the tab contents.
- **Replica** - as with the single sequence window, this creates a second window so you can look at different tabs of an alignment at the same time.
- **Add Seqs** – lets you choose disk files to add more sequences to the alignment.
- **Blocking -** this adjusts the "blocking" of the sequence residues to make them easier to visualize.
- **Dots –** substitutes a dot wherever a residue in a sequence matches the consensus.
- **Line Wrap –** toggles between linear and line-wrapped modes,
- **Groups –** lets you edit the colors used in the display. You can choose different groups to color e.g. by chemical type, charge, propensity to form beta sheets or by consensus similarity etc.
- **Width –** lets you control the cell width.

Text Tab

This displays the plain text view of the alignment. The format is essentially identical to the raw `ClustalW` output, except it honors the font, line length, consensus calculation parameters and match characters in the MacVector preferences setting.



Pairwise Tab

This displays all combination of sequences aligned against each other as pairwise alignments.

Matrix Tab

This shows the % similarity and % identity of each sequence to every other sequence in the alignment, displayed as a matrix.

Picture Tab

This displays a graphical view of the alignment, tuned for printing on black and white laser printers. If you want to save the image to a disk file, you should choose **File | Export Tab Contents As...**, a variation of **File | Save** that lets you save the data displayed in the tab view, rather than the underlying multiple sequence alignment document. This works for the text based tabs as well, Alternatively, you can either choose **File | Print** and then click on the **Save as PDF** option or choose **Edit | Copy** and then use the Apple utility **Preview.app** to create a new document from the clipboard (in **Preview**,

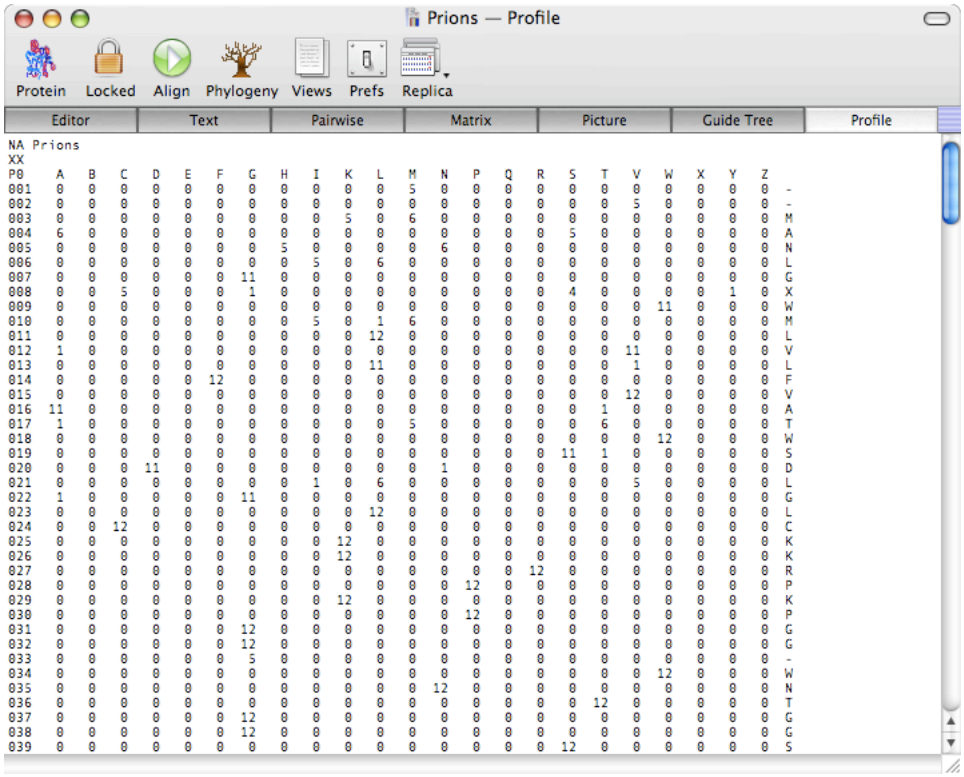
choose **File | New from Clipboard** and then save that document to a disk file).

Guide Tree Tab

This displays the guide tree used by `ClustalW` to choose the order in which it assembles the multiple alignment after the pairwise alignment phase. (If the tab is not active, you may need to open the Views dialog and select that option).

Profile Tab

The Profile tab displays the frequency of occurrence of each residue at each position in the alignment. Although primarily designed to simply display the information along with the consensus sequence, the text is in *Transfac* format, so you can save this information (using **File | Export Tab Contents As...**) and use it in other applications (including the MacVector Nucleic Acid Toolbox in the case of DNA sequences).



Interactions

The data in the various tabs are updated essentially in real time. The data in each tab is refreshed whenever the tab is activated after the underlying alignment has been edited. To see this, click on Replica and set one window to the Editor tab and the second window to the Text tab. In the Editor tab, delete some of the spaces from the beginning of the `NIGHT MNKY` sequence. Nothing should happen to the text in the second window. Now

click on that window. The window activates and the Text tab is refreshed to reflect the change to the sequence. The same will happen to most of the other tabs.

The reason for the delay is purely to avoid performance problems with large alignments. Because some of the text views can take some time to be refreshed, if they were updated every time you typed a character, updating the windows could be extremely slow, preventing you from being able to easily edit the alignment. This approach ensures the views are only updated when you are ready to look at them.

Additional Algorithms

In addition to `ClustalW`, MacVector 13 also supports the popular `Muscle` and `T-Coffee` algorithms. You can invoke the different algorithms either by choosing **Analyze | Align Multiple Sequences Using | <algorithm name>** menu item, or by clicking and holding on the Align toolbar button and selecting the algorithm from the resulting popup menu. MacVector remembers the last algorithm you ran, so if you prefer to always run `Muscle` for example, after running it once simply clicking on the Align button will always start the `Muscle` algorithm.

Sequence Assembly Overview

MacVector has an Assembly module that must be purchased separately from the main application. The module is tightly integrated into MacVector such that they appear as a single application. To determine if you have Assembler, choose **MacVector | About MacVector...** and a dialog will open;

If the dialog shows that the product is “MacVector with Assembler” then the Assembler module is active.

There is a simple `Contig Assembly Tutorial.pdf` document that introduces the basic functionality of the Assembler module. You can find this in the `/MacVector/Documentation/` folder.

Protein Analysis

Translation

Protein sequences can be imported into MacVector using the same approaches used for DNA (from existing files in a wide variety of formats, direct typing/pasting into a new protein sequence document, or by downloading from *Entrez*) or they can be created by translation of an existing MacVector DNA sequence.

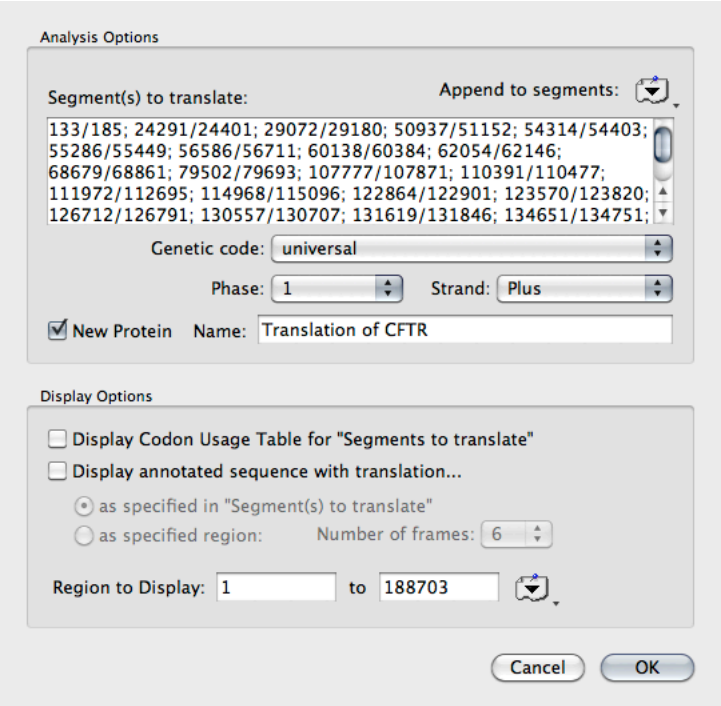
There are three primary ways of setting up a translation. Each uses the same **Analyze | Translation...** menu item.

(i) Select a region of a target sequence that you want to translate (typically by clicking on a result object from one of the Open Reading Frame analysis functions, or by clicking on a CDS feature in the Map tab) then choose **Analyze | Translation...** The translation dialog will be filled out with the selection.

(ii) Type in your own region(s) to translate in the dialog box.

(iii) Select a CDS feature from the Append to Segments popup menu that displays all the sequence features.

Try this: Open the sequence `/MacVector/Tutorial Files/Align To Reference/CFTR/CFTR.nucl`. This is the genomic region containing the human cystic fibrosis transmembrane regulator gene. Switch to the Map tab and click on one of the CFTR segments – you should see that all (26) segments become highlighted. Choose **Analyze | Translation...** and in the following dialog box should appear;



Note that the Segment(s) to translate edit box gets pre-filled with the locations of all of the exons described by the CDS feature. Make sure the New Protein checkbox is select and click OK to generate the translated protein.

Protein Analysis Functions

Protein sequences can be analyzed using the identical workflows to DNA sequences, with most of the functionality being accessible through the **Analyze** menu.

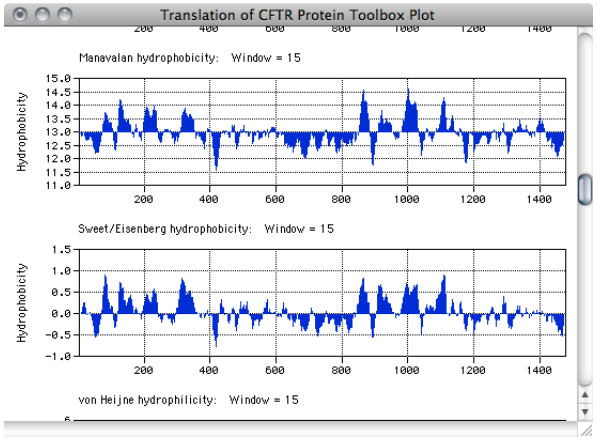
Reverse Translation – this lets you create a DNA sequence by reverse translation from a protein sequence.

Proteolytic Enzyme – the direct equivalent of the Restriction Enzyme analysis for DNA, this lets you find potential proteolytic cleavage sites in a protein.

Protein Subsequence – the equivalent of the Nucleic Acid Subsequence analysis, you can use this to find a variety of different protein motifs a number of data files are shipped with MacVector in the `/Subsequences/` folder, including `Long Prot Motifs`, `Short Prot Motifs`, `protein patterns` and `protein subsequence`.

Protein Analysis Toolbox – this is a collection of algorithms that generate profile plots displaying the likely hydrophobic, hydrophilic, antigenic and secondary structure regions of the protein. There is also a text output listing the amino acid composition of the protein, pI and a variety of other information such as molecular weight and absorbance.

Try This: If you still have the CFTR translation window open, invoke **Analyze | Protein Analysis Toolbox...** and select all of the Plot options (hold down the `<option>` key and click in an empty checkbox to toggle all the selections to "on"). When you click **OK**, a result window will appear. If you scroll down this you'll see some of the hydrophobic plots quite clearly show the 8 transmembrane segments of the CFTR protein;



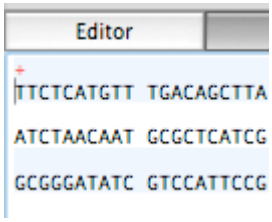
Pustell Protein Matrix – a dot-plot showing similarity between two proteins or repeat regions within a single protein.

BLAST Internet Search – directly equivalent to the DNA function for identifying matches in the online NCBI databases.

Getting The Most Out of MacVector

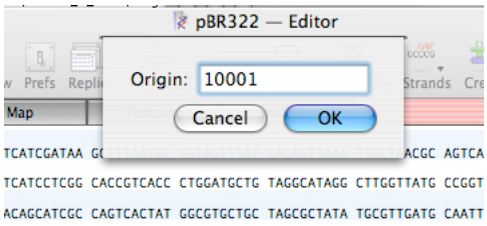
Setting The Numbering Origin

MacVector has always had the ability to set the numbering origin to a residue within the sequence by clicking on and dragging the small red cross that usually appears at the beginning of the sequence.

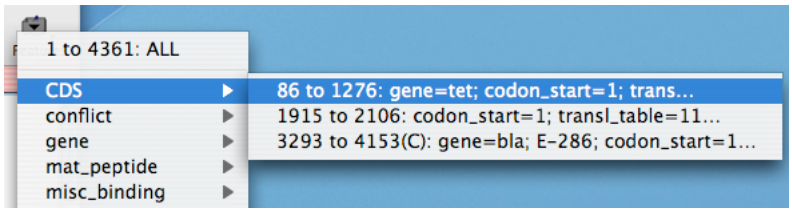


Dragging the cross to another location designates that as the “plus 1” residue – all residues before that position will be given negative numbers.

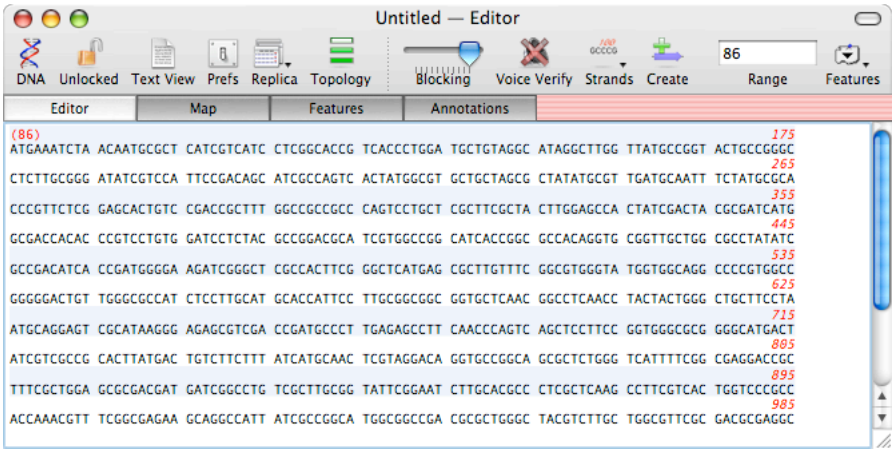
You can also set the first residue to a positive number. To set this, double-click on the red cross and enter a new start value in the sheet that appears.



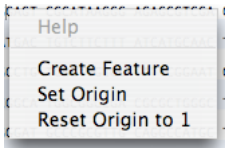
This is particularly useful if you want to work on a smaller more manageable region of a large chromosome but wish to retain the original numbering. To help with this, if you copy a section of a larger sequence and paste the copy into a new MacVector sequence window, the original numbering is retained. For example, using `pBR322` click on the Features popup menu and select the tetracycline resistance CDS;



This selects the region from 86 to 1276 in the editor. Now choose **Edit | Copy**, followed by **File | New from Clipboard**. A new window appears with the numbering origin set to 86. (You can also accomplish this by choosing **File | New | Nucleic Acid** and then **Edit | Paste** into the new window).



If you want to quickly reset the origin to "1", you can right-click (or **<ctrl>**-click) in the sequence area to bring up a context sensitive menu and choose **Reset Origin to 1**;

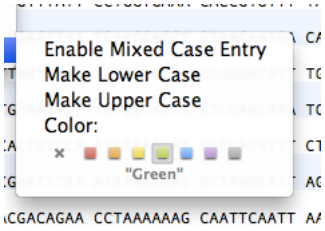


Setting the Circular Origin

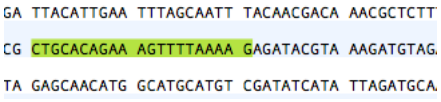
If you are working with a circular sequence, then you can change the location where the sequence is "split" in the editor. This also changes the Map view so that the new position is located at 12 o'clock. Again, locate the flashing caret at the desired location, right-click in the sequence area and now choose **Set Circular Origin** from the popup menu.

Coloring in the Editor

MacVector 12 added the ability to color interesting regions of a sequence. The simplest way to do this is the highlight a region of the sequence, then choose **Edit | Transformations | Color:** and choose a color from the menu;



The selected sequence is then colored appropriately;



Behind the scenes, this actually creates a feature that will also show up in the Map tab (see below). When this color mode is enabled, any visible feature

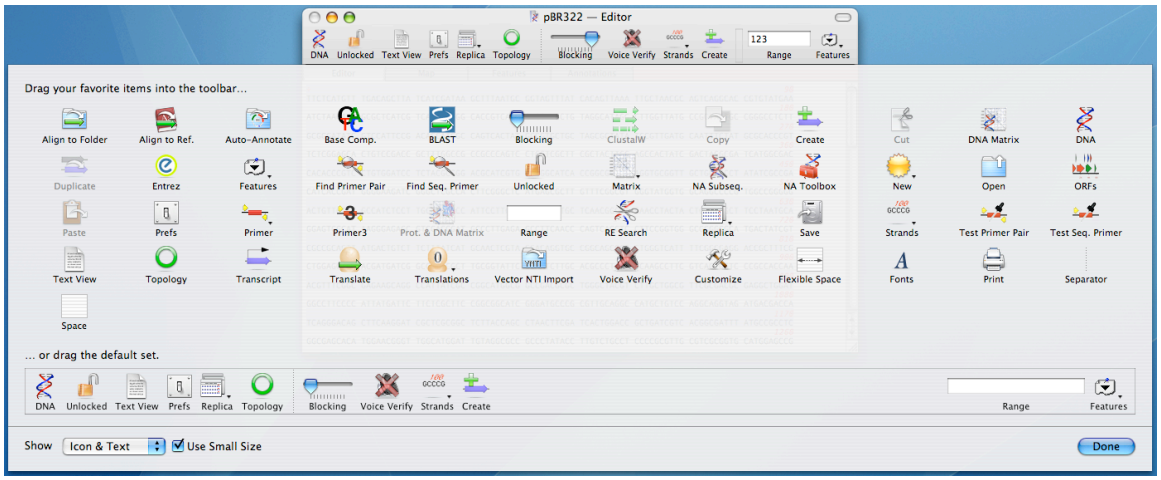
will be displayed using its "Fill" color in the Editor tab. You can control this behavior using the **MacVector | Preferences | Colors** pane which lets you turn on the coloring for all sequences and also choose between coloring the background or coloring the sequence residues.

Mixed Case Entry In The Editor

MacVector 12 added the ability to view and edit sequences using lower case as well as the traditional upper case. You can change the case of any selected residues using **Edit | Transformations | Make Lower Case**. If you want to type in new residues in mixed case, choose **Edit | Transformations | Enable Mixed Case Entry**. Changing the case of the residues does not affect any MacVector algorithms (e.g. gaattc, GaAtTC and GAATTC are all recognized as *Eco*RI sites), but the case is displayed in all text output windows so that you can quickly identify your region(s) of interest in those results.

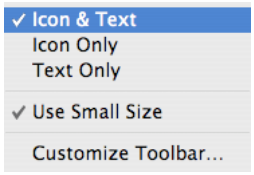
Toolbar Customization

In any MacVector window you can either right-click or **<ctrl>**-click in the toolbar area to bring up the standard OS X toolbar customization dialog. You can use this feature to remove buttons you don't use or to add other buttons such as the Print button or any of the analysis functions. Each tab has its own toolbar layout, so you can customize each tab independently.

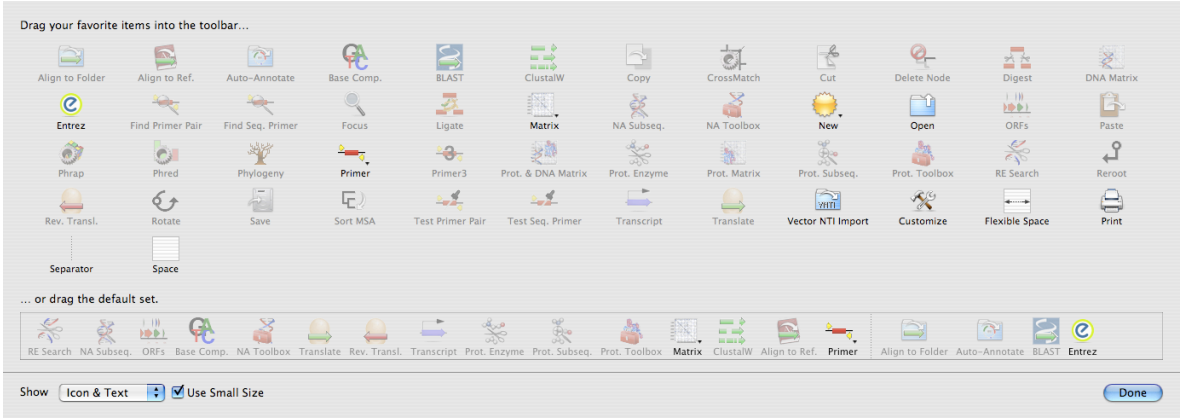


Customizing the Analysis Toolbar

You can control which buttons appear in the Analysis Tools toolbar, as well as the size of the toolbar and even the order the buttons appear using the standard OS X toolbar customization interface. First, **make sure that you have no other MacVector windows open**. If you have a window open, that will be used as the target of the customization as the OS considers that to be the active window. Now right-click on the toolbar (or **<ctrl>-**click if you do not have a right-clickable mouse) and a popup menu will appear.



This lets you control how the buttons appear in the toolbar. If you click on the Customize Toolbar option a dialog sheet appears;

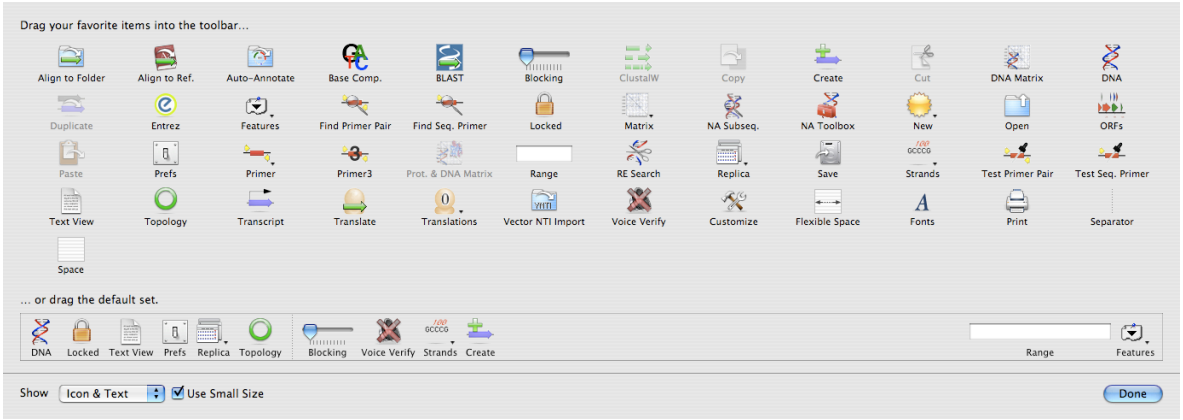


You can see that there are many more buttons available than we use in the default set. Some of the default buttons (e.g. the Primer and Matrix buttons) display a popup menu letting you choose which analysis function you want to use. However, you can add individual buttons for each function if you wish. If you use Primer3 a lot, you might want to add that to the toolbar. There are also buttons for many of the functions in the File and Edit menus (New, Open, Save, Print, Cut, Copy, Paste, Digest, Ligate) and also for the **Phylogenetic Analysis** sub menu (Focus, Reroot etc). Assembler users also have the Phred, Phrap, CrossMatch and Bowtie tools as options.

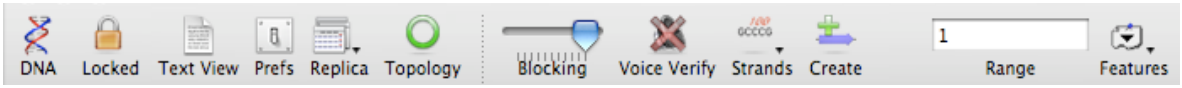
Simply drag the appropriate icons onto (or off) the toolbar until you have it as you want. You can also add separators and space icons to help you organize the toolbar. When finished, simply click Done and the toolbar will take on your new customizations.

Customizing Sequence Window Toolbars

The analysis buttons can also be added to the sequence window toolbars if you want to have ready access to one-click analyses. Open a nucleic acid sequence and right click on the toolbar;



This time, only the analysis icons relevant for nucleic acid sequences are shown. There are also additional icons that are applicable only to the actual tab you have selected (e.g. the Range and Voice Verify icons for the sequence Editor tab). You could for example change the Editor tab toolbar from the the default;

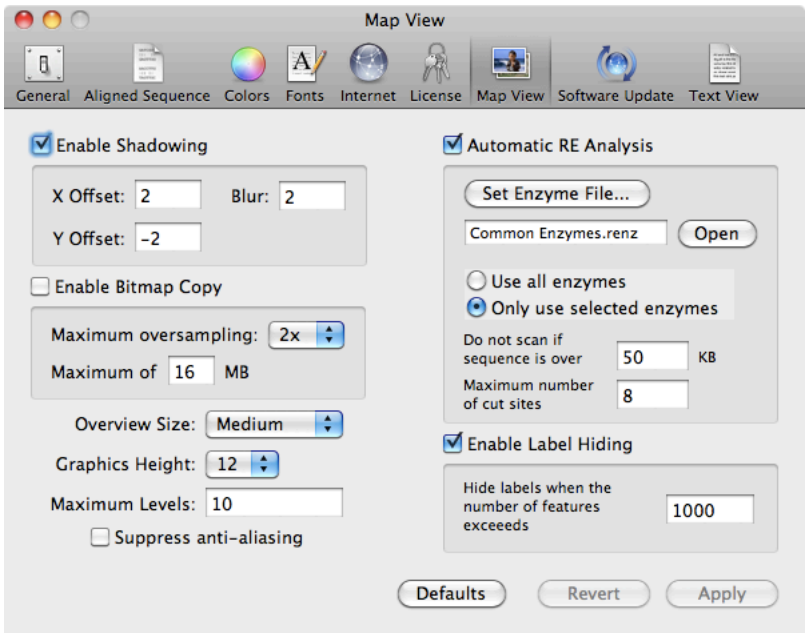


To something like this if you like to do a lot of editing and analysis;

Note that each tab maintains its own unique toolbar. Some functions are available only in certain tabs and you will typically be doing different operations in different tabs so you can customize the toolbars to most closely match your workflows.

Configuring Automatic Restriction Enzyme Searching

One particularly useful feature in MacVector is that sequences can be automatically scanned for restriction enzymes and the cut sites displayed in the Map tab. This feature is controlled by a preferences dialog that can be displayed by either (a) selecting **Options | Map View Options...** or (b) clicking on the Prefs button in the Map tab toolbar or (c) choosing **MacVector | Preferences** and selecting the Map View pane.



The right hand Automatic RE Analysis group controls the setting for the analysis.

- **Automatic RE Analysis** - deselect this to turn off the automatic restriction enzyme display in the map view. By default it is always set to be on.
- **Set Enzyme File...** - click on this button to select the enzyme file you want to use. By default, this is set to `/MacVector/Restriction Enzymes/Common Enzymes`. The name of the selected file is displayed in the box immediately below the button.
- **Open** - this button will open the selected restriction enzyme file in a restriction enzyme editor window. You can select/deselect enzymes in this window and the Map tab graphical view will refresh to display the locations of the selected enzymes in the file (see later).
- **Use all enzymes/Only use selected enzymes** - these radio buttons determine how MacVector should treat the selected file. Their use is pretty obvious - if you select Use all enzymes every enzyme in the selected file will be used in the search, otherwise only the selected enzymes will be used.
- **Do not scan if sequence is over** - the restriction enzyme scan is relatively fast - a full scan of the *E. coli* genome takes just a few seconds on an average Macintosh. However, the graphical display can take much longer if trying to layout 1 million restriction enzyme sites on a circular chromosome, so this setting provides an option to turn off the scan and display for larger sequences. The default is set to 50kb.
- **Maximum number of cut sites** - this lets you screen out restriction enzymes that cut too frequently in your sequence. This not only reduces the calculation/display speed, but it also reduces on-screen

clutter. The default is six sites, but you might want to reduce this if you routinely use relatively large sequences.

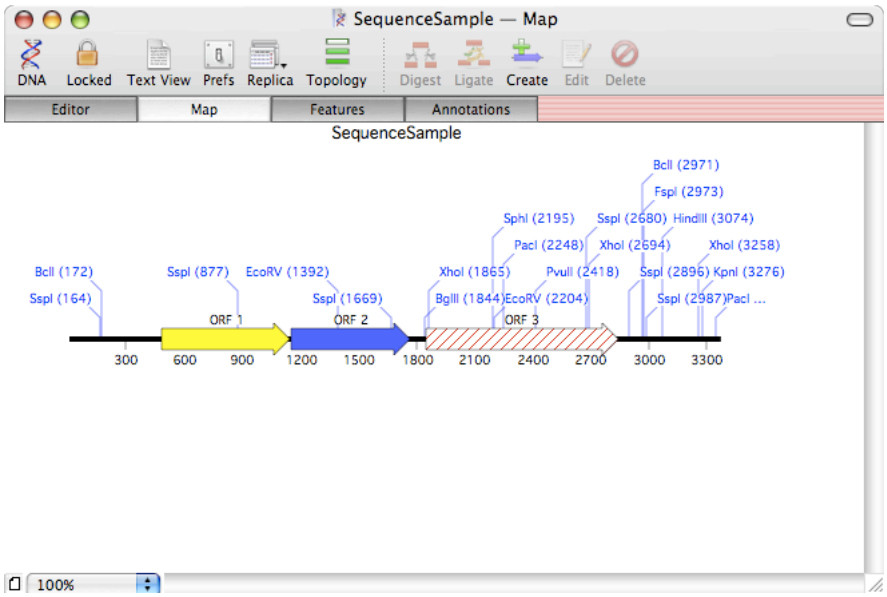
One other relatively new feature is the ability to suppress the display of labels for very large crowded maps. This is particularly useful when viewing whole genomes. You will find that this greatly improves display performance for maps with over a thousand features and avoids the appearance of many overlapping labels. Click the Enable Label Hiding checkbox to turn on this feature and set the maximum number of labels to display in the appropriate edit box.

Working with the Default Restriction Enzyme file

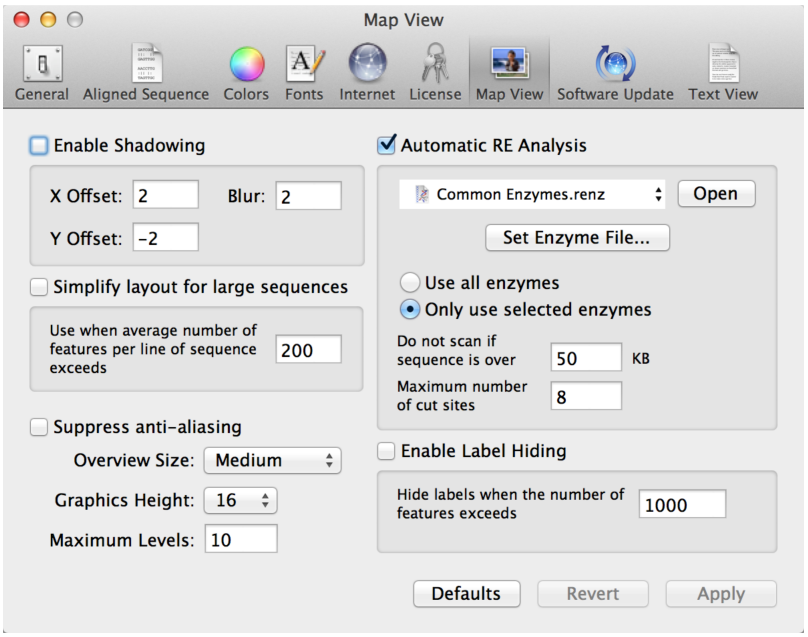
All open sequence windows with the Map tab selected will respond dynamically to selection changes in the selected restriction enzyme file. Try the following tutorial;

Open `/Applications/MacVector/Tutorial Files/Align To Reference/Sequence Confirmation/SequenceSample`.

Make sure you have the Map tab selected. If you haven't changed anything from the defaults after installation, it should look something like this;

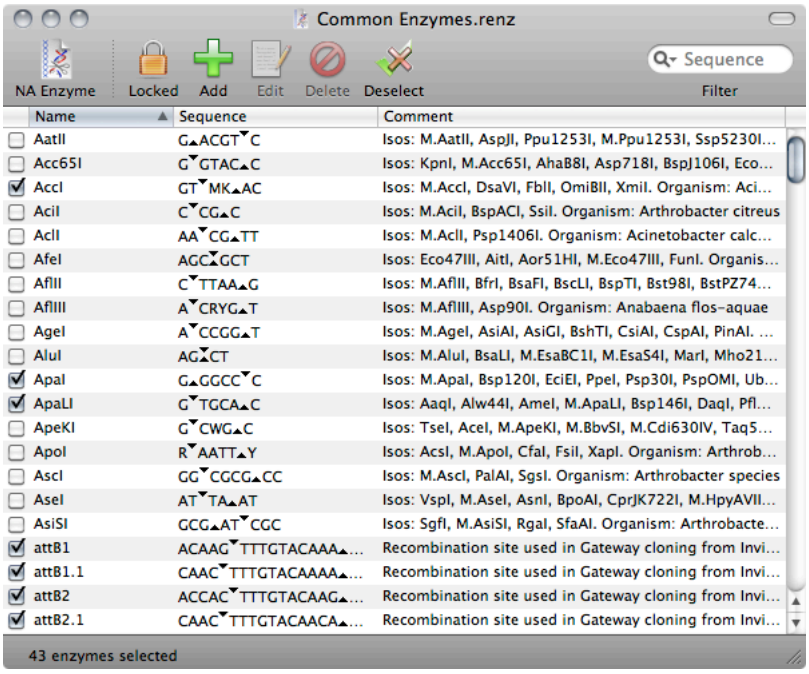


The restriction enzymes displayed in the Map view are controlled by the Map View Options dialog. Click on the Prefs toolbar button;



If you do not see `Common Enzymes` as the selected enzymes file, click on Set Enzyme File, navigate to the `/Applications/MacVector/Restriction Enzymes/` folder and choose the `Common Enzymes` file.

Click on the Open button. You should see the Common Enzymes restriction enzyme window open. Click on the close button to dismiss the Preferences pane and then bring the Common Enzymes window to the front;



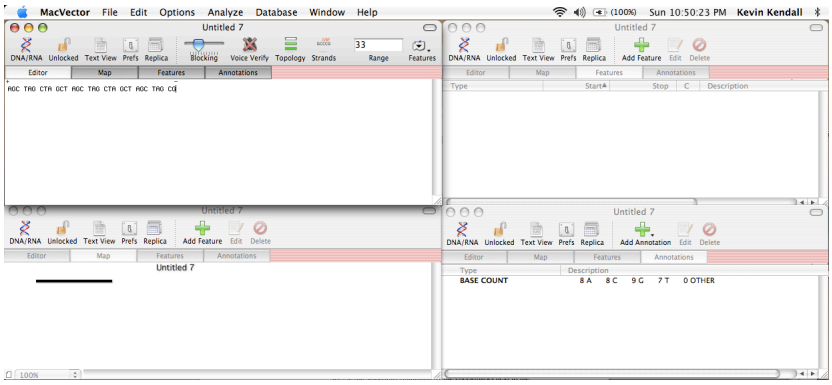
Position the windows so you can see both the Common Enzymes and SequenceSample windows. Click on the checkboxes next to the enzyme names to toggle the selections. As you click on the checkboxes, the

selection toggles and the Sequence\Sample Map view refreshes in the background to show the new selected restriction enzymes.

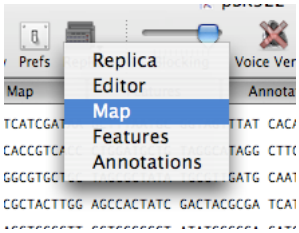
Using Replica Windows

You can click on the Replica icon on the toolbar of any sequence window to display a new window that shows an independent view of the same sequence. The windows are interactive, so that changes in one window are reflected in the other(s). To get a good feel for this interactivity, try the exercise below;

Choose **File | New | Nucleic Acid** to create a new DNA window. Switch to the Editor tab and type in a few residues. Now click on the Replica button to create a second window. In that window, click on the Replica button again. Repeat until you have four windows. Now select a different tab in each window and arrange them so that you can see each window clearly. Try to make them look something like this;



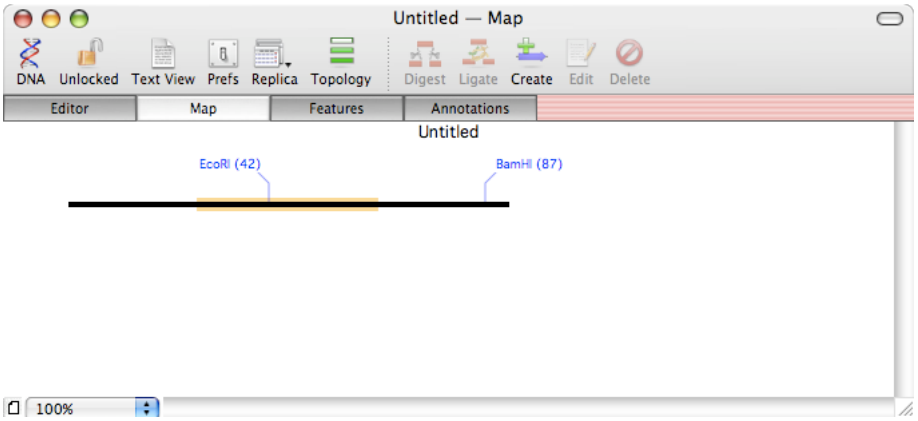
Note that you can choose which tab you want a replica window to open with by selecting the name of the tab in the drop down menu that appears when you click on the Replica button;



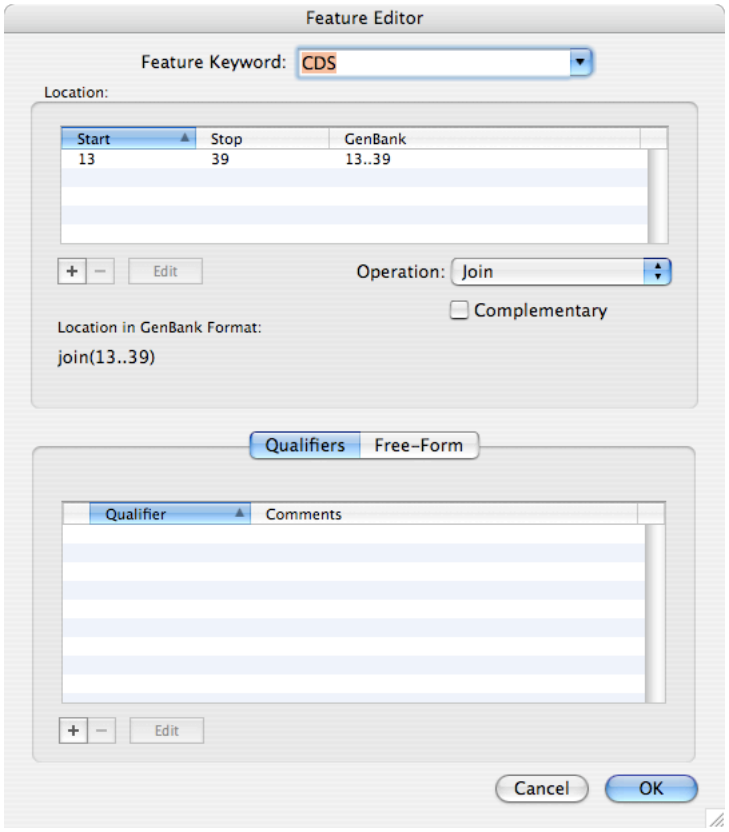
With the Editor tab window front most, continue typing in the window. Note how the Annotations tab window updates the Base Count in real time. Type the sequence "GAATTC" in the Editor tab window - note how the Map tab refreshes to show an *Eco RI* restriction enzyme site as you type the final "C" in the sequence.

Type a few more random residues, then type the sequence "GGATCC" - a *Bam HI* site should be shown in the Map tab window.

Select a few (10-20 or so) residues in the Editor tab window using the mouse. Switch to the Map tab window - you will see that the selected region is now shown highlighted;

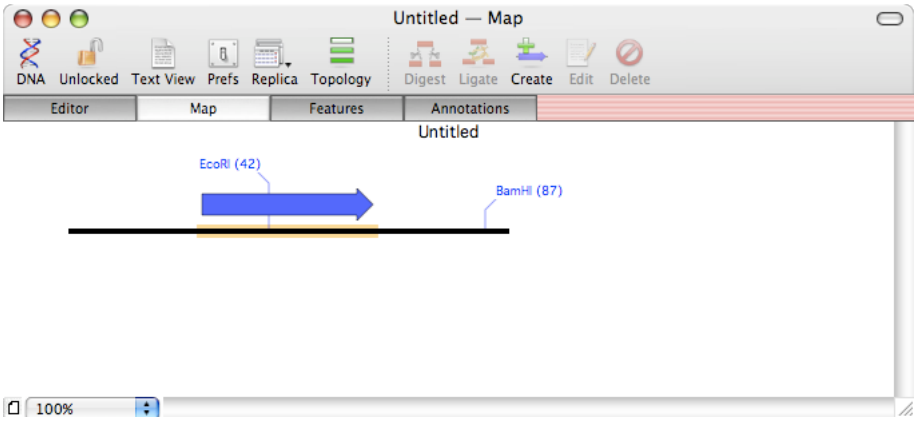


Lets create a new feature from this selection. Click on the Create toolbar button. The Feature Editor opens;

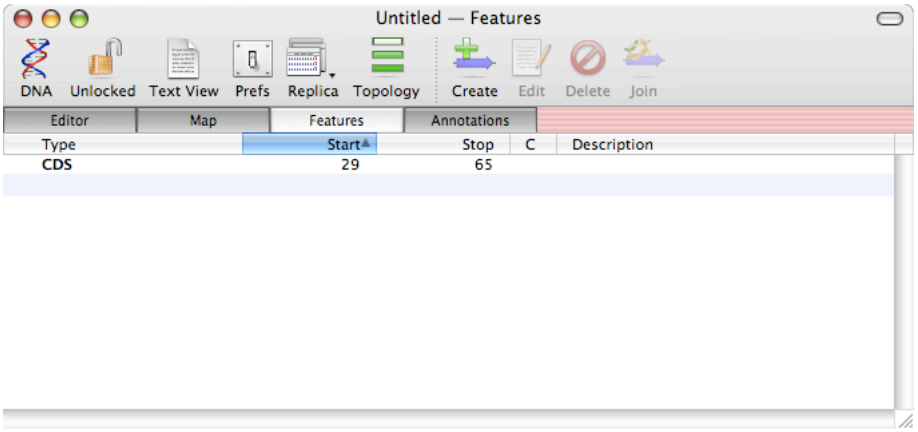


Note how the Start and Stop location of the selection are already filled out for you in the Location list. We'll look at the use of the Feature Editor in more detail later, but for now, choose `CDS` from the Feature Keyword drop down menu, then click the OK button.

A new blue arrow appears in the Map tab window, showing the location of the new feature you created;



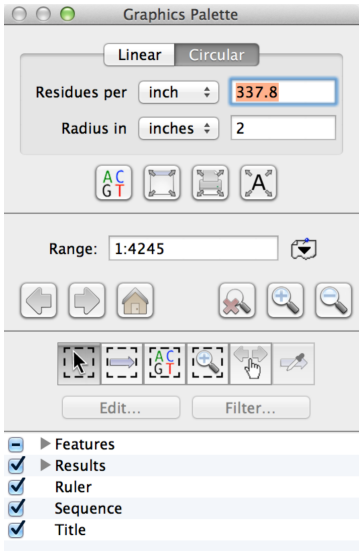
The Feature tab window also shows the new feature;



Now go back to the Editor tab window. Click near the beginning of the sequence to position the flashing insertion cursor before the start of the feature you created. Type a few random DNA residues. Note how the Map tab window and the Features tab window dynamically update and the new feature moves to the right as you type. MacVector always maintains the correct positions of features whenever you cut, paste or type into a sequence.

The Floating Graphics Palette

Whenever a Map view is active in the front most MacVector window, a floating Graphics Palette window appears. (If this is not displayed, choose **Window | Show Graphics Palette** to make it visible).



You can use this to turn on and off sequence, features and results objects in the Map display using the lower tree view. However, there is a lot of additional functionality controlled by the palette, which had a significant makeover for MacVector 12.0.

Linear/Circular – the tabs at the top let you toggle the Map between linear and circular views of the sequence. The circular option is disabled if the underlying sequence does not have circular topology.

Residues per inch/Line wrap – this section controls the scale of the Map and how "wide" a line is.

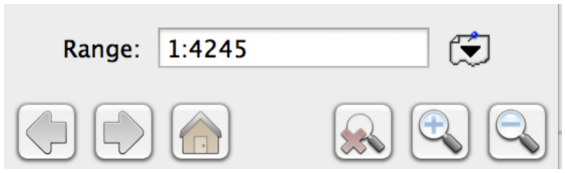
A row of 4 buttons provides "quick layout" functionality:

Zoom to Sequence – this adjusts the Residues per Inch so that individual sequence residues are visible. It does this without affecting the current zoomed section.

Fit to Window – adjusts the Residues per Inch and Line Wrap so that the current zoomed section fits in the current window.

Fit to Page – adjusts the Residues per Inch and Line Wrap so that the current zoomed section fits in the current printed page.

Fit Residues – similar to Zoom to Sequence, but this resets the zooming so that the entire sequence is displayed.



The range section lets you zoom to display a specific section by typing in the box or selecting a feature from the popup features menu.

Left Arrow/Right Arrow – these let you "nudge" the zoomed section to either side. You can also use the left/right keyboard arrow keys. For unzoomed circular sequences, these will rotate the graphic on screen.

Home – if you ever get "lost" in a sequence, this will center the sequence on the screen.

Reset zoom – restores the display to show the full sequence. Equivalent to a double-click in a blank area of the Map.

Zoom In/Zoom Out – this pair of buttons zooms in and out of the sequence in two-fold increments. You can also use the up/down keyboard arrow keys.

There is a row of six "mode" buttons that controls what happens when you click and drag with the mouse in the Map window.

Select Zoom – this is the default and the mode used by older versions of MacVector – you can click on features or sites to select them and if you click, hold and drag, the display resets to "zoom in" to the segment you selected.

Select Features – clicking and dragging selects all the features or sites that are touched by the selection rectangle.

Select Sequence – clicking and dragging selects just the sequence touched by the selection rectangle.

Magnify – in this mode a click magnifies the display 2-fold. Hold down the `<option>` key to reduce the magnification 2-fold.

Slide – this mode lets you drag the current zoomed region to the left or right. You can also use this to rotate circular sequences so that any arbitrary location is set to the 12 o'clock position.

Copy Feature Appearance - this mode is only available if you have one or more features selected. Once selected, if you then click on a different feature, all the selected features will change appearance to match that feature

Click Cloning

MacVector makes it particularly easy to construct new DNA molecules from existing clones and vectors by selecting restriction enzyme sites, copying the intervening fragment, then pasting into a target molecule. Lets run through a couple of examples. For this we will use the files `pUC19` and `S. coelicolor cosmid SC5A7`. Choose **File | Open** and navigate to the `/Applications/MacVector/Tutorial Files/Click Cloning/` folder. Select them both (hold down the **<shift>** key) and click on the Open button.

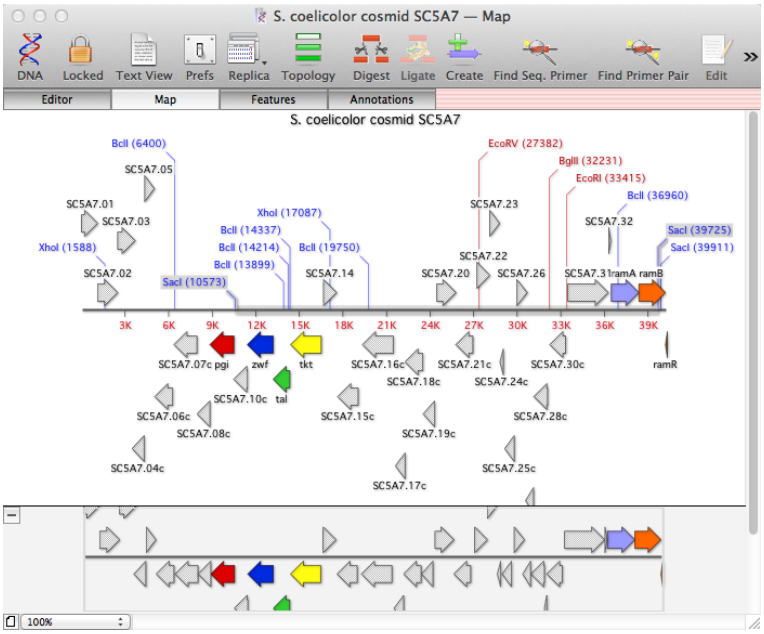
A simple cloning experiment

In the `S. coelicolor cosmid SC5A7` window, click on the Map tab. You will see a graphical map of the cosmid with the locations of the cuts sites displayed. There should also be a floating graphics palette window with the title `S. coelicolor cosmid` available. If you don't see this, look in the **Windows** menu and select **Show Graphics Palette**.

Click on the Fit to Window button in the floating palette window. The display should resize so that you can see the entire cosmid on a single line. At this stage you may notice that the two *Sac*I sites at the extreme right end of the cosmid have truncated labels – to display the entire label, click in the Line Wrap edit box of the floating palette and increase the wrap to 9.

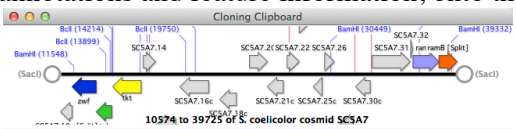
You should be able to see a *Sac*I site at 10573. Click on the site label and it should select.

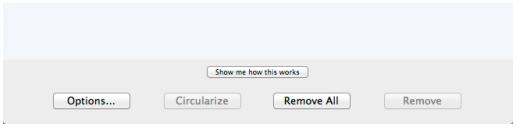
Hold down the **<shift>** key and select the *Sac*I site at 39725 (near the end of the cosmid). Now both *Sac*I sites should be selected.



Choose **Digest** from the **Edit** menu (or click on the Digest toolbar icon). This copies the fragment, along with sticky end information and all overlapping

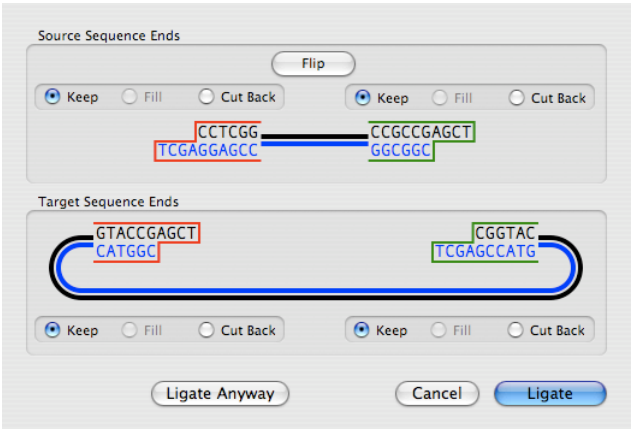
annotations and feature information, onto the Cloning Clipboard.





Switch to the pUC19 sequence window, make sure the Map tab is selected and click on the single *Sac*I site at 406.

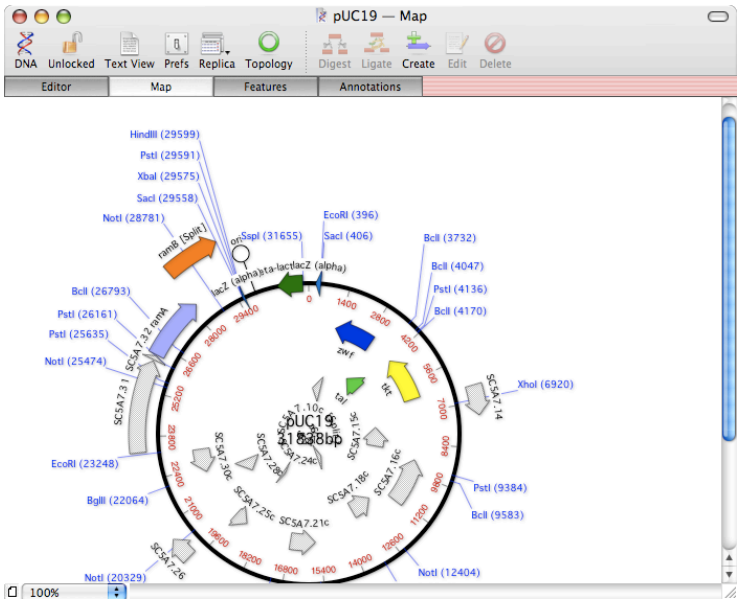
Select **Ligate** from the **Edit** menu (or click on the Ligate toolbar button or drag the fragment from the Cloning Clipboard). You should be prompted to unlock the pUC19 sequence. This is to prevent you from accidentally modifying an important sequence file. A ligation dialog will appear displaying the structure of the sticky ends.



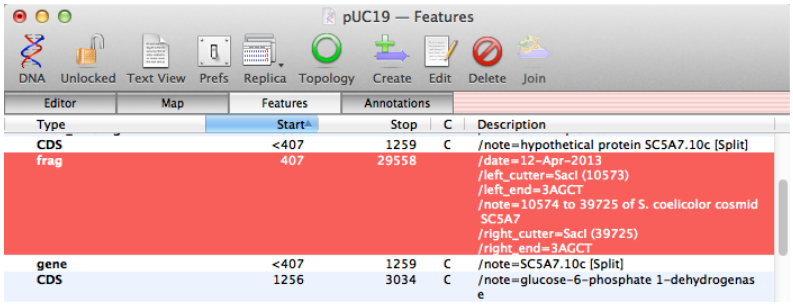
The dialog shows red and green outlines around the source and target sticky ends. This indicates that both sets of ends are compatible and can be ligated together.

Click on the Ligate button.

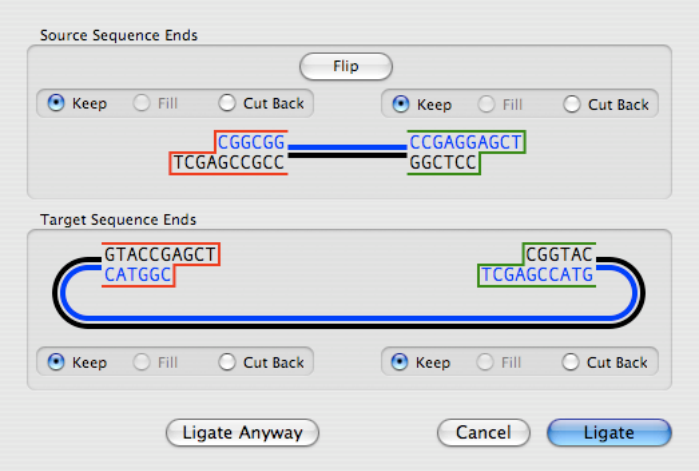
The fragment from the clipboard becomes inserted into the *Sac*I site of pUC19. Note that the pasted fragment retains all of the features and feature appearance of the source DNA molecule (compare the two screenshots).



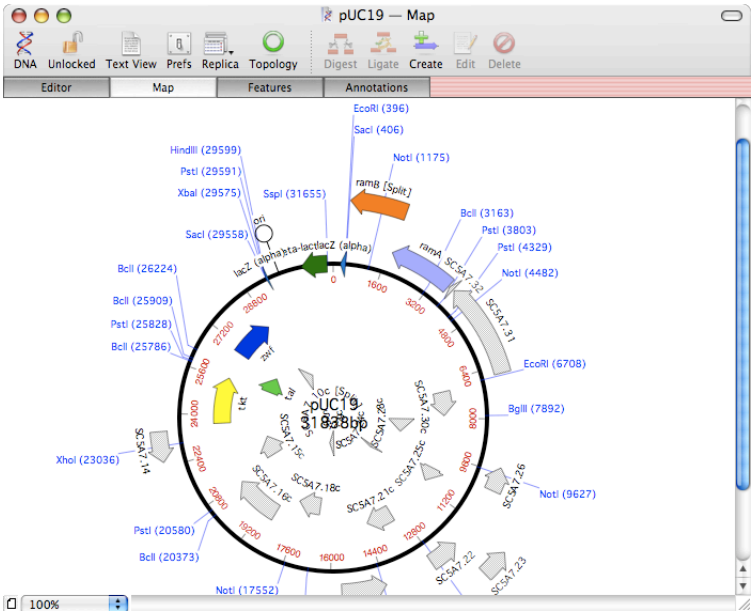
Note the direction that the fragment was inserted. Switch to the Features tab and note the new Frag feature that has been created. This shows the date of ligation, the source of the fragment, enzymes used and the sequence of any overhanging ends.



Now choose **Edit | Undo Ligate** – pUC19 is restored back to its original appearance. Once again, click on the *Sac*I site and choose **Edit | Ligate**. This time, click on the Flip button in the ligation dialog. You should now see that the strands in the source fragment have changed colors – the upper strand is now blue and the lower strand is black. This provides a visual indication that the fragment is "upside down" i.e. it has been reversed and complemented.

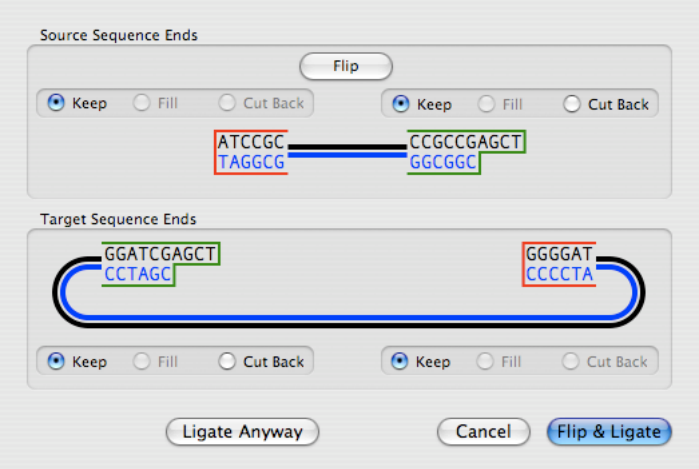


Now when you click on the Ligate button, the source fragment is inserted in the opposite orientation;



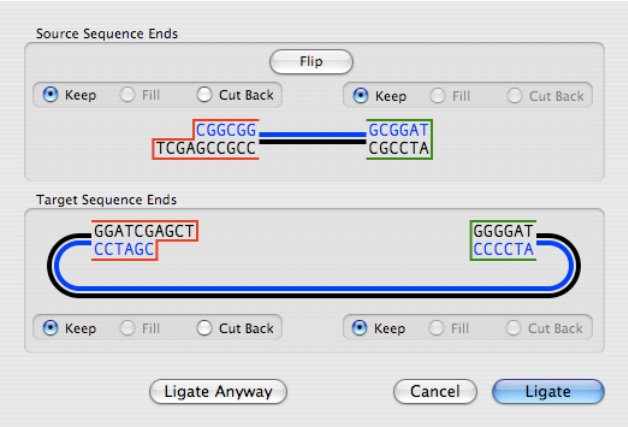
Two Enzyme Directed Cloning

The ligation dialog is sufficiently aware of the compatibility of sticky ends to know when a fragment needs to be flipped to be able to be ligated. To illustrate this, first undo the last ligation into pUC19 (or close without saving and re-open pUC19 from disk). Next, switch to the S. coelicolor Cosmid window and select the *Eco*RV (27382) – *Sac*I (39725) fragment (remember to hold the <shift> key down) and choose **Edit | Digest** (or click on the Digest button). Switch to pUC19, select the *Sac*I (406) and *Sma*I (414) sites and choose **Edit | Digest** (or click on the Ligate button).



The ligation dialog outlines compatible ends in colors to indicate what will happen when you click on the Ligate button. The left end of the source sequence is always outlined in red, with the right end outlined in green. In this case the source *Eco*RV and destination *Sma*I are both blunt cutters and outlined in red, with the compatible *Sac*I sites outlined in green. Because the colors are diagonally opposed, this indicates that the fragment needs to be flipped before it can be ligated. The blue Ligate button text has changed to Flip & Ligate to indicate this.

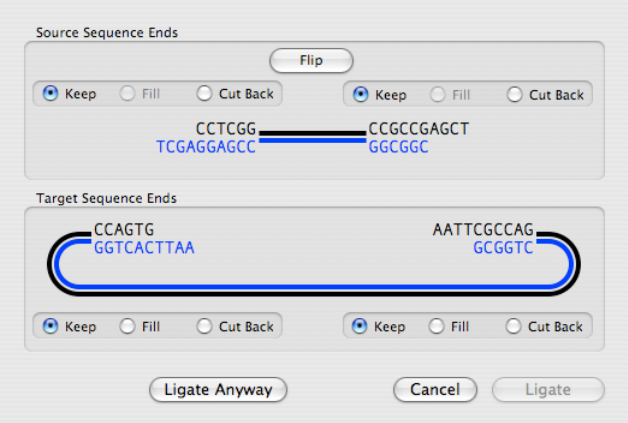
If you click on the Flip button, the dialog changes so that the source sequence is reversed and complemented – the blue and black strands are swapped and the ends switch so that the sticky SacI site is now at the left end of the source. In addition, the text of the Ligate button has changed from Flip & Ligate to simply Ligate to indicate that no additional flipping is required.



Manipulating Sticky Ends

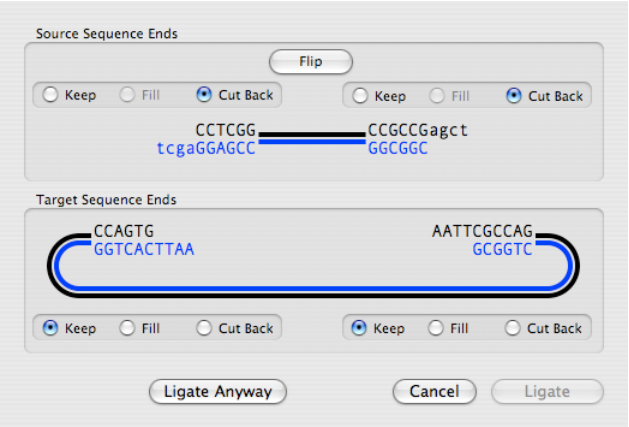
It is often necessary to fill in or cut back the ends produced by restriction enzymes so that fragments can be blunt end ligated into a vector. The ligation dialog lets you fill in 5' overhangs and/or cut back 5' or 3' overhangs to blunt the ends of either the source or target sequence ends.

We will illustrate this using the *Sac*I fragment from earlier, but this time we will try to insert it into the *Eco*RI site of pUC19. Undo your last ligation into pUC19 (or close without saving and re-open pUC19) then switch to the S. coelicolor Cosmid window. Select the large *Sac*I fragment (10573-39725), Digest it, switch to pUC19, select the *Eco*RI (396) site and choose Ligate.



This time there are no outlines around the ends and the Ligate button is disabled because the ends are not compatible. Note that the source fragment has 3' overhangs that cannot be filled in, they can only be cut back, and so only the Cut Back radio buttons are enabled. Conversely, the target fragment has 5' overhangs, so both the Fill and Cut Back radio buttons are active.

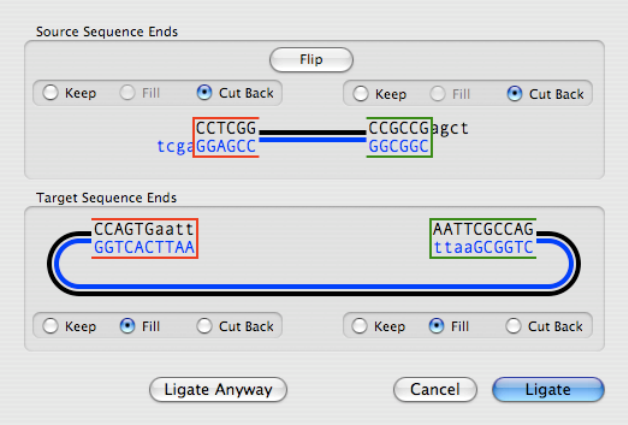
Click on the Cut Back radio button for the source fragment right end.



Note how the AGCT 3' overhang is now displayed in lower case – this indicates that you want to remove those residues.

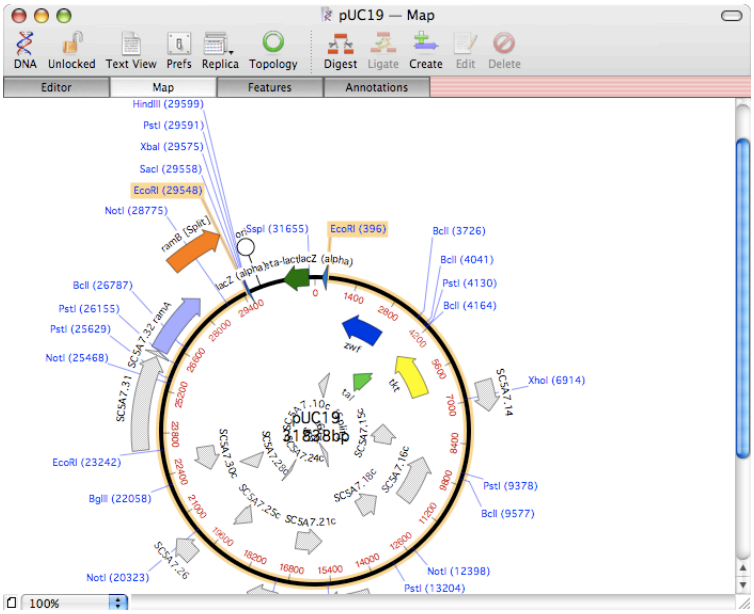
Select the Cut Back radio button for the left end as well.

Now click on the Fill radio button for the left target end.



The end gets filled with lower case residues and the end becomes outlined to indicate it is now compatible with the source left end. You can also see that by filling in the end the target sequence will have the sequence …GAATT at the end. The cut back source will start with CCT... so the ligated molecule should have the sequence GAATTC at the junction, which will regenerate an *Eco*RI site. When you click on the Fill button at the target right end, that will be compatible with the second source end and the Ligate button becomes enabled.

Click on Ligate – the resulting molecule does indeed have *Eco*RI sites regenerated at the junctions of the two molecules (locations 396 and 29548)(shown highlighted in the screenshot for clarity).



Additional Information

Although the examples shown here used the Digest and Ligate buttons, you can accomplish the same things using the **Edit | Copy** and **Edit | Paste** functions. The differences are that (a) **Copy** places a copy of the graphical

image on the clipboard in PDF format as well as copying the DNA sequence and (b) **Paste** only displays the **Ligation** dialog if the source and target ends are incompatible.

If you click on the Ligate Anyway button, MacVector will ignore the end information and treat all of the cut sites as if they were blunt ended.

Whenever you paste a fragment into a target molecule in this way, a special "frag" feature is created in the target with a note describing the source of the segment of DNA, the date of ligation, enzymes used and the sequence of any overhanging ends. You can use this to keep track of the history of your constructs.

Understanding the Feature Editor

The Feature Editor has had a number of changes over the years with the aim to bring MacVector more in line with the GenBank standard for annotating sequence features.

GenBank Feature Format Primer

GenBank is the primary US repository of DNA and Protein sequences, curated by the NCBI (National Center for Biotechnology Information) at the National Institutes of Health (NIH) in Bethesda Maryland. Sequences maintained by GenBank have two main types of annotations. MacVector arbitrarily splits these up into "Features", which we define as annotations that have a defined location on the sequence (such as a gene or a site), and "Annotations" which are general data associated with a sequence such as an accession number, publications or authors.

GenBank features have a defined type such as `CDS`, `mRNA`, `promoter` etc - only a limited number of types are allowed. You can find a full description of the GenBank file format at;

`http://www.ncbi.nlm.nih.gov/collab/FT/`

Each feature can have one or more qualifiers associated with it. The permissible qualifiers depend on the type of feature. For example, here is the definition for the gene feature type;

```
Feature Key           gene


Definition            region of biological interest identified as a gene
                      and for which a name has been assigned;

Optional qualifiers   /allele="text"
                      /citation=[number]
                      /db_xref="<database>:<identifier>"
                      /experiment="text"
                      /function="text"
                      /gene="text"
                      /inference="TYPE[ (same species)][:EVIDENCE_BASIS]"
                      /label=feature_label
                       /locus_tag="text" (single token)
                      /map="text"
                      /note="text"
                      /old_locus_tag="text" (single token)
                      /operon="text"
```

```
/product="text"
/pseudo
/phenotype="text"
/standard_name="text"
/trans_splicing

Comment         the gene feature describes the interval of DNA that
                corresponds to a genetic trait or phenotype; the feature is,
                by definition, not strictly bound to it's positions at the
                ends;  it is meant to represent a region where the gene is
                located.
```

Note how the qualifiers always start with a forward slash. Most are followed by an equals sign and text in quotes. Some qualifiers have no additional information (e.g. the `/pseudo` qualifier in this example).

You can submit sequences to GenBank using the **Sequin** utility, available for download from;

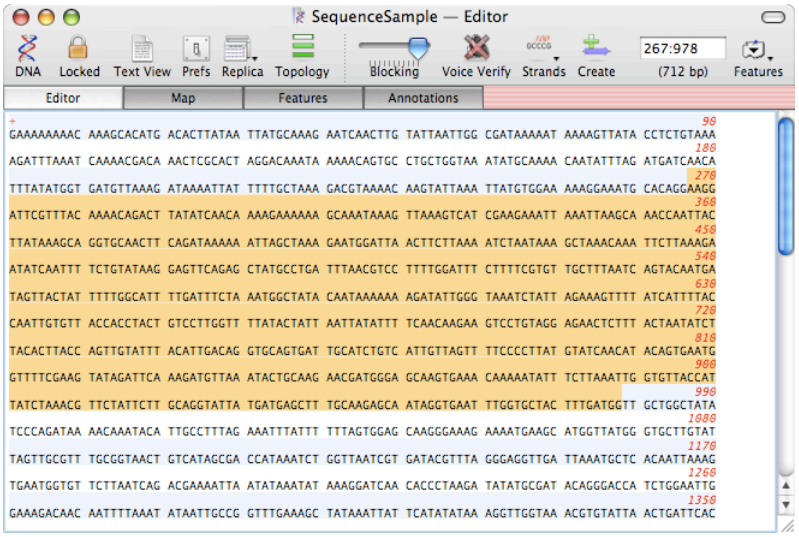
`http://www.ncbi.nlm.nih.gov/Sequin/`

MacVector has some file saving features designed to simplify the use of **Sequin** that are described in more detail later.

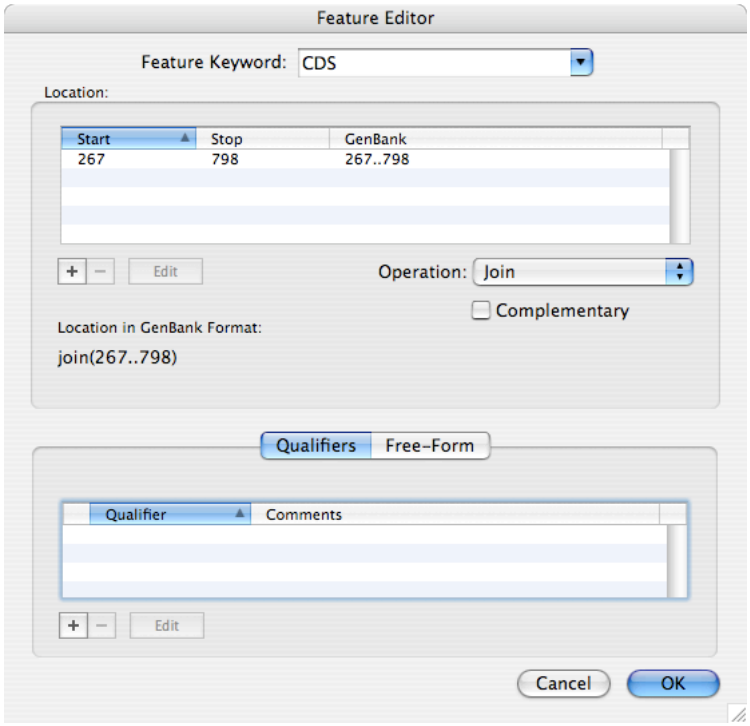
Adding a New Feature

You can easily create new features in MacVector. To illustrate this, you could open or create any sequence file, but we will use `/MacVector/Tutorial Files/Align To Reference/Sequence Confirmation/SequenceSample` if you want to follow along.

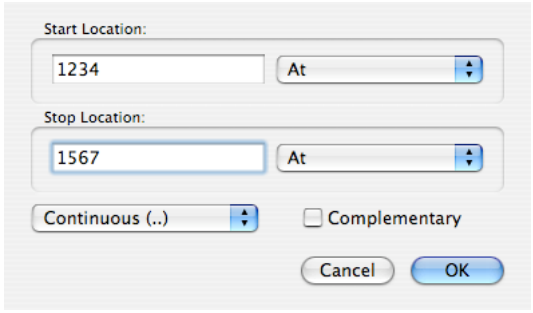
There are a number of ways of automatically selecting a region of a sequence from analysis results, but for this example we will just select a random range of the sequence in the Editor tab;



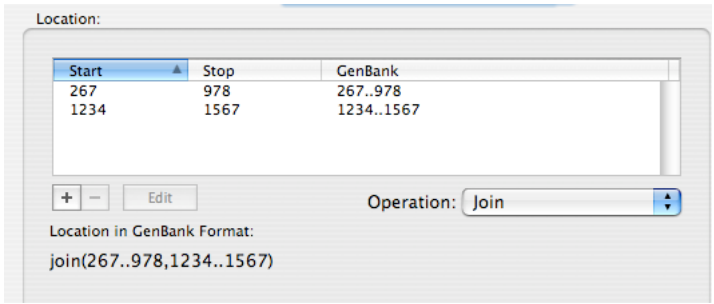
Simply click on the Create toolbar button to create a features based on the current selection. The Feature Editor will open (you will likely get prompted to unlock the sequence before you can add a feature);



By default, the region you have selected appears as the first entry in the Location list. You can enter additional regions by clicking on the + button. This brings up a sheet in the dialog that lets you define the additional segment for the feature;



Currently, you must enter the coordinates of each additional segment by hand. Most features just have one segment, but certain features (such as `CDS` features in intron containing eukaryotic genes) make heavy use of segments. If you do add additional segments, once you click OK the Feature Editor will reflect the combination of segments;



Note that the Location in GenBank Format field displays text showing how GenBank would represent the feature co-ordinates - you can use this to double-check that the feature has been created correctly and also to understand a little more about how GenBank represents the location of features on a sequence.

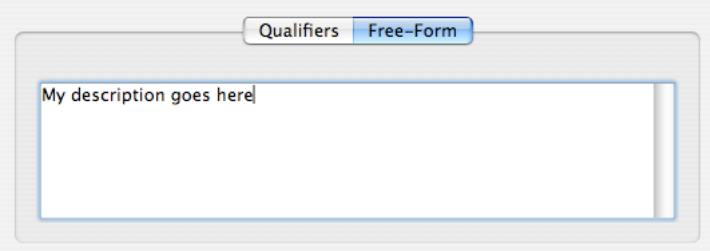
In MacVector 13 you can also create multi-segmented features by creating a series of individual features, then selecting them all in the Features tab and clicking on the Join button

You can select one of the lines the Location list and click Edit to change the co-ordinates of that segment.

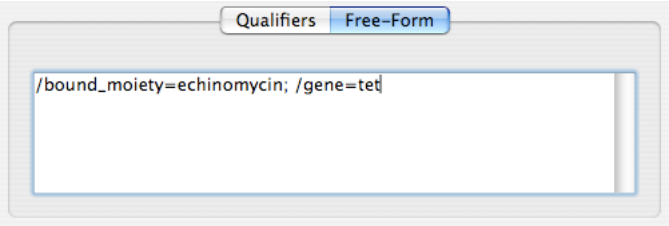
You can click on the Feature Keyword popup menu to choose a different feature type for this feature, or you can type the name of a feature type in the box (partial names will jump to that section of the popup menu). Note that MacVector lets you type invalid keywords into the box, but will not enable the OK button until you have selected a valid keyword.

Feature Qualifiers

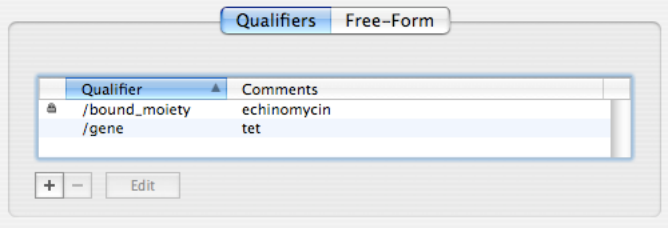
There are two ways of adding name/text information to a Feature. If you have used earlier versions of MacVector, you may be used to just adding a simple text description to a feature. You can do this in MacVector 11 using the Free-Form tab. Any text you enter here will normally get added to the Feature as a `/note=` qualifier.



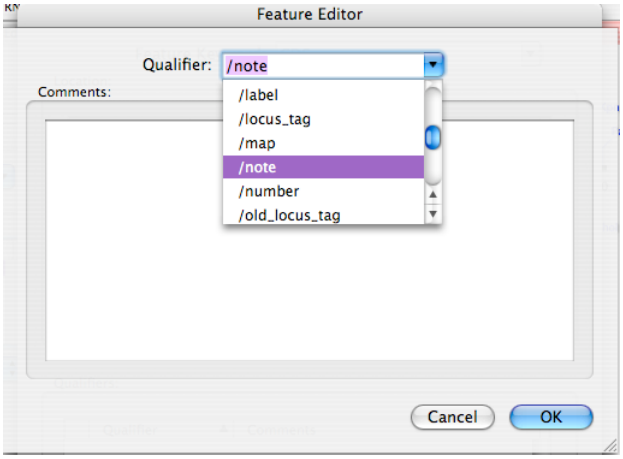
However, if you are comfortable with the GenBank format, you can type in qualifiers as free form text;



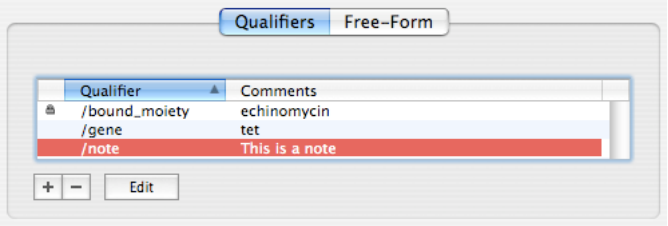
If you want to follow the GenBank recommendations for describing features with qualifiers and are not comfortable with the qualifier format, use the Qualifiers tab. Adding feature qualifiers is very simple. Click on the + button underneath the Qualifiers list;



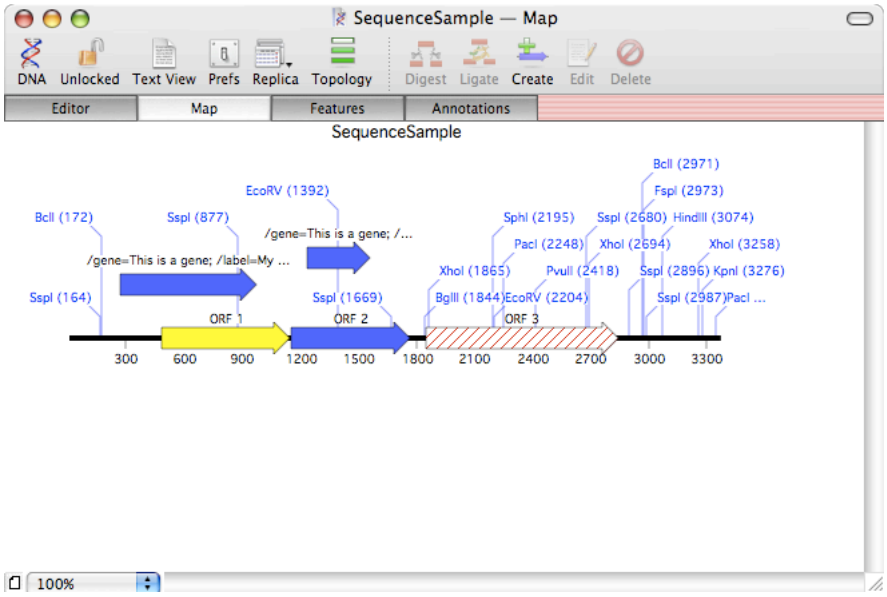
The Qualifier sheet opens. Select an appropriate qualifier from the drop down list. You can type the first letter of a qualifier to quickly select it. Note that only the qualifiers that are allowed by the GenBank specification for the feature type are available in the list.



Then type any text into the comments box that you want associated with the feature. You can continue to add different qualifiers to the list;

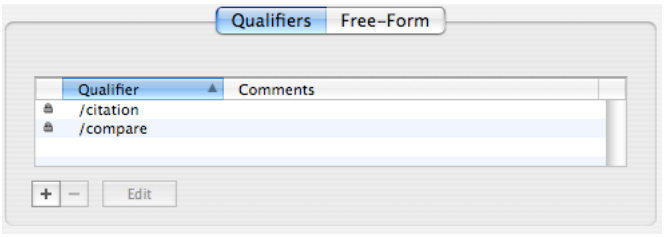


Add `/gene` and `/label` qualifiers, click OK and return to the Map tab. You will see that the feature has been created and displayed using the default symbol for that feature type;



Mandatory Qualifiers

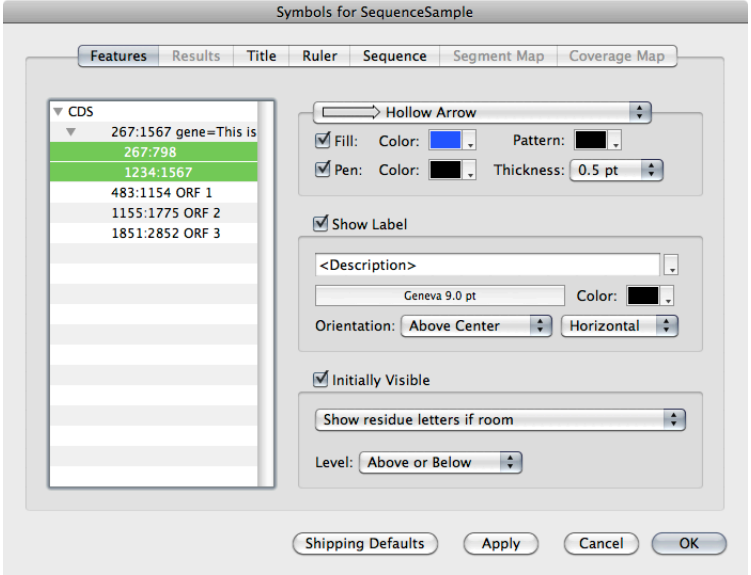
Some feature types have mandatory qualifiers – in those cases, MacVector will automatically add the appropriate qualifier(s) when you select the feature type. For example, try creating a feature of type conflict. You will see that two mandatory qualifiers are placed in the Qualifiers list. There is a small lock icon placed next to them, indicating they cannot be removed.



You should edit each qualifier and add appropriate comments before saving the feature, although MacVector will not force you to do this.

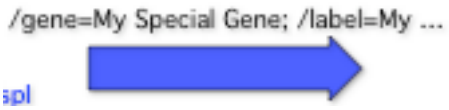
Displaying Qualifiers in the Map using Meta Tags

By default, the Map tab displays the full description as the label for the feature. However, as shown above, this can be very inconvenient, leading to rather long truncated labels. MacVector lets you use the comments associated with individual qualifiers as the label. Double click on the new `CDS` feature we created earlier to open the Symbol Editor for that feature;

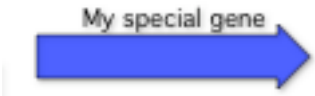


Note in this case we created a segmented feature, so both features are initially selected in the Features list. Note how the Show Label field is set to the text `<Description>`. This is an example of a "meta-tag" that MacVector uses to display information dynamically retrieved from the feature. In this case, the text `<Description>` gets replaced by the entire description associated with the feature. This will be a concatenation of all the qualifiers and their comments. To replace the displayed text with just the comments associated with one qualifier, change `<Description>` to `<gene>` and click on the OK button.

The label for the new feature changes from



to



Similarly, if you use `<label>` as the Show Label text, the comments from your `/label` qualifier will be used.

These meta-tags are very useful for showing specific information in the Map view labels. You can also click on the popup menu to the right of the label text box to display a list of valid meta tags. In addition to qualifier names the following additional meta-tags can be used:

`<Description>` or `<Desc>` - display the entire description associated with a feature. If the description contains just a single `/note` qualifier, then the only the value of the `/note` qualifier is substituted.

`<Start>` or `<start>` - substitutes the co-ordinate of the start location of the feature.

`<Stop>` or `<stop>` - substitutes the co-ordinate of the stop location of the feature.

`<Type>` - substitutes the name of the feature type (e.g. CDS or misc_feature).

`<Size>` or `<Length>` - substitutes the length of the feature in residues.

`<Total>` - substitutes the total number of cuts for this Restriction Enzyme

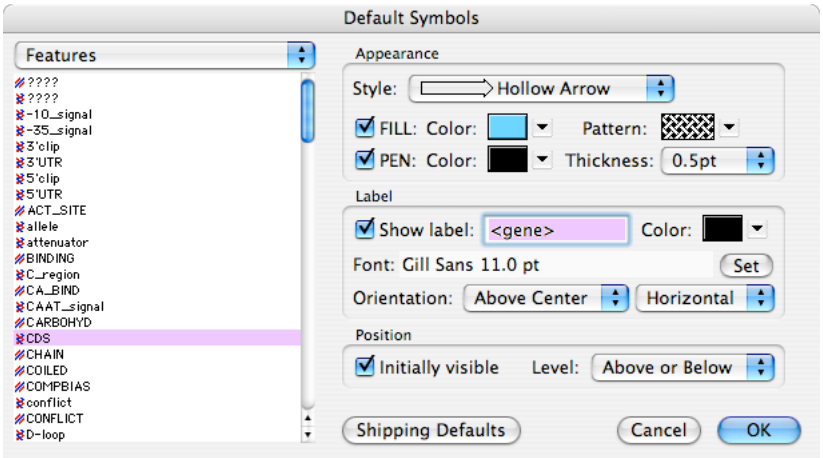
`<Index>` - substitutes the number of this Restriction Enzyme cut.

You can combine meta-tags along with free text in the Show Label field to display a lot of useful information. So, for example, typing the text `<gene> (<start> - <stop>)` will show our new feature as;

Similarly using the text `<Description> (<Start>) [<Index>/<Total>]` for Restriction Enzyme Results will create labels like this;

Setting Symbol Defaults

If you consistently use the same set of qualifiers for your sequence, or if you primarily use sequences downloaded from the NCBI, where most sequences are consistently annotated, you might want to always show features labeled by their appropriate qualifier. To do this, hold down the **<option>** key and select **Options | Default Symbols** – this opens the default Symbol Editor. (Without the **<option>** key, this opens the Symbol Editor for the current sequence). For our example, make sure you have the Features popup selected, then scroll down to the `CDS` feature. Finally, type the new meta-tags you want in the Show Label edit box and use the opportunity to change any other default appearance information you want;



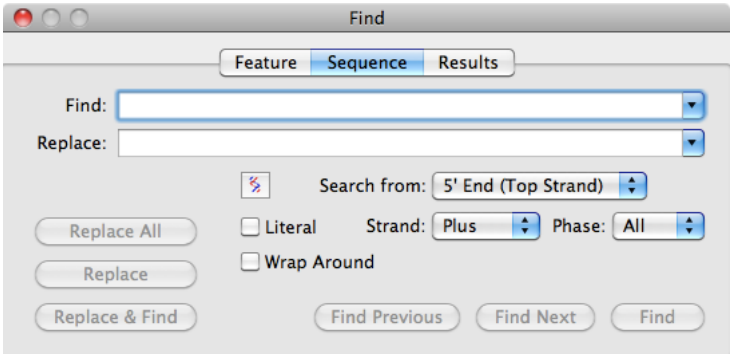
Click OK. From now on all NEW sequences you create (including any sequences you download from the NCBI, or read from disk in non-MacVector format) will use this information to display the feature in the Map tab. Existing sequences saved in MacVector format will override this information with their own default settings.

Searching Sequences and Features

MacVector has a Find dialog that lets you search not only the residues in a sequence but also the features associated with a sequence. Lets look at a few examples;

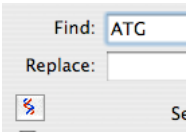
Searching Sequence Residues

Open the file `/Applications/MacVector/Sample Files/Human Mitochondrial Genome`. Select the Editor tab, then choose **Edit | Find**. The new Find dialog opens set to the Sequence tab.

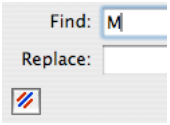


If you've used the Find function in previous versions of MacVector, this should be very familiar to you. You can search for matches in the sequence in a variety of ways. The default is to use the IUPAC codes to search for matching residues, so the sequence AGY will find AGT or AGC. If you actually want to find the sequence AGY, then select the Literal checkbox. The remaining parameters are fairly obvious, except for the little DNA icon button.

MacVector allows you to not only search DNA with a DNA sequence, but it also lets you search DNA with a Protein sequence. This button lets you tell MacVector what sort of sequence you are using in the search. To see it working, type ATG into the Find edit box;



Now click on the DNA button - it changes to a Protein icon and the Find text changes from ATG to M.



Click on the Find button in the dialog - you will see the first ATG in the Human Mitochondrial Genome sequence highlight in the background. Clicking on Find Next will move to the next ATG in the sequence. This works for any amino acid - the search uses the currently selected Genetic Code which can be changed if necessary using **Options | Modify Genetic Codes**.

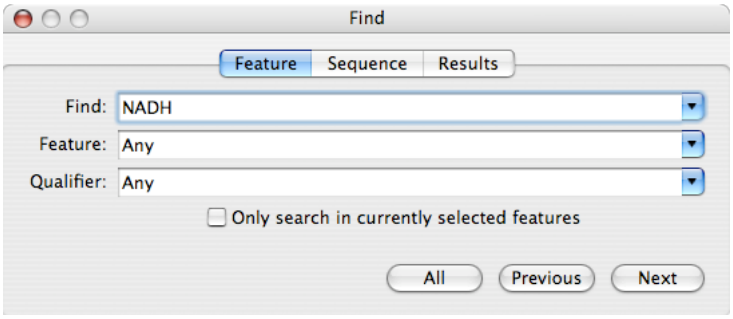
The same approach works in reverse so you can search Protein sequences for matches to DNA residues.

Searching Features

You can search for matching text within the features associated with a sequence.

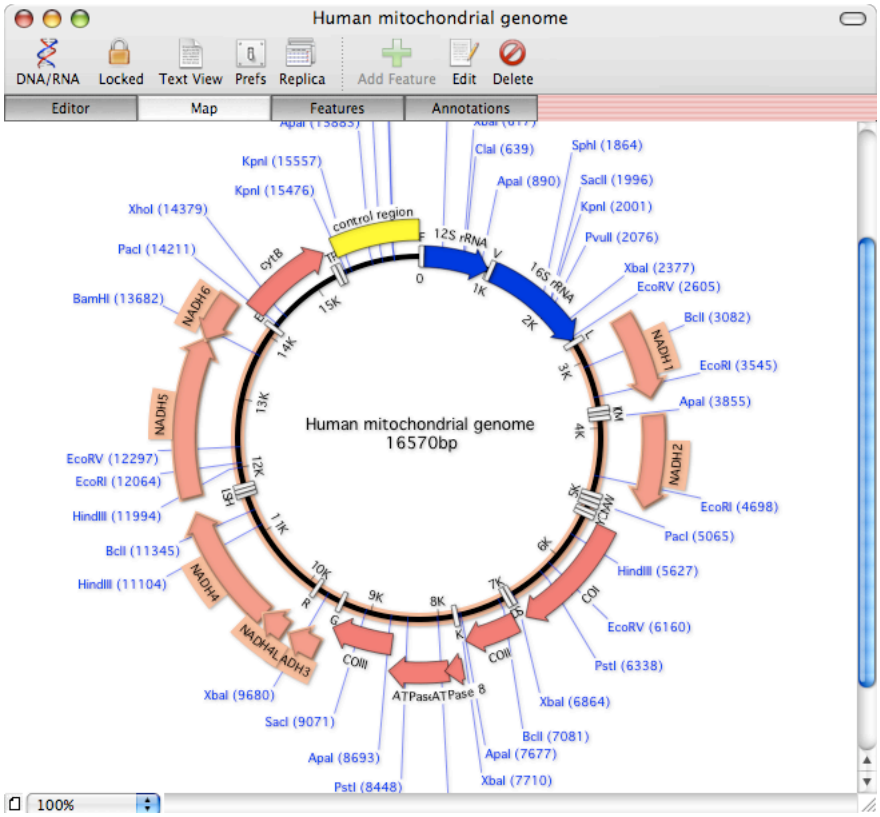
Simple Search

Switch to the Map tab of Human Mitochondrial Genome. Select **Edit | Find**. The Find dialog will be brought to the foreground with the Features tab selected;

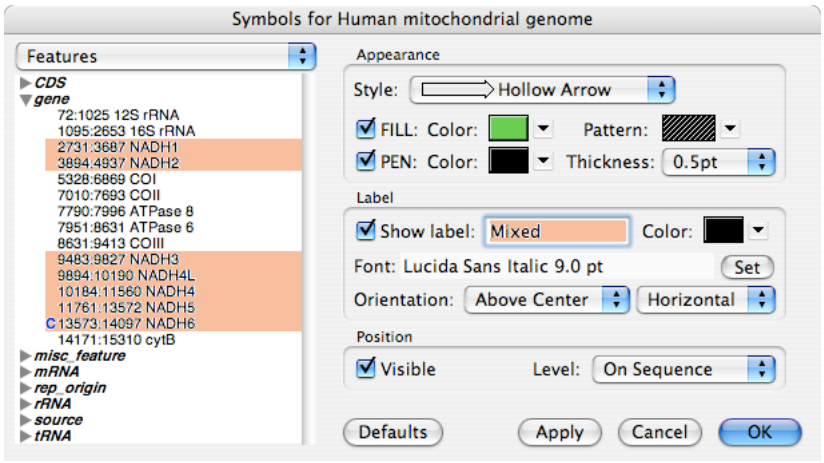


Type `NADH` into the Find edit box and then click All. This finds and then selects all features of any type that contain the text "NADH". The Map tab

will update to show all of the features selected that contain the text "NADH";



Once selected, you can double-click on one of the selected features and change the appearance of all of the features at one time;

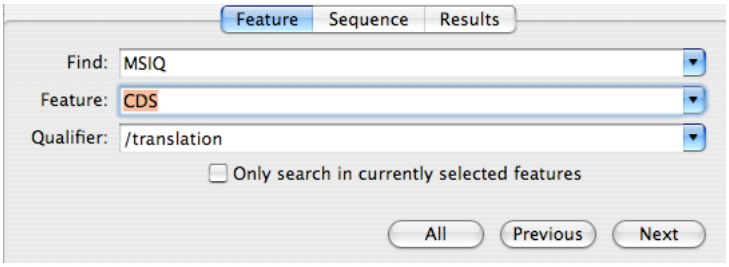


After clicking OK, the Map tab will update to change all of the selected features to the new settings.

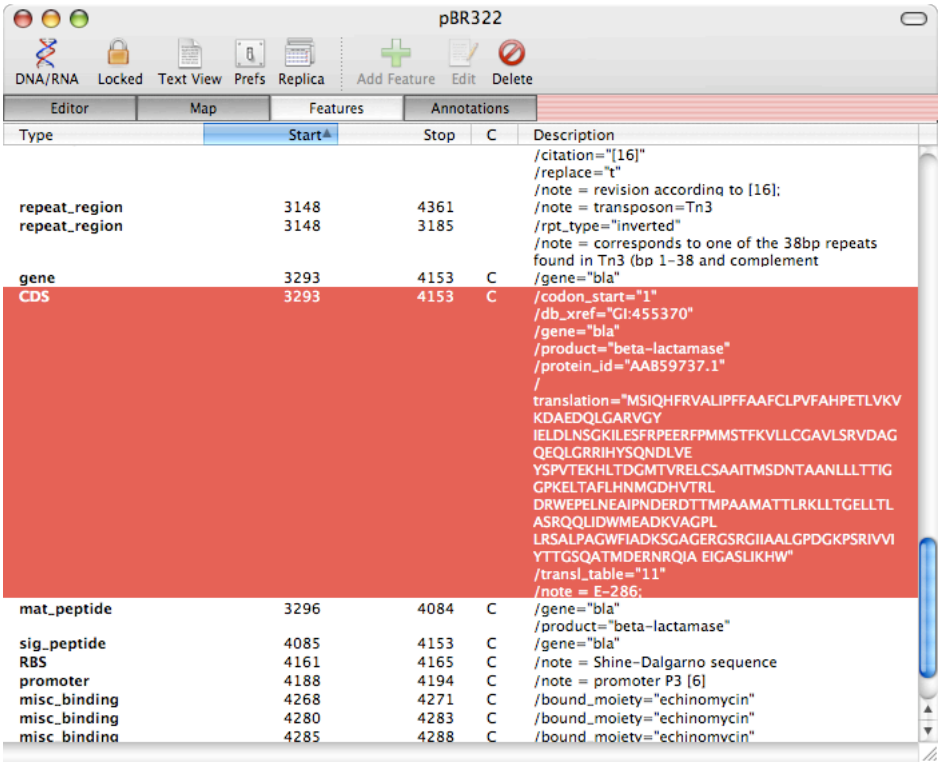
Advanced Search

You can search within a subset of the features by defining the feature type and/or qualifier type to be used in the search. Lets illustrate this with the sequence `pBR322` (again, you can find this in the `MacVector/Sample`

`Files/` folder). Flip to the Features tab and open the Find dialog. We'll look for a specific translation product in one of the `CDS` features so type `MSIQ` in the Find box (this is the translated protein sequence we are going to search for) then select the `CDS` feature type and the `/translation` qualifier.



Click on All - just one feature gets selected in the Features tab, but you should see that it has a `/translation` qualifier that starts with the sequence "MSIQ".

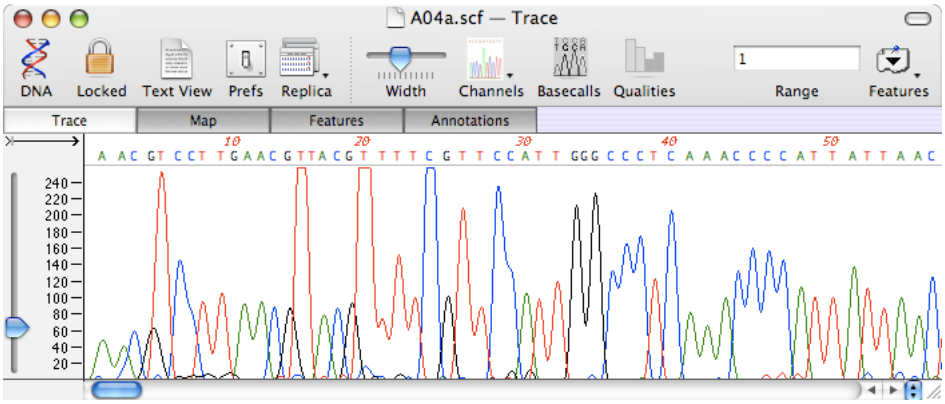


The Find dialog also lets you select matching features one at a time using the Next and Previous buttons. In addition, you can build up complex queries slowly by only searching within those features that are already selected.

The Trace ("Chromatogram") Window

Trace Tab

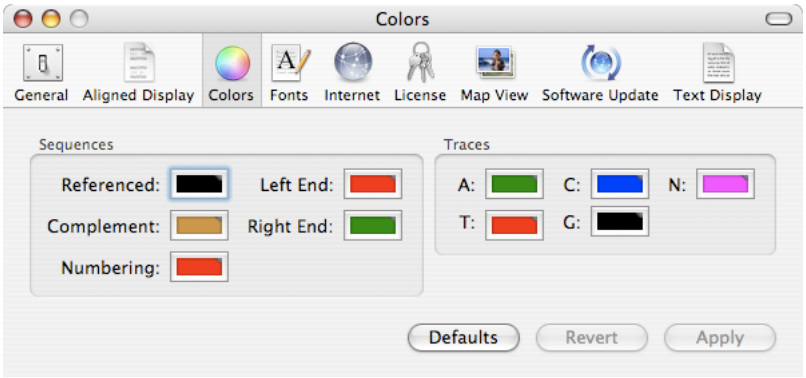
The Trace Editor window is very similar to the regular sequence window except that the Editor tab now has been replaced with the Trace tab.



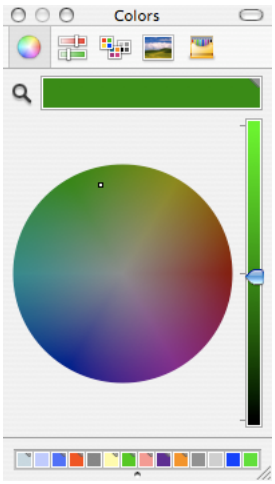
Most of the toolbar buttons have direct equivalents in the single sequence Editor tab. The only exceptions are the Width, Channels, Basecalls and Qualities buttons. The Width slider lets you widen or compress the traces so you can more closely examine adjacent peaks or zoom out for a better view of a larger section of the trace. The Channels button has a drop down menu that lets you toggle the different channels on and off. Basecalls and Qualities are only available if you have the optional Assembler module installed. These will display the base calls and quality values associated with the trace if they were present in the source file.

Trace Colors

You can control the colors used to represent the A, C, G and T channels. This is particularly useful for users with red-green color blindness. To change the colors used, open the MacVector Preferences Pane by selecting the **MacVector | Preferences** menu item, then switch to the Colors tab.



Click on one of the color buttons to bring up the standard system color selector that allows you to set the new color.



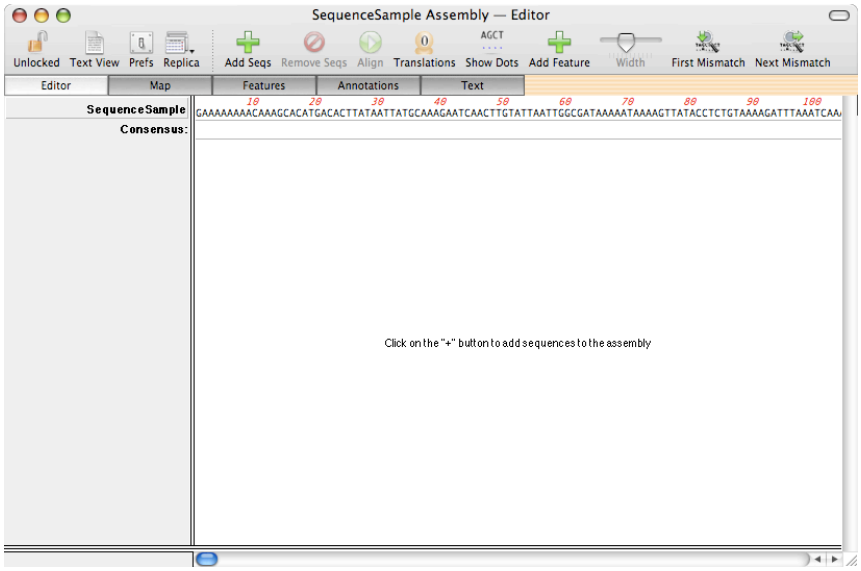
Note that the same preference pane also lets you choose the colors used for complementary sequences and numbering in the sequence editor and ligation dialog.

The Align To Reference Window

If you are not familiar with the Sequence Confirmation functionality in MacVector, there is an excellent stand-alone tutorial that can be found in `/MacVector/Documentation/Sequence Confirmation Tutorial.pdf`. The functionality was updated in MacVector 11 to let you align cDNA files to a genomic sequence – you can use this new functionality to quickly identify splice sites (described in more detail below).

Sequence Confirmation

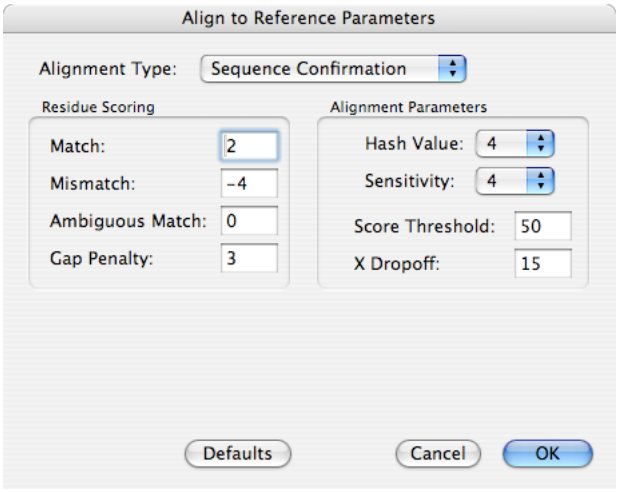
The Sequence Confirmation window also uses tabs. Open the file `/MacVector/Tutorial Files/Align To Folder/Sequence Confirmation/SampleSequence`. Choose **Analyze | Align To Reference**;



The Map, Features and Annotations tabs are essentially identical to the tabs used in the single sequence window. If you look at them now, you will see that they have exactly the same information as is present in the original SequenceSample window. Note that the data from the original has been copied into this new document; this is not a "Replica" like we saw earlier.

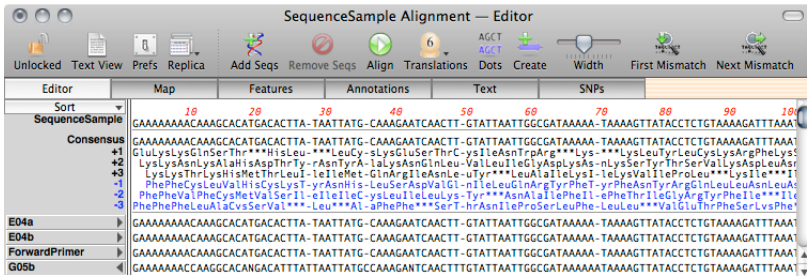
Before we look at the window in more detail, lets populate the window with trace files and align them. Click on the Add Seqs button, navigate to the `/MacVector/Tutorial Files/Align To Reference/Sequence Confirmation/Trace Files/` folder, select all of the files in the folder (**command-A** will do that) and click OK. The files will be imported into the Align To Reference window and the Editor tab will refresh;

Now align the traces with the **SequenceSample** reference by clicking on the **Align** toolbar button. Make sure you have the Alignment Type set to **Sequence Confirmation**;



Click on **OK** and after a few seconds the Editor tab will refresh to show the aligned sequences;

- **Replica** - as with the other windows in MacVector, this opens a second window so you can have a different tab open viewing the same underlying data.
- **Align** - this displays the alignment dialog.
- **Translations** - If you click on this the display will change to show the three or six frame translation of the reference (top line) sequence displayed directly under the consensus sequence;



In six-frame mode, the complementary strand translations are shown in blue to help clarify the display. The translations use the currently selected genetic code and can be display as single characters or as the three-character amino acid code. You can toggle this setting in the Options | Format Annotated Display dialog.

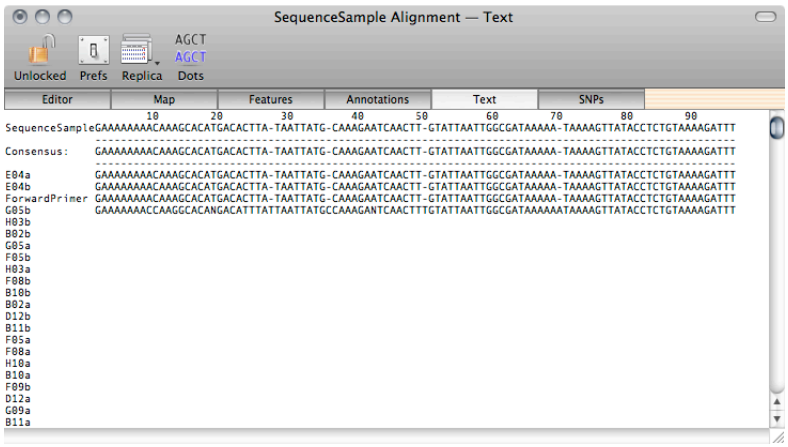
Note that the translations skip any gaps inserted into the reference sequence and treat it as a single ungapped sequence. The same is true for the automatic restriction enzyme searching in the Map tab - all gaps are ignored so that `GAA-TTC` will still be reported as an *Eco RI* site.

- **Show Dots** - this substitutes dots at any location where the aligned sequences or the consensus match the reference sequence. This lets you quickly identify residues where the alignments differ from the reference.

- **First Mismatch** - as the name implies, if you click this, MacVector will find the first mismatch between the reference and consensus sequences, move to that location and highlight the appropriate residues;

- **Next Mismatch** - this button searches from the current cursor/selection location to the find the next mismatch (5' - 3').

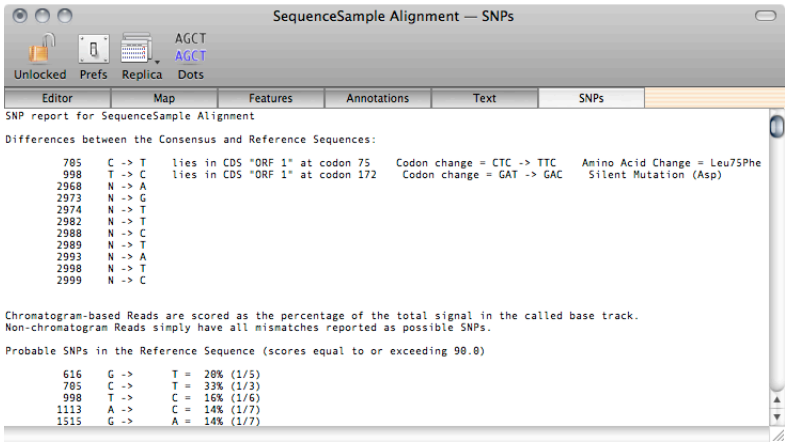
Text Tab



The Text tab is a standard MacVector plain text window, meaning that it can be printed and text selected and copied to the clipboard. The line length can be controlled by clicking on the **Prefs** button. The display matches the Editor tab for the order of Reads, position and content of the consensus, and the **Show Dots** consensus match function. The display is updated to match any edits in the Editor whenever the tab is switched in or when a replica window is activated.

SNPs Tab

This lists the details of any differences between the consensus sequence and the reference (including any amino acid changes in CDS features) and also between individual reads and the reference.

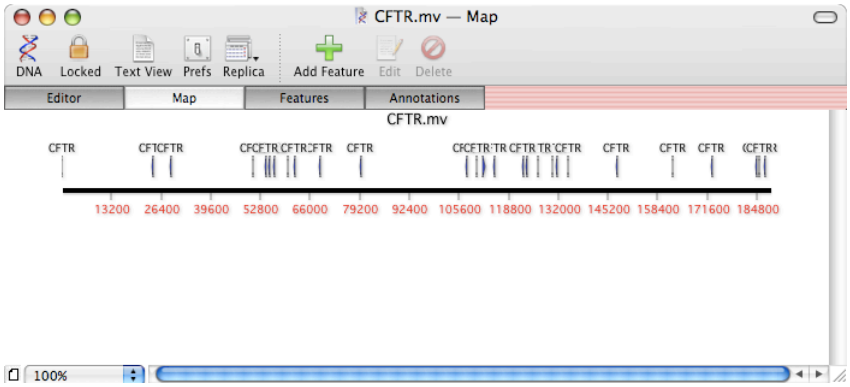


cDNA Alignment

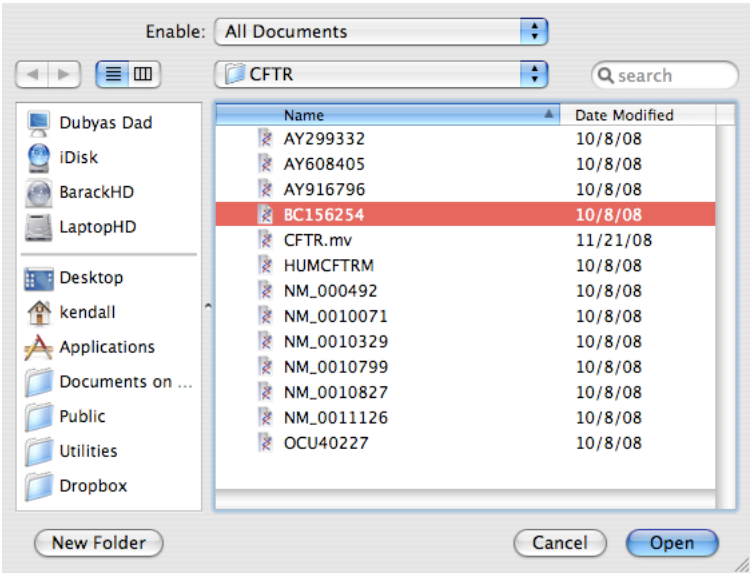
As well as being able to align traces against a template, Align to Reference can align sequences against a reference with unlimited length gaps. That makes it ideal for aligning cDNA sequences against a genome, even using the raw chromatogram files from random cDNA clones.

MacVector also has some useful shortcuts for annotating the alignments – lets look at a simple example.

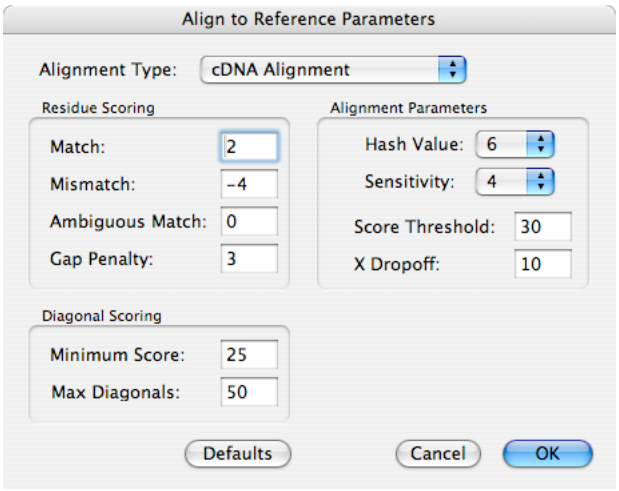
Choose **File | Open** and select `/MacVector 13/Tutorial Files/Align To Reference/CFTR/CFTR.mv`. This is the human cystic fibrosis transmembrane regulator genome locus. If you switch to the Map tab you will see that the coding region is annotated as a series of individual segments; (you may need to click on the Fit to Window button in the floating graphics palette to get all the features to display on a single line)



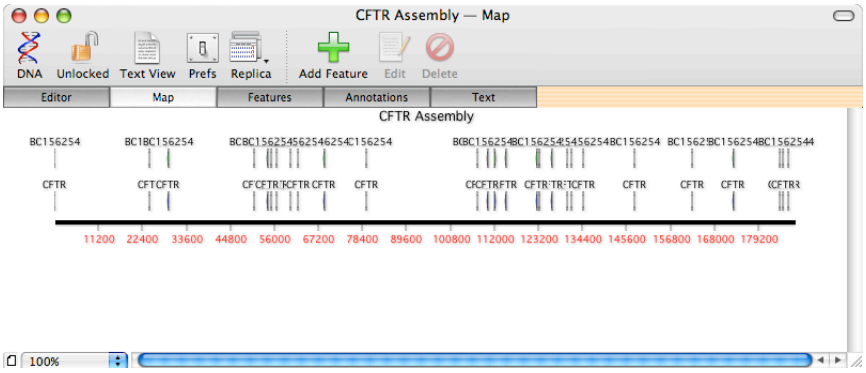
Choose **Analyze | Align to Reference** then click on the Add Seqs button and select `BC156254` in the file browser;



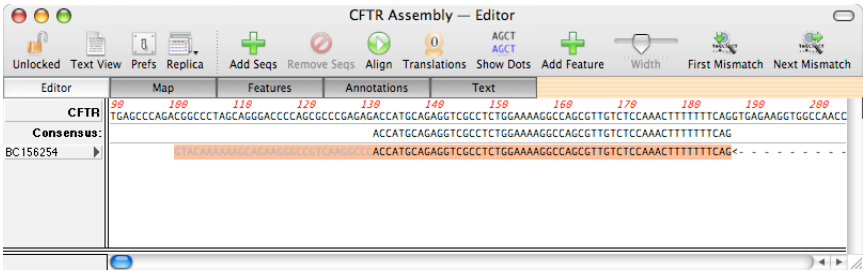
Click on the Align button and make sure you have the cDNA Alignment type selected with the default parameters;



After the alignment has completed, switch to the Map tab.



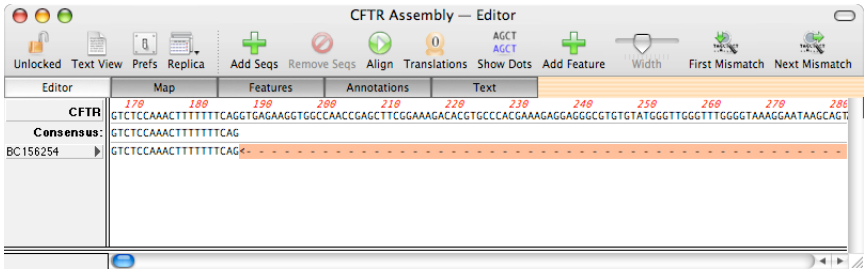
This displays a graphical overview of the alignment – it can be seen that BC156254 aligns in 27 segments, where each segment corresponds exactly to one of the CDS segments already annotated in the CFTR sequence. We can see that the alignment is correct to the base pair by looking at the Editor tab. Switch to the tab, then scroll so that the first segment is visible in the window. Double-click anywhere in the aligned segment. Note how the selection is extended to the ends of the aligned segment (the gray residues represent a region of BC156254 that has no similarity to CFTR);



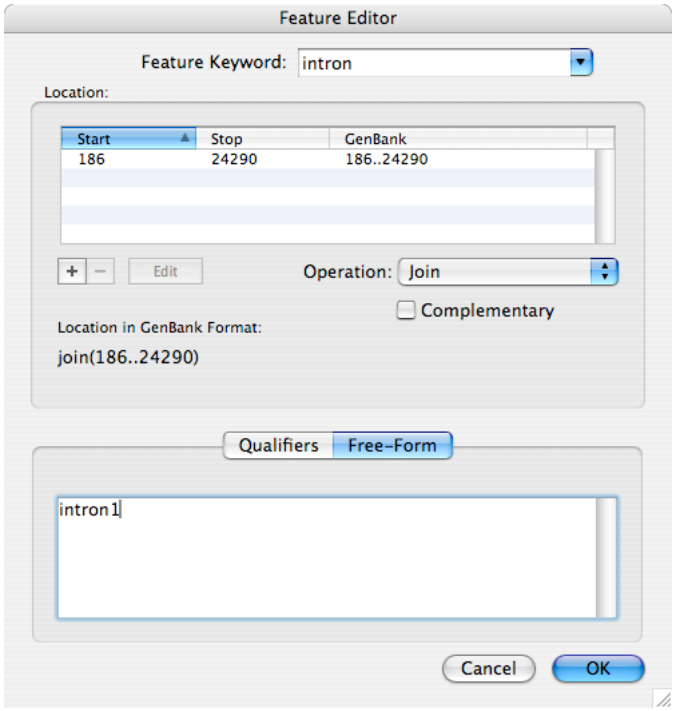
If you switch back to the Map tab (or open a second Replica window set to the Map tab) you will see that the first BC156254 segment is now selected. This selection works both ways – you can click on a segment in the Map tab and have the corresponding region select in the Editor tab. NOTE: You must hold down the **<option>** key and click on a graphical segment to select

individual segments. Without the **<option>** key, a click will select ALL of the segments.

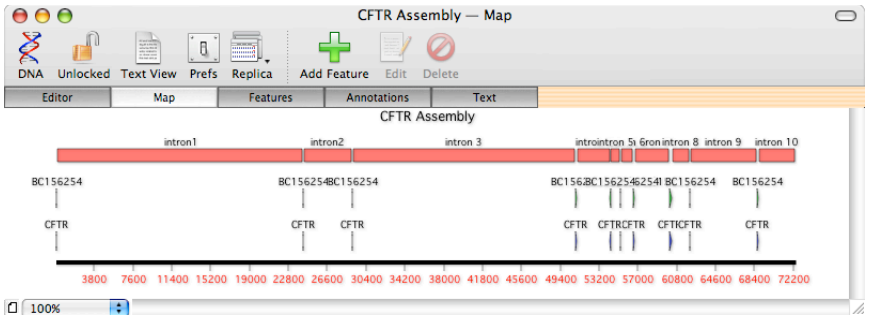
Note that the gap representing the “intron” in the BC156254 alignment is displayed using <- - - - - ->. You can double-click in this region to select the “intron” region between two aligned segments;



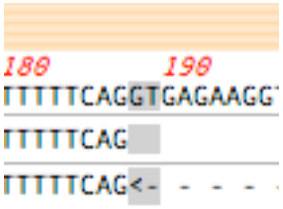
With the region selected, it is easy to add the selection as an `intron` feature to the alignment. Click on the Add Feature button, choose `intron` as the feature type and give the feature a name (e.g. “intron 1”).

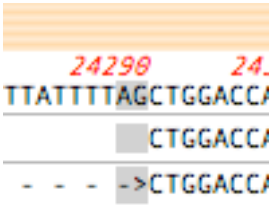


After dismissing the dialog, hold down the **<option>** key and press the right arrow key – the Editor refreshes and the next “exon” segment is selected. Press **<option>-rightArrow** again and the next “intron” segment becomes selected. You can then create a second `intron` feature and continue this way until you have annotated all of the introns in the alignment (to move backwards, hold down the **<option>** key while pressing the left arrow key);

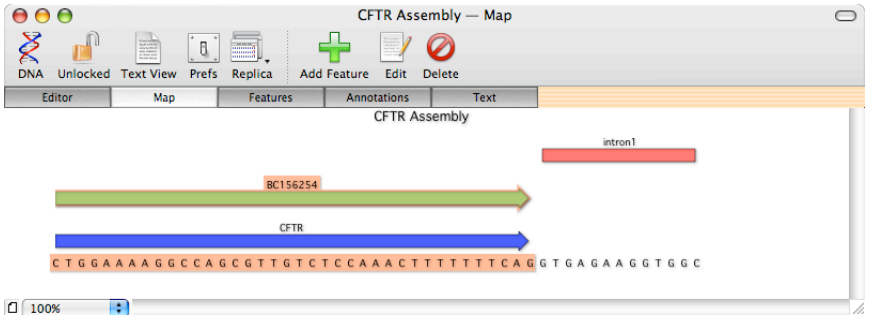


The algorithm tries to follow the GT..AG rule for introns when identifying the ends of the aligned segments. Studies have found that the majority of splice sites have the sequence GT at the 5' (donor) end of an intron and AG at the 3' (acceptor) end. This can be seen looking at the ends of the first intron;



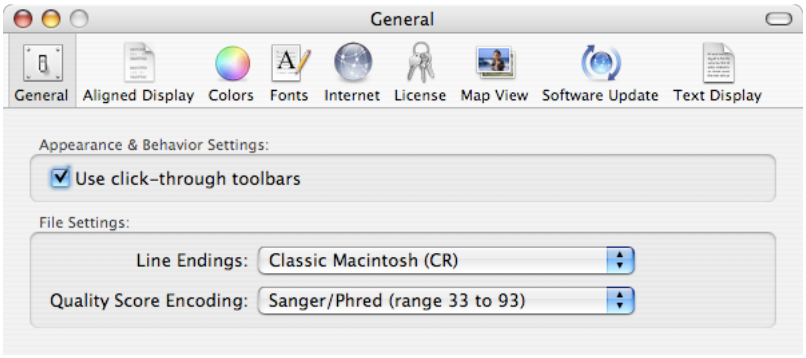


If you zoom the Map tab in to residue level, you can see that this exactly corresponds to the annotated CFTR `CDS` feature;



Preferences Panel

MacVector uses an OS X style Preferences Panel. This can be accessed by choosing the **MacVector | Preferences** menu item. Individual preference panes can be displayed by choosing the appropriate **Options** menu item or by clicking on the Prefs button in many of the tabbed views. The Preferences Panel consists of a number of individual panes. In all cases, any changes you make get applied when you either click on the Apply button, change panes, or dismiss the dialog. Normally, any windows you have open will automatically then update with the new settings.

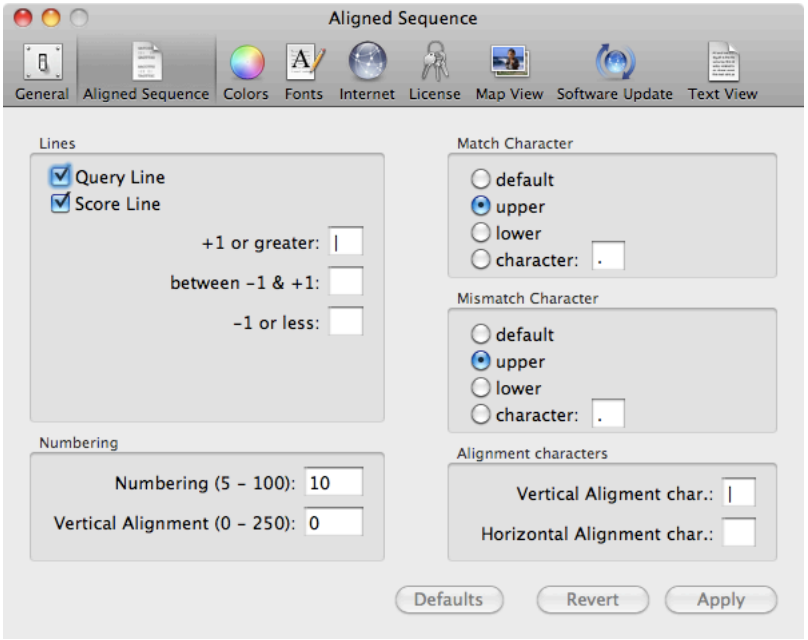


General

This contains a few settings that affect the general behavior of MacVector.

Aligned Display

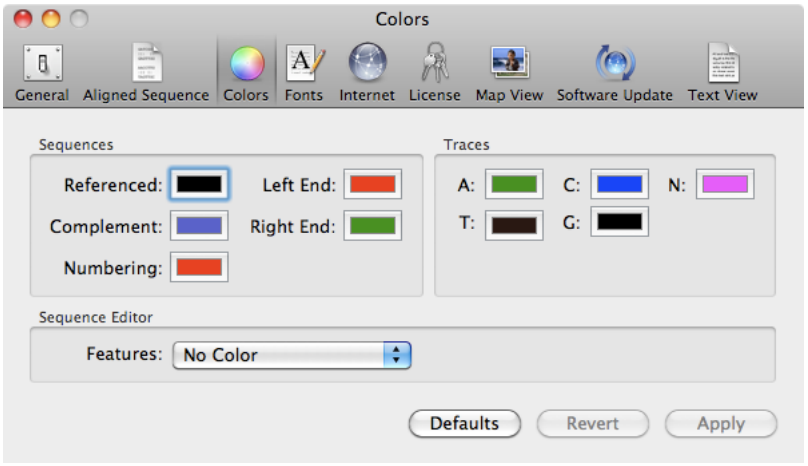
This is the equivalent of the **Options | Format Aligned Display...** dialog from earlier versions of MacVector.



These settings largely affect the layout and display of text-based alignment result windows.

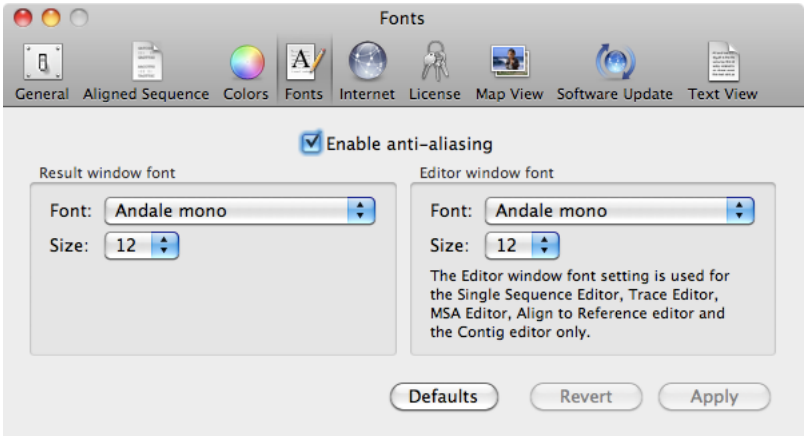
Colors

This is the pane we saw earlier that lets you change the colors used in various places throughout MacVector.



Fonts

This pane lets you control the fonts used by MacVector in the sequence editors and the text result windows.



Note: If you are having problems with misalignments when you print text windows, change the font here to Andale Mono.

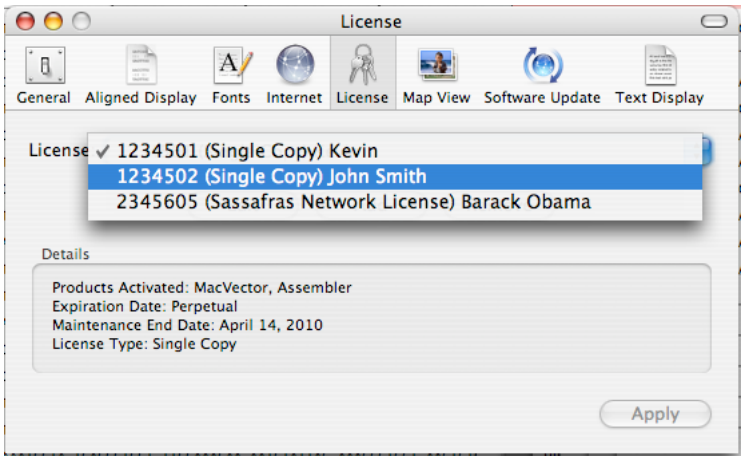
Internet

You can configure MacVector to use a local BLAST server if it conforms to the NCBI QBLAST specification. This pane lets you redirect BLAST queries to a local server, but most users should never need to change these settings. If you are interested in this functionality, contact MacVector, Inc and we can provide you with scripts and executables to get this working on your own network.

License

MacVector 10 introduced a new software-based licensing approach to replace the old hardware USB-dongle implementation. In MacVector 10.5 we extended this to allow you to easily switch between different licenses. This pane lets you edit existing licenses (to change the activation code for

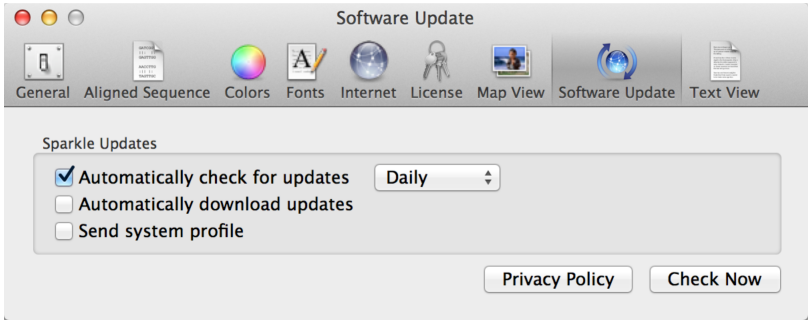
example), add new licenses or switch between licenses using a simple popup menu.



Map View

This is the Map tab preferences pane described in an earlier section.

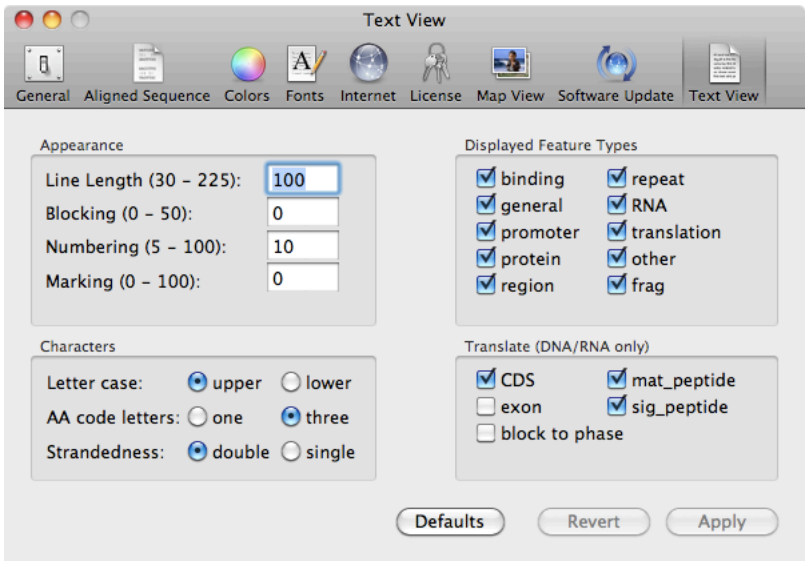
Software Update



MacVector 13 has a semi-automated automated software update facility. The default is for MacVector to check daily for updates. You can also configure MacVector to be fully automatic and install updates whenever it finds them. In addition, MacVector will check with the MacVector web site to display other information that may be of interest. That might include workarounds for bugs, availability of new documentation, utility programs or data files.

Text Display

This is the equivalent of the **Options | Format Annotated Display...** dialog from earlier versions of MacVector.



While this largely affects the text result windows, the AA code letters are used throughout MacVector wherever amino acid sequences are display in the same view as DNA.

Primer Design with Primer3

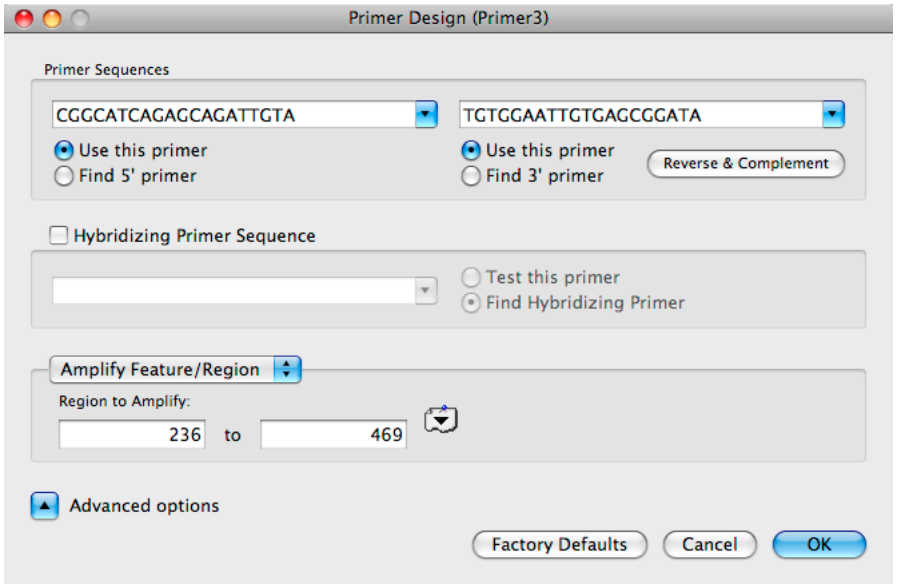
This is the main interface to use if you want to work with pairs of primers. If you are more interested in designing an individual primer, you should use the **Quicktest Primer** functionality. There is an excellent in-depth `Primer Design Tutorial` available in the `/MacVector/Documentation/` folder.

The Primer3 interface has been designed to be as easy as possible to design primers, yet to be very powerful as well, taking advantage of the popular Primer3 algorithm. For example you can design a set of primers in as few as three mouse clicks. Using the default values will produce a ranked list of 5 sets of primers to amplify a region using optimum default alues for the primer Tm and %GC. However, by changing advanced parameters you can also tune these and other values to obtain the best primers for non-standard situations.

All testing and design of primers is done using the same dialog, and can even be mixed to allow you to test one fixed primer and find another to match.

How to design a set of primers with Primer3

The simplest way to design primers with MacVector is to select a particular region of your sequence and design primers either side of this, so as to amplify this selection. With these default settings primers are chosen from a 200 bp long region on each side of the region to be amplified.



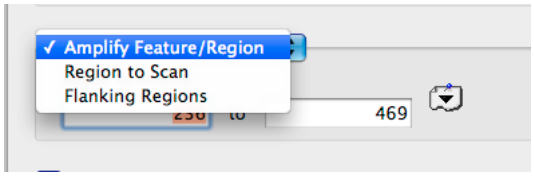
At it's simplest this can be done by:

(1) select a feature in the Map tab, or select a region in the Editor tab in a sequence window.

(2) choose **Analyze | Primers | Primer Design (Primer3)**

(3) accept the defaults and run the analysis by clicking OK.

However, there are other approaches that are controlled by a simple popup menu;.



Amplify Feature/Region

This is the default approach described above. Primers will be designed to amplify the target region.

Region to Scan

This lets you choose a region within which the amplified product should lie. You can select the sizes of products you desire, and Primer3 will design primers to amplify products of this size from anywhere in that region.

Flanking Regions

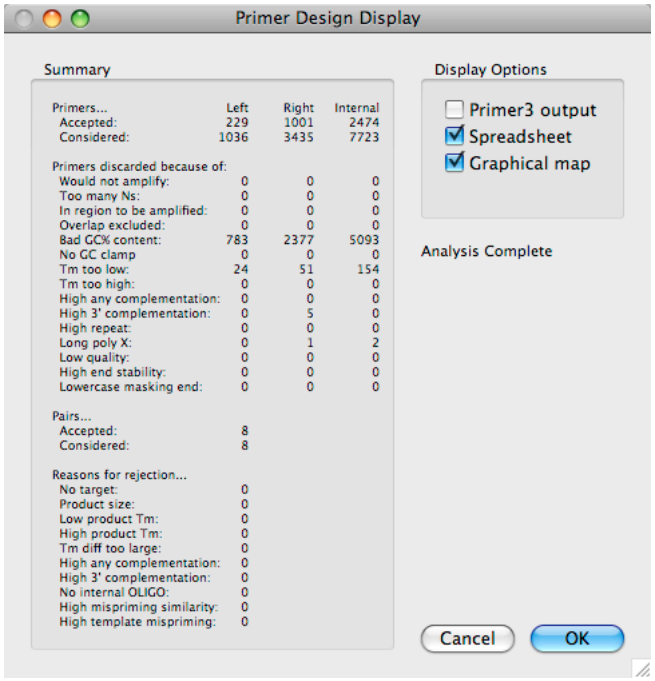
You can specify two flanking regions, so that the left and right primers will be located within these regions. This is similar to the default Amplify Feature/Region option, but offers more flexibility.

Testing Specified Primers

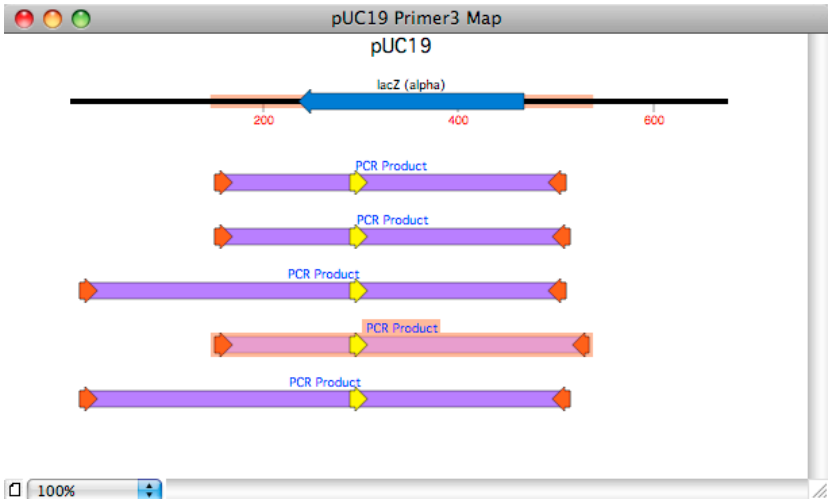
You can also test primer sequences, either as a pair, or if you have just one primer and you need to design a matching primer, then you can type/copy in one sequence, and design a suitable matching primer. You can also do this in combination with selecting an area to amplify. For example, if you want to amplify a gene, and you absolutely want to use a specific left primer at a certain point, you can select the gene in the Map view then paste or type your sequence for the primer into the left primer edit box. When you click OK, MacVector will test the left primer, and then design suitable matched right primers. To test a pair of primers, you can just type or paste them into the left and right boxes.

Primer3 Result Windows.

When you first run Primer3 you will be presented with a results dialog that summarizes the number of primers found, the number that were rejected, the number of suitable pairs of primers found, and reasons why others were rejected. If you are also designing internal primers you will see similar results for these too. If suitable pairs are found, then you will also be presented with options to see these results.

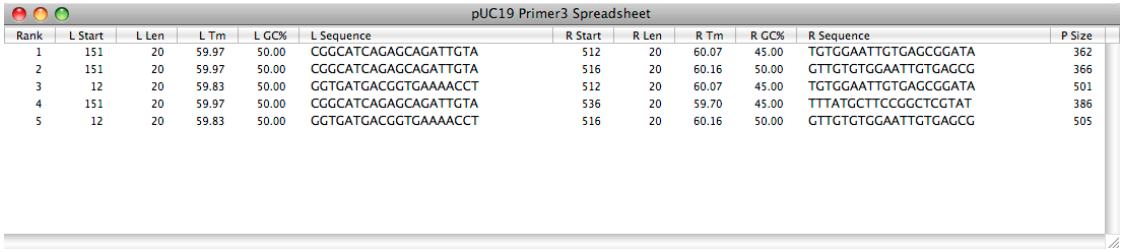


Graphical Map



The Graphical Map option displays the found primer pairs, any internal hybridization primers and the predicted product/amplicon. You can click on any graphical object in the window and the corresponding region will be selected in the parental sequence.

Spreadsheet

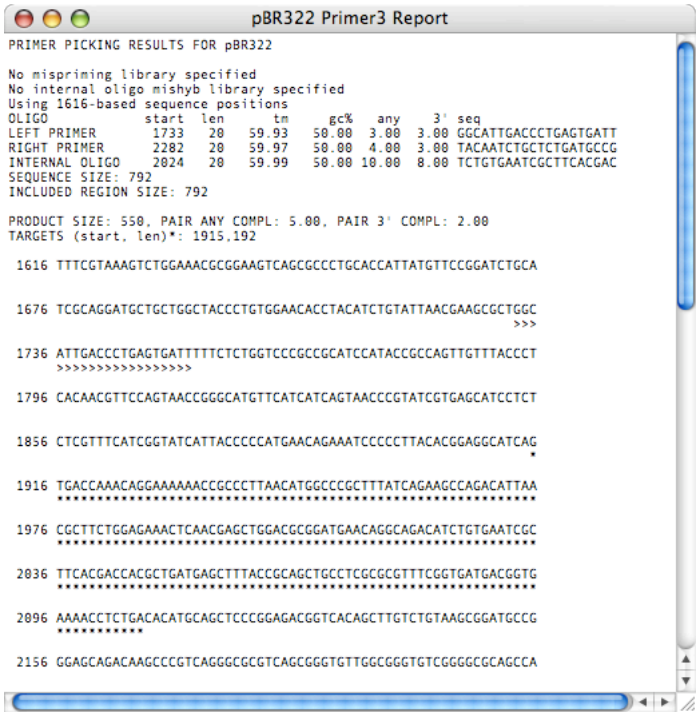


Rank	L Start	L Len	L Tm	L GC%	L Sequence	R Start	R Len	R Tm	R GC%	R Sequence	P Size
1	151	20	59.97	50.00	CGGCATCAGAGCAGATTGTA	512	20	60.07	45.00	TGTGGAATTGTGAGCGGATA	362
2	151	20	59.97	50.00	CGGCATCAGAGCAGATTGTA	516	20	60.16	50.00	GTTGTGTGGAATTGTGAGCG	366
3	12	20	59.83	50.00	GGTGATGACGGTGAAAACCT	512	20	60.07	45.00	TGTGGAATTGTGAGCGGATA	501
4	151	20	59.97	50.00	CGGCATCAGAGCAGATTGTA	536	20	59.70	45.00	TTTATGCTTCCGGCTCGTAT	386
5	12	20	59.83	50.00	GGTGATGACGGTGAAAACCT	516	20	60.16	50.00	GTTGTGTGGAATTGTGAGCG	505

The Spreadsheet option displays a window containing a text list of the primer sequences and all their appropriate statistics. You can select and then copy from the cells in the spreadsheet, allowing you to then paste individual sequences, entire lines, or the whole sheet to send to your oligonucleotide synthesis service.

In the spreadsheet view you can sort on any of the shown results. By default the view is sorted by the score that each pair is given. So the first primer pair is the most suitable and so on. However, you can also sort on Tm and %GC by clicking on the header of each of those columns. When you highlight a primer in the spreadsheet that primer feature is also highlighted in the graphical map, and on the sequence in the graphical map. This also works in the opposite direction and when you select the primer feature, it gets highlighted in the spreadsheet view. Finally in all these situations if you move back to the original sequence editor the sequence is also highlighted there.

Primer3 Output

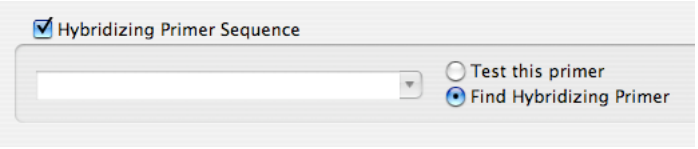


You can also see a raw results output of `Primer3`. By default this is not selected.

Real Time (RT) Primer design

Many real time PCR techniques use a third oligonucleotide that will anneal inside the desired product. This allows you to track the progress of the PCR amplification using one of the many proprietary technologies available. You can use the new Primer3 module to design these. While the default settings will usually produce the optimal third hybridizing primer, you also have control of the parameters used to design the internal primer. You can also design a internal primer suitable for use with a pair of existing primers, or start from scratch and design all three.

To scan for a suitable internal hybridization primer, simply select the Hybridizing Primer Sequence check box.



One of the advantages of being able to test or design combinations of primers is that it is easy to design a internal primer to go with a pair of existing external primers by pasting those primers into the left and right primer boxes and letting Primer3 search for a suitable internal primer. Conversely, you can also use this to design external primers to suit an existing internal primer you already have by pasting a primer into the

Hybridizing Primer Sequence edit box and selecting the Test this primer checkbox.

Auto Annotation

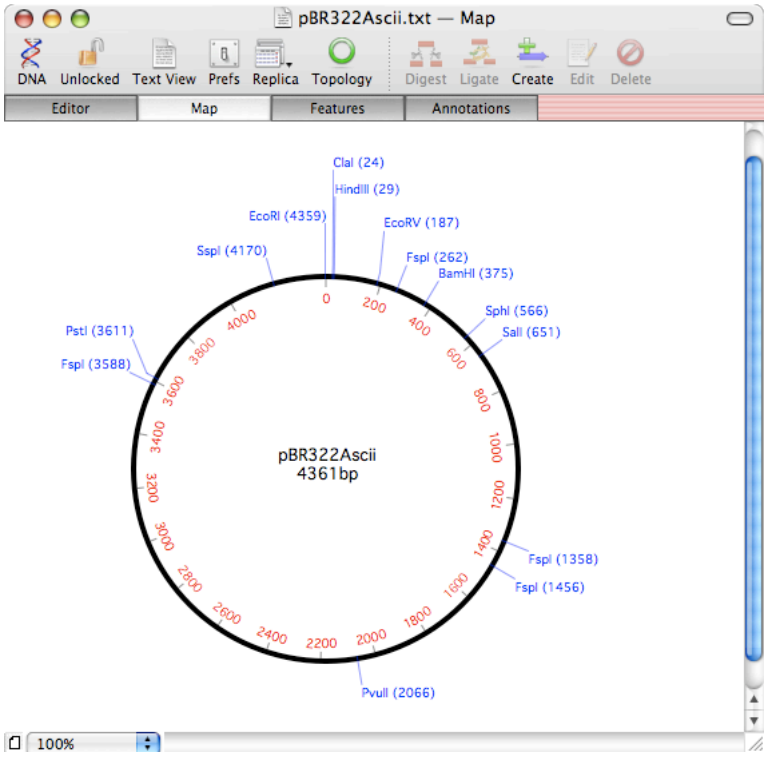
The Auto Annotation function is designed to simplify the process of consistently annotating a collection of related sequences. Whether you download most of your sequences from GenBank, or receive plain sequences from a central sequencing facility, it is tedious to have to manually annotate each feature or adjust the appearance information to match the other related sequences you have been working with.

The idea behind this function is that you can maintain a carefully curated folder (or multiple folders) containing annotated sequences of interest and when you receive a new sequence from any source, simply scan it against the folder and it will be automatically annotated with matching features from the curated sequences. The algorithm uses sequence similarity to identify matching features and has a certain amount of fuzziness that you can control to handle minor sequencing errors.

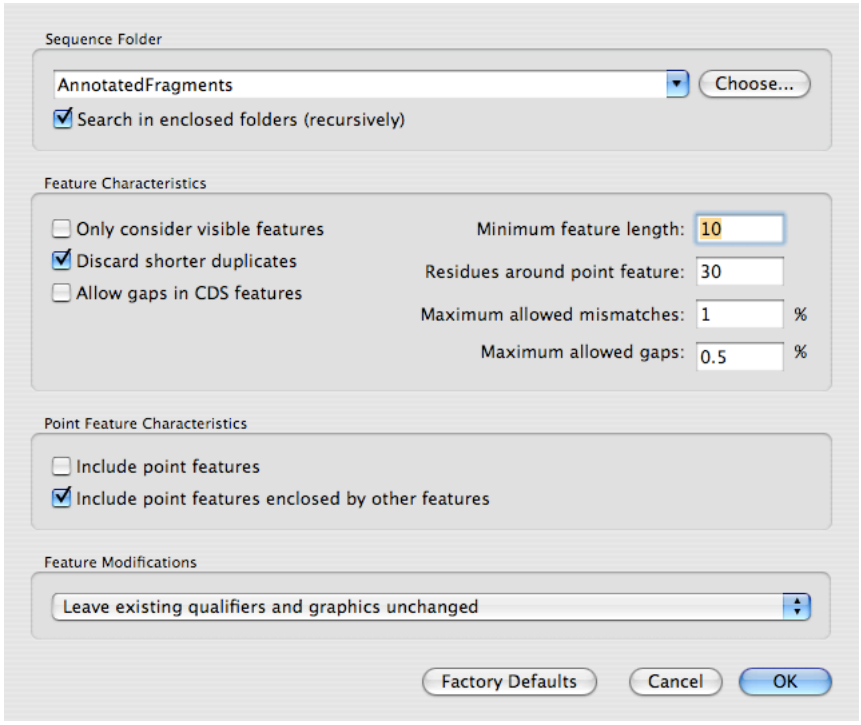
Scanning for Vector Features

MacVector has a large collection of plasmid vectors in the `/MacVector/Common Vectors/` folder, including a large number of vectors formatted to match the New England Biolabs, Invitrogen and Promega catalogs. There is also a folder containing a number of fragments of DNA, each containing a common vector feature. You can use this folder as a first pass to annotate a bare DNA sequence;

Open the sequence `/MacVector/Tutorial Files/AutoAnnotation/pBR322Ascii.txt`. As the name suggests, this is actually an unannotated copy of the classic cloning vector pBR322. The Map tab shows that there are no features on the sequence. If you don't see a circular sequence, click the Topology button to tell MacVector that this is a circular plasmid vector.



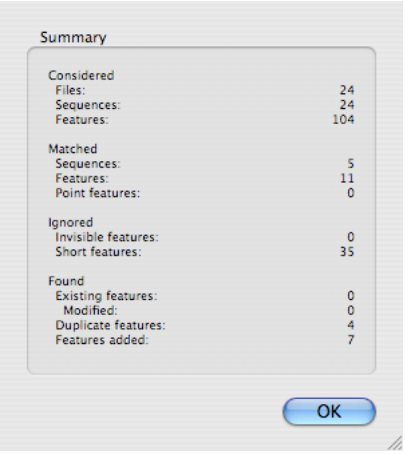
Now choose **Database | Auto-Annotate Sequence**. Click on the Choose button at the top of the sheet and select the `/Applications/MacVector 13/Common Vectors/Annotated Fragments/` folder.



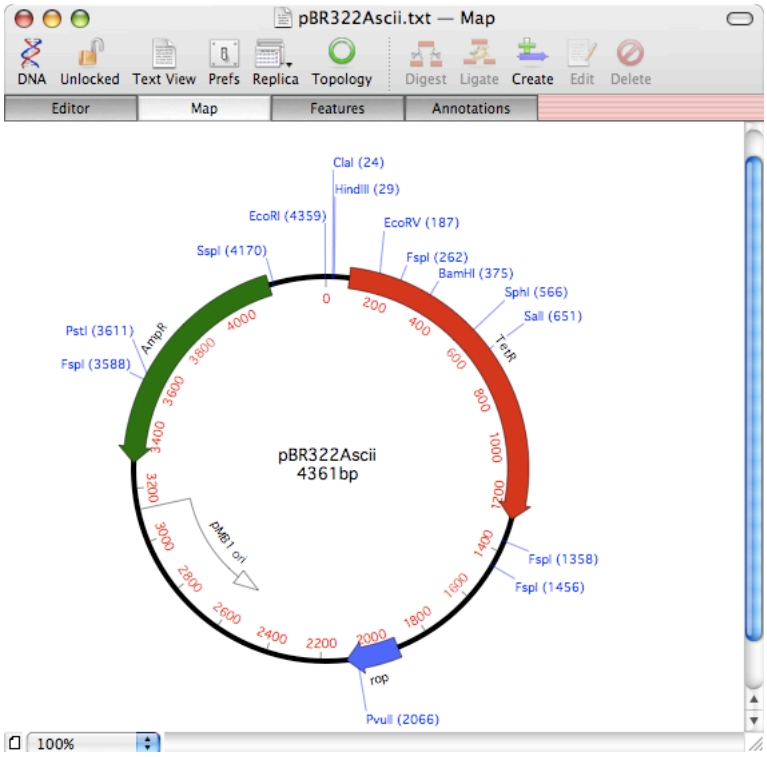
You can leave the rest of the parameters with the default settings.

Note that Sequence Folder is a combo box – it remembers all of the previous places you have searched. This makes it easy to toggle between all your favorite folders with different searches. These history combo boxes are used in numerous places through MacVector to simplify selection of folders and analysis data files like Restriction Enzyme files or comparison matrices. In addition, if you are using Mac OS X 10.5 or later, you can drag and drop folders or files onto the combo box to select that folder or file.

Click on the OK button to run the analysis. A progress sheet will appear which (depending on the speed of your machine) should disappear within a second or so to be replaced by a result summary;



When you click OK, you will see that the Map has been updated to show several features, including Tetracycline and Ampicillin resistance genes and the pMB1 origin of replication;



Parameters

The basic operation of the Auto-annotation function is fairly obvious once you see it in action, but some of the parameters need discussion so you can understand the subtle effects they can have.

The basic algorithm takes each folder sequence and looks for annotated features on the sequence. It then takes the sequence corresponding to each feature and performs a pairwise comparison with both strands of the sample sequence. If a valid match is found, it copies the feature to the appropriate location on the sample sequence, including all feature appearance information such as colors, fonts, shading, visibility etc.

Feature Characteristics

Only consider visible features

Many annotated molecules have a large number of features that are typically hidden to avoid cluttering the display. For example, the version of pBR322 present in the `/MacVector 13/Sample files/` folder has over 50 annotated features, but only 4 are displayed in the Map to keep the visual appearance clean. If you select this checkbox, only visible features in the folder sequences will be used in the analysis.

Discard Shorter Duplicates

As you build up a collection of annotated sequences, you will often find that you have related features that are essentially identical, but differ

slightly in their end locations. If you select this checkbox, only the longest of these features will be retained.

Allow gaps in CDS features

The algorithm does not require an exact match between the folder feature and the sample sequence. It can handle mismatches and also gaps inserted into either sequence to maintain an alignment. This is very useful to help overcome sequencing errors. However, if you have a gap in a CDS feature, it basically invalidates the open reading frame due to frameshifts, so you may want to ignore CDS features that need gaps inserting to match the sample sequence. However, if you suspect there may be sequencing errors in the sample sequence, check this box to make sure the any CDS features are identified.

Minimum Feature Length

Because the algorithm use sequence similarity to identify matching features, very short features (e.g. 4 base pair annotated restriction sites) would appear very frequently throughout the sample sequence. To suppress the spurious appearance of short features, set this to a reasonable value, typically from 10 to 30, to ensure the annotated features are indeed unique to the sequence.

Residues around Point Feature

As point features do not have a sequence range associated with them, MacVector artificially uses the sequence on either side of the point to determine if it should be included in the annotation. The default value of 30 means it takes the 15 residues on either side of the point and scans the sample sequence for a match to that.

Maximum Allowed Mismatches

This is simply the maximum percentage of mismatches allowed for a feature to be considered a match to the sample. The default value of 1% means that no more than 1 residue in 100 can be different.

Maximum Allowed Gaps

Set this to 0 to discard any matches that contain a gap. The default of 0.5% means no more than 1 gap can be inserted per 200 residues.

Point Feature Characteristics

Include Point Features

Check this to have point features considered. The default is to ignore them.

Include point features enclosed by other features

Use this if you only want point features to be added to your sample sequence if they are enclosed by a regular feature. This is useful for when the point feature may like near the edge of a matching region

(e.g. an exon if comparing a cDNA with a genomic sequence) and would not be annotated by the normal point feature search.

Feature Modifications

The popup menu here has three settings;

Leave existing qualifiers and graphics unchanged

Features in the folder that match existing features in the sample sequence (i.e. it has the same type and start and stop location) will be ignored and any existing features will always be left untouched.

Replace qualifiers and graphics for existing features

If a folder feature matches an existing feature the qualifiers and description associated with the sample sequence feature will be replaced by the data from the folder feature. In addition the feature appearance information will be replaced by a copy from the folder feature.

Replace only graphics for existing features

If this is selected, the graphics is replaced in the sample sequence, but its qualifiers and description are left intact. This is particularly useful for cleaning up sequences downloaded from entrez. The feature data remains intact, but you can apply your favorite graphics to the features.

Frequently Asked Questions

How does MacVector licensing work?

There are four main types of MacVector licenses;

Standard Licenses

These replace the old purple USB keys. The licensing is now entirely done by software. You are given a new serial number and can install MacVector on as many machines as you like, but only one copy of MacVector with that serial number can be running on a local network at one time.

Personal Licenses

This is a cost-effective alternative that is essentially identical to the Standard License except that the serial number can only be installed on one specific machine.

KeyServer "Network" Licenses

These are unchanged from earlier versions of MacVector. When you start MacVector it contacts a local server at your institution to checkout a license from a common pool. If all the licenses are in use, MacVector will

not run and you are placed in a queue to be notified when a license becomes available.

There is also KeyServer License variant called a *Roaming License*. This is identical to a regular KeyServer license except that if the KeyServer cannot be contacted, MacVector will fall back to using a temporary *Standard License*, as long as you have successfully contacted the KeyServer within the previous 3 weeks. This allows you to use MacVector at home or when traveling without needing a VPN internet connection to your institutional KeyServer.

Site Licenses

These are unlimited usage licenses assigned to an institution that do not require a KeyServer for operation. These licenses must be renewed each year.

All four types of license require that you enter a License Owner, Serial Number and Activation Code. The activation code contains embedded information that distinguishes the different types of license, the products activated and your maintenance end date. You can activate a license at any time by choosing **Options | Activate License**. If successful, you will get a dialog displaying your encoded license details;

Maintenance End Date

The maintenance end date for your license is particularly important. Each release of MacVector has an internal "Build Date". As long as your maintenance end date is after that date, you will be able to run that version of MacVector. However, if your maintenance expired prior to the build date, you will not be able to run that version unless you purchase a new maintenance contract with MacVector, Inc. This lets us release new versions of MacVector publicly, so you can always visit our web site and download and run the latest version without having to wait for us to send you a CD.

Note that (except for Site Licenses) all MacVector licenses are perpetual, so you can always run the version you originally purchased. In addition, all new sales and upgrades come with a 12 month maintenance contract, so you are entitled to use any version of MacVector we release during that time.

Why is MacVector rejecting my license Activation Code?

The most likely reason is that you have typed or copied the details incorrectly. In particular, you must type or paste the License Owner information EXACTLY as sent to you. If you were sent license details with the owner set to "Kevin J Kendall" you must enter it like that. "Kevin Kendall" or "kevin j kendall" will not work.

What happens to the license if someone forgets to shut down MacVector?

If you are using an individual license, MacVector will not start up if it detects another computer running on the network using the same serial number. However, if a user forgets to exit MacVector and leaves the computer, that license can be used by another user once the first computer goes to sleep. It is a good idea to make sure your computer will go to sleep in a reasonably short period of time if you are using MacVector in a lab with other users of the same serial number. If you subsequently wake up your computer and another user is now running MacVector with that serial number, you will be warned about the conflict and MacVector will enter a disabled mode. You will be permitted to save any existing files, but you cannot open new files or start a new analysis.

How do I save a graphical image to a file?

One consequence of the tabbed approach used by MacVector is that the Map tab is now considered to be just another view of the sequence. This has a tremendous advantage over earlier versions of MacVector as you can run an analysis from any of the tabs, not just the Editor window. However, the disadvantage is that the **Save/Save As** functions now work on the underlying sequence, so choosing Save As attempts to save the sequence document. To get around this either;

(a) Choose **File | Export Tab Contents As..** to save the graphical image as a PDF file..

(b) Choose **File | Print** and in the dialog that opens, click on the PDF button, then choose **Save As PDF...** from the menu.

(c) Choose **Edit | Copy**, then open the Apple Preview application (you can find this in the `/Applications/` folder) and select **File | New from Clipboard**. The graphic will appear in a new window - you can then save the graphic in any of the numerous formats supported by Preview.