MacVector 17.5

for Mac OS X

Getting Started
with
MacVector 17

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

MacVector, Inc.
Software for Scientists

MacVector Resources

Tutorials

You can find these installed in `/Applications/MacVector/Documentation/`. They are also 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 (12.6) at

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

Discussion Forums

Check out the user forums at;

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

Copyright Statement

This version of the getting started workshop was published in Jan 2020.

Contents

Introduction

MacVector 17.5 is the culmination of a series of significant updates to MacVector over the last eleven 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, agarose gel simulation, support for Next Generation Sequencing data analysis and numerous other enhancements. MacVector is a 64-bit signed and Apple-notarized application that it is fully compatible with the latest macOS Mojave (including running in "Dark Mode") and should be fully supported on all foreseeable upcoming macOS releases. 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. Any 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 create new documents from the data, or open sequences directly from *Entrez* over the Internet.

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 many 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: 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.

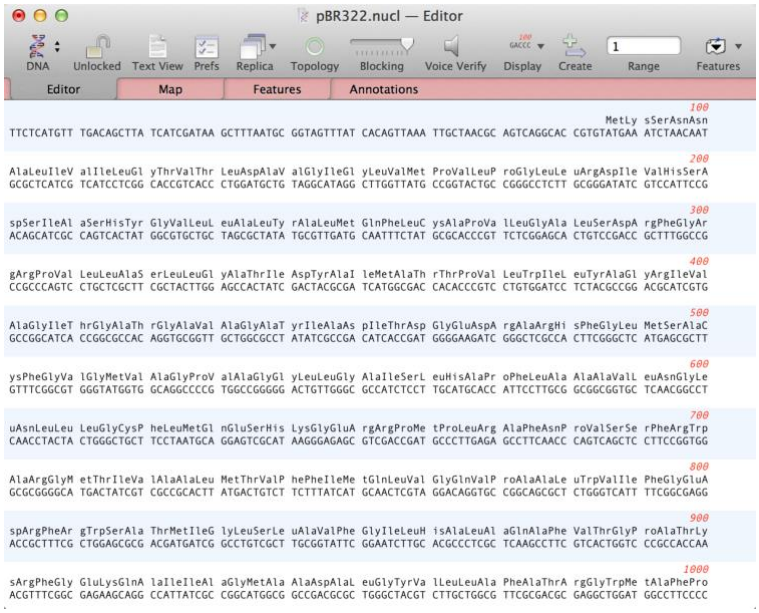
Note: You can select multiple results and open them all in separate Sequence windows by clicking To Desktop.

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 see a single window with a set of tabs offering different views of the sequence.



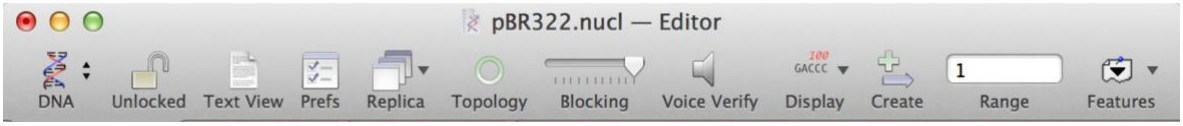
The tab bar is color-coded, where pastel red indicates a nucleic acid sequence and blue indicates protein.

- **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.

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

The key to understanding the tabbed approach is that although the tabs are each displaying different data, they are all looking at the same underlying sequence document. In particular, selections are shared between the views so that if a feature is selected in the Features tab, the graphical representation of that will be selected in the Map tab and the corresponding sequence residues will be selected in the Editor tab. You can explore this in more detail by clicking on the Replica button so that you can view more than one tab at a time.

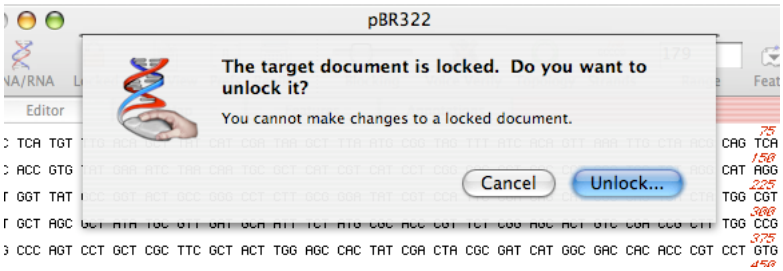
The Editor Tab



Common Toolbar Items

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

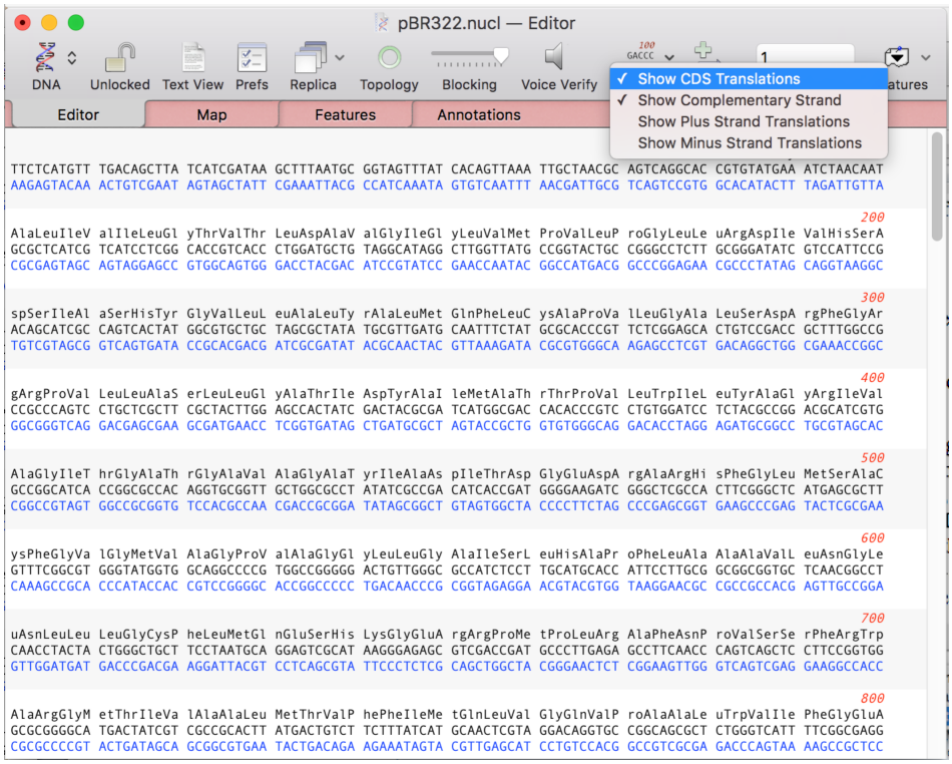
- **DNA/RNA** - use this button to toggle the molecule between DNA and RNA. (Protein windows show a non-functional Protein icon).
- **Locked** - the padlock helps prevent you from inadvertently modifying a sequence - you are prompted to unlock the sequence if you try to perform a destructive edit.



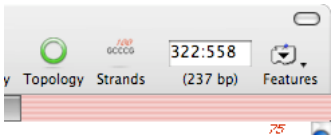
- **Text View** - this opens a separate window with the "Annotated Sequence" text view. If you have a selection in the Editor, only the selected region is displayed.
- **Prefs** - this opens a Preferences Pane dialog, usually set to the Fonts pane. Clicking on this is equivalent to selecting the **MacVector | 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. 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.



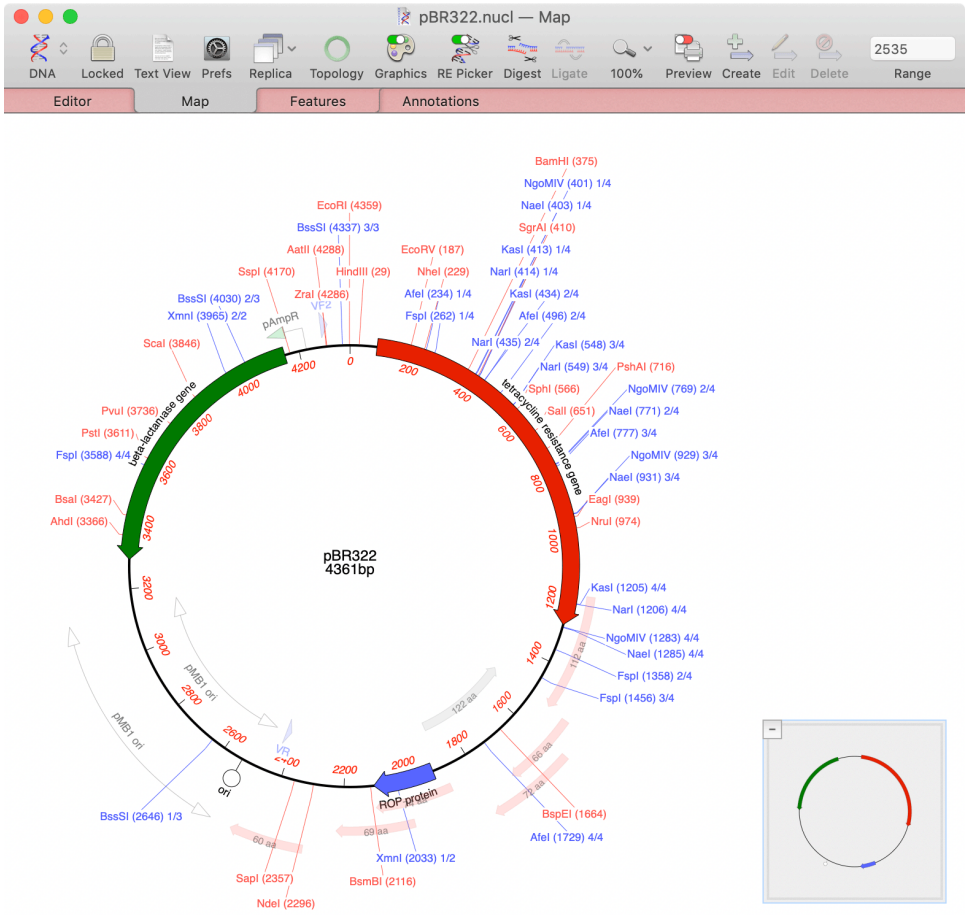
- **Create** - this button opens the Feature editor with the start and stop locations of the new feature set to the current selection.
- **Range** - shows the current selection as a range. You can also type into this box to either jump to a specific location or to select a particular range by separating 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> or <enter>` keys. Any of these will move the focus to the main sequence editor pane without losing any selection.

- **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 and also (if set in the **MacVector | Preferences | Scan DNA** pane) unannotated open reading frames, missing common features such as antibiotic resistance genes and other plasmid markers and primer binding sites. The enzymes used for the analysis are controlled by the **MacVector | Preferences | Scan DNA** pane. 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

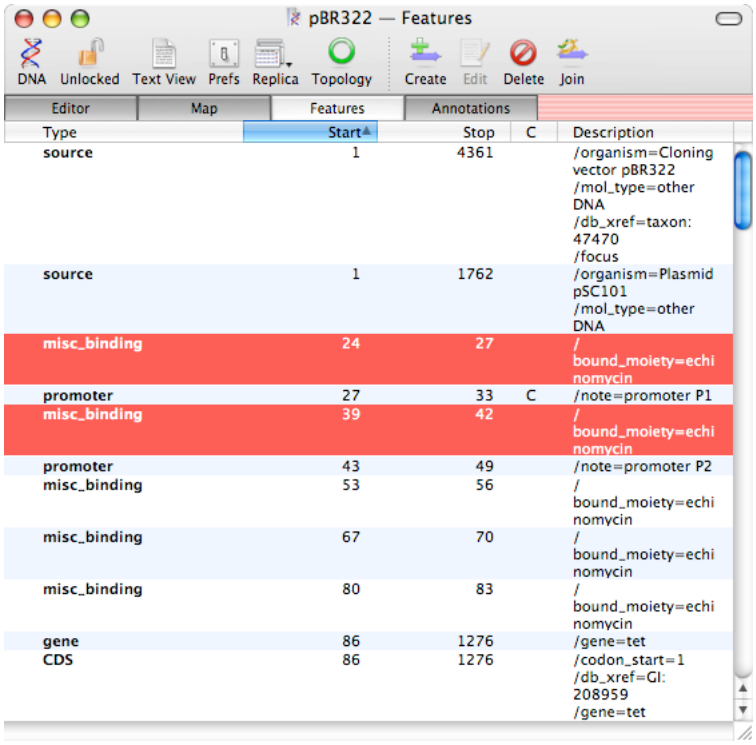
- **Graphics** - this button shows or hides the floating graphics palette. The little red/green slider reflects the current visibility of the window.
- **RE Picker** – this button shows or hides the floating Restriction Enzyme Picker window.
- **Digest** - this button is enabled whenever two restriction 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 Feature Editor, loaded with the existing details of the feature you have selected. With MacVector 16 and later, this is a dialog with tabs to edit both the Genbank representation of the feature (type, qualifier, location, strand etc) and the visual appearance (Symbol) of the feature.
- **Delete** - this deletes the selected feature. This physically deletes it from the underlying sequence document. 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.

- **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

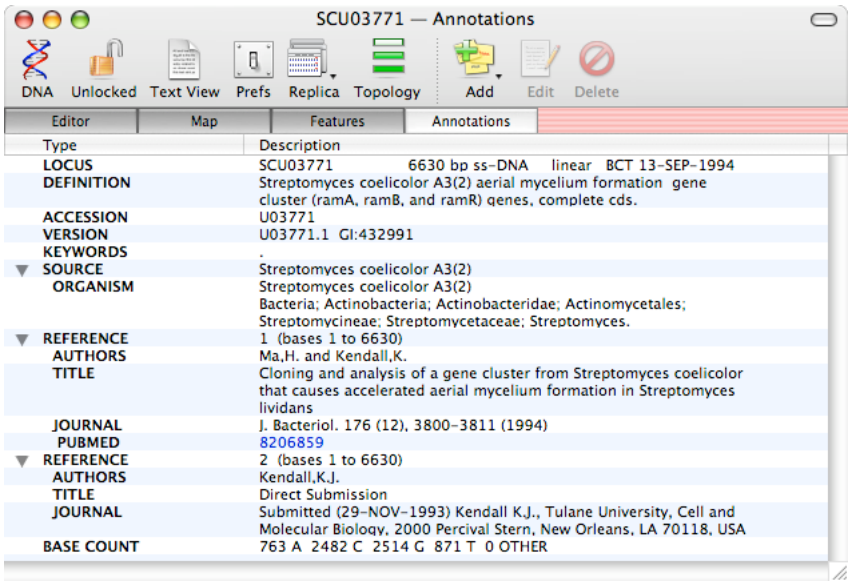


Features Specific Toolbar Items

- **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.

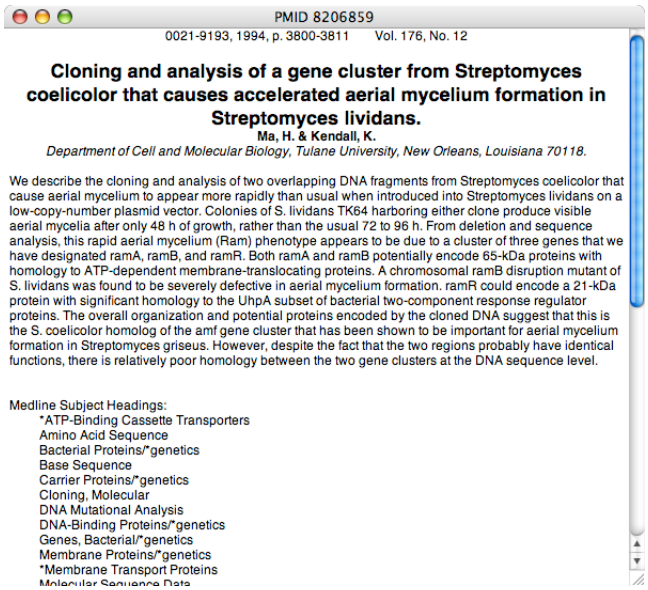


Annotations Specific Toolbar Items

- **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 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 | Export Tab Contents As…** to save abstracts in Rich Text Format (RTF).

Understanding the File Menu

Because of the many different file types and data types supported by MacVector, the **File** menu has a number of important extra items;

New – displays a sub-menu with each of the MacVector document types e.g. Nucleic Acid Sequence, Protein Sequence or Agarose Gel.

New From Starting Point… - opens the Starting Point dialog where you can keep shortcuts to mall your favorite folders.

New From Clipboard – parses the current data on the clipboard (also known as a "pasteboard" in Apple parlance) and creates a new document containing that data. The function will parse text data on the clipboard and if it conforms to GenBank or EMBL formatted data, will add the appropriate features and annotations to the sequence.

Open Recent – displays a sub-menu letting you quickly open documents you have recently worked on.

Close Window – equivalent to clicking the red Close button. Will prompt to save if the document is dirty. Hold down the `<option>` key to close ALL open windows.

Save – this simply saves the document in the default MacVector format. If the document is untitled and has never been saved before, you will be prompted for a new name. It doesn't matter which tab is currently active, this always saves the entire sequence document.

Save As… - this lets you save the document in the default MacVector format, but under a different name. The original document on disk is unchanged.

Export… - use this to save the sequence data in a non-MacVector format. Depending on the document type (DNA, Protein, Alignment, Assembly etc) there may be many format options that can be selected from a popup menu in the Save dialog.

Import Features – use this to add features from a GFF, GTR or BED file. Note it is very important that the data you are adding uses the same co-ordinate system as the sequence you are annotating!

Export Consensus As… – when working with alignments of any type, this lets you specifically save the consensus of that alignment.

Export Selected Contigs To… - used in Assembler – saves the consensus sequences of the selected contigs to a fasta or fastq file.

Export Selected Reads To… - from an Align to Reference or Contig Editor, this lets you export selected reads in fasta or fastq file format.

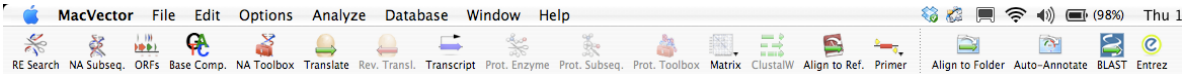
Export Selected Reads as SCF Chromatograms To… - from an *Align to Reference* or *Contig* Editor tab, this lets you export selected ABI/chromatogram sequences in the Staden Chromatogram Format, complete with edits and any quality data that has been added.

Export Tab Contents As… - this exports the data displayed in the currently selected tab in a format appropriate for the displayed data. So, the Map tab can be exported in a variety of graphics formats, the Features and Annotations tabs can be export in comma separated values (.csv) and tab separated values (.tsv) suitable for importing into MS Excel and plain text views can be exported in a variety of text formats.

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.

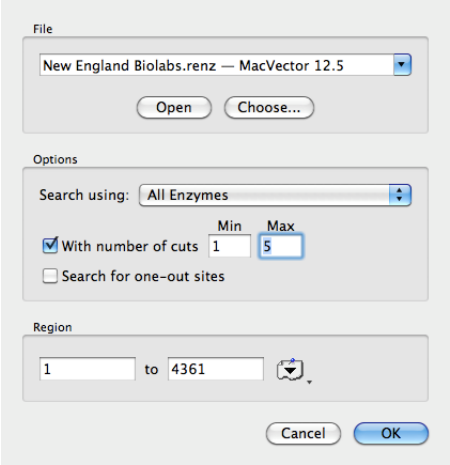
Performing a Restriction Enzyme Analysis

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

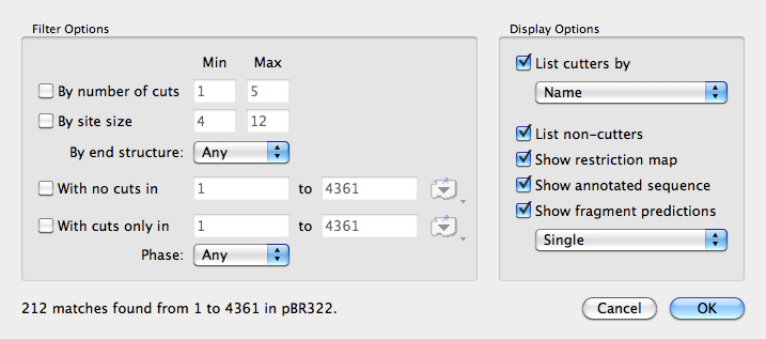
Open `/Applications/MacVector/Sample Files/pBR322.nucl`.

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

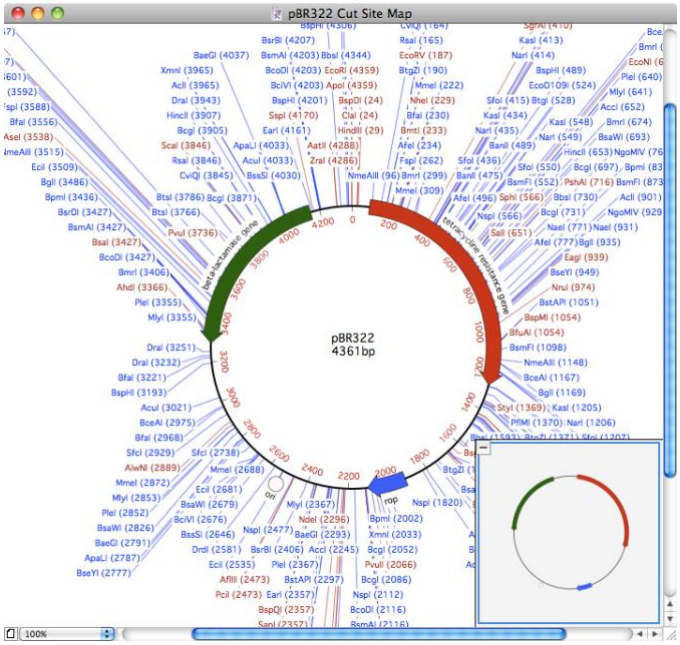
The setup dialog appears;



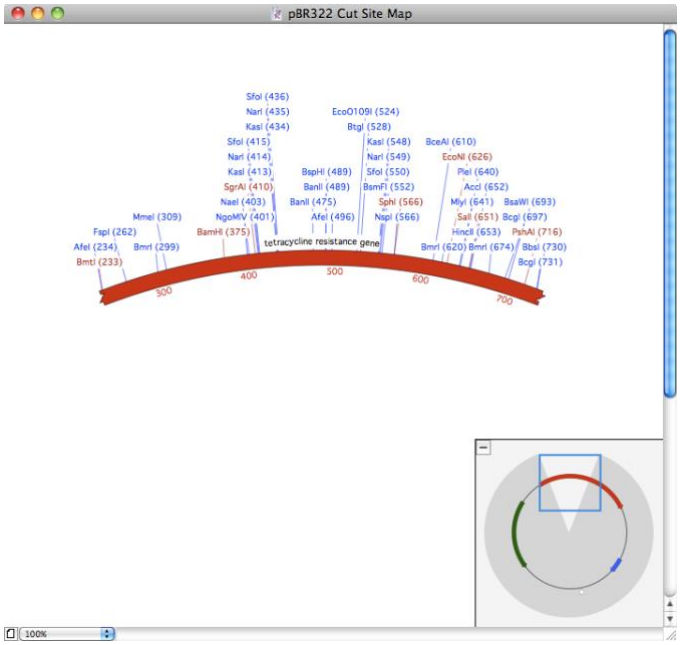
For this analysis, we'll Search using **All Enzymes**, but select the checkbox to restrict the number of cuts to between 1 and 5. After clicking OK the analysis runs and very quickly the filter dialog appears;



The Display Options section controls the result tabs 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;

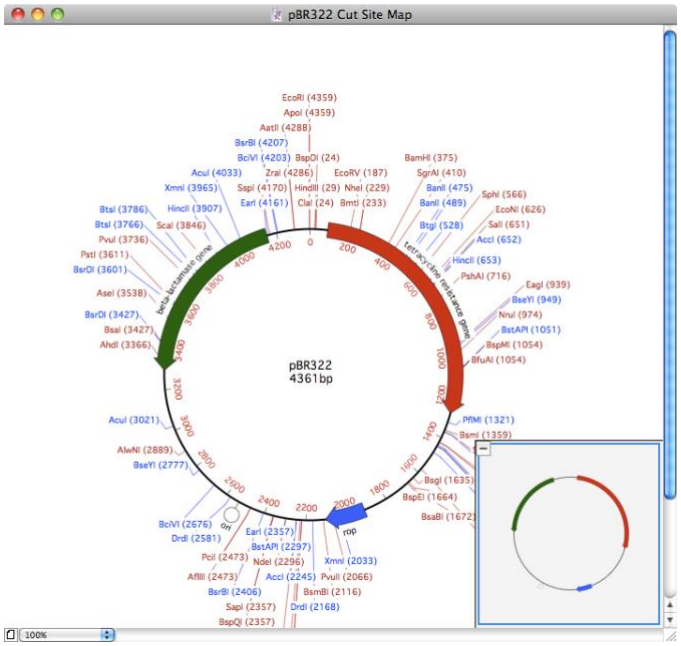
pBR322 Enzyme Cutters

Sequence: pBR322 Range: 231 to 738

Enzyme	#Cuts	Positions		
AccI	2	652		
AfeI	4	234	496	
BamHI	1	375		
BanII	2	475	489	
BbsI	3	730		
BceAI	3	610		
BcgI	3	697		
BcgI	3	731		
BmrI	5	299	620	674
BmtI	1	233		
BsaWI	5	693		
BsmFI	4	552		
BspHI	4	489		
BtgI	2	528		
EcoNI	1	626		
Eco0109I	4	524		
FspI	4	262		
HincII	2	653		
KasI	4	413	434	548
MlyI	4	641		
MmeI	4	309		
NaeI	4	403		
NarI	4	414	435	549
NgoMIV	4	401		
NspI	4	566		
PleI	4	640		
PshAI	1	716		
SalI	1	651		
SfoI	4	415	436	550
SgrAI	1	410		
SphI	1	566		

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.

Primer Database Search – similar to the NA Subsequence Search but tuned for primer searches. By default this uses the primers defined in `Primer Database.nsub`, but you can edit that file or supply your own. Understands tails on primers and mismatches within the primer binding sites.

Primer Design/Test (Pairs) – This uses the popular Primer3 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 (Individual) – 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.

Find 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 the optimum 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. You can also use this to align large NGS datasets to a reference with the caveat that currently the reads are loaded into a memory. For a 16 GB RAM laptop, this practically limits you to ~2 million MiSeq reads or ~5 million HiSeq reads.

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. You can even scan NGS data files (e.g. in fastq format) and retrieve matching hits into a separate fastq file, or even a pair of files if your original data was paired-end reads.

Database | Auto Annotation – if you have a sequence that has little or no existing annotation, 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.

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

Cloning Tools

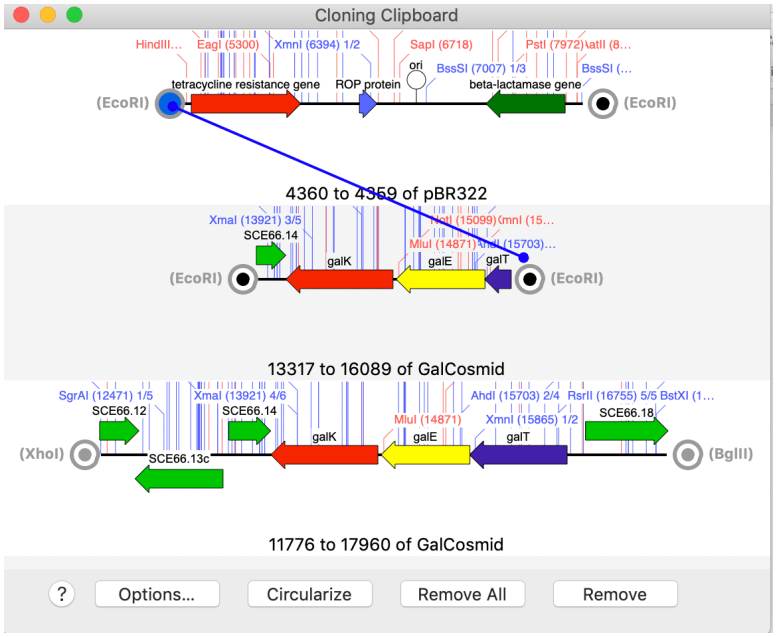
MacVector has a number of different interfaces where you can generate new annotated DNA sequences by replicating workflows you might perform in the lab using restriction enzymes, Gateway cloning, Gibson or Ligase Independent assembly or other approaches.

Copy and Paste

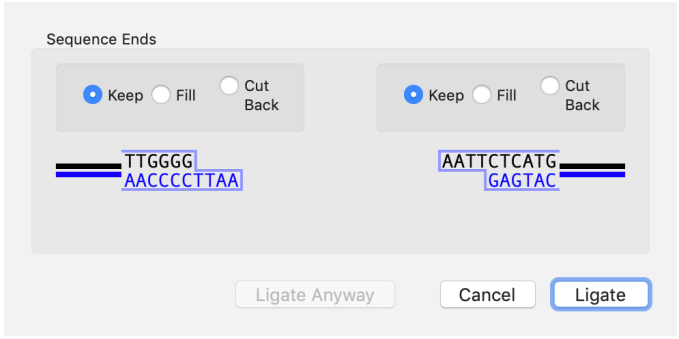
The simplest approach to creating new constructs using restriction enzymes is to <shift>-select restriction enzymes in the Map tab of a source molecule and choose **Edit | Copy** to copy that sequence (plus its annotations and the structure of its ends) to the clipboard. You can then switch to a target molecule, select a matching restriction enzyme(s) site there and use **Edit | Paste** to paste the insert. Because MacVector is "sticky-end" aware, it will not let you paste a fragment into a vector with incompatible ends, though it will let you fill or cut-back the ends if required;

Cloning Clipboard

Rather than use **File | Copy**, you can use **File | Digest** (or use the Digest button on the toolbar) to copy a fragment onto MacVector's own *Cloning Clipboard*. This is a separate window that collects all your recent digested fragments and lets you join them together using a simple drag and drop interface;



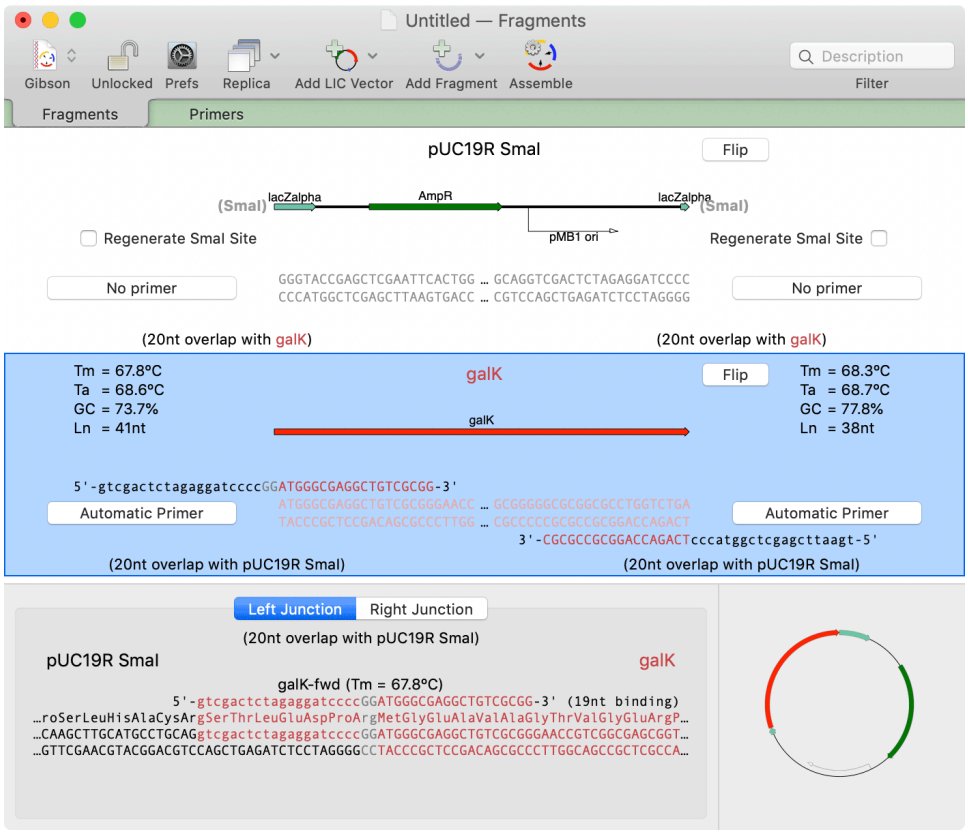
In this case, you always get a ligation dialog, so you can confirm that the ends are exactly as you would want them;



This approach has the advantage that you can assemble multiple fragments (e.g. for Gateway clonings) and many users feel more comfortable being able to view the exact steps taken to create a new construct.

Gibson and Ligase Independent Assembly

MacVector also has an interface to enable to you design and/or document Gibson and LIC assembly experiments. The interface is similar to the popular “NEBuilder” online design tool but is highly interactive, allowing you to drag fragments from other windows in MacVector, add spacer sequences, or your own custom primers and even see the effect of your changes on CDS features at the junctions of the fragments.



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). This is an incredibly powerful tool as it can find matching sequences in large fasta or fastq files and allow you to retrieve them for further analysis.

Pustell Matrix (Dot Plot) – 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)

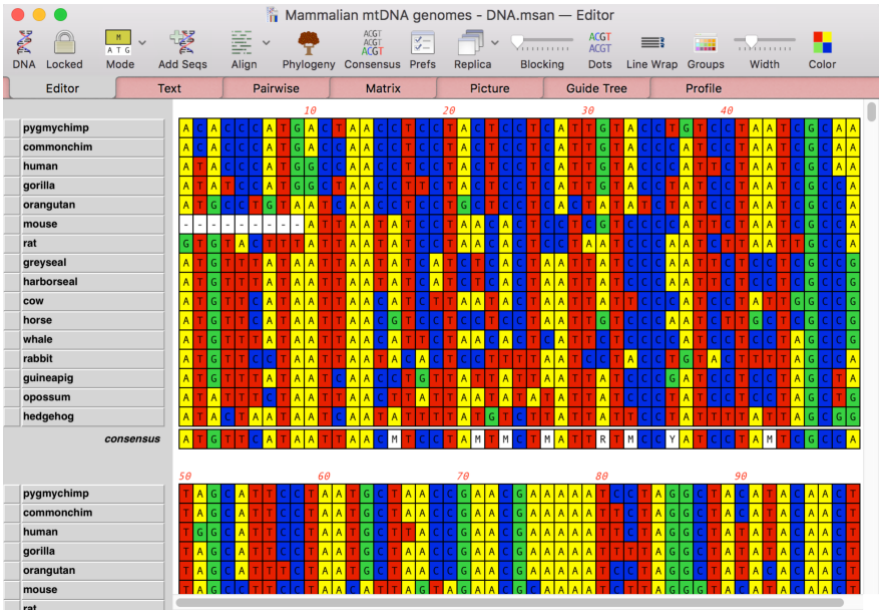
and automatically "flip" source sequences to match the reference it is ideal for re-sequencing applications such as sequence confirmation, mutagenesis analysis or simply comparing clones or related sequences to a reference. There is also a mode that can align cDNA clones to a genome, taking splice sites into account. This can also handle NGS fasta or fastq files, though there are some limitations regarding RAM usage with very large data files.

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, or create *de novo* assemblies using either Velvet or SPAdes.

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/mtDNA Genomes - DNA.msan`

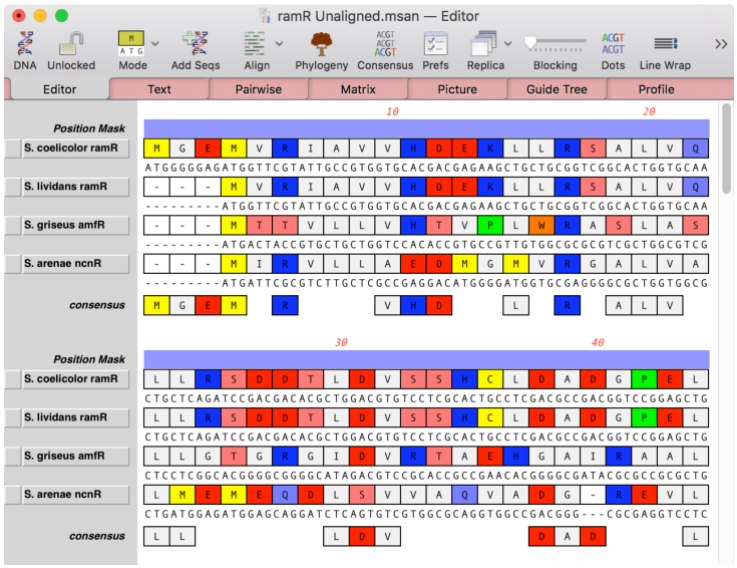


The Editor tab provides the only editable view of multiple sequence alignments. You can cut/copy/paste and perform a variety of editing functions in this tab. All of the other tabs provide read-only views of the alignment, but they do update in response to edits in the Editor tab. In the same way that 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.
- **Mode** – this lets you turn on/off the consensus line, and also lets you toggle the display of DNA sequences between DNA, virtualAA and DNA+virtual DNA. When the virtualAA residues are shown, all alignment algorithms take place on the translations of the DNA sequence, rather than on the DNA itself. This lets you align DNA sequences based on their amino acid translations;

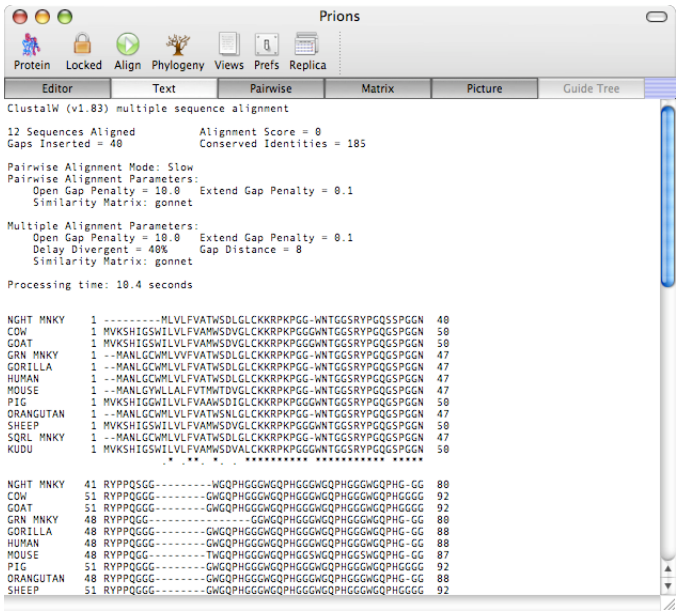


- **Align** – use this to automatically align the sequences using the ClustalW, Muscle or T-Coffee algorithms.
- **Phylogeny** – invokes the built-in MacVector phylogenetic reconstruction functionality.
- **Consensus** - this reviews the results from the last time an alignment algorithm was run and lets you quickly create a consensus sequence in a separate window.
- **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 pair-wise 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

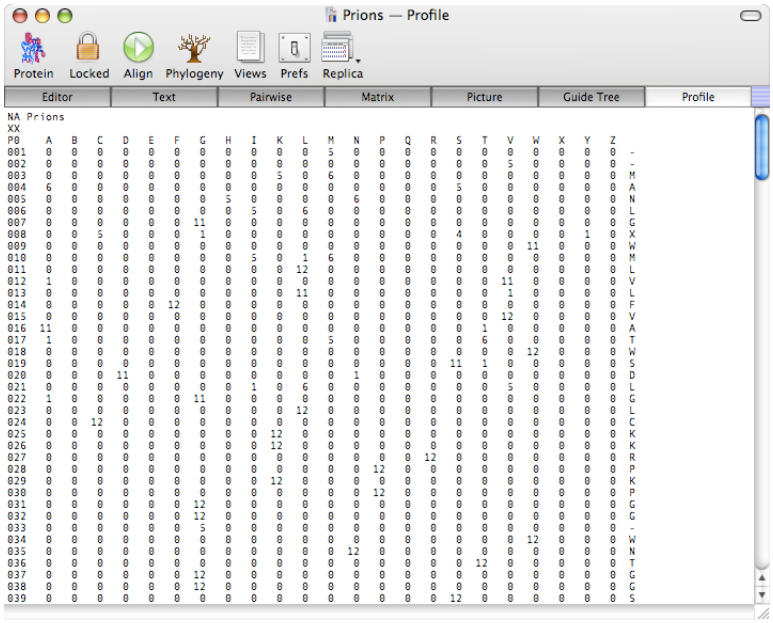
choose **File | Print** and then click on the **Save as PDF** option or choose **Edit | Copy** and then use the Apple utility **Preview.app** to create a new document from the clipboard (in **Preview**, choose **File | New from Clipboard** and then save that document to a disk file).

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.

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 residues from one of the sequences. 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 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.

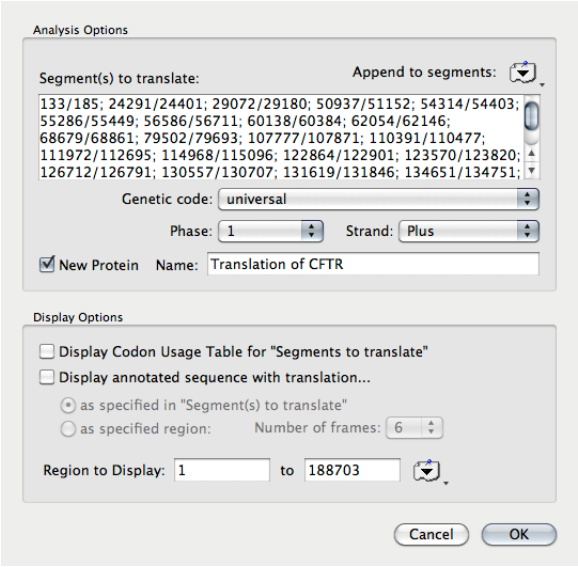
There are three primary ways of setting up a translation. Each uses the same **Analyze | Translation...** menu item.

(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. There are a variety of options to control how the codons are chosen so that you can find the least ambiguous DNA sequence for use in probes or primers, or generate a sequence for optimal expression in a specific host.

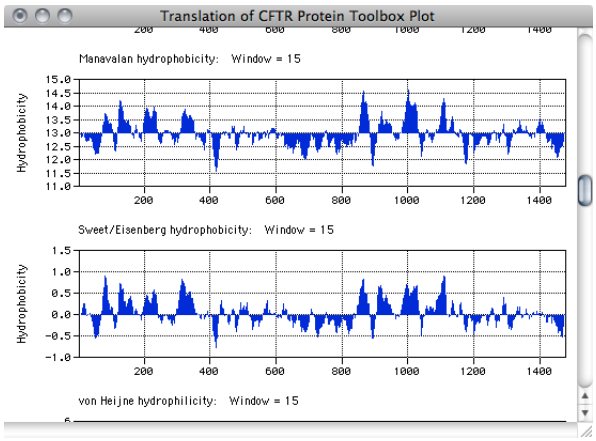
Proteolytic Enzyme – the direct equivalent of the Restriction Enzyme analysis for DNA, this lets you find potential proteolytic cleavage sites in a protein.

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.

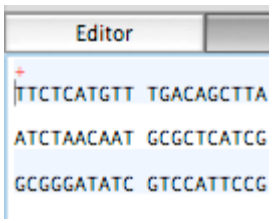
BLAST Internet Search – directly equivalent to the DNA function for identifying matches in the online NCBI databases.

Internet InterProScan Search – this contacts the popular online InterProScan server and runs a search to identify all of the functional domains in your protein sequence.

Getting The Most Out of MacVector

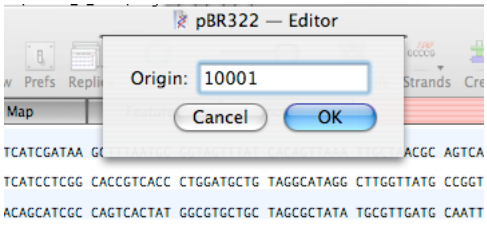
Setting the Numbering Origin

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.

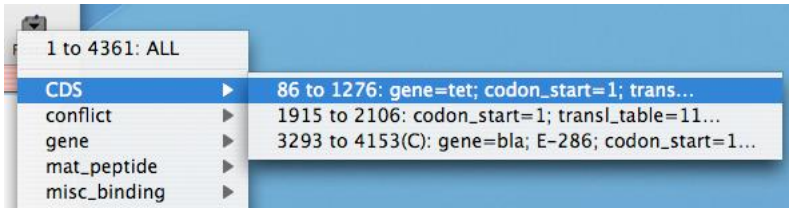


Dragging the cross to another location designates that as the "plus 1" residue – all residues before that position will be given negative numbers.

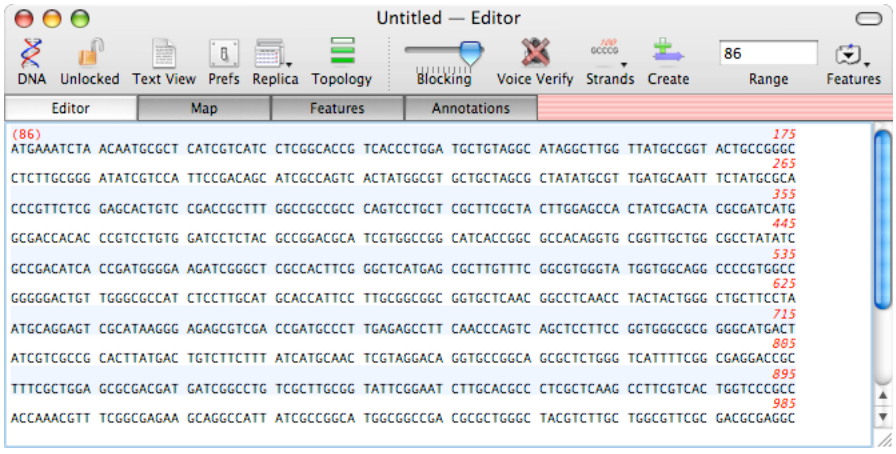
You can also set the first residue to a positive number. To set this, double-click on the red cross and enter a new start value in the sheet that appears.



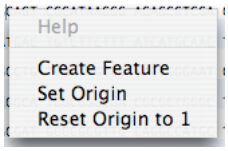
This is particularly useful if you want to work on a smaller more manageable region of a large chromosome but wish to retain the original numbering. To help with this, if you copy a section of a larger sequence and paste the copy into a new MacVector sequence window, the original numbering is retained. 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**;

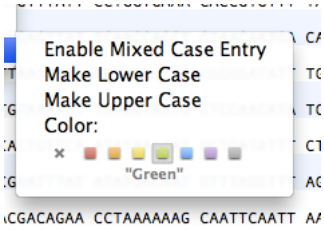


Setting the Circular Origin

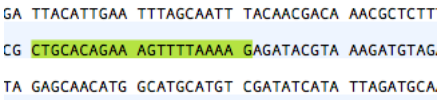
If you are working with a circular sequence, then you can change the location where the sequence is "split" in the editor. This also changes the Map tab so that the new position is located at 12 o'clock. Again, locate the flashing caret at the desired location, right-click in the sequence area and now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

Coloring in the Editor

You can 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;



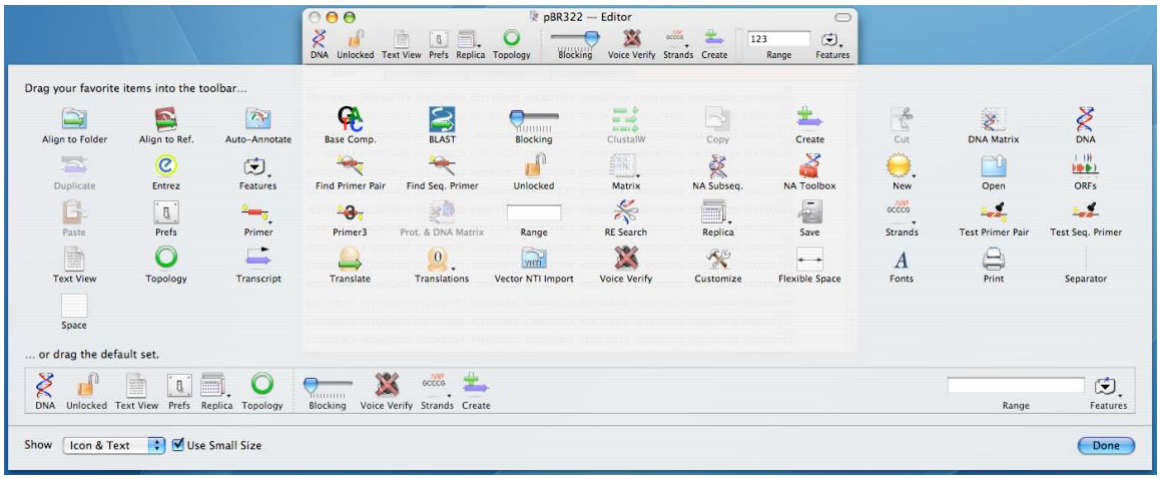
Behind the scenes, this actually creates a feature that will also show up in the Map tab (see below). When this color mode is enabled, any visible feature will be displayed using its "Fill" color in the Editor tab. You can control this behavior using the **MacVector | Preferences | Colors** pane which 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

You can 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.

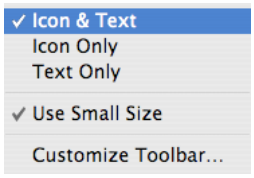
Toolbar Customization

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.

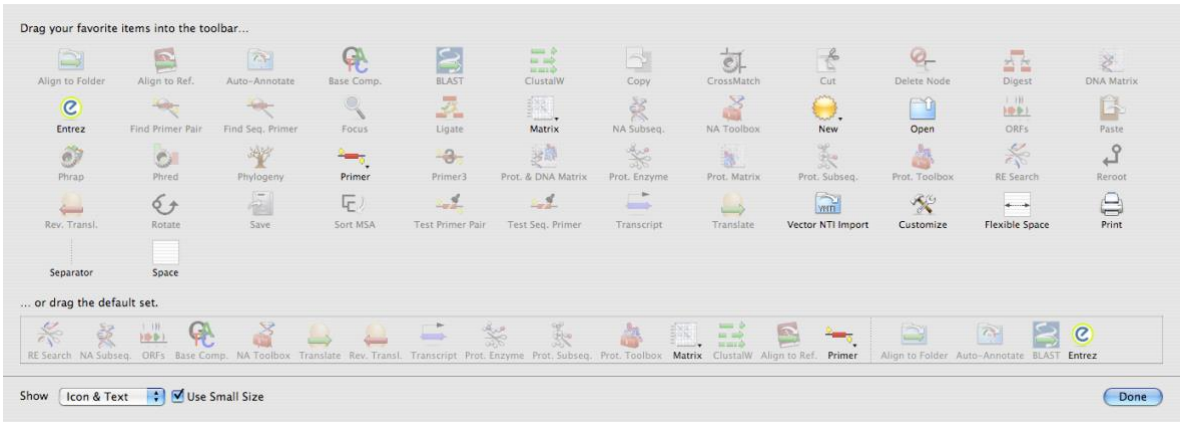


Customizing the Analysis Toolbar

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;

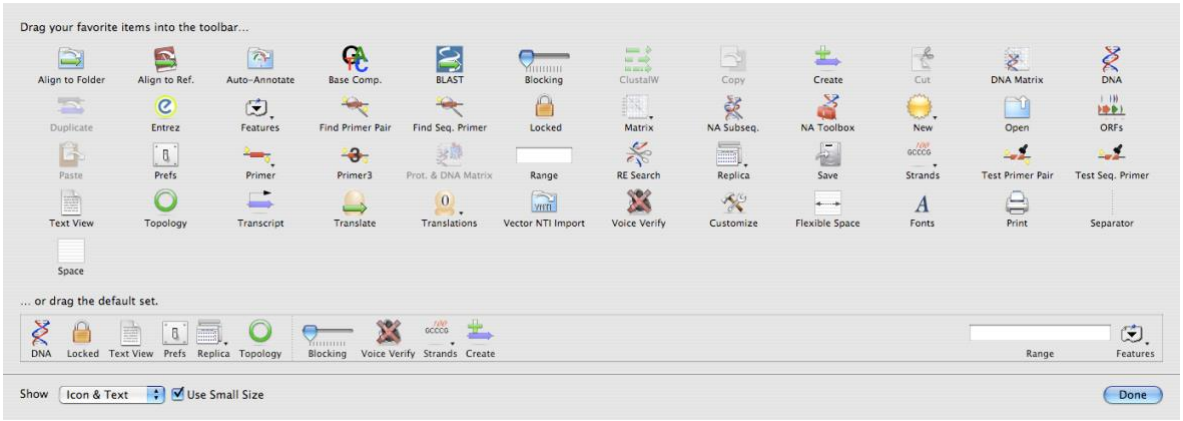


You can see that there are many more buttons available than we use in the default set. Some of the default buttons (e.g. the Primer and Matrix buttons) display a popup menu letting you choose which analysis function you want to use. However, you can add individual buttons for each function if you wish. 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.

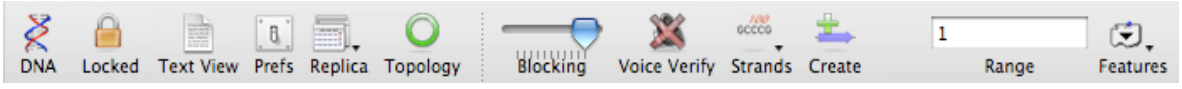
Simply drag the appropriate icons onto (or off) the toolbar until you have it as you want. You can also add separators and space icons to help you organize the toolbar. When finished, simply click Done and the toolbar will take on your new customizations.

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;



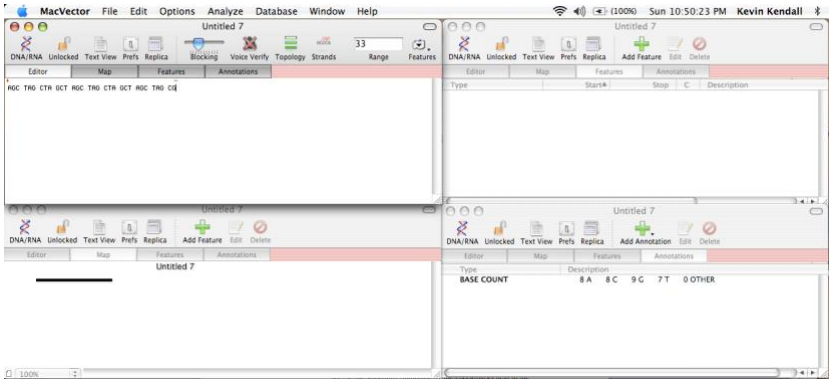
To something like this if you like to do a lot of editing and analysis;

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.

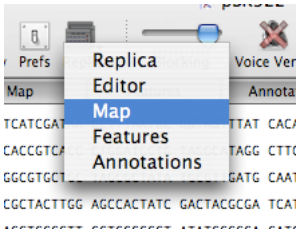
Using Replica Windows

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;



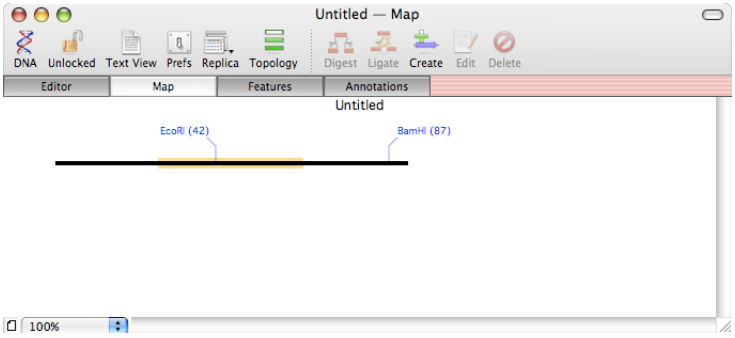
Note that you can choose which tab you want a replica window to open with by selecting the name of the tab in the drop down menu that appears when you click on the Replica button;



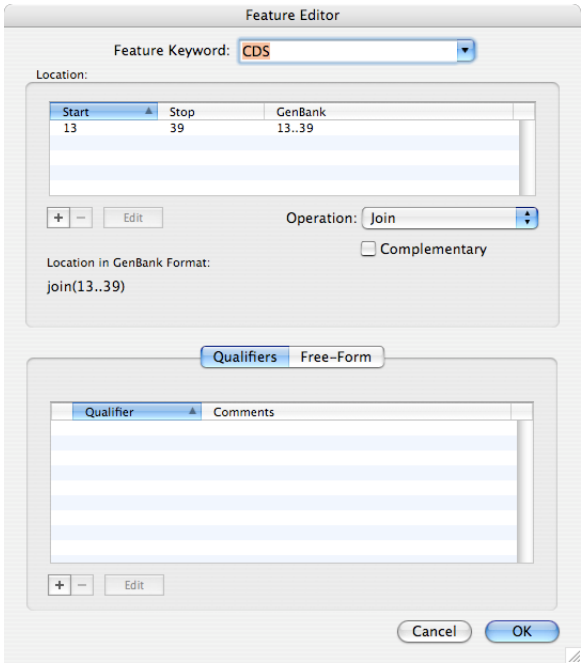
With the Editor tab window front most, continue typing in the window. Note how the Annotations tab window updates the Base Count in real time. Type the sequence "GAATTC" in the Editor tab window - note how the Map tab refreshes to show an *Eco RI* restriction enzyme site as you type the final "C" in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

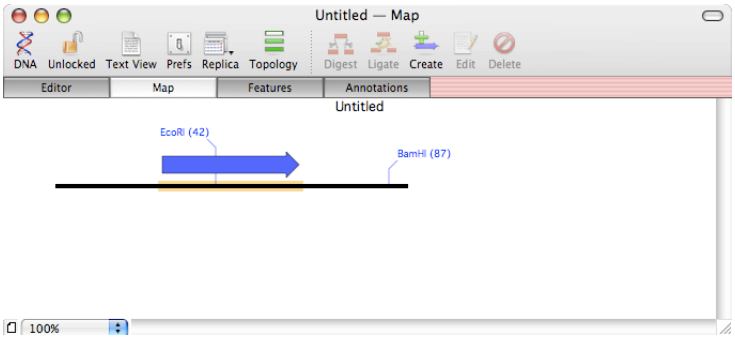


Lets create a new feature from this selection. Click on the Create toolbar button. The Feature Editor opens;

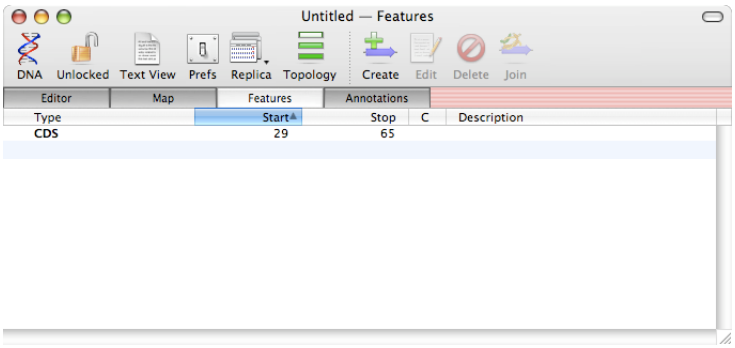


Note how the Start and Stop location of the selection are already filled out for you in the Location list. We'll look at the use of the Feature Editor in more detail later, but for now, choose `CDS` from the Feature Keyword drop down menu, then click the OK button.

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



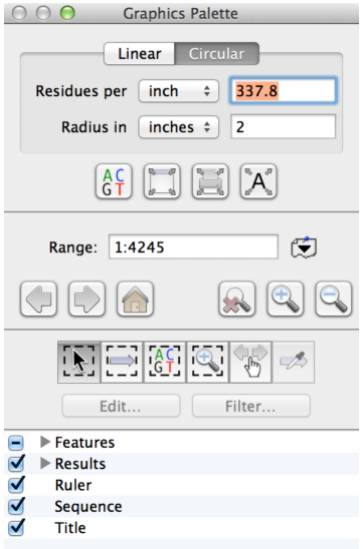
The Feature tab window also shows the new feature;



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.

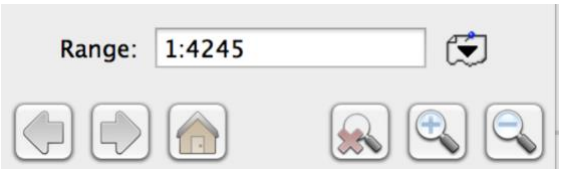
A row of 4 buttons provides "quick layout" functionality:

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.

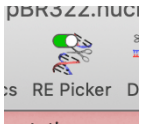
Magnify – in this mode a click magnifies the display 2-fold. Hold down the `<option>` key to reduce the magnification 2-fold.

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.

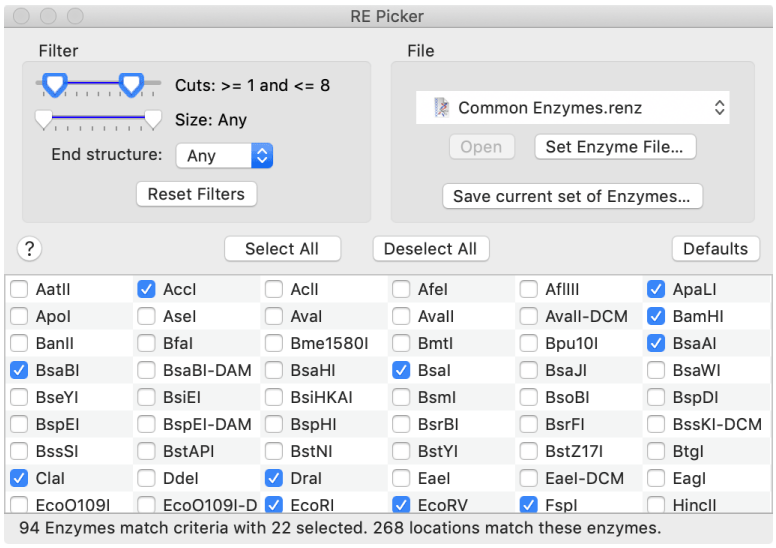
Copy Feature Appearance - this mode is only available if you have one or more features selected. Once selected, if you then click on a different feature, all the selected features will change appearance to match that feature

The Floating RE Picker

By default, there is floating window that displays all of the restriction enzymes available and lets you control which are displayed in the Map tab. It is only displayed when the Map tab is selected. You can quickly show/hide the window using the RE Picker toolbar item;



Whenever a DNA sequence is opened, MacVector automatically scans ALL of the restriction enzymes in the default file (set in the **MacVector | Preferences** Scan DNA tab) against the sequence, even if they are not selected in the source file. The RE Picker thus knows how many sites each enzyme has in the target sequence and provides controls to let you filter the displayed sites based on a number of parameters;



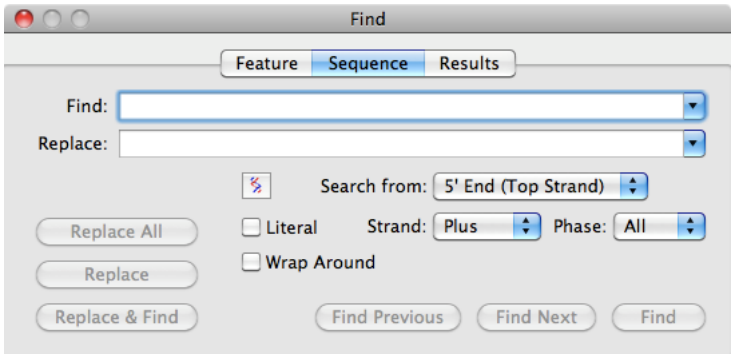
The Filter section lets you adjust the list based on the number of cuts in the target molecule (including non-cutters), the size of the site, or the end type of the cut. The File section lets you change the source file used for the analysis and even lets you save the currently displayed set of enzymes into a new file. In all cases, only the selected enzymes are actually displayed in the Map view.

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;

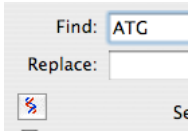
Searching Sequence Residues

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.

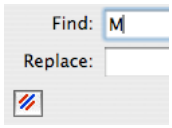


If you've used the Find function in previous versions of MacVector, this should be very familiar to you. You can search for matches in the sequence in a variety of ways. The default is to use the IUPAC codes to search for matching residues, so the sequence AGY will find AGT or AGC. If you actually want 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**.

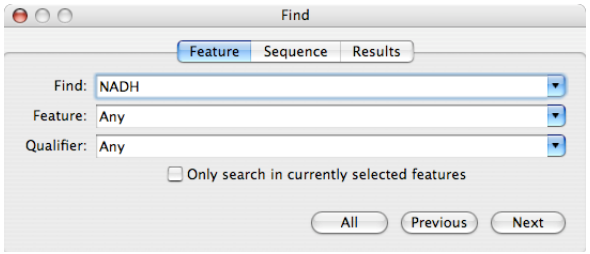
The same approach works in reverse so you can search Protein sequences for matches to DNA residues.

Searching Features

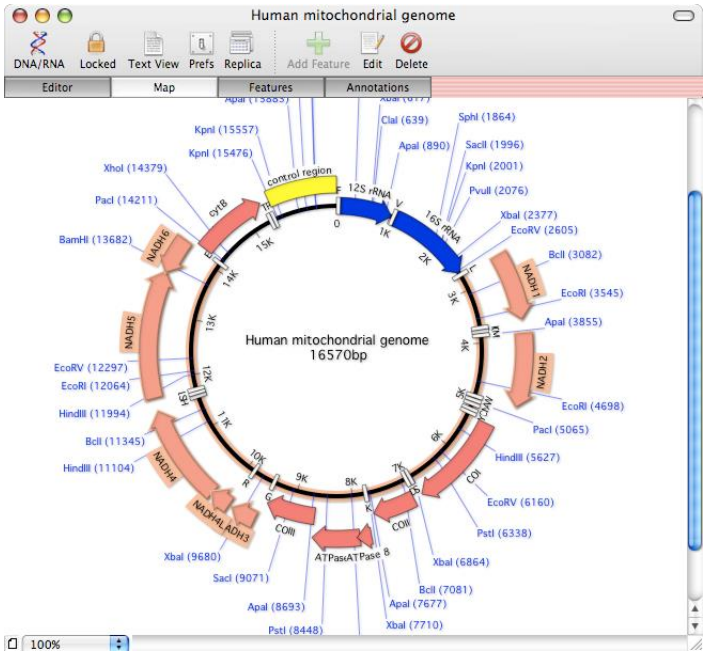
You can search for matching text within the features associated with a sequence.

Simple Search

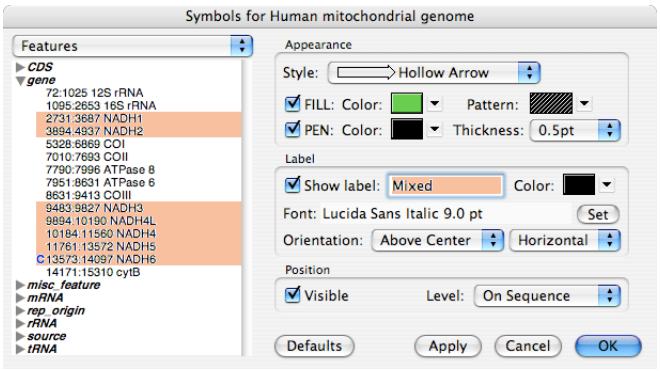
Switch to the Map tab of Human Mitochondrial Genome. Select **Edit | Find**. The Find dialog will be brought to the foreground with the Features tab selected;



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";



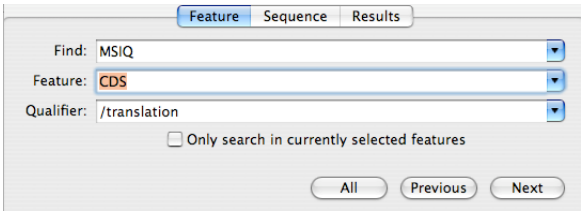
Once selected, you can double-click on one of the selected features and change the appearance of all of the features at one time;



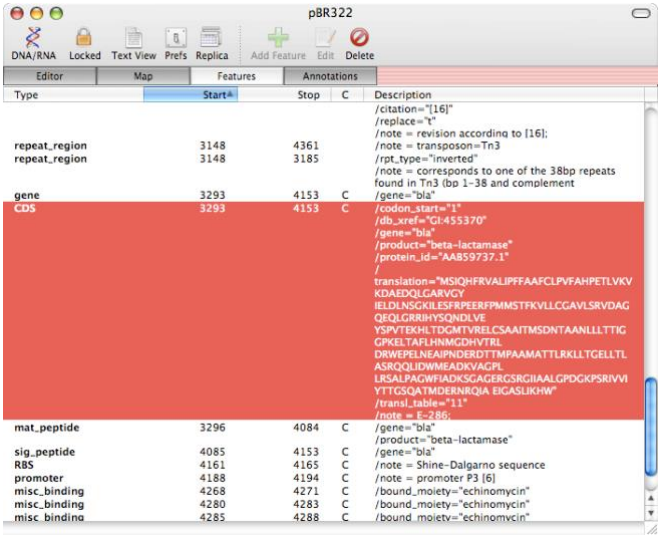
After clicking OK, the Map tab will update to change all of the selected features to the new settings.

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.

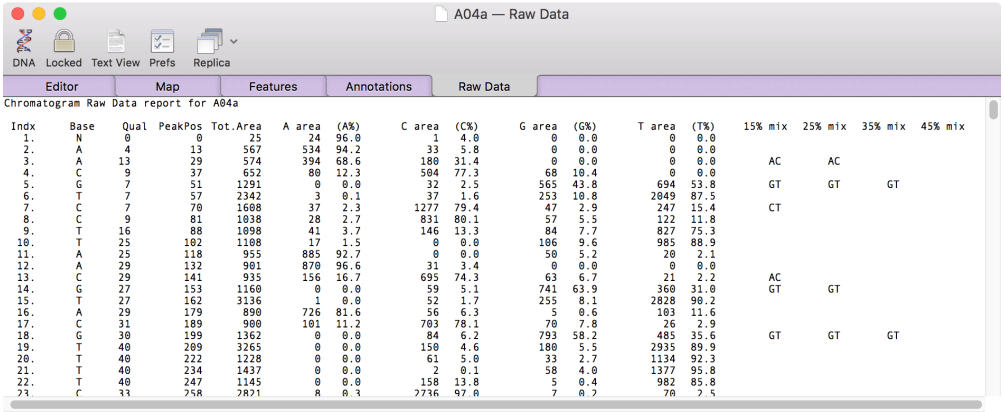
The Trace ("Chromatogram") Window

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.

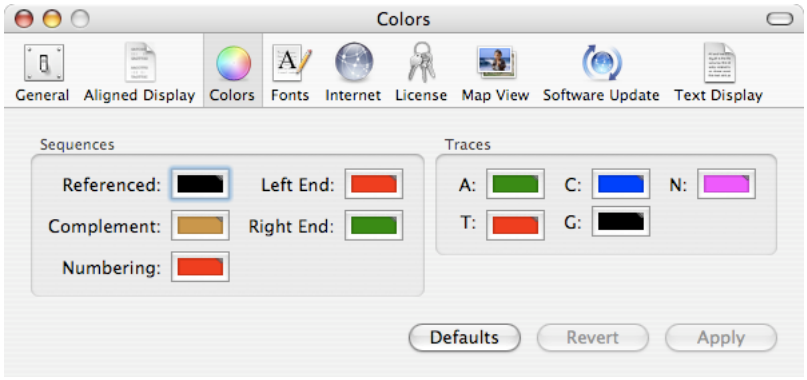
Most of the toolbar buttons have direct equivalents in the single sequence Editor tab. The only exceptions are the Width, Channels, Basecalls and Qualities buttons. The Width slider lets you widen or compress the traces so you can more closely examine adjacent peaks or zoom out for a better view of a 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.

Note that there is also a Raw Data tab. This lists a variety of numerical information about the trace data, including areas under the curve for each called residue and a rough estimate of the likelihood that peaks might represent heterozygotes.

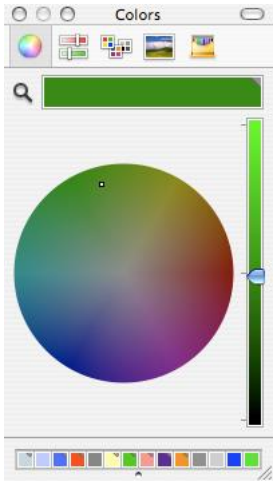


Trace Colors

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.



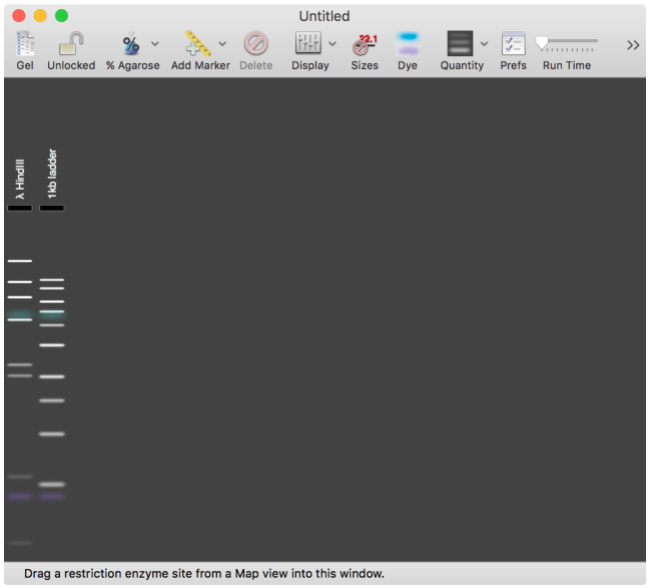
Click on one of the color buttons to bring up the standard system color selector that allows you to set the new color.



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 Agarose Gel Window

You can create simulated agarose gels by choosing the **File | New | Agarose Gel...** menu item.



Adding Digests

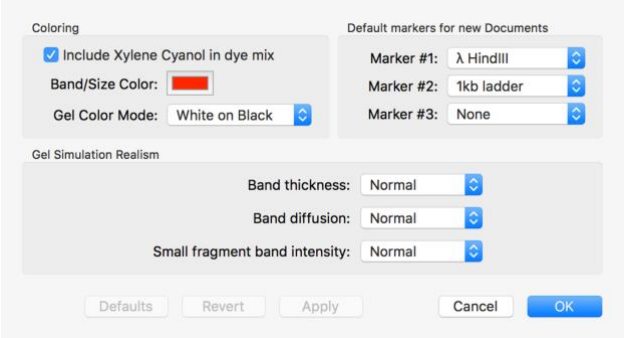
Simply drag any restriction enzyme label from any Map tab of a sequence window and drop it on the Agarose Gel window to see the predicted pattern for that molecule cut with that enzyme.

You can do this with any combination of enzymes to get double, triple or more digests.

You can click and drag on any gel track to rearrange the columns and resize the entire window to make the image larger.

Toolbar Buttons

- **Locked/Unlocked** – gels can be saved, so if you open one up, this button prevents you accidentally overwriting your data.
- **% Agarose** – choose a percentage between 0.5% and 1.5%.
- **Add Marker** – this lets you choose a marker to add to the gel, using a dropdown menu. You can also add your own specific markers to the list.
- **Delete** – removes the selected track(s) from the gel.
- **Display** – the default is to display a realistic image of the gel. However, you can choose a schematic representation, where all the bands are given equal intensities, or a combination of the two.
- **Sizes** – select this to view the sizes of each band on the gel. You can also simply mouse-over a band to see the size in the status bar along the bottom,
- **Dye** – turns on and off the bromophenol blue + xylene cyanol dye fronts.
- **Quantity** – use this to increase or decrease the amount of DNA loaded in the selected track(s)..
- **Prefs** – opens up the preferences dialog;



Here you can control the use of xylene cyanol in the dye mix, the color of the fonts, the default markers used in new gel documents and the realism of the gel. In addition, you can toggle the gel from the normal white on black to be black on white – very useful for printing!

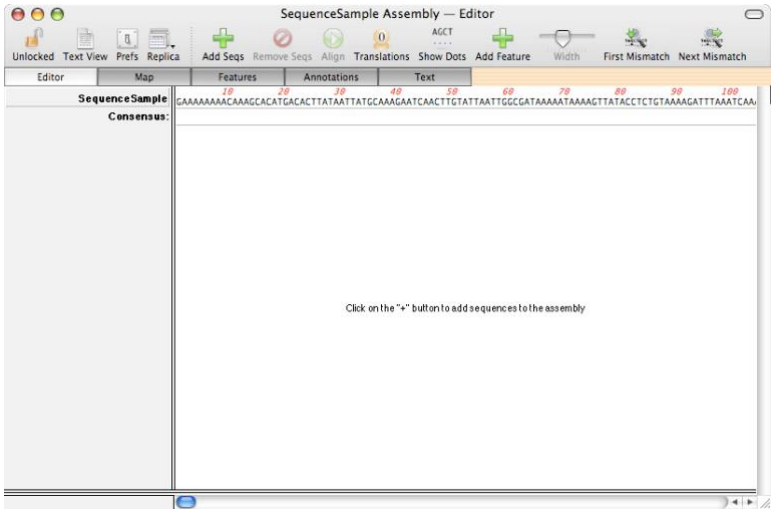
- **Run Time** – this lets you increase the run time to see better resolution of larger bands. Normally, the gel simulation is only run until the smallest band (or the dye front) is close to the bottom of the gel to ensure no bands are missing.

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`. You can also align cDNA files to a genomic sequence – you can use this functionality to quickly identify splice sites (described in more detail below).

Sequence Confirmation

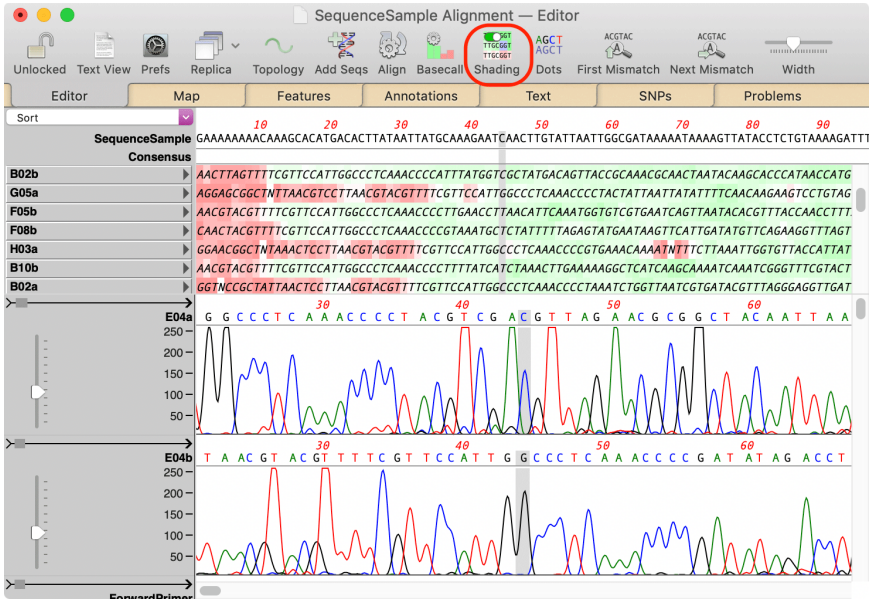
The Sequence Confirmation window also uses tabs. Open the file `/MacVector/Tutorial Files/Align To Folder/Sequence Confirmation/SampleSequence`. Choose **Analyze | Align to Reference**;



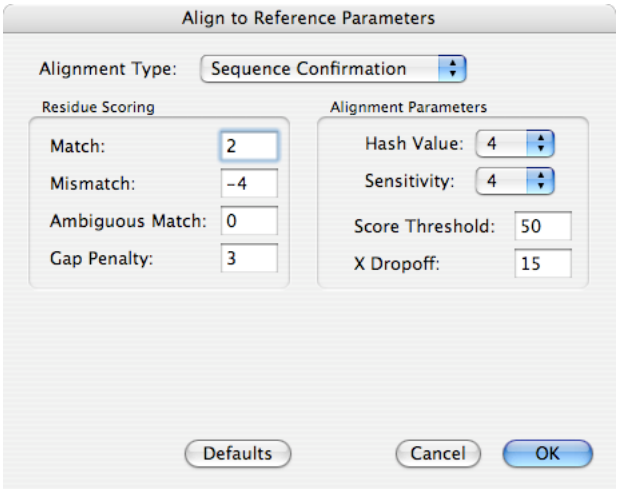
The Map, Features and Annotations tabs are essentially identical to the tabs used in the single sequence window. If you look at them now, you will see that they have exactly the same information as is present in the original SequenceSample window.

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. With the Shading button selected, the quality values of each residue are indicated by a red-green color scheme.



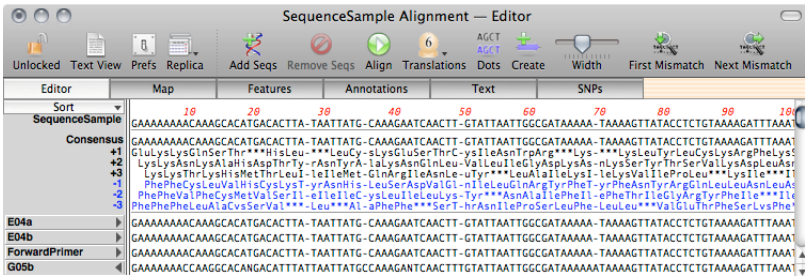
Now align the traces with the **SequenceSample** reference by clicking on the **Align** toolbar button. Make sure you have the Alignment Type set to **Sequence Confirmation**;



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

Toolbar Buttons

- **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.
- **Topology** – MacVector can align reads across the circular origin or sequences.
- **Align** - this displays the alignment dialog.
- **Basecall** – this runs the *phred* basecall tool which re-interprets the chromatogram peaks and adds quality values.
- **Shading** – turns the background quality shading on and off.
- **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.

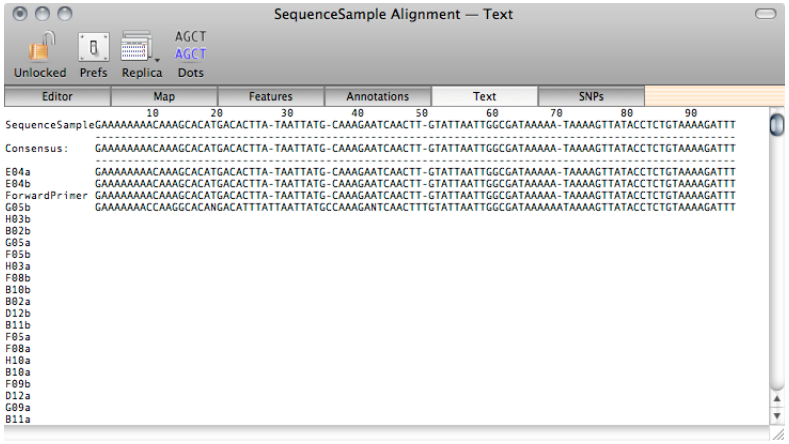
Note that the translations skip any gaps inserted into the reference sequence and treat it as a single ungapped sequence. The same is true for the automatic restriction enzyme searching in the Map tab -

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').

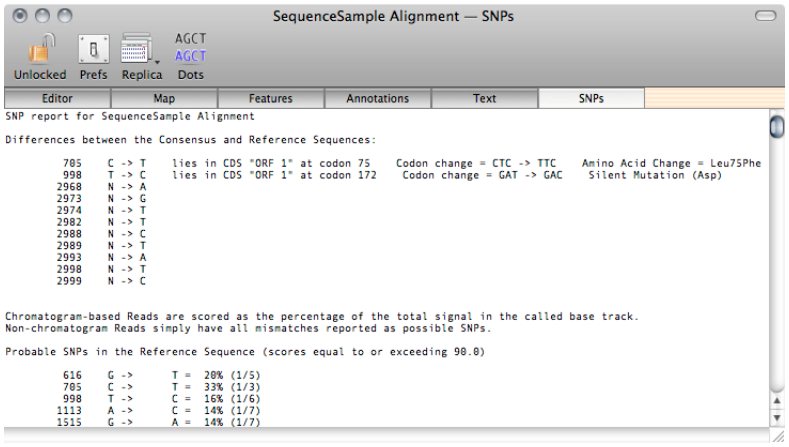
Text Tab



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.

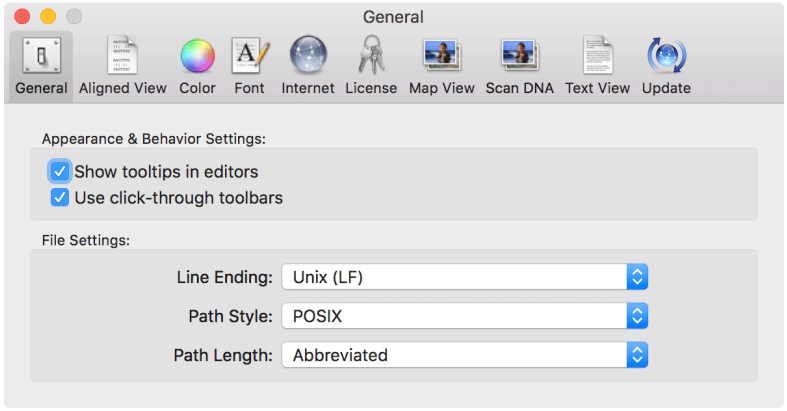
SNPs Tab

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.



Preferences Panel

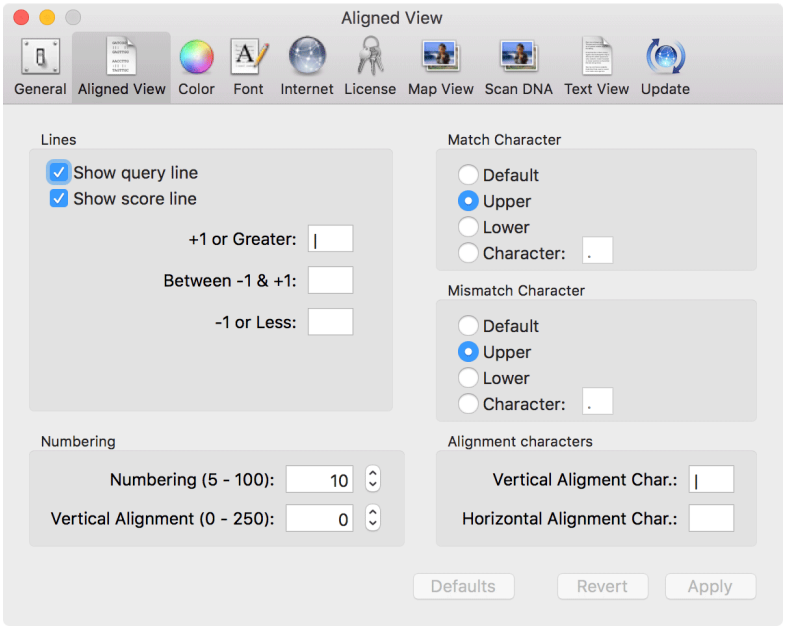
MacVector uses an OS X style Preferences Panel. This can be accessed by choosing the **MacVector | Preferences** menu item. Individual preference panes can be displayed by choosing the appropriate **Options** menu item or by clicking on the Prefs button in many of the tabbed views. The Preferences Panel 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.

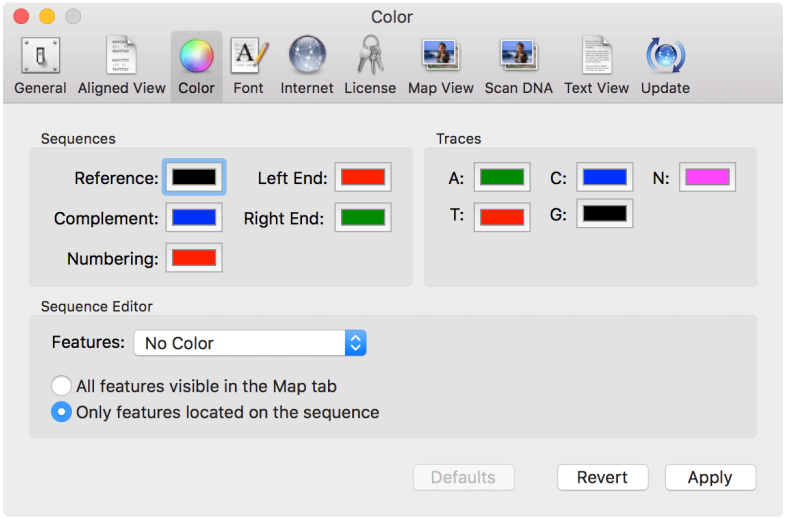
Aligned Display



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

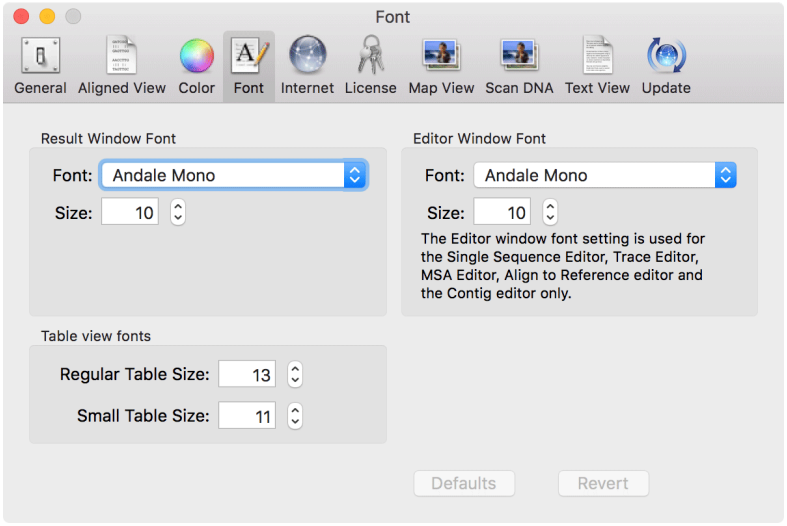
Colors

This is the pane 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.



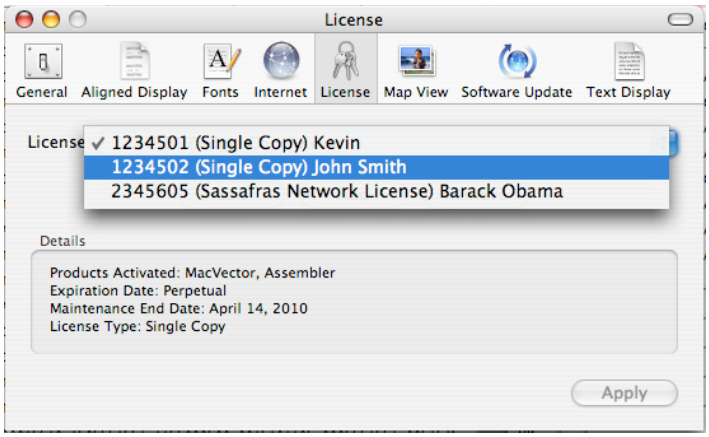
Note: If you are having problems with misalignments when you print text windows, change the font here to Andale Mono.

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.

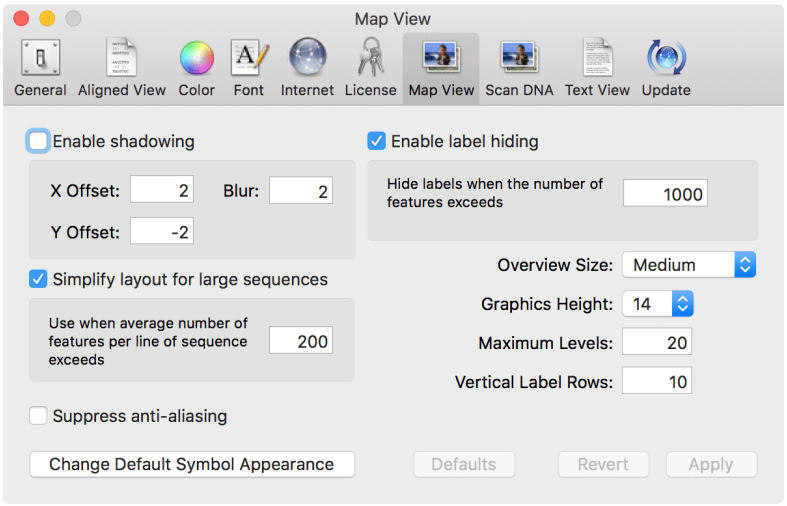
License

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.



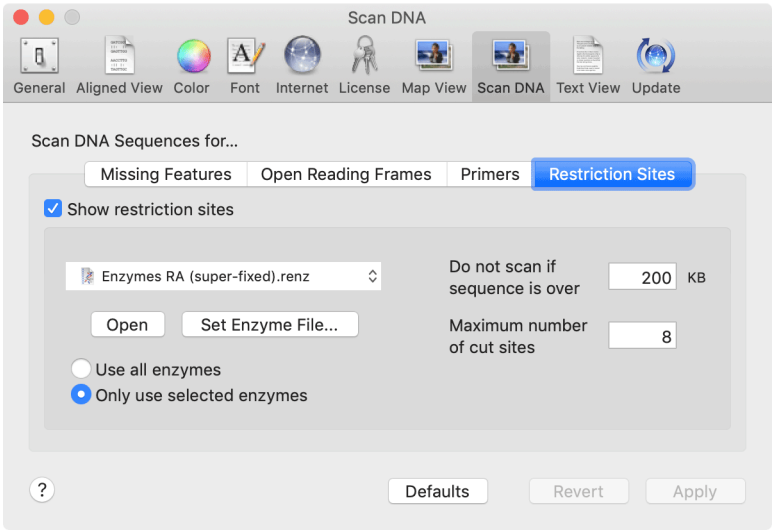
Map View

This pane lets you control all of the global preferences that affect the Map view. These include effects such as shadowing and anti-aliasing and controls that affect the display and layout of features and their labels for large sequences.

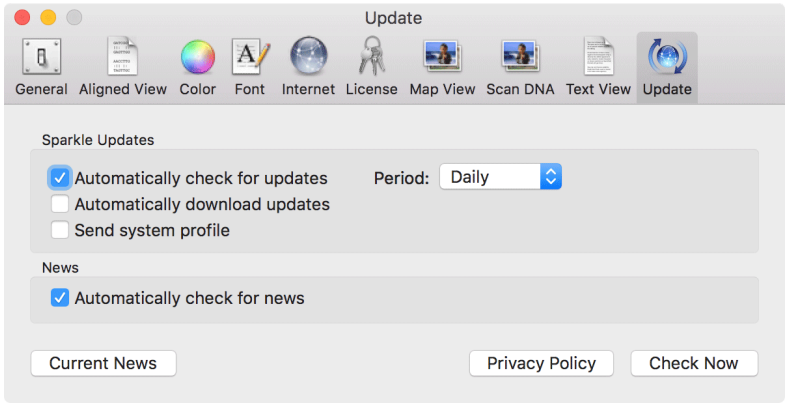


Scan DNA

This tab controls a variety of algorithms that are run against DNA sequences whenever you open a file, or change a sequence. The Restriction Sites tab controls how the restriction enzymes are displayed while the Open Reading Frames tab lets you automatically view and ORFs that are not already annotated as features on the sequence. The Missing Features tab scans the DNA for common features (replication origins, antibiotic resistance genes and other interesting features) that are typically found in cloning vectors. The Primers tab scans the sequence for matches to a default `Primer Database.nsub` file. Each of the algorithms can be turned off with a simple checkbox if you want an uncluttered display.

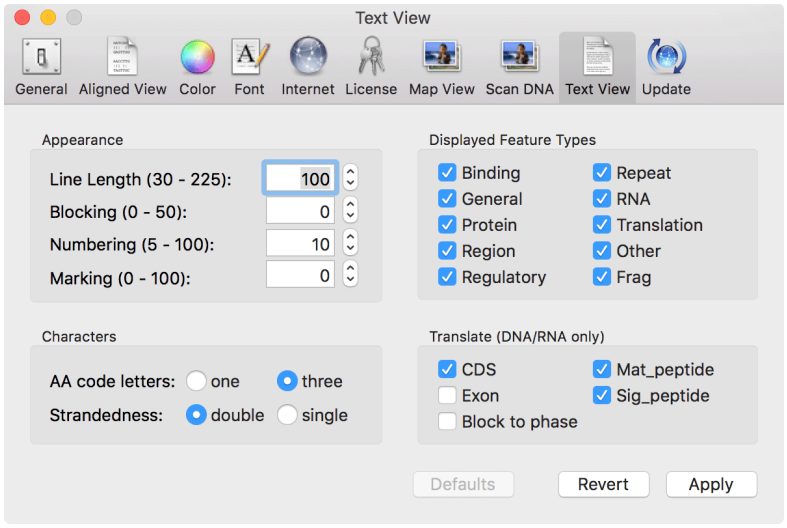


Update



MacVector 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 including workarounds for bugs, availability of new documentation, utility programs or data files.

Text View



This pane controls the appearance of the many plain text output tabs that are generated by MacVector's algorithms. In addition, the AA code letters are used throughout MacVector wherever amino acid sequences are displayed in the same view as DNA.