User Guide to

MegAlign Pro

DNASTAR, Inc. 2023

Table of Contents

Welcome to MegAlign Pro

MegAlign Pro is Lasergene's application for generating multiple and pairwise alignments of protein and/or nucleic acid sequences using your choice of alignment methods. Easily navigate between synchronized views with a click of a button, and rearrange views and toolbars by simply dragging and dropping them where you want them to be.

If you're not sure whether you need to do a sequence alignment or a sequence assembly, we recommend reading our "disambiguation" topic Sequence alignment vs. sequence assembly.

These support pages contain written help and tutorials, as well as video tutorials.

- For help **INSTALLING** Lasergene, see our separate Installation Guide.

- For written **TUTORIALS** with free data, see MegAlign Pro Tutorials.

- Click here for a list of all topics that include **VIDEOS**.

- To **PRINT** the current page of the User Guide, click the printer icon in the top right corner ( ).

- To download a **PDF** of the entire User Guide, scroll to the bottom of the table of contents on the left, and press **Download as PDF**.

Sequence alignment vs. sequence assembly

Sequence alignment and **sequence assembly** are very different workflows, but the terms are often used incorrectly. Adding to the confusion, both workflows can involve an optional "reference" sequence, and both can use some of the same sequence file types. This topic defines each workflow and recommends the best Lasergene application to use in different scenarios.

Are your sequences from different organisms or strains?

If so, you will want to perform **sequence alignment** with MegAlign Pro.

Sequence alignment is used to investigate the evolutionary relatedness of two or more organisms or strains, and the creation of a phylogenetic tree is often the end goal. MegAlign Pro lets you align DNA, RNA or protein sequences from a single gene up to an entire genome. The sequence files used in this workflow typically have formats like *.fasta*, *.seq*, *.pro*, *.genbank*.

Using a reference sequence in multiple sequence alignment is optional, but can be used to guide a MAFFT alignment.

Do your sequences consist of sequencing reads from one individual DNA sample?

If so, you will want to perform **sequence assembly** using SeqMan Ultra for Sanger trace data reads (e.g., *.abi*) and SeqMan NGen for all other read types (e.g., *.fasta*, *.fas*, *.fastq*). In this workflow, longer consensus sequences ("contigs") are assembled from short fragments of a single DNA sample.

Both SeqMan Ultra and SeqMan NGen allow you to assemble the reads *de novo* or by using a "reference" sequence as a template. Using a reference in sequence assembly allows the software to assemble DNA fragments into larger contigs faster and more accurately than assembling them *de novo*.

As with MegAlign Pro, the use of a reference sequence in sequence assembly also allows you to explore variants between the sequence reads/contigs and the reference sequence. Variants resulting from sequence assemblies in Lasergene are viewed using graphical views and tables in SeqMan Ultra and ArrayStar.

MegAlign Pro Tutorials

MegAlign Pro tutorials appear in the table of contents just under corresponding help topics. In addition, each tutorial is also listed below.

Miscellaneous tutorials:

- Add and align multi-segment sequences

- Map features

Multiple alignment tutorials:

- Perform a Clustal Omega alignment

- Perform a MUSCLE alignment with multi-segment sequences

- Perform a genomic alignment with Mauve

Pairwise alignment tutorials:

- Follow a multiple alignment with Global pairwise alignments

- Align transcripts to genes using Local and Global pairwise alignments

- Use Local pairwise alignment to find a gene within a genome

MegAlign Pro also plays an important role in the tutorial Phage display assembly, translation and prediction, located in the Protean 3D User Guide. This integrated workflow starts in SeqMan Ultra, continues through SeqNinja and MegAlign Pro, and ends in Protean 3D.

Create or Open a Project

Before you can add sequences to a project and align those sequences, you first need to create a MegAlign Pro project. If you already have a project from MegAlign Pro (or its predecessor, MegAlign), you can also open that to view or make changes. For details about each option, click the links below.

- Open an existing MegAlign Pro project (*.msa*)

- Create a new MegAlign Pro project

- Import an alignment from another application

- Open a Newick-format tree file,

If you create or open multiple projects, see Navigate between projects to learn how to switch between them.

Open an existing MegAlign Pro project

To open an existing MegAlign Pro project (*.msa*):

Do any of the following:

- From the **Welcome** tab, click **Open Project** on the left. Then use the options on the right to **Open alignment project** or open a recent document.
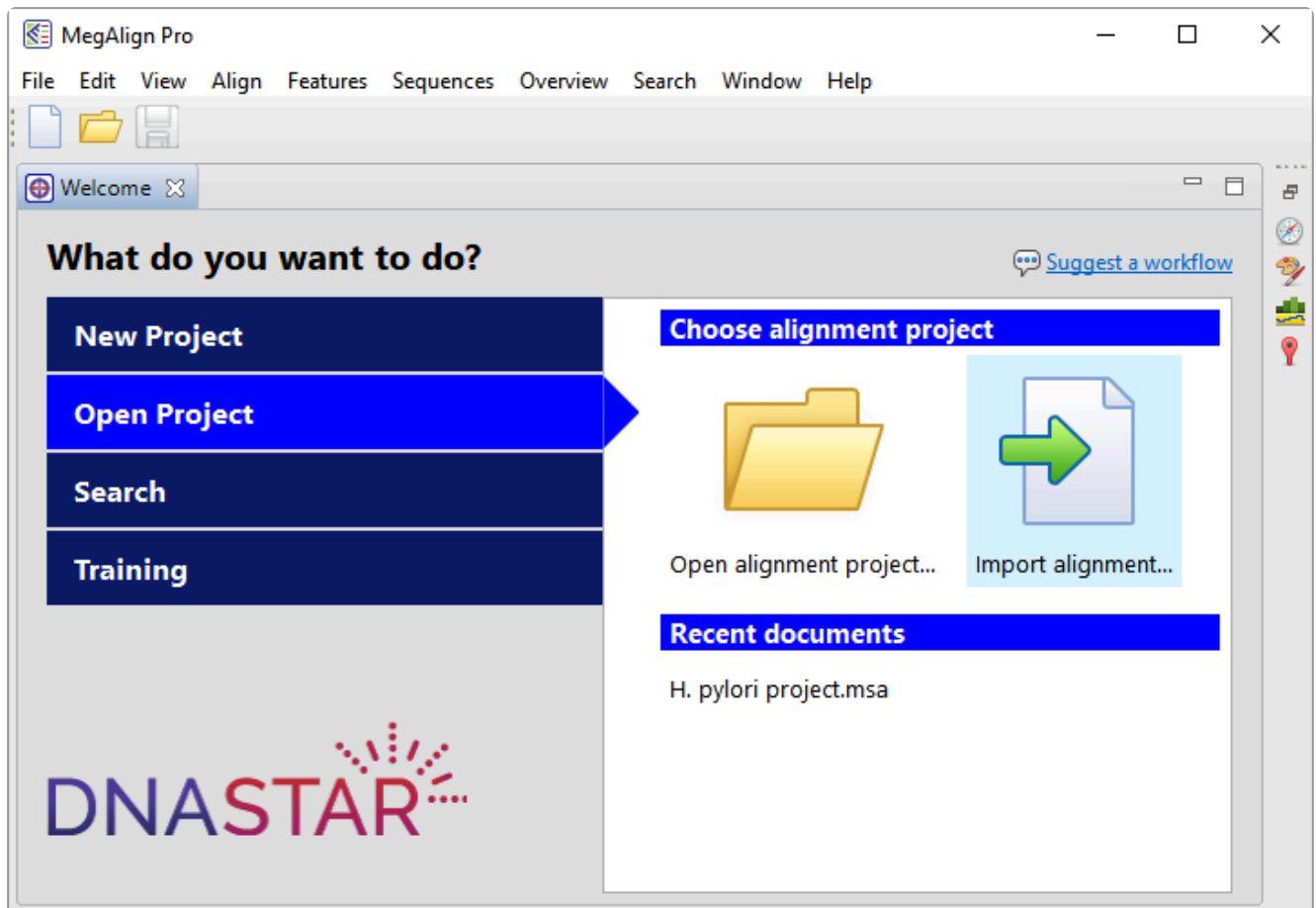
- Drag an *.msa* file from the file explorer or desktop and drop it on an empty MegAlign Pro project window or tab.

- Choose **File > Open** from the menu.

- Click the **Open a file** tool ( ).

- Press **Ctrl/Cmd+O**.

> ✳ **Note:** If you moved or renamed any of the sequences prior to reopening a project, you may

be prompted to supply the new name(s) or navigate to the location(s) of those sequences.

The following video (which launches, by design, at the 3:05 minute mark) demonstrates how to open existing MegAlign Pro and classic MegAlign projects.

Create a new MegAlign Pro project

To create a new MegAlign Pro project:

Do any of the following:

- From the **Welcome** tab, click **New Project** on the left. Then choose one of the options on the right:
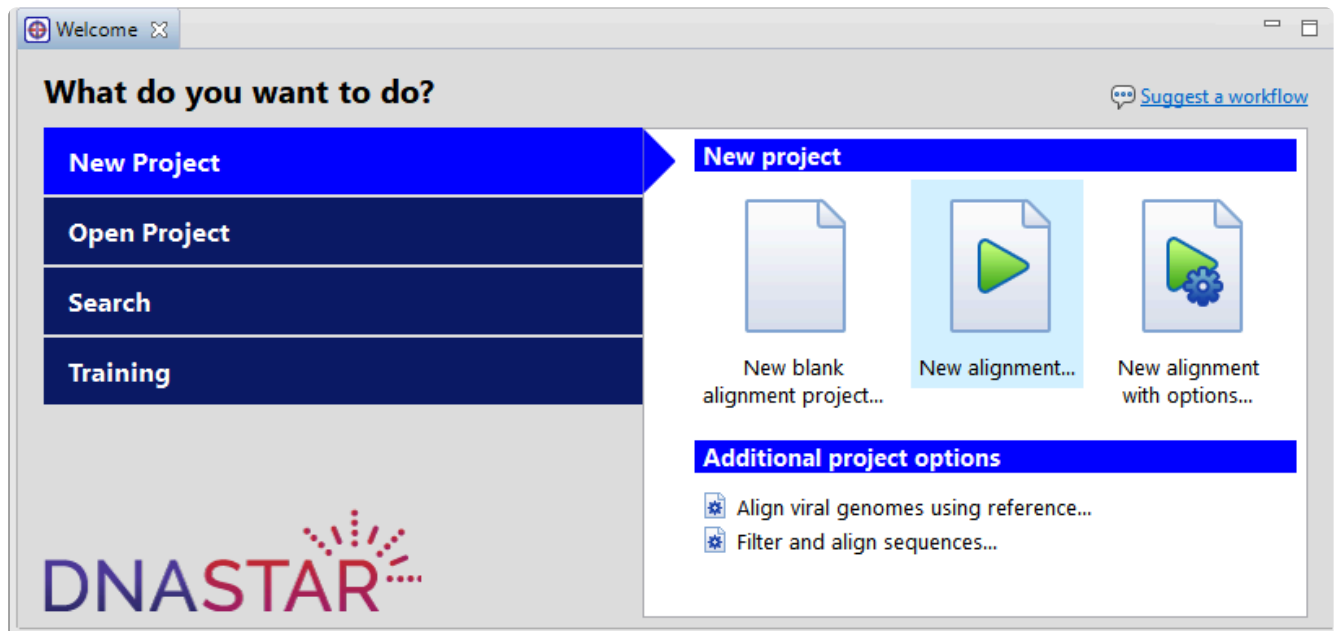
- ◦ **New blank alignment project** – Opens an empty project window.

- ◦ **New alignment** – Launches a file chooser. Choose files and press **Open** to create a new MegAlign Pro project with the default alignment already in progress.

- ◦ **New alignment with options** – Launches a file chooser. Choose files and press **Open** to create a new MegAlign Pro project with the Align Sequences dialog open. Specify the alignment algorithm, change other settings as desired, then press **Align** to begin the alignment.

- ◦ **Align viral genomes using reference** – Launches a file chooser. Choose files and press **Open** to create a new MegAlign Pro project and open the Align Sequences dialog for the MAFFT multiple alignment algorithm. Use the **Reference Sequence** drop-down menu to specify the sequence to use as the reference and change any other desired settings. Press **Align** to begin the alignment.

- ◦ **Filter and align sequences** – Creates an initially blank project, prompts you to select sequences from a file chooser, and then opens the Filter Sequences dialog. Set up any desired filters and press the **Filter** button. See Add only sequences that meet filtering thresholds for instructions. Only the sequences that meet the specified criteria will be added to the project.

- Choose **File > New Alignment Project** from the menu. Three options are available:

  - **Blank Document** – Creates an empty document, regardless of whether any sequences or subsequences are currently selected. See Add ungapped sequences to a project for an alternative method.

  - **With Selected Sequences** – Creates a new document containing the currently selected sequences. If no sequences are selected, this option is disabled.

  - **With Selected Subsequences** – Creates a new document containing the currently selected subsequences. If no subsequences are selected, this option is disabled.

- To create a new project and add sequences to it simultaneously, drag sequence files from a file explorer and drop them on the project tab row of the MegAlign Pro window.

- Click the **Create a new alignment project** tool (   ).

- Press **Ctrl/Cmd+N**.

If you elected to open a blank project, see Sequences to learn how to add sequences to it.

Import an alignment from another application

To import an alignment from another applicaton:

1. Do either of the following:

    • From the **Welcome** tab, click **Open Project** on the left. Then use the option **Import alignment** on the right.

    • Choose **File > Import Alignment**.

2. Navigate to the file you wish to open. A wide variety of formats are supported, including gapped FASTA (*.fasta*, *.fas*, *.fap*, etc.), Geneious alignment (*.geneious*), MEGA alignment file (*.meg*), MegAlign file (*.meg*), MSF file (*.msf*), Vector NTI alignment project file (*.apr*) formats.

3. Click **Open**.

> ✳ **Note:** If you attempt to open a file that is not an alignment file, MegAlign Pro will instead attempt to import any sequences from that file.

The following video (which launches, by design, at the 3:05 minute mark) demonstrates how to open existing MegAlign Pro and classic MegAlign projects.

Navigate between projects

You can open multiple projects within a single instance of MegAlign Pro. Each project is displayed in a separate tab and is initially assigned a unique name (*Untitled[#].msa*).

To make a project the active project:

Click on its project tab.

To open an additional MegAlign Pro window:

Select the **Window > New Window** command.

To navigate between open projects:

One instance of MegAlign Pro may include multiple open projects. Each window is displayed in a separate tab.

One tab, the Welcome screen, is always present by default unless it has been manually closed. To make another project active, click on its tab.

If you cannot see all the window tabs in a project, press the **>>** icon that appears to the right of any visible window tab(s). This icon is displayed when multiple view windows are open and the main MegAlign Pro window is too small to display all the tabs. Click on the icon to open a yellow popup "window chooser" with a line for typing in text at the top, and a list of open projects below.
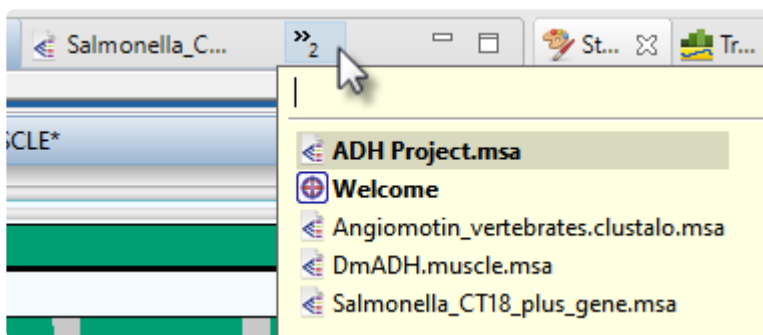
Use your mouse to click on the window you want to make active. If you do not have a mouse, make a selection using the keyboard arrow keys, then press **Enter**. To instead search for a project using a filter, type text into the text field and press **Enter**. The text field supports the "?" and "*" wildcards.

MegAlign Pro Interface

The following table describes components of the MegAlign Pro user interface, along with links to more in-depth information about each item:

Item	Description
Menus	The MegAlign Pro menus are: **MegAlign Pro** (available only on Macintosh), **File**, **Edit**, **View**, **Align**, **Features**, **Sequences**, **Overview**, **Distance**, **Tree**, **Window** and **Help**. Some menu commands can also be initiated by keyboard shortcuts, tools, and/or context menu (right-click) options.
Tools	The MegAlign Pro toolbar is situated just below the menus. To toggle between showing/hiding the toolbar, choose **View > Show (Hide) Toolbar**.
Projects	In MegAlign Pro, each project is displayed in a separate tab with an *.msa* extension.
Overview	The Overview provides a way to navigate within the sequences, no matter what type of sequences or alignment is used. In the case of Mauve-generated alignments consisting of multiple alignment blocks, the Overview also allows you to explore the relationships between aligned blocks.
Sequences view	The Sequences view shows the sequences and any enabled tracks (see Work with Tracks).

Distance view	The Distance view, available only after an alignment has been performed, shows a matrix (i.e., table) of numbers representing distances between each pair of sequences for the alignment block selected in the Overview. Sequence distances are used as input in creating the Tree View. In the case of multi-block alignments, each alignment block will have a separate matrix.
Tree view	The Tree view, available only after an alignment has been performed, shows evolutionary relationships predicted from the multiple sequence alignment, and is calculated using the current Distance metric. In the case of multi-block alignments, each alignment block will have a separate tree.
Panels	The Style, Tracks, Features, Places and Details panels allow you to specify how to display items in the views, or show details about items in the views.
View splitter	The view splitter is used to divide the Sequences view or Overview vertically into multiple panes.
Header/ footer	A header and footer are collapsible sections above and below the selected reference sequence. They contain tracks that apply to all those sequences and their alignment, if any. The Overview provides a header, and the Sequences view provides a header and footer.
Sequence ruler	Sequence rulers pertain only to the sequences just above them, and use pre-alignment (ungapped) numbering. Gaps are not counted in this ruler.
Show/hide icons	The gray circles to the left of each sequence name in the Sequences view and Overview contain a plus or minus sign. These toggle between displaying and hiding the detail tracks.
Console	The Console panel displays information about an alignment in progress, or a log for a completed alignment.
Console toolbar	This toolbar is located in the upper right of the Console and pertains only to the Console.

Welcome screen

The Welcome "project" opens when you launch MegAlign Pro and is a central location for opening projects, searching NCBI databases, getting help, and much more. As you add projects, the Welcome "project" will remain as the leftmost tab unless you manually move or close it. We recommend leaving the Welcome tab in place, since it has many time-saving shortcuts to useful commands.
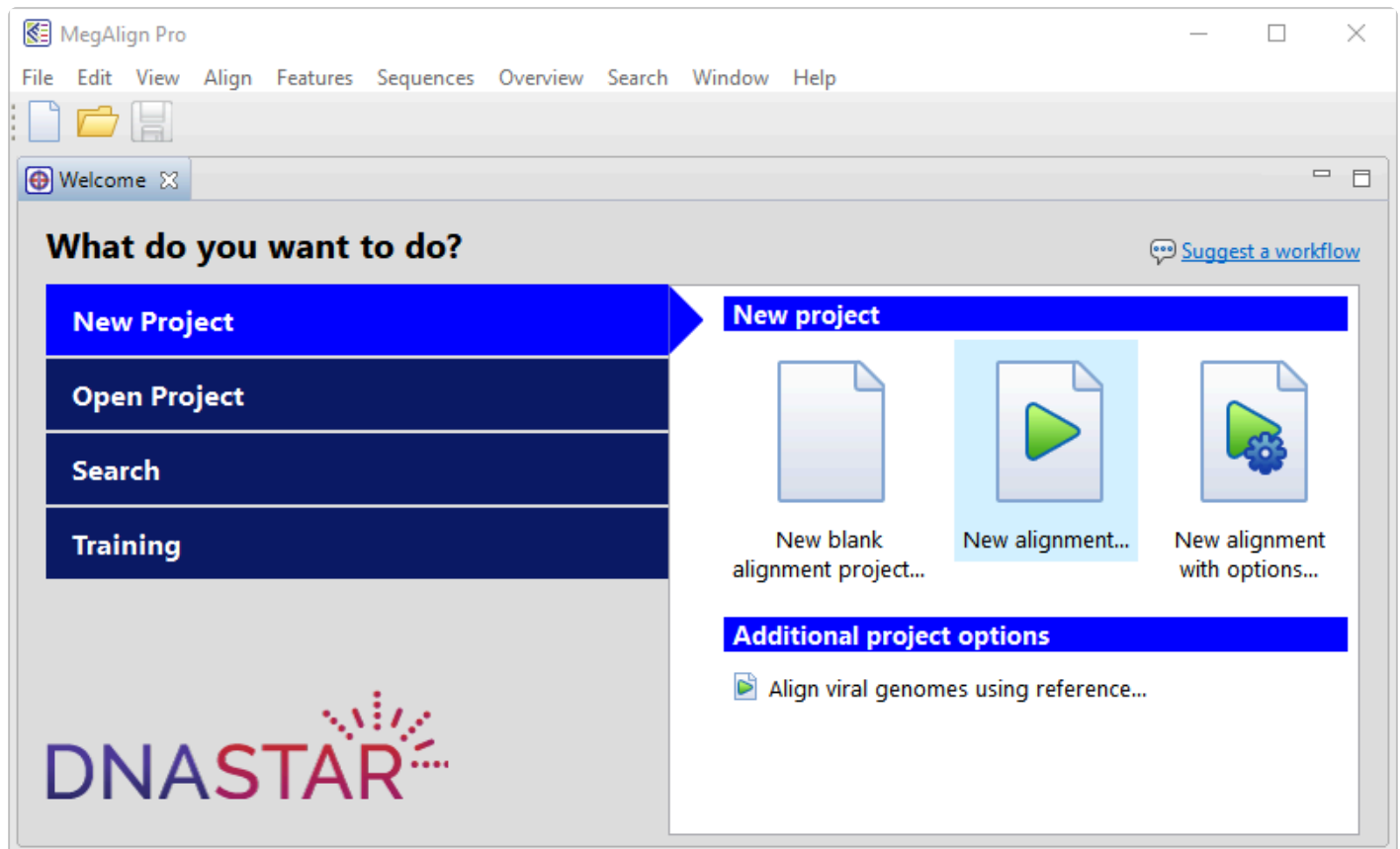
Tab on left	Option on right	Shortcut to this procedure:
New project	**New blank alignment project**	**File > New Alignment Project > Blank Document**. (see [Create or Open a Project](#)).
	New alignment	**File > New Alignment Project > Blank Document** plus **File > Add Sequences** plus **Align** using default method and options. (see [Create or Open a Project](#)).
	New alignment with options	**File > New Alignment Project > Blank Document** plus **File > Add Sequences** plus **Align with Options** (allows you to choose the method and edit options). (see [Create or Open a Project](#)).
	Align viral	**File > New Alignment Project > Blank Document** plus **File > Add Sequences**

	genomes using reference	plus **Align with Options** set specifically to use the MAFFT alignment algorithm. (see Create or Open a Project).
Open project	**Open alignment project**	**File > Open** (see Create or Open a Project).
	Import alignment	**File > Import Alignment** (see Create or Open a Project).
	Recent documents	**File > Recent Documents**. (see Create or Open a Project).
Search	**Search for sequences**	**Search > Search for Sequences** (see Search for sequence matches in BLAST).
	Search for text in NCBI nucleotide	**Search > Search for text in NCBI nucleotide** (see Search for text or accession number matches in Entrez).
	Search for text in NCBI protein	**Search > Search for text in NCBI protein** (see Search for text or accession number matches in Entrez).
Training	**User guide**	**Help > User Guide**. Opens this User Guide.
	Tutorials	Opens this User Guide to the MegAlign Pro Tutorials topic.

Views

The left side of the MegAlign Pro window is comprised of one or more "views." To learn more about the types of views and how to change their appearances, click the links below.

View types:

- Overview
- Sequences view
- Pairwise view
- Distance view
- Tree view
- Text view
- Table view

To learn how to select, display and customize views:

See Customize the look and layout.

To access a different instance of a view:

Some DNASTAR applications, including MegAlign Pro, allow multiple instances of a view to be created. For instance, if you select a pair of sequences and use **Align > Align Pairwise** — and repeat that three times — MegAlign Pro will create three separate Pairwise views. As shown in the image below, the view tab shows how many views of that type are in that tab folder.

To display a different instance:

1. Hover over the view tab until you see a popup showing each instance by name.

2. Click on the desired instance.
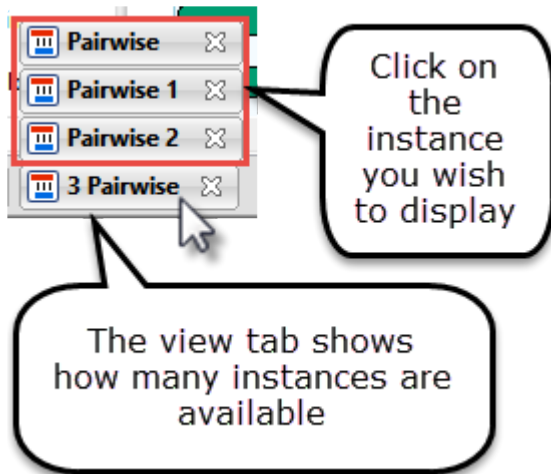
Click on the instance you wish to display

The view tab shows how many instances are available

Overview

The Overview provides a way to navigate within the sequences, no matter what type of sequences or alignment is used. In the case of genomic alignments generated by [Mauve](#), the Overview also allows you to explore the relationships between multiple aligned blocks. Selections and vertical ordering within this view are synchronized with other views in MegAlign Pro. Similarly, selections and ordering done elsewhere are reflected in the Overview.

To access the view, click on the **Overview** tab or choose **View > Overview > Show**.
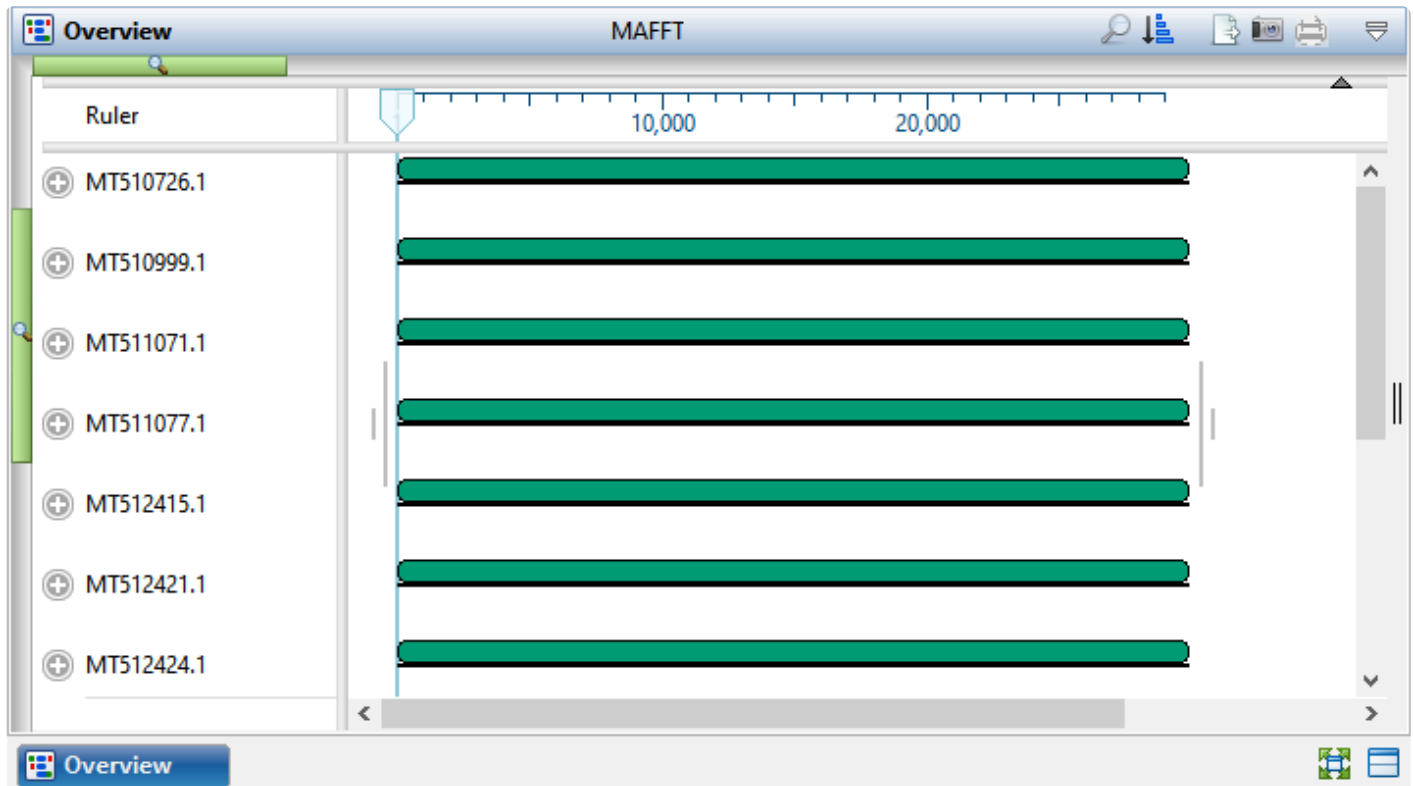
By default, the Overview displays a header, followed by one or more sequences with labels to their left. Each sequence is depicted as a "sequence/alignment block," and the length is proportional to other sequences in the group. Each sequence row can display a variety of optional data tracks.

The Overview displays information in different ways, depending on the alignment stage and the workflow. In all cases, sequences are represented as either gray or colored blocks. Any sequence that is not part of an alignment is shown in solid gray. Sequences included in an alignment are usually drawn in dark green, although multi-block alignments, such as those generated by Mauve, are drawn in up to seven colors.

See the following topics for images and details of the Overview in different situations:

- [Prior to performing an alignment](#)

- [After aligning multiple genomes with Mauve](#)

- [After performing a non-Mauve alignment or a Mauve non-genome alignment](#)

- [After performing an alignment and entering additional sequences](#)

Information bar:

The blue/gray bar at the top of the view shows information pertaining to the current selection.

Selection	Example	Information displayed
Nothing, and sequences have not been aligned	(blank)	(blank)
Nothing, and at least some sequences have been aligned	Clustal Omega	The name of the alignment.
One sequence	isoH-NM_001032097.2 = 2,024	The sequence name and its length in bp.
Multiple sequences	2 of 4 aligned sequences 2 of 5 unaligned sequences	The number of sequences selected, the total number of sequences, and the alignment status.
A feature or any portion of one sequence	isoF-NM_001032098.2721>1257 = 537	The sequence name and selected range.
A portion of multiple sequences (selected in the Sequences view)	54>342 = 289	The selected range and length in bp.

Tasks pertaining to the Overview:

Task	How to…
Change settings for this view	To change settings, including showing/hiding connecting lines and changing spacing between rows, use the [Overview section](#) of the [Style panel](#).
Customize the look and layout of the view, zoom in or split the view	See [Customize the look and layout](#).
Set an unselected alignment block in the Overview as the active block	Double-click on an inactive alignment block, or right-click on an inactive alignment block and choose **Set As Active Block**. The corresponding sequence blocks will line up vertically. The alignment block containing the sequence block will become active.

Change settings related to the Overview	use the Overview section of the Style panel.
Make an alignment block the active block and move it to the left of the Overview	Make a block active by double-clicking a sequence block. Then choose **Overview > Set Start of Sequences to Active Block** or the context (right-click) option of the same name. The blocks in each sequence are rotated so that the active sequence block is leftmost.
Center the Overview display over the active block	Choose **Overview > Show Active Block**.
Return alignment blocks to their default order	Choose **Overview > Reset Start of Sequences** or the context (right-click) option of the same name. In the Overview, the individual sequence ruler coordinates are updated to reflect the change.Note that resetting the start of sequences presumes that alignment consists of circular genomic sequences.
Search for a sequence by name	Click the **Search** tool ( ) on the right of the header area to open a search area. See Search for a sequence using a text query for details.
Change the order of sequences in this view	Use the **Sort sequences** tool ( ). See Change the order of sequences.
Informally compare other sequences in a Mauve alignment to one particular sequence	See Visually compare other sequences to one sequence.
Export data from the view	Click the **Export Data** tool ( ), a shortcut to the **File > Export Data** subcommands. See Export data to a file.
Save an image of the view	Click the **Export Image** tool ( ), a shortcut to the **File > Export Image** subcommands. See Export an image of the view.
Print the view	Click the **Print** tool ( ). See Print Images for details.

Overview prior to performing an alignment

The Overview has different appearances, depending on the stage and workflow.

Prior to alignment, each sequence in the Overview appears as a green rectangle ("sequence block") with no gaps. The length of each block is proportional to the ungapped length of its corresponding sequence.
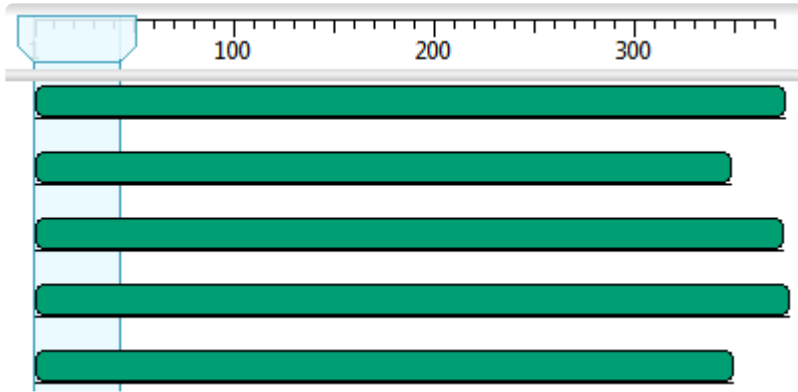
Overview after aligning multiple genomes with Mauve

The Overview has different appearances, depending on the stage and workflow.

A Mauve genome alignment results in a set of locally colinear blocks (LCBs), each of which is a conserved region across multiple sequences. In MegAlign Pro, these may be referred to as "blocks" or "alignment blocks". The Overview displays the part of a sequence in an alignment block as a "sequence block." The sequence blocks in the same alignment block have the same color, and by default are shown connected by lines.
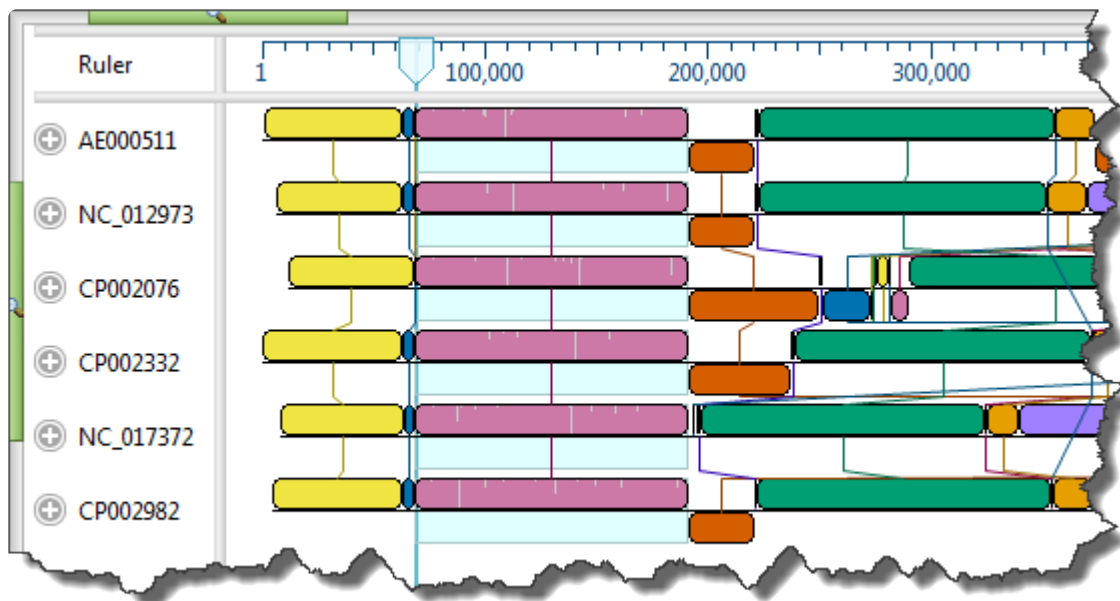
To orient the alignment around an active block, double-click on the sequence block. If necessary, individual sequences flip (i.e., are reverse complemented) to orient properly for the new alignment. Sequence rulers reorient along with the strands, and display the new coordinates. Gaps in the active block are shown by the presence of gray bars.

- Only one block is active at a time, and the current block for all the sequences blocks for are always shown in the forward orientation (i.e., on the top strand).

- Sequence blocks are usually displayed only if they include at least two sequences, though an exception is made when there is only one sequence in the project.

- Some sequences may not be included in the active block.

- In order to keep the current blocks in the same orientation, it may be necessary to reverse complement one or more entire genomes.

- The interior of the active block (only) displays a histogram of residue count, with gray areas representing gaps. These histograms reflect the proportion of non-gapped positions in a window corresponding to the number of positions spanned by a single pixel at a given zoom level.

- Blocks appearing out of order compared to other rows can signify an inversion. Blocks in reverse order compared to other rows can signify that the sequence is from the opposite strand compared to the other rows.

- If an alignment block is "hanging down" compared to other blocks in the same track, this signifies that the feature is encoded on the opposite strand. Having blocks represented on the lower strands are the result of rearrangements, such as inversions, and are not meant to imply that the alignment is double-stranded.

- The contents of the Distance and Tree views are specific to the active block only.

After aligning sequences end-to-end, vertical blue bars are used to show the beginnings and ends of the component sequences. These bars are displayed regardless of the type of alignment performed.
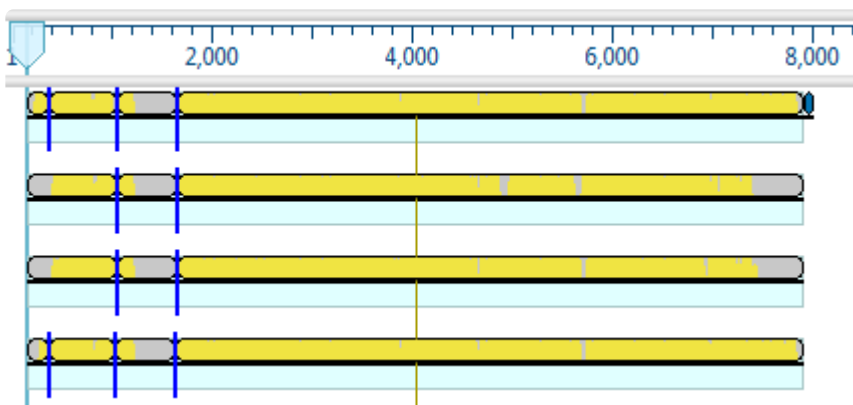
Overview after other alignments

The Overview has different appearances, depending on the stage and workflow.

After aligning with a method other than Mauve, or after performing a Mauve alignment that results in a single aligned block, any blocks in the Overview will be green and will have the same lengths.

The interior of a sequence block displays a histogram of residue count, with gray areas representing gaps. These histograms reflect the proportion of non-gapped positions in a window corresponding to the number of positions spanned by a single pixel at a given zoom level.
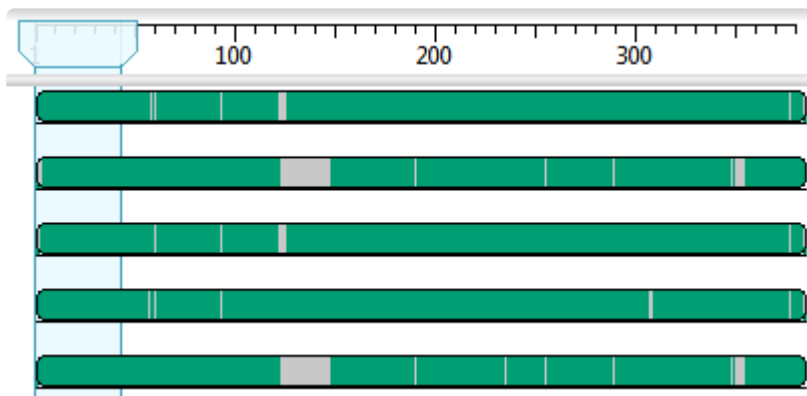
After aligning sequences end-to-end (see Add Multi-Segment Sequences), vertical blue bars are used to show the beginnings and ends of the component sequences (i.e., the segments). These bars are displayed regardless of the type of alignment performed.
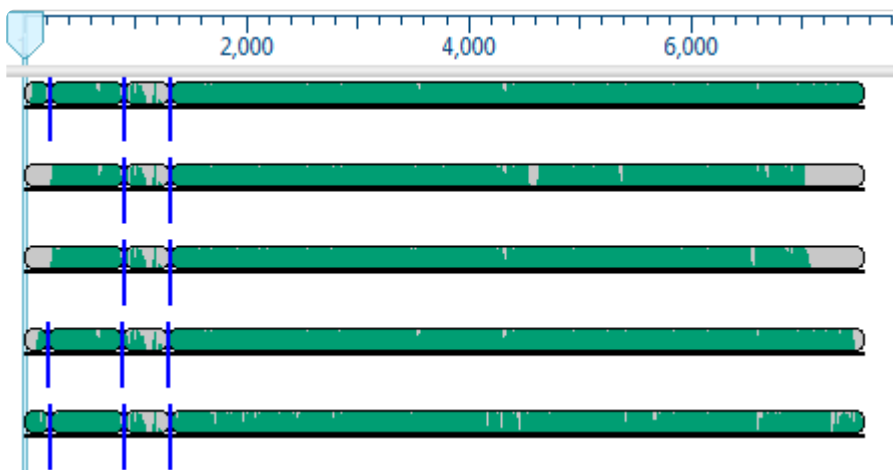
Overview after aligning and entering additional sequences

The Overview has different appearances, depending on the stage and workflow. In all cases, sequences are represented as either gray or colored blocks. Any sequence that is not part of an alignment is shown in solid gray.

After performing an alignment, any new sequences added to the project are shown as gray blocks under a bar with the words "Unaligned Sequences." The bar displays the number of unaligned sequences in parentheses, and a moveable pane divider separates the unaligned sequences from the alignment. The unaligned sequences can be shown/hidden by clicking the Unaligned Sequences bar. The Unaligned Sequences bar is not displayed if there is only one sequence in the project.
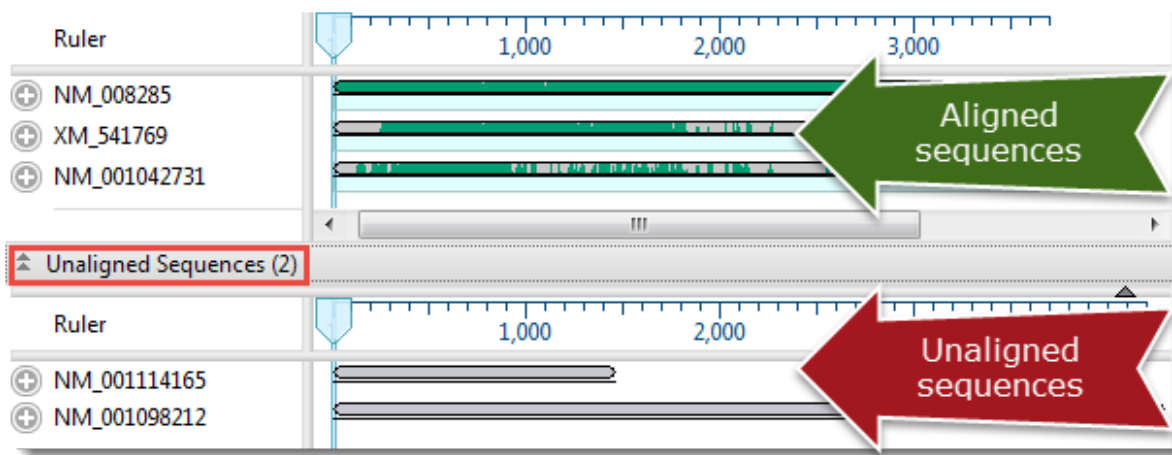
- To learn how to add the unaligned sequences to an existing alignment, see Merge unaligned sequences into an existing alignment.

- For information on performing a completely new alignment consisting of a combination of one or more unaligned sequences and one or more previously aligned sequences, see Align selected unaligned sequences with an existing alignment.

Using the Overview as a sequence navigation tool

The light blue vertical column in the [Overview](#) is a "viewport" showing the region currently being displayed in the [Sequences view](#). There are a number of methods—all performed within the Overview—for using the slider to navigate to a particular region in the Sequences view:

- In any project, move the slider to a desired region by dragging it with the mouse.

- In a single-block project, double-click anywhere on a block to center the slider over that position.

- In a multi-block project, double-click on an un-highlighted block to make it the active block and to move the slider to its left end.

- In a multi-block project, double-click anywhere on the active block to center the slider over that location.

Once you have centered the viewport in the desired location, use the horizontal zoom slider in the Sequences view to increase or decrease the width of the view.

Sequences view

The Sequences View displays added sequences and aligned sequences. Sequence names selected in this view are simultaneously selected all other views in MegAlign Pro. Similarly, sequence names selected elsewhere will be highlighted in the Sequences View.

To access the view, click on the **Sequences** tab or choose **View > Sequences > Show**.
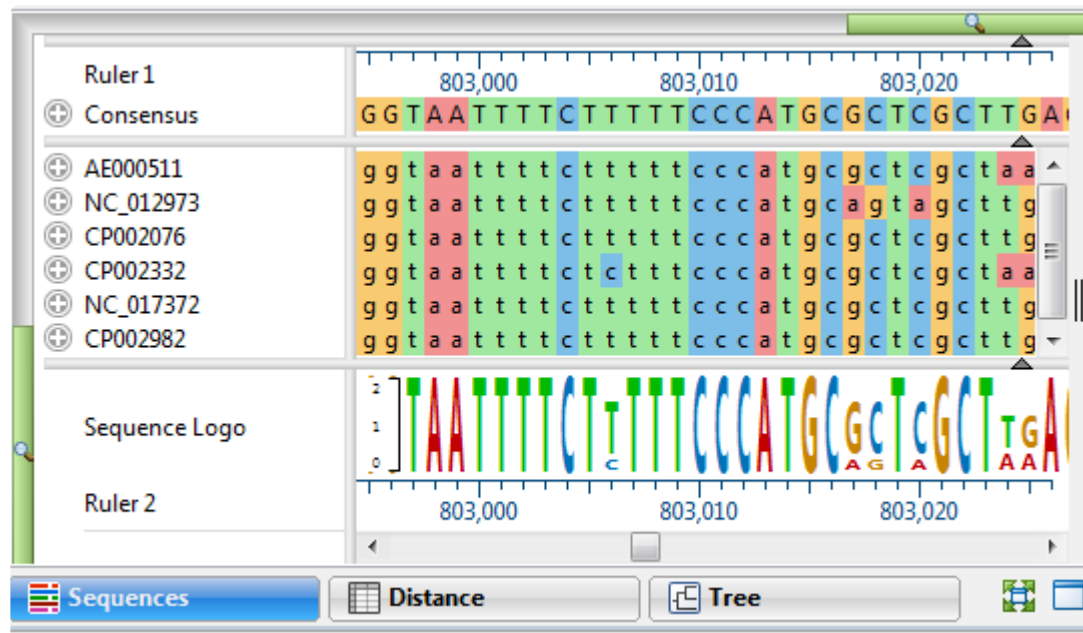
By default, the Sequences view consists of three sections: the header, the footer, and sequences area, the last of which consists of nucleotide or amino acid codes for each sequence on the right, and their labels on the left. Each section can display a variety of optional data tracks. The Sequences view has a different appearance depending on the zoom level.

When viewing a multi-block alignment produced by Mauve and using a minimal horizontal zoom, the currently-selected alignment block will appear the same color as in the Overview. Flanking sequences in adjacent blocks will appear gray. At higher zoom levels, the residues for the active block will be shown in bright colors, while flanking sequence residues are represented by "dimmed" (dull) colors.
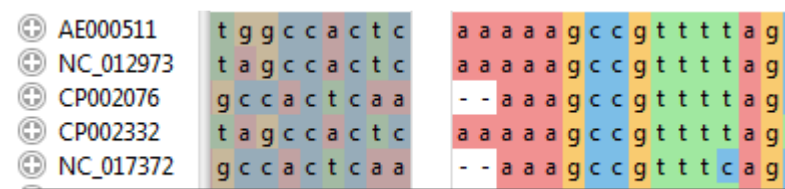
The dimmed residues are displayed as unaligned sequences, and therefore appear ungapped and without a consensus. To view the dimmed residues as aligned sequences, select the corresponding block in the Overview.

Information bar:

The blue/gray bar at the top of the view shows information pertaining to the current selection.

Selection	Example	Information displayed
Nothing, and sequences have not been aligned	(blank)	(blank)
Nothing, and at least some sequences have been aligned	Clustal Omega	The name of the alignment.
One sequence	isoH-NM_001032097.2 = 2,024	The sequence name and its length in bp.
Multiple sequences	2 of 4 aligned sequences 2 of 5 unaligned sequences	The number of sequences selected, the total number of sequences, and the alignment status.
A feature or any portion of one sequence	isoF-NM_001032098.2721>1257 = 537	The sequence name and selected range.
A portion of multiple sequences (selected in the Sequences view)	54>342 = 289	The selected range and length in bp.

Tasks pertaining to the Sequences view:

Task	How to…
Specify the look and colors used in the Sequences view	Use the Sequences section and Multiple Alignment section of the Style panel. Available color schemes are described here.
Customize the layout of the view, zoom in or split the view	See Customize the look and layout.
Show a specific	The Overview's "viewport indicator" highlights the portion of the sequence displayed in the Sequences view. Drag the indicator to scroll the Sequences view along the sequence. Use the

	zoom sliders in the Sequences view to control the width of the viewport, and thus, the amount of visible sequence.
sequence range in the view	Note that if you split the Sequences view, multiple viewports will appear in the Overview.
Show/hide a vertical guideline	The guideline is displayed, by default, and is a navigational aid to help in guiding your eye between the sequences and the rulers. To hide the guideline in all open projects, click **Sequences > Show Guideline** to remove the checkmark. Even after hiding the guideline, an abbreviated version of the guideline will remain visible in the header and footer.
Show/hide the header, sequences area, or footer in the Sequences view	The display of the header, sequences area or footer can be toggled on/off by clicking on the small dark gray arrows near the upper right corners of each section.
Search for a sequence by name	Click the **Search** tool ( ) on the right of the header area to open a search area. See Search for a sequence using a text query for details.
Informally compare other	See Visually compare other sequences to one sequence. This method offers several advantages compared to using a reference (see row below): 1) If you have already computed variants, those data remain accessible; 2) sequences don't move in the alignment. However, if there are many

sequences in a non-Mauve alignment to one particular sequence	sequences, the comparison sequence can scroll off screen.
Compare other sequences in the alignment to a designated reference sequence	See [Specify a reference sequence](#). In contrast to the "comparison" method in the row above, this offers the advantage that the comparison sequence can be seen at the same time as any other sequence in the alignment, even if there are many sequences in the alignment. In addition, you can choose to compute variants for the new reference, or leave them uncomputed.
Change the order of sequences in this view	Use the **Sort sequences** tool ( ). See [Change the order of sequences](#).
Export data from the view	Click the **Export Data** tool ( ), a shortcut to the **File > Export Data** subcommands. See [Export data to a file](#).
Save an image of the view	Click the **Export Image** tool ( ), a shortcut to the **File > Export Image** subcommands. See [Export an image of the view](#).
Print the view	Click the **Print** tool ( ). See [Print Images](#) for details.

Pairwise view

The Pairwise view displays sequences after performing a [pairwise alignment](#) or a [BLAST or Entrez database search](#). Sequence names selected in this view are simultaneously selected all other views in MegAlign Pro. Similarly, sequence names selected elsewhere will be highlighted in the Pairwise view.

To access the Pairwise view, do either of the following:

- [Perform a pairwise alignment](#). In this case, the Pairwise view opens automatically to display the results. To open the view at a later time, click on the **Pairwise** tab or choose **View > Pairwise > Show**. If there are multiple Pairwise tabs, hover over each tab to see a tooltip showing which sequences were involved in that alignment.

- [Perform a BLAST search](#). In this case the [Text](#), [Table](#) and Pairwise views are displayed together in a single popup window that is separate from the main MegAlign Pro window. To reopen the Pairwise view (and the associated views) at a later time, click on the Status link in the [Jobs panel](#).
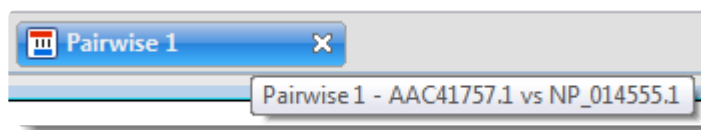
In the left of the header, the **Align** (left) drop-down menu is used to specify the target sequence, which should be the longest sequence of the two. The **with** (right) drop-down menu is used to specify the query sequence. Specifying a different sequence in either menu will initiate a new pairwise alignment; the original pairwise alignment and its statistics will be replaced with information from the new alignment.

Summary area:

The summary area describes the alignment engine and settings that were used to create the pairwise alignment.

If you [aligned a short sequence to a chromosome-length sequence](#) using the **Align to Chromosome: DNASTAR** method, the area will look like the image below.

	NC_000001.10	PER3_CDS	Exons Aligned	Min Length	Max Length
	1>249250621	1>3633	Count	Exon	Exon
Alignment	7844938>7902815	1>3633	21	52	698
Selection	7845118>7883417	129>1957	14	52	202

Align to Chromosome: DNASTAR DNA alignment [Minimum mer size: 19 Minimum match percentage: 90]

The columns will include the following information:

Value	Description
Exons Aligned Count	The number of exons in the short sequence that MegAlign Pro was able to align to the chromosome while honoring the **Minimum mer size** and **Minimum match percentage** thresholds.
Min Length Exon	The length of the shortest exon in the alignment.
Max Length Exon	The length of the longest exon in the alignment.

If you used any [pairwise alignment method](#) other than **Align to Chromosome: DNASTAR**, the area will look like the image below.

The columns will include the following information:

Value	Description
% Identity	Number of matching residues divided by the sum of all residues and gap characters. The quotient is multiplied by 100 to give the similarity as a percent.
% Gaps	The gap length divided by the consensus length. The quotient is multiplied by 100 to give the similarity as a percent.
Identical	Number of residues that are identical in both sequences.
Gap Count	Number of gaps (of any length) in either sequence of an alignment.
Gap Length	Number of gap characters in an alignment.
Score	The cumulative score for an alignment derived by adding the matrix-derived match scores and subtracting gap penalties.
Length	Number of residues in the consensus, including gaps that match in both sequences.

Sequences area:
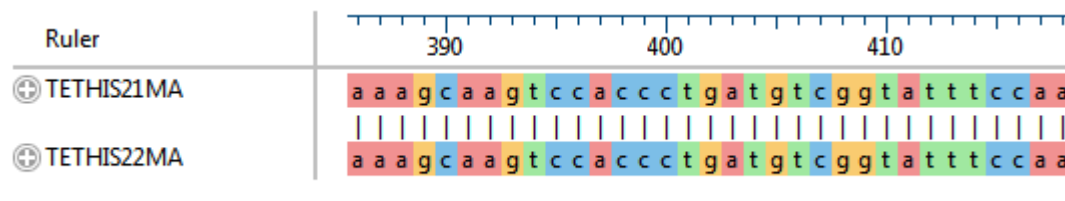
The Sequences area consists of nucleotide or amino acid codes for each sequence on the right, and their labels on the left. The order is always "target (or reference) sequence" first, and "query sequence" second. This section has a different appearance depending on the zoom level. A variety of optional data tracks can be displayed in this section.

If you used the **Align to Chromosome: DNASTAR** method, you will likely want to display the **Features** track for both the reference and query sequences in the Sequences area. You can then use the Features panel to limit the display to only certain types of features, such as CDS features. This lets you visually verify that the features start and end in the same places for both sequences. Features will usually be separated by large stretches of intron sequence.
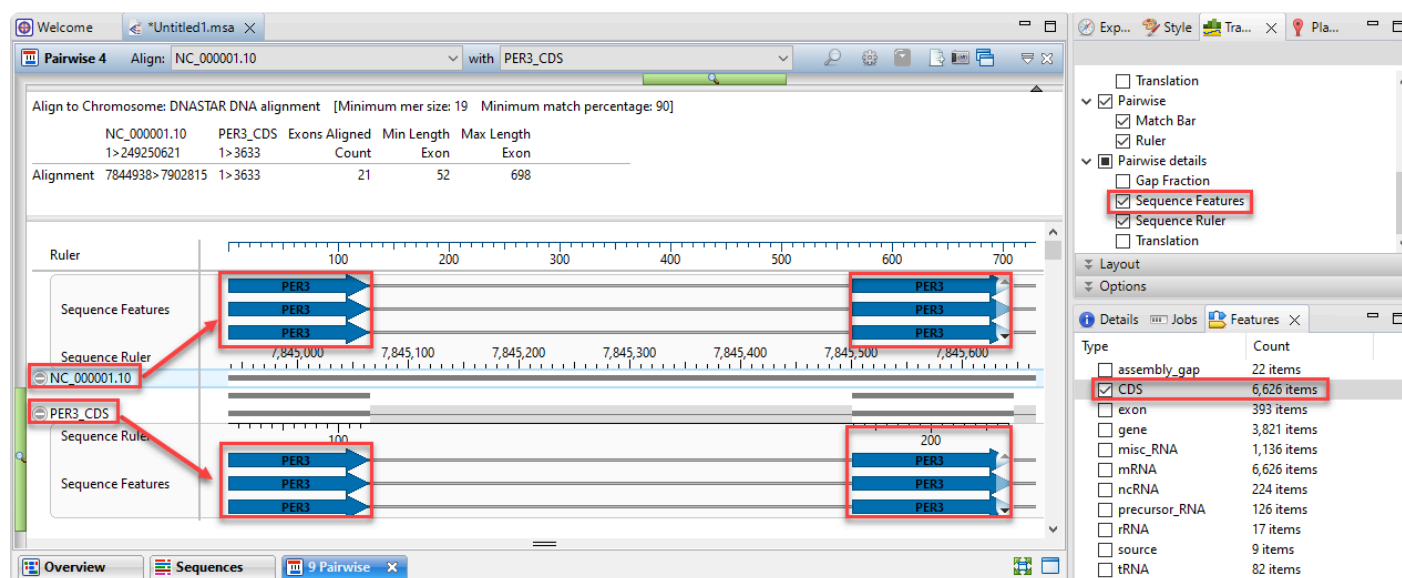
Tasks pertaining to a Pairwise view that resulted from a pairwise alignment:

In all but one case, the following tasks pertain solely to a Pairwise view that resulted from performing a pairwise alignment. If the Pairwise view resulted from a BLAST search, the only task available is **Export image** (second row from the bottom).

Task	How To
Customize the look	Use the Pairwise Alignment section of the Style panel. Available color schemes are

and colors used in the view	described here.
Customize the layout of the view	See Customize the look and layout.
Copy data from the view to paste into an e-notebook or text editor	First select the range by clicking on the first residue and **Shift** +clicking on the last residue. Then right-click on the selection and choose **Copy**.
Search for a sequence by name	Click the **Search** tool (🔍) on the right of the header area to open a search area. See Search for a sequence using a text query for details.
Change pairwise alignment options	Click the **Change alignment options in this view** tool (⚙), a shortcut to **Align > Pairwise**. Set options as desired and press **Align**.
Create a new pairwise subalignment	Select a portion of the alignment, then click the **Create new pairwise alignment from selection** tool (⬛). The subalignment will open in a new Pairwise view.
Export data from the view	Click the **Export Data** tool (📄), a shortcut to the **File > Export Data** subcommands. See Export data to a file.
Save an image of the view	Click the **Export Image** tool (🖼), a shortcut to the **File > Export Image** subcommands. See Export an image of the view. The summary and sequences areas are both exported.
Make a copy of the current view	Click the **Clone this view** tool (🗗). This makes an exact copy of the current Pairwise view, including sequences, alignment type, settings and current selection. This is useful if you want to change one factor for comparison to the original alignment.

Distance view

The Distance view shows a matrix (i.e., table) of numbers representing distances computed for each pair of sequences for the selected aligned block. Sequence distances are used as input in creating the phylogenetic tree shown in the Tree view. Selections within the Distance view are synchronized with other views in MegAlign Pro. Similarly, selections made elsewhere will be highlighted simultaneously in the Distance view.

The Distance view is only accessible after you have performed a multiple alignment, and becomes inaccessible again if you unalign the sequences. After performing an alignment, you can access the view by clicking on the **Distance** tab or by choosing **View > Distance > Show**.
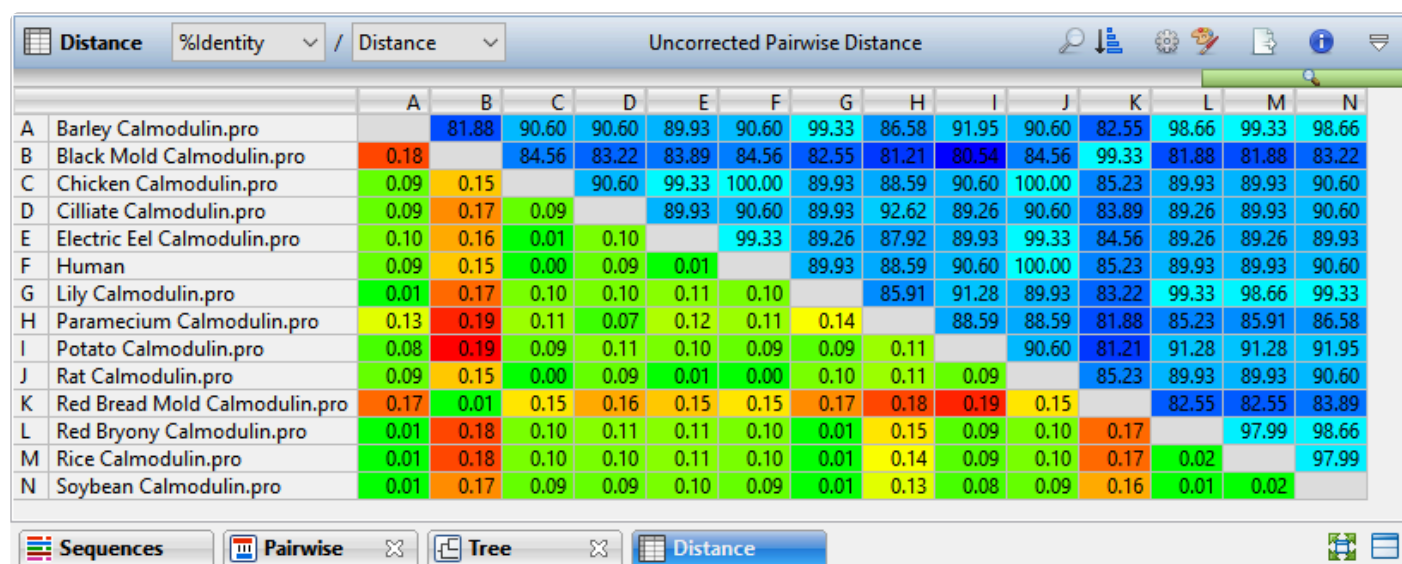
The upper-right and lower-left triangular-shaped sections of the matrix can show different information depending on **Values** selections made in the Distance section of the Style panel or using the drop-down menus in the upper left of the Distance view itself. For instance, in the image above, **%Identity** appears in the upper-right and **Distance** in the lower-left. Note that no numbers appear on the diagonal, since a sequence cannot differ from itself. In certain situations, it may not be possible to calculate distances for every pair of aligned sequences, in which case "NA" is displayed. In other cases, "NA" may be an artifact of the computational algorithm; for example, an attempt to calculate the log of a negative number.
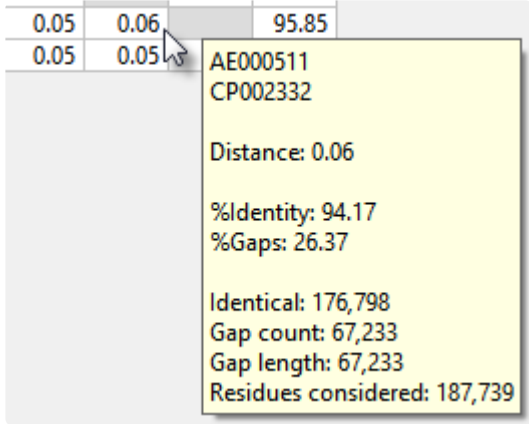
If **Identity** is displayed, note that larger numbers correspond to greater differences between sequences. Uncorrected distances are always ≤1.0. However, the Kimura and Tamura-Nei values can exceed 1.0 because they reflect the inferred actual number of nucleotide substitutions or amino acid replacements per site, not the observed distance. When the divergence between a pair of sequence is greater than about 0.75, it may not be possible to calculate a value for corrected distance using the Kimura and Tamura-Nei metrics.
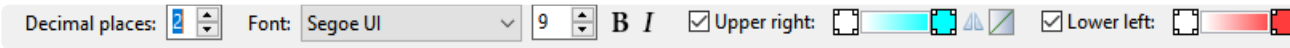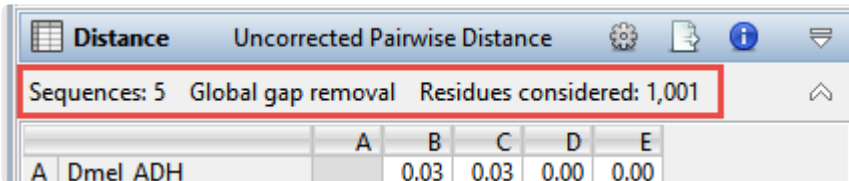
Information bar:

The blue/gray bar at the top of the view starts on the left end showing the color of the **Block** (multi-block Mauve alignments only). The rest of the information bar consists of information pertaining to the current selection.

Selection	Example	Information displayed
Nothing	Uncorrected Pairwise Distance	The distance metric currently specified.
One sequence or a row/column header (letter)	AAC41757 = 375	The name of the sequence and length in bp.
One table cell	2 of 5 unaligned sequences	The number of sequences involved in the cell and their alignment status.

Tasks pertaining to the Distance view:

Task	How To
Customize the look and layout of the Distance view	See Customize the look and layout. Statistical metrics and fonts are set using the Distance section of the panel.
Display a tooltip with the name of the sequence corresponding to a cell	Hover over any row/column header or any cell. For projects using **Pairwise gap removal** (see Gap treat options), the tool tip also contains the number of residues considered for that pair.
Select all cells in the table	Choose **Distance > Table > Select All**.
Use editing commands with the Distance table	Choose **Distance > Table > Copy (Select All, Rename, Rename with Fields)**.
Freeze	Choose **Distance > Table > Freeze Scrolling Region**. In a large table, this command allows you to com

scrolling	rows near the top with those further down.
Search for a sequence by name	Click the **Search** tool ( ) on the right of the header area to open a search area. See Search for a sequ using a text query for details.
Change the order of sequences in this view	Use the **Sort sequences** tool ( ). See Change the order of sequences.
Change the parameters used to generate the distance table	Click the **Change Analysis Parameters** tool ( ) or choose use **Distance > Parameters** to open the A Options dialog. See Distance metric options and Gap treatment options for detailed information on this d
Expand the header to show style tools	Click the **Show Style** tool ( ) to add a row below the current header that allows you to control the same settings that can also be controlled via the Distance section of the Style panel. Decimal places: 2 ⇕ Font: Segoe UI ▾ 9 ⇕ **B** *I* ☑ Upper right: 🔲▬🔲◪◩ ☑ Lower left: 🔲▬ To minimize the row, click the tool again or use the **Collapse** tool (inverted 'v') on the right of the row.
Export data from the view	Click the **Export Data** tool ( ), a shortcut to the **File > Export Data** subcommands. See Export data to
Save an image of the view	Click the **Export Image** tool ( ), a shortcut to the **File > Export Image** subcommands. See Export an i of the view.
See additional details in the view header	To add a second row of header information, click the **Show Quick Details** tool ( ). **Distance** Uncorrected Pairwise Distance 🔳 📄 ℹ ⇳ Sequences: 5 Global gap removal Residues considered: 1,001 ⌃ A B C D E A Dmel_ADH 0.03 0.03 0.00 0.00 • **Sequences** – The number of sequences in the alignment or in the active block. • **Global/Pairwise gap removal** – The gap treatment specified in the Distance section. • **Residues considered** – Total length of alignment minus all columns containing even a single This field shows both the minimum and maximum number of residues considered, if they diffe banners at the tops of the Distance and Tree views will show a range here if different number residues are considered for different pairs of sequences. This happens frequently when **Pairw**

	gap removal is checked. However, if the minimum and maximum values differ greatly, we recommend selecting **Global gap removal** instead

Distance metric options

To change the distance metric used to generate the [Distance view](#) table, click the **Change Analysis Parameters** tool (⚙) in the [Distance view](#). This opens the Analysis Options dialog where you can specify the **Metric** and **Gap Treatment**. Another way to access the same dialog is to use **Distance > Parameters**.
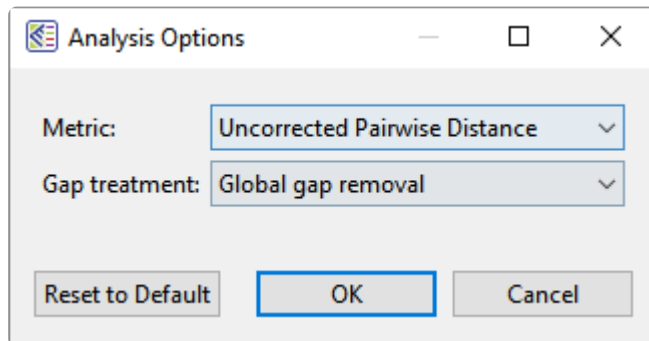
Use the **Metric** drop-down menu to change the metric used to calculate distance. Note that metric selection is disabled for projects without sequences. Otherwise, choices include:

- **Uncorrected Pairwise Distance** – (the default metric) – The number of identical bases (or residues) divided by the number of bases being compared, ignoring any positions with gaps. Uncorrected pairwise distance can be converted to %Identity using the formula: %ID = 100 * (1 – distance).

- **Kimura** – (protein sequences only) – The [Kimura](#) model (1983), which should not be used for very divergent sequences, uses the formula: D = – ln (1 – p – 0.2 p2), where p is the uncorrected pairwise distance. The Kimura distance approximates the "**PAM distance**" used in MegAlign.

- **Scoredist** – (protein sequences only) – This model, developed by [Sonnhammer and Hollich](#) (2005) computes the alignment score between two sequences using the BLOSUM62 scoring matrix. The units for the ScoreDist function are "percent accepted point mutations" (PAM). Scores are converted to distances and normalized by the average scores of the two sequences matched to them. This approach works even for very divergent sequences, providing that there are overlapping residues.

- [**Tamura-Nei**](#) (1993) – (nucleotide sequences and low degrees of divergence only) – An estimate of divergence using the TN93 model of the evolutionary process. This model attempts to separately account for different rates of transversion mutations (e.g., purine ↔ pyrimidine) and the two categories of transition mutations (i.e., purine only A ↔ G; pyrimidine only C ↔ T). The frequencies of each type of nucleotide do not need to be the same.

Note that you must press the **OK** button before MegAlign Pro will initiate the calculation or display new results in the Distance view..

See the topic [Gap treatment options](#) for information on that portion of the dialog.

Gap treatment options

To change the distance metric used to generate the Distance view table, click the **Change Analysis Parameters** tool (⚙) in the Distance view. This opens the Analysis Options dialog where you can specify the **Metric** and **Gap Treatment**. Another way to access the same dialog is to use **Distance > Parameters**.
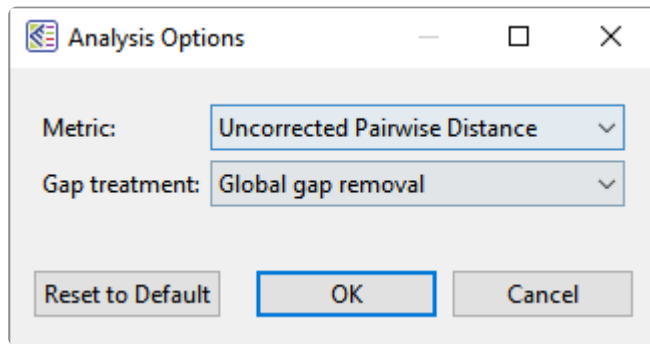
Use the **Gap Treatment** drop-down menu to specify when gaps should be ignored. Choices include:

- **Global gap removal** – (default) This option ignores any column that contains a gap in *any* of the aligned sequences in the project. This is the only enabled option for MegAlign Pro projects created prior to the 12.0 release (spring 2014).

- **Pairwise gap removal** – This option ignores any column that has a gap in either of the two sequences being compared.

Note that you must press the **OK** button before MegAlign Pro will initiate the calculation or display new results in the Distance view..

To learn about the other part of this dialog, see Distance metric options.

Tree view

The Tree view shows evolutionary relationships estimated from the multiple sequence alignment, and is calculated using the current tree-building algorithm and settings, and the currently-selected Distance metric. This view is only accessible after you have performed a multiple alignment, and becomes inaccessible again if you unalign the sequences. This topic is an introduction to the Tree view. To learn how to make, customize, reroot or export trees, refer to the main topic Phylogenetic Trees.

To access the Tree view, use the **View > Tree > Tree** command or click on the **Tree** tab. The Tree view initially displays the following graphic, prompting you to choose a tree-building algorithm.
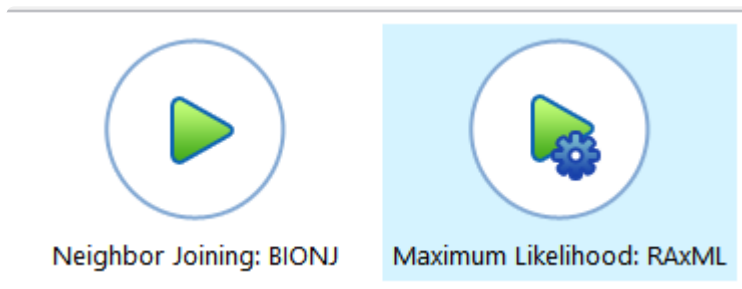
Once you choose an algorithm and MegAlign Pro builds the tree. The phylogeny is displayed as a rooted tree reflecting the evolutionary relationships between the aligned sequences. The following image shows a tree with some styling changes applied to it, including collapsing some of the branches to more easily view the sequences of interest.
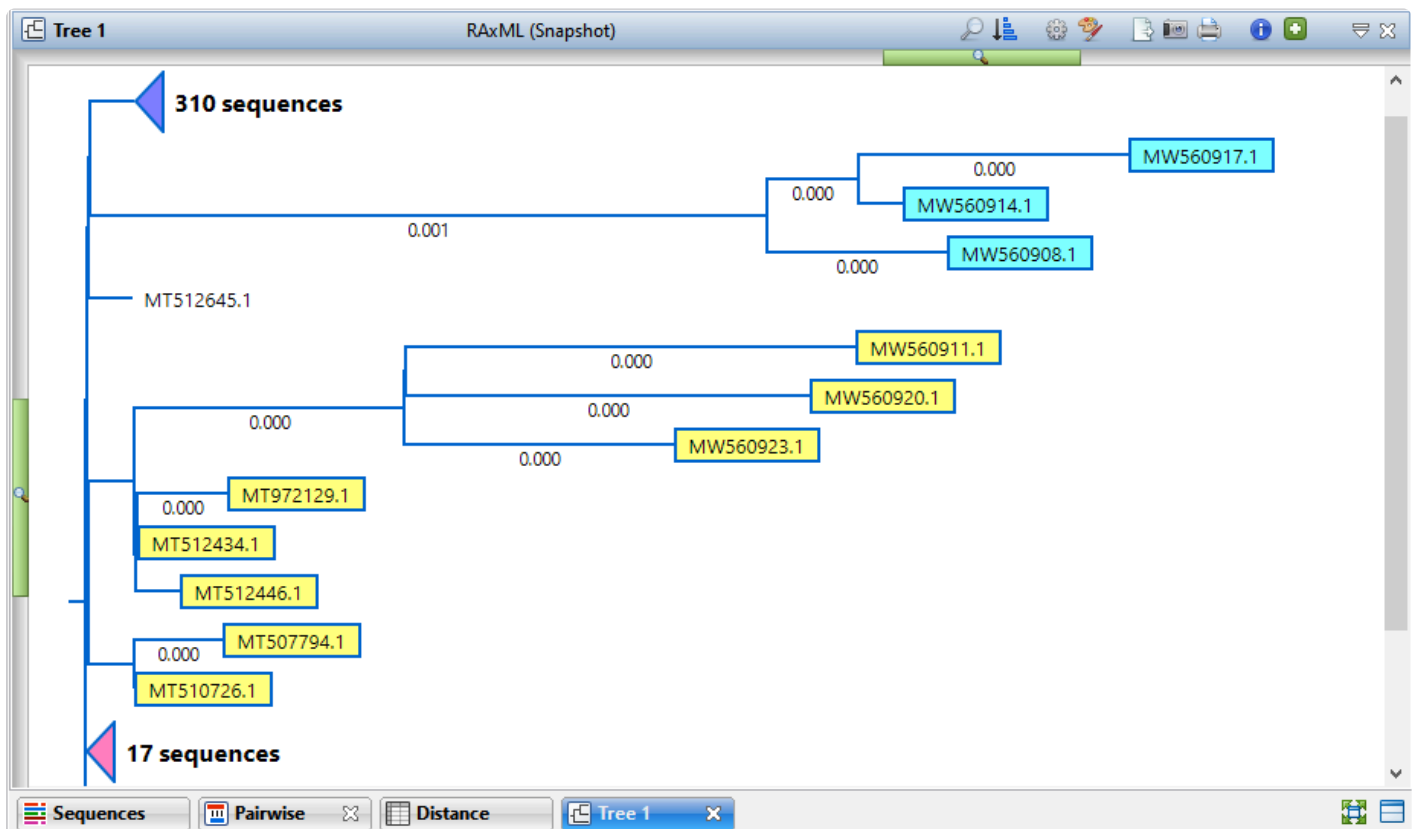
Some information about the view:

- In this view, sequences are ordered top-to-bottom in the vertical formats (see Tree section) and left-to-right in the horizontal ones.

- Selections within the Tree view are synchronized with other views in MegAlign Pro. Similarly, selections made elsewhere will be highlighted simultaneously in the Tree view.

- Distance values are reflected in the lengths of the branches and their numerical values are are displayed by default. A distance scale appears at the bottom left of the image.

- Changes made to the alignment (e.g., realigning with a different method, or with additional sequences) or to the distance metric cause the tree to be recalculated.

- In certain situations, it may not be possible to calculate distances for every pair of aligned sequences. In these cases, sub-clusters of related sequences may be shown together on separate trees, or listed together as a single node. See Distance view for more information.

- Due to the difference in alignment methods, trees built in MegAlign (discontinued in 2019) and MegAlign Pro may appear different. In addition, trees built in MegAlign are rooted, whereas trees built in MegAlign Pro are always unrooted, regardless of the method used for the alignment. The presence or absence of "roots" can cause a discrepancy in the vertical ordering of sequences between the two applications.

Information bar:
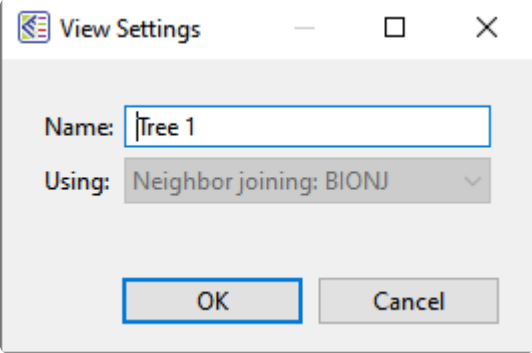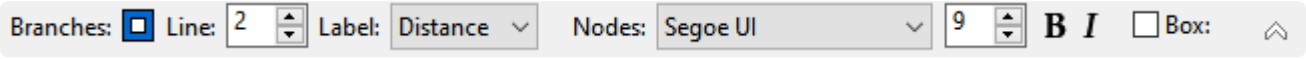
The blue/gray bar at the top of the view starts on the left end showing the color of the **Block** (multi-block Mauve alignments only). The rest of the information bar consists of information pertaining to the current selection.

Selection	Example	Information displayed
Nothing	BIONJ	The name of the tree-building algorithm.
One sequence	AAC41757 = 375	The name of the sequence and length in bp.
More than one sequence	2 of 5 unaligned sequences	The number of sequences selected and their alignment status.

Tasks pertaining to the Tree view:

Task	How To
View more	Drag the vertical zoom slider upward. This will cause the font size for branch labels to shrink, and

of a large tree	eventually to disappear, allowing more space to be devoted to the tree branch layout.
Collapse branches for better viewing of the tree	Hover over any node where tree branches meet and click the minus sign (⊖) to collapse those nodes. The minor branches will collapse into a triangle accompanied by text showing the number of collapsed sequences in that node. To control the size and appearance of the triangle, use the **Collapsed branches** area at the bottom of the [Tree section](#) of the [Style panel](#). To expand the node, hover over the triangle and click the "plus sign" (⊕).
Flip the order of leaves on a selected branch	Select a branch and then choose **Tree > Flip > Selected Branch**. This command does not affect the statistics underlying the tree, but is solely an aesthetic choice, such as when exporting the image for publication.
Flip the entire tree	Choose **Tree > Flip > Whole Tree**. This changes the tree to a horizontal "mirror image" of the original layout. Text is still shown in the usual format. This command does not affect the statistics underlying the tree, but is solely an aesthetic choice, such as when exporting the image for publication.
Designate a particular branch as the "root" of the tree	Select any branch and choose **Tree > Root On**. The true distance value of the actual branch with the displayed root is the sum of the two displayed branches flanking that displayed root.
Redraw the tree according to the default settings	Use **Tree > Reset View of Phylogenetic Tree** or **Ctrl/Cmd+Shift+R**.
Make a selection in the Tree view	See [Make a Selection](#).
Search for a sequence by name	Click the **Search** tool (🔍) on the right of the header area to open a search area. See [Search for a sequence using a text query](#) for details.
Change the order of sequences in this view, or change	Use the **Sort sequences** tool (🗐). See [Change the order of sequences](#).

other views to match the order in the tree	
View current tree algorithm and name the tree	Click the **View settings** tool (⚙). This opens the View Settings dialog where you can see the method employed to draw the tree and type in a new **Name** if desired. *[View Settings dialog]* Name: Tree 1 Using: Neighbor joining: BIONJ OK Cancel
Change the appearance of the tree	Use **View > Style > Tree** to open the [Tree section](#) of the [Style panel](#). Alternatively, click the **Show Style** tool (🎨) to add a row below the current header that replicates certain settings in that section. In either case, see the [Tree section](#) topic for information on changing settings. Branches: ☐ Line: 2 ⬍ Label: Distance ⌄ Nodes: Segoe UI ⌄ 9 ⬍ **B** *I* ☐ Box: ⌃ To minimize the row, click the tool again or use the **Collapse** tool (inverted 'v') on the right of the row.
Export data from the view	Click the **Export Data** tool (📄), a shortcut to the **File > Export Data** subcommands. See [Export data to a file](#).
Export an image of the view	Click the **Export Image** tool (🖼), a shortcut to the **File > Export Image > (Tree Name)**. See [Export an image of the view](#) for details about export options.
Print the view	Click the **Print** tool (🖨). See [Print Images](#) for details.
Add a "details" header to the view	Click the **Show Quick Details** tool (ℹ). The header contains a subset of the following information: • **Block** – (multi-block Mauve alignments only) The color of the alignment block currently selected in the [Overview](#). • **Alignment** – The alignment method used.

- **Distance Metric** – The distance metric (calculation method) specified using **Distance > Parameters**.

- **Sequences** – Number of sequences in the alignment or in the active block.

- **Global/Pairwise gap removal** – The gap treatment specified in the Distance section of the Style panel.

- **Residues considered** – Total length of alignment minus all columns containing even a single gap. This field shows both the minimum and maximum number of residues considered, if they differ.

Save an image of the current BIONJ tree	Use the **Take snapshot** tool (⬛). In the save dialog, choose whether to save the image as a PNG, JPG, PDF, or editable PowerPoint document (*.pptx*). This tool is only available in the Tree view for BIONJ trees.
Change the method and/or parameters used to generate the tree	Click the **Compute Phylogeny** tool (➕). This is a shortcut to the **Tree > Compute Phylogeny** command and allows you to change the tree calculation algorithm and related parameters. For detailed information, see Phylogenetic Trees.

Text view

To access the Text view, perform a BLAST search. When the search is complete, the Text, Table and Pairwise views are displayed together in a single popup window that is separate from the main MegAlign Pro window. To reopen all three views at a later time, click on the **Status** link in the Jobs panel.

An example Text view is shown below:
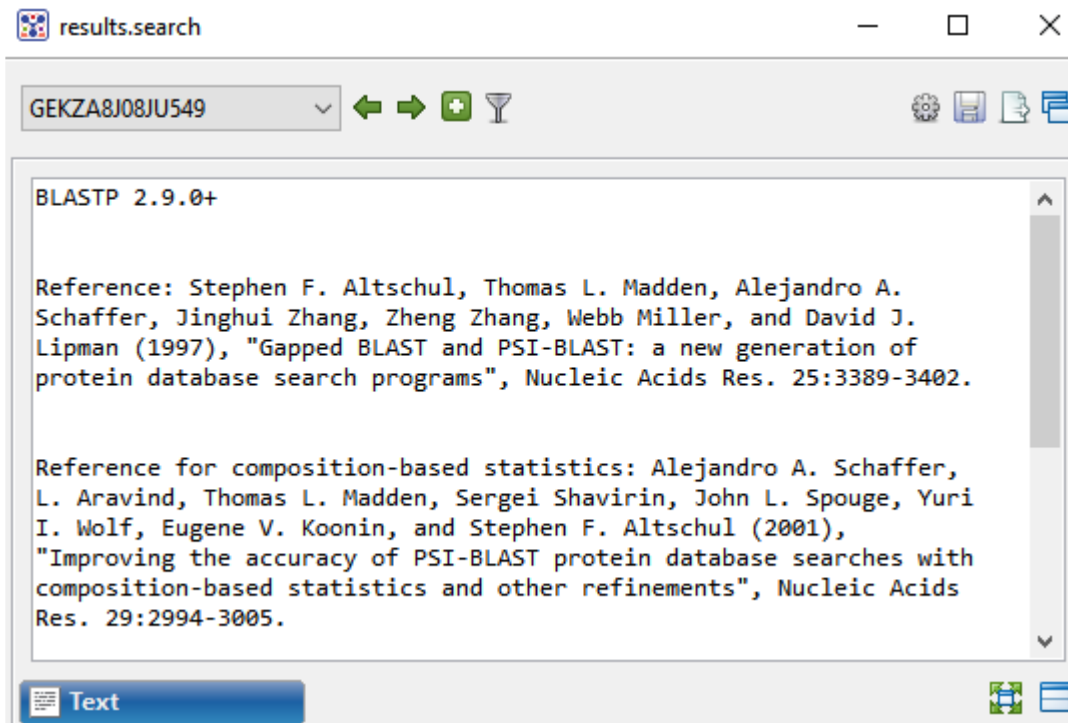
In order, the view shows the following:

- The version of BLAST used in the search (e.g., "BLASTP 2.9.0+").

- References for the BLAST search engine.

- A description of the BLAST database used in the search, including the number of sequences.

- Information about the Query sequence, including its length and the Request ID (RID) number.

- A table of sequences producing significant alignments.

- Pairwise alignments of each significant match with summary information on the match.
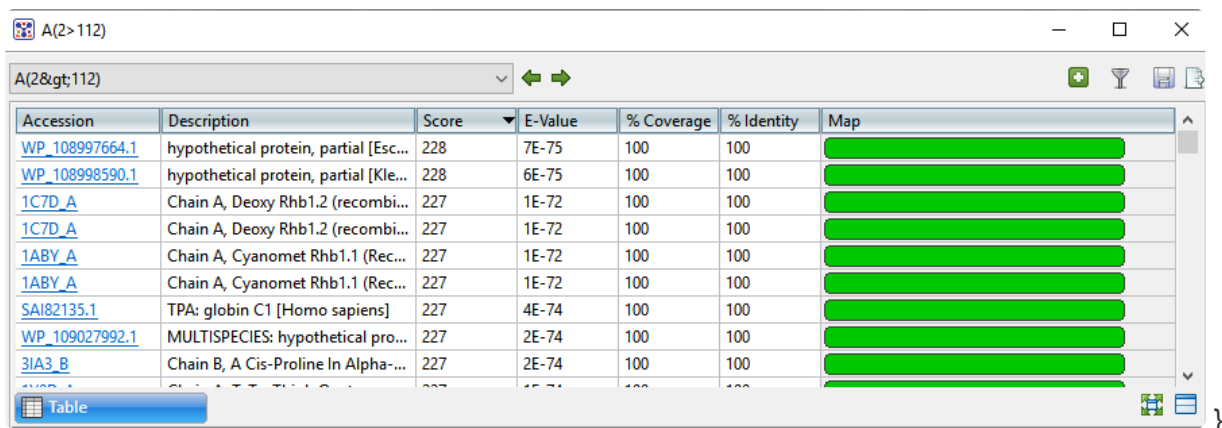
Tasks pertaining to the Text view:

Task	How to…
Make selections in the view	See [Make a Selection]().
Copy all text	Right-click anywhere in the view and choose **Copy**.
Select all text	Right-click anywhere in the view and choose **Select All**.

Table view

To access the Table view, perform a BLAST search. When the search is complete, the Text, Table and Pairwise views are displayed together in a single popup window that is separate from the main MegAlign Pro window. To reopen all three views at a later time, click on the **Status** link in the Jobs panel.

If no significant matches were identified, the Table view will be empty. An example of a populated Table view is shown below:

Columns in the view's table are described below. Only a subset of these columns will be available, depending on whether you did a sequence or text search.

Column	Description
Accession	The NCBI "Accession" number for the match. Press the link to open the corresponding page on the NCBI website.
Description	The NCBI "Definition" field text.
Score	The similarity score for the match. (See the NCBI definition for **raw score**). In general, a higher score denotes a better match.
E-Value	The chance that the match is expected to occur by chance. (See the NCBI definition for **E value**). In general, a lower score denotes a better match.
%Coverage	The percentage of the query sequence that aligns to a match in the BLAST database. Example: Query has 596 bases, of which 565 align to a match in the NCBI database. **%Coverage** = 565/596 = 95%.
%Identity	The percentage of bases that match when the query and result sequence are aligned. (NCBI definition). Example: Query has 565 bases that align to the match and 3 bases that don't match. **%Identity** = 562/565 = 99.5%.

Map	Graphic showing areas of agreement between the query and the match.
Length	(Entrez search only) The length of the sequence in bp.
Organism	(Entrez search only) The name of the organism.

Tasks pertaining to the Table view:

Task	How To…
If a BLAST or Entrez search was performed	
Launch the NCBI page for matching database entry	Click the hyperlink in the table's Accession column.
Copy all information from a search result row	Right-click on the row and choose **Copy**.
Save the sequence match in *.gbk* format	Right-click on the row and choose **Download sequences**. In the ensuing Save As dialog, choose a location in which to save the sequence. By default, the sequence is named using its Accession number.
Add sequences or matches	Use the **Add sequences or matches** tool ( ). Sequences are added back into the host application, not the job's document.
Filter results from a sequence or text search	Use the **Filter results** tool ( ) and choose **Filter** to launch the Filter dialog. See See Add only sequences that meet filtering thresholds for instructions on using this dialog. This dialog lets you limit matches in the table by setting up custom filters.
Export data from the view	Use the **Export data** tool ( ), described in the topic Export data from the Text, Table or Pairwise views.
Only if a BLAST search was performed	
To save results	Click the **Save as** tool ( ). A save dialog opens and allows you to save the search results as a *.search* format file.
To display a pairwise alignment using a particular search result	If both the Table and Pairwise views are open, click on the desired row in the Table view to see the corresponding pairwise alignment in the Pairwise view. As an alternative, open the Pairwise view manually and choose the desired target from the **with** drop-down menu.

Variants view

To see the Variants view, you must first perform a multiple alignment. When the alignment is complete, the Variants view opens showing a **Compute Variants** button. Click the button to calculate the variants.

If you have specified a reference sequence, variants are calculated against the reference. If no reference sequence was specified prior to alignment, variants are calculated against the consensus sequence. Pressing the button results in a table of variants being displayed in the view.

Once the Variants view has been enabled, you can access it by clicking the **Variants** tab or by choosing **View > Variants > Show**.

Each row in the Variants table represents a variant in an aligned sequence.

> ✳ **Note to SeqMan Ultra users:** In SeqMan Ultra, raw sequencing reads that are generated

from the same source DNA are aligned and variants are calculated by assessing the vertical
depth of aligned reads in a column. In MegAlign Pro, the aligned sequences are all verified
but do not share the same source DNA, so variants can be calculated from a single aligned
sequence. Therefore, the MegAlign Pro variants table reports variants in each aligned
sequence, but does not compute variants vertically down columns.

The table columns are described below. Some columns may only be available for amino acid sequences
and others for nucleotide sequences.

Column Name	Description
Peptide Impact	
Ref Aliphatic	The reference's relative volume occupied by aliphatic side chains (alanine, valine, isoleucine and leucine) as calculated using the method of (Gasteiger et al. 2005). The aliphatic index predicts regional stability by calculating the relative volume occupied by aliphatic side chains and is a positive indicator of globular protein thermostability. Detailed information about this method is available on the ExPASY website.
Ref Charge	The reference's pH-dependent sum of charges in molecules, assuming a pH of 7. DNASTAR developed this method using the pK tables of Lehninger, 2005.
Ref Hydropathy	The reference's hydropathy value as calculated using the approach of Kyte and Doolittle, 1982. Residue hydropathy assignments are derived from water-vapor transfer free energies and the interior-exterior distribution of residue side-chains.
Ref Instability	An estimate of the stability of the reference protein in a test tube, calculated using the approach of Guruprasad et al., 1990. This approach predicts regional instability by calculating the weighted sum of dipeptides that occur more frequently in unstable proteins when compared to stable proteins. Detailed information about this method is available on the ExPASY website. An instability index less than 40 predicts a stable protein, whereas values higher than 40 denote a potentially unstable protein.
Ref pI	The reference's isoelectric point, the pH at which the residue carries no net electrical charge. This is calculated by using the pKa tables from Lehninger et al., 2005 to identify uncharged regions at a given pH, which have an increased potential to aggregate if surface exposed.
Var Aliphatic	The variant's relative volume occupied by aliphatic side chains (alanine, valine, isoleucine and leucine) as calculated using the method of (Gasteiger et al. 2005). The aliphatic index predicts regional stability by calculating the relative volume occupied by aliphatic side chains and is a positive indicator of globular protein thermostability. Detailed information about this method is available on the ExPASY website.
Var Charge	The variant's pH-dependent sum of charges in molecules, assuming a pH of 7. DNASTAR developed this method using the pK tables of Lehninger, 2005.

Var Hydropathy	The variant's hydropathy value as calculated using the approach of Kyte and Doolittle, 1982. Residue hydropathy assignments are derived from water-vapor transfer free energies and the interior-exterior distribution of residue side-chains.
Var Instability	An estimate of the stability of the variant protein in a test tube, calculated using the approach of Guruprasad et al., 1990. This approach predicts regional instability by calculating the weighted sum of dipeptides that occur more frequently in unstable proteins when compared to stable proteins. Detailed information about this method is available on the ExPASY website. An instability index less than 40 predicts a stable protein, whereas values higher than 40 denote a potentially unstable protein.
Var pI	The variant's isoelectric point, the pH at which the residue carries no net electrical charge. This is calculated by using the pKa tables from Lehninger et al., 2005 to identify uncharged regions at a given pH, which have an increased potential to aggregate if surface exposed.
ΔAliphatic	**Var Aliphatic** minus **Ref Aliphatic**.
ΔCharge	**Var Charge** minus **Ref Charge**.
ΔHydropathy	**Var Hydropathy** minus **Ref Hydropathy**.
ΔInstability	**Var Instability** minus **Ref Instability**.
ΔpI	**Var pI** minus **Ref pI**.
Variant	
Codons	When a CDS feature is present on the reference sequence at the position of a variant, this column displays information about the codon change. All information in this column is a compilation of data from other available data columns, regardless of whether or not those columns are currently displayed. The information appears as: **Ref Protein Pos** followed by **Ref Translation:Ref Codons → Var Translation:Var Codons**. Example: 5 S:TCA → L:TtA.
Cons Pos	The nucleotide position where the variant occurs in the gapped consensus, as measured from the left end of the alignment block.
Feature Distance	The distance from the nearest known feature, if any. A positive number denotes that the position is upstream of the nearest known feature, while a negative number means that it is downstream of the feature. A zero value means that the position is inside the span of a coding feature (including introns).
Feature Name	Name of the containing or nearby coding feature, if any.
Impact	The impact of the variant or indel on the genome, displayed as one of the following values: • **Synonymous** – No amino acid changes. • **Non Synonymous** – Amino acid substitution only.

- **Nonsense** – Amino acid to translational stop.

- **Frameshift** – An indel within a coding region and which is not a multiple of 3, thereby changing the reading frame.

- **No Start** – A change that disrupts the start codon.

- **No Stop** – A change that converts a stop codon to an amino acid, and thereby extends the reading frame.

- **Inframe Insertion** – An insertion within a coding region whose length is divisible by 3.

- **Inframe Deletion** – A deletion within a coding region whose length is divisible by 3.

    If sorting by the **Impact** column, the column is ordered alphabetically, except that Noncoding appears at the bottom of the list.

Length	The total length of the variant, whether it consists of a 1-base SNP or a multi-base insertion or deletion.
Ref Bases	The base(s) found at that position in the reference sequence.
Ref Coding Pos	The DNA position in the coding sequence for the reference protein.
Ref Codons	The corresponding reference-sequence codons, for variants aligned with a coding feature in the reference sequence.
Ref Pos	The nucleotide position where the variant starts, as measured from the 5' end of the reference sequence.
Ref Protein Pos	The amino acid position in the reference protein where the variant has been detected. The reference protein is derived from a translation of the cDNA features in the reference sequence.
Ref Translation	The translation of the **Ref Codons**. Uses the genetic code, if any, identified by the coding feature. Otherwise, uses the standard genetic code.
Sequence	The name of the sequence that contains the variant, when compared to the designated reference sequence.
Strand	The strand of the coding feature on the reference sequence. A right-facing arrow (>) denotes the forward strand, while a left-facing arrow (<) denotes a reverse-oriented strand.
Type	Specifies the variation type. **SNP** is only used for nucleotide sequences and **Sub** (stitution) is only used for protein sequences. **Del** (etion) or **Ins** (ertion) can be used for either sequence type and is followed by the length of the deletion/insertion.
Var Bases	The base(s) found at that position in a non-reference sequence.

Var Codons	The codons in the variant sequence containing the variant residues, for variants aligned with a coding feature in the reference sequence.
Var Pos	The nucleotide position where the variant starts, as measured from the 5' end of the sequence containing the variant.
Var Translation	The translation of the **Var Codons**. Uses the genetic code, if any, identified by the coding feature. Otherwise, uses the standard genetic code.

Tasks pertaining to the Variants view:

Task	How to…
Select or deselect all variant rows in the table	Select all variants using **Variants > Select > Select All**. Remove the selection using **Variants > Select > Select None**. Selection is often followed by using a copy command.
Select individual variant rows	Use the mouse or use **Shift+click** or **Cmd/Ctrl+click**. Selection is often following by using a copy command.
Select all rows *except for* the currently-selected rows	After first selecting one or more rows (see above), choose **Variants > Select > Invert Selection**. Selection is often following by using a copy command.
Sort variants	Click on a column header to sort variants by the values in that column. Some columns are sorted alphabetically, some numerically, and some with custom orders. To sort by multiple columns, press the **Alt** key (Windows) or the **Option** key (Mac) while clicking additional column headers. Note that in Lasergene 17.4 (the first version featuring this view), sorting by sequence name may be slow.
Navigate to a variant	Click the variant in the variants table. The Sequences view and Overview will navigate to that variant.
Navigate to a variant and show the Sequences view	Double-click the variant in the variants table. The Sequences view will appear. The Sequences view and Overview will navigate to that variant.
Choose columns that should appear in the view and their order	Press the **Choose or rearrange columns** (⚙) tool in the top right of the view. Items in the list on the right are the columns currently being displayed and the order in which they appear. Make selections and use the up/down and right/left arrows to add, subtract, and reorganize these columns as desired.

Export data from the view	Click the **Export Data** tool ( ), a shortcut to the **File > Export Data** subcommands. See [Export data to a file](#).

Panels

Settings panels allow you to optimize the look, contents and layout of the MegAlign Pro window. To learn what can be done within each panel and the expandable "sections" within each panel, click the links below.

- Style
- Tracks
- Features
- Places
- Details
- Console

The expand bars used to open panel areas have a "hamburger menu" on the right end of the bar.
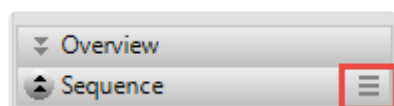
Hamburger Menu Command	Description
Available for all expand bars	
Collapse Others	Collapses all expand bars in the panel except the one whose menu was used to choose this command.
Collapse All	Collapses all expand bars in the panel.
Available only for expand bars in the Style panel	
Apply Sequence Style	Lets you choose and apply a file of custom settings for the section controlled by this expand bar. See Style panel for more information.
Save Sequence Style	Saves the current settings in the section controlled by this expand bar. See Style panel for more information.
Save Sequence Style as Default	Saves the current settings in the section controlled by this expand bar to use as default settings in future projects. See Style panel for more information.
Restore DNASTAR Default Sequence Style	Restores the default DNASTAR settings for the section controlled by this expand bar. See Style panel for more information.

Explorer panel

The Explorer panel is used to control which phylogenetic tree is displayed in the Tree view, and allows you to create additional trees or delete them.

To open the Explorer panel:

Click on the **Explorer** tab ( ). If the panel is not already visible, reveal it using any of the following methods:

- Choose **View > Explorer > Explorer** or **View > Reset Panel Layout**.

- Press **Ctrl/Cmd+Alt+E**.

The only section currently available is the Trees section, which is described separately.

Trees section

The Trees section of the Explorer panel is used to control which phylogenetic tree is displayed in the Tree view, and allows you to create additional trees or delete them.

> ✳ **Note**: If you are instead seeking information on the Tree section of the Style panel, click here.

To open the Trees section, click on the **Explorer** tab (⊘); if the **Trees** section is closed, expand it by clicking the **Trees** bar. Alternatively, choose **View > Explorer > Explorer** or **View > Reset Panel Layout**.

Prior to performing a multiple alignment, the Trees section is empty. Once an alignment has taken place, the Trees section contains a table with one row for each phylogenetic tree that has been generated. The table contains three columns:
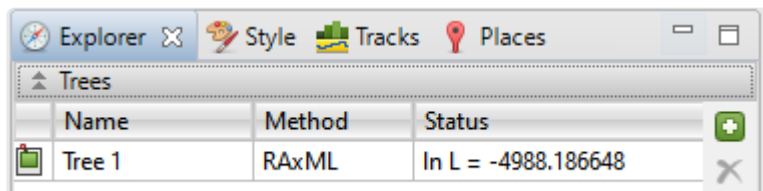
- **Name** – The default tree generated during multiple alignment is always named **Tree**. Other trees are assigned names during the use of the **Tree > Compute Phylogeny** command.

- **Method** – The algorithm used to create the tree. The default tree is always created using **Neighbor joining: BIONJ**, which is abbreviated to **BIONJ**. Other trees may use that method or the **Maximum likelihood: RAxML** method, which is abbreviated as **RAxML**.

- **Status/Completion** – Time the tree took to complete.

To the right of the table are two tools:

Icon	Name	Purpose
➕	**Computes a new phylogeny for the aligned sequences**	To create a new tree and add it to the table. This is a shortcut to the **Tree > Compute Phylogeny** command. See the topic Phylogenetic Trees for instructions on adding a tree.
✕	**Remove phylogeny**	To delete a selected tree from the list.

Style panel

The Style panel has expandable sections with options affecting the Overview, Sequences, Distance and Tree views, as well as multiple and pairwise alignments.

To open the Style panel:

Click on the **Style** tab (Style). If the panel is not already visible, reveal it using any of the following methods:

  • Choose **View > Style > Style** or **View > Reset Panel Layout**.

  • Press **Ctrl/Cmd+Alt+Y**.

  • Right-click on a track and choose **Show Style Panel**.

  • Use **Distance > Table > Show Style Panel**.
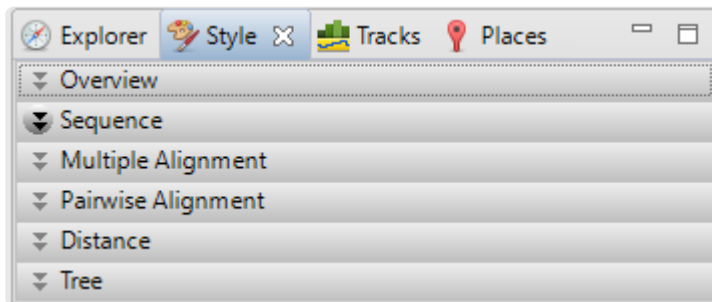
Click the links below to learn about each of the sections in the Style panel:

  • Overview section
  • Sequence section
  • Multiple Alignment section
  • Pairwise Alignment section
  • Distance section
  • Tree section

To save styles and apply them to future projects:

Once you have set up styles, you can save them and then apply them to future projects, either by default or only when you specify that they be applied. You can choose whether to save and apply styles for just one particular view or for all views.

Note that some of the instructions below refer to the "hamburger" menu. To access commands on these

menus, click on the expand bar of interest in the Style panel. Then click on the tool consisting of three horizontal lines.
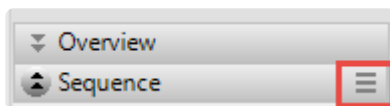
Task	How To
Save all current styles to a file that can be applied to future projects	Use **View > Document Style > Save Style**.
Save the styles from one particular view to a file that can be applied to that view in future projects	Click the "hamburger" menu on the right of a Style panel expand bar and choose **Save (Expand bar name) Style**.
Save all current styles to a file that will be applied as the default for future projects	Use **View > Document Style > Save Style as Default**.
Save the current styles from one particular view to a file that will be applied as the default for that view in future projects	Click the "hamburger" menu on the right of a Style panel expand bar and choose **Save (Expand bar name) Style as Default**.
Apply styles that were saved earlier to the current project	Use **View > Document Style > Apply Style**
Apply styles that were saved for this view previously and apply them to the same view in the current project	Click the "hamburger" menuon the right of a Style panel expand bar and choose **Apply (Expand bar name) Style**.
Restore all DNASTAR default styles to the current project	Use **View > Document Style > Restore DNASTAR Default Style**.
Restore the DNASTAR default style for one particular view to that view in the current project	Click the "hamburger" menu on the right of a Style panel expand bar and choose **Restore DNASTAR Default (Expand bar name) Style**.

Overview section

To access options affecting the <u>Overview</u> display, click on the <u>Style panel</u> expand bar entitled **Overview**, or choose **View > Style > Overview**.

Task	How to…
Increase/ decrease the amount of vertical space between sequence blocks	Use the **Spacing** slider. This command does not affect the sizes of the alignment blocks, but only the amount of white space between them. To make the alignment blocks increase/ decrease in size, instead use the Overview's <u>zoom in/out</u>.
Change font options for the ruler numbers in the Overview	Use the **Font** drop-down menu to choose a font for the ruler numbers in the Overview. Recently selected fonts (if any) appear at the top of the menu, and other available fonts appear below the row of dashes. Enter the preferred font size by typing a number between 4-72 in the box at right, or by using the corresponding up and down arrows (⮝⮟). See note below for additional clarification.
Select font options for other items in the Overview	To set a different font for sequence and track labels, first check the box next to **Track names font**. If desired, use the corresponding drop-down menu to change the font from the current value. Enter the preferred font size by typing a number between 4-72 in the box at right, or by using the corresponding up and down arrows (⮝⮟). p(banner tip). **Note:** When **Track names font** is unchecked and tracks are open, **Font** and font size (described above) affect everything in the view. When **Track names font** is checked and tracks are open, **Font** and **Font size** affect only the rulers, while **Track names font** and **Track names font size** affect sequence labels and <u>track</u> labels.

Display or hide the lines connecting blocks from one track to another	Check or uncheck the **Show connecting lines** checkbox. This option is only enabled for multi-block (Mauve) alignments.
Compare other sequences in a Mauve alignment to one specified sequence	Choose the comparison sequence from the **Compare to** drop-down menu. This option is only enabled for multi-block (Mauve) alignments, and changes the lines that connect the blocks. All blocks for the specified comparison sequence will be placed on the forward strand. Blocks for all other sequences can appear on either strand.

✳ **Note:** Once you set up styles, you can save them and apply them to future projects, either by default or when you specify.

Sequence section

To access options affecting the Sequences view and Pairwise view, click on the Style panel expand bar entitled **Sequences**, or choose **View > Style > Sequences**.

Task	How to…
Select font options for the ruler numbers in the view	Use the **Font** drop-down menu to choose from available fonts. Recently selected fonts (if any) appear at the top of the menu, and other available fonts appear below the row of dashes. Enter the preferred font size by typing a number between 4-72 in the box at right, or by using the corresponding up and down arrows ( ). To make the font **Bold** and/or *Italic*, press the **B** and/or *I* buttons, respectively.
Specify the case for all sequence letters in the view	By default, all sequence letters in the Sequence view initially appear in upper case, regardless of what case is used in the sequence files. To change this, use the **Case** dropdown menu. Choices are **Upper**, **Lower** or **Original case**. If you choose **Original case**, the letters will be shown as they appear in the original sequence files. The selection you make here applies to all sequences in the MegAlign Pro project, not only to the selected sequence(s).
Select font options for other items in the view	First check the box next to **Track names font**, then use the corresponding drop-down menu to change the font from the current value. Enter the preferred font size by typing a number between 4-72 in the box at right, or by using the corresponding up and down arrows ( ). To make the font **Bold** and/or *Italic*, press the **B** and/or *I* buttons, respectively.
Toggle between displaying as DNA/RNA	Use the **Display as** drop-down menu to choose **DNA** or **RNA**.

Note: Once you set up styles, you can <u>save them and apply them to future projects</u>, either by default or when you specify.

Multiple Alignment section

To access options affecting the display of multiple alignments in the Sequences view, click on the Style panel expand bar entitled **Multiple Alignment**, or choose **View > Style > Multiple Alignment**.

Task	How To
Designate the sequence you want to compare the other sequences to	From the **Compare to** drop-down menu, choose **Consensus** or any of the aligned sequences by name. This menu is disabled until you have performed an alignment.
Specify the wrapping style for rows	Use the **Layout** drop-down menu. Choose between: • **Linear** – All information for a sequence appears on a single row with no line breaks. A horizontal scrollbar is provided. • **Automatic wrapping** – The alignment is wrapped to fit within the view. Wrapping adjusts automatically if the view's width changes. • **Fixed wrapping** – The alignment is wrapped at the specified number of

	residues. Enter a number in the **Size** box to wrap after a different number of residues.
	Fixed wrapping ▾ Size: 50 ⬍
	Vertical splits are not available in either type of wrapped view. However, any existing splits are restored if wrapping is turned off.
Specify how to color the sequences and the consensus to facilitate visual comparisons	Use the **Alignment coloring** drop-down menu to s. Sequence comparisons can only be performed with aligned sequences. Choose between: • **Color all residues** – All residues appear in the color schemes chosen in **Foreground color** (dark 'T' icon) and **Background color** (white 'T' icon; see row below). • **Color only differences from reference** – Only residues that differ from the consensus are colored. • **Color only matches to reference** – Only residues that are identical to the reference sequence specified above are colored. • **Show only differences from reference** – Only residues that differ from the consensus are displayed and colored. All other residues are shown as uncolored dashes. If you chose this option, you should select the reference sequence from the **Reference** drop-down menu. The reference can be the consensus or any of the aligned sequences. • Choose either **Color only conserved residues** or **Color only non-conserved residues** to color according to whether or not residues are conserved. A residue is "conserved" if all sequences have the same residue in that column. Otherwise, it is "non-conserved." • If you prefer to ignore gaps when determining whether or not a residue is "conserved," check the box next to **Ignore gaps**. The **Ignore gaps** checkbox is only enabled if you selected **Color only conserved/non-conserved residues** above.
Specify the color of sequence characters	Use the **Foreground color** tools to the right of ⊤. To display the letters in a single solid color of your choice, check the upper box. Then click on the 'A' tool to the right of the checkbox and choose a color from the color picker. To instead choose a preset color scheme, check the lower box and make a selection from the drop-down menu.

Color scheme choices are described in Available color schemes.

Specify the background color (boxes around each letter)	Use the **Background color** tools to the right of ▢T. To use a single solid color of your choice, check the upper box. Then click on the 'highlighter pen' tool to the right of the checkbox and choose a color from the color picker. To instead choose a preset color scheme, check the lower box and make a selection from the drop-down menu (see image in row above). Color scheme choices are described in Available color schemes.
Specify the background color for gaps	Use the **Background color for gaps** tool to the right of ▢.To use a single solid color of your choice, check the box next to **Solid**. Then click on the 'highlighter pen' tool to the right of the checkbox and choose a color from the color picker. If you do not wish to use a background color for gaps, check the **Default** box.
Specify a font style for residues that differ from the reference	To display a specialized font, check the box next to **Differences from reference** and then choose **B** for **bold**, **I** for *italic*, and/or **U** for underline. Multiple selections can be made, if desired. Note that the font style display can be used alone, or in addition to text color and background color options described above. If you do not wish to use a specific font for differences from the reference, leave the box unchecked.
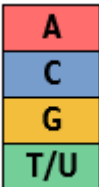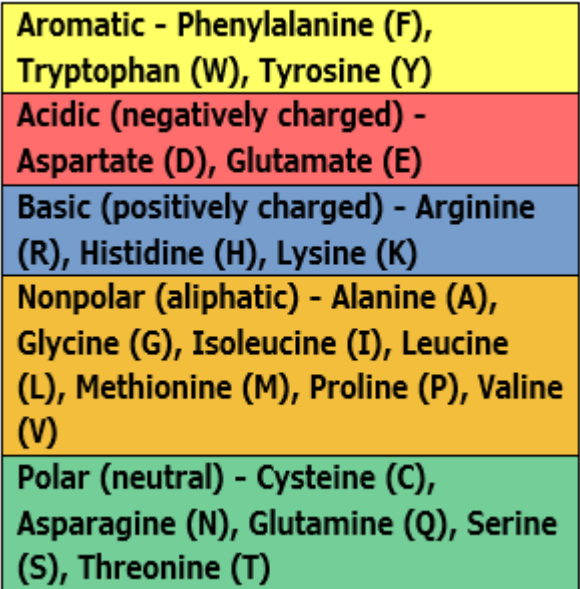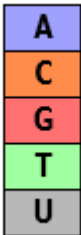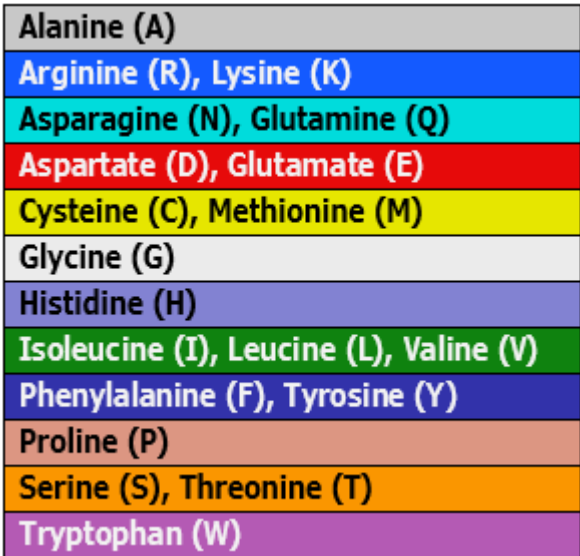
> ✳ **Note:** Once you set up styles, you can save them and apply them to future projects, either by default or when you specify.

Available color schemes

The Multiple alignment and Pairwise alignment sections allow you to select color schemes for sequences in the Sequences and Pairwise views, respectively. Each available color scheme is described below:

Color Scheme	Description	Nucleotide Legend	Amino Acid Legend
Color by Chemistry	The default color scheme used by multiple DNASTAR applications. Amino acids are colored according to their side chain chemistry. While Color by Chemistry can be used for nucleotide sequences too, we recommend instead using a **Solid** background color for nucleotides and then selecting a **Comparison** like **Color only non-conserved residues**, **Color only differences from reference**, etc.	A C G T/U	Aromatic - Phenylalanine (F), Tryptophan (W), Tyrosine (Y) Acidic (negatively charged) - Aspartate (D), Glutamate (E) Basic (positively charged) - Arginine (R), Histidine (H), Lysine (K) Nonpolar (aliphatic) - Alanine (A), Glycine (G), Isoleucine (I), Leucine (L), Methionine (M), Proline (P), Valine (V) Polar (neutral) - Cysteine (C), Asparagine (N), Glutamine (Q), Serine (S), Threonine (T)
Physiochemical: Shapely	Matches the RasMol amino acid and RasMol nucleotide color schemes, which are, in turn, based on Robert Fletterick's "Shapely models."	A C G T U	Alanine (A) Arginine (R), Lysine (K) Asparagine (N), Glutamine (Q) Aspartate (D), Glutamate (E) Cysteine (C), Methionine (M) Glycine (G) Histidine (H) Isoleucine (I), Leucine (L), Valine (V) Phenylalanine (F), Tyrosine (Y) Proline (P) Serine (S), Threonine (T) Tryptophan (W)

Physiochemical: Zappo	Colors residues according to their physico-chemical properties, and is also used in JalView (Waterhouse et al., 2009).	A C G T U	Aliphatic/hydrophobic - Alanine (A), Isoleucine (I), Leucine (L), Methionine (M), Valine (V) Aromatic - Phenylalanine (F), Tryptophan (W), Tyrosine (Y) Conformationally special - Glycine (G), Proline (P) Cysteine (C) Hydrophilic - Asparagine (N), Glutamine (Q), Serine (S), Threonine (T) Negatively Charged – Aspartate (D), Glutamate (E) Positively Charged - Arginine (R), Histidine (H), Lysine (K)
Physiochemical: UGene	Unipro UGENE coloring scheme.	A C G T/U gap	Positive (KRH) Aromatic (FWY) Large aliphatic hydrophobic (ILM) Small hydrophobic (ST) GP EX

Spectral Colors: Taylor	Taken from Taylor (1997) and also used in JalView (Waterhouse et al., 2009).	A C G T/U	Alanine (A) Arginine (R) Asparagine (N) Aspartate (D) Cysteine (C) Glutamate (E) Glutamine (Q) Glycine (G) Histidine (H) Isoleucine (I), Leucine (L) Lysine (K) Methionine (M) Phenylalanine (F) Proline (P) Serine (S) Threonine (T) Tryptophan (W) Tyrosine (Y) Valine (V)
Clustal X	Taken from Thompson et al. (1994). This color scheme is only available for protein alignments.	N/A	Hydrophobic (ACFHILMVWY) > 60% conservatism Negatively charged (DE) > 50% conservatism Postively charged (KR) > 60% conservatism Polar (STQN) > 50% conservatism Cysteine > 85% conservatism Glycine 85% conservatism Proline 85% conservatism Aromatic (FYW) 50% conservatism

Grayscale – Percentage Identity	This option uses only black/white/gray and is suitable for publications requiring grayscale images.	0-40% 41-60% 61-80% 81-100%	0-40% 41-60% 61-80% 81-100%
Percentage Identity	This option uses only shades of slate blue.	0-40% 41-60% 61-80% 81-100%	0-40% 41-60% 61-80% 81-100%

Pairwise Alignment section

To access options affecting the display of pairwise alignments in the Pairwise view, click on the Style panel expand bar entitled **Pairwise Alignment**, or choose **View > Style > Pairwise Alignment**.

Task	How To
Specify the wrapping style for rows	Use the **Layout** drop-down menu. Choose between: • **Linear** – All information for a sequence appears on a single row with no line breaks. A horizontal scrollbar is provided. • **Automatic wrapping** – The alignment is wrapped to fit within the view. Wrapping adjusts automatically if the view's width changes. • **Fixed wrapping** – The alignment is wrapped at the specified number of residues. Enter a number in the **Size** box to wrap after a different number of residues. Vertical splits are not available in either type of wrapped view. However, any existing splits are restored if wrapping is turned off.
Visualize	Check the **Show context** box. A local alignment reports the highest scoring contiguous

flanking segments of sequences that are not within the aligned segment	segment of alignment between two sequences, even if the full extent of one or both of the sequences is not included in the final alignment. Local alignments are ideal for finding short sequences within a larger sequence. When you check **Show context**, the flanking segments appear in a dimmer color scheme than the alignment. As an example, consider a local alignment of a long sequence to a short one. Any positions that aren't part of the alignment are considered to be context and are not displayed in the Pairwise view unless **Show context** is selected. In the case below, the segment "GGGC" in on the left of Seq1 is context, as is the trailing "TTAAAA". The internal part of Seq2 that spans a gap in Seq1 might be considered unaligned, but is actual context, because it's not internal to the alignment. Seq1 GGGC **CCC –– TTA** TTAAAA Seq2 –––– **CCC C C TTA** ––––––
Specify how to color the sequences and the consensus to facilitate visual comparisons	Use the **Alignment coloring** drop-down menu to s. Sequence comparisons can only be performed with aligned sequences. Choose between: • **Color all residues** – All residues appear in the color schemes chosen in **Foreground color** (dark 'T' icon) and **Background color** (white 'T' icon; see row below). • **Color only differences from reference** – Only residues that differ from the consensus are colored. If you chose this option, you should select the reference sequence from the **Reference** drop-down menu. The reference can be the consensus or any of the aligned sequences. • **Color only matches to reference** – Only residues that are identical to the consensus are colored. If you chose this option, you should select the reference sequence from the **Reference** drop-down menu. The reference can be the consensus or any of the aligned sequences. • **Show only differences from reference** – Only residues that differ from the consensus are displayed and colored. All other residues are shown as uncolored dashes. If you chose this option, you should select the reference sequence from the **Reference** drop-down menu. The reference can be the consensus or any of the aligned sequences. • Choose either **Color only conserved residues** or **Color only non-conserved residues** to color according to whether or not residues are conserved. A residue is "conserved" if all sequences have the same residue in that column. Otherwise, it is "non-conserved." • If you prefer to ignore gaps when determining whether or not a residue is

	"conserved," check the box next to **Ignore gaps**. The **Ignore gaps** checkbox is only enabled if you selected **Color only conserved/non-conserved residues** above.
Reference	Choose which sequence to use as the "Reference" in the pairwise comparison. Choices are **Target (top sequence)** or **Query (bottom sequence)**.
Specify the color of sequence characters	Use the **Foreground color** tools to the right of ⊤. To display the letters in a single solid color of your choice, check the upper box. Then click on the 'A' tool to the right of the checkbox and choose a color from the color picker. To instead choose a preset color scheme, check the lower box and make a selection from the drop-down menu. Color scheme choices are described in <u>Available color schemes</u>.
Specify the background color (boxes around each letter)	Use the **Background color** tools to the right of ⊤. To use a single solid color of your choice, check the upper box. Then click on the 'highlighter pen' tool to the right of the checkbox and choose a color from the color picker. To instead choose a preset color scheme, check the lower box and make a selection from the drop-down menu (see image in row above). Color scheme choices are described in <u>Available color schemes</u>.
Specify the background color for gaps	Use the **Background color for gaps** tools to the right of ▬. To use a single solid color of your choice, check the box next to **Solid**. Then click on the 'highlighter pen' tool to the right of the checkbox and choose a color from the color picker. If you do not wish to use a background color for gaps, check the **Default** box.
Specify the appearance of the match bar track	Use the **Match bar** section to customize display of the optional <u>Match bar track</u>. This track is used to show similarities and differences between the query and target sequences. These options are mirrored in the <u>Match Bar track options</u> section. Choices made in either location are instantly reflected in the other location. • To choose how to show the connection between the query and the target, use the **Show identity as** drop-down menu. Choosing **Residues in target case** or **Residues in query case** cause the connection to appear as a copy of the residue

from the target or query sequence. Choosing **Vertical bars** causes the connection to appear as a symbol, depending on the match status. See the small table below this one for a description of the options in the **Show identity as** drop-down menu.

- To add a background color behind each residue, check the **Color background** check box and use the adjacent drop-down menu to choose between **Same as query** and **Same as target**.

Match Bar Symbol	Interpretation	Weight table value
Vertical line	Match	1
Colon	Favorable mismatch	(0,1)
Period	Neutral mismatch	0
Space	Unfavorable mismatch	< 0

✳ **Note:** Once you set up styles, you can save them and apply them to future projects, either by default or when you specify.

Distance section

The Distance section affects the appearance of the <u>Distance view</u>. To access this section, click on the Style panel expand bar entitled **Distance**, or choose **View > Style > Distance**.

> ✱ **Note:** In Lasergene 17.1 and earlier, this section was used to specify distance parameters. Beginning in Lasergene 17.2, these are now set using the **Distance > Parameters** command.

Task	How to…
Specify which statistical values appear in the table	Use the **Values** drop-down menu in the **Upper right** and **Lower left** sections to choose from: • **Distance** – A number between 0 and 1, with 0 being an exact match and 1 indicating no matching bases at all. • **%Identity** – 100% * (Number of identical bases, excluding ambiguous bases) / (The number of aligned pairs of bases considered during computation of distance). • **%Similar** – 100% * (Number of matching bases, including ambiguous bases that may be a possible match) / (The number of aligned pairs of bases considered during computation of distance). • **%Dissimilar** – 100% – **%Similar** (see above). • **%Gaps** – 100% * (Gap count / (Gaps + Length)).

	• **Identical** – Number of identical bases, excluding ambiguous bases.
	• **Similar** – Number of matching bases, including ambiguous bases that may be a possible match.
	• **Dissimilar** – Total number of bases – **Similar** (see above).
	• **Gap count** – Number of aligned positions in the two sequences in which at least one is a gap.
	• **Gap length** – Length of regions in which at least one of the two sequences has a gap.
	• **Length** – The number of aligned pairs of bases considered during computation of distance.
	Notes:
	• You can alternatively choose the distance metrics to display using the drop-down menus in the upper left of the [Distance view](#).
	• Similarity is measured on aligned pairs of amino acids, with the help of the BLOSUM62 substitution matrix. A pair is similar if the substitution matrix has a positive value for the pair. A positive value in BLOSUM62 indicates a more likely substitution. The substitution matrix chosen for the alignment method, if any, does not affect the similarity measure in the distance table.
	• Statistics are calculated across the entire alignment, regardless of where the sequences overlap.
	• With global gap removal, positions ignored because they're global gaps are not considered in the statistics. All and only pairwise gaps are reflected.
Add heat map coloration to the table	Check the **Heat map** boxes corresponding to the **Upper right** and/or **Lower left** sections of the table.
	• To select a color for either end, click on the end of the colored bar corresponding to a section and make a selection from the color picker.
	• To flip the direction of the gradient, click on the left of the two corresponding **Flip gradient** tools (⬌). If the gradient went from white to dark blue, for example, it will now go from dark blue to white. By default, the gradient progresses linearly from 0-100%.
	• To better see lower or higher values, click on the right of the two corresponding

	Flip gradient tools ( ) and select **Distinguish lower values better** or **Distinguish higher values better**. To return to a linear progression, choose **Normal**.
Specify the Distance view font style	Use the **Font** menu to choose between seven available fonts. Recently selected fonts (if any) appear at the top of the menu, and other available fonts appear below the row of dashes.
Specify the Distance view font size	Type a number to the right of **Font** or use the corresponding up and down arrows ( ) to select a font size. Allowed values = 4 to 72. To make the font **Bold** and/or *Italic*, press the **B** and/or *I* buttons, respectively.
Specify the number of decimal places displayed in the Distance view	Type a number to the right of **Decimal places**, or use the corresponding up and down arrows ( ) to specify the number of decimal places. Allowed values = 1 to 7.

> ✳ **Note:** Once you set up styles, you can save them and apply them to future projects, either by default or when you specify.

Tree section

To access options affecting the Tree view, click on the Style panel expand bar entitled **Tree**, or choose **View > Style > Tree**.

Task	How to…
Specify the layout of the view	Click on any of the four orientations to show the phylogenetic tree with that layout (left, up, down, right).
Specify the font style for selected tree leaves	Select one or more leaves, then use the **Font** drop-down menu to choose from available fonts. Choose **More fonts** to open a popup listing additional fonts. To specify the font size, type a number to the right of **Font** or use the corresponding up and down arrows (⯅⯆). Allowed values are 4 to 72. To make the font **Bold** and/or *Italic*, press the **B** and/or *I* buttons, respectively.
Place a colored box around selected leaves	Select one or more leaves, then check the box next to **Box**. Click the first square to the right of **Box** to specify the background color of the box. Click the second square to specify a color for the box's border. Use the text box on the right to specify a thickness for the border.
Specify a color and thickness for selected branches	Select one or more branches, then click the color box to the right of **Line** and make a selection. Enter a number on the right to specify the thickness of the branch or use the corresponding up and down arrows (⯅⯆). Allowed values are 1 to 20.
Specify whether	Select one or more branches, then make a selection from the **Branch label** drop-down

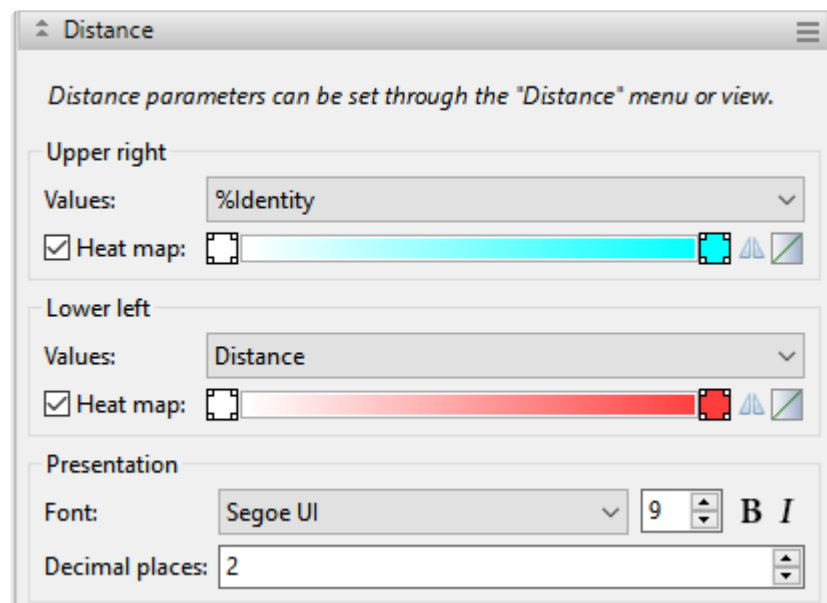|	
and how to label selected branches	menu. Options include **None**, **Distance** and **Name**.
Specify the number of decimal places for the lengths of selected branches	Select one or more branches, then type a number to the right of **Decimal places**, or use the corresponding up and down arrows ( ). Allowed values are 1-7. Note that branch lengths will only change if distances were already being displayed for those branches, as described above.
Specify the style of selected triangles that are used to represent collapsed branches	Collapse tree branches that are not of interest by hovering over a corresponding branch node and clicking the "minus sign" ( ). Uncollapse them by hovering over a triangle and pressing the "plus sign" ( ). Collapsed branches are represented by a triangle with text next to it showing the number of leaves in the branch. To change the background color or outline color for a selected triangle, click the first or second box (respectively) to the right of **Triangle** and make selections from the color chooser. Use the text box on the right to specify the width of the outline. To cause the triangle to be displayed at the same width as the collapsed branches had been, check the box next to **Width shows maximum leaf distance**. The width of the triangle before zooming is the minimum of 1) the distance from the collapsed node to the leaf farthest by distance and 2) a small constant that causes a triangle appear instead of a line. When the box is unchecked, the width of the triangle is affected only by zooming. To make the height of the triangle proportional to the number of branches it represents, check the box next to **Height varies with sequence count**. When the box is unchecked, the height of the triangle is affected only by zooming.

✳ **Note:** Once you set up styles, you can <u>save them and apply them to future projects</u>, either by default or when you specify.

Tracks panel

The Tracks panel lets you manage the tracks displayed in the Overview and the Sequences view.

To open the Tracks panel:

Click on the **Tracks** tab (Tracks). If the panel is not already visible, reveal it using any of the following methods:

- Choose **View > Tracks > Tracks** or **View > Reset Panel Layout**.

- Press **Ctrl/Cmd+Alt+T**.

- Right-click on a track in the Overview or Sequences View and choose **Show Tracks Panel**.

To change settings in the Tracks panel:

Expl... Style Tracks ✕

- ∨ ◼ Overview header
  - › ☐ Biophysics
  - › ☐ Immunogenicity
  - ☑ Reference
  - ☑ Reference Features
  - ☑ Reference Ruler
  - ☑ Ruler
- ∨ ◼ Overview details
  - › ☐ Biophysics
  - › ☐ Immunogenicity
  - ☑ Sequence Features
  - ☑ Sequence Ruler
- ∨ ◼ Sequences header or footer
  - › ☐ Biophysics
  - › ☐ Immunogenicity
  - ☐ Consensus
  - ☐ Consensus Match
  - ☑ Reference
  - ☑ Reference Features
  - ☑ Reference Ruler
  - ☑ Ruler 1
  - ☑ Ruler 2
  - ☑ Sequence Logo
- ∨ ◼ Sequence detail
  - › ☐ Biophysics
  - › ☐ Immunogenicity
  - ☐ Gap Fraction
  - ☑ Sequence Features
  - ☑ Sequence Ruler
- ∨ ☑ Pairwise
  - ☑ Match Bar
  - ☑ Ruler
- ∨ ◼ Pairwise details
  - ☐ Gap Fraction
  - ☑ Sequence Features
  - ☑ Sequence Ruler

The uppermost part of the Tracks panel lets you specify which tracks are visible, and configure their options and appearance. Depending on the sequence type (nucleotide vs. amino acid) and alignment status, different sub-sets of track types will be listed. For instance, prior to alignment, **Consensus**, **Consensus Match**, **Sequence Logo** and **Gap Fraction** are hidden from the list.

By default, some tracks on the list are checked (i.e., available to display in the Sequences view) while others are unchecked. Put check marks next to tracks that you wish to make available for display, and remove check marks for tracks that you wish to hide. Tracks are added to the views in the same order in which their boxes are checked in the Tracks panel. This means that the position of a track may change when its box is unchecked and then checked again.

Click the links below to learn about each of the expandable sections in the Tracks panel:

- [Layout](#)
- [Options](#)

To see the detail tracks that have been check-marked for display:

Expand one or more sequences in the Overview and/or Sequences view by clicking the relevant plus sign icon. ( )

Layout section

To access options affecting the height of a track and the space above and/or below the track, click on the Tracks panel expand bar entitled "Layout," or choose **View > Tracks > Layout**. Changes made in this section are applied to all currently selected tracks.
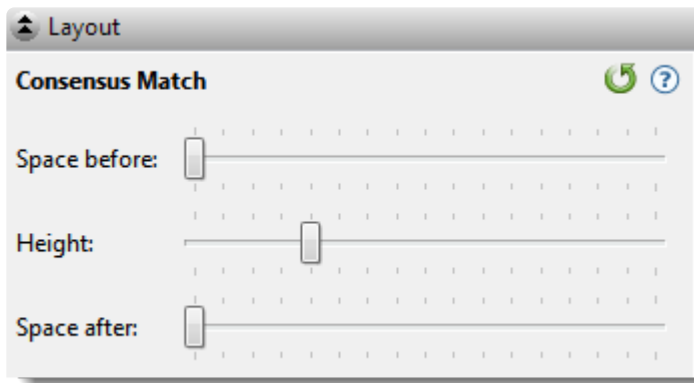
Task	How to…
To control the amount of blank space above and/or below each of the selected track(s)	Use the **Space before/after** sliders.
To control the heights of the selected tracks	Use the **Height** slider. This option is enabled only when all selected tracks support user-scalable height.
To return the sliders in this dialog to their default settings	Click the **Restore default settings** tool ( ).

Options section

To access settings for the track type currently selected in the [Tracks panel](#), click on the Tracks panel expand bar entitled "Options," or choose **View > Tracks > Options**.

The **Options** section varies in appearance depending upon the selection. For discussions of each track type and its corresponding Options section, see [Tracks](#).

Features panel

The Features panel lets you specify which feature types should be visible in the <u>Overview</u> and the <u>Sequences view</u> once a sequence has been expanded.

<u>To open the Features panel:</u>

Click on the **Features** tab (Features). If the panel is not already visible, reveal it using any of the following methods:

- Choose **View > Features** or **View > Reset Panel Layout**.

- Right-click on a track in the Overview or Sequences view and choose **Show Features Panel**.

<u>To change settings in the Features panel:</u>

Type	Count
☑ CDS	9,289 items
☑ gene	9,712 items
☑ misc_feature	90 items
☑ misc_RNA	2 items
☑ rRNA	38 items
☑ source	6 items
☑ STS	3 items
☑ tRNA	216 items
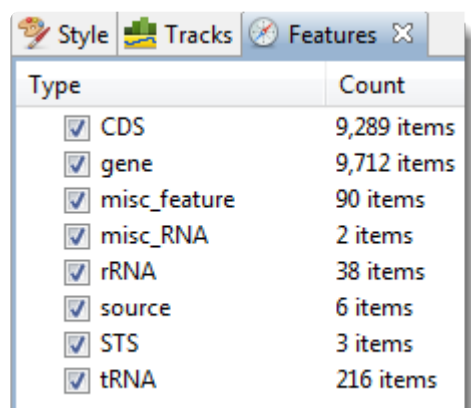
The **Type** column shows the feature type. By default, all feature types except **Source** are checked.

The **Count** column shows the number of each feature type in the project. It includes all sequences, whether aligned or unaligned, selected or unselected. The count includes copies of features mapped to the consensus (e.g., with **Features > Map Features to Active Consensus**) and other sequences.

Task	How to…
To display/hide a particular feature type	Check/uncheck its box in the **Type** column.
To display/hide all features	Right-click anywhere in the Features panel and select **Show All** or **Hide all**.

Places panel

The Places panel lets you quickly navigate to a sequence location or range that you have visited before. The following video shows how to bookmark favorite locations using the Places panel:

To open the Places panel:

Click on the **Places** tab (📍 Places). If the panel is not already visible, reveal it using any of the following methods:

- Choose **View > Places > Places** or **View > Reset Panel Layout**.

- Press **Ctrl/Cmd+Alt+P**.

To change settings in the Places panel:
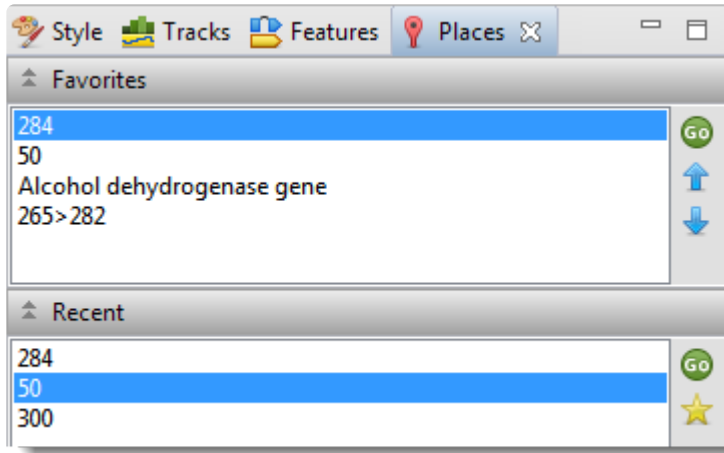
Click the links below to learn about each of the expandable sections in the Places panel:

- [Recent section](#)

- [Favorites section](#)

Favorites section

The Favorites section is similar to the bookmarks folder for a web browser. It contains a list of previously-visited locations that you have decided you may wish to visit again.

To access the Favorites section:

Click on the Places panel expand bar entitled "Favorites," or choose **View > Places > Favorites**.
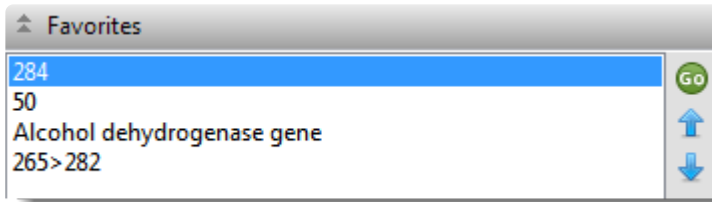
Entries are nominated for this list from within the Recent section of the Places panel or by chooseing **Sequences > Add Place to Favorites**. Favorites are listed in the same order in which they were added, by default, though you can change the order manually (see table below). The Favorites list is saved whenever you save a project.

Task	How to…
To return to a location in this section	Double-click on the item or make a selection from the list and press the **Go to this gene or location** button (Go).
To move a bookmarked location further up the Favorites list	Make a selection from the list and press the **Move this place up** arrow (⬆).
To move a bookmarked location further down the Favorites list	Make a selection from the list and press the **Move this place down** arrow (⬇).
To rename a selected item	Click on an item in the Favorites section to select it. Then right-click on the same item and choose **Rename Favorite**. Type a name into the box and click **OK**.
To remove a selected item from the Favorites list	Click on an item in the Favorites section to select it. Then right-click on the same item and choose **Remove Favorite from This List**.
To remove **all** items from the list	Right-click in the Favorites section and select **Clear Favorites**.

Recent section

When you perform a search for a gene or numerical sequence location using **Edit > Go To Position**, each found location is automatically stored in the Recent section. Access this section by clicking on the Places panel expand bar entitled "Recent," or by choosing **View > Places > Recent**.

Recent locations are listed in the same order in which they were added, and the list is saved whenever you save a project.

Task	How to…
To return to a location on the list	Double-click on the location name, or make a selection from the list and press the **Go to this gene or location** button ( ).
To save a location on the list to the Favorites section (similar to bookmarking a page on the Internet)	Make a selection from the list and press the **Add this place to Favorites** button ( ). Alternatively, make a selection from the list, then right-click on the item and choose **Add This Place to Favorites**.
To remove all items from the Recent list	Right-click in the Recent section and select **Clear Recent**.

Details panel

The Details panel shows details of the application, selection or project.

To open the Details panel:

This panel is represented by a blue "information" icon (🛈). If the panel is not already visible, reveal it using any of the following methods:

- Choose **View > Details** or **View > Reset Panel Layout**.

- Right-click on a track in the [Overview](#) or [Sequences view](#) and choose **Show Details Panel**.

Interpreting the Details panel display:
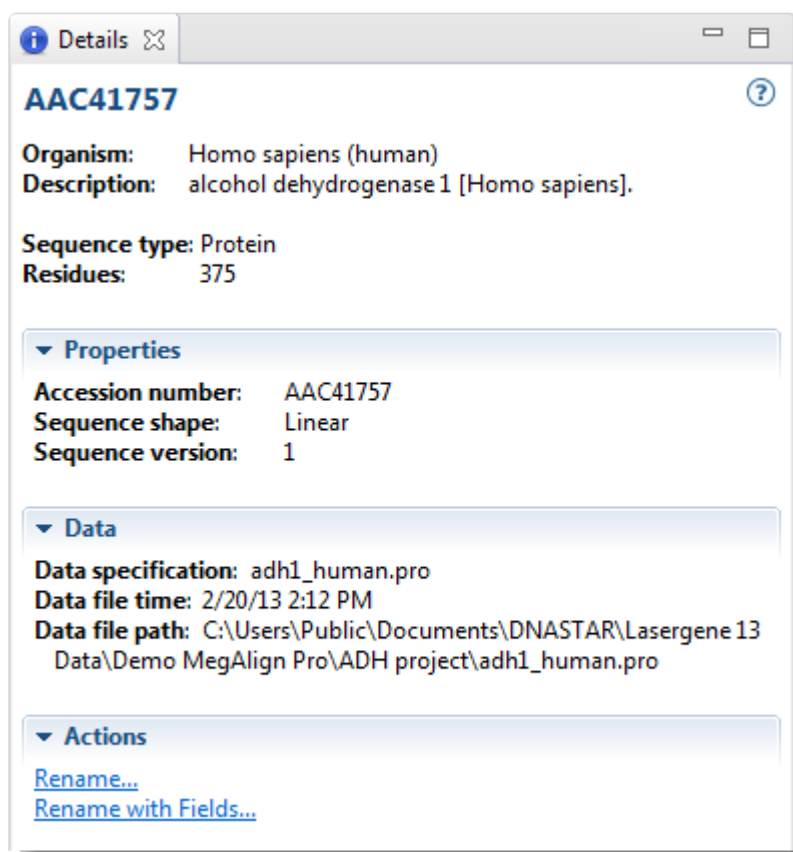
Details are available for all tracks in the Overview and Sequences View, as well as the [Tracks panel](#). The panel shows actions to get started when there is no project open, or when the active project is empty. If there is a selection, the panel shows details about the selected item(s). Otherwise, the panel shows details about the project. For detailed information, see the following sub-topics:

- [When no document, or a blank document, is open](#)

- [When a non-empty document is open](#)

- [When one sequence is selected](#)

- [When a portion of a sequence is selected](#)

- [When multiple sequences are selected](#)

- [When a portion of multiple sequences is selected](#)

- [When one track is selected](#)

- [When multiple tracks are selected](#)

- [When one feature is selected](#)

- [When multiple features are selected](#)

- [When a portion of a pairwise alignment is selected](#)

When no document, or a blank document, is open

When no document or only a blank document (one containing no sequences) is open, the Details panel displays links to help you get started in MegAlign Pro.

Heading	One or more links are available. The first three options are available for blank documents, but unavailable when no document is open:
Getting Started in MegAlign Pro Create a new alignment project Open a file Import an alignment Add sequences Add sequences end-to-end Watch a training video	• Create or open a project • Import an alignment • Add sequences to the project (or add them end-to-end). • Watch a training video
Recent Documents ▼ Recent Documents RP1 project.msa ADH Project.msa Drosophila project.msa ADH project.msa	Open a recent document.

When a non-empty document is open

If a document is open and contains sequences (whether they have been aligned or not) and nothing is selected, the Details panel contains the following information:

Heading	
ADH Project.msa Sequence type: Protein	• The project filename. • **Sequence type** – DNA, RNA, or Protein.
Alignment ▼ Alignment **Alignment:** Clustal Omega **Sequences:** 5 **Sequence type:** Protein **Min. sequence length:** 348 **Max. sequence length:** 377	This section is only displayed for, and only pertains to, sequences in the project that have already been aligned. • **Alignment** – Alignment algorithm used (e.g., MAFFT, MUSCLE). • **Sequences** – Number of sequences in the alignment. • **Sequence type** – DNA or protein. • **Min.** and **Max.** sequence lengths – Lengths of the shortest and longest sequences in the alignment.
Unaligned Sequences ▼ Unaligned Sequences **Sequences:** 5 **Sequence type:** Protein **Min. sequence length:** 348 **Max. sequence length:** 377	This section is only displayed for, and only pertains to, sequences in the project that are not included in the alignment. • **Sequences** – Number of unaligned sequences. • **Sequence type** – DNA, RNA, or Protein. • **Min.** and **Max.** sequence lengths – Lengths of the shortest and longest unaligned sequences.
Actions ▼ Actions Add sequences Add sequences end-to-end Align with options	• **Add sequences** – Shortcut to **File > Add Sequences** (see Add ungapped / gapped sequences to a project). • **Add sequences end-to-end** – Shortcut to **File > Add Sequences End-to-end** (See Add multi-segment sequences to a project). • **Align with options** – Shortcut to **Align > Align with Options**.

When one sequence is selected

If a single sequence has been selected in any view, the [Details panel](#) may contain a subset of the following information:

Heading ![AAC41757] Organism: Homo sapiens (human) Description: alcohol dehydrogenase 1 [Homo sapiens]. Sequence type: Protein Residues: 375	The default name, if present, appears in blue above 'Organism.' • **Organism** – The organism from which the sequence originated, if available. • **Description** – A description of the selected sequence, if available. • **Sequence type** – **DNA**, **RNA**, or **Protein**. • **Nucleotides/Residues** – Number of nucleotides or residues in the selected sequence.
Properties ▼ Properties Accession number: AAC41757 Sequence shape: Linear Sequence version: 1	• **Accession number** – The NCBI accession number of the selected sequence. • **Sequence shape** – **Linear**, **Circular**, or **Unknown**. • **Sequence version** – The NCBI version number (the part following the decimal after the accession number).
Data ▼ Data Data specification: adh1_human.pro Data file time: 2/20/13 2:12 PM Data file path: C:\Users\Public\Documents\DNASTAR\Lasergene 13 Data\Demo MegAlign Pro\ADH project\adh1_human.pro	• **Data specification** – The name of the sequence file and the "sequence name within the file," followed by the file extension. • **Data file time** – The time the sequence file was created or modified. • **Data file path** – The path to the original sequence file on your computer.
Actions	• **Rename** – Shortcut to **Edit > Rename** (See

Rename sequences manually).

- **Rename with Fields** – Shortcut to **Edit > Rename with Fields** (See Rename sequences automatically using specified data fields).

When a portion of a protein sequence is selected

When a portion of a protein sequence has been selected in any view (e.g., by double-clicking on a feature, or by dragging across the sequence with the mouse) except the Pairwise view, the Details panel contains the following information:

Heading	
Selected Region Sequence: Electric Eel Calmodulin.pro Selection: 1>149 Alignment selection: 1>149	• **Sequence** – The name of the sequence from which the region was selected.This is not shown for pairwise alignment selections. • **Alignment/Sequence selection** – **(Pairwise) Alignment selection** displays the inclusive coordinates of the beginning and end of the selection. When the sequence is not part of an alignment, (i.e., is listed with 'unaligned sequences'), then **Alignment selection** is replaced with **Sequence selection**. In this case, the coordinates are the ungapped sequence coordinates rather than the alignment coordinates.
Biophysical Properties **▾ Biophysical properties** MW (g/mol): 16,809.573 1µg to pmol: 59.49 Net charge (pH=7): -23.875 pl: 4.087 Avg. hydropathy (GRAVY): -0.65 Aliphatic index: 65.503 Instability index: 28.207 1 mg/mL to A$_{280}$: 0.177 1 A$_{280}$ to mg/mL: 5.641 ε$_{280}$ (M^{-1} cm^{-1}): 2,980 1 mg/mL to A$_{280}$ (red.): 0.177 1 A$_{280}$ to mg/mL (red.): 5.641 ε$_{280}$ (M^{-1} cm^{-1}) (red.): 2,980	Displays the physical properties of a continuous selection, including: • **MW [g/mol]** – The molecular weight of the selected region. • **1µg to pmol** – The number of picomoles of protein in 1 microgram. • **Net charge (pH=7)** – The pH-dependent sum of charges in a population of molecules. MegAlign Pro calculates the net charge for protein regions and amino acid residues, and assumes a pH of 7. • **pl** – The isoelectric point (the pH at which the residue carries no net electrical charge), calculated from pKa tables from Lehninger et al. (2005). This calculation is available only for amino acid selections. • **Average hydropathy (GRAVY)** – The Kyte-Doolittle hydropathy value, calculated using the method of Kyte, J. and Doolittle, R.F. (1982). This calculation is available only for amino acid selections.

- **Aliphatic index** – The relative volume occupied by aliphatic side chains (alanine, valine, isoleucine and leucine). This calculation is available only for amino acid selections and is made using the method of Gasteiger et al. (2005).

- **Instability index** – An estimate of the stability of the protein in a test tube. A protein whose instability index is smaller than 40 is predicted as stable, while a value above 40 predicts that the protein may be unstable. This value is calculated using the method of Guruprasad et al. (1990).

- **1 mg/mL to A$_{280}$** – Calculates protein concentration from absorbance, where **A$_{280}$** is the absorbance (optical density) for the residue. This calculation is affected by presence of disulfide bonds, and assumes the disulfide bonds are oxidized (formed). This calculation is made using the method of Gasteiger et al. (2005).

- **1 mg/mL to A$_{280}$ (red.)** – Same as above, but assumes all disulfide bonds have been reduced (broken).

- **A$_{280}$ to 1 mg/mL** – Calculates the absorbance of a protein concentration, where **A$_{280}$** is the absorbance (optical density) for the residue. This calculation is affected by presence of disulfide bonds and assumes the disulfide bonds are oxidized (formed). This calculation is made using the method of Gasteiger et al. (2005).

- **A$_{280}$ to 1 mg/mL (red.)** – Same as above, but assumes all disulfide bonds have been reduced (broken).

- **ε$_{280}$ [M^{-1}cm^{-1}]** – The extinction coefficient (amount of light a residue absorbs). This calculation assumes all Cys residues are reduced and all disulfide bonds are oxidized (formed). This calculation is available only for amino acid selections and is made using the method of Gasteiger et al. (2005).

- **ε$_{280}$ [M^{-1}cm^{-1}] (red.)** – Same as above, but assumes all disulfide bonds have been reduced (broken).

Note: For **aliphatic index**, **absorbance** and **extinction coefficient**, the

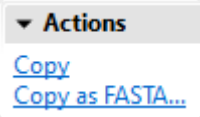|	
	sequence is treated as independent residues. There is no algorithmic dependence on length.
Actions ▼ **Actions** Copy Copy as FASTA...	• **Copy** – Shortcut to **Edit > Copy** (See Copy, Paste and Delete). • **Copy As FASTA** – Shortcut to **Edit > Copy As FASTA** (See Copy, Paste and Delete).

When a portion of a nucleotide sequence is selected

When a portion of a sequence has been selected in any view (e.g., by double-clicking on a feature, or by dragging across the sequence with the mouse) except the Pairwise view, the Details panel contains the following information:

Heading	• **Sequence** – The name of the sequence from which the region was selected.This is not shown for pairwise alignment selections.
Selected Region Sequence: XM_541769 Alignment selection: (1,538>1,555)	• **Alignment/Sequence selection** – **(Pairwise) Alignment selection** displays the inclusive coordinates of the beginning and end of the selection. When the sequence is not part of an alignment, (i.e., is listed with 'unaligned sequences'), then **Alignment selection** is replaced with **Sequence selection**. In this case, the coordinates are the ungapped sequence coordinates rather than the alignment coordinates.
Actions ▼ Actions Copy Copy as FASTA...	• **Copy** – Shortcut to **Edit > Copy** (See Copy, Paste and Delete). • **Copy As FASTA** – Shortcut to **Edit > Copy As FASTA** (See Copy, Paste and Delete).

When multiple sequences are selected

If two or more sequences have been selected in any view, the Details panel contains the following information:

Heading	
2 sequences Sequence type: Protein Min. sequence length: 348 Max. sequence length: 375	• The number of sequences selected (in blue). • **Sequence type** – **DNA**, **RNA** or **Protein**. • **Min. and Max. sequence lengths** – Lengths of the shortest and longest sequences selected, in nucleotides or residues.
Actions ▼ Actions Rename with Fields...	• **Rename with Fields** – Shortcut to **Edit > Rename with Fields** (see Rename sequences automatically using specified data fields).

When a portion of multiple sequences is selected

When a portion of multiple aligned sequences has been selected (e.g., by triple-clicking on a feature, or by dragging across the consensus sequence with the mouse), the Details panel contains the following information:

Heading	
Selected Region Sequences: 4 Alignment selection: 2,148>2,173	• **Sequences** – The number of sequences in the alignment. • **Alignment selection** – The inclusive coordinates of the beginning and end of the selection.
Actions ▼ Actions Copy Copy as FASTA...	• **Copy** – Shortcut to **Edit > Copy** (see Copy, Paste and Delete). • **Copy As FASTA** – Shortcut to **Edit > Copy As FASTA** (see Copy, Paste and Delete)

When one track is selected

If a single track has been selected in the Overview or Sequences view, the Details panel contains the following information.

Heading	• Type of track (in blue). • **Summary** – Basic information about what is displayed in the selected track.
Actions 	Only a subset of links may be available, depending on the selected track. • **Rename** – Shortcut to **Edit > Rename** (See Rename sequences manually). • **Rename with Fields** – Shortcut to **Edit > Rename with Fields** (See Rename sequences automatically using specified data fields). • **Show track layout** – Shortcut to **View > Tracks > Layout**. • **Show track options** – Shortcut to **View > Tracks > Options**.

For additional information about using tracks, see Tracks panel and Tracks.

When multiple tracks are selected

If two or more tracks have been selected in the <u>Overview</u> or <u>Sequences view</u>, the <u>Details panel</u> contains the following information.

Heading	
2 tracks ⑦ **Summary:** Multiple tracks are selected. They can be moved or hidden together.	• Number of tracks currently selected (in blue). • **Summary** – Description of options available for the multiple track selection.
Actions	
▼ Actions Show tracks panel	• **Show tracks panel** – Shortcut to **View > Tracks > Tracks**.

For additional information about using tracks, see the topics <u>Tracks panel</u> and <u>Tracks</u>.

When one feature is selected

If a single feature has been selected in a track in the Overview or Sequences view (e.g., by clicking one time on the feature), the Details panel contains the following information:

Heading	
alcohol_DH_class_I_II_IV Type: Region Left: 3 Right: 375 Length: 373	• Feature name (in blue). • **Type** – Feature type. • **Left** and **Right** –The inclusive coordinates of the beginning and end of the feature range. • **Length** – Length of the selected feature in nucleotides or residues.
Location ▼ Location 3>375	Range of the feature in nucleotides or residues.
Properties ▼ Properties db_xref: taxon:10090 organism: Mus musculus chromosome: 3 map: 3 strain: C57BL/6	Properties shown vary by feature and may include the organism, gene name NCBI feature qualifiers, etc.
Notes ▼ Notes note: class I, II, IV alcohol dehydrogenases; cd08299	• **note** – The NCBI feature qualifier of the same name.
Actions ▼ Actions Select Subsequence Select Subsequences Across Alignment *You can also double- or triple-click on this feature.*	• **Select Subsequence** – Shortcut to **Edit > Select Subsequence** (see Make Selections). Double-clicking on the feature fulfills the same function. • **Select Subsequences Across Alignment** – Shortcut to **Edit > Select Subsequences Across Alignment** (see Make Selections). Triple-clicking on the feature fulfills the same function. This item is not available for features in unaligned sequences.

For additional information about working with features, see Features panel and Features.

When multiple features are selected
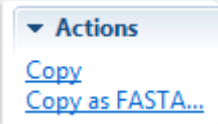
If two or more features have been selected in a track in the <u>Overview</u> or <u>Sequences view</u>, the <u>Details panel</u> contains the following information:

Heading	
3 features Type: Region, source, Protein Left: 1 Right: 374 Min. Length: 368 Max. Length: 374	• Number of features currently selected (in blue). • **Type** – List of all feature types represented in the selection. • **Left** and **Right** – The inclusive coordinates of the beginning and end of the feature range. • **Min.** and **Max. Length** – Lengths of the shortest and longest features in the selection.
Properties ▼ Properties db_xref: taxon:10090, GeneID:15465, MGI:107619	Properties, if shown in this case, will vary by feature and may include the organism, gene name NCBI feature qualifiers, etc.

For additional information about working with features, see <u>Features panel</u> and <u>Features</u>.

When a portion of a pairwise alignment is selected

When a portion of a sequence has been selected in the Pairwise view (e.g., by double-clicking on a feature, or by dragging across the sequence with the mouse), the Details panel contains the following information:

Heading	
Selected Region ? Pairwise alignment selection: 19>55	• **Selected Region** – (in blue) Unchangeable default value. • **Pairwise alignment selection** – The inclusive coordinates of the beginning and end of the selection.
Actions ▼ Actions Copy Copy as FASTA...	• **Copy** – Shortcut to **Edit > Copy** (See Copy, Paste and Delete). • **Copy As FASTA** – Shortcut to **Edit > Copy As FASTA** (See Copy, Paste and Delete).
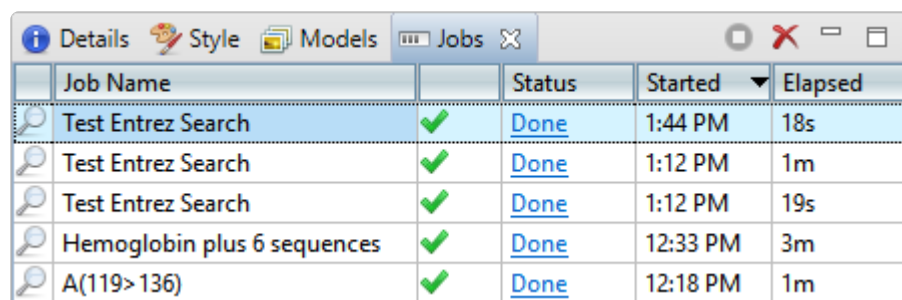
Jobs panel

After initiating an NCBI BLAST or Entrez search, you can monitor its progress and download its eventual output from the Jobs panel. This panel is not visible in a project until you have started running at least one search.

To open the Jobs panel:

Do any of the following:

- Click on the **Jobs** tab.
- Press **Ctrl+Alt+J** (Win) or **Option+Cmd+J** (Mac).
- Use **View > Jobs** or **Search > Show Jobs**.

The Jobs panel consists of a header and a table with the following columns:

Column	Description
Job Name	Text searches are all named with the field and term of the first search criterion followed by "plus # terms," where # represents the number of additional criteria. Sequence searches use the name specified in the Job screen of the Search wizard. The same **Job Name** can be used for more than one prediction.
Status	A word and icon indicating the status of the search. • **Started** – **Run** or **Run Job** were pressed in the Search wizard • **Queued** – ▶ – Submitted and placed into the queue • **Running** – ▶ – Running • **Done** – ✔ – Job complete; results available • **Failed** – ✖ – Job failed • **Stopped** – ✖ – Job stopped by user
Started	The date and time the job began.
Elapsed	The total elapsed time between starting and completing the job.

Tasks that pertain to the Jobs panel:

Task	How To
To open the results of one or more jobs	Click on the **Done** (Done) or **Expiring** links for that job. See below this table for more information on job expiration. You can also open a saved *.search* file using **Search > Open Search File**.
To stop one or more jobs that are in progress	Select one or more rows with a Status of **Started**, **Queued** or **Running**. Then press the **Stop** tool (🛑) in the header above the table. Once stopped, a job cannot be restarted from the Jobs panel. You must instead initiate a new search.
To delete a finished job	Select one or more rows with a Status of **Done**, **Failed** or **Stopped**. Then press the **Delete** tool (❌) in the header above the table.
To open/reopen the progress dialog for a job	Select one or more rows with a Status of **Started**, **Queued** or **Running**. Then right-click and choose **Open Progress dialog**.

Sequence (BLAST) and text (Entrez) search jobs are intended to be ephemeral and expire after 7 days. The Jobs panel notifies you of upcoming expiration dates.

All jobs are recorded in the location specified in the Temporary File preferences dialog. If you are a Macintosh user and elect to keep the default save location, the operating system may remove jobs after a reboot. To avoid this, we recommend choosing a different location in which to save temporary files.
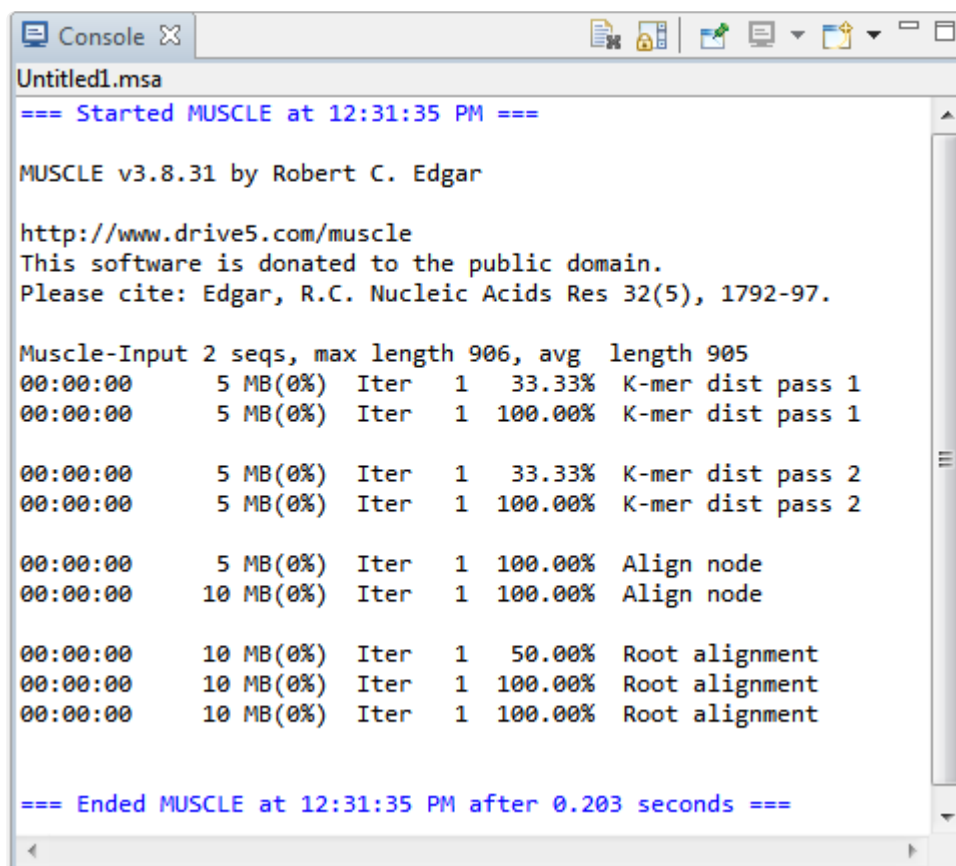
Console panel

The Console panel displays information, messages and warnings about the operation in progress, or an operation that was recently performed.

To display the console for a particular project:

Click the tab with the project name, then open its console using the **View > Console** command. Each view will have a different console.

To use the Console panel:

Each console has a maximum number of characters. If this maximum is exceeded, the oldest characters disappear in order to make room for new characters.

Right-click on the Console tab to access commands allowing you to move, hide, rearrange, or restore panels and views. (See Customize the layout of other window components).

The following tools are available in the upper right of the Console panel:

Tool	Name	Description
	Clear Console	Clears all text from the active Console.
	Scroll Lock	Stops auto-scrolling caused by alignment output. When alignment output is long, this may fail due to a limit on the amount of text in the console.
	Word Wrap	Controls whether or not lines wider than the console panel should be wrapped.
	Pin Console	Keeps the current Console active even if another project is selected.
	Display Selected Console	Consoles can be displayed for one or more open projects. If more than one alignment is in progress, clicking the arrow to the right of this icon opens a drop-down menu from which the desired Console can be brought to the front.
	Open Console	Opens a new Console. If more than one alignment is in progress at the same time, you can open a new Console and pin it to a particular run using Pin Console (above).
	Minimize	Minimizes the Console.
	Maximize	Maximizes the Console.

Customize the Look and Layout

The following topics describe the many ways you can customize the look and layout of MegAlign Pro.

- [Use the Style panel to change the look of a view](#) (font styles and sizes, colors, number of decimal places, etc.)

- [Customize the appearance and layout of views](#)

- [Customize the layout of other window components](#)

- [Apply information tracks to a view](#)

Once you have customized the look and layout as desired, you can:

- [Save Styles for individual views and apply them to other projects](#)

- [Save the entire layout and apply it to other projects](#)

Customize the appearance and layout of views

The following table lists ways in which you can customize the appearance and layout of <u>views</u>.

Task	How to…
Show a view	Use **View > (View Name) > Show**.
Change the style (fonts, colors, calculation metrics and more)	Use the <u>Style panel</u> and its subsections. Customization for font style and size applies to all text in a view. Individual rows cannot have different styles.
Move one or more sequence rows as a group	Sequence rows in some views may be rearranged one at a time or as contiguous groups using drag and drop. First, select the desired rows using **Ctrl/Cmd+click** or **Shift+click**. Next, drag the selected row(s) to the desired location and release the mouse button. Any ordering changes made to one view are simultaneously made to the other.
Detach a view	Only certain views can be detached. To do this, right-click on the view tab and choose **Detach (View Name)**. Detaching a view turns it into a free-floating window. Note that each detached view has menus in the upper left corner. For example, if you detach a Tree view and then wish to use commands such as **Tree > Root On**, you must use the menu in the detached view rather than the menu in the main MegAlign Pro window.
Reattach all detached views to their original locations	Use **View > Reattach All Views**.
Show the active view in full-screen mode	Use the **View > Full Screen** command or the **Full screen** tool ⬚ at the bottom right corner of the view, or press **Ctrl/Cmd+Shift+F**.
Exit from full-screen mode	Click the **Esc** key, use the **Exit Full Screen** tool (⬚) at the bottom of the full-screen view, or press **Ctrl/Cmd+Shift+F**.
Configure the number and location of visible views.	Use the **View > Document Layout** command or the **Change layout** tool (image varies) in the bottom right corner of the view to access possible layouts. The following options are available: • **One** – ⬚ One – To show only the current view.

	• **Two Horizontal** (<kbd>Two Horizontal</kbd>) – To show the current view on top and one additional view below it. • **Two Vertical** (<kbd>Two Vertical</kbd>) – To show the current view on the left, and one additional view to the right. • **Three Horizontal** (<kbd>Three Horizontal</kbd>) – To show the current view on top and two additional views below it. • **Three – Left** (<kbd>Three - Left</kbd>) – To show the current view on the left and two additional views stacked one on top of the other on the right. • **Three – Right** (<kbd>Three - Right</kbd>) – To show the current view on the right and two additional views stacked one on top of the other on the left. • **Three – Top** (<kbd>Three - Top</kbd>) – To show the current view on the top and two additional views side-by-side on the bottom. • **Three – Bottom** (<kbd>Three - Bottom</kbd>) – To show the current view on the bottom and two additional views side-by-side on the top.
Zoom in or out on a view	Some views have independent vertical and horizontal zoom in/out sliders. Grab any green slider with your mouse and drag left/right or up/down to zoom in or out in the view. For details, see the table just below this one.
Split a view vertically	Use **Overview > Split View** or **Analysis > Split View**. The split can be moved by dragging it left or right with the mouse. Alternatively, you can drag the vertical splitter tool (‖), found at the right edge of the view, in a leftward direction.
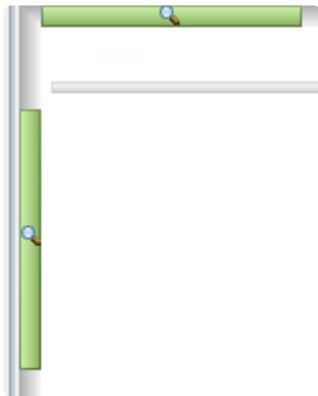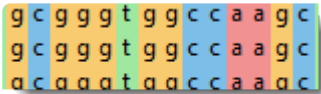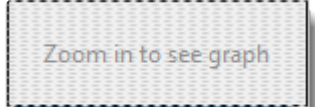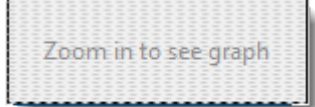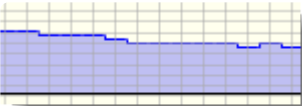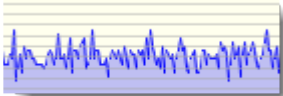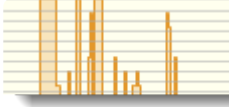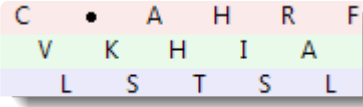
	When the Analysis view is split vertically, the Overview provides a separate viewport indicator for each section. Splits are ignored when printing or exporting an image, although the layout is preserved.
Remove the right-most vertical split	Use **Overview > Remove Split** or **Analysis > Remove Split**. Or drag the divider as far as you can to the side of the view and drop it there.
To view two or more documents simultaneously on a split screen	By default, only one document at a time is displayed, even if several documents are open. To view two or more documents simultaneously on a split screen, drag a project tab in a downward, right or left direction. When you see a black arrow, release the mouse button. Documents can be re-docked in their original positions by reversing this process.
Return to the default window display	Choose **View > Reset Panel Layout**.

Three of MegAlign Pro's views (Sequences, Tree and Overview) all have independent vertical and horizontal zoom in/out sliders. The following table uses the Sequences view as an example to show the look of the sequences and some of the detail tracks at different zoom levels.

Track	Maximum Zoom	Medium Zoom	Minimum Zoom
Sequences		Once the letters disappear, only the background color, if any, is shown.	At minimal zoom levels, sequences are shown as colored bars with white gaps.
Sequence Logo		Once the logo letters disappear, only the background color, if any, is shown.	Zoom in to see graph. Some tracks cannot be displayed at low zoom levels, and prompt you to zoom in.
Consensus Match			Zoom in to see graph

GC Content			Some tracks cannot be displayed at low zoom levels, and prompt you to zoom in.
Gap Fraction			
Translation			

Customize the layout of other window components

Controls for this application's view, panels and expand bars are described in the following table. For information that pertains only to views, also see [Customize the appearance and layout of views](#).

Task	"How To"
Show or open a hidden item	Use the associated **View** menu command, click on a closed expand bar, click on a plus sign, or click the **Restore** icon ( ).
Show/Hide the side panels	Double-click on the project name to toggle between showing and hiding the side panels on the right of the window.
Display panel sub-sections	Click on a closed expand bar (medium gray bar) or use the associated **View** menu command. If you can't see all the information in an expanded section, close other expanded sections and/or increase the size of the whole panel as described elsewhere in this table.
Hide panel sub-sections	Click on an open expand bar.
Maximize an item	use the **Maximize** icon (.
Change the relative sizes of panes	Drag the divider bar between them to the left, right, up or down.
Hide or close a visible item	Use the **Close** icon ( ) or **Minimize** icon ( ), click on an open expand bar, or click on a minus sign. To close one or more tabs, right-click on a project, view or panel tab and choose **Close**, **Close Others**, **Close Tabs to Right**, or **Close All**. Some of these options are only available for certain tabs.
Move a view, panel or project to another location	Right-click on the view tab and choose **Move (View Name) to (Location)**. In some cases, you can use your mouse to grab a project, view or panel tab, then drag and drop it in the desired location.
Scroll vertically	Use the vertical scrollbar or the mouse wheel.
Scroll horizontally	Use the horizontal scrollbar, or hold down the **Alt** key (Win) or **Option** key (Mac) and drag the mouse.

Return to the default side panel layout	Choose **View > Reset Side Panel Layout**. This command restores the default layout of panels, including opening any default panels that have been closed; and restoring panels that have been minimized or maximized. It will not reattach a detached view. In that case, use **View > Reattach All Views** instead.
Return to the default layout for the entire window	Choose **View > Document Layout > Restore DNASTAR Default Layout**.

Save the layout and apply it to other projects

You can change the layout of MegAlign Pro's views and panels as described in <u>Customize the look and layout</u>.

Once you have arrived at a layout you like, you can save it and then apply it to future projects, either by default or only when you specify that it be applied.

Task	How To
Save the current layout to a file that can be applied to future projects	Use **View > Document Layout > Save Layout**.
Save the current layout to a file that will be applied as the default layout for future projects	Use **View > Document Layout > Save Layout as Default**.
Apply a layout that was saved earlier to the current project	Use **View > Document Layout > Apply Layout**.
Restore the DNASTAR default layout to the current project	Use **View > Document Layout > Restore DNASTAR Default Layout**.

Sequences

You can add the following types of sequences to a MegAlign Pro project:

- [Ungapped sequences, including SeqNinja files](#)

- [Gapped sequences](#)

- [Only sequences that meet filtering thresholds](#)

- [Multi-segment sequences](#)

- [Only the portion of a sequence that corresponds to a feature](#)

- [Sequences from a saved MegAlign Pro project](#)

The following video showcases several different ways to add sequences to a MegAlign Pro project:

Once sequences have been added, you can:

- [Specify a reference sequence](#)

- [Manually specify sequence type](#) (DNA, RNA, protein)

- [Reorder sequences](#)

- Rename sequences [manually](#) or [automatically, using specified data fields](#)

- [Reverse complement sequences](#)

- [Trim an individual sequence](#)

- [Remove sequences from a project](#)

Add ungapped sequences to a project

To add standard (ungapped) sequences, or additional sequences, to an existing project:

You can add standard DNA, RNA or protein sequences from a wide variety of file types, including SeqNinja .*star* files. (See Advantages of using SeqNinja files).
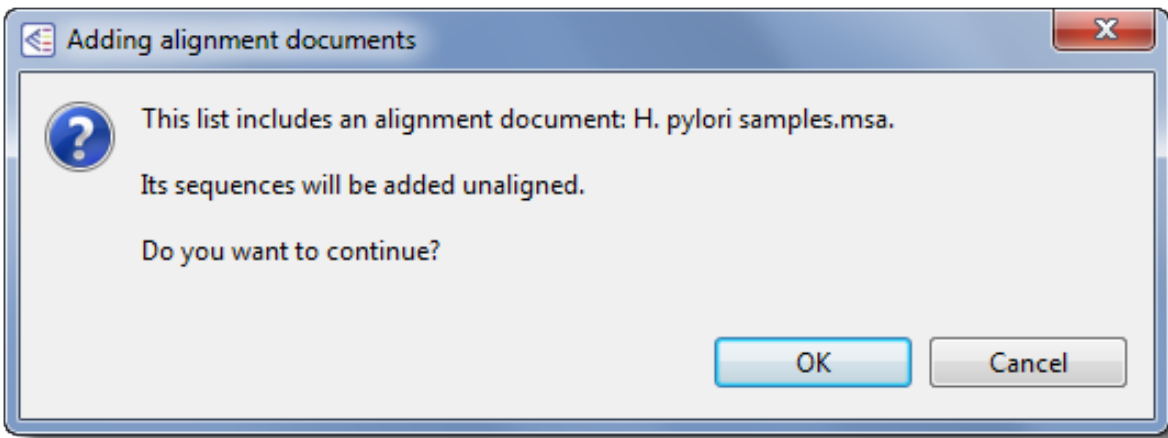
> ✳ **Note:** To add only a filtered subset of the selected sequences, instead use the procedure in Add only sequences that meet filtering thresholds.
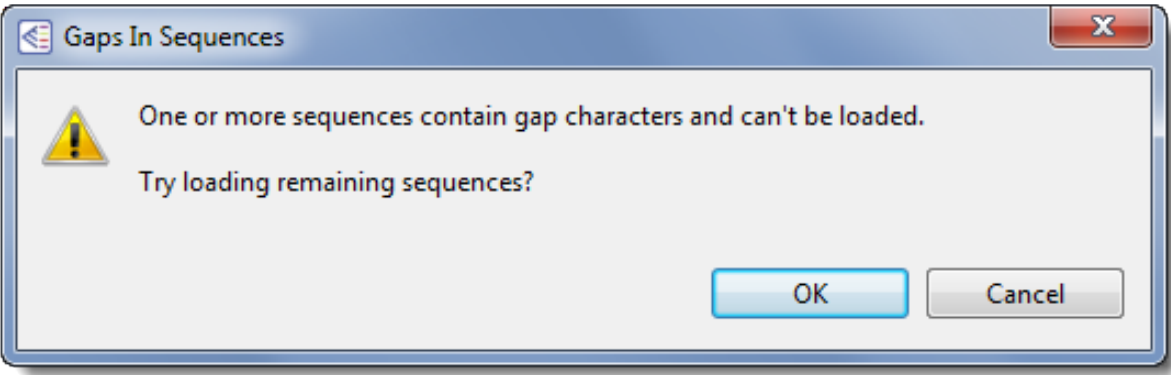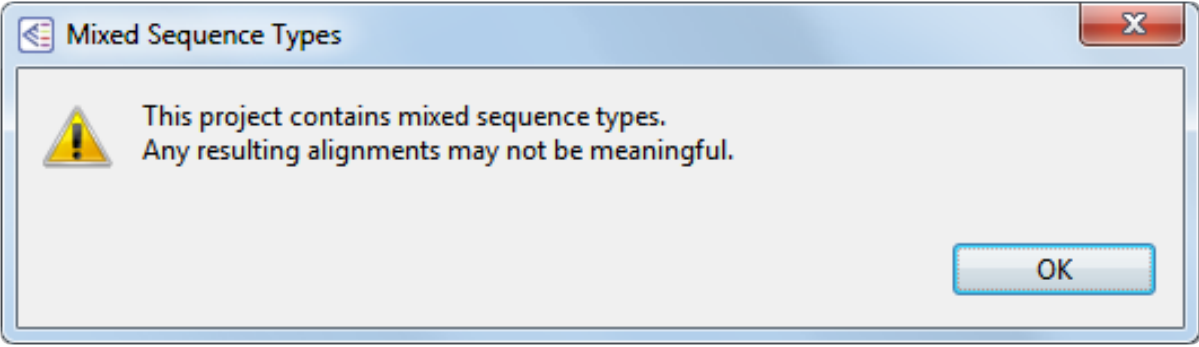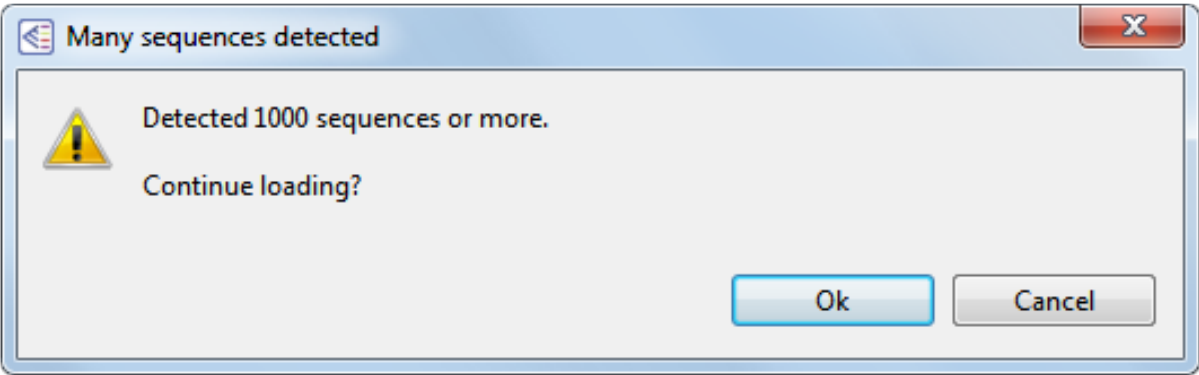
- Choose **File > Add Sequences**, select the **Add Sequences** button (ACG+), or press **Ctrl/Cmd+E**. Then choose one or more sequences from the file browser.

- Drag & drop one or more sequence files or folders of sequence files (but not MegAlign Pro .*msa* projects) on an open project.

A project can consist of any combination of file types, including single and multi-segment sequences, but all sequences must be either DNA, RNA or protein.

> ✳ **Note:** Addition of gapped or multi-segment sequences is discussed separately.

The following table shows error messages that you might encounter when adding sequences to a project.

Situation	Message and response
Adding a MegAlign Project (.msa file)	**Adding alignment documents** This list includes an alignment document: H. pylori samples.msa. Its sequences will be added unaligned. Do you want to continue? OK Cancel Click **OK** to add unaligned sequences from the saved project to the current project. This is one way to remove gaps from sequences so that they can be used in MegAlign Pro.

Adding a sequence that contains gap characters (e.g., dashes or question marks)	**Gaps In Sequences** ⚠ One or more sequences contain gap characters and can't be loaded. Try loading remaining sequences? [OK] [Cancel]
	Click **OK** to add only those sequences without gaps to the project. (There is a separate method for adding gapped sequences).
Attempting to enter mixed sequence types (both protein and nucleotide) to the Sequences View	**Mixed Sequence Types** ⚠ This project contains mixed sequence types. Any resulting alignments may not be meaningful. [OK]
	Click **OK** to add all the sequences to the project. Afterwards, manually remove any sequences that are not of the desired category.
Attempting to load more than 1000 sequences	Initial warning: **Many sequences detected** ⚠ Detected 1000 sequences or more. Continue loading? [Ok] [Cancel] After clicking **OK**, a progress bar is displayed:
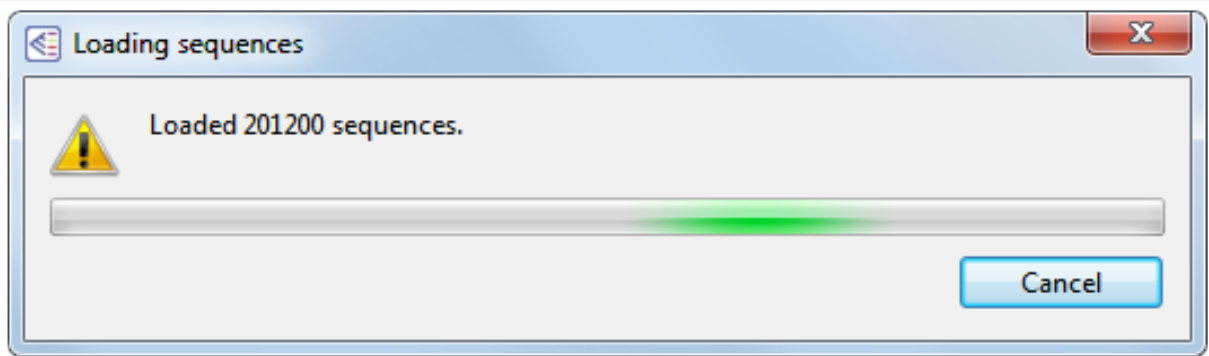
Press **OK** to continue loading all of the sequences. MegAlign Pro will display a progress bar (see lower image, left) with a running count of the number of sequences loaded so far. At any point, you may cancel the entire sample upload by pressing **Cancel**. This terminates the entering step without adding the new file(s). Otherwise, simply wait for the samples to finish loading.

Entering SeqNinja projects (.star) or FASTA files with ambiguous extensions (e.g., .fasta)	
	MegAlign Pro can usually determine whether the sequence consists of DNA, RNA or protein. If it cannot, MegAlign Pro will prompt you to specify file type by selecting a button from the dialog.
Attempting to add a SeqNinja project (.star) that contains errors	
	Click **OK**. Launch SeqNinja, fix any errors in the project and save it before adding the file to MegAlign Pro.
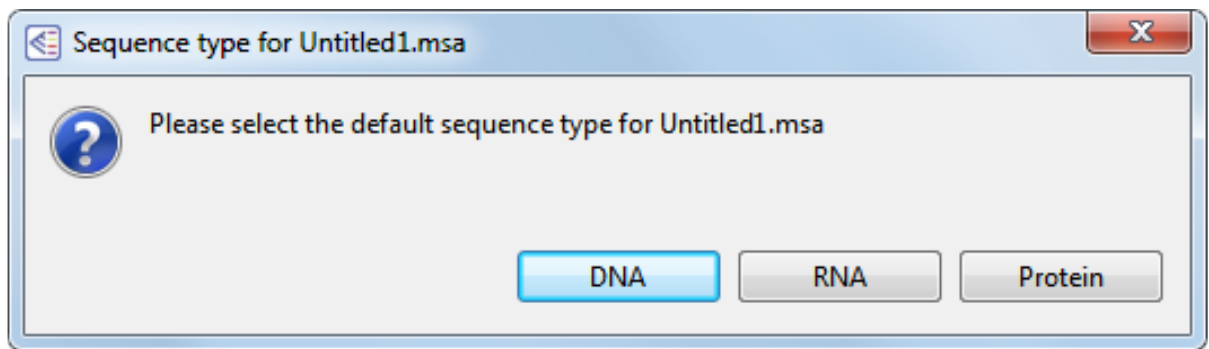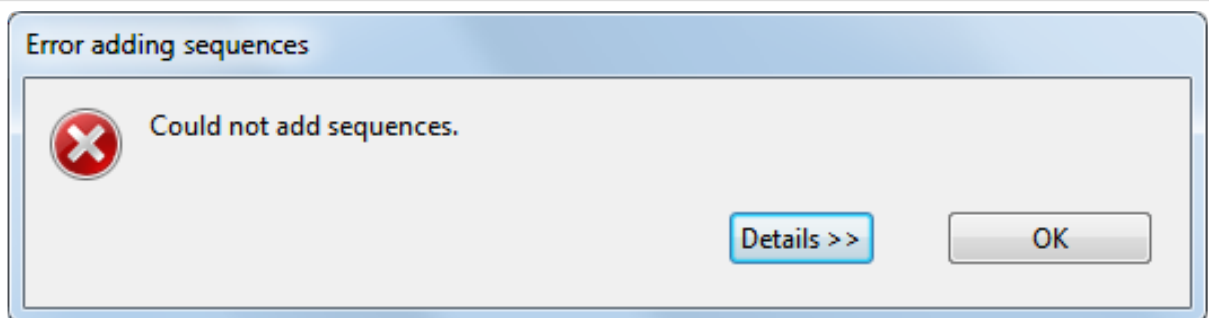
Add gapped sequences to a project

To import gapped sequences from an existing MegAlign Pro file (*.msa*):

The following procedures both remove gaps from the sequences and place the updated sequences in the "Unassembled Sequences" area of the project.

- Click on the **Add Sequences** button ( ) or select **File > Add Sequences** and choose the *.msa* file.

- Use **File > Open** and choose the *.msa* file. Once the project is open, select all the sequences, then use **Align > Unalign Selected**.

> ✳ **Note :** MegAlign Pro uses a dash (-) to represent gaps. MegAlign Pro does accept sequences where gaps are represented by letters such as "n" or "x." Sequences with non-IUPAC symbols other than dashes may be accepted by MegAlign Pro as input, but may not be recognized by certain alignment algorithms. Of MegAlign Pro's methods, MAFFT is the most tolerant of non-IUPAC symbols.

To import gapped nucleotide sequences from another type of file (not *.msa*):

When the gapped sequences are not part of a MegAlign Pro file, you can use SeqMan Pro to remove the gaps. This procedure is safe to use for nucleotide (DNA or RNA) sequences only.

1. Launch DNASTAR SeqMan Pro, then drag and drop the gapped sequence file(s) onto the SeqMan Pro window.

2. Select all sequences in the Unassembled Sequences using **Ctrl+A** (Win) or **Cmd+A** (Mac).

3. Use **Contig > Export Sequences > Single File**.

4. Select either **FastA** or **GenBank Flat File** format and choose a file name and location. Make sure that **Include gaps** is left unchecked, and then click **Save**.

The sequence(s) in the newly saved file can now be added in the usual way.

> ❗ **WARNING:** Do not use this method for protein sequences, as all letters that are not also in the IUPAC nucleic acid code will be lost.

Add only sequences that meet filtering thresholds

You can specify a group of sequences to add and then use filtering options to only add a subset of those to the project.

To add only sequences that meet filtering thresholds:

Open the filter dialog by doing any of the following:

- From the Welcome screen, select **New Project** on the left and **Filter and align sequences** on the right.

- Within any project, choose **File > Add Sequences with Filter**.

- Within any project, press the black triangle to the right of the **Add Sequences to Project** tool ( ) and choose **Add Sequences with Filter**.

- After searching NCBI's BLAST or Entrez databases, press the **Filter results** tool ( ) in the header above the Table view and then choose **Filter**.

To set up the filters:

1. If the Choose Sequences to Add dialog appears, select one or more sequences or a folder of sequences and click **Open**. This step is skipped if you are filtering search results, as MegAlign Pro will be using the sequence matches from the table.

2. The Filter or Filter Sequences dialog opens. Press the **<select>** drop-down menu to choose from among some of the filter types below. Available options depend on the circumstances under which you accessed the filter dialog.

Filter type	Filters sequences to allow only sequences…
Accession	Matching a particular Accession number.
Ambiguous residues	With more or less than the desired number/percent of ambiguous residues.
% Coverage	Where a specified portion of that sequence aligns to a match in the database. Example: Query has 596 bases, of which 565 align to a match in the NCBI database. **%Coverage** = 565/596 = 95%.

Description	Whose definition matches NCBI "Definition" field text.
% Identity	Where a specified percentage of bases align to a match in the database. ([NCBI definition](#)). Example: Query has 565 bases that align to the match and 3 bases that don't match. **%Identity** = 562/565 = 99.5%.
Score	Meeting the similarity score criteria for the match. (See the [NCBI definition for **raw score**](#)). In general, a higher score denotes a better match.
E-value	That are <u>not</u> likely to match only coincidentally. (See the [NCBI definition for **E value**](#)). In general, a lower score denotes a better match.
Length	Matching a specified length in base pairs.
Organism	Matching a specified species name.
Probability	Meeting criteria based on probability of inclusion.
Sequence	Having a particular string of amino acids or nucleotides, with options to set the maximum number of mismatches or specify that the string can be found on either strand.
Sequence length	Shorter or longer than a specified length.
Sequence name	Having/lacking a certain string of letters/numbers in their names. Wildcards are supported.
Sequence number	Incling only every 'n' number of sequences between two specified sequences, where 'n' is entered as the **Interval**.

3. A variety of text or number boxes and additional drop-down menus will populate to the right of your selection. Enter the desired criteria for that filter row.

4. To add another filter row, press the plus icon (**+**). To delete a filter row, press the minus icon (**−**). The image below shows an example with two filter rows.

5. Once you have set up all desired filter rows, go to the top of the dialog and choose whether to **Match all of the following** or **Match any of the following** queries.

6. After you finish setting up filters, do any of the following:

   • To apply the filters and open the sequences that met the criteria, press **Apply** or **Filter**.

   • To add all sequences without filtering, press **Don't Filter**. This option is not available when filtering BLAST or Entrez matches.

   • To clear all filters and start over, press **Clear** or **Clear Filter** in the top right of the popup.

   • To exit the dialog without filtering, press **Cancel**.

Add multi-segment sequences to a project

Sequences that consist of more than one chromosome, contig, or fragment are called multi-segment sequences. By entering a group of sequences into MegAlign Pro using a special command, the entire data set will be concatenated end-to-end and treated as a single multi-segment sequence.

Suppose you want to string together the consensus sequences from all the contigs in an assembly, or chromosomes in a genome, or exons in a gene, etc. If you enter these sequences using **File > Add Sequences** or the corresponding button ( ), you will end up aligning the parts to one another. To avoid this, use the following procedure:

To add sequences end-to-end:

Choose **File > Add Sequences End-to-end**, select the **Add sequences end-to-end in project** tool ( ), or press **Ctrl/Cmd+Shift+E**. The following dialog appears:
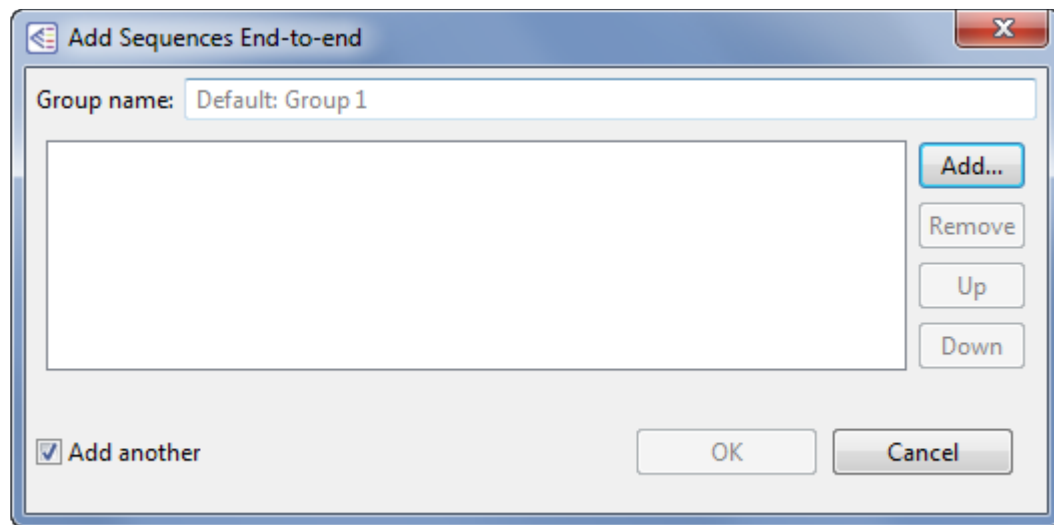
- To name the group, type a name in the **Group name** box. This serves as a substitute for a regular sequence name in all the MegAlign Pro views. If you do not explicitly type a name into the **Group name** box, a default name will be computed from the items in the list.

> ✱ **Note:** Like other sequence names, a Group name can be changed later from within the MegAlign Pro views by right-clicking on it and choosing **Rename** or **Rename with Fields**.

- To add a sequence file to the left-hand pane, click **Add**. Alternatively, you may simply drag files from any file explorer & drop them into the main part of the dialog. Any supported MegAlign Pro file type may be added.

- Each group of multi-segment sequence should have its segments arranged in the same order relative to other groups. To reorganize sequences, select a sequence in the left pane and then use the **Up** or **Down** buttons to move it to the desired position.

- To remove a sequence from the left-hand pane, select it and click **Remove**.

- To add another group of sequences after you save the current group, leave the **Add another** box checked. Otherwise, remove the check mark.

Click **OK** to keep your changes and add the sequence group(s) to the project, or **Cancel** to leave the dialog without adding the sequence group(s) to the project.

Unless you unchecked the **Add another** box, a new Add Sequences End-to-end dialog will appear after you check **OK**. If you don't have any additional groups to add, you can exit by pressing **Cancel**.

To edit an existing end-to-end group:

Select an aligned or unaligned sequence group in any view and choose **Edit > Edit Sequences End-to-end**.

If you select an aligned sequence and then apply changes in this dialog, the sequence will be removed from the alignment and placed in the "Unaligned Sequences" area. Because of this possibility, you must respond **OK** to the following prompt before the dialog will open:
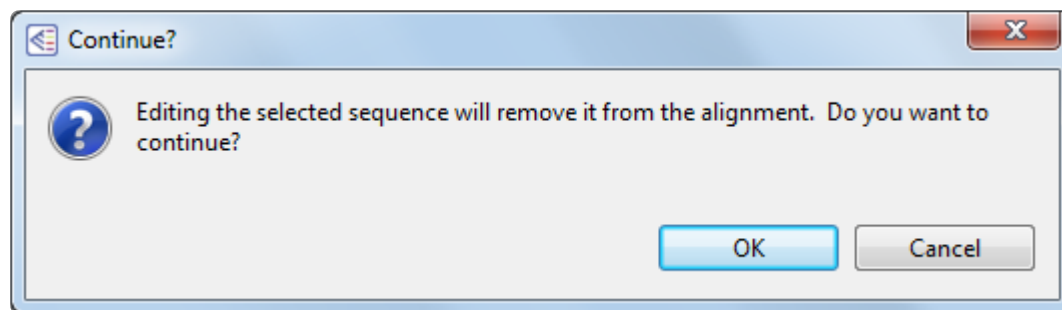
Try it! – Add and align multi-segment sequences

The following video tutorial describes how to add multi-segment sequences and then align them using the MUSCLE multiple alignment algorithm.

If you have not yet downloaded and extracted the tutorial data, click here to download it. Then decompress (unzip) the file archive using the method of your choice.

Add only the portion of sequence that corresponds to a feature

This topic describes two methods that allow you to align only the portions of sequences the correspond to a particular gene or other feature. Other portions of the sequences are discarded or removed prior to alignment, making the alignment for that gene both faster and more accurate.

If you only need to trim a few sequences, you can do this manually. See Trim an individual sequence for instructions.

If you need to trim a larger number of sequences, we recommend doing this in SeqNinja, a batch file editor that is included in the same Lasergene package as MegAlign Pro. You can use SeqNinja's Extract Features as Sequences template to batch extract the gene features (e.g., CDS, etc.) and save these as a separate set of sequences. You then add the trimmed sequences to MegAlign Pro and align them. If you are new to SeqNinja, Parts A and B of this tutorial show how to batch extract genes from multiple sequence files.

Add sequences from a saved project to an active project

To add sequences from a saved project to an active project:
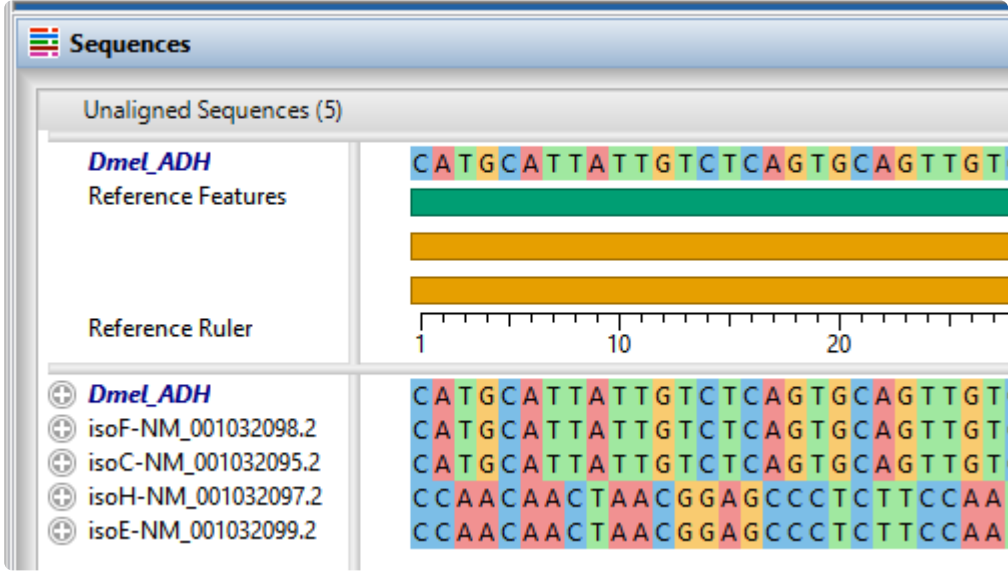
Choose **File > Add Sequences** and select a MegAlign Pro project file (*.msa*). All sequences in the *.msa* file will be added to the "Unaligned Sequences" area of the active project.

Specify a reference sequence

This topic describes several reasons you might want to use a reference and procedures for specifying, changing, and removing the reference.

Note that a reference sequence in MegAlign Pro is different from a reference in SeqMan Ultra or SeqMan NGen. In SeqMan NGen or SeqMan Ultra, the reference is a long DNA sequence that is used as a template for guiding the assembly of small DNA fragment reads into contigs. By contrast, any sequence of any length or type (DNA, RNA, protein) can be used as a reference in MegAlign Pro.

Task	How To
Specify a reference sequence for guiding a [MAFFT alignment](#)	Use any of the following methods: • Select the sequence you would like to use as the reference. Then choose **Sequences > Use as Reference** or right-click on the sequence name and choose the same command from the context menu. Then choose **Align > Align Using MAFFT**. • Choose **Align > Align with Options**. In the **Using** menu, choose **MAFFT**. In the **Reference Sequence** menu, choose the sequence you want to use as the reference. Click **Align**. • From the Welcome screen, choose the **New Project** tab on the left, then choose **New alignment with options** on the right. From the ensuing dialog, choose the sequences you wish to align and press **Open**. The Align Sequences dialog then pops up and is already set to **MAFFT**. In the **Reference Sequence** menu, choose the sequence you want to use as the reference. Click **Align**.
Specify a reference for any other reason	Select the sequence you would like to use as the reference. Then choose **Sequences > Use as Reference** or right-click on the sequence name and choose the same command from the context menu. Then choose the **Align > Align Using** command for the algorithm you want to use. You can specify the reference either before or after any alignment type. Below are two common reasons for specifying a reference sequence: • For performing [variant analysis](#) for any type of alignment. For instance, you might download an annotated reference sequence that you found online and use this workflow to compare your experimental sequences to it. • To [map the features](#) from an annotated reference sequence to other sequences in the alignment.
Switch to another	Use any of the methods above, but choose a different sequence. Note that if you perform a MAFFT alignment using "Sequence A" as the reference and later specify "Sequence B" as the

reference	reference, the variants calculated against "Sequence A" are not lost. You can redesignate "Sequence A" as the reference again without recomputing its variants.
Stop using a reference	One reason you might choose to stop using a reference is if you want to compute variants by comparing other sequences to the consensus, rather than to the reference. To remove the reference, select its name in a graphical view and choose **Sequences > Stop Using as Reference**. Alternatively, you can right-click on the reference sequence name and choose the same command from the context menu. The reference sequence will be removed from the **References** areas at the top of the Overview and Sequences view, but the sequence will still appear in blue bold-italic font in the list of sequence names. To remove this formatting, see the row just below.
Show/hide the reference used in the last alignment (after you have stopped using it as a reference)	If you align sequences with a reference selected and later stop using the reference (see the row just above), the reference sequence will be removed from the **References** area at the tops of the Overview and Sequences view, but the former reference sequence will still appear in blue bold-italic font in the list of sequence names. To remove this special formatting, choose **Sequences > Unmark Reference of Last Alignment**. To return to the special formatting, choose **Sequences > Mark Reference of Last Alignment**.
Change how the reference sequence is displayed in the views	After you specify a reference sequence using one of the methods above, the reference sequence, reference features and reference ruler can be displayed as tracks in the Sequences view and/or Overview. By default, the sequence that you specify as the reference is displayed in bold blue italics to differentiate it from the other sequences. To change the way the reference is displayed, you can select the corresponding **Reference**

	track in the Tracks panel, then modify the track options.
Visually compare sequences without changing the reference	See Visually compare other sequences to one sequence.

Manually specify sequence type

When you add a sequence to a project, MegAlign Pro uses clues in the sequence file to ascertain whether it is DNA, RNA or protein. When uncertain, it will prompt you to supply the missing information as you add the sequences. If MegAlign Pro guesses incorrectly (usually due to ambiguous information in the sequence file) or you make an erroneous selection, you can change the sequence type at any time using any of the methods below:

- Choose **Edit > Change Source Sequence Type**. In the popup, select either **DNA**, **RNA** or **Protein**.
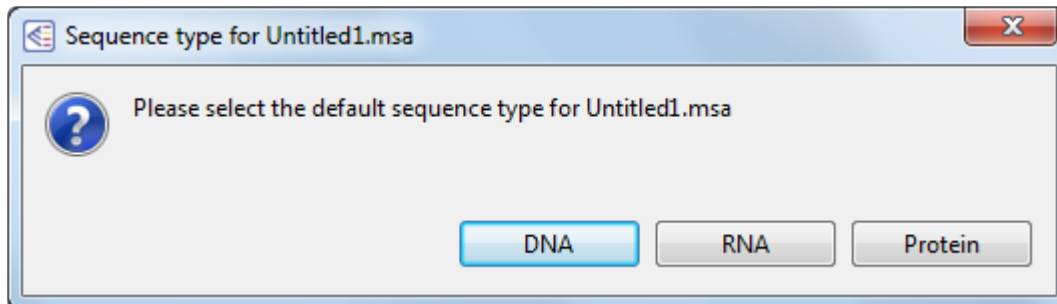
- If sequences in the Sequences view are currently displayed as DNA (there are T's in the sequences), switch to RNA using **Sequences > Display As > RNA**. All T's will be replaced by U's.

- If sequences in the Sequences view are currently displayed as RNA (there are U's in the sequences), switch to DNA using **Sequences > Display As > DNA**. All U's will be replaced by T's.

In all cases, the sequence type for all sequence in the project will be changed. MegAlign Pro does not support aligning a mix of DNA, RNA and/or protein sequences.

Change the order of sequences

Whether preparing an image of a view for publication, or simply for ease of analysis, you may want to change the order of sequences in one or more views. For instance, you may wish to order the tree to create a pleasing image or to draw the viewer's focus to a certain spot, or to allow for easier comparison.

The following methods allow you to reorder sequences in a view. Note that some commands require you to first select one or more sequences before using the command.

Task	Menu command	Views where shortcut is available via the Sort Sequences tool (⬇️)	Affects sequence order in…	Notes
Order selected sequences manually	In the Sequences view or Overview, drag and drop the sequence(s) in the desired position.	N/A	Overview and Sequences view	
Order sequences automatically by name or length	**Sequences > Order Sequences Ascending** (or **Descending**) **> By Name** (or **Length**).	Overview, Sequences view and Distance view	Overview and Sequences view	
Order sequences automatically by pairwise statistics	**Distance > Order Sequences Ascending** (or **Descending**) **By Pairwise > Distance** (or **%Identity, %Similar, %Dissimilar, %Gaps, Identical, Similar, Dissimilar, Gap count, Gap length** or **Length**).	Overview, Sequences view and Distance view	Overview, Sequences view and Distance view	
Order sequences	**Distance > Order Sequences by**	Distance view	Overview, Sequences	If nothing is selected, you will instead see the command described in the next row.

by distance from the top sequence in the Distance view table	**Distance from Top Sequence**		[view](#) and [Distance view](#)	
Order sequences by distance from a selected sequence	**Distance > Order Sequences by Distance from Selection**	[Distance view](#)	[Overview](#), [Sequences view](#) and [Distance view](#)	If anything is currently selected, you will instead see the the command described in the previous row.
Order sequences in the Tree view by depth or distance	**Tree > Order Ascending** (or **Descending**) **By > Default Order** (or **Depth** or **Distance**)	[Tree view](#)	[Tree view](#)	**Default Order** and **Depth** both sort in order of depth (i.e. by the sum of the branch lengths). Depth measures branching; the children of a subtree are sorted by the maximum number of branches to a leaf. **Distance** sorts in order of the distance between samples in the dataset. The children of a subtree are sorted by the maximum distance (i.e. length in one dimension) from the subtree node to a leaf. In this set of commands, **Ascending** and **Descending** are arbitrary terms referring to the display structure of the tree. **Ascending** tends to form a diagonal from the upper-left to lower right while **Descending** tends to form a diagonal from lower-left to upper-right.
Order sequences in all views to match the order in the Tree view	**Tree > Order Sequences Like Tree**	[Tree view](#)	[Overview](#), [Sequences view](#) and [Distance view](#)	

Rename sequences manually

To manually rename a selected sequence:

1. Select a single sequence name from any MegAlign Pro view, then right-click on it and choose **Rename Sequence(s)**. Alternatively, select the sequence name and choose **Edit > Rename Sequence(s)**.

   The Choose Label dialog opens with the **Display** drop-down menu set to **Custom label**.
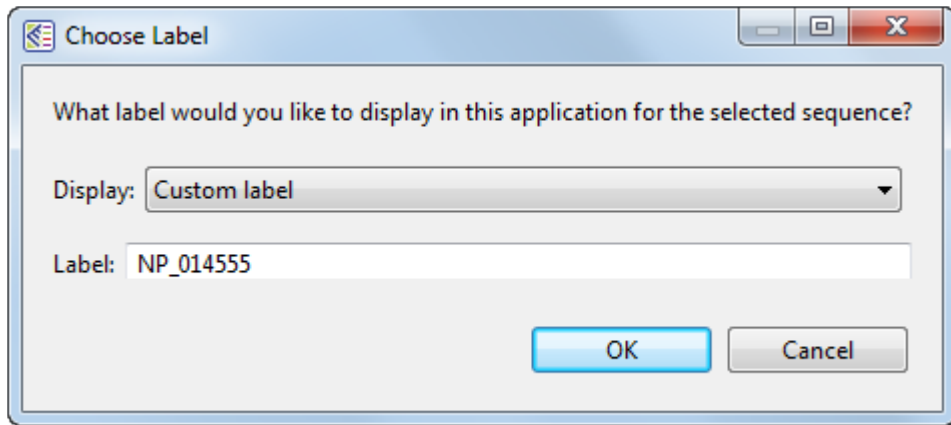
2. Type the desired name into the **Label** text box.

3. Click **OK** to save changes and exit from the dialog, or **Cancel** to leave the dialog without saving any changes.

> ✳ **Note:** When you export data for a renamed sequence, the new name may be exported or not, depending on the situation. See Retain custom names during export for details.

Rename sequences automatically using specified data fields

The following procedure lets you create labels for one or more selected sequences automatically, using specified fields such as sequence length, organism, or description.

<u>To rename sequences using specified data fields:</u>

1. Select one or more sequence names from any MegAlign Pro view.

2. Right-click on the selection and choose **Rename Sequence(s) with Fields**, or instead choose **Edit > Rename Sequence(s) with Fields** from the main menu. Making no selection prior to using the main menu command is equivalent to selecting all sequences.

    The Choose Label dialog opens with the **Display** drop-down menu set to **One or more fields**.
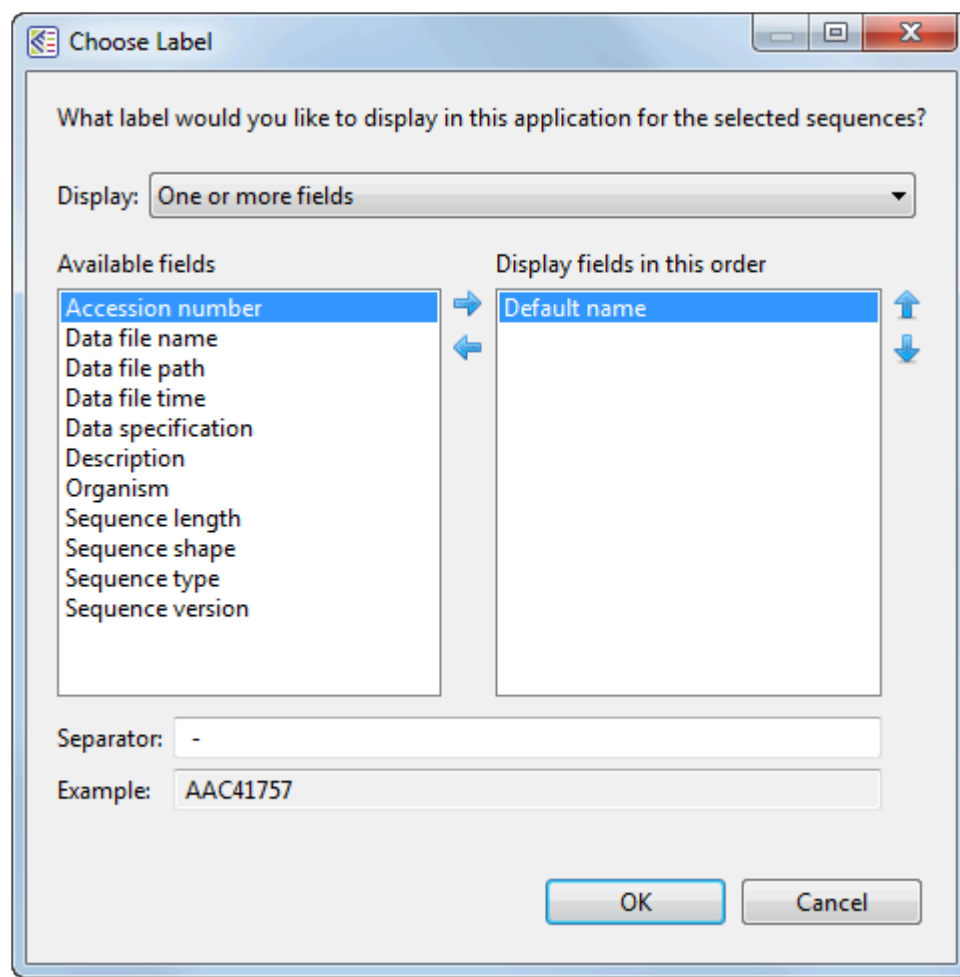
The left pane shows available fields, and the right pane shows fields that will be displayed as part of

the sequence name. Initially, **Default name** is the only field listed in the right pane. An example showing the appearance of the name using the specified field(s) appears in the **Example** box at the bottom of the dialog.

3. Select and organize the fields you wish to display as part of the sequence name:

   - To add or remove items, move them between panes using the left and right arrows (⬅️ ➡️). Use the up and down arrows on the right (⬆️⬇️) to specify the order in which displayed fields should appear.

   - By default, a hyphen (-) is included between each of the displayed fields. If desired, type in a different **Separator**. Or remove all text in the **Separator** box to include nothing between the fields.

   - The **Example** text box shows what the label will look like with the current selections. This text can be manually edited, if desired.

4. When you are satisfied with the choices made in this dialog, click **OK** to save changes and exit from the dialog. Or click **Cancel** to leave the dialog without saving any changes.

✱ **Note:** When you export data for a renamed sequence, the new name may be exported or not, depending on the situation. See Retain custom names during export for details.

Reverse complement one or more sequences

To reverse complement one or more sequences:

1. Select one or more sequences in the Overview or Sequences view.

2. Do either of the following:

   • Press the **Reverse Complement** tool ( ) in the toolbar.

   • Right-click on the selection and choose **Reverse Complement**.

If the selected sequences were *not* yet aligned, they will be reverse-complemented and renamed with the suffix "-rc" to denote the change.

If the selected sequences *were* part of an existing alignment, the following popup will be displayed.
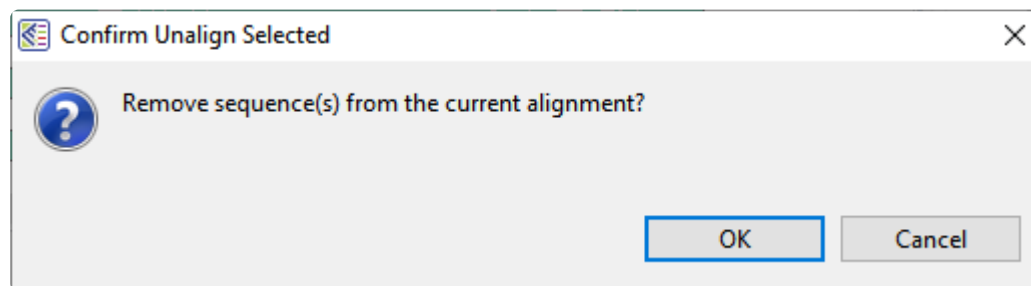
If you choose **OK**, the selected sequences will be removed from the alignment, reverse-complemented and will have the suffix "-rc" added to them. They will then be placed in the Unaligned sequences area.

Trim an individual sequence

Individual sequences in the project can be trimmed before or after alignment based on features or by direct selection. This is different from trimming the entire alignment, which affects all sequences in the project.

To trim an individual sequence before or after performing an alignment:

Select a sequence in the Overview or Sequences view, then do any of the following:

- Press the **Trim Sequence** tool (✂) from the toolbar.

- Choose **Sequences > Trim Sequence**.

- Right-click on the sequence and choose **Trim Sequence**.

The Trim Sequences dialog opens.
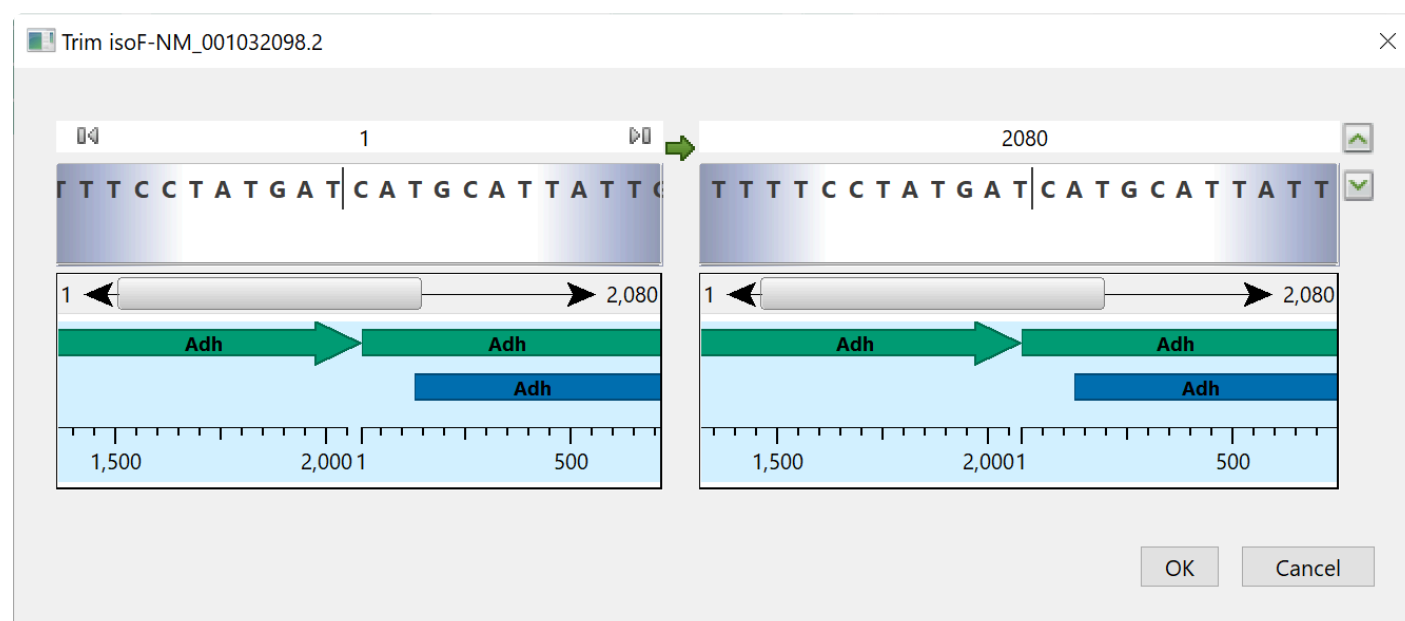
The left and right halves of the dialog are used to specify the desired left and right sequence ends. On each side, the uppermost section is a textbox showing the currently-selected endpoint. Initially, these boxes are populated with the first and last nucleotides or amino acids in the sequence.

The following table shows tasks that can be done from within this dialog, or that affect its appearance.

Task	How To
To reveal/hide the sequence	Click the green "up" arrow on the upper right of the popup.

To reveal/ hide the ruler and any features that are present	Click the green "down" arrow on the upper right of the popup.
To specify the desired range	Use any of the following methods: • Drag the left or right sequence wheels (gray/white areas under the text boxes) to scroll to a position of interest. A parenthesis denotes each currently selected endpoint. Click on a base or amino acid to snap to that position or drag the sequence wheels left or right to move the parenthesis to a different position. If the sequence is long, you can scroll along it more quickly by using the sliders under each wheel or by dragging the features/ruler wheels below the sliders. • Type a position into the text boxes above the left and right ends of the sequence. You can reference a numerical position, the left end ("lend") or right end ("rend") of the sequence, or a simple numeric expression. For example, `lend+12` references the base that is 12 places to the right of the left end (i.e., position 1), while `815-27` references the base that is 27 places to the left of position 815. • Select the range corresponding to a feature by scrolling until the feature is visible under both sequence wheels. Then double-click on the same feature name under both the left and right wheels.
To skip to the left or right ends of the sequence	Click in a range text box, then use the gray left/right arrow icons that appear within the box. Alternatively, type in `lend` (left end) or `rend` (right end).
To view feature information	Hover above a feature (if any are present) to see information about it, such as its left and right coordinates.
To reverse complement the current sub-range	Click the green arrow between the two range textboxes.

IUPAC ambiguity codes are recognized both in the sequence and in the range boxes. For example, typing AAS (where S = C or G) into a range box would cause SeqNinja to look for the first instance of AAC or AAG in the sequence. Conversely, typing AAC into a range box would cause SeqNinja to look for the first instance of AAC, as well as any combination of bases and ambiguity codes that would allow for AAC (e.g.,

AAS, WWM, etc.).

Remove sequences from a project

The following procedure causes the selected sequences to be completely removed from the project, not simply moved to the "Unaligned Sequences" area.

To remove sequences from the project:

1. Select one or more of the aligned or unaligned sequences in the Overview or Sequences View by clicking, **Ctrl/Cmd+clicking** or **Shift*+clicking** on the sequence name(s).

2. To remove sequences, either:

   - ◦ Right-click on the list of sequence names, and select **Remove Sequence(s)** from the menu.

   - ◦ Press the **Delete** key or use **Edit > Delete**.

3. Respond to the confirmation message that appears, if any. Different versions of the message appear, depending whether you have selected aligned and/or unaligned sequences. One example is shown below.
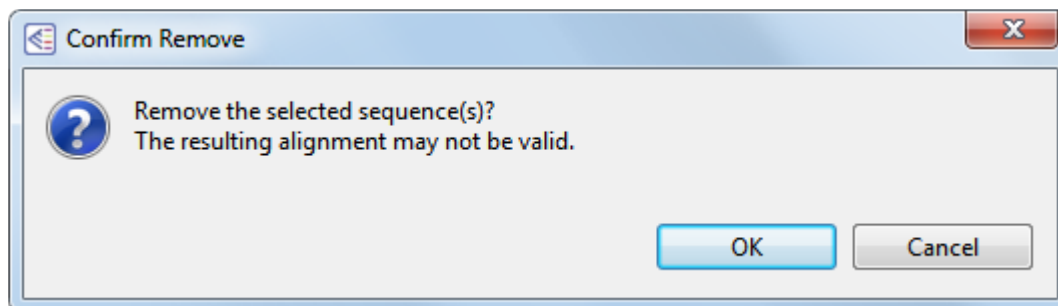
Notes:

- During multiple alignment, "guide trees" are created that govern the order in which sequences are aligned to each other, starting with the most similar pair. Guide trees, in turn, are based upon all possible pairwise relationships among the member sequences.

- When one or more sequences are removed from an alignment, any column containing 100% gaps is removed from all of the remaining aligned sequences. No new guide tree (representing the remaining sequences) is calculated. Therefore, the resulting alignment may not be as good as one produced by performing a completely new alignment with the remaining sequences.

- When features have been mapped to the consensus, removing sequences from the alignment or unaligning (using **Align > Unalign All**) will cause all features to be removed from the consensus sequence. This is because feature mapping may no longer be accurate, since the consensus

sequence may have different gapping.

Advantages of using SeqNinja files

SeqNinja is DNASTAR's tool for advanced and batch editing of genome sequences and features. SeqNinja allows you to convert file formats, batch edit and export features, merge and split sequence data, and transfer features between genome versions, automating the process of adjusting sequence coordinates between SeqMan NGen assemblies. The SeqNinja project – which can be created easily using the provided templates – has the extension ".star," and uses one or more data files as inputs for its calculations. When you run the *.star* program, the output format can be saved in a variety of popular file types, such as FASTA, GenBank, Lasergene, etc.

While you can certainly add the SeqNinja output files to MegAlign Pro, it is also possible to add the *.star* project as input. To add either of these, see Add ungapped sequences to a project.

Adding one or more SeqNinja projects (*.star*) to a MegAlign Pro project is an indirect way of adding sequences. Sequences specified by a SeqNinja project file are all those that would be written if the project were run using SeqNinja. Using SeqNinja project files as MegAlign Pro input can save a lot of time and effort. The advantage of adding a .star project to MegAlign Pro is that it will always call up the most up-to-date version of the input data for use in making its output. So if you change, add, or delete the input data that the *.star* file points to, the .star file will update this information in MegAlign Pro.

The following example demonstrates the advantages of adding a *.star* file to MegAlign Pro:

Assume you have a directory of genomes for a phylogenetic study. You want to extract a specific gene, or set of genes, from each genome to build an alignment, and then a phylogenetic tree. Over the course of your work, new genomes and new versions of existing genomes become available. You want to generate a current alignment and tree that reflects those additions or changes. Here are three scenarios of the steps needed to do this:

Good: Without using SeqNinja:

1. Manually extract the genes of interest.

2. Generate output files (e.g., FASTA, GenBank, etc.).

3. Add the files to MegAlign Pro.

4. Run the alignment.

5. **Repeat Steps 1-4 every time the input data changes.**

Better: Using SeqNinja output (e.g., FASTA, GenBank) or a SeqNinja project file as MegAlign Pro input:

1. Create a SeqNinja program to automatically extract the genes of interest.

2. Run the program and generate output files (e.g., FASTA, GenBank, etc.).

3. Add the output files *or* the SeqNinja project (.star) to MegAlign Pro.

4. Run the alignment.

5. **Repeat Steps 2-4 every time the input data changes**

Best: Creating a SeqNinja project to use with MegAlign Pro:

For detailed information about creating SeqNinja projects, launch the SeqNinja application and select **Help > Help Contents**.

When you use a SeqNinja Project (.star) file as input for MegAlign Pro, note that the project is read as a "set of sequences," and is not actually executed, meaning no sequence files are written. This type of project has the following limitation:

Any input files referenced in the program must already exist. In other words, the program should not reference an input file that doesn't exist until it is created earlier in the same program. Instead:

- Replace all the references to the input file with a variable (e.g., $A), or…

- Create a SeqNinja program to create the necessary files and run the program. Then, in the SeqNinja program you wish to add to MegAlign Pro, reference those already-created files.

Tracks

"Tracks" are rows of information that can be displayed or hidden, as desired, in the Overview and Sequences view. Tracks can pertain to a header, footer, or individual sequence row. The following video provides a brief overview of how tracks are used in MegAlign Pro:
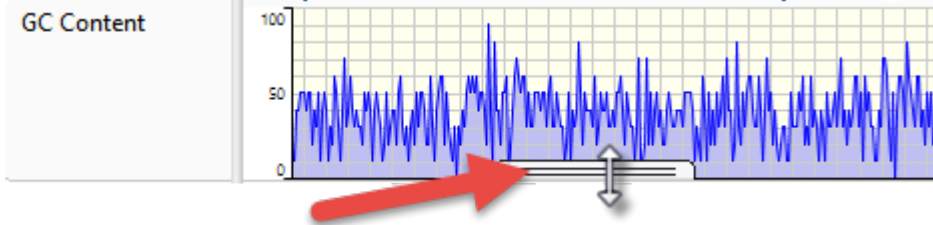
The following tracks can be applied, removed or (in most cases) edited:

- Ruler
- Features
- Reference
- Consensus
- Consensus Match
- Sequence Logo
- GC Content
- Gap Fraction
- Numeric
- Pairwise Match Bar
- Translation

The following table shows tasks that affect the display of tracks within the Overview and Sequences views:

Task	How to…
Create tracks	Track data can be imported automatically with some types of sequence data (e.g., Feature tracks), or can be calculated by the application itself (e.g., Ruler and Numeric tracks).

Display tracks in the views	To display any type of track, you must check the box next to the name of the track in the Tracks panel. This is the only condition needed to view a Ruler track. For any type of track other than a Ruler, you must also expand the track as described below.
Expand/collapse tracks for selected sequences	Do any of the following: • Click the plus/minus sign to the left of the sequence name (⊕ / ⊖). • To expand tracks for selected sequences, choose **Sequence > Expand Sequence Rows** or **Overview > Expand Overview Rows**; or right-click on the selection and choose **Expand**. • To collapse tracks for selected sequences, choose **Sequence > Collapse Sequence Rows** or **Overview > Collapse Overview Rows**; or right-click on the selection and choose **Collapse**.
To expand/ collapse tracks for all sequences in a particular view	Ensure that no sequences are selected. • To expand tracks for all sequences, choose **Sequence > Expand All Rows** or **Overview > Expand All Rows**; or right-click in the view and choose **Expand All**. • To collapse tracks for all sequences, choose **Sequence > Collapse All Rows** or **Overview > Collapse All Rows**; or right-click in the view and choose **Collapse All**.
Hide tracks	Do one of the following: • To hide a specific track from all views, uncheck the box next to the name of the track in the Tracks panel. • To hide a specific track in one view only, right-click on the track within that view and choose **Hide Track**. Alternatively, select a track and use **Edit > Delete** or press the **Delete** key. • To hide all tracks in one view only, click the minus sign (⊖) to the left of the sequence name in that view, or use **Analysis > Collapse Sequence Rows** or **Overview > Collapse Sequence Rows**. • To hide all tracks in all views, do a combination of the above.
Access the options for a particular type of track	Do either of the following: • Right-click on a track in either view and choose **Show Track Options**.

	• Click **View > Tracks > Options** or open the Tracks panel and click the Options bar (⬍ Options), then do either of the following to specify the track of interest by clicking on its name in the [Tracks panel](#).
Change the order of all tracks in a view	Open a sequence row to view its track. Then drag a track label (e.g. "Features") to the desired position among the other tracks and drop it there. Note that if you turn off some tracks in the Tracks panel and add them again later, they will appear in the order in which their boxes were checked.
Increase/decrease the vertical height of a visible track	Hover the mouse over the bottom of the track in the Overview or Sequences view, then grab the resize bar and drag it up or down.
See how the zoom level can affect the appearance of tracks	See [Customize the appearance and layout of views](#).

Ruler tracks

To display rulers:

The visibility of ruler tracks in the Overview, Sequences view and Pairwise view is dependent on:

- Whether or not tracks are visible in those views. To display hidden tracks, click on the plus sign next to the sequence of interest in the view.

- Which boxes you check in the Tracks panel.

  ◦ Sequence rulers can be displayed in MegAlign Pro's graphical views by checking the following boxes in the Tracks panel:

    ▪ **Overview details > Sequence Ruler** – In the Overview, a sequence ruler running from left to right (i.e., larger numbers on the right), denotes a sequence that has been reverse complemented with respect to its original orientation. MegAlign Pro does not permit manual reverse-complementing of sequences, but the Mauve multiple alignment algorithm reverse-complements automatically, where necessary. In a Mauve multi-block alignment, changing the active block can cause entire sequences to be reverse-complemented.

    ▪ **Sequence detail > Sequence Ruler**

    ▪ **Pairwise > Ruler**

  ◦ If you have specified a Reference sequence, reference rulers can be displayed in the views by checking these boxes in the Tracks panel:

    ▪ **Overview header > Reference Ruler**

    ▪ **Sequences header or footer > Reference Ruler**

To change Ruler options:

Currently, the Ruler tracks do not have editable options. However, you can add white space above and/or below the ruler using sliders in the Layout section.

Features tracks

<u>To display sequence features:</u>

The visibility of feature (annotation) tracks in the <u>Overview</u>, <u>Sequences view</u> and <u>Pairwise view</u> is dependent on:

- Whether or not tracks are visible in those views. To display hidden tracks, click on the plus sign next to the sequence of interest in the Overview or Sequences view.

- Which boxes you check in the <u>Tracks panel</u>.

Sequence features can be displayed in MegAlign Pro's graphical views by checking the following boxes in the Tracks panel:

- **Overview details > Sequence features**
- **Sequence detail > Sequence Features**
- **Pairwise details > Sequence Features**

Reference sequence features can be displayed in the views by checking these boxes in the Tracks panel:

- **Overview header > Reference Features**
- **Sequences header or footer > Reference Features**

Note that after performing a multiple alignment, checking the **Consensus** box will also cause Consensus features imported with **Features > Map Features to Active Consensus** to be displayed.

To view a tooltip with details about a feature, hover your mouse over the feature in any view. Alternatively, select one or more features by clicking on them and then view information about them in the <u>Details panel</u>.

The same default color scheme is used for features in both MegAlign Pro and <u>GenVision Pro</u>.

Feature Type	Color
gene	green
CDS	blue
RNA	orange
repeat	yellow
transcription	red-purple
all others	orange-red

<u>To change Features options:</u>

To learn how to access the options section for this track, see Tracks.
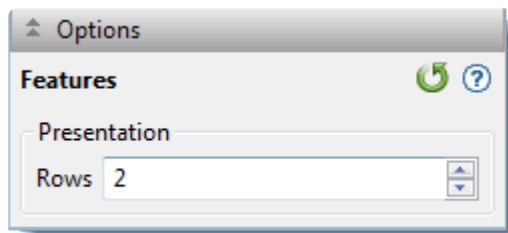
Specify the number of feature rows to display by typing a number in the **Rows** box or by using the up/down arrows. If features are overlapping, this indicates a need to increase the number of rows.

The example below shows three rows of features. The semi-opaque scrollbar to the right allows you to scroll up/down to see additional features.
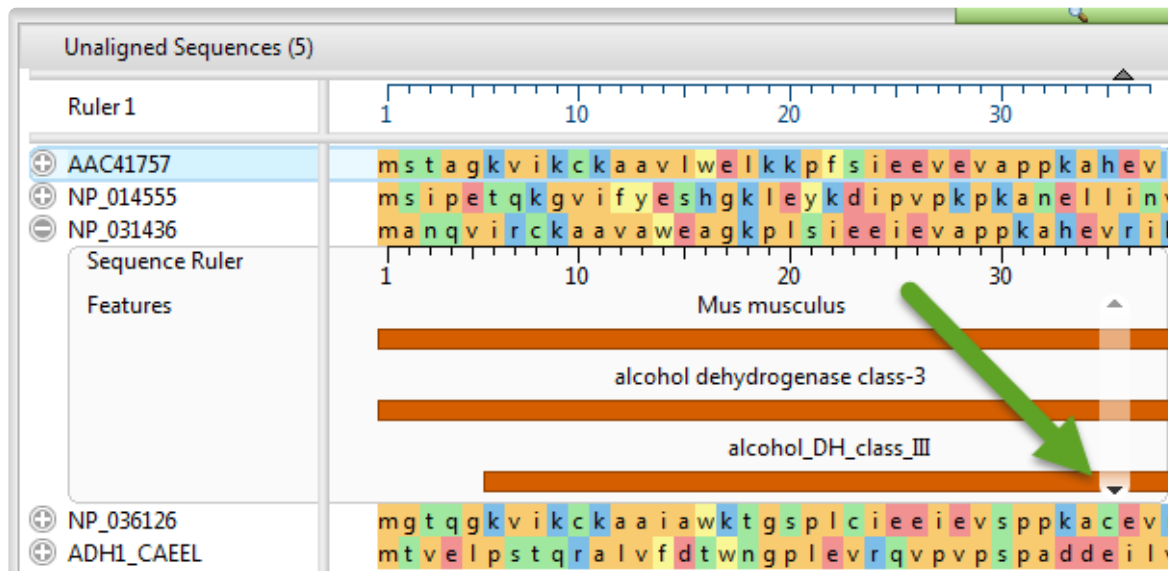
You can also add white space above and/or below features tracks using sliders in the Layout section.

Click if you wish to return to the default values.

Reference tracks

MegAlign Pro provides the option to specify a reference sequence and then use it to guide a MAFFT alignment or act as the comparison sequence for variant analysis. By default, the reference sequence is displayed as a track in the Overview and/or Sequences view headers. The sequence name is shown in bold blue font.

To show/hide this reference sequence, check or uncheck the **Reference** boxes in the Tracks panel. Checking the box will not do anything unless you have already specified a reference sequence.

To change Reference options:

These options let you change the font style and color for the reference sequence's name in the Sequences view and Overview.
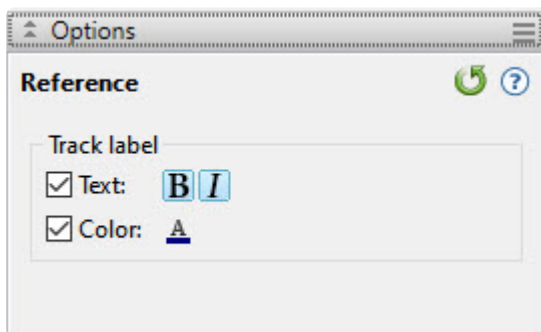
- To control bolding and italicization, check the **Text** box and use the **B** and/or **I** buttons. Uncheck the box if you want to use normal font.

- To control the font color, check the **Color** box and click on the **A** to open the color picker. Uncheck the box if you want the font color to be black.

You can also add white space above and/or below the reference track using sliders in the Layout section.

Consensus track

To display the consensus:

After any type of alignment, a consensus sequence is automatically displayed in the Sequences view header. To display or hide this consensus sequence, check or uncheck the **Consensus** box in the Tracks panel. This box is only visible after performing an alignment.

Consensus characters in upper case denote that the called character comprises > 50% of the total count for the column. This therefore disqualifies any character that shares the same frequency with another character.
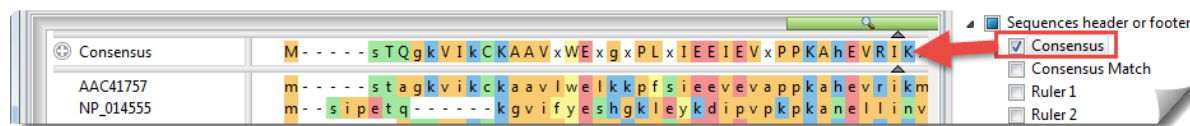
To change Consensus options:

Currently, the Consensus track does not have editable options.
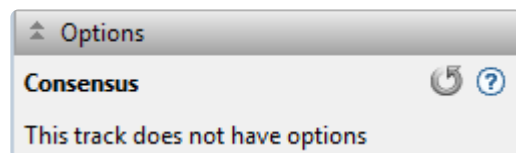
Consensus Match track

To display the Consensus Match:

To display or hide the Consensus Match track in the Sequences view footer, perform an alignment and then check or uncheck the **Consensus Match** box in the Tracks panel. This box is only visible after performing an alignment.

How Consensus Match is calculated:

The Consensus Match histogram is calculated by dividing the total score for the called consensus character by the number of sequences at the position. As agreement increases, the bar height increases and appears in a lighter shade of green. The histogram is not calculated for positions where there are consensus gaps. Hover over any bar in the histogram to display an information balloon showing the consensus match percentage, as demonstrated in the image below.
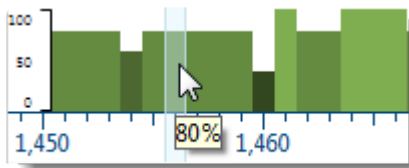
The following procedure is used to calculate the consensus sequence and the match percent for a given histogram column:

1. Score the characters in the column - Each occurrence of a non-ambiguous character in the column is scored as "1" for that character. If the character is an ambiguity code, a fractional count is added to each of the relevant characters. For example, in a nucleotide sequence, the character R would add "1/ 2" to the counts for A and G. Likewise, B would add "1/3" to the counts for C, G, and T.

2. Use the counts to determine the consensus character - If there is a single un-ambiguous character that is the most frequent, the consensus is called for that character. If there are two or more characters whose scores are tied for the maximum, the consensus is called as "X" or "n," for proteins or nucleic acids, respectively. If there is a two-way tie, and one of the characters is a gap, the non-gap character will be called.

3. Determine the case of the consensus character - If the count for the called character is > 50% of the total count for the column, upper case is used. Otherwise, lower case is used.

4. Calculate the consensus match for the histogram; The histogram calculation uses the "maximum count" value, which will either be the count of the single predominant character (the unambiguous consensus) or will be the "tied" value that led to the calling of an ambiguity code. The final consensus match is the total score for the called character divided by the number of sequences at the position.

Examples:

Characters in the DNA alignment column	Results		
	Consensus call	Consensus match calculation	Histogram bar size (as % of total available height)
A C G Y	c	1.5/4	38%
A A C C G T	n	2/6	33%

To change Consensus Match options:

To learn how to access the options section for this track, see Tracks.
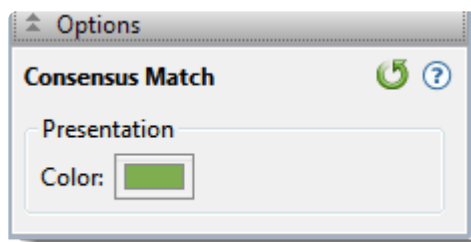
- By default, the Consensus Match histogram is shown in tones of green. To choose another color, click on the color box to the right of **Color**.

Click if you wish to return to the default values.

Sequence Logo track

To display the Sequence Logo track:

To display or hide the Sequence Logo track in the Sequences view footer, perform an alignment and then check or uncheck the **Sequence Logo** box in the Tracks panel. This box is only visible after performing an alignment.

How Sequence Logo is calculated:

The Sequence Logo is a visualization of the nucleotide or amino acid at a given position, based on a formula for the Shannon uncertainty measure given by Schneider and Stephens (1990). The height ranges from 0-2 for nucleotides and 0-4.3 for proteins, and the height of each letter code is proportional to its frequency. The most common letter code is at the top of each position. The accuracy of this calculation is proportional to the number of sequences in the alignment, and the track may not be displayed if there are too few sequences.

To avoid implying conservation in the presence of gaps, the Sequence Logo algorithm adjusts the height of the letter stack such that the total height of the symbols is multiplied by the fraction (# of residues / (# of residues + # of gaps)).

Hover over the logo to reveal a tooltip with the information content of each letter at the position.
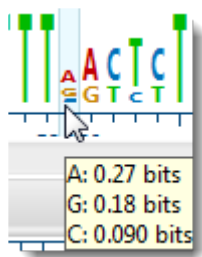
To change Sequence Logo options:

To learn how to access the options section for this track, see Tracks.
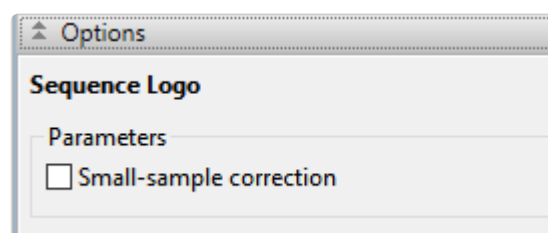
When there are relatively few sequences without gaps represented at a given position, we recommend

checking **Small sample correction** . This correction has no effect when there are sufficient numbers of sequences to provide a good estimate of information content.

Click ⟳ if you wish to return to the default values.

GC Content track

To display the GC Content track:

GC Content is a graph track with a scale from 0-100 showing the proportional amount of G or C residues in a sliding window of user-defined width.
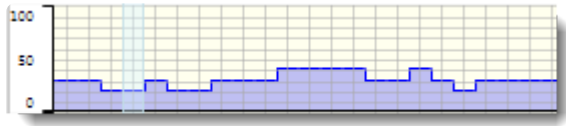
The visibility of GC Content tracks in the Sequences view is available only for nucleotide (DNA) sequences, and is dependent on:

- Whether or not tracks are visible in the Sequences view. To display tracks, click on the plus sign next to the sequence of interest.

- Whether or not the **GC Content** box is checked in the Tracks panel. This box is only visible for nucleotide sequences.

To change GC Content options:

To learn how to access the options section for this track, see Tracks.
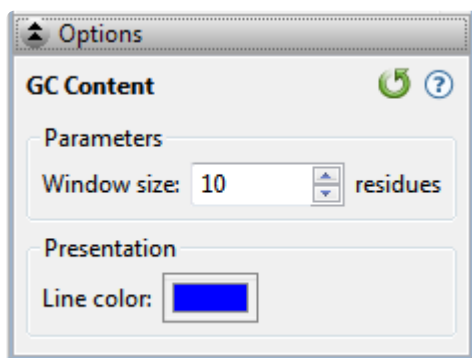
- Calculations are performed in a sliding window. Type a number in the **Window size** box to specify the number of consecutive residues to use in the GC content calculation. The window size may be any integer from 1-5,000, and the default is 25 residues. The window shifts along sequences one position at a time, and will therefore overlap for widths greater than 1.

> ✱ **Note:** the GC value at each residue is computed using a window centered over the residue in the ungapped sequence. An even window width causes the window to be slightly off-

> center. As the window approaches either end of the sequence, one side will stop moving, and the window will shrink.

- By default, GC Content is shown in tones of blue. To choose another color, click on the color box to the right of **Line color**.

Click ⟳ if you wish to return to the default values.

Gap Fraction track

To display the Gap Fraction track:

Gap Fraction is a graph track with a scale of 0-1 that shows the proportion of gapped positions in a sliding window centered on each position.
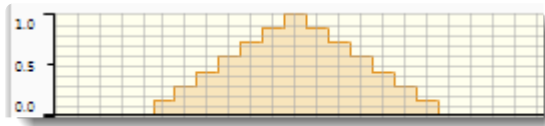
The availability of a Gap Fraction track in the Sequences view or Pairwise view is available only for aligned sequences, and is dependent on:

- Whether or not tracks are visible in the views. To display tracks, click on the plus sign next to the sequence of interest.

- Whether or not the **Gap Fraction** box is checked in the Tracks panel. This box is only visible after alignment has taken place.

To change Gap Fraction options:

To learn how to access the options section for this track, see Tracks.
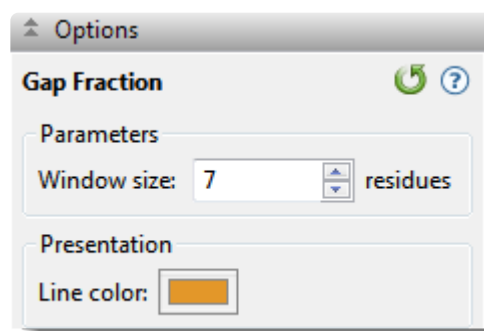
- Calculations are performed in a sliding window. Type a number in the **Window size** box to specify the number of consecutive residues to use in the gap fraction calculation. The window size may be any integer from 1-50,000, and the default is 7 residues. The window shifts along sequences one position at a time, and will therefore overlap for widths greater than 1.

> ✳ **Note:** The gap fraction value at each position is computed using a window centered over the position in the gapped sequence. An even window width causes the window to be slightly off-center. As the window approaches either end of the sequence, one side will stop

> moving, and the window will shrink.

- By default, the Gap Fraction graph is shown in gold. To choose another color, click on the color box to the right of **Line Color**.

Click if you wish to return to the default values.

Numeric tracks

A numeric track displays a plot of values along one or more sequences in the project.

To display numeric tracks:

The visibility of numeric tracks in the Sequences view is dependent on:

- The origin of the project. For example, the Model 'n' Cα Distances track is only available for the query sequence in a project sent to MegAlign Pro using NovaFold's **Send template alignment to MegAlign Pro** link. It is not available for template sequences in those projects, nor for MegAlign Pro projects that do not originate from NovaFold.

- Whether or not tracks are visible in the Sequences view. To display tracks, click on the plus sign next to the sequence of interest.

- Whether or not corresponding box is checked in the Tracks panel.

When the track is visible, it consists of a graph. For example, the Model 'n' Cα Distances track is a graph with a y-axis from 0-16. It shows the predicted pairwise distances between Cα atoms in a predicted model and its theoretical ideal structure. These are calculated by NovaFold using a support vector machine trained to recognize patterns in template coverage, secondary structure, solvent accessibility, and simulated structural diversity. A separate track is available for each model. Models tend to be more accurate in regions with lower distance values.
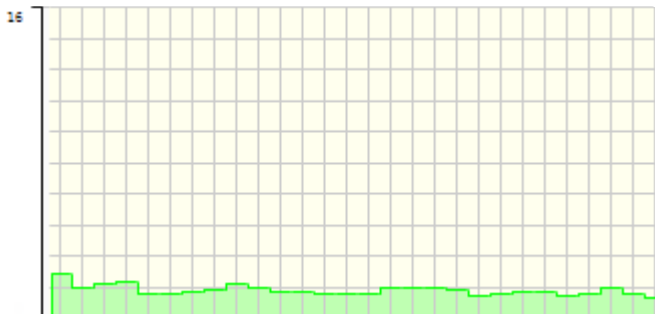
To change numeric track options:

To learn how to access the options section for this track, see Tracks.
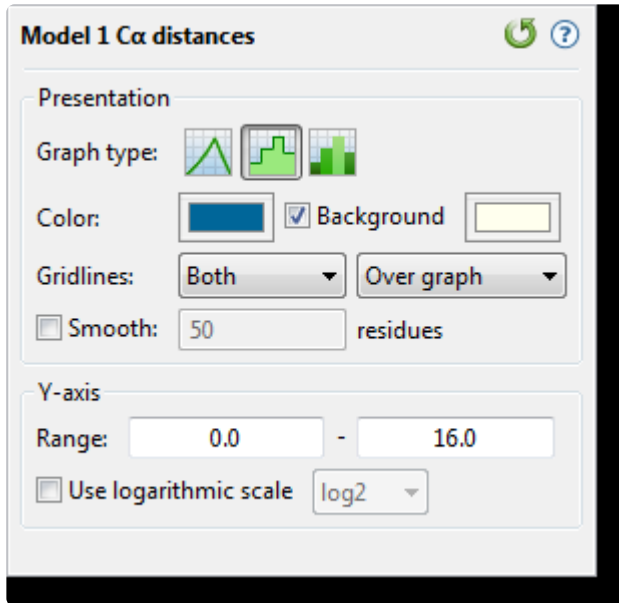
- Select from three different **Graph types** (**Line**, **Skyline**, **Bar**) by clicking on the corresponding icon.

- By default, numeric graphs are displayed in green with a pale yellow background. To choose another graph color, click on the box to the right of **Color** to make a selection from a color chart. To choose another background color, click the color box to the right of the **Background** checkbox. To turn off the background color, uncheck the checkbox.

- Use the first **Gridlines** dropdown menu to select which gridlines to display: **Both** (horizontal & vertical), **Horizontal** (only) or **None**. Use the second dropdown menu to choose whether gridlines should be visible above (**Over graph**) or below (**Under graph**) other parts of the graph.

- Check the **Smooth** box if you wish to average the values over a specified number of residues. Smoothing is computed by averaging values in a sliding window of the specified number of residues. Enter the number of residues in the adjacent text box (maximum=5000). Inputting a high value will smooth the graph curves, while a low value will highlight local differences.

- To change the range of values displayed, enter new y-axis endpoints in the **Range** text boxes. The default range may be too large or too small for some sections, causing the graph to appear empty or peaks to be cut off. You may need to raise or lower the second value, in particular, in order to see the graphical display clearly.

- Check the **Use logarithmic scale** box and select an option from the corresponding drop-down menu if you want to use a logarithmic, rather than a linear scale. Options are **log2** or **log10**.

> ✳ **Note:** The log of 0 is not defined. Therefore, log values for the track are calculated as follows: For $y > 0$, **y' = log(y)**; otherwise, it is displayed as a **0**.

Click if you wish to return to the default values.

Match Bar track

To display the Match Bar track:

The Match Bar track is used to show similarities and differences between the query and target sequences as calculated using the Blosum 62 substitution matrix. This track is available in the Pairwise view only after performing a pairwise alignment.
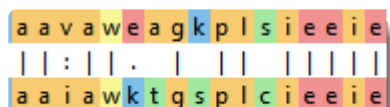
The visibility of this track in the Pairwise view is dependent on:

- Whether or not tracks are visible in the Pairwise view. To display tracks, click on the plus sign next to the sequence of interest.

- Whether or not the **Match Bar** box is checked in the Tracks panel. This box is only visible after performing a pairwise alignment.

To learn how to access the options section for this track, see Tracks.
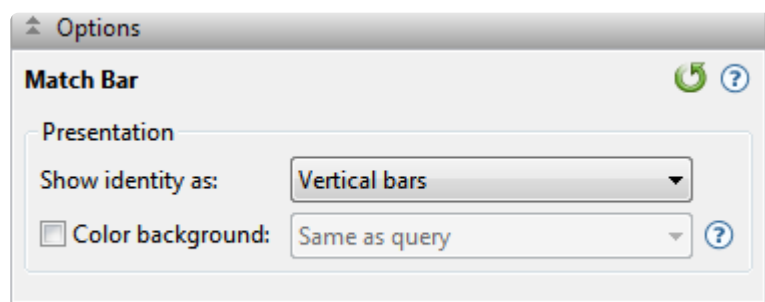
- Use the **Show identity as** drop-down menu to choose how to show the connection between the query and the target. The identity calculation is an estimate for the significance of the change. Choosing **Residues in target case** or **Residues in query case** cause the connection to appear as a copy of the residue from the target or query sequence. Choosing **Vertical bars** causes the connection to appear as a symbol, depending on the match status.

Symbol	Interpretation	Weight table value
Vertical line	Match	1

Colon	Favorable mismatch. The change is likely to have little or no impact on the protein because the amino acids have similar chemistry. An example of a favorable mismatch is a change from valine to isoleucine.	(0,1)
Period	Neutral mismatch. This change may be important or not, depending on the situation. An example of a neutral mismatch is a change from glutamate to lysine, both of which carry a charge. If the residue was on a non-binding protein surface, the change would qualify as a "favorable mismatch" (see above). By contrast, if it was on a binding site or in a hydrophobic region, the change might qualify as an "unfavorable mismatch" (see below). Thus, the change is categorized as a "neutral mismatch."	0
Space	Unfavorable mismatch. The change is from polar to non-polar, or vice versa, and will nearly always have a noticeable effect. Unfavorable mismatches may indicate proximity to another residue that had a similar change.	< 0

• To add a background color behind each residue, check the **Color background** check box and use the adjacent drop-down menu to choose between **Same as query** and **Same as target**.

✳ **Note:** These options are mirrored in the Pairwise alignment section of the Style panel. Choices made in either location are instantly reflected in the other location.

Click ⟳ if you wish to return to the default values.

Translation track

To display the Translation track:

Translation tracks in the Sequences view are available only for nucleotide (DNA) sequences, and are dependent on:

- Whether or not tracks are visible in the Sequences view. To display tracks, click on the plus sign next to the sequence of interest.

- Whether or not the **Translation** box is checked in the Tracks panel. This box is only visible for nucleotide sequences.

Translation track conventions:

- Stop codons are represented by dots.

- For purposes of translation, alignments are assumed to be linear.

- Translations are not displayed for incomplete codons at the right end of the block.

- Translation frames 4-6 are counted from the right end (the 3' end of the bottom strand).

- In multi-block Mauve alignments, translation is not done across the boundaries of blocks.

To change Translation options:

To learn how to access the options section for this track, see Tracks.
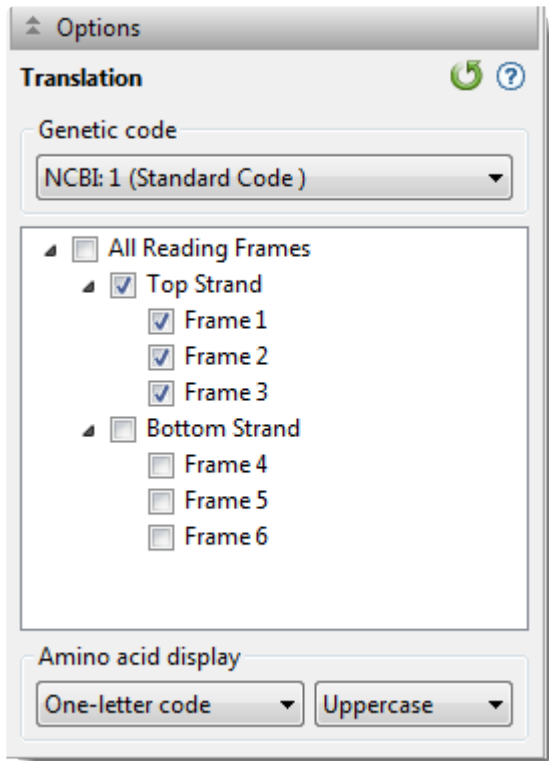
- Use the upper drop-down menu to choose from a variety of NCBI and Lasergene translation codes.

- Put checkmarks next to the reading frames you wish to display, and remove checkmarks from those you do not wish to display. Click **All Reading Frames** to show or hide all reading frames. Click **Top Strand** or **Bottom Strand** to show or hide all the reading frames on that strand.

- Use the left-hand drop-down menu to choose between **One-letter** or **Three-letter** amino acid abbreviations.

- Use the right-hand drop-down menu to choose between **Uppercase** and **Lowercase** letters.

Click if you wish to return to the default values.

Biophysics tracks

The Tracks panel lets you choose which biophysics tracks to apply to the Sequences view. The following links describe the methods and explain how to change their Track Options settings, if you desire to do so:

- Amphiphilicity
- Charge Density
- Disorder (JRONN)
- **Hydropathy:**
    - Hopp-Woods
    - Kyte-Doolittle
    - Parker
- **Secondary structure:**
    - Chou-Fasman
    - Deleage-Roux
    - Garnier-Robson
- **Stability:**
    - Aliphatic index
    - Instability index
    - Isoelectric precipitate

Amphiphilicity (Eisenberg)
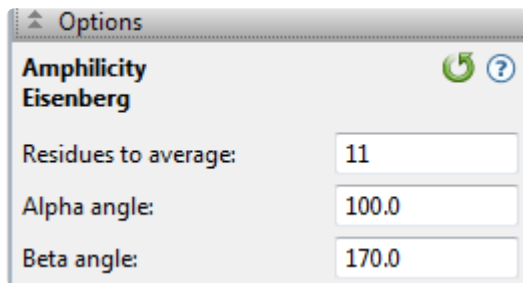
The Amphiphilicity (Eisenberg) track predicts the Eisenberg Moment using the approach of Eisenberg et al., 1984. Hydrophobic moments are semi-empirical quantities based on computation and experimental measurements, which describe the distribution of hydrophilic and hydrophobic residue groups in a protein. They describe asymmetry of hydrophobicity, or amphiphilicity.

To apply this track to the sequence:

In the Tracks panel, check the box next to *Amphiphilicity (Eisenberg)*. The track will now be visible in the Sequences view.

To edit track options:

Select the track in the Tracks panel. Open the Track Options section, which appears as follows:

- **Residues to Average** - the number of residues to be averaged for the hydrophobic moment and hydrophobicity. The default is 11.

- **Alpha Angle** and **Beta Angle** - the angles used in moment and amphipathic region plots to assign emergence angles for successive side chains. Only change these options if you are looking for a structure with atypical α-helix and ß-sheet angles, such as a 3-10 helix. The defaults are 100 and 170, respectively.

Click if you wish to return to the default values.

Charge Density (Lehninger)
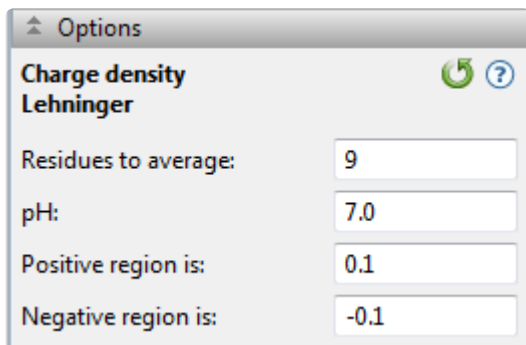
The Charge Density (Lehninger) track predicts regions of positive and negative charge by summing charge over a specific range of residues. DNASTAR developed this method using the pK tables of Lehninger, 2005. Since charged residues tend to lie on the surfaces of proteins, this method aids in predicting surface characteristics. A set of data is generated that can be displayed three ways: a line plot graph of average charge, a region graph of negative charge and another region graph of positive charge. The method is sensitive to small changes in parameter settings, so you may wish to set up multiple plots to compare how parameter changes affect each plot.

To apply this track to the sequence:

In the Tracks panel, check the box next to *Charge Density (Lehninger)*. The track will now be visible in the Sequences view.

To edit track options:

Select the track in the Tracks panel. Open the Track Options section, which appears as follows:

- **Residues to Average** – the size of the window used to average the charge on consecutive residues. Inputting a high value will smooth the curve, while a low value will make the method more sensitive to charge differences between adjacent amino acids. The default is 9.

- **pH** – desired pH for calculating pK values for each residue. The default is 7.0.

- **Positive Region is** - the threshold value for a region to be considered positive. The default is 0.1.

- **Negative Region is** – the threshold value for a region to be considered negative. The default is -0.1.

Click if you wish to return to the default values.

Disorder (JRONN)

The Disorder (JRONN) track predicts disordered regions of proteins, which are unlikely to form well-defined three-dimensional structures. These regions may affect both solubility and the ability to crystallize. DNASTAR developed this track based on the regional order neural network (RONN) analysis method of Troshin et al., 2011, and Yang et al., 2005.

To apply this track to the sequence:

In the Tracks panel, check the box next to *Disorder (JRONN)*. The track will now be visible in the Sequences view.

To edit track options:

There are no editable options for this track.

Hydropathy – Hopp-Woods

The Hydropathy – Hopp-Woods track predicts protein antigenic determinants by searching protein sequences for the area of greatest local hydrophilicity. Hopp and Woods, 1981, make two assumptions in using hydrophilicity to find antigenic determinants: 1) antigenic determinants are usually found on regions with a high degree of exposure to solvents and 2) antigenic determinants commonly possess charged hydrophilic side chains. Each residue is assigned a hydrophilicity value and these values are averaged over a hexamer window.

To apply this track to the sequence:

In the Tracks panel, expand *Hydropathy* and check the box next to *Hopp-Woods*. The track will now be visible in the Sequences view.

To edit track options:

Select the track in the Tracks panel. Open the Track Options section, which appears as follows:
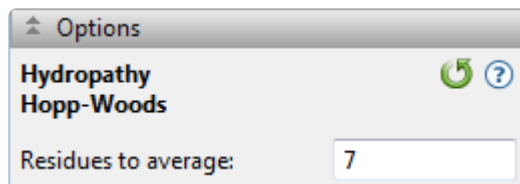
- **Residues to Average** – the size of the window over which hydropathy values are averaged. The MegAlign Pro default is 7, while the original authors use 6. There is very little difference in antigenic site prediction with this change, except that the larger default value requires a slightly larger antigenic site. The authors recommend setting this value between five and eight. If you set the value below five, too many regions may be falsely predicted. Conversely, values set above eight can create a smooth curve, making it difficult to locate areas of potential antigenic sites.

Click if you wish to return to the default value.

Hydropathy – Kyte-Doolittle

The Hydropathy – Kyte-Doolittle track predicts regional hydropathy of proteins from their amino acid sequences, using the approach of Kyte and Doolittle, 1982. Hydropathy values are assigned for all amino acids and are then averaged over a user-defined window. The average is plotted at the midpoint of the window. Residue hydropathy assignments are derived from water-vapor transfer free energies and the interior-exterior distribution of residue side-chains.

To apply this track to the sequence:

In the Tracks panel, expand *Hydropathy* and check the box next to *Kyte-Doolittle*. The track will now be visible in the Sequences view.

To edit track options:

Select the track in the Tracks panel. Open the Track Options section, which appears as follows:
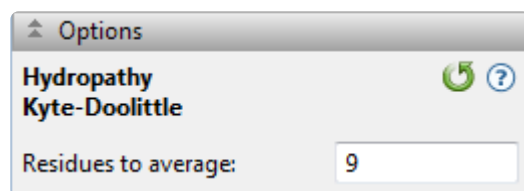
- **Residues to Average** – the size of the window over which hydropathy values are averaged. It is recommended you select 7 to 11 residues. The default is 9.

Click if you wish to return to the default value.

Hydropathy – Parker

The Parker track uses the method of Parker et al., 1986, which is a hydrophilicity scale based on high-performance liquid chromatography retention times of model synthetic peptides. Hydrophilicity measures have been used extensively in the prediction of antigenic amino acid residues. The Parker method uses a modified Hopp-Woods algorithm together with a new set of hydrophilicity options.

To apply this track to the sequence:

In the Tracks panel, expand *Hydropathy* and check the box next to *Parker*. The track will now be visible in the Sequences view.

To edit track options:

Select the track in the Tracks panel. Open the Track Options section, which appears as follows:
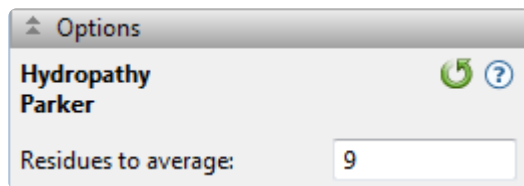
- **Residues to average** – number of residues to be averaged when constructing each point in the plot. The default is 9.

Click if you wish to return to the default value.

Secondary Structure – Chou-Fasman

The Secondary Structure – Chou-Fasman track predicts secondary structure of proteins from the crystallographic structures of their amino acid sequences.

Secondary structure prediction methods from the1970s and 1980s rely on the propensities of individual amino acids observed in either alpha helices or beta strands/sheets. For a local region in the protein sequence, these propensity values are summed for each individual class. The secondary structure for central residue of the region is predicted by comparing whether propensity sum is greater than the threshold.

Chou and Fasman, 1978, identified several protein classes, with different data sets used to train the model for each class. The class for 29 Proteins, later augmented to 64 Proteins (Chou, 1990), is the more general model. The other four classes are more appropriate for specific types of protein folds: Alpha, Beta, Alpha plus Beta and Alpha/Beta (i.e., some combination of the two structural elements).

To apply this track to the sequence:

In the Tracks panel, expand *Secondary Structure* and check the box next to *Chou-Fasman*. The track will now be visible in the Sequences view.

To edit track options:

Select the track in the Tracks panel. Open the Track Options section, which appears as follows:

Choose which conformation parameter table to use:

- **29 Proteins** – to use parameters from only 29 protein structures.

- **64 Proteins** – to use parameters from all 64 protein structures.

- Alpha Class, Beta Class, Alpha plus Beta Class or Alpha/Beta Class* – to use the Chou classification method parameter tables.

If you select **29 Proteins** or **64 Proteins**, enter numbers in the Threshold parameters section:

- **Alpha region threshold** – P_α has a default of 103.

- **Beta region threshold** – P_β has a default of 105.

These values are constants which limit whether a region is defined as α-helix or ß-sheet. If P_α > 103 and P_α > P_β a region is predicted as helical. If P_β > 105 and P_β > P_α a region is predicted as a beta sheet. The default values are from Chou and Fasman (1978).

Click ![icon] if you wish to return to the default value.

Secondary Structure – Deleage-Roux

The Secondary Structure – Deléage-Roux track predicts secondary structure using the approach of [Deléage and Roux, 1987](). The method uses an independent prediction of the protein's structural class to bias the prediction of its secondary structure. MegAlign Pro begins its calculations by predicting the class of the protein from its amino acid composition. Class designations are identical to those of [Nakashima et al., 1986](), except for α/ß class proteins. MegAlign Pro divides α/ß class proteins into two subclasses, α/ß-α and α/ß-ß, according to their distance from the α and ß classes. The letter R is used by Deléage and Roux to designate the "irregular" class of proteins. Next, secondary structure is predicted from the frequency of occurrence of each residue in the four states (helix, sheet, turn and coil) using the secondary structure database of [Kabsch and Sander]().

To apply this track to the sequence:

In the [Tracks panel](), expand *Secondary Structure* and check the box next to *Deleage-Roux*. The track will now be visible in the [Sequences view]().

To edit track options:

Select the track in the Tracks panel. Open the [Track Options section](), which appears as follows:
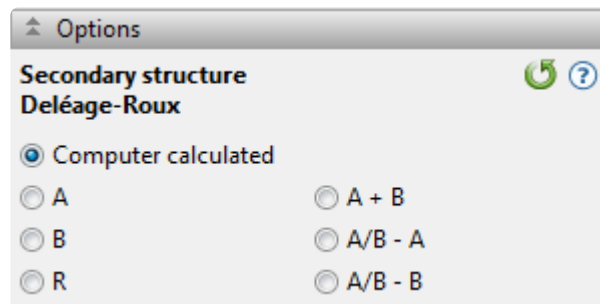
- If you know the structural class for your protein (**A**, **B**, **A+B**, **A/B – A**, **A/B – B**, **R**), choose it by clicking on a button.

- Use the default, **Computer Calculated**, if you do not know to which class your protein belongs.

Click if you wish to return to the default value.

Secondary Structure – Garnier-Robson

The Secondary Structure – Garnier-Robson track , from [Garnier et al. (1978), Garnier and Robson (1990), and Garnier, Gibrat and Robson (1996)](#), predicts protein structure from the amino acid sequence. The method examines the propensity of a given residue to exist in a certain structure. It is a statistical approach which uses examination of known crystallographic structures to define regions as α-helical (H), ß-pleated sheet (E - extended chains), ß-turns (T) and coil (C).

To apply this track to the sequence:

In the [Tracks panel](#), expand *Secondary Structure* and check the box next to *Garnier-Robson*. The track will now be visible in the [Sequences view](#).

To edit track options:

Select the track in the Tracks panel. Open the [Track Options section](#), which appears as follows:
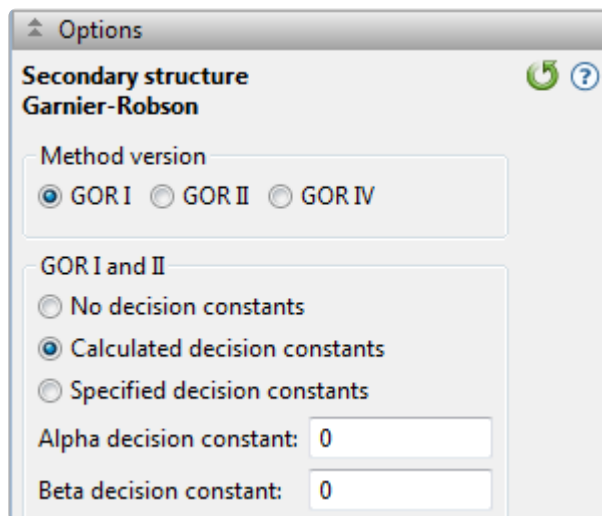
Specify which version of the GOR method you wish to use: **GOR I**, **GOR II** or **GOR IV**. GOR II uses the same algorithm as GOR I, but is based on updated propensity tables. GOR IV does not predict turns, but only predicts the probability of areas being helices, sheets or coils. GOR IV is the most modern secondary structure prediction method in MegAlign Pro.

If you choose GOR I or GOR II, you may select from among three options related to the decision constant:

- **No Decision Constants** - if you do not want MegAlign Pro to make assumptions about the global α-helix and ß-sheet content.

- **Calculated Decision Constants** - to use computer-derived constants from global α-helix and ß-sheet

probabilities. Constants are based on three protein classes: proteins with less than 20%, 20-50%, and greater than 50% α-helix or ß-sheet.

- **Specified Decision Constants** - if you have prior knowledge of circular dichroism data. If you select this button, you need to input values for **Alpha Decision Constant** and **Beta Decision Constant**.

Click if you wish to return to the default value.

Stability – Aliphatic Index

The Stability – Aliphatic Index track (Gasteiger et al. 2005) predicts regional stability by calculating the relative volume occupied by aliphatic side chains. This is a positive indicator of globular protein thermostability. This method is only available for amino acid selections. Detailed information about this method is available on the ExPASY website.

To apply this track to the sequence:

In the Tracks panel, expand *Stability* and check the box next to *Aliphatic Index*. The track will now be visible in the Sequences view.

To edit track options:

Select the track in the Tracks panel. Open the Track Options section, which appears as follows:
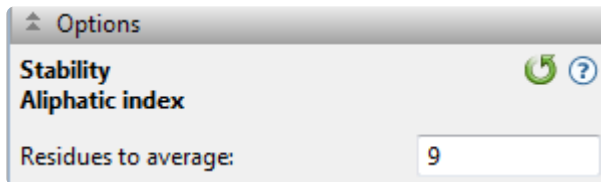
- **Residues to average** – number of residues to be averaged when constructing each point in the plot. The default is 9.

Click if you wish to return to the default value.

Stability – Instability Index

The Stability – Instability Index track estimates the stability of the protein in a test tube. The Instability Index is calculated using the approach of Guruprasad et al., 1990,, which predicts regional instability by calculating the weighted sum of dipeptides that occur more frequently in unstable proteins when compared to stable proteins. Detailed information about this method is available on the ExPASY website. An instability index less than 40 predicts a stable protein, whereas values higher than 40 denote a potentially unstable protein.

To apply this track to the sequence:

In the Tracks panel, expand *Stability* and check the box next to *Instability Index*. The track will now be visible in the Sequences view.

To edit track options:

Select the track in the Tracks panel. Open the Track Options section, which appears as follows:
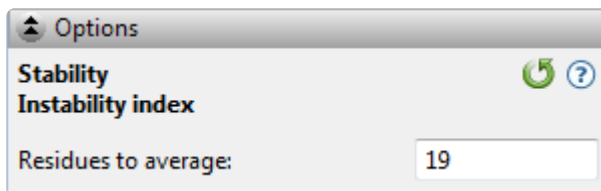
- **Residues to Average** – number of residues to be averaged when constructing each point in the plot. The default is 19. The closer a calculation is to this value, the more likely the chance of aggregation. Over small windows, this method gives an idea of the potential for local instability.

Click if you wish to return to the default value.

Stability – Isoelectric Precipitate
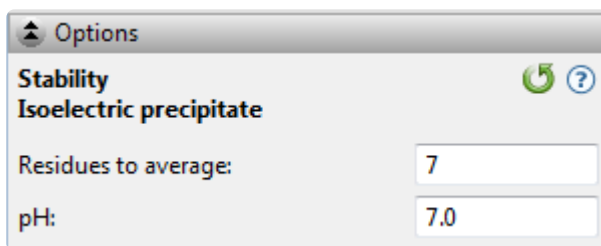
The Stability – Isoelectric Precipitate track predicts regional instability by identifying uncharged regions at a given pH, which have an increased potential to aggregate if surface exposed. This method uses the pKa tables from Lehninger et al., 2005.

To apply this track to the sequence:

In the Tracks panel, expand *Stability* and check the box next to *Stability – Isoelectric Precipitate*. The track will now be visible in the Sequences view.

To edit track options:

Select the track in the Tracks panel. Open the Track Options section, which appears as follows:

- **Residues to average** – number of residues to be averaged when constructing each point in the plot. The default is 7.

- **pH** – MegAlign Pro highlights regions with a pI ± 0.5 pH units from the pH specified here. The default is 7.0.

Click ⟳ if you wish to return to the default settings.

Immunogenicity tracks

The Tracks panel lets you choose which immunogenicity tracks to apply to the Sequences view. The following links describe the methods and explain how to change their Track Options settings, if you desire to do so:

- **Antigencity:**
  - Welling
  - Jameson-Wolf

Copyright DNASTAR, Inc. 2022   Page 195 of 384

Antigenicity – Welling

The Welling track predicts antigenicity using the method of Welling et al., 1985. Previous methods assumed that antigenic regions were mainly hydrophilic regions on the surface of a protein. By contrast, the Welling method is based on the percentage of each amino acid present in known antigenic determinants compared to the percentage of the amino acids in the average composition of a protein.

To apply this track to the sequence:

In the Tracks panel, expand *Immunogenicity*, then *Antigenicity*, and check the box next to *Welling*. The track will now be visible in the Sequences view.

To edit track options:

Select the track in the Tracks panel. Open the Track Options section, which appears as follows:
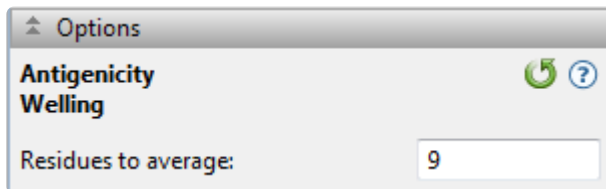
- **Residues to average** – number of residues to be averaged when constructing each point in the plot. The default is 9.

Click if you wish to return to the default value.

Antigenicity – Jameson-Wolf

The Jameson-Wolf track predicts potential antigenic determinants by combining existing methods for protein structural predictions, using the approach of Jameson and Wolf , 1988. Method results appear as multiple peaks in the Antigenic Index plot, with each peak signifying a potential antigenic determinant. The Antigenic Region graph summarizes the antigenic regions using a threshold value of 1.0. The Jameson-Wolf output is affected by parameter changes made to the Hopp-Woods, Chou-Fasman and Garnier-Robson methods.

To apply this track to the sequence:

In the Tracks panel, expand *Immunogenicity*, then *Antigenicity*, and check the box next to *Jameson-Wolf*. The track will now be visible in the Sequences view.

To edit track options:

Select the track in the Tracks panel. Open the Track Options section, which appears as follows:
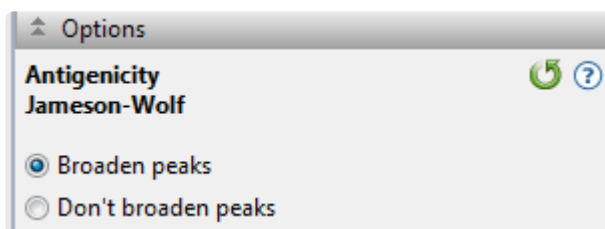
- **Broaden Peaks** - causes the following smoothing procedure to be applied: For values A[i] that exceed 1.0, where i is a sequence position, find the position p of the peak value within the following 9 residues. If neither structure prediction method predicts helix at position p, apply A[p+n] = A[p+n] + 0.2 * (5 - abs(n)) * A[p] for n = -4 to 4.

- **Don't Broaden Peaks** - causes the method to produce the default index of antigenicity by combining values for hydrophilicity (Hopp-Woods, H[<2 > 0.5 <1 > 0 <-1 > -0.5 <-2 >]), surface probability (Emini, S[<1 > 1.0 <0 >]), flexibility (Karplus-Schulz, F[<1 > 1.0 <0 >]), and the secondary structure predictions of Chou-Fasman (CF[T=2, t=1, 0]) and Garnier-Robson (GR[T=2, t=1, 0]) using the following formula: A = 0.3H + 0.15S + 0.15F +0.2CF + 0.2GR.

Click if you wish to return to the default value.

Features

MegAlign Pro provides a wizard for copying features from one sequence to another, a procedure known as "feature mapping." To learn more about how to map features or remove existing features, see:

- Map features

  ◦ Features, Options and Output wizard screens

  ◦ Files output after mapping features

  ◦ Mapping features in MegAlign Pro vs. SeqNinja

- Remove features

If there are feature types that are not of interest to you, you can tell MegAlign Pro to hide them in future documents. This is done using the **Features** tab of the Set Preferences dialog.

Map features (copy features to another sequence)

MegAlign Pro's feature mapping lets you map a single annotation or all annotations from a source sequence to a target sequence. The sequences involved must have been previously aligned. During the process, you may optionally filter annotations so as to include/exclude specific gene types. Feature mapping is most commonly used to map desired annotations from a "completely annotated" sequence to a closely-related but "incompletely annotated" one.

The following video provides a short overview on feature mapping:

To map features from one or more source sequences to a target sequence:

1. In Column 1 of the table below, locate the row that describes the target sequence—the sequence that will be "receiving" features from one or more other sequences.

2. As directed in Column 2, select one or more features or one or more sequences that will act as the source(s) for the features.

   • To select an individual feature, open the tracks for the sequence and click on the feature. To select two or more features, select them using **Ctrl+click** (Win) or **Cmd+click** (Mac).

   • To select one or more sequences from the Overview or Sequences View, use click, **Shift+click**, **Ctrl+click** (Win), or **Cmd+click** (Mac).

3. As directed in Column 3, use the specified **Features** menu command, or the corresponding "right-

click" context menu command.

Column 1: Target sequence	Column 2: Sequence or feature selection	Column 3: Command name
All sequences other than the designated reference	All features or selected features in the designated reference sequence (see Compute Variants)	**Features > Map (Selected) Feature(s) from Reference**
Active consensus	One or more features; or one or more source sequences. Note that feature segments falling entirely within a gap will not be copied.	**Features > Map (Selected) Feature(s) to Active Consensus**
Sequence in the top row of the view	A single feature; or one or more source sequences, none of which may be the sequence in the top row of the view.	**Features >) Map (Selected) Feature(s) to Top Sequence**
Sequence that is higher on the list than the source sequences	One or more features; or the target sequence and one or more source sequences.	**Features > Map (Selected) Feature(s) to Top Selected Sequence**
Sequence that is lower on the list than the source sequences	One or more features; or the target sequence and one or more source sequences.	**Features > Map (Selected) Feature(s) to Bottom Selected Sequence**
Sequence in the bottom row of the view	One or more features; or one or more source sequences, none of which may be the sequence in the bottom row of the view.	**Features > Map (Selected) Feature(s) to Bottom Sequence**

4. If more than two sequences are selected, the following popup will appear. Select **Yes** to launch the Map Features dialog and map all features from the specified source sequences to the target sequence; or **No** to abort feature mapping.
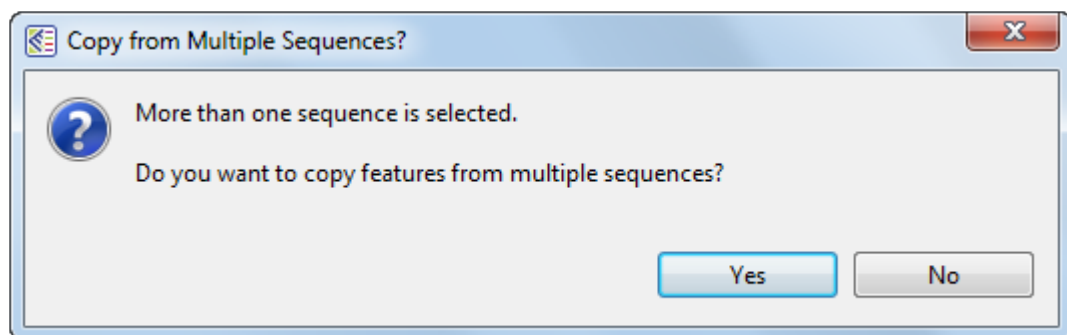
If you initiate feature-mapping for one or more individual features, MegAlign Pro maps the specific feature(s) you requested without further input.

When mapping a single feature, or mapping to the consensus, mapping happens automatically, without

launching a wizard. If you initiate feature-mapping in other situations, a wizard may appear. This Map Features wizard has three screens—Features, Options and Output described in a separate topic. To navigate between the three wizard screens, either use the **Back** or **Next** buttons or click on the screen name in the left panel of the wizard.

Things to note when mapping features:

- It is possible to map multiple copies of features to any sequence, including the consensus. For example, if you first map a single feature to the consensus, then later map all features from the same source sequence, the first feature that was added will appear twice on the consensus. Be aware of this, in case you wish to avoid adding multiple copies of the same feature.

- When mapping features to the consensus in Mauve multi-block alignments, only features within the active block are mapped. When mapping features to a sequence, features in all the blocks are mapped.

- The project directory will be organized into sub-directories when 1) there are multiple source sequences, or 2) any of the source sequences are multiple end-to-end sequences, or 3) the target includes multiple end-to-end sequences.

Features screen

The Features screen is the first wizard screen of the Map Features dialog, and allows you to choose the features you want to include or exclude in the mapping.

> ✳ **Note:** This dialog is optional, and you may click **Next**, if desired, to skip to the next screen. However, note that this will cause all features except the source features to be selected. In order to select all features, it is necessary to change the drop-down menu to **All features**, as described below the following image.
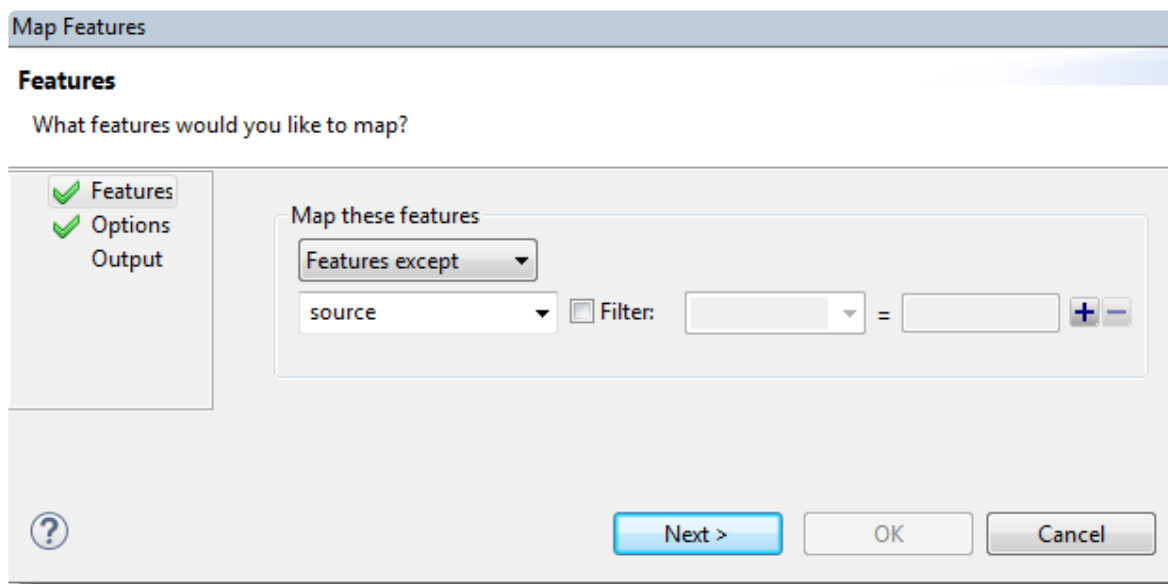
- Use the upper drop-down menu to elect to map **All features**, **Features matching** those you will specify in the subsequent row(s) of the dialog, or **Features except** those specified in the subsequent row(s).

- Use the lower drop-down menu to select the feature type to match (or *not* match), e.g., **gene**, **CDS**, **exon**, **intron**, **mRNA**, **tRNA**, **promoter**, **misc_binding**, etc. To specify more than one feature type, use the plus (+) button to add additional "Features" rows. If you want to further limit the search to features matching (or *not* matching) particular qualifiers, check the **Filter** box, then choose a qualifier (**gene**, **/product**, **/locus_tag**, **/note** or **/db_xref**) from the drop-down menu to its right. In the right-most textbox, enter the text that the qualifier must match or *not* match (e.g., */gene = thrL*) in order for the feature to be removed in the output file. You may use wildcards in this box if you wish (e.g., */gene = thr**).

The image below shows an example of how the screen might look after specifying the desired options:
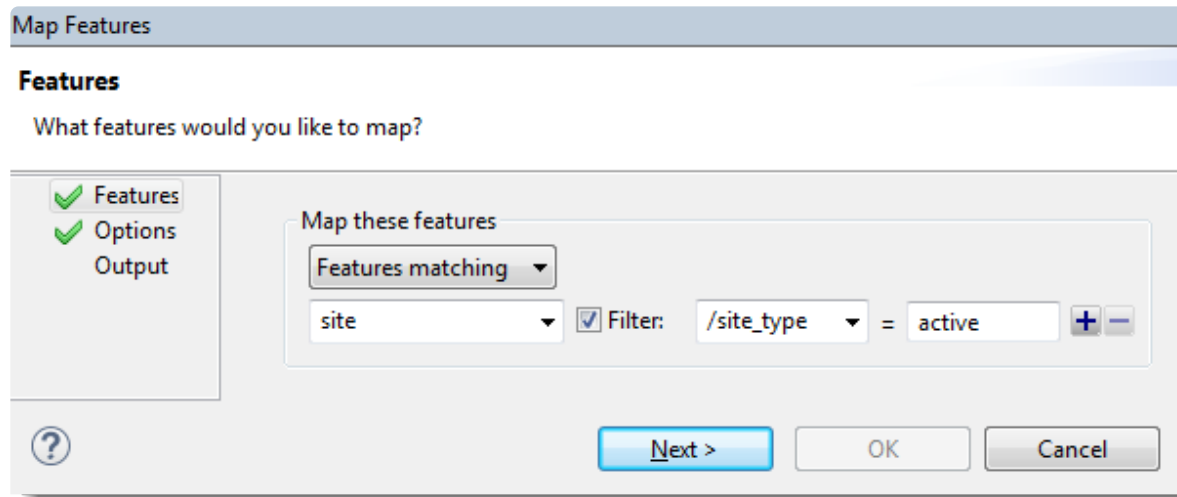
Click **Next** to move to the Options screen.

Options screen

The Options screen is the second wizard screen of the Map Features dialog, and lets you specify mapping options.

- Mapping – Use the text boxes or sliders to specify percent cutoffs for:

  - **Minimum feature coverage**, where:

$$\text{feature coverage} = \frac{\text{portion of feature contained within the aligned interval}}{\text{feature length}} \times 100\%$$

  - **Minimum sequence identity**, where:

$$\text{sequence identity} = \frac{\text{number of identical aligned residues}}{\text{length of feature in gapped alignment space}} \times 100\%$$

In this calculation, non-common gaps are counted in both source and target sequences.

Feature coverage is less sensitive to insertions and substitutions than is sequence identity. The default for both Mapping items is 80%.

- Translation – The genetic code specified in this dialog is used for translation, as well as for describing the type of sequence change in the Map Features Report file (see next bullet point). By default, MegAlign Pro translates using the genetic code identified in the source data (e.g., the GenBank /transl_table qualifier), if available; otherwise, it uses the standard genetic code (NCBI:1). To specify a different code, check **Override genetic code** and make a selection from the corresponding drop-down menu. This option does not appear for protein alignments.

- Feature Report – After running the Map Features program, the output is sent to the Map Features Report file. To add a column to that report containing a SeqNinja expression describing the bases in the mapped feature in terms of the source and target sequences, check the box next to **Include specific sequence changes**.

- Unmapped Interval Report – To make a report listing regions that may need further annotation, check the box next to **Generate report of intervals without mapped features**. Then use the text box **Minimum length of interval to report** to specify a minimum length for a stretch of sequence where features were not mapped.

Click **Next** to proceed to the [Output screen](#).

Output screen

The Output screen is the third wizard screen of the Map Features dialog, and lets you specify where to store reports and logs related to feature mapping.

- In the **Save project in** box, specify the directory that will contain the project directory. To use a different directory than the one listed, use the **Browse** button.

- In the **Project name** box, type in a name for the project.

The project directory will be created in the specified folder, with the specified name. The project directory cannot be an existing directory. Therefore, if you decide to repeat the mapping using different options, you will need to change the project directory before re-running.

Press **OK** to start feature mapping, or **Cancel** to leave the dialog. To learn about files output after feature mapping, click here.

Output files

After running the Map Features step, the output consists of several files.

File Name	Notes
PROJECT_map_features_console.txt	Copy of the text that appeared for this mapping in the Console side panel in the MegAlign Pro user interface.
PROJECT_map_features_report.tsv	See the text and table below this one for more information.
PROJECT_map_features_unmapped.tsv	Report of intervals in the target sequence to which no features were mapped. Generated only if requested in the Map Features Options screen under **Unmapped Interval Report**.

The *report.tsv* file is a tab-separated report file that can be opened in [Notepad++](#) (freeware) or any spreadsheet utility. This report file contains a row for each feature that you directed MegAlign Pro to map. The "Variation" column summarizes the most significant variation of the mapped feature against the source feature:

Category	Label in report	Comments
Not mapped	Not_mapped	Can occur when the feature doesn't map at all, or when the feature maps, but not well enough to pass the mapping thresholds.
Indels	Deleted_end_5prime Deleted_end_3prime Deleted_end Deleted_ends	[Non-protein sequences] First residue and/or last residue cannot be mapped with the current settings, or one/both align with a target gap.
	Indel	Both insertions and deletions have occurred.
	Insertion	An insertion has occurred.
	Deletion	A deletion has occurred.
Substitutions	No_start	[CDS only] The initial codon in the mapped feature is not a start codon.
	Nonsense	[CDS only] A substitution has resulted in an interior stop codon.
	No_stop	[CDS only] The last codon in the mapped feature is not a stop codon.
	Non-synonymous	[CDS only] Non-synonymous substitution occurred.
	Synonymous	[CDS only] Synonymous substitution occurred.
	Substitution	[Non-CDS] A substitution has occurred.
No change	Identical	

✳ **Note:** If more than one item in the above table applies, the topmost item is reported.

Mapping features in MegAlign Pro vs. SeqNinja

The ability to map features exists in both MegAlign Pro and DNASTAR's [SeqNinja](). Here are some reasons you might want to map features in one application versus the other:

Task	Use MegAlign Pro	Use SeqNinja
View the alignment and the mapped features	x	
Configure alignment options	x	
Map features in an end-to-end alignment. For instance, to perform a Mauve alignment of end-to-end chromosomes with some rearrangement.	x	
Map VCF features from a *.vcf* file.		x
Map features into a table, then edit the table before generating the resulting sequence or feature file		x

Remove features

While features can only be mapped when all sequences involved have been aligned, features can be removed from a sequence before or after aligning. Feature removal is not reversible, so use caution when removing features, especially those from non-consensus sequences.

To remove individual features from a sequence or from the consensus:

Click the plus sign to the left of consensus or sequence to reveal the associated Features track. Select one or more features to remove using **Ctrl+click** (Win) or **Cmd+click** (Mac), then right-click on the selection and choose **Remove Selected Feature(s)** from the context menu, or **Features > Remove Selected Feature(s)** or **Edit > Delete** from the main menu. You will be asked to confirm feature deletion.

To remove all features from one or more selected sequences:

- For regular sequences, select one or more sequences using **Ctrl+click** (Win) or **Cmd+click** (Mac), then right-click on the selection and choose **Features > Remove Features from Selected Sequence(s)**; or right-click on the consensus sequence and choose the same command from the context menu.

- For the consensus sequence, select it and choose **Features > Remove Features from Active Consensus**; or right-click on the consensus sequence and choose the same command from the context menu.

- For the reference sequence, select it and choose **Features > Remove Features from Reference Sequence**; or right-click on the reference sequence and choose the same command from the context menu.

Try it! – Map features

Both MegAlign Pro and SeqNinja allow you to copy a single annotation, or all annotations from a source sequence to a target sequence. During the process, you may optionally filter features so as to include or exclude specific gene types.

In this tutorial, you will experiment with feature mapping using two closely-related *Drosophila simulans* sequences: a "source" sequence (*.gbk*) that contains features and a "target" sequence (*.fas*) that lacks features.

Begin with Part A: Add, rename and align sequences.

Part A: Add, rename and align sequences

Before mapping features in MegAlign Pro, you first need to add and align the sequences.

1. If you have not yet downloaded and extracted the tutorial data, click here to download it. Then decompress (unzip) the file archive using the method of your choice.

2. If a MegAlign Pro project is already open, choose **File > New Alignment Project**. Otherwise, launch MegAlign Pro.

3. Choose **File > Add Sequences**, or click on the **Add sequences to project** tool ( ).

4. Navigate to and open the demo data folder *Lasergene 'x' Data\Demo MegAlign Pro\Drosophila project*, where 'x' denotes the Lasergene version. This folder contains two similar sequences for *Drosophila simulans* chromosome X.

5. Use **Ctrl+A** (Win) or **Cmd+A** (Mac) to both sequences and click **Open**.

6. To rename the first sequence, right-click on the sequence name in the Sequences view or Overview and select **Rename**. Type in the new **Label** (see table below) and press **OK**. Do the same for the second sequence.

Type	Original sequence label	New label
Feature source sequence (.gbk)	CM002914	**D. simulans PRJNA170244 chrX**
Target sequence (.fas)	gi|62998559|gb|CM00168.1|	**D. simulans PRJNA12464 chrX**

7. If sequences are not already ordered as in the image below, drag one above the other so they are.
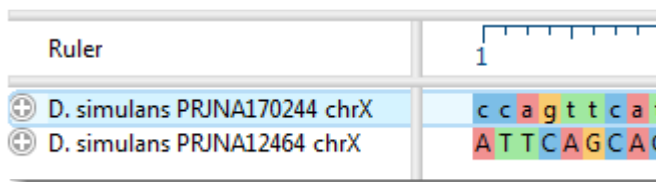
8. To align sequences with Mauve using the default parameters, choose **Align > Align Using Mauve** or click the **Align** tool arrow ( ) and choose **Align Using Mauve**.

9. Use **File > Save** to save the project with the name *Drosophila project.msa*.

Proceed to Part B: Map a single feature.

Part B: Map a single feature

In this part of the tutorial, you will map a single selected feature from the source sequence to the target sequence. In order to choose the appropriate mapping command, you will need to know where the source and target sequences appear in relation to one another in the list on the left of the views. If you followed the steps in Part A, the source sequence should be at the top of the two-member list.

1. To verify this, open the tracks panels by clicking both gray plus signs (⊕) in the Overview or Sequences view. If you followed the steps in Part A, the upper sequence should contain features and the lower sequence should not.
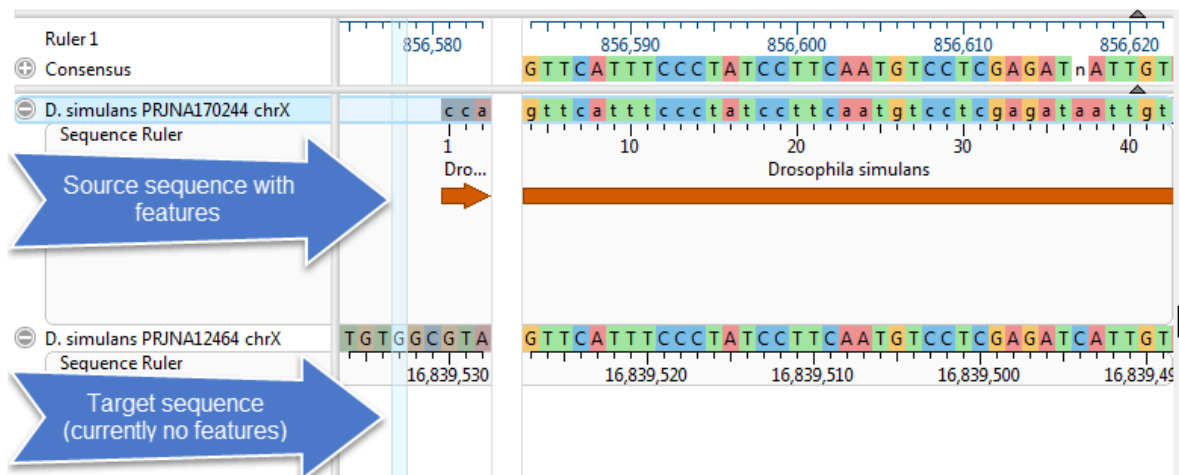
2. In the Overview, double-click on any block to make it the focus block.

   In the Sequences view, note that the letters (or bars, if zoomed out) with highly-colored backgrounds denote the active block selected in the previous step. Letters or bars with dull-colored backgrounds indicate an adjacent, non-focus block. Colored arrows represent features.
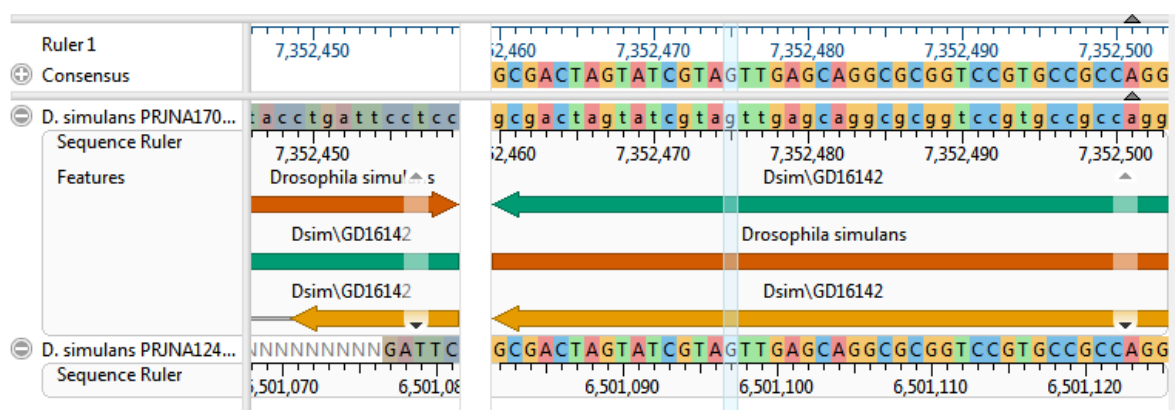
3. Keeping within the active block section, right-click on any feature and choose **Map Selected Feature to Bottom Sequence**. Observe that the feature is copied to the corresponding location in the lower

sequence. Repeat this with other features, if desired. If you don't see any features, go back to Step 2 and double-click on a different block to make it active.

4. To prepare for the next part of the tutorial, remove the feature(s) you just added to the target sequence by right-clicking on the lower sequence name and choosing **Features > Remove Features From Selected Sequence**. When prompted, click **Yes**. Observe that the target sequence has returned to its original state.

Proceed to [Part C: Map all features](#).

Part C: Map all features

In this part of the tutorial, you will map all features from the source sequence to the target sequence.

1. Right-click on the name of the source (upper) sequence and select **Features > Map Features to Bottom Sequence**. The Map Features wizard launches.

2. In the <u>Features screen</u>, use the drop-down menu to select **All Features**, then press **Next**.

3. In the <u>Options screen</u>, keep the defaults and press **Next**.

4. In the <u>Output screen</u>, type in the **Project name** "*All features*" and use the **Browse** button to specify where save the project.

5. Click **OK** to initiate feature mapping. When mapping is finished, click **OK** and **Close**. The output sequence will be examined later in this tutorial.

6. To prepare for the next part of the tutorial, remove the features you just added to the target sequence by right-clicking on the lower sequence name and choosing **Features > Remove Features From Selected Sequence**. When prompted, click **Yes**.

Proceed to <u>Part D: Map a filtered set of features</u>.
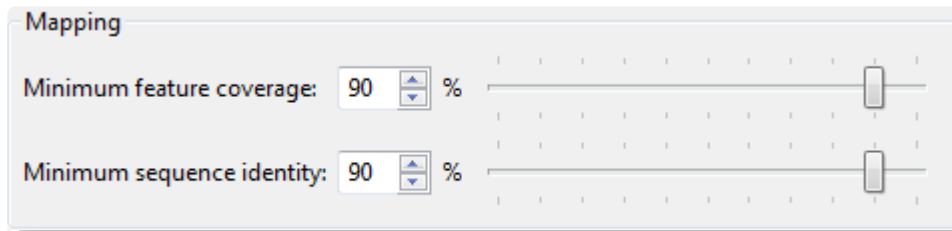
Part D: Map a filtered set of features

In this example, you will map only CDS features which have a note containing the letters RB.

1. Right-click on the name of the source (upper) sequence and select **Features > Map Features to Bottom Sequence**. The Map Features wizard launches.

2. In the Features screen, select the features you want to include or exclude in the mapping:

    a. From the upper drop-down menu, choose **Features matching**.

    b. From the menu just below that, select **CDS**.

    c. To restrict mapping to features that have a note containing the letters "RB", check the **Filter** box and select **/note** from the drop-down menu to its right. In the text box, type in ***RB***. The asterisk wildcards are used so that any text string containing RB will be included.

    d. Click **Next**.
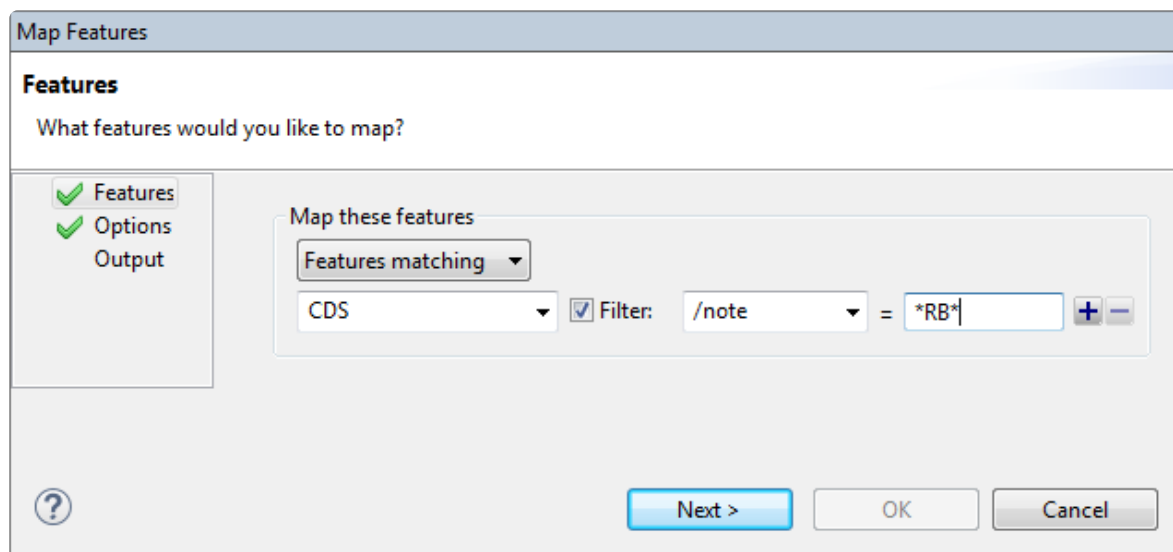
3. In the Options screen, the **Mapping** sliders and corresponding text boxes are used to specify the feature coverage and sequence identity thresholds that are required to map a feature. Change both items to **90%** to make mapping extra stringent. Leave other items at their default settings and click **Next**.

4. In the Output screen, type in the **Project name** "*CDS-RB features*" and save the project in the same location as the other output file.

5. Click **OK** to initiate feature mapping.

Proceed to Part E: Compare output files.

Part E: Compare output files

1. When the mapping initiated in <u>Part D</u> is finished, click the hyperlink to the output directory.
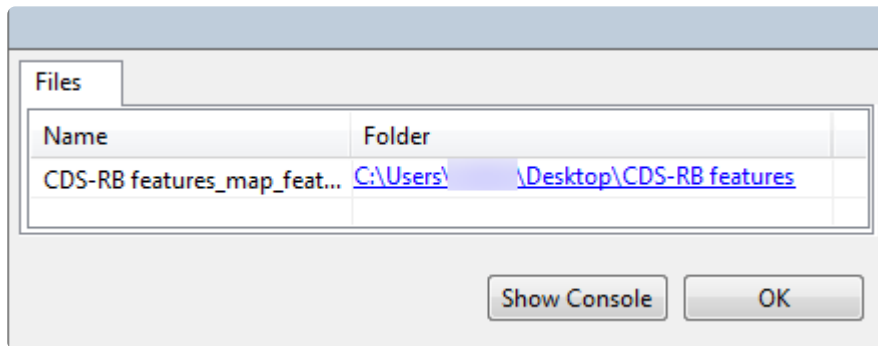
2. Open both *All features_map_features_1_report.tsv* and *CDS-RB features_map_features_1_report.tsv* in any suitable spreadsheet editor.

3. Observe that the output file *All features* has many more data rows than the file *CDS-RB features*. That's because in the latter case, non-CDS features and features without a note containing the letters "RB" were omitted from feature mapping.

4. Of the remaining rows, look in the **Variation** column and see that some features were mapped and others weren't. This assignment was based on the %Coverage and %Identity thresholds.

5. In each spreadsheet, find rows in which the **Coverage** and **%Identity** columns are between 80 and 90%. In which mapping project were these features mapped, and in which were they "Not_mapped"?

6. When you are finished, close both spreadsheets and return to the MegAlign Pro window.

7. Once in MegAlign Pro, click **OK** and then **Close** to close the Map Features-related dialogs.

Proceed to <u>Part F: Export the annotated target sequence</u>.

Part F: Export the annotated target sequence

1. In MegAlign Pro, hover over any target sequence annotation to read the entry. Verify that the sample entry matches your search criteria by checking the following:

   - Is the feature a CDS?

   - Does its original note contain the letters "RB"?

   Also observe that a new note has been added about the mapping you just performed, including the accession number of the source sequence and Coverage and Identity statistics.
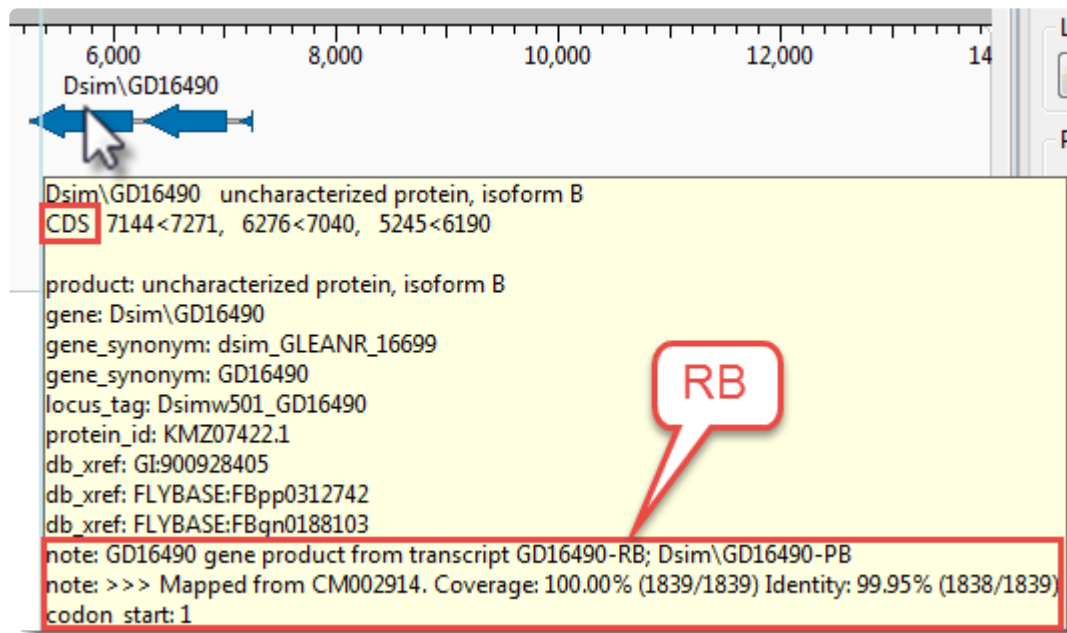
   To export the newly-annotated target sequence, select it and choose **File > Export Data > Sequences**. To preserve the annotations, select a format that supports them (e.g., *.gbk*, *.meg*).

2. Open the exported file in a text editor and observe that the notes about feature mapping have been preserved.

This is the end of the Map Features tutorial.

Search

See the following topics to learn how to:

- [Move the cursor to a specific position on a sequence or alignment](#)

- [Search for or within a sequence using a text query](#)

- [Search for nucleotide or residue strings](#)

- [Search for disagreements between sequences](#)

Move the cursor to a specific position

To go to a specified position:

Choose **Edit > Go To (Pairwise) Position**, use **Ctrl/Cmd+J**.

- If the project contains only aligned or only unaligned sequences, type in a number and click **OK**.
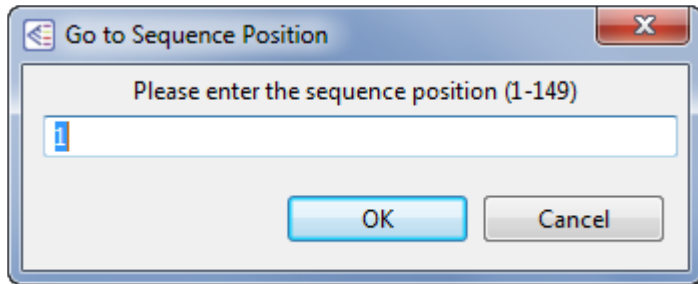
- If the project contains a mixture of aligned and unaligned sequences, type in a number and use the drop-down menu to choose whether to go to the position in the aligned or unaligned sequences. Then click **OK**.
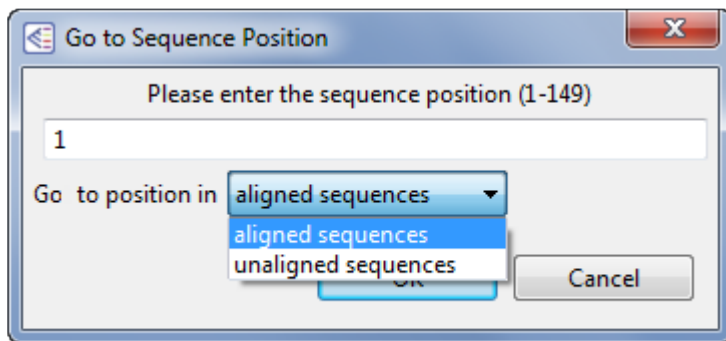
To go to a specific position in a Mauve alignment:

In the Overview, double-click on the desired alignment block to make it active. In the case of a multi-block sequence, double-click on the block again to move the cursor to the selected position. This centers the viewport on the left edge of the block and moves the visible range in the Sequences view to the same area.

Search for or within a sequence using a text query

MegAlign Pro supports searching for sequences by their name or by any text stored as part of the sequence (e.g., the description or the name of a feature).

By default, the search tools are hidden. To display them, click on the **Search** tool (🔍) in the upper right of most views. This causes the search tools to be displayed on the top left of that view.

1. From the **Find** drop-down menu:

     • If you are in any view other than the Pairwise view, choose **Sequence name**.

     • If you are in the Pairwise view, choose in which sequence to search: **In target** or **In query**.

2. Type a text query into the text box to the right of the drop-down menu. The query should be some part of the sequence name, a feature name, part of the sequence description, etc. If your query is wider than the search box, you can hover over the search box to view the entire query.
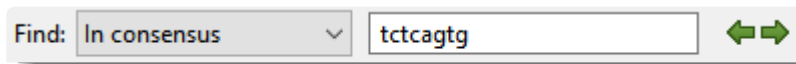
3. Use the green right/left arrows to jump to the next/previous find.

4. (optional) If you wish to limit or expand the search, use the check boxes in the search bar:

     • Check **Match case** to find only exact matches using the same case as the query term.

     • Check **Wrap around** if you would like to start searching the sequence again from the beginning or end once you have searched through the sequence using the green arrows.

     • Check **All fields** if you would like to search for text contained in the sequence file (e.g., in the description or a feature name). If you only want to search for sequences with the query in the sequence name, leave this box unchecked.

Search for nucleotide or residue strings

MegAlign Pro supports searching for strings of nucleotides or residues that match a typed query of IUPAC code letters.

By default, the search tools are hidden. To display them, click on the **Search** tool (🔍) in the upper right of most views. This causes the search tools to be displayed on the top left of that view.

1. If you plan to search within one particular sequence, select it in any view. If you plan to search within the consensus, skip this step and continue to Step 2.

2. From the **Find** drop-down menu, choose **In consensus** to search only the consensus sequence or **In selected sequence** to search within the sequence selected in Step 1.

3. Type the query into the text box to the right of the drop-down menu in the form of a string of IUPAC codes. Use nucleotide codes for DNA/RNA sequences and amino acid codes for protein sequences.

   If the box is disabled, this signifies that you have chosen **In selected sequence** without yet making a sequence selection. To enable the text box, select a sequence.

4. Use the green right/left arrows to jump to the next/previous find.

Search for disagreements between sequences

MegAlign Pro supports searching for positions where the same letter code does not occur in all sequences. This type of search looks only at IUPAC letter codes and ignores gaps. In other words, it will not look for positions in which one sequence has a letter code and the others have gaps, or vice versa.

By default, the search tools are hidden. To display them, click on the **Search** tool (🔍) in the upper right of most views. This causes the search tools to be displayed on the top left of that view.

1. From the **Find** drop-down menu, choose **Difference**.

   Find: [Difference ∨] ⬅➡
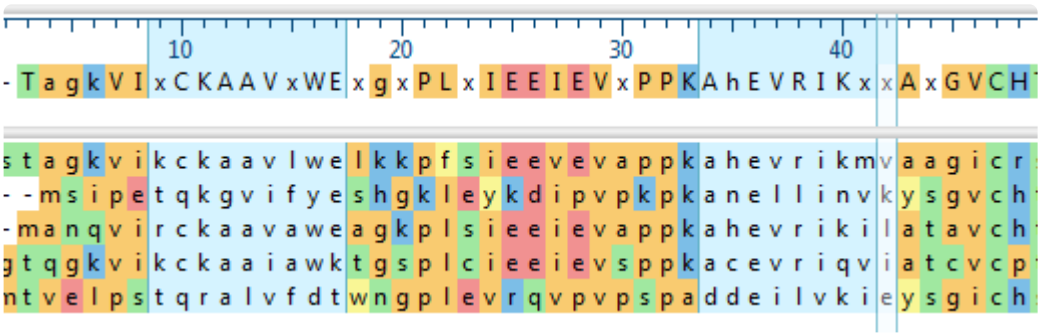
2. Use the green right/left arrows to jump to the next/previous find.

Make a Selection

Depending on the view, you can select one or more sequences, any portion of a sequence, or a specific portion of sequence corresponding to a feature. The active selection is highlighted in light blue. The selection can then be copied to the clipboard, copied in a specific format, exported to a file, exported to MegAlign, or realigned using different parameters and/or alignment method. This functionality is useful for sending a selected sequence or region of sequence to a new alignment project as an unaligned sequence. In this way, you can create a new document using a subset of the sequences in the current document, thereby avoiding having to find all of the input files. You can also use the selected region of one or more sequences to create a new document.

To select…	Do this…
All sequences	Click on the name of any sequence, then use **Ctrl/Cmd+A** or **Edit > Select All**. • In a Mauve alignment, this command extends the selection across the entire active block. • If a portion of sequence or alignment was originally selected, the entire sequence or alignment will be selected. • In the Distance view, this command selects the entire distance matrix. • If a subsequence was selected in one or all aligned sequences prior to using this command, this action will extend the subsequence selection to the full sequences.
Multiple sequences	Use **Ctrl/Cmd+click** or **Shift+click** to select multiple sequences.
One entire sequence	Either: • Click on the sequence name. • Drag the mouse across all residues in a sequence track while holding down the left mouse button. • Double-click on a feature that spans the entire sequence.
A portion of all aligned sequences	Drag the mouse across residues in the consensus sequence while holding down the left mouse button. To toggle selection of items on or off, hold down the **Ctrl/Cmd** key while dragging the mouse, or while clicking on the items to toggle on or off.

A portion of one sequence	Drag the mouse across residues in a sequence track while holding down the left mouse button.
The portion of a single sequence associated with a section of track (e.g. a Gap Fraction peak, etc.)	Drag the mouse across the open track while holding down the left mouse button.
One or more tracks	Click on a track to select it. To add additional tracks to the selection, use **Ctrl/Cmd+click** or **Shift+click**.
One or more features and corresponding sequence	Click on a feature to select it. To add additional features to the selection, use **Ctrl/Cmd+click**.
The portion of a single sequence associated with a selected feature	To expand a selection consisting of one or more features to a selection containing just the sequence corresponding to the feature(s), do any of the following: • Choose **Edit > Select Subsequence**. • Click on the feature to select it, then right-click and choose **Select Subsequence**. • Double-click on the feature under that sequence. • In the Details panel, click on the Action **Select Subsequence**.
The portion of all aligned sequences associated with a feature	To expand a selection consisting of one or more features or subsequence within a single sequence, to the corresponding subsequences across all aligned sequences or subsequences, either: • Choose **Edit > Select Subsequences Across Alignment**. • Triple-click on the feature under any of the sequences in the alignment.
Additional continuous items, when the first	Use **Ctrl/Cmd+Shift** and click on the other endpoint.

endpoint has been selected	
The current selection, but expanded to encompass additional bases	You can extend the length of an existing selection by **Shift+clicking** on the new endpoint. The selection will expand to use the chosen base as the new endpoint. If there are multiple selected ranges, the one closed to the chosen base will be expanded.
One or more items in the Tree view	In the Tree view, to select one sequence, click on it. To select multiple sequences, use **Ctrl/Cmd+click**. To select all sequences in a clade, click the horizontal "stem" for that branch.
All variants in the Variants view	Choose **Variants > Select > Select All**. Remove the selection using **Variants > Select > Select None**.
Individual variants in the Variants view	Drag the mouse or by use **Shift+click** or **Cmd/Ctrl+click**. To select all rows *except for* the currently-selected rows, choose **Variants > Select > Invert Selection**.

Search for Sequences Online

This application lets you search National Center for Biotechnology Information (NCBI) databases for matches to text or sequence that you specify.

- Query part or all of a nucleotide or protein sequence against one of NCBI's sequence databases using the Basic Alignment Search Tool (BLAST).

- Query text or accession numbers against entries NCBI's Entrez database.

Once a search has been initiated, you can use the Jobs panel to monitor the search, stop an in-progress search, or open search results in the Text and Table views.

Tools available in the Search wizards:

Some or all of the following optional tools may be available in the bottom left corner of the Search wizards:
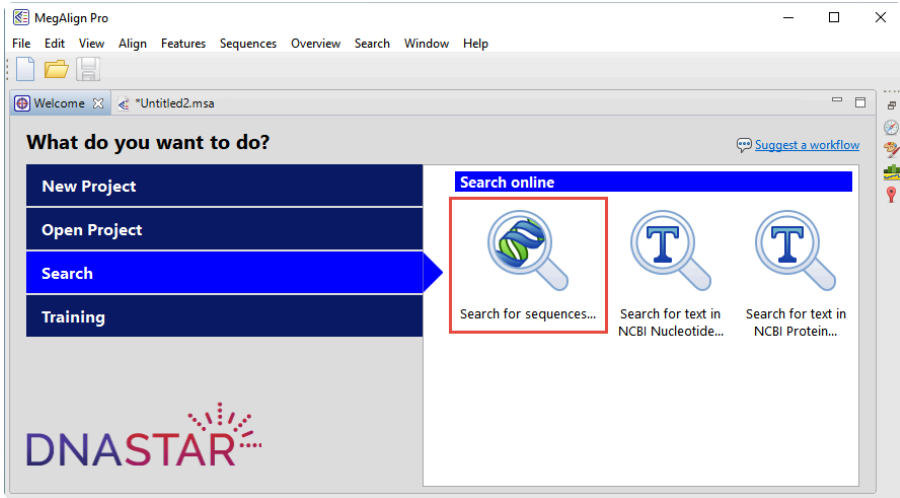
1. To display the help for this topic, press the question mark icon ( ).

2. To enter an NCBI Access ID, press the **Set Access ID** ( ) tool. This causes the Access ID dialog to pop up. See Access ID dialog for information.

3. To submit an idea for a new feature, press the **Suggest a feature** ( ) tool.

4. To view information about your service plan, press the **Show Service Plan** ( ) tool. For sequence or text search services, the ensuing dialog describes the state of your service plan and any warning or error text messages. The license control has three states: **Active**, **Expiring**, and **Expired**.

Search for sequence matches (BLAST)

The BLAST Search wizard is used to search a query sequence against one of NCBI's databases. The sequence can be part or all of an active sequence, another sequence file on your computer, or typed/pasted-in text. The wizard consists of four consecutive screens, each described in a separate topic:

- Query screen
- Search screen
- Options screen
- Job screen

To launch the BLAST Search wizard, use one of the following commands or tools. Some commands let you launch the wizard with the workflow pre-selected and the Search list already pre-populated.

Command, shortcut or tool	Description
Search > Search for Sequences	
Use the "Change search strategy" tool (⚙) from the Text, Table or Pairwise views	
From the Welcome screen, choose **Search** and then **Search for sequences**. 	Launches the wizard at the Query screen.
Search > Sequence Search with Selection Press **Ctrl/Cmd+B**	Launches the wizard at the Query screen, whose table is pre-populated with the currently-selected portion of sequence.
Search > Search for All Sequences	Launches the wizard at the Query screen, whose table is pre-populated with all sequences in the document.

Search wizard: Query

The Query screen is the first of four screens in the [BLAST Search wizard](#). The Query screen is where you add the query sequence, i.e., the sequence for which you wish to locate matches in the NCBI BLAST database.

At the top of the screen, next to "Search with," select the **Nucleotides** or **Proteins** button.

To add one or more query sequence files or a folder of sequences:

Press the **Add** or **Add Folder** buttons. In the file explorer, navigate to and select the desired file(s)/folder(s), and then click **Open**. If you add a multi-file sequence or multiple sequences, each sequence will appear on a separate row in the table.

To paste or type in the query sequence:

Press **Enter**.

Enter

Enter sequences in FASTA format

>Hemoglobin (Alpha Chain)
VLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTK
TYFPHFDLSHGSAQVKGHGKKVADALTNAVAHVDDMPNAL
SALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPA
VHASLDKFLASVSTVLTSKYR

Range

Start: _____   End: _____

☐ Add another

OK   Cancel

In the Enter dialog:

1. Type or paste in the desired sequence following the "dimmed" example in the dialog. A header line is optional.

2. (optional) If you wish to use only a range of this sequence, specify the **Start** and **End** positions.

3. (optional) If you want to add another query sequence after this one, check the **Add another** box. After clicking **OK** to the first sequence, a new Enter dialog will open to add the next query. Otherwise, leave this box unchecked.

4. Press **OK**. If you add additional sequences, each sequence will appear on a separate row in the table.

To remove a file that has already been added:

Select the unwanted file from the list and click **Remove**.

To learn about the optional tools in the bottom left corner of the screen, see the bottom of this topic.

Once you have set up the query sequence(s) as desired, do either of the following:

• To choose the BLAST program to use and database to search, press **Next** to continue to the Search screen.

• To run the search directly, press **Run**.

- To close the Search wizard without starting a search, press **Cancel**.

Search wizard: Search

The Search screen is the second of four screens in the BLAST Search wizard. The Search screen is used to choose the BLAST program and database.

- Specify whether you wish to search for **Nucleotides** or **Proteins**.

- Select the desired BLAST program from the **Using** drop-down menu. For information on the different BLAST programs, see the BLAST Program Selection Guide PDF. Available options depend on the combination of the selections you made in **Search with** (Query screen) and **Search for** (this screen).

  ◦ If you selected **Proteins** on both screens, the **Using** menu contains the unchangeable default of **Similar proteins (blastp)**.

  ◦ If you selected **Nucleotides** on both screens, Choices are **Highly similar sequences (megablast)**, **Similar nucleotides (blastn)**, and **Similar translated nucleotides (tblastx)**.

  ◦ If you selected **Nucleotides** in the Query screen and Proteins in the Search screen, the **Using** menu contains the unchangeable default of **Similar proteins (blastx)**.

  ◦ If you selected **Proteins** in the Query screen and **Nucleotides** in the Search screen, the **Using** menu contains the unchangeable default of **Similar translated nucleotides (tblastn).**

- Select the desired BLAST database from the **In** drop-down menu. Available choices depend on whether you selected **Nucleotides** or **Proteins** in this screen.

To learn about the optional tools in the bottom left corner of the screen, see the bottom of this topic.

Once you have made the desired selections, do any of the following:

- To go back to the Query screen, press **Back**.

- To add additional parameters to the search, press **Next** to continue to the Options screen.

- To start the search immediately, press the **Run Now** button ( ) or **Run** button. The Jobs panel in the application will open and contain information on the status of the search. During the search, a progress dialog is displayed where you can cancel the search using the **Stop** button, hide the dialog using the **Hide** button, or set up another query using the **Settings…** button. Use the Jobs panel to monitor an online search, stop an in-progress search, or open search results in the Text, Table and Pairwise views.

- To close the Search wizard without starting a search, press **Cancel**.

Search wizard: Options

The Options screen is the third of four screens in the BLAST Search wizard, and has two tabs: **Selection** and **Matches**. These tabs allow you to customize miscellaneous parameters relating to the BLAST search.

- Click the **Selection** tab to make it active.

In the **Selection** tab:

- To filter the database you're searching using a text term, enter a text string in the **Entrez** query field.

- To filter low-complexity regions in the query, check the box next to **Ignore low complexity in query sequence**. The **blastn** search program will use **DUST** for filtering; all other programs will use **SEG**.

- To mask the query while producing seeds used to scan database (but not for extensions), check the box by **Ignore low complexity during database lookup**.

- To mask any letters that were lower-case in the FASTA input, check the box next to **Ignore lower case letters**.

- Click the **Matches** tab to make it active.

In the **Matches** tab:

- ◦ Enter a number next to **Maximum sequences**, the maximum number of aligned sequences to
  display. The actual number of alignments may be greater than this number. The default is **100**.

- ◦ Enter a number beside **Expect threshold**, the expected number of chance matches in a
  random model. The default is **10**.

- ◦ Enter a number next to **Maximum regional matches** to limit the number of matches to a query
  range. The default is **0**.

To learn about the optional tools in the bottom left corner of the screen, see the bottom of this topic.

Once you have made the desired selections, do any of the following:

- • To continue filtering, press **Next** to continue to the Job screen.

- • To start the search immediately, press the **Run Now** button ( ) or **Run** button. The Jobs
  panel in the application will open and contain information on the status of the search. During the
  search, a progress dialog is displayed where you can cancel the search using the **Stop** button, hide
  the dialog using the **Hide** button, or set up another query using the **Settings…** button. Use the Jobs
  panel to monitor an online search, stop an in-progress search, or open search results in the Text,
  Table and Pairwise views.

- To close the Search wizard without starting a search, press **Cancel**.

Search wizard: Job

The Job screen is the fourth of four screens in the BLAST Search wizard, This screen allows you to overwrite the default search name with a custom name and initiate the BLAST search.

The text box contains an automatically-generated name for the search job. If you wish, you can overwrite this default name with a custom name.

To learn about the optional tools in the bottom left corner of the screen, see the bottom of this topic.

After entering a custom job name (or keeping the default name), do either of the following:

- To start the search , press the **Run** button. The Jobs panel in the application will open and contain information on the status of the search. During the search, a progress dialog is displayed where you can cancel the search using the **Stop** button, hide the dialog using the **Hide** button, or set up another query using the **Settings…** button. Use the Jobs panel to monitor an online search, stop an in-progress search, or open search results in the Text, Table and Pairwise views.

- To close the Search wizard without starting a search, press **Cancel**.

Search for text or accession number matches (Entrez)

The Entrez Search wizard is used to <u>search</u> a text query or accession number against NCBI's Entrez databases for nucleotides or proteins.

To launch the Entrez search dialog, choose **Search > Text Search in > NCBI Nucleotide** or **Search > Text Search in > NCBI Protein**. Alternatively, open the Welcome tab, choose **Search** and then click either **Search for text in…** link.
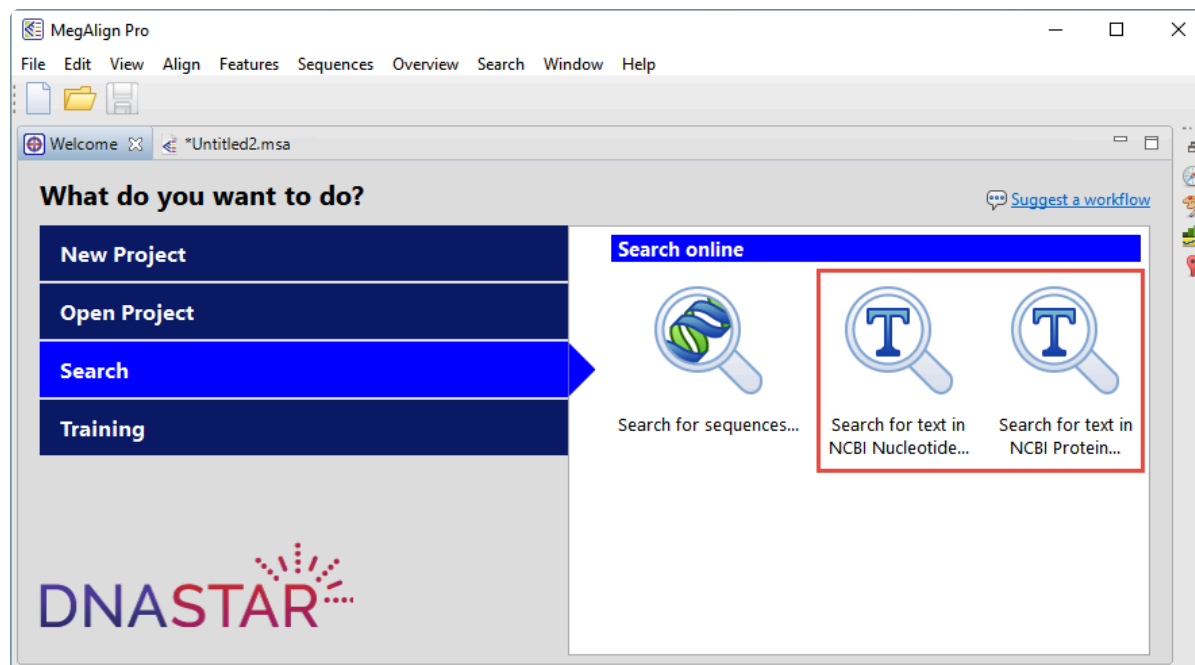
The **Visual** view is displayed by default and is described below.

> ✳ **Note:** The Visual view is only for non-nested queries using the same type of operator (AND, OR). To perform a more complicated search, press the **Text** tab near the bottom left corner and create the search in the <u>Text view</u>.

1. Use the **<select>** drop-down menu to choose a criterion. The field headings vary depending upon the online database that you are using. Choosing **All fields** will cause the application to search every word of every listing for the query term in the database.
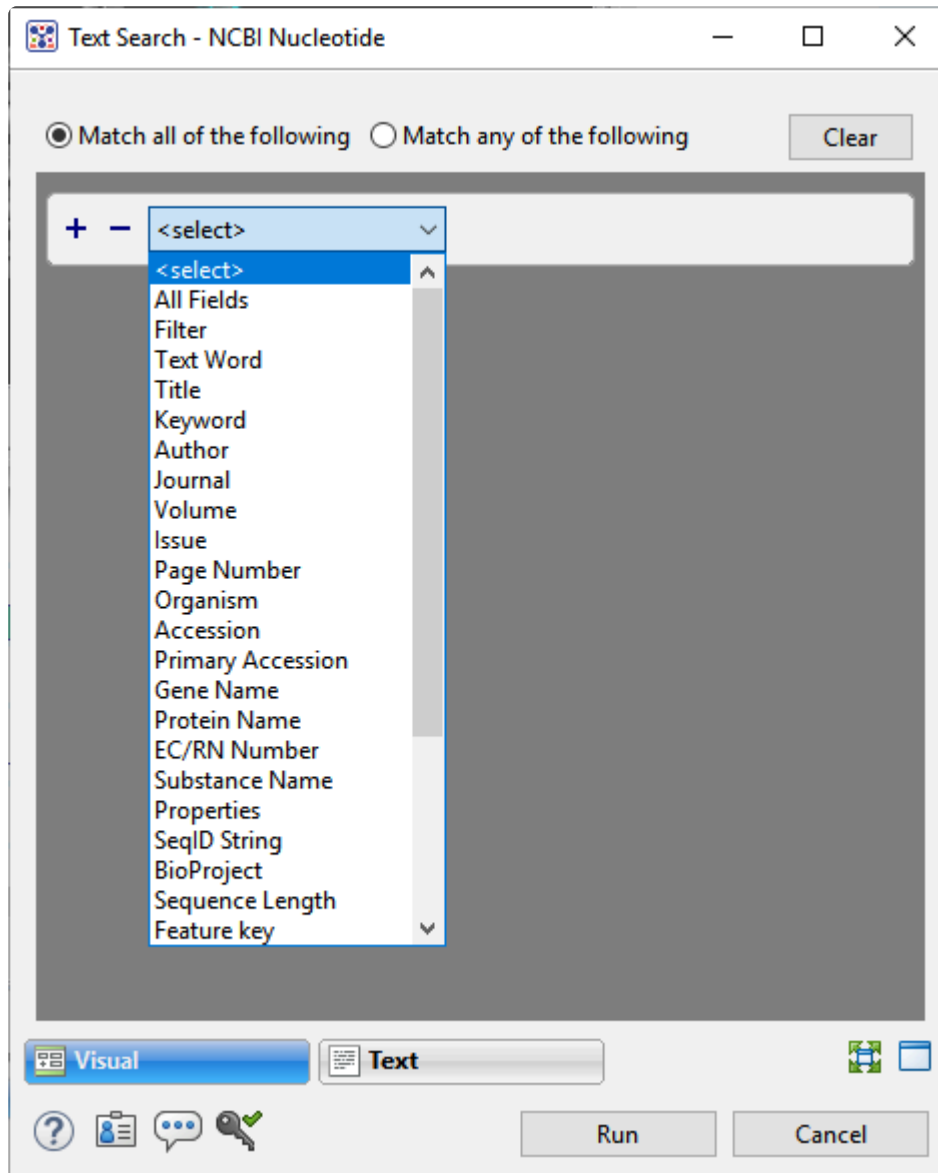
Once you make a selection, a second drop-down menu and an empty text box appear to the right.

- In the second drop-down menu on the row, choose **contains** or **doesn't contain** (see image below).

- In the text box, type the text that <u>should</u> be found in any results (in the case of **contains**) or <u>should not</u> be found in any results (in the case of **doesn't contain**).

2. To add another criterion, press the plus icon (**+**). To delete a criterion, press the minus icon (**–**).
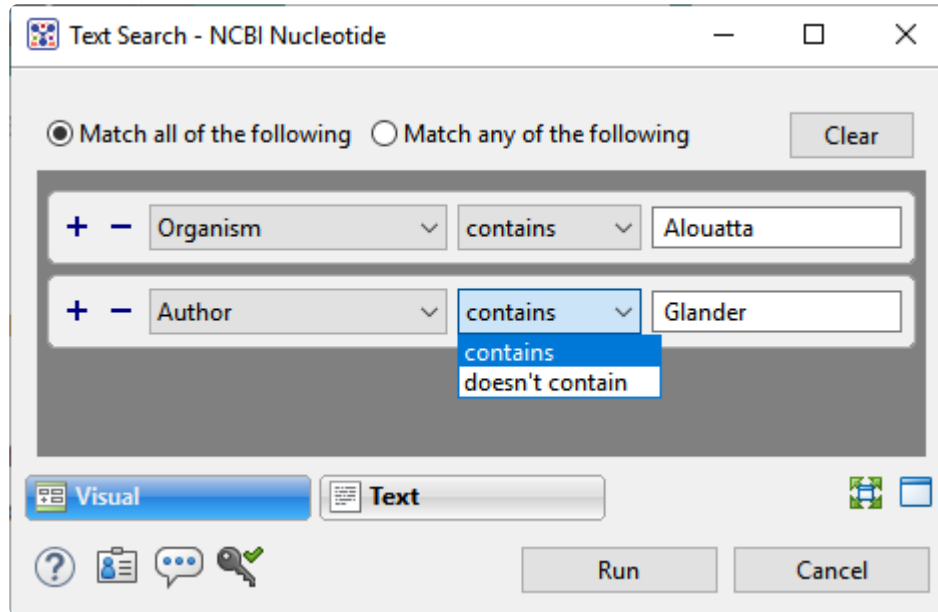
3. Once you have set up all desired criteria, go to the top of the dialog and choose whether to **Match all of the following** or **Match any of the following** criteria.

4. To learn about the optional tools in the bottom left corner of the screen, see the bottom of this topic.

5. Click **Run** to send the query to the database. To instead exit the dialog without searching, press **Cancel**.

During the search, a progress dialog is displayed. Once the search is completed, any matches are displayed in the Jobs panel.

Filter search results

After doing a BLAST or Entrez search, the Table view shows a list of potential matches. To filter this table to include only matches that meet additional criteria, click the **Filter results** tool (⧩) in the header and then click **Filter**. This opens the Filter dialog. The dialog offers different options, depending whether you did a BLAST or Entrez search.
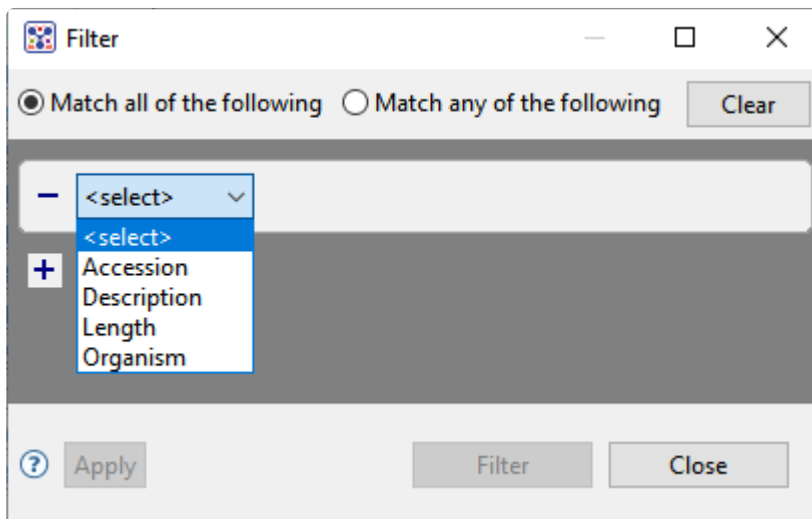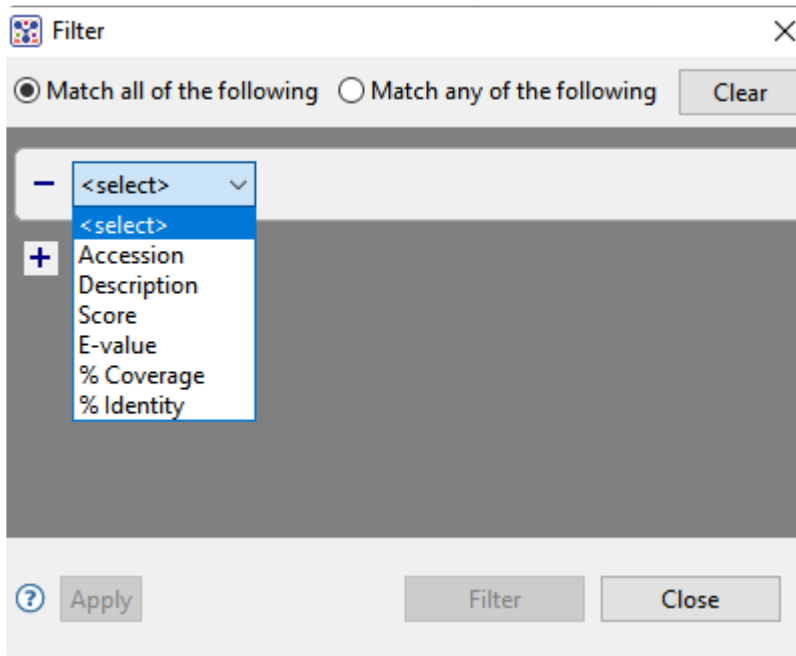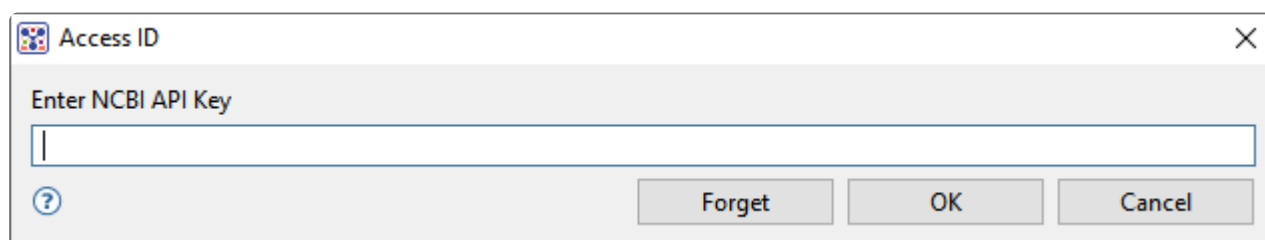
See Add only sequences that meet filtering thresholds for details about how to set up filters in this dialog.

Access ID dialog

As of December 2018, NCBI requires an API key for faster access to NCBI services, or to run more searches concurrently. To get a key, start by setting up an NCBI account. After registering your account, click on your user name and request your key using **API Key Management**. Once you have your key, you can enter it in Lasergene applications to enable NCBI's premium search access.

To add your key to a Lasergene application:

1. Begin an Entrez database search using **Search > Text Search in > NCBI Nucleotide** or **Search > Text Search in > NCBI Protein**.

2. Press the **Set Access ID** ( ) tool in the bottom left corner of the dialog; or, from the Table view, press the **Export data** tool and choose **Set Access ID**. In either case, the Access ID pop up appears.

- If an NCBI API key is not already present in the text box, enter your key and press **OK**.

- If an NCBI API key is present but you no longer use it, press **Forget** to remove it from the application's memory. Then enter the desired key and press **OK**.

- To exit without saving your changes, press **Cancel**.

Choose the Alignment Type

Which is better, multiple or pairwise alignment?

This question is difficult to answer because it very much depends on how the alignment is going to be used.

Mechanistically, the best sequence alignment is the one that produces the fewest number of mismatches. That metric can be misleading, especially if minimizing the score entails extreme amounts of gapping. Consider an example where the goal is to identify some particular conserved domains or a large insertion. Here the placement of gaps outside of the regions of interest may well be of limited concern. Note that with MegAlign Pro, you can select the interesting regions identified by a multiple alignment and copy them as subsequences to a new document for further analysis.

Now consider a situation where a multiple sequence alignment is used to represent the actual relatedness of a group of sequences. Here the alignment is essentially a model, typically of an evolutionary process. In this case the "best" alignment is the one that is most plausible in the light some biological theory or model. One way that this visualized, of course, is to use the alignment to make an evolutionary tree.
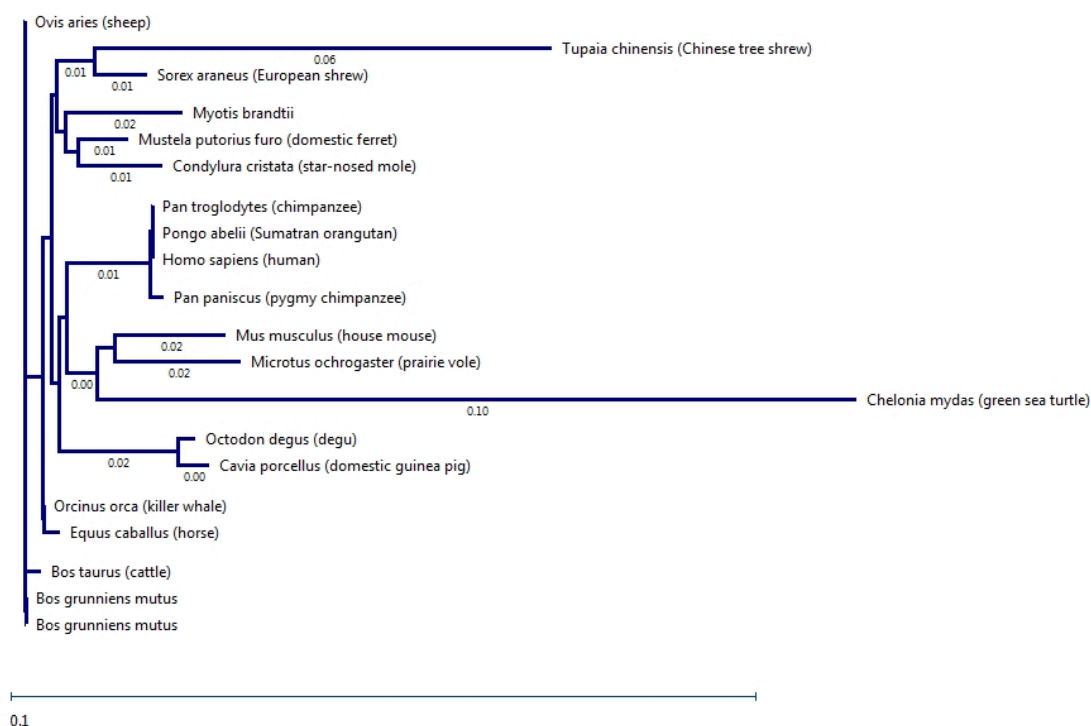
When should I use a pairwise alignment?

The answer to this question may seem obvious: use pairwise alignment when you are only interested in two sequences. Also, sometimes pairwise alignment is simply more suitable than multiple alignment. Additionally, there are situations where a multiple sequence alignment (MSA) might help identify pairs for

sequences or sub-sequences that are worth a more detailed, pairwise comparison.

Beyond workflow considerations, there are some fundamental differences between the two categories of alignment that might make pairwise alignment a better option for some sequence comparisons. Due to the nature of progressive multiple aligners (including Clustal, MUSCLE and MAFFT), the final sequence alignment can contain inappropriately placed gaps, which adversely affect the interpretation of the results. To understand this, consider how progressive multiple alignments work. The process invariably begins with a single pairwise alignment, adding gaps as necessary in order to minimize the number of mismatches. As the aligner proceeds, additional gaps are added as single sequences and groups of sequences aligned during an earlier stage of the process are included in the growing multiple sequence alignment. During this phase, gaps may be added but are never removed.

This "once a gap, always a gap" approach is a potential drawback that is shared by all progressive multiple alignment algorithms. The heart of the problem is that gap placement (and therefore the alignment) might be affected by the order in which sequences are aligned to each other because sequences added later in the process might be incorrectly aligned. All of the multiple alignment engines used by MegAlign Pro use a "guide tree" based on pairwise similarities of sequences to determine the order in which to align sequences. The first pair chosen consists of the two that are least distant on the guide tree. If the nearest neighbor to this pair is more distant than some other pair are to each other, that pair gets aligned to each other. If not, the neighbor is aligned with the first pair and gaps are added as necessary. In later rounds there may be no singleton sequences left, just clusters of two or more sequences that got aligned. Imagine a case where one of a group of early aligned sequences should have been added later, or where a close relative was added too late. It's hard to know when this has happened unless you have some *a priori* information, such as knowledge of the evolutionary relationship of your group of sequences.

The bottom line is that when you examine just a pair from a multiple sequence alignment you may not see the same results as a pairwise alignment of just the two. So the direct approach under these circumstances might give a better picture of the relatedness of the pair.

Perform a Multiple Alignment

MegAlign Pro supports the following multiple alignment methods: Clustal Omega, Clustal W, MAFFT, MUSCLE and Mauve. Click here for descriptions and a comparison of the different methods.

To learn how to perform, modify, and troubleshoot multiple alignments, see the following topics:

- Perform an initial multiple alignment
- Multiple alignment methods and options
- Modify a multiple alignment
- Unalign aligned sequences

Multiple alignment tutorials:

The following tutorials all use free data that can be downloaded from the DNASTAR website.

- Try it! – Perform a Clustal Omega alignment
- Try it! – Perform a MUSCLE alignment with multi-segment sequences
- Try it! – Perform a genomic alignment with Mauve

The following video is a quick introduction to performing multiple alignments in MegAlign Pro:

Perform an initial multiple alignment

To perform the initial multiple sequence alignment:

1. Select three or more sequences to align, or select nothing if you want to align all sequences. (Note: If you choose just two sequences, or if only two are present, you will be prompted to instead do a pairwise alignment.) All sequences must belong to the same category (DNA/RNA or protein). Selected sequences can include any combination of already-aligned sequences and sequences in the "Unaligned Sequences" area. Both types are visible in the Overview (see image) and Sequences view.
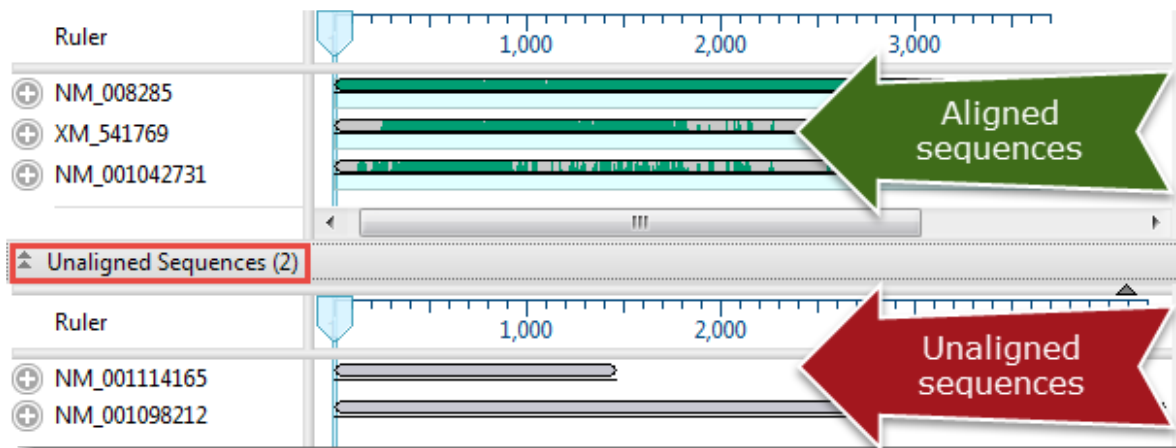
2. Hover the mouse over the **Align/Realign** tool ( ) to view a tool tip showing the currently-selected algorithm.
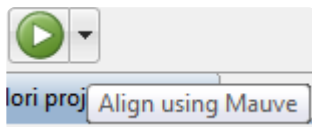
3. Choose the desired multiple alignment method and perform the alignment. MAFFT [auto] is the default method for nucleotide alignments, MAFFT [l-ins-i] is the default for protein sequence alignments, and MAUVE is the default if the sequences are >1Mb in size.

   • If the method and options are compatible with the current sequence type, and you want to use it, press the **Align/Realign** tool. In general, we recommend using Mauve for genome scale alignments and the other alignment methods for gene-level alignments. Since Mauve works by initially breaking alignment projects into a series of smaller sub-alignments, we also suggest using Mauve if there is insufficient memory to align sequences with the other methods, or if another type of alignment is taking an excessive amount of time to complete.

   • To choose a different method, click on the small black arrow to the right of the **Align** tool to choose from available methods. Or use the main menu to choose **Align > Align Using (Method**

Name). If the sequences have already been aligned, note that the menu command names change to **Align > Realign Using (Method Name)**.

- To choose a different method and/or change method options before aligning sequences, click on the small black arrow to the right of the **Align** tool and choose **Align with Options**; or select **Align > Align with Options** from the main menu. For a detailed description of the Alignment Options dialog for different methods, see Clustal Omega, Clustal W, MAFFT, MUSCLE or Mauve alignment options.

Except in the case of divergent sequences (see note, below), we recommend using the DNASTAR default settings the first time you perform a multiple sequence alignment. Later, you may wish modify some of the Alignment Options before realigning the sequences. After you customize options in the Alignment Options dialog and perform an alignment using the new settings, they become the new defaults, but only for this project, and only as long as the project remains open. For subsequent alignments in the same project, you can use the **Align** tool (▶) to perform the alignment using your custom settings.

4. (optional) During the alignment, MegAlign Pro displays a progress window. If you need to interrupt the alignment, click **Cancel**. To view a console window showing the start time and progress of the alignment, click **Show Console**.

- ◦ If an alignment fails, you will receive a message with recommendations on how to obtain a successful alignment (e.g., by modifying options or choosing a different alignment method). See Troubleshooting 'wrong type' or 'too long' errors or Troubleshooting 'fatal error' in Clustal Omega alignments.

  ◦ If an alignment finishes successfully, a table of pairwise distances and a phylogenetic tree are created and displayed in the Distance view and Tree view, respectively. If the Console is not already open, and you wish to view alignment statistics and other information there, select **View > Console**. To reset the sequences to the unaligned state (i.e., as if they were just added to MegAlign Pro), use **Align > Unalign All**.

Notes regarding multiple alignments

When performing a multiple alignment, take the following tips into consideration.

Sequence order matters:

The order in which sequences appear in the Overview and Sequences view may affect the results of the multiple sequence alignment. If you are not satisfied with an alignment, try reordering sequences and running the alignment again.

Divergent sequences

The MUSCLE multiple alignment algorithm assumes that the sequences to be aligned have a certain degree of relatedness. Groups of relatively divergent sequences, especially those with very large data sets, may require considerably more computer resources, particularly RAM. When working with these types of data, we recommend using **Align > Align with Options** and changing Maximum iterations to 1 or 2. This will both reduce the amount of memory needed, and ensure that the sequences are sufficiently related to allow for the alignment process to be successful.

Additional alignment methods:

Two additional multiple alignment methods are available in DNASTAR's MegAlign application: Jotun Hein and Clustal V. Please open the classic MegAlign application and select **Help > Contents** for more information on these methods.

High-throughput sequencing reads:

We caution against attempting a multiple alignment in MegAlign Pro using high-throughput sequencing reads. Analysis of these reads should instead be handled through DNASTAR's SeqMan NGen and ArrayStar applications.

Known issues with Mauve:

- An issue in Mauve can cause an alignment to stall if sequences are ordered one way, but not another. If this issue occurs, you may see up to three error messages after beginning a Mauve alignment: "progressiveMauve-[OS].exe has stopped working," "Do you want to send more information about the problem?," and "Unexpected error while running the alignment." The last message provides a **Details** button which reveals the text: "Mauve quit unexpectedly. Reported error: Unrecognized file format." In addition, a progress bar will open, but the alignment will not finish. To circumvent this issue, **Cancel**

the alignment, drag the sequences into a different order in the Sequences View or Overview, and then try the alignment again.

- Sequences aligned with Mauve may give slightly different results on Windows and Macintosh. These differences originate with the Mauve algorithm and not with the implementation in MegAlign Pro.

Troubleshooting "wrong type" or "too long" errors

Error message that sequences are the wrong type or are too long:

After beginning an alignment, an error message will appear if the sequences are not of the appropriate type or are too long for the selected algorithm. If performing a MUSCLE or Clustal Omega alignment, an error message usually indicates that the sequences are too long or that multi-segment sequence groups are ordered incorrectly. To remedy the situation, you can align using Mauve, or correctly re-enter any incorrect multi-segment sequences, or break up the sequences into shorter lengths. This last task can be performed very quickly and easily using DNASTAR SeqNinja.
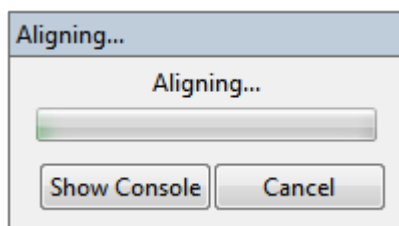
Message that an unexpected error occured:

During a multiple sequence alignment, you may receive the following warning:
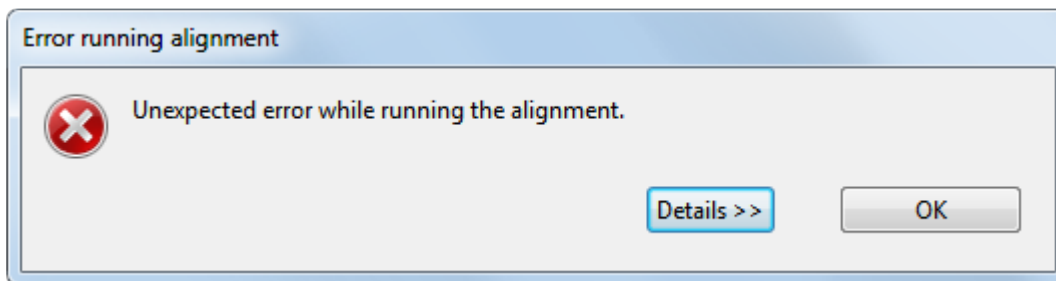
To reveal more information, click on the **Details** button.
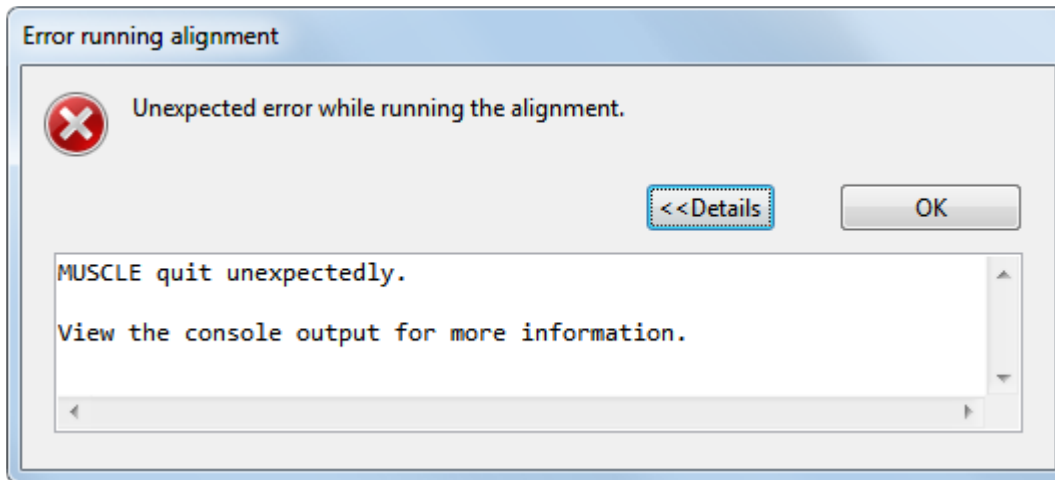
For additional details, select **View > Console** from the menu. An example console is shown below:
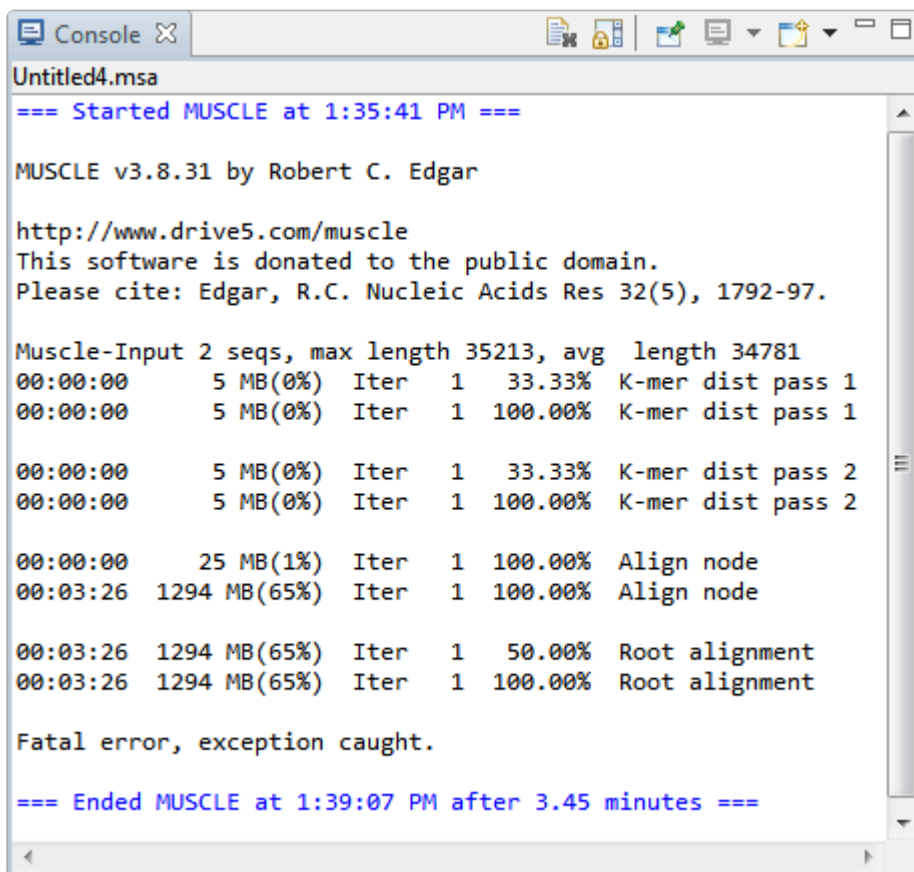
Notice that in this console, the message "Fatal error, exception caught" appears near the bottom. Either that message or an "Out of memory" message are known to occur with particularly long, highly divergent sequences.

> ✳ **Note:** If you have a question about a console error message (or anything else related to the MegAlign Pro software), please contact DNASTAR Support.

Before attempting the alignment again, we recommend using **Align > Align with Options** and checking the box next to **Use Brenner's Alignment**. Checking the box causes MegAlign Pro to use Brenner's method for the root alignment. This method uses less memory, and enables the alignment of long and divergent sequences, but with a concurrent loss in accuracy.

Troubleshooting "fatal error" in Clustal Omega alignments

Clustal Omega requires that each sequence in an alignment contain at least one residue that is not an ambiguity code.

- DNA sequences must contain at least one A, C, T or G.
- RNA sequences must contain at least one A, C, U, or G.
- Protein sequences must contain at least one A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W or Y.

If there are any sequences that do not meet this requirement, Clustal Omega will terminate with a "fatal error" message (visible in the MegAlign Pro Console), and no alignment will be created.

To address this issue, you can do either of the following:

- Remove the sequence and try aligning again with Clustal Omega.

- Include the sequence, but switch to MAFFT or MUSCLE as the alignment method. Neither of these algorithms requires sequences to contain at least one non-ambiguous residue.

Multiple alignment methods and options

MegAlign Pro offers both gene-level and genome-level multiple alignment algorithms.

<u>Gene-level alignment of either protein or nucleotide sequences:</u>

In general, the three gene-level aligners have higher accuracy than the genome-level (Mauve) aligner. They offer editable options for speed, capacity, algorithm, etc., and are the only methods available for "profile" alignments. The disadvantages to gene-level aligners is that sequences must be on same strand, and that large rearrangements (e.g., inversions, translocations) are not allowed.

- **Clustal Omega** – [Clustal Omega](#) ([Sievers F et al.](#), 2011) was developed at University College Dublin, and is the most advanced version of Clustal. It can align hundreds of thousands of sequences in just a few hours. This method has few editable options, but has high, 64-bit capacity on both Windows and Macintosh.

- **Clustal W** – Clustal W aligns sequences using the method of [Thompson et al.](#) (1994). Clustal W was designed to create more accurate alignments than Clustal V when alignments include highly diverged sequences. However, Clustal W does not always handle end-gaps ideally. Also, note that true Clustal W performance is only achieved when you choose the default "Slow-Accurate" option rather than the "Fast-Approximate" option.

- **MAFFT** – MAFFT (Multiple Alignment Fast Fourier Transform; [Katoh M & Kumar M](#), 2002) was developed by the [Computational Biology Research Center](#) and generously donated to the public domain. This method has many editable options, and provides a variety of algorithms for different scenarios and a choice of very slow to very fast speeds. It has 32-bit capacity on Windows, and 64-bit on Macintosh. See the CBRC's [MAFFT page](#) for additional references. As of the Lasergene 17.3 release (August 2021), MegAlign Pro uses MAFFT version 7. This powerful aligner can typically align thousands of viral genomes, for example, in under two minutes. The ability to align to a specified reference sequence was added in the Lasergene 17.3.1 release (January 2022). The ability to align to a reference is ideal for the small percentage of MegAlign Pro users that need an extremely high capacity alignment algorithm.

- **MUSCLE** – The [MUSCLE alignment algorithm](#) was developed by Dr. Robert Edgar ([Edgar RC](#), 2004 & 2004), who very kindly donated it to the public domain. It is one of the faster aligners, and has numerous selectable options. MUSCLE features iterative cycles that guide tree refinement and realignment. Like MAFFT, it has 32-bit capacity on Windows, and 64-bit on Macintosh.

<u>Genome-level alignment of nucleotide sequences:</u>

In comparison to the methods above, the genome-level aligner allows large rearrangements. However, it only allows nucleotide sequences, and fine-scale gapping may not be as good as in gene-level aligners.

- **Mauve** was developed in the Genome Evolution Laboratory at the University of Wisconsin-Madison ([Darling AE, Mau B, and Perna NT](), 2010) and is licensed under a [GNU General Public License](). Mauve has high capacity and uses MUSCLE to create multiple alignments for each block that contains more than a single sequence. The Mauve algorithm is currently the only MegAlign Pro alignment method that is:

  - suitable for aligning very long sequences up to genome-length.

  - capable of producing an alignment when one or more of the sequences are rearranged relative to one another.

  - capable of producing a multi-block alignment (see [Overview]()).

  - only available for nucleotide sequences.

Clustal Omega alignment options

The Clustal Omega algorithm is for gene level alignment of either protein or nucleotide sequences. To run a Clustal W alignment, select two or more sequences and choose **Align > (Re)Align Using Clustal Omega**.

If you wish to change method options, instead choose **Align > Align with Options**. The dialog is the same whether the sequences are protein or nucleotide.
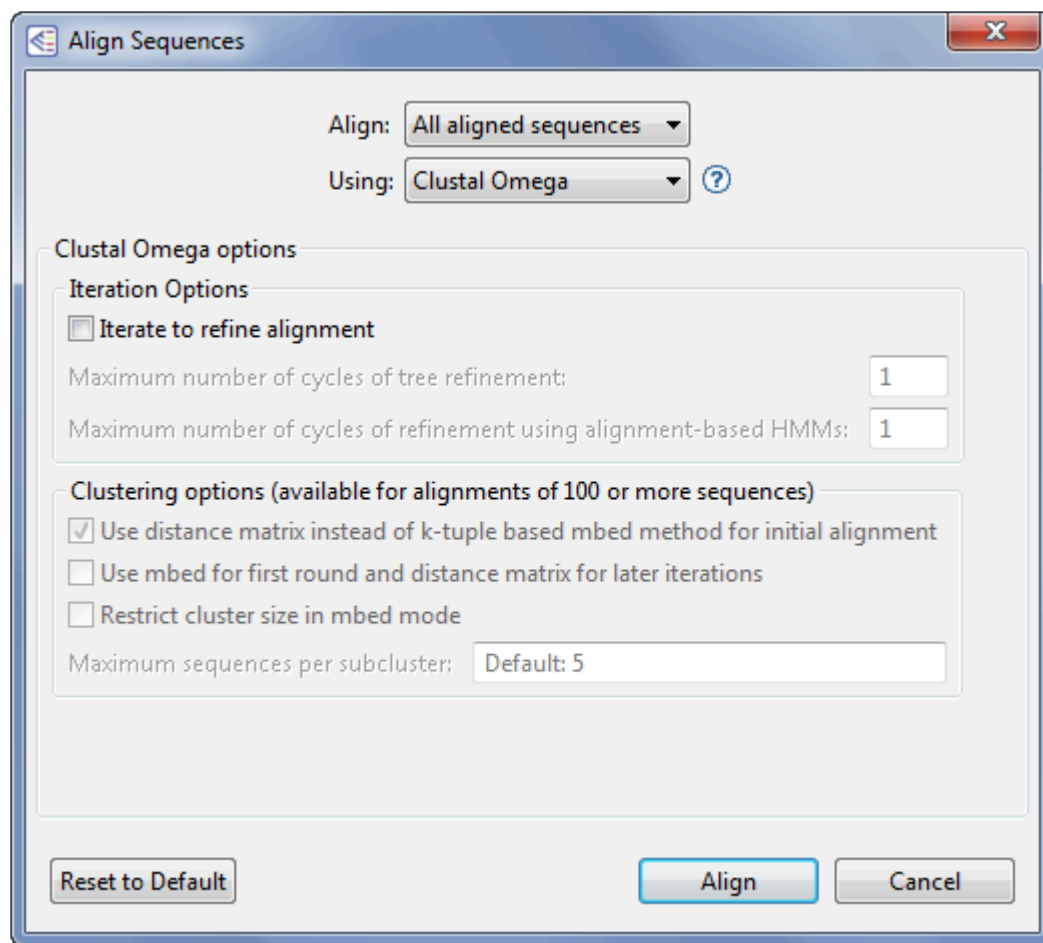
Change settings as desired:

> ✳ **Note:** Certain options in this dialog may initially be disabled. For example, some options are only enabled if there are ≥ 100 sequences in the project, while others are enabled only after options higher up in the dialog are checked. See below for further information.

- Use the **Align** drop-down menu to select sequences to align or realign.

- In the **Using** drop-down menu, choose **Clustal Omega**.

- To refine the alignment by specifying a threshold for iteration cycles, check the box next to **Iterate to refine alignment**. When this box is checked, two refinement options become available. You may enter values for neither, either, or both options.

  - ◦ Enter a value for **Maximum number of cycles of tree refinement** to specify a cutoff for the maximum number of iterations. The default is 1.

  - ◦ Enter a value for **Maximum number of cycles of refinement using alignment-based HMMs** to set a cutoff for the number of HMM iterations. MegAlign can often improve the placement of gaps by comparing the alignment to a Hidden Markov Model (HMM) of the alignment. The default is 1. The HMM is very slow, so be aware that entering a large number may cause a commensurate decrease in alignment speed.

  In both cases, MegAlign Pro will recompute the pairwise distances, recalculate the guide-tree, and then realign the sequences while staying within the maximum cycle threshold.

- If your project contains ≥ 100 sequences, three clustering options are available. All three options are disabled for projects containing fewer than 100 sequences.

  - ◦ To force clustering using a distance matrix based comparison, check **Use distance matrix instead of k-tuple based mbed method for initial alignment**.

  - ◦ To use mbed for the first iteration and a distance matrix for other iterations, check **Use mbed for first round and distance matrix for later iterations**.

  - ◦ If you check the second box, but not the first, you may optionally set a threshold for the cluster size during mbed iterations by checking **Restrict cluster size in mbed mode**. Once you have checked that box, you may elect to over-ride the default by entering a value for Maximum sequences per subcluster. The default is <number of sequences to be aligned> minus '1,' up to a maximum value of 100.

The following table shows how the MegAlign Pro settings correspond with settings in [Clustal Omega](#) online.

MegAlign Pro setting	Clustal Omega setting
Iterate to refine alignments	Does not correspond with the command-line, but enables the two settings directly below.
Maximum number of cycles of tree refinement	max-guidetree-iterations=(integer) iterations=(integer)
Maximum number of cycles of refinement using alignment-based HMMs	max-hmm-iterations=(integer) iterations=(integer). The value used for "iterations" is the maximum of the values specified for this option and the option in the preceding row.

Use distance matrix instead of k-tuple (enabled)	full
Use mbed for first round and distance matrix for later iterations (enabled)	full-iter
Maximum sequences per subcluster	cluster-size=(integer)

After making your changes:

- Choose **Align** to use the entered options to perform a multiple sequence alignment.

- Use the **Reset to Default** button if you would like to reset all values to the MegAlign Pro defaults.

- Select **Cancel** to leave the dialog without saving any changes or performing the alignment.

Clustal W alignment options

The Clustal W algorithm is for gene level alignment of either protein or nucleotide sequences. To run a Clustal W alignment, select two or more sequences and choose **Align > (Re)Align Using Clustal W**.

If you wish to change method options, instead choose **Align > Align with Options**. The dialog has slightly different options, depending whether the sequences are protein or nucleotide (nucleotide shown below).

Change settings as desired:

- Use the **Align** drop-down menu to select sequences to align or realign.

- In the **Using** drop-down menu, choose **Clustal W**.

- Use the **Scoring matrix** drop-down menu to choose between these options:

Matrix	Description		
Nucleotide sequences only			
IUB	The default scoring matrix used by BESTFIT for the comparison of nucleic acid sequences. X's and N's are treated as matches to any IUB ambiguity symbol. All matches score 1.9; all mismatches for IUB symbols score 0.		
Clustal W	The previous system used by Clustal W, in which matches score 1 and mismatches score 0. All matches for IUB symbols also score 0.		
Protein sequences only			
BLOSUM	([Henikoff & Henikoff](), 1992). BLOSUM is ideal for carrying out similarity searches.	PAM	([Dayhoff et al.](), 1978). Widely used since the late 1970s.
GONNET	Derived from PAM matrices ([Dayhoff et al.](), 1978) but more sensitive, and based on a much larger data set.		
ID	Scores two identical amino acids as 1, and anything else as -10,000.		

- In the **Iteration method** drop-down menu, choose between **None**, **All alignment steps**, and **Final alignment only**. If you choose an item other than **None**, type the number of **Maximum Iterations** into the text box on the right.

- Enter the desired Multiple alignment parameters:

   ◦ **Gap open penalty**, the numerical penalty for introducing a gap of any length when calculating alignments. This penalty does not take into account the size of the gap. The default is 10.

   ◦ **Gap extend penalty** – This option affects the lengths of gaps, and must be zero or a negative number. Lowering the magnitude of the gap extension penalty may allow for longer gaps. The default is 0.20.

   ◦ **Delay divergent sequences** – This setting aligns more closely related sequences, delaying the alignment of more distantly related ones. Input the percent identity required to delay the addition of a sequence. The default is 30.

   ◦ **Transition weight** – This option is only available for nucleotide sequences and gives purine-purine or pyrimidine-pyrimidine substitutions (i.e., A-G or C-T) a weight between 0 and 1. One denotes a match, while zero denotes a mismatch. Assign a high score for closely related sequences and a lower score for distantly related ones. The default is 0.50.

- Enter the desired Pairwise alignment parameters:

- ◦ Choose whether MegAlign Pro should use the **Slow, accurate** or **Fast, approximate** version of Clustal W when making the guide tree.

- ◦ Type in the **Gap open penalty**, the numerical penalty for introducing a gap of any length when calculating alignments. This penalty does not take into account the size of the gap. The default is 10.

- ◦ **Gap extend penalty** – This option affects the lengths of gaps, and must be zero or a negative number. Lowering the magnitude of the gap extension penalty may allow for longer gaps. The default is 0.1.

After making your changes:

- • Choose **Align** to use the entered options to perform a multiple sequence alignment.

- • Use the **Reset to Default** button if you would like to reset all values to the MegAlign Pro defaults.

- • Select **Cancel** to leave the dialog without saving any changes or performing the alignment.

MAFFT alignment options

The MAFFT algorithm is for gene level alignment of either protein or nucleotide sequences. If you want to align a large number of sequences in MegAlign Pro, we recommend using this algorithm. As of the Lasergene 17.3 release (August 2021), MegAlign Pro uses MAFFT7. This powerful aligner can typically align thousands of viral genomes, for example, in under two minutes. The ability to align to a specified reference sequence was added in the Lasergene 17.3.1 release (January 2022).

To run a MAFFT alignment, select two or more sequences and choose **Align > (Re)Align Using MAFFT**. If you wish to change method options, instead choose **Align > Align with Options**. Options vary depending whether the sequences are protein or nucleotide (nucleotide version shown below).

Change settings as desired:

- Use the **Reference Sequence** drop-down to select either **None** or one of the sequences that have been chosen for alignment. This option is ideal for the small percentage of MegAlign Pro users that need an extremely high capacity alignment algorithm.

- Use the **Align** drop-down menu to select sequences to align or realign.

- In the **Using** drop-down menu, choose **Clustal W**.

- Specify the **Gap open penalty**, the numerical penalty for introducing a gap of any length when calculating alignments. This penalty does not take into account the size of the gap. The default is -2.0.

- **Gap extension penalty** – This option affects the lengths of gaps, and must be zero or a negative number. Lowering the magnitude of the gap extension penalty may allow for longer gaps. The default

is 0.

- In the **Algorithm** drop-down menu, either retain the default of **Choose algorithm depending upon size [auto]**, or specify a particular alignment algorithm.

Description	Name	No. of Seqs	Strategy
Choose algorithm depending upon size [auto]			This is the default method for most nucleotide sequences. (MAUVE is the default method for sequences >1Mb in size.
Very slow	global homology	< 200	g-ins-i
Very slow	one conserved domain	< 200	l-ins-i. This is the default method used for protein sequences.
Very slow	multiple conserved domains	< 200	e-ins-i
Slow	iterative refinement, iterations specified by user	N/A	fft-ns-i
Medium	Iterative refinement, two iterations	N/A	N/A
Fast	Progressive	N/A	fft-ns-2
Very fast	Progressive	> 2000	fft-ns-1

- in the **Scoring matrix** drop-down menu, choose the desired matrix:

Matrix name	Description
Nucleotide sequences	
1PAM / k=2	For closely related sequences.
20PAM / k=2	For moderately related sequences.
200 PAM / k=2	For distantly related sequences.
Protein sequences	

BLOSUM30, 45, 62, 80	(Henikoff & Henikoff, 1992). The BLOSUM series of matrices contain the same values as in some of MegAlign Pro's other alignment methods, except that the protein ambiguity codes B, Z, and X are excluded in this case. These matrices are ideal for carrying out similarity searches. Choose a larger number BLOSUM matrix for less divergent sequences.
JTT100, 200	(Thorne JL et al., 1998) These matrices are similar to PAM matrices and were generated using an algorithm similar to the approach of Dayhoff et al. (1978), but based on a larger set of protein sequences.

After making your changes:

- Choose **Align** to use the entered options to perform a multiple sequence alignment.

- Use the **Reset to Default** button if you would like to reset all values to the MegAlign Pro defaults.

- Select **Cancel** to leave the dialog without saving any changes.

MUSCLE alignment options

The MUSCLE algorithm is for gene level alignment of either protein or nucleotide sequences. To run a MUSCLE alignment, select two or more sequences and choose **Align > (Re)Align Using MUSCLE**.

If you wish to change method options, instead choose **Align > Align with Options**. Options vary depending whether the sequences are protein or nucleotide (protein version shown below).

Change settings as desired:

- Use the **Align** drop-down menu to select sequences to align or realign.

- In the **Using** drop-down menu, choose **MUSCLE**.

- **Substitution matrix** – (protein sequences only) Choose from a list that includes one or more members of the VTML, BLOSUM GONNET and PAM series of matrices.

- **Gap open penalty** – This option affects the chance of gaps being introduced into alignments, and must be zero or a negative number. Decreasing the magnitude of the gap open penalty may lead to an excessive amount of gapping. The default is -400 for DNA; -420 for RNA; -2.9 for protein.

- **Gap extension penalty** – This option affects the lengths of gaps, and must be zero or a negative number. Lowering the magnitude of the gap extension penalty may allow for longer gaps. The default is 0.

- **Center penalty** – This must be zero or a negative number. The default is 0 for DNA; -300 for RNA; and -0.52 for protein.

- **Maximum iterations** – This option, which must be a positive integer, affects the number of rounds of refinement used to produce an alignment. Depending on your data, the alignment might finish before the specified number of rounds have been performed. The default is 8.

- **Use Brenner's Alignment** – Checking the box causes MegAlign Pro to use Brenner's method for the root alignment. Brenner's method uses less memory than the default root alignment, but the alignment will not be as accurate as if the box were left unchecked. For this reason, it is not usually necessary to check the box unless you received certain types of error messages during a previous alignment attempt.

- **Disable anchors optimization** – The default for MUSCLE is for anchor optimization to be enabled (i.e., for this box to be unchecked). One side effect of anchor optimization is that the selected number of **Maximum iterations**, if greater than two, may be over-ruled during sequence processing. If you instead wish to run the exact number of iterations selected in **Maximum iterations** (above), add a checkmark to this box.

- Use the drop-down menus to choose the desired **Cluster method** for **Initial iterations** and **Later iterations**.

Cluster method	Description
UPGMA	Fast O(N2) approximation to the textbook UPGMA algorithm.
UPGMB (default)	Variant of UPGMA that is used, by default, for both iteration 1 and 2.
Neighbor Joining	Better at making phylogenetic trees, but slower, and usually worse for making guide trees for progressive alignment. Not recommended for most alignments.

- Choose the desired **Distance measures**.

Distance measure	Description
Initial iterations	

Kmer 4-6 [default for DNA/RNA; not available for protein]	
Kmer 6-6 [default for protein]	
Kmer 20-3	
Kmer 20-4	
Kmer bit 20-3	
Later iterations	
Kimura % identity [default]	
Log % identity	

- Select the desired **Sequence weights** for **Initial iterations** and **Later iterations**.

Sequence weight	Description
None	Do not weight sequences.
ClustalW (default)	
Henikoff	
Henikoff PSI-BLAST	
gsc	
Threeway	

The following table shows how the MegAlign Pro settings correspond with settings in [MUSCLE](#) online.

MegAlign Pro setting	MUSCLE setting
Substitution matrix	matrix
Gap open penalty	gapopen
Gap extension penalty	gapextend
Center parameter	center
Maximum iterations	maxiters
Use Brenner's Alignment	brenner
Disable anchors optimization	anchors
Cluster method: Initial iterations Later iterations	 cluster1 cluster2

Distance measure: Initial iterations Later iterations	distance1 distance2
Sequence weight: Initial iterations Later iterations	weight1 weight2

After making your changes:

- Choose **Align** to use the entered options to perform a multiple sequence alignment.

- Use the **Reset to Default** button if you would like to reset all values to the MegAlign Pro defaults.

- Select **Cancel** to leave the dialog without saving any changes.

Mauve alignment options

The Mauve alignment algorithm is for genome level alignment of nucleotide sequences. To run a Mauve alignment, select two or more nucleotide sequences and choose **Align > (Re)Align Using Mauve**.

If you wish to change method options, instead choose **Align > Align with Options**.
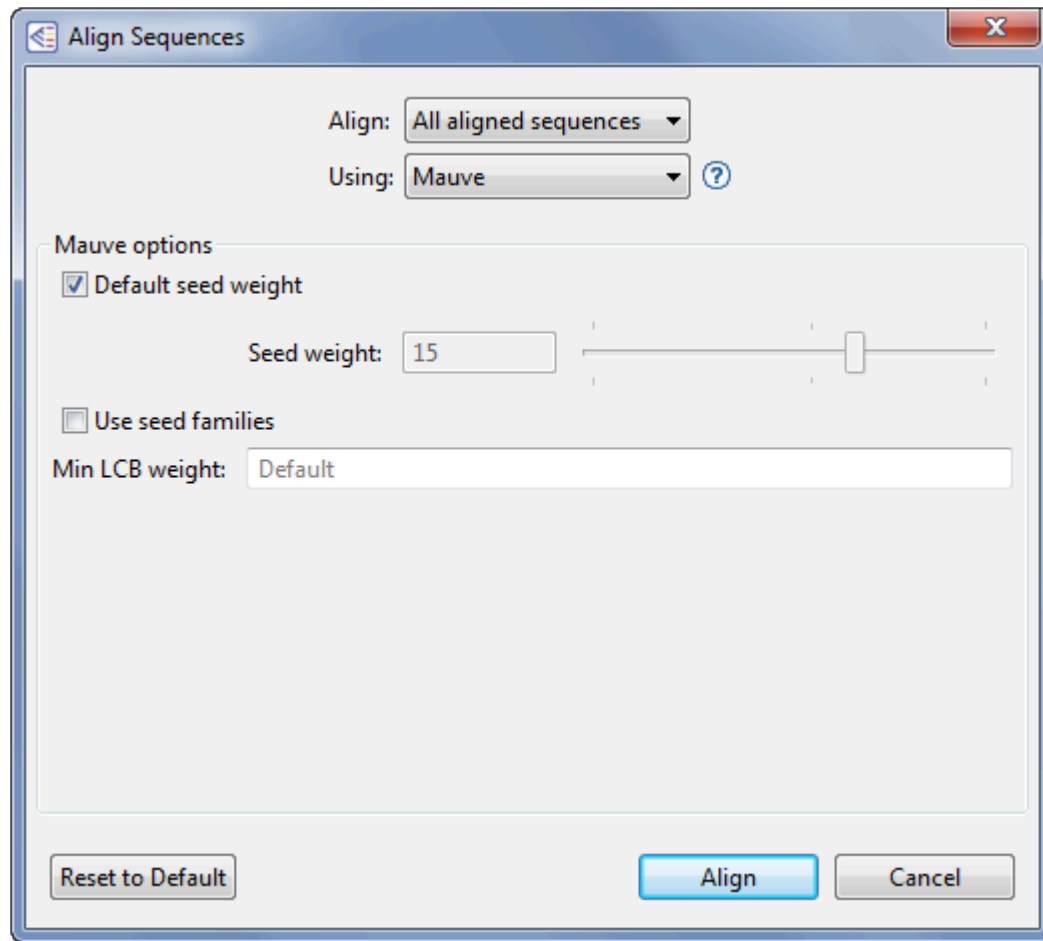
Change settings as desired:

- Use the **Align** drop-down menu to select sequences to align or realign.

- In the **Using** drop-down menu, choose **Mauve**.

- If the **Default seed weight** box is checked, as it is by default, the optimal seed weight (number of matches required per k-tuple) is calculated automatically. You may uncheck the box and either use the slider or manually enter a number in the **Seed weight** box to specify a particular seed weight between 3 and 21. The default is 15.

- Check the **Use seed families** box to use multiple spaced seed patterns to identify potential homology. When working with divergent genomes, checking this box can significantly improve the sensitivity and

accuracy of the alignment.

- Whether or not you have checked either of the boxes in this dialog, you may override the default **Min LCB weight**, if desired, by entering a value in the box. Increasing the **Min LCB weight** may reduce the number of spurious rearrangements, and is advisable in many cases.

The following table shows how the MegAlign Pro settings correspond with settings in Mauve online.

MegAlign Pro setting	Mauve setting
Default seed weight (disabled)	seed-weight
Use seed families (enabled)	seed-family
Min LCB weight (non-default value entered)	weight

After making your changes:

- Choose **Align** to use the entered options to perform a multiple sequence alignment.

- Use the **Reset to Default** button if you would like to reset all values to the MegAlign Pro defaults.

- Select **Cancel** to leave the dialog without saving any changes.

The following videos demonstrate how to perform a multiple alignment using the Mauve alignment algorithm:

The following video shows a multiple alignment using the Mauve algorithm:

Modify a multiple alignment

The following table describes ways to modify an [initial multiple alignment](#):

Task	See this topic
To realign a range of aligned sequences using different parameters or a different alignment engine	[Subalign sequences](#)
To merge two existing completed alignments together	[Merge two existing alignments](#) ("Profile alignment")
To add all/selected sequences from the "Unaligned Sequences" area into the current alignment, while retaining all of the gaps of the original, pre-merge alignment	[Merge unaligned sequences into an existing alignment](#)
To realign existing aligned sequences PLUS selected unaligned sequences	[Align selected unaligned sequences with an existing alignment](#)

The following video illustrates some of these modification procedures:

Subalign sequences

While sequences such as full-length calmodulin proteins typically align well throughout the entire length of the polypeptides, full-length alignments of other sequences may not generate such neat blocks of similarity. For example, multiple sequence alignments of full-length beta-tubulins typically show "ragged" carboxyl termini, where the different tubulins terminate at varying positions with respect to the consensus. The "ragged" termini are caused by considerable length variance of the beta tubulins and carboxyl termini that are poorly conserved. Therefore, once you have performed an initial multiple alignment for two or more sequences, you may wish to realign a section using different parameters or a different alignment engine.

To perform a sequence subalignment:

1. Make a single selection across the aligned sequences. One way to do this is to drag the mouse along the consensus sequences. Another is to triple-click on a feature. Either selects a "vertical slice" of all the sequences in the current alignment. If you plan to use Mauve as the alignment algorithm, note that you can only select and subalign sequences within the active block.

2. Choose **Align > Realign Subsequences** to open a dialog populated for a sequence subalignment.

3. Choose the desired alignment method from the drop-down menu, then set other options as desired.
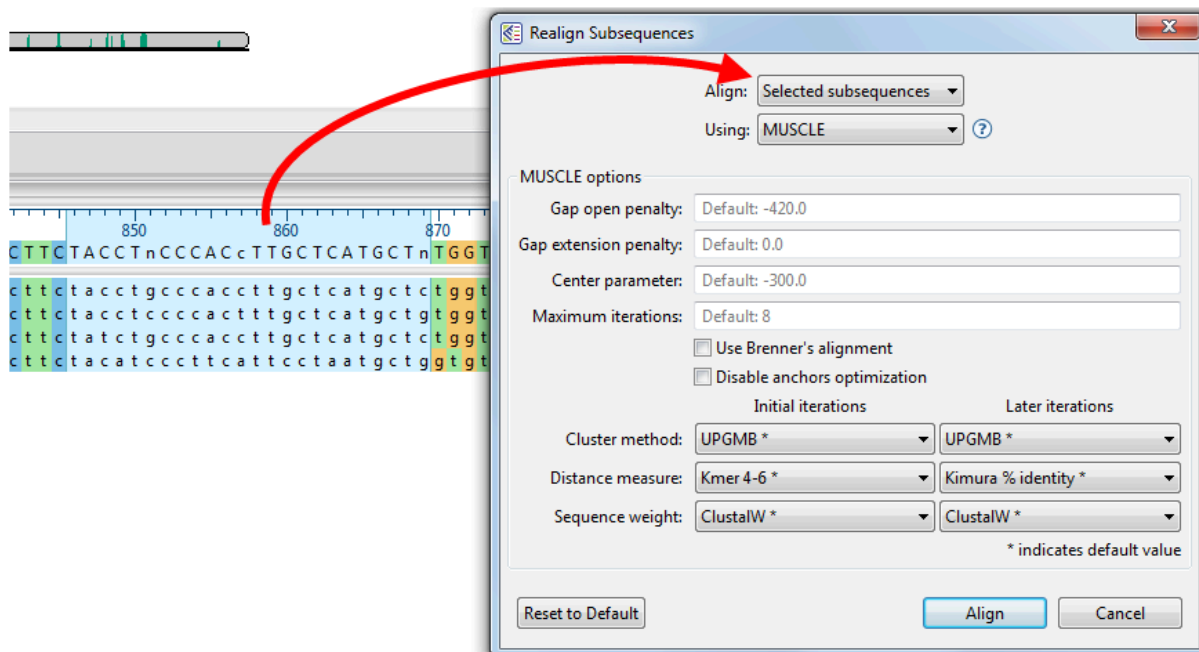
4. Press **Align**.

✳ **Notes**:

- Subalignments are not supported when the document contains any end-to-end sequences.

- Occasionally, the new subalignment may no longer be the same as the optimal alignment your alignment method would generate from scratch for this particular set of sequence coordinates. You may want to realign the shorter sequences to evaluate this possibility.

- If you are not sure exactly which segment of an alignment you need for your new alignment, you may select the region liberally, generate the new alignment, and continue to select sub-regions of the alignment until you are satisfied.

- You can create a new alignment from a selection of any length, but you cannot create a new alignment from discontiguous segments of your current alignment.

Merge two existing alignments ("Profile alignment")

To merge two existing alignments:

A "profile alignment" occurs when you merge a MegAlign Pro alignment (*.msa*) that was saved earlier with the current project. Profile alignments merge two existing multiple alignments without removing any of the existing gaps. However, new gaps may be automatically inserted to reconcile the new alignment. Mauve alignments are not supported in this workflow.

1. Create a project, align the sequences using any alignment method except Mauve, and then save the project in MegAlign Pro (*.msa*) format.

2. Create a second project and perform an alignment using any method except Mauve. Both projects must use the same sequence type (nucleotide vs. amino acid), but it is not necessary to use the same alignment method that was used in Step 1. If the projects do not contain the same sequence types, you will receive an error message:
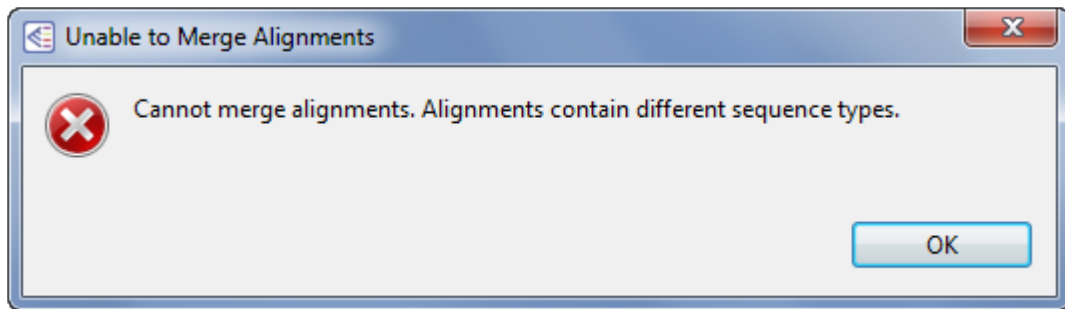
3. (optional) The next step will over-write the alignment in the second (current) project. Therefore, you may wish to save the alignment before proceeding further.

4. From within the second (current) project, select **Align > Merge Alignment into Existing Alignment**.

5. In the ensuing Open dialog, navigate to the saved project and click **Open**. A warning will be displayed if you attempt to merge an unsaved project with the active project.

6. Choose the desired alignment method from the drop-down menu, then set other options as desired. Note that Mauve and MAFFT alignment methods are not supported for this workflow.

7. Press **Align**.

> ✳ **Note:** Any unaligned sequences from the merged project are automatically transferred to the "Unaligned Sequences" area of the current project.

Merge unaligned sequences into an existing alignment

To merge unaligned sequences into the current project's existing alignment:

In contrast to the "Profile alignment" method, this method retains existing gaps, and adds new gaps, if necessary.

1. Add sequences to the project and align some or all of them. If all sequences were aligned, add more sequences to the project.

2. Before proceeding, check that the "Unaligned Sequences" bars in the Overview and Sequences view display a number ≥ 1 in the parentheses. If you wish, you may show or hide the unaligned sequences by clicking the arrows on the left of each bar.
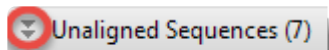
3. (optional) An upcoming step will overwrite the current alignment. Therefore, you may wish to save the original alignment before proceeding further.

4. Specify which unaligned sequences to add to the current alignment.

5. Select **Align > Merge Unaligned into Existing Alignment** or press the **Align sequences to an existing alignment** tool (   ). The menu command and tool are enabled only if the project contains an alignment and at least one unaligned sequence.

6. Choose the desired alignment method from the drop-down menu, then set other options as desired. The Mauve alignment method is not supported in this workflow.

7. Press **Align**.

Align selected unaligned sequences with an existing alignment

In some cases, you may have already aligned sequences, then decide to completely realign the same sequences plus certain selected unaligned sequences.

To align selected unaligned sequences with an existing alignment:

1. Add sequences to the project and align some or all of them. If all sequences were aligned, add more sequences to the project.

2. Before proceeding, check that the "Unaligned Sequences" bars in the Overview and Sequences view display a number ≥ 1 in the parentheses. If you wish, you may show or hide the unaligned sequences by clicking the arrows on the left of each bar.

3. (optional) An upcoming step will over-write the current alignment. Therefore, you may wish to save the original alignment before proceeding further.

4. Specify which unaligned sequences to include in the new alignment by clicking on them or by using **Ctrl/Cmd+click** or **Shift+click**.

5. Select **Align > Realign with Selected Unaligned**.

6. Choose the desired alignment method from the drop-down menu, then set other options as desired.

7. Press **Align**.

# Unalign aligned sequences

To unalign all sequences in an alignment:

1. Choose **Align > Unalign All**.

2. When the following message appears, click **OK**.
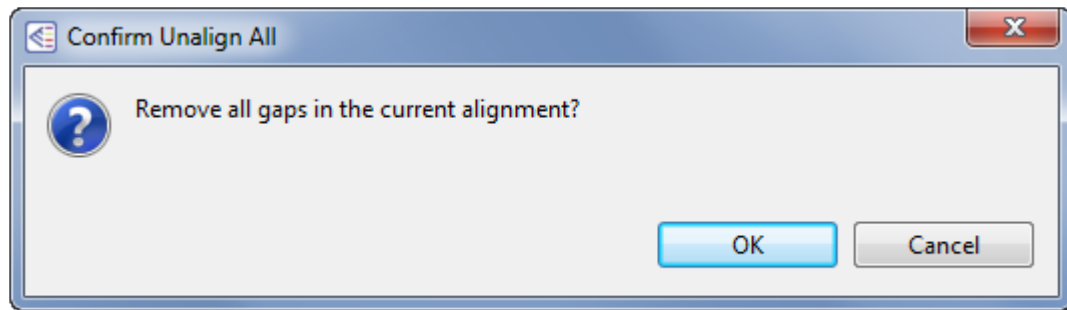
Execution of this command causes gaps to be removed from all of the previously aligned sequences. The sequences are then automatically relocated to the "Unaligned Sequences" areas of the Overview and Sequences view. In some cases, a progress bar may be displayed during the unalignment procedure.

To unalign particular aligned sequences:

1. Select one or more of the aligned sequences in the Overview or Sequences view.

2. Choose **Align > Unalign Selected**.

3. When the following message appears, click **OK**.
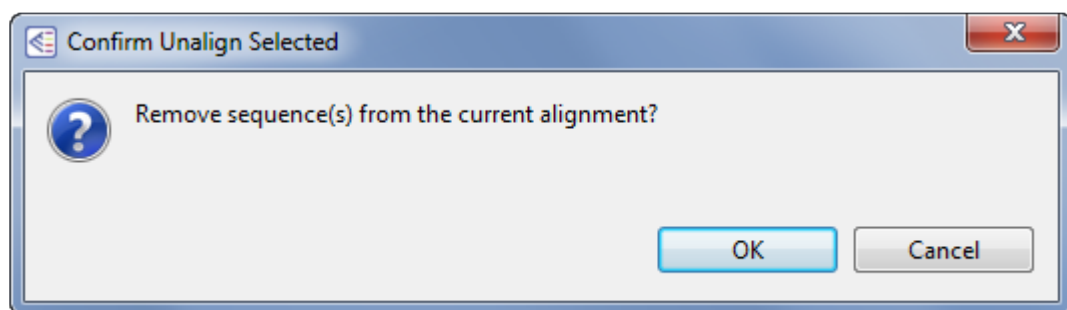
Depending on the size of the alignment, a progress dialog entitled Removing All-Gaps Columns may be displayed during the unalignment process.

Execution of this command causes gaps to be removed from all of the previously aligned sequences. The sequences are then automatically relocated to the "Unaligned Sequences" areas of the Overview and

Sequences view. In some cases, a progress bar may be displayed during the unalignment procedure.
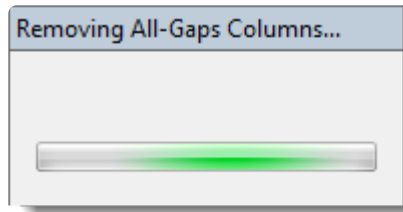
Detect and remove outliers from the alignment

After performing a multiple sequence alignment, you may wish to detect and remove outliers from the alignment.

1. From the Distance view, select a reference sequence.

2. Choose **Distance > Order Sequences by Distance from Selection**. By default, a heat map color scheme is applied. This can be customized in the Distance section of the Style panel.

3. From the Distance table, select the sequences you want to remove from the alignment. Note that the rows nearest to the bottom of the table differ the most from the reference sequence. The Distance table metrics and heat map coloration can help you decide which sequences to remove.

4. Choose **Align > Unalign Selected**.

Try it! – Perform a Clustal Omega alignment

This tutorial will use the Clustal Omega algorithm to multiply align 143 DNA sequence files consisting of regional variations of the Human Papilloma Virus L1 gene (HPV-L1). HPV has affected humans for hundreds of thousands of years and is known to cause cervical cancer. Some regional variations of the gene are more pathogenic than others. For example, HPV-16 is the most pathogenic variation in Western populations.

The L1 gene is not the cause of variation in pathogenicity. However, variations of the L1 gene are used to tell various strains apart and group them according to relatedness. The results of this tutorial reveal a deeply branched tree revealing an ancient phylogeny for HPV within the human genome.

Begin with Part A: Add and align sequences.

Part A: Add and align sequences

In this part of the tutorial, you will be adding 143 DNA sequences to the MegAlign Pro project and aligning them with the Clustal Omega algorithm using the default settings.

1. If you have not yet downloaded and extracted the tutorial data, click here to download it. Then decompress (unzip) the file archive using the method of your choice.

2. If a MegAlign Pro project is already open, choose **File > New Alignment Project**. Otherwise, launch MegAlign Pro.

3. Choose **File > Add Sequences**, or click on the **Add sequences to project** tool ( ).

4. Open the folder *HPV project*. Use **Ctrl+A** (Win) or **Cmd+A** (Mac) to select all 143 DNA sequences and click **Open**.

5. When prompted, select **DNA**.
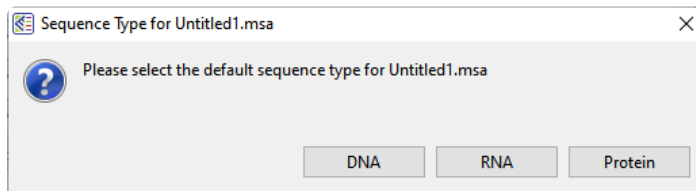
6. In the Overview at the top left of the MegAlign Pro window, note the appearance of a series of gray blocks, one for each sequence. The gray blocks have the lengths of their corresponding sequences. Solid gray blocks and varying block lengths are both indicative of sequences that have not yet been aligned.
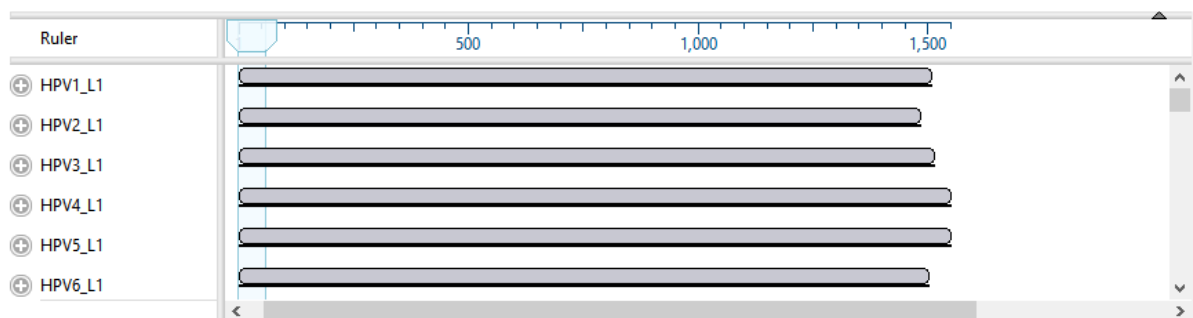
7. To align sequences with Clustal Omega using the default parameters, choose **Align > Align Using Clustal Omega** or click the **Align** tool arrow ( ) and choose **Align Using Clustal Omega**. Sequences now appear mostly in green, with gray representing gaps.

Proceed to Part B: Navigate using the Overview.

Part B: Navigate using the Overview

Observe that the sequence blocks in the Overview are now the same lengths. Sequence gaps are represented by gray coloring within the blocks.
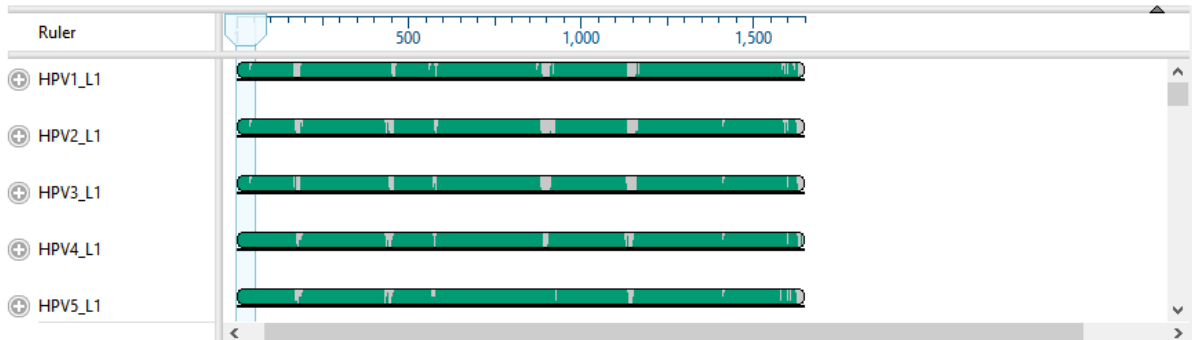
The light blue vertical column shown at the left of the green blocks, above, is a "viewport" showing which region is currently being displayed in the Sequences view. The width of the viewport will vary depending on the zoom levels in both this view and the Sequences View. You can use the mouse to drag the viewport to whichever region you wish to see in the Sequences view.

- Using the mouse, slide the viewport to the far right, and then back to the left. At the same time, notice how the Sequence view slides in the same way.

- To navigate to a particular residue in the Sequences view, double-click anywhere on a block in the Overview. The viewport will move so that it is centered on the selection.

- Use the horizontal zoom slider in the Sequences view to zoom in and out. As you do this, notice how the width of the viewport in the Overview changes.

Proceed to Part C: Change sequence names and rendering in the Sequences view.

Part C: Change sequence names and rendering in the Sequences view

In Part B, you looked at the Sequences view while learning the function of the "viewport." You will now explore the Sequences view in more depth. The Sequences view displays entered sequences and aligned sequences. Sequences selected in this view are simultaneously selected in all other MegAlign Pro views.

1. In the Sequences view, observe that the names of each sequence originated from the accession numbers that were embedded in the input files.
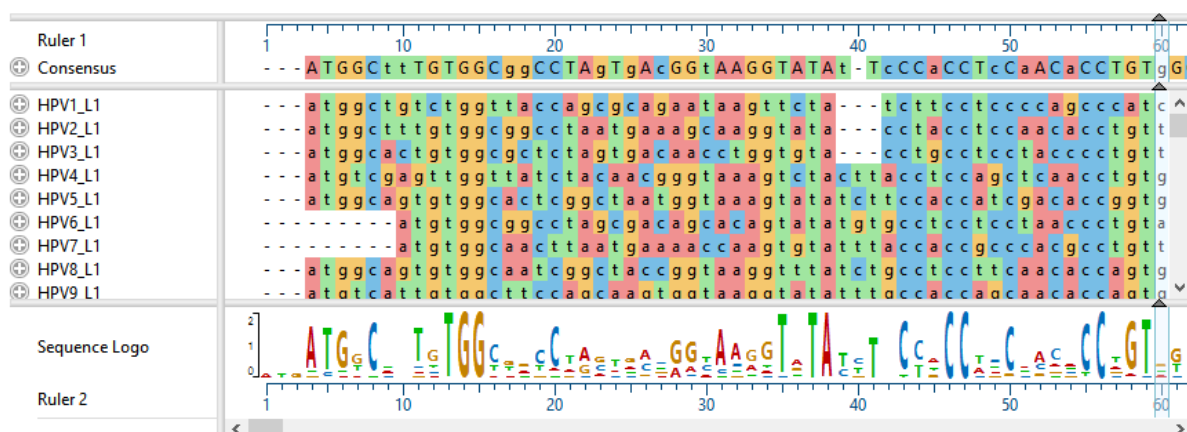
To also add the sequence length as part of the name:

   a. Use **Ctrl+A** or **Cmd+A** to select all of the sequence names.

   b. Right-click on the highlighted group and select **Rename Sequences with Fields**, or choose **Edit > Rename Sequences with Fields** from the main menu.

   c. Select **Sequence length** from the left pane and click the right arrow to add it.
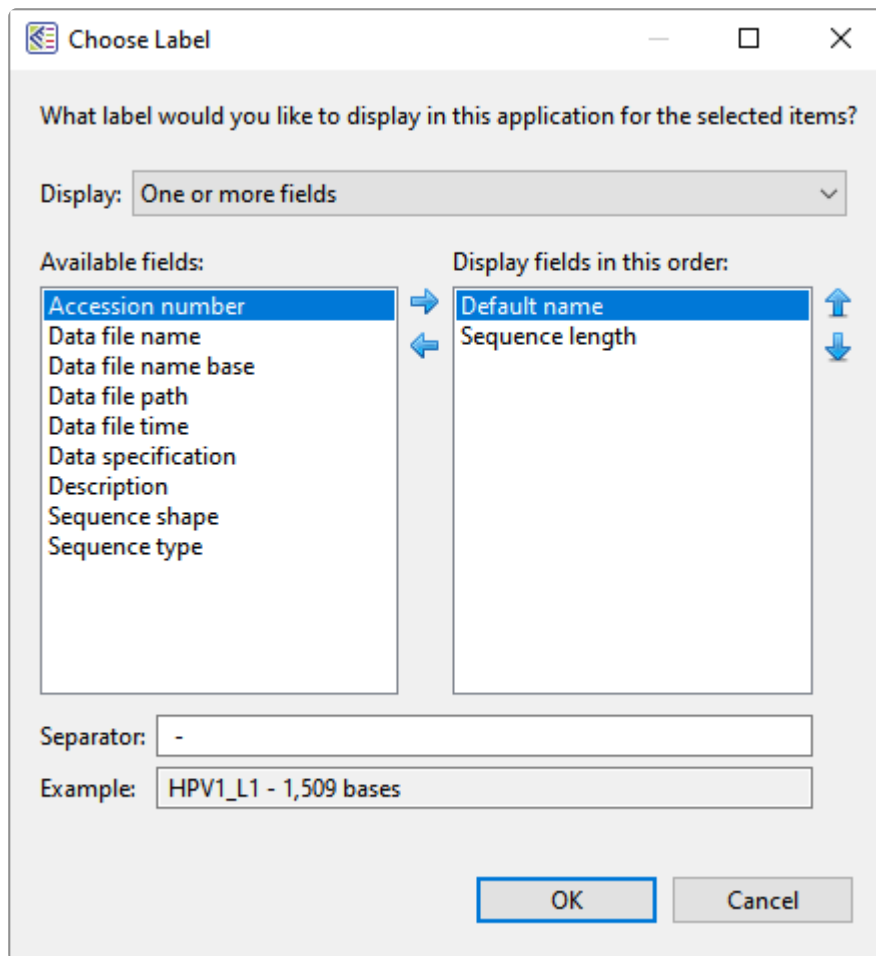
d. Click **OK**.

The sequence labels now reflect their lengths in base pairs:

2. Remove the current selection by double-clicking on any white space, e.g., white space in the header or footer area of the Overview or Sequences view.

3. Use **File > Save** to save the project as *HPV project.msa*.

4. In the Sequence view, click on any plus icon (⊕) to reveal a details track containing a gap-aware ruler. These sequences do not contain annotations, but if they did, the annotations would also be visible.

5. The appearance of the sequences can be changed using settings in the Style panel's Sequence and Multiple Alignment sections located to the right of the views. Experiment with the items in this area to see how they affect the look of the Sequences view. Some items must first be enabled by checking the associated checkbox.
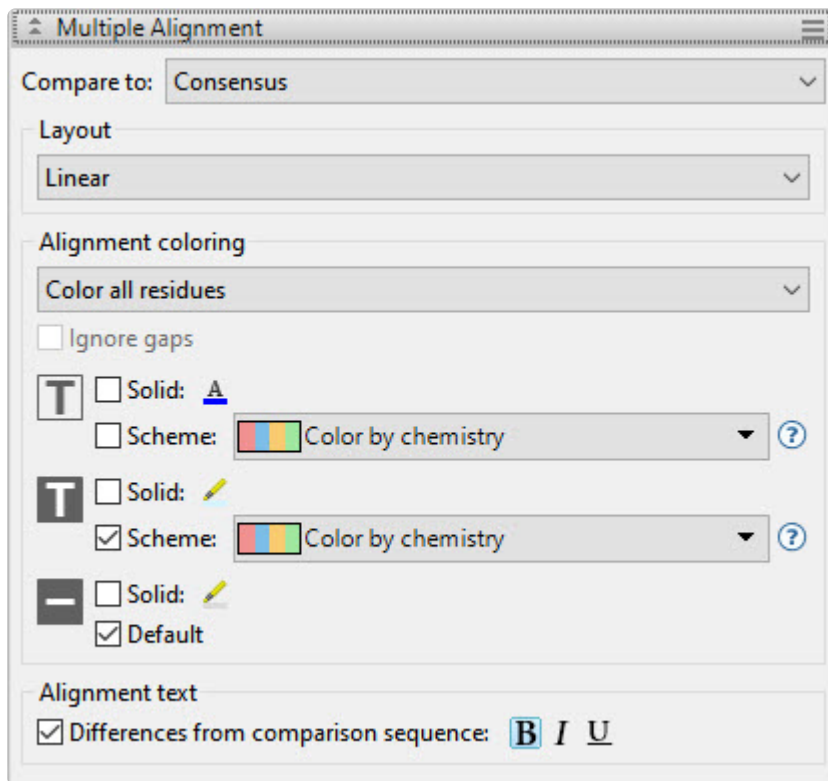
When you have finished trying out different color schemes and fonts, we recommend changing them back to the original settings (shown in the image above) for ease of following the remainder of this tutorial.

6. Imagine that you want to see the Consensus Match track, but it is not included in the default display. To display the track, click the **Tracks** tab in the upper right of MegAlign Pro (or select **View > Tracks > Tracks**) and check the box next to **Consensus Match**.
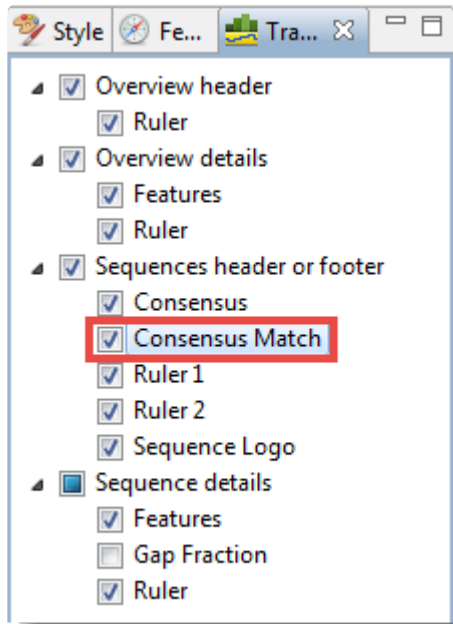
7. Locate the **Consensus Match** histogram in the footer of the Sequences view. Hover over a bar to see a pop-up showing the percent of sequences that match the consensus at that residue.
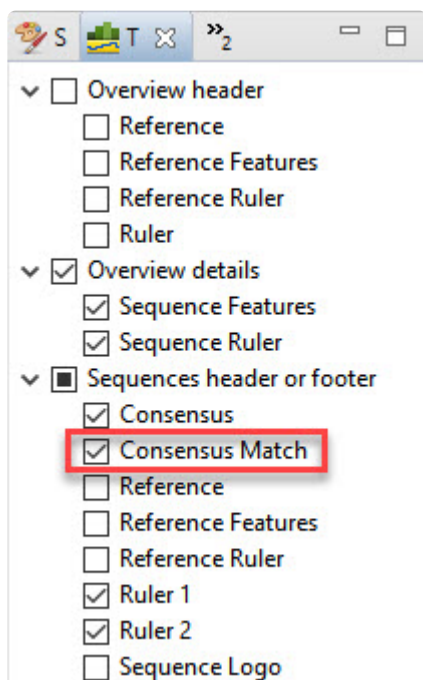
8. Hide the footer by clicking on the small gray arrow near its top right corner ( ). The Sequence Logo, Ruler and Consensus Match will disappear. Click on the arrow again to redisplay the footer. (Note: For future reference, you can close other panes the same way, including the footer and the main parts of the Sequence view and Overviews).

Proceed to Part D: Change metrics and rendering in the Distance view.

Part D: Change metrics and rendering in the Distance view

The Distance view shows a matrix (i.e., table) of numbers representing distances between each pair of sequences. Sequence distances are used as input in creating the Tree view, which you will explore in Part E. Selections within this view are synchronized with other views in MegAlign Pro.

1. To display the Distance view, click on the **Distance** tab at the bottom of the MegAlign Pro window.

   The Distance view matrix is calculated, by default, using Uncorrected Pairwise Distance. The default display shows **%Identity** results in the upper right and **Distance** results in the lower left.

|		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA
A	HPV1_L1 - 1,509 bases		56.69	56.85	59.92	59.35	58.47	58.15	58.47	60.40	57.74	59.11	59.92	57.58	60.81	61.05	60.73	59.27	58.31	60.73	60.40	60.73	59.84	59.44	60.08	60.40	58.71	55.81
B	HPV2_L1 - 1,485 bases	0.43		68.63	57.10	57.58	67.34	64.52	59.03	57.42	68.47	67.10	57.42	65.24	58.55	56.69	66.61	57.82	68.15	58.23	57.98	56.77	56.29	56.94	56.94	58.71	64.03	90.16
C	HPV3_L1 - 1,515 bases	0.43	0.31		58.31	58.79	67.18	66.37	58.95	59.84	82.66	68.79	59.60	67.58	59.44	58.15	68.31	58.31	70.97	58.79	59.84	58.15	57.50	58.79	58.63	59.60	69.03	68.47
D	HPV4_L1 - 1,551 bases	0.40	0.43	0.42		65.08	61.45	61.21	62.10	63.06	58.79	60.89	63.06	60.65	65.97	62.82	62.34	61.85	62.34	65.56	64.92	65.32	61.13	61.94	64.19	65.16	60.08	57.58
E	HPV5_L1 - 1,551 bases	0.41	0.42	0.41	0.35		60.89	61.29	80.48	72.74	59.35	60.00	82.34	60.16	77.66	71.61	60.56	71.05	61.29	78.23	77.74	78.15	69.52	69.35	77.02	76.37	60.65	57.50
F	HPV6_L1 - 1,503 bases	0.42	0.33	0.33	0.39	0.39		68.79	60.48	59.19	66.85	85.73	60.00	77.58	60.00	60.08	70.81	60.00	68.47	59.84	60.32	59.35	58.63	59.03	59.19	60.73	69.35	67.18
G	HPV7_L1 - 1,518 bases	0.42	0.35	0.34	0.39	0.39	0.31		59.52	60.97	66.45	69.92	61.21	70.08	61.69	61.69	71.53	61.61	69.35	60.81	61.85	59.60	60.32	60.24	60.89	61.21	68.31	64.03
H	HPV8_L1 - 1,545 bases	0.42	0.41	0.41	0.38	0.20	0.40	0.40		70.24	60.97	59.68	81.77	60.08	77.74	70.81	61.61	70.48	59.84	77.58	76.21	76.85	68.31	69.84	76.85	76.21	60.40	59.11
I	HPV9_L1 - 1,524 bases	0.40	0.43	0.40	0.37	0.27	0.41	0.39	0.30		59.11	59.68	72.02	60.65	72.74	78.95	61.45	78.95	61.29	72.26	72.82	73.31	73.55	74.84	71.85	73.15	60.48	56.21
J	HPV10_L1 - 1,512 bases	0.42	0.32	0.17	0.41	0.41	0.33	0.34	0.39	0.41		67.42	59.11	67.50	60.65	57.02	68.95	57.82	70.48	59.35	61.13	60.00	57.90	58.95	59.19	59.84	69.84	68.15
K	HPV11_L1 - 1,506 bases	0.41	0.33	0.31	0.39	0.40	0.14	0.30	0.40	0.40	0.33		61.21	77.82	61.05	60.08	71.61	60.48	69.19	60.40	61.61	59.27	59.68	60.16	59.76	60.56	67.58	66.85
L	HPV12_L1 - 1,554 bases	0.40	0.43	0.40	0.37	0.18	0.40	0.39	0.18	0.28	0.41	0.39		61.13	78.31	70.81	61.85	70.32	61.53	79.11	78.23	76.77	67.82	68.87	76.94	77.34	59.76	58.79
M	HPV13_L1 - 1,500 bases	0.42	0.35	0.32	0.39	0.40	0.22	0.30	0.40	0.39	0.32	0.22	0.39		61.29	61.13	71.77	61.77	69.60	59.84	61.21	61.29	60.97	60.89	59.52	59.19	68.63	64.92
N	HPV14_L1 - 1,557 bases	0.39	0.41	0.41	0.34	0.22	0.40	0.38	0.22	0.27	0.39	0.39	0.22	0.39		72.02	64.11	72.58	62.82	84.27	86.05	86.85	70.65	71.94	77.98	84.19	61.21	58.79
O	HPV15_L1 - 1,524 bases	0.39	0.43	0.42	0.37	0.28	0.40	0.38	0.29	0.21	0.43	0.40	0.29	0.39	0.28		63.06	80.81	60.65	72.10	71.53	70.97	75.32	75.32	70.97	71.21	60.16	56.53
P	HPV16_L1 - 1,518 bases	0.39	0.33	0.32	0.38	0.39	0.29	0.28	0.38	0.39	0.31	0.28	0.38	0.28	0.36	0.37		61.37	70.24	62.82	63.39	60.97	61.45	61.45	61.69	63.87	70.16	67.02
Q	HPV17_L1 - 1,524 bases	0.41	0.42	0.42	0.38	0.29	0.40	0.38	0.30	0.21	0.42	0.40	0.30	0.38	0.27	0.19	0.39		61.85	71.77	71.13	71.29	75.24	75.08	70.73	70.73	59.03	57.98
R	HPV18_L1 - 1,524 bases	0.42	0.32	0.29	0.38	0.39	0.32	0.31	0.40	0.39	0.30	0.31	0.38	0.30	0.37	0.39	0.30	0.38		61.85	62.74	60.40	60.40	60.97	61.53	62.42	71.29	67.26
S	HPV19_L1 - 1,554 bases	0.39	0.42	0.41	0.34	0.22	0.40	0.39	0.22	0.28	0.41	0.40	0.21	0.40	0.16	0.28	0.37	0.28	0.38		83.79	83.95	69.92	69.92	77.50	86.37	60.65	59.19
T	HPV20_L1 - 1,551 bases	0.40	0.42	0.40	0.35	0.22	0.40	0.38	0.24	0.27	0.39	0.38	0.22	0.39	0.14	0.28	0.37	0.29	0.37	0.16		86.21	70.40	70.40	79.27	83.71	61.77	57.90
U	HPV21_L1 - 1,557 bases	0.39	0.43	0.42	0.35	0.22	0.41	0.40	0.23	0.27	0.40	0.41	0.23	0.39	0.13	0.29	0.39	0.29	0.40	0.16	0.14		69.68	71.05	78.39	83.06	60.48	57.66
V	HPV22_L1 - 1,533 bases	0.40	0.44	0.43	0.39	0.30	0.41	0.40	0.32	0.26	0.42	0.40	0.32	0.39	0.29	0.25	0.39	0.25	0.40	0.30	0.30	0.30		78.79	70.16	69.27	60.81	56.21
W	HPV23_L1 - 1,521 bases	0.41	0.43	0.41	0.38	0.31	0.41	0.40	0.30	0.25	0.41	0.40	0.31	0.39	0.28	0.25	0.39	0.25	0.39	0.30	0.30	0.29	0.21		70.48	69.92	60.48	55.97
X	HPV24_L1 - 1,539 bases	0.40	0.43	0.41	0.36	0.23	0.41	0.39	0.23	0.28	0.41	0.40	0.23	0.40	0.22	0.29	0.38	0.29	0.38	0.22	0.21	0.22	0.30	0.30		77.18	60.24	57.82
Y	HPV25_L1 - 1,554 bases	0.40	0.41	0.40	0.35	0.24	0.39	0.39	0.24	0.27	0.40	0.39	0.23	0.41	0.16	0.29	0.36	0.29	0.38	0.14	0.16	0.17	0.31	0.30	0.23		61.13	59.27
Z	HPV26_L1 - 1,512 bases	0.41	0.36	0.31	0.40	0.39	0.31	0.32	0.40	0.40	0.30	0.32	0.40	0.31	0.39	0.40	0.30	0.40	0.31	0.39	0.38	0.40	0.39	0.40	0.40	0.39		64.84
AA	HPV27_L1 - 1,458 bases	0.44	0.10	0.32	0.42	0.43	0.33	0.36	0.41	0.44	0.32	0.33	0.41	0.35	0.41	0.43	0.31	0.43	0.33	0.41	0.42	0.42	0.44	0.44	0.42	0.41	0.35	
AB	HPV28_L1 - 1,509 bases	0.43	0.33	0.13	0.42	0.42	0.34	0.34	0.42	0.42	0.16	0.33	0.41	0.33	0.40	0.43	0.31	0.41	0.30	0.42	0.41	0.43	0.42	0.43	0.42	0.42	0.31	0.32
AC	HPV29_L1 - 1,512 bases	0.41	0.31	0.22	0.44	0.41	0.33	0.33	0.40	0.38	0.19	0.32	0.40	0.32	0.40	0.42	0.32	0.40	0.30	0.40	0.40	0.41	0.42	0.42	0.41	0.40	0.30	0.32
AD	HPV30_L1 - 1,506 bases	0.41	0.33	0.30	0.38	0.39	0.34	0.33	0.39	0.39	0.30	0.32	0.38	0.32	0.39	0.38	0.30	0.39	0.29	0.40	0.38	0.40	0.40	0.39	0.39	0.39	0.31	0.35
AE	HPV31_L1 - 1,455 bases	0.39	0.33	0.30	0.39	0.38	0.27	0.28	0.39	0.37	0.31	0.28	0.38	0.28	0.37	0.38	0.21	0.38	0.30	0.39	0.37	0.38	0.38	0.39	0.38	0.38	0.29	0.32
AF	HPV32_L1 - 1,512 bases																											

2. Click on the **Style** tab in the upper right of the MegAlign Pro window, then open the **Distance** section. Experiment with changing the **Values** and **Presentation** settings and observe how these changes affect the look of the Distance view.
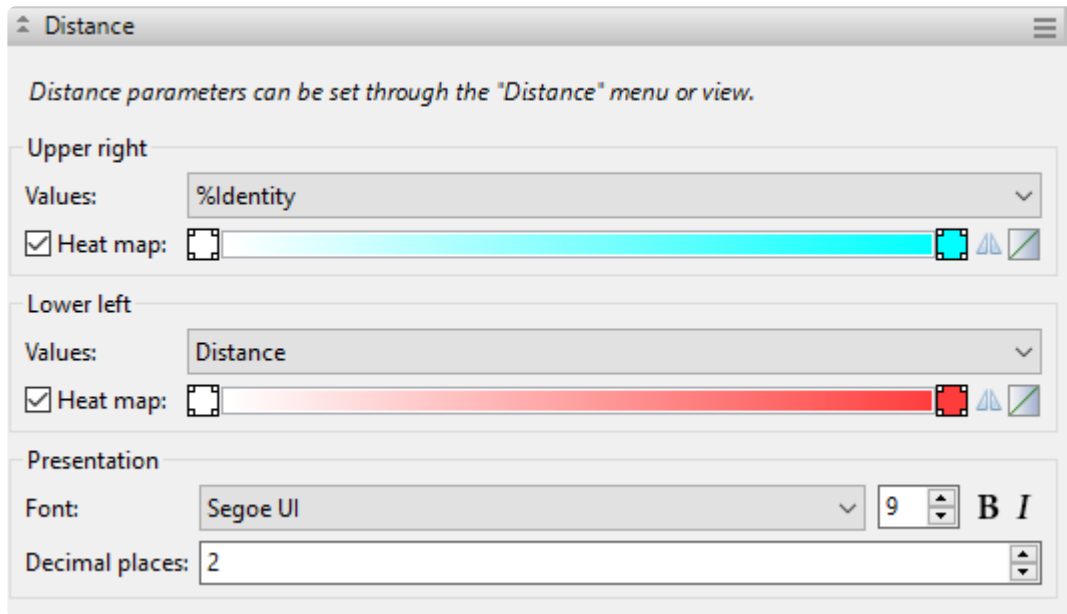
3. Open the Distance Parameters dialog by clicking on the **Change analysis parameters** tool (⚙) in the top right of the Distance view, or by choosing **Distance > Parameters**.
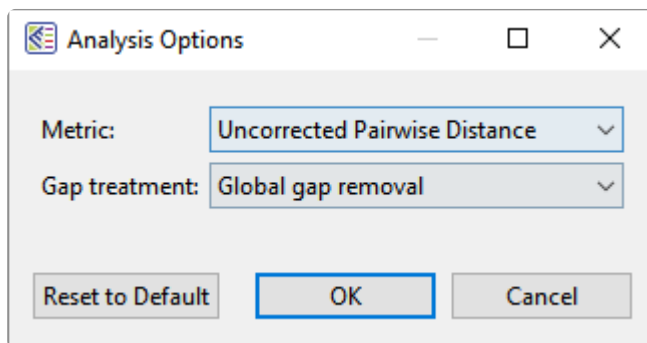
Change the distance **Metric** to **Tamura-Nei (1993)**. For greatest accuracy, we recommend using an evolutionarily adjusted metric such as Tamura-Nei (1993) rather than "Uncorrected Pairwise Distance. Press **OK** to recalculate the Distance view using the new settings.

4. Try selecting, copying, and exporting the distance table.

- To select a portion of the table, use the mouse. To select the entire table, use **Edit > Select All** or **Ctrl/Cmd+A**.

- To copy the selected information to the clipboard, use **Edit > Copy** or **Ctrl/Cmd+C**. The information can be pasted into common text or spreadsheet applications.

- To export the distance matrix in your choice of available formats, press the **Export data** tool (🗇) or choose **File > Export Data > Distance Matrix**.

Proceed to Part E: Change the appearance of the Tree view.

Part E: Change the appearance of the Tree view

The Tree view shows evolutionary relationships predicted from the multiple sequence alignment.

1. Click on the **Tree** tab at the bottom of the MegAlign Pro window. Press **Neighbor Joining: BIONJ** and then **OK** to calculate the tree.
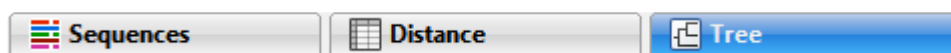
The Tree view is calculated using the current Distance metric which is currently the default value **Tamura-Nei (1993)**.
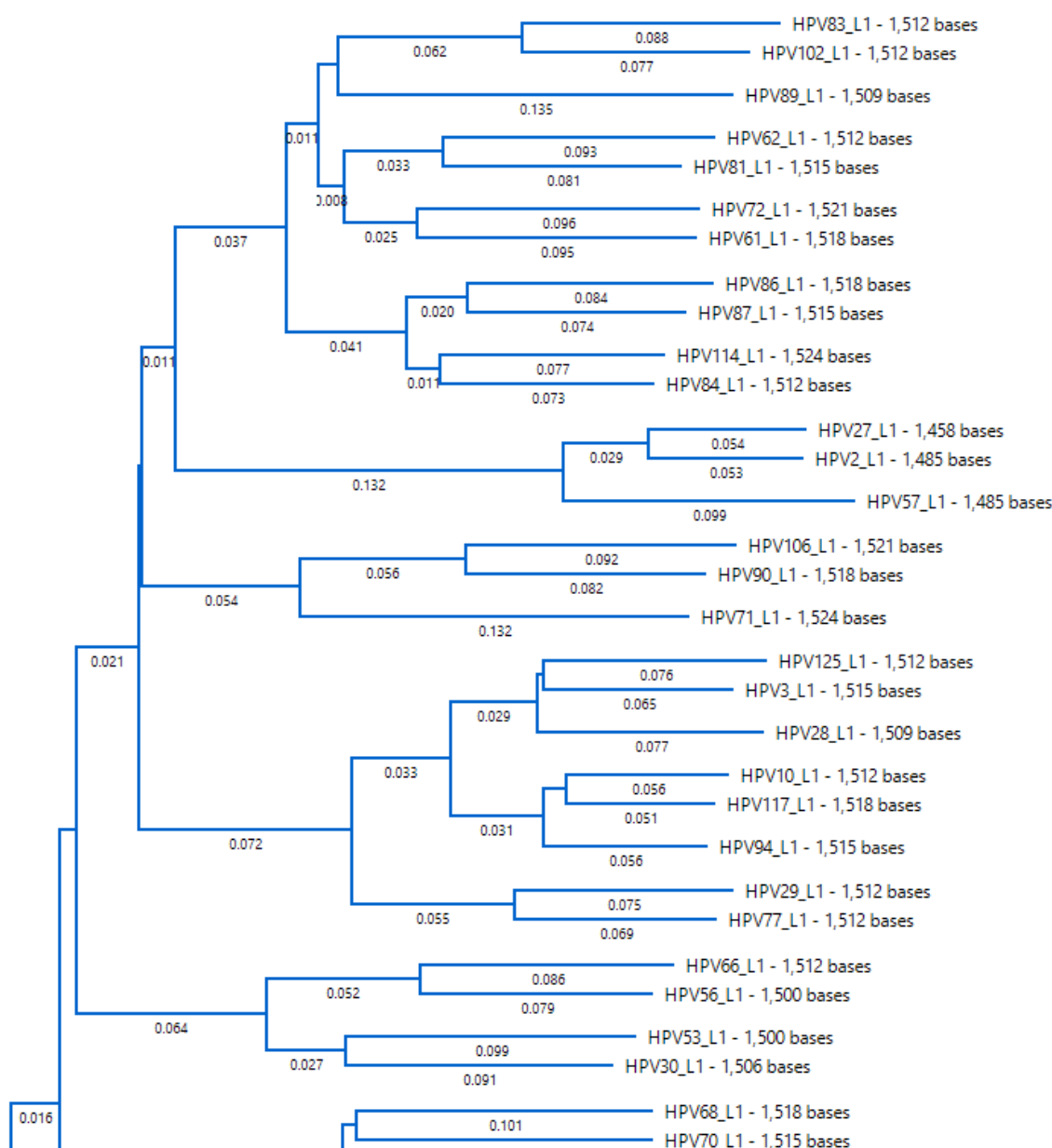
Note that the MegAlign Pro tree is always unrooted, regardless of the methods used for the multiple alignment and distance calculations.

2. Hover your mouse over a node where branches meet and click on the minus sign symbol to collapse those branches. The branches collapse and are replaced by a triangle and text showing the number of branches that were collapsed.
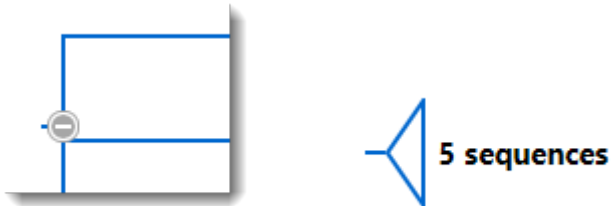
3. Select some of the uncollapsed branches by clicking on any vertical line towards the left of the tree. The selected branches are highlighted in blue.

4. Now's your chance to get creative. Change the look of the Tree view however you want using the Style panel's **Tree** section. You can use the drop-down menus, change background and outline colors for various elements (by pressing the color box icons), and/or type different numbers into the text boxes. For a description of the many tree customization options, visit the topic Tree section.
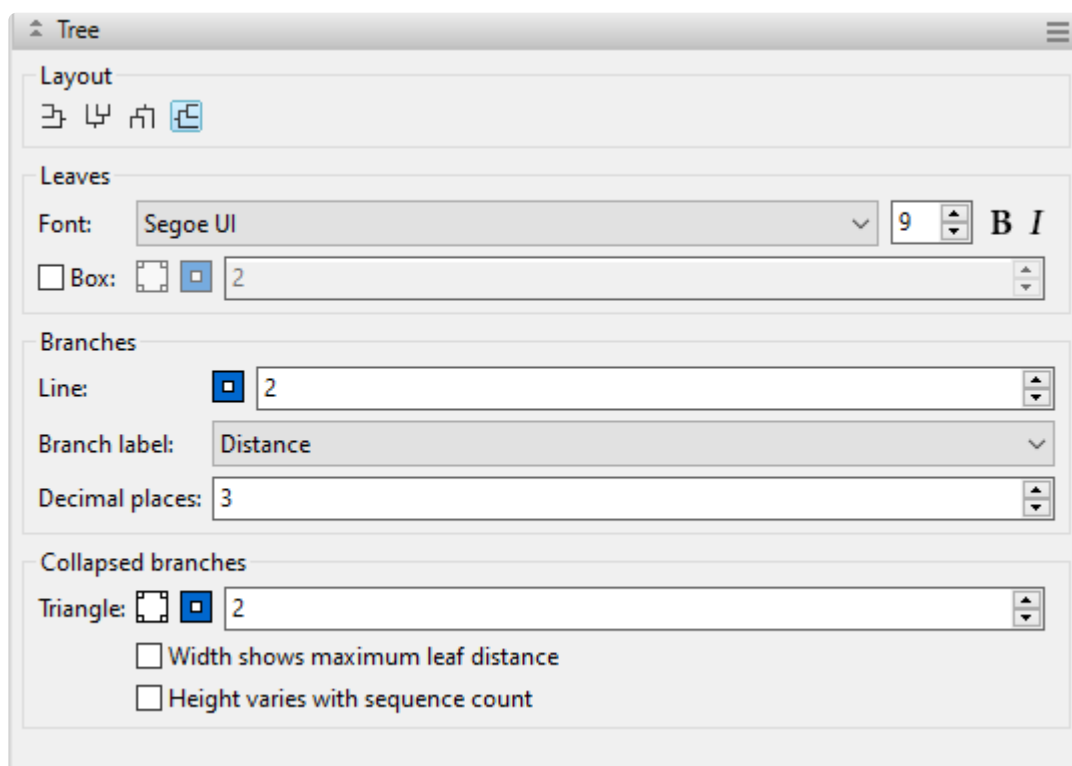
5. If desired, save the project using **File > Save**, or by clicking on the **Save** tool ( ).

This is the end of the Clustal Omega alignment tutorial.

Try it! – Perform a MUSCLE alignment with multi-segment sequences

Sequences that consist of more than one chromosome, contig, or fragment are called multi-segment files.

Suppose you want to string together the consensus sequences from all the contigs in an assembly, or chromosomes in a genome, or exons in a gene, etc. If you enter these sequences using **File > Enter Sequences** or the corresponding button ( ), you will end up aligning the parts to one another. By entering these sequences into MegAlign Pro using the procedure in this tutorial, the entire set of sequences in the multi-segment file (or set of files) will instead be treated as an end-to-end concatenation.

Begin with Part A: Add and align multi-segment files.

Part A: Add and align multi-segment files

1. If you have not yet downloaded and extracted the tutorial data, <u>click here to download it</u>. Then decompress (unzip) the file archive using the method of your choice.

2. If a MegAlign Pro project is already open, choose **File > New Alignment Project > Blank Document**. Otherwise, launch MegAlign Pro.

3. Choose **File > Add Sequences End-to-end**, or click on the **Add sequences end-to-end in project** tool (![icon]).
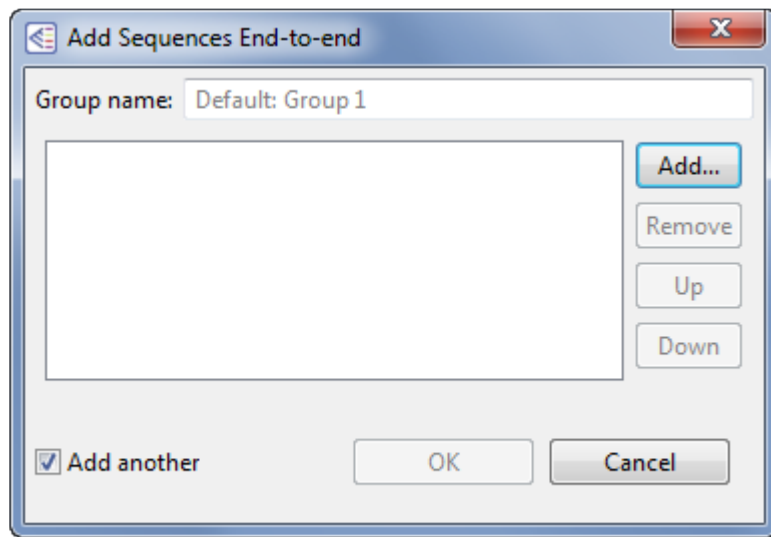
4. Click the **Add** button, then navigate to and open the tutorial data folder *Rp1 exon groups*. The nucleotide files in this folder represent the RP1 exons for several mammal species. Each exon consists of 3-4 segments.

5. If needed, click either once or twice on the file name header to ensure the sequences are in ascending alphabetical order, as in the image below.
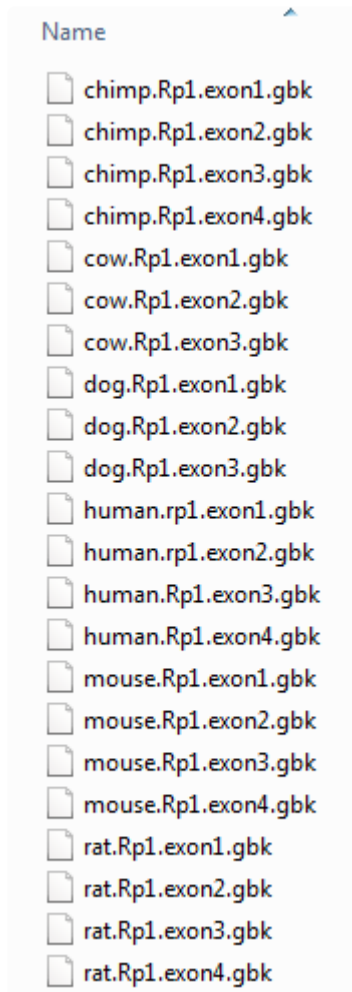
6. Use **Shift+click** to select only the four "chimp" segments, and then click **Open**. In the Add Sequences End-to-end dialog, the files will appear in ascending alphabetical order.
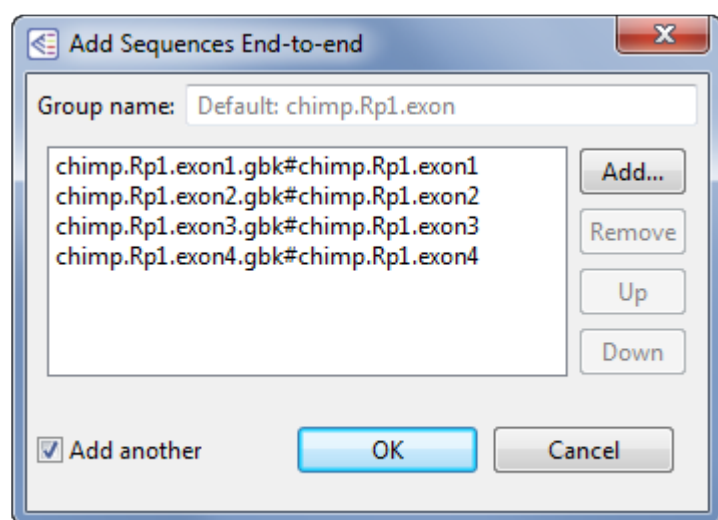
7. In the **Group name** text box, replace the default name with **Chimp**, and then click **OK**.

8. Follow steps 3-5 four more times to create separate groups for the **Cow**, **Dog**, **Human**, and **Mouse** sequences. In each case, replace the generic **Group name** with the informal species name.

9. For Group 6, add the rat sequences and replace the default name with **Rat**, but do *not* yet click **OK**. For demonstration purposes, you will deliberately be ordering these sequences incorrectly.

10. Select the "exon1" sequence by clicking on it once, and then click three times on the **Down** button until the sequence appears at the bottom of the list. Similarly, select the "exon4" sequence, the click twice on the **Up** button until the sequence appears at the top of the list.
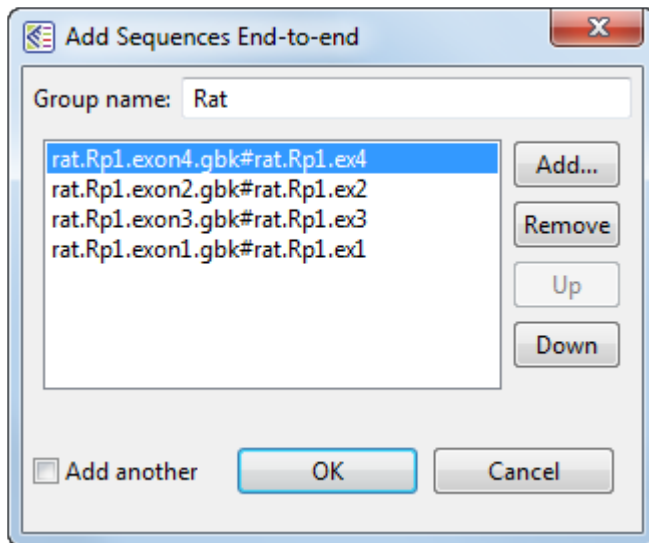
11. Remove the checkmark next to **Add another**, and then click **OK**.

Looking at the Overview, observe that each species is represented by multiple gray blocks of different lengths. Blue dividers show where one segment ends and another begins.
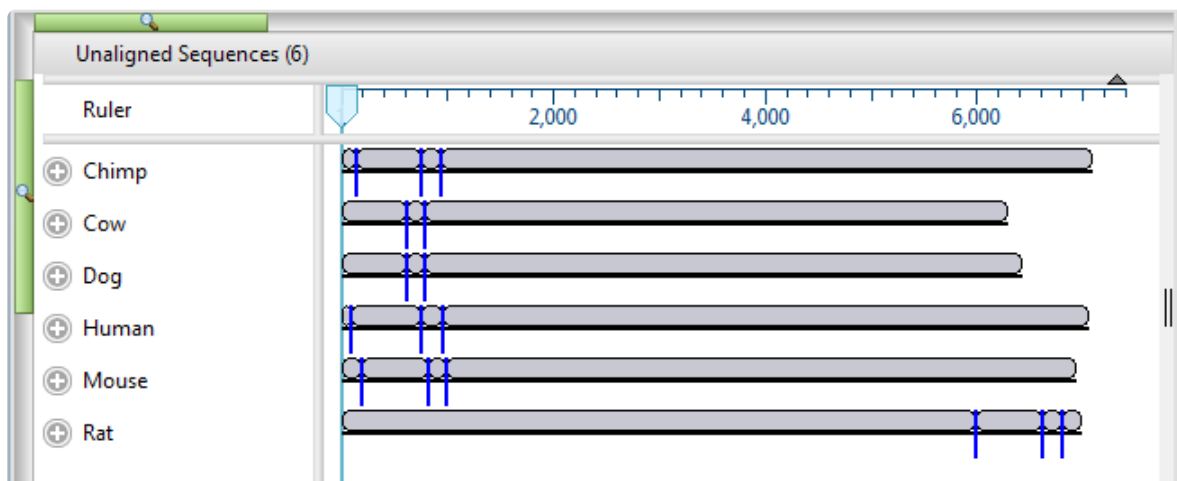
12. To make the blocks appear smaller or larger, use the horizontal and/or vertical zoom sliders ().

13. Select **Align > Align Using MUSCLE** to align the sequences using the MUSCLE algorithm. The

alignment will likely take several minutes to complete.

14. While the alignment is being performed, try using the **View > Console** command. This opens a log showing the progress of the alignment. Before proceeding, close the Console by clicking the 'x' in its tab.

15. Use **File > Save** to save the project under the name *RP1 project.msa*.

Proceed to <u>Part B: Replace a sequence in the Overview and view the phylogenetic tree</u>.

Part B: Replace a sequence in the Overview and view the phylogenetic tree

In the Overview, look for the longest segment and the longest gap region (shown in gray) in each sequence. Note that both of these are located on the left for "Rat," and on the right for all the other groups. These clues indicate that the segments for "Rat" were likely added in the wrong order. (As you recall, these were deliberately added these in the wrong order in Part A.)

In this part of the tutorial, you will remove the "Rat" group and add it again--this time in the proper order. You will then re-align the sequences.

1. Single-click on the name "Rat" to select it.

2. Right-click on "Rat" and choose **Remove Sequence**. When asked to confirm, click **OK**.

3. Click the black triangle to the right of the **Add Sequences to Project** tool and select the **Add sequences end-to-end in project** tool ( ).

4. Click the **Add** button and add the rat sequences again. This time, leave the sequences in ascending order (exon1, exon2, exon3, exon4).

5. Change the default **Group name** to **Rat**, and uncheck the **Add another** box. Click **OK**.

   Observe that the Rat sequence has been added to a new area of the view called "Unaligned Sequences." Click on the expand arrows to reveal the sequence, shown in the image below.
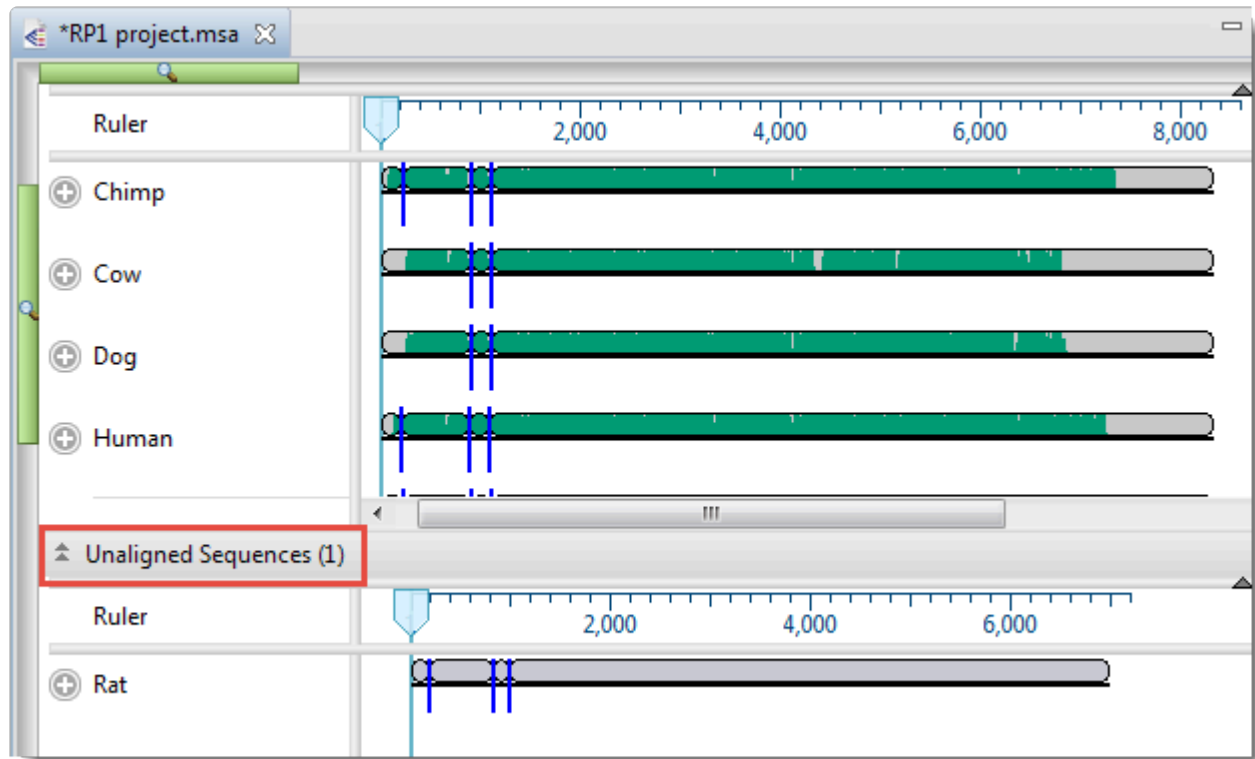
6. To perform the MUSCLE alignment with the "Rat" group in the correct order, select the name "Rat" on the left of the Overview and choose **Align > Realign with Selected Unaligned**. This command tells MegAlign Pro to perform a novel alignment that includes all the sequences in the existing alignment plus the selected unaligned sequences.

7. Leave the settings at their default values, and choose **Align**. As before, the alignment will likely take several minutes to complete.

   Once the alignment has finished, note that the gap regions in the Overview now appear more randomly distributed, and that all six organisms now have the longest block on the right.

8. Click on the **Tree** tab at the bottom of the MegAlign Pro window to display the Tree view. Choose the **Neighbor Joining:BIONJ** method. In the popup dialog, keep the default values and click **OK**. The tree appears, graphically illustrating plausible evolutionary relationships for the seven species.
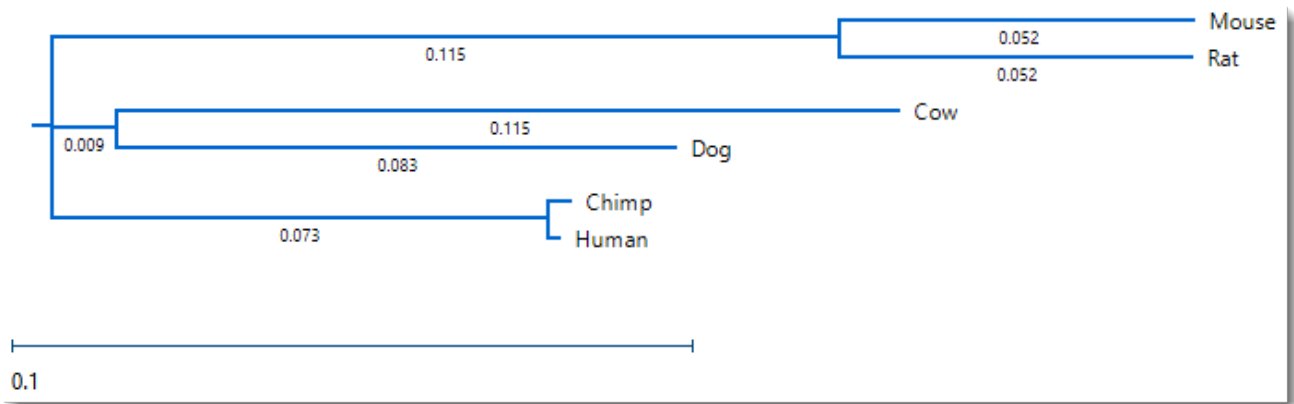
9. Change the distance metric by choosing **Distance > Parameters**. Note that the original **Metric** was **Uncorrected Pairwise Distance**. Change the selection to **Tamura-Nei (1993)** and press **OK**.

    Unlike the Try it! – Perform a Clustal Omega alignment tutorial, where changing the method caused a visible change in the tree organization, the tree retains the original structure, and has only slight changes in the branch length.
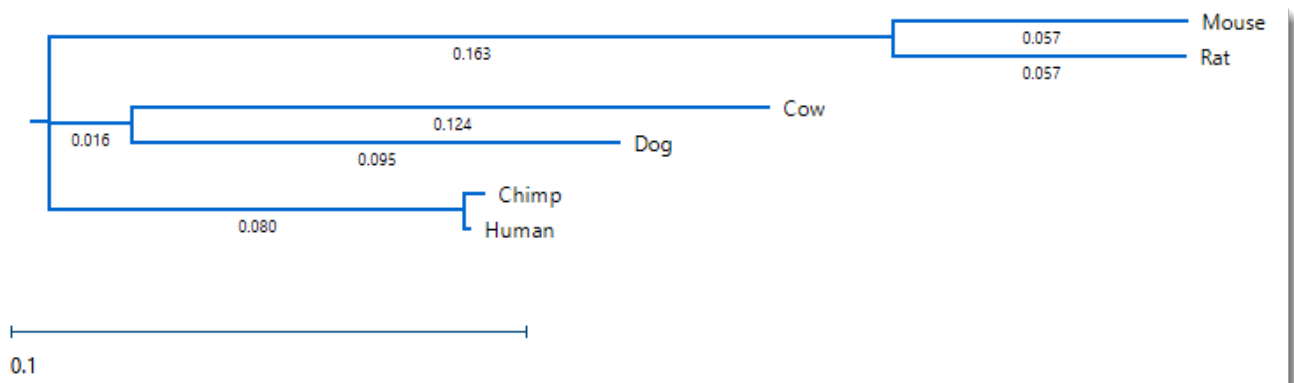
10. Save the project using **File > Save**, or by clicking on the **Save** tool ( ).

This is the end of the MUSCLE alignment tutorial.

Try it! – Perform a genomic alignment with Mauve

In this tutorial, you will be adding six nucleotide sequences to MegAlign Pro using drag & drop, and then aligning them with the Progressive Mauve algorithm. Drag and drop is a feature available in all Lasergene applications, and is useful for entering multiple sequences.

Begin with Part A: Add genomic sequences using drag & drop.

Part A: Add genomic sequences using drag & drop

1. If you have not yet downloaded and extracted the tutorial data, click here to download it. Then decompress (unzip) the file archive using the method of your choice.

2. If a MegAlign Pro project is already open, choose **File > New Alignment Project**. Otherwise, launch MegAlign Pro.

3. **Outside** MegAlign Pro, use the Windows Explorer or Macintosh Finder to locate and open the tutorial data folder *H. pylori genomes*. Arrange and size the file explorer and the MegAlign Pro window so that you can see both simultaneously.

4. In the file explorer, use **Shift+click** to select the four files.

5. Use the mouse to drag & drop the selected files onto any "white space" in the Overview or Sequences view.

   The sequences appear in the Overview as solid gray blocks of unequal length, indicating that the sequences have not yet been aligned.

6. Use your mouse to drag and drop the sequence names (on the left side of the window) so they appear in the same order as in the image below. This will ensure that your block colors and positions match those referenced in later sections of this tutorial.
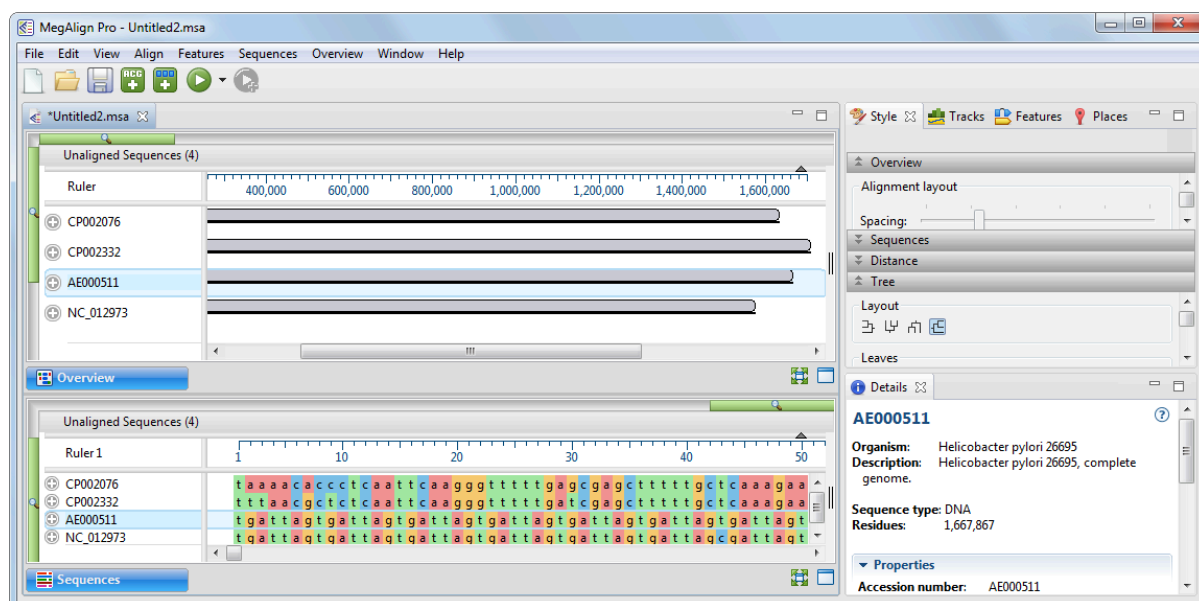
7. Use **File > Save** to save the project under the name *H. pylori project.msa*.

Proceed to Part B: Perform a Mauve alignment using modified parameters.

Part B: Perform a Mauve alignment using modified parameters

In this part of the tutorial, you will explore how to view and change parameters prior to performing a multiple sequence alignment.

1. Choose **Align > Align Using Mauve** and align the sequences using the default parameters. Because this is a genomic alignment, the alignment will take longer than those done in the first two tutorials (approximately 5-10 minutes).

   After genomic sequence alignment with Mauve, MegAlign Pro displays genomes in the Overview as Locally Co-linear Blocks, usually abbreviated to "blocks" or "LCBs." Blocks represent multiple-aligned sections of at least two of the current set of genomic sequences. Discontinuity in numbering is displayed as a break between blocks. Connecting lines show the correspondence between blocks. Blocks hanging below the black line, signify that the hanging blocks are on the opposite strand compared to the other blocks.
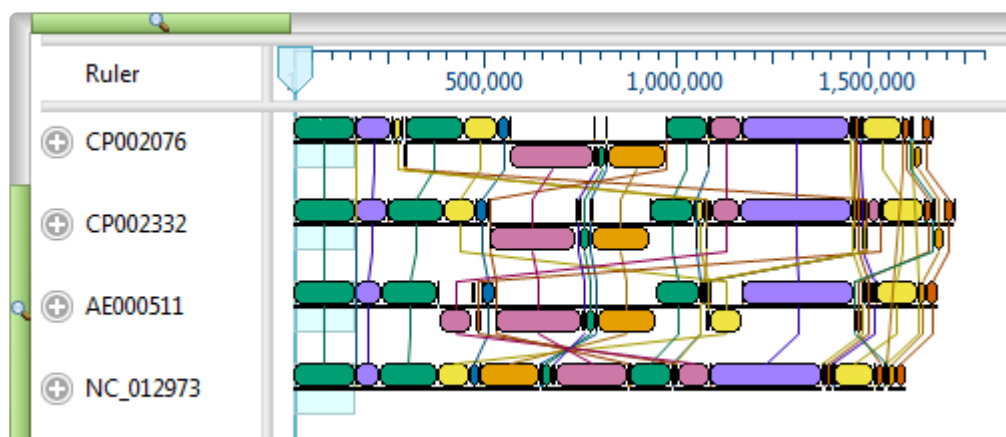
   The default settings used in creating this alignment are suitable for most projects, including this one. However, you will make a change to Mauve's **Seed weight** parameter in order to see how this affects the alignment results.

   In Mauve, **Seed weight** is the minimum weight of the seed pattern used during the first pass of a multiple alignment. When aligning divergent genomes, lower seed weights may provide better sensitivity. However, if the value is set too low, the result will be reduced sensitivity. In the following steps of this tutorial, you will lower the **Seed weight** slightly, from the 15 (the default) to 13.

2. Choose **Align > Align with Options**. This dialog allows you to change methods and/or parameters. Note that **Mauve** is already selected in the menu at the top of the dialog, since it is the last method used.

3. By default, the **Default seed weight** box is checked, meaning that the optimal seed weight (number of

matches required per k-tuple) is calculated automatically. Uncheck the box and either use the slider to change the **Seed weight** to 13, or type **13** in the **Seed weight** text box.
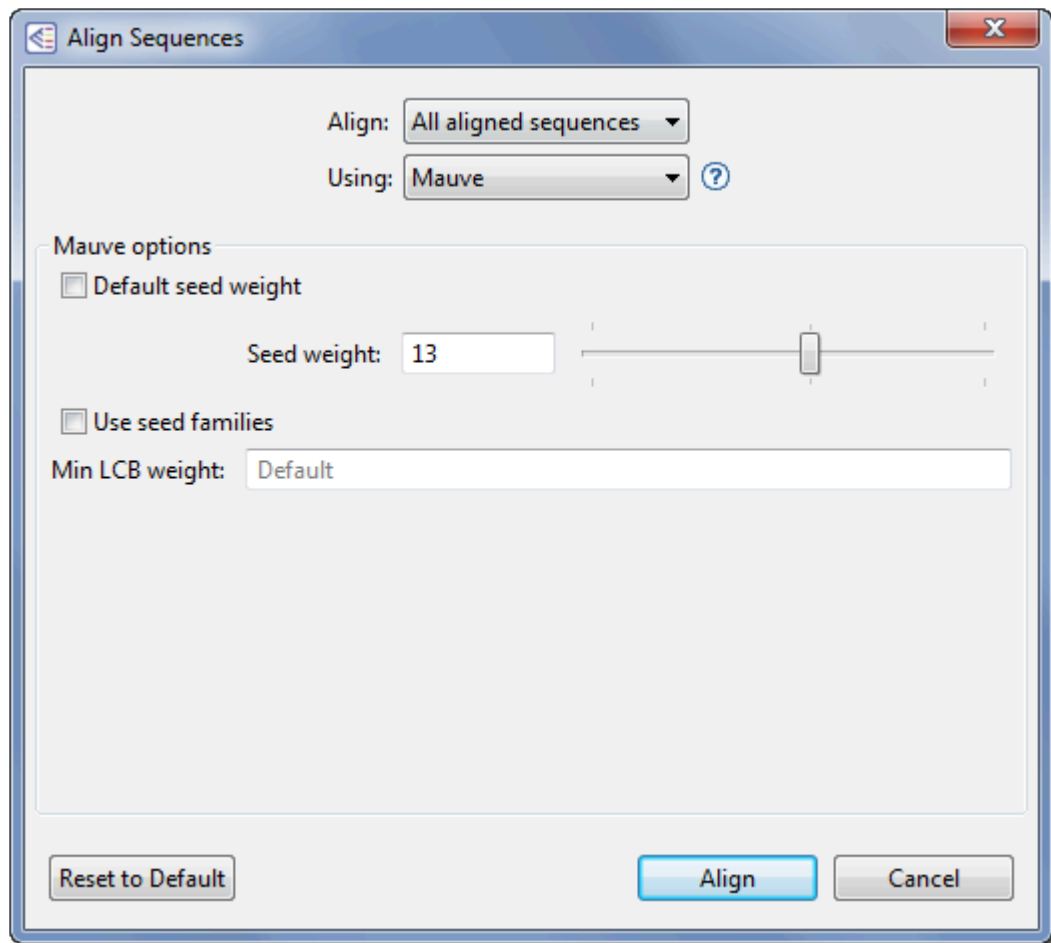
4. Choose **Align** to perform the Mauve alignment again using the selected parameters.

Proceed to .

Part C: View a genomic alignment in the Overview

Once the alignment has finished, notice how the Overview has changed in response to the parameter change, including changes to the colors of the blocks, and differences in which blocks are hanging below the line (i.e., on the opposite strand).

1. Locate the medium-width block whose right edge is close to the 1,000,000 mark on the ruler. The block will not necessarily be gold, as shown here, but the location should match the location of checkmarks in the image below. Double-click on the block in any row to make it the active block.
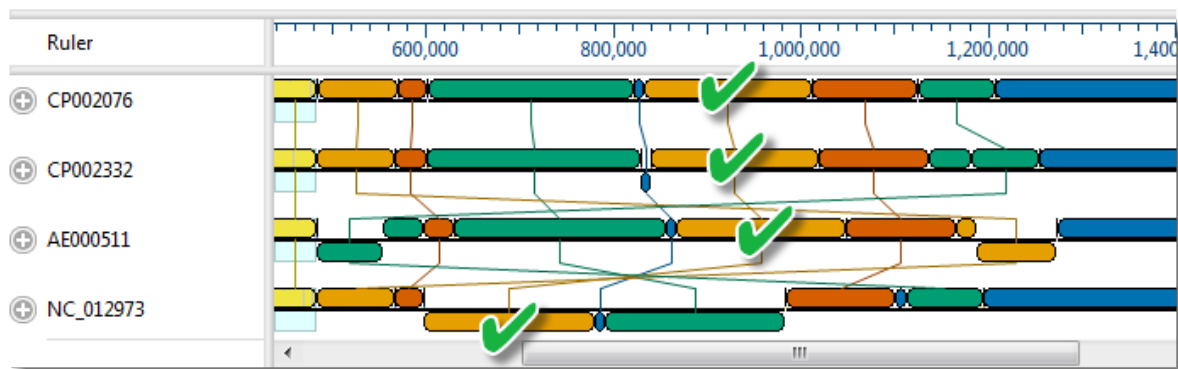
Observe that the active blocks are now in a single, highlighted column, with each block having equal width. The viewport is located on the left-most end of the blocks.
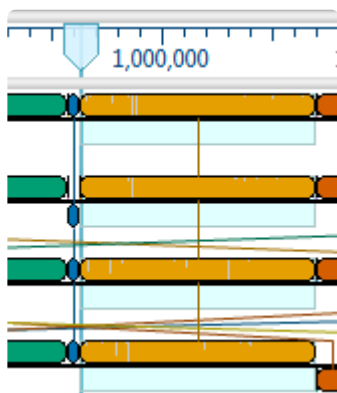
Notice how genome *NC_012973* flipped in orientation when the new block was made active. To allow this genome's active block to match the orientation of the active blocks in other rows, MegAlign Pro automatically reverse-complemented all the blocks on the opposite strand.

2. In the Overview, slide the viewport to the right edge of the active blocks. You can use the horizontal green slider above the view to expand this area so you can see it more clearly.
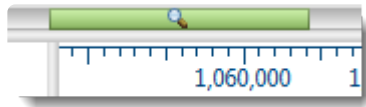
3. Use either set of scroll arrows to center the Sequences view on the right edge of the active block, as shown below. (The exact location of the edge may differ from that shown in the image). Observe that residues outside and to the right of the active block are shown in a dimmed/dull version of the color scheme.
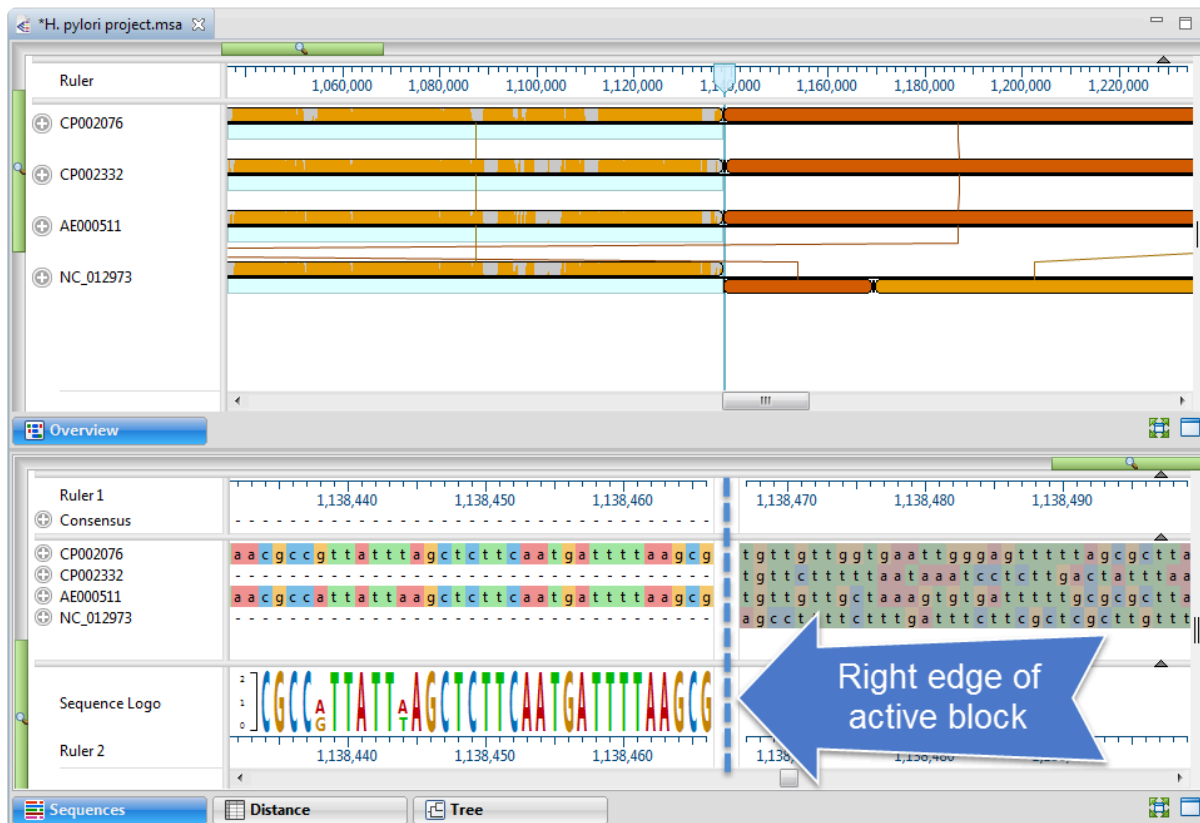
4. Select **View > Style > Overview** to open the Overview section. Experiment with changing the parameters.
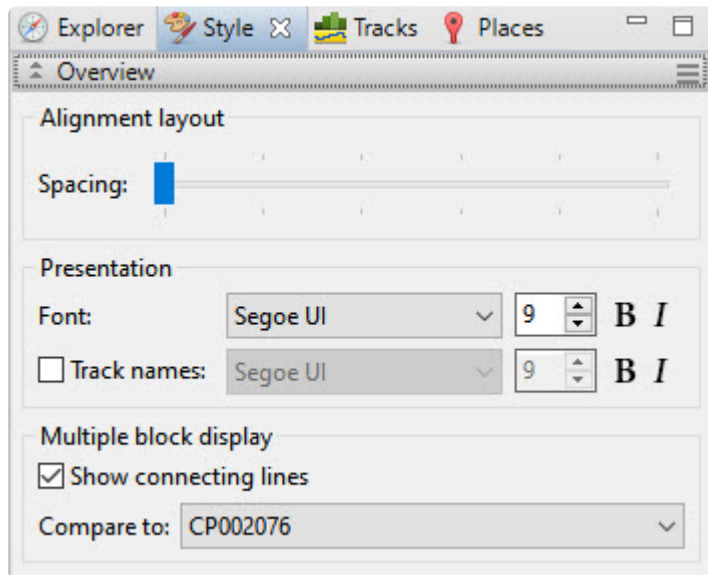
- To increase/decrease the amount of white space between rows of blocks, use the **Spacing** slider. Unlike the zoom in/out sliders, this setting affects only the white space between the blocks, and does not affect the sizes of the blocks.

- Change the font and font size. If you wish to change the track names font, check the box to enable changing the font and size.

- To toggle between displaying or hiding the lines connecting blocks from one row to another, check or uncheck the **Show connecting lines** checkbox.

- If you select a different **Compare to** sequence, change it back to CP002076 before you are finished.
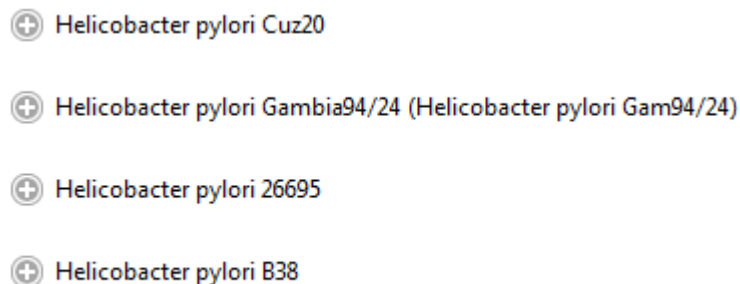
Proceed to Part D: View the phylogenetic tree.

Part D: View the phylogenetic tree

> ✳ **Note:** The images in this section show examples of how the trees might appear. Your trees may have different arrangements or different branch lengths from the ones shown.

1. The current names of each sequence originated from the accession numbers that were embedded in the input files. Before viewing the phylogenetic tree, you can change these names to show the organism from which each sequence originated:

   a. Use **Ctrl+A** or **Cmd+A** to select all of the sequence names.

   b. Right-click on the highlighted group and select **Rename Sequences with Fields**, or choose **Edit > Rename Sequences with Fields** from the main menu.

   c. Select **Default name** from the right pane and click the left arrow (blue) to remove it. Then select **Organism** from the left pane and click the right arrow to add it. Click **OK**.

   The sequence labels now reflect the species names:

   ⊕ Helicobacter pylori Cuz20

   ⊕ Helicobacter pylori Gambia94/24 (Helicobacter pylori Gam94/24)

   ⊕ Helicobacter pylori 26695

   ⊕ Helicobacter pylori B38

2. To display the Tree View for the active blocks, click on the Tree tab at the bottom of the MegAlign Pro window.

   ▤ **Sequences**   ▦ **Distance**   ⊞ **Tree**
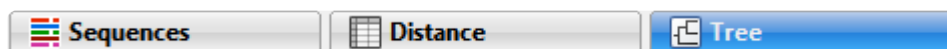
   The contents of the Tree view (and also the Distance view) are specific to the currently selected block only.
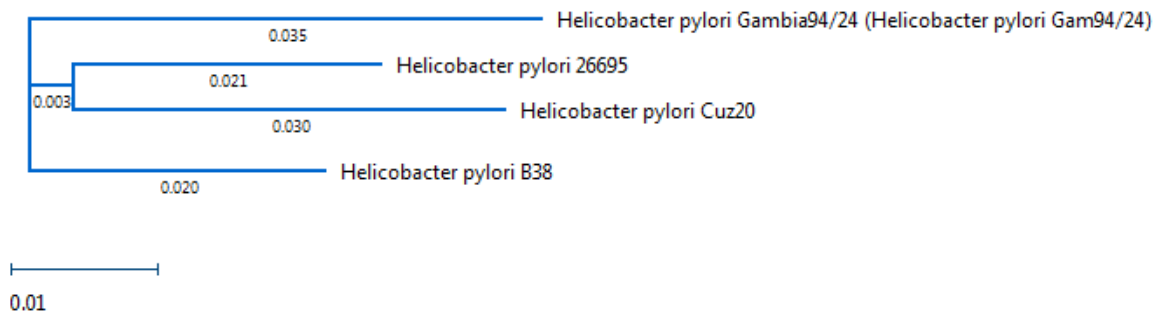
3. In the Overview, double-click on different blocks in the Overview and see how the branch lengths, or (sometimes) the tree itself, change.
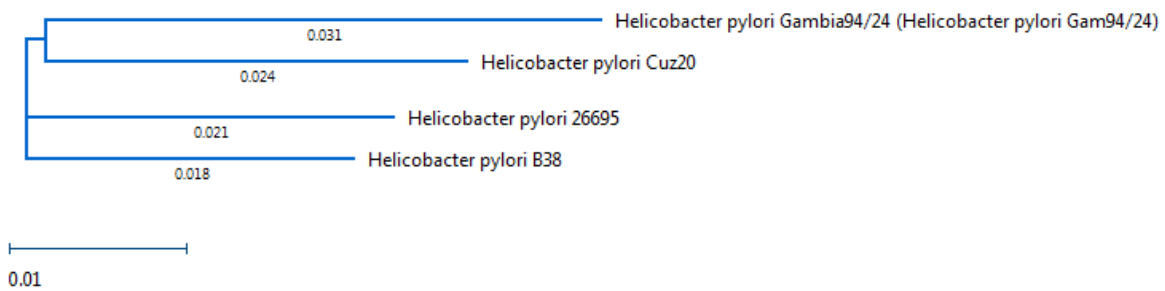
4. To order sequences in the other views so they match the tree, select **Tree > Order sequences like tree**.

5. If desire, save the project using **File > Save**, or by clicking on the **Save** tool ( ).

This is the end of the Mauve alignment tutorial.

Visually Compare Other Sequences to One Sequence

You can visually compare sequences in the alignment to one of the aligned sequences by using settings in MegAlign Pro's Style panel to highlight differences or matches.

By default, the comparison is against the reference sequence, if you have specified one. Otherwise, the default is to compare sequences to the consensus. However, you can choose any sequence as the comparison sequence, and you can switch between comparisons without affecting subsequent alignments, the variant table, or the positions of the sequences in the Sequences view.

When might you want to do a comparison? Imagine that you have aligned a thousand CoV-2 genomes, and computed variants. All the viral strains are similar, and none of them is "special" enough to merit use as the reference for alignment or variant computation. In this case, a visual comparison is probably all you need.

To visually compare sequences in the Sequences view:

The following comparison applies to any alignment other than Mauve. Results are displayed in the Sequences view.

1. Add sequences to the project.

2. If desired, specify a reference by right-clicking on a sequence and choosing **Use as Reference**.

3. Align the sequences using any algorithm. If you use **Align > Align with Options** and choose MAFFT, the Align Sequences dialog will give you a second chance to specify a reference.

4. To make residues of interest stand out in the Sequences view, use the Multiple alignment section of the Style panel (**View > Style > Multiple Alignment**).
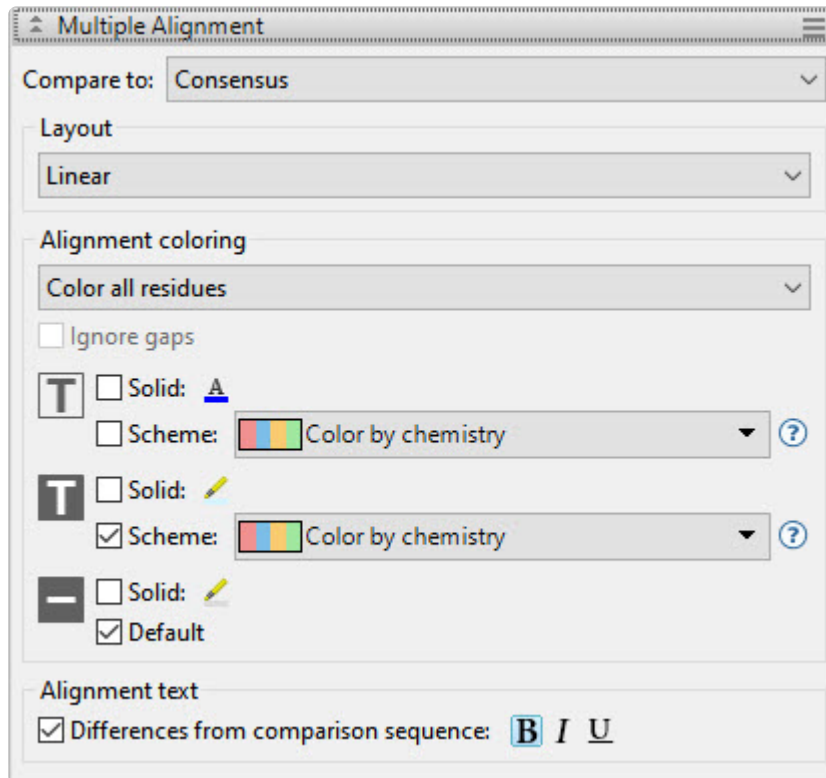
- If a different comparison sequence is desired, use the **Compare to** drop-down menu.

- Use the **Alignment coloring** drop-down menu to choose what to color and display. For example, you could choose to **Color only differences from comparison sequence**, **Color only matches to comparison sequence**, or **Show only differences from comparison sequence**.

- If desired, check the box **Differences from comparison sequence** and choose one or more of the font-style options (**Bold**, *Italic*, Underline). This accentuates mismatches even more, even if you've coloring conserved or not conserved residues.

In the Sequences view, similarities/differences will appear as you specified in the Multiple alignment section.

To change which sequence is compared against in the Sequences view:

You can either:

- Choose a different comparison sequence, using the **Compare to** drop-down menu in the Multiple alignment section of the Style panel (**View > Style > Multiple Alignment**).

- Set a reference sequence by right-clicking a sequence in a view, and choosing **Use as Reference**.

Choosing a different comparison sequence affects only the Sequences view. It doesn't cause sequences to

be moved, impact the variant table, or affect subsequent alignments.

By contrast, specifying a reference sequence automatically sets it as the comparison sequence for the Sequences view. It also moves that sequence to the header of the view, making it easier to compare against any other sequence in the body of the view. The variant table will change to show either the variants based on the new reference, or a button for computing those variants. In addition, if you align the sequences in the project again and use MAFFT, the reference will be used for that alignment.

In the example below (image), the alignment was done with MAFFT using "Rat" as the reference. "Red Bread Mold" was chosen as the comparison sequence and **Show only differences from comparison sequence** was selected in the Style panel. Note that "Rat" continues to be used as the reference sequence (blue box). With this type of automatic coloring selected, it is easy to see that "Red Bread Mold" is the active comparison sequence (red), since it appears in normal black text throughout. In all other sequences, including the reference sequence, only differences from the comparison sequence are displayed (yellow).
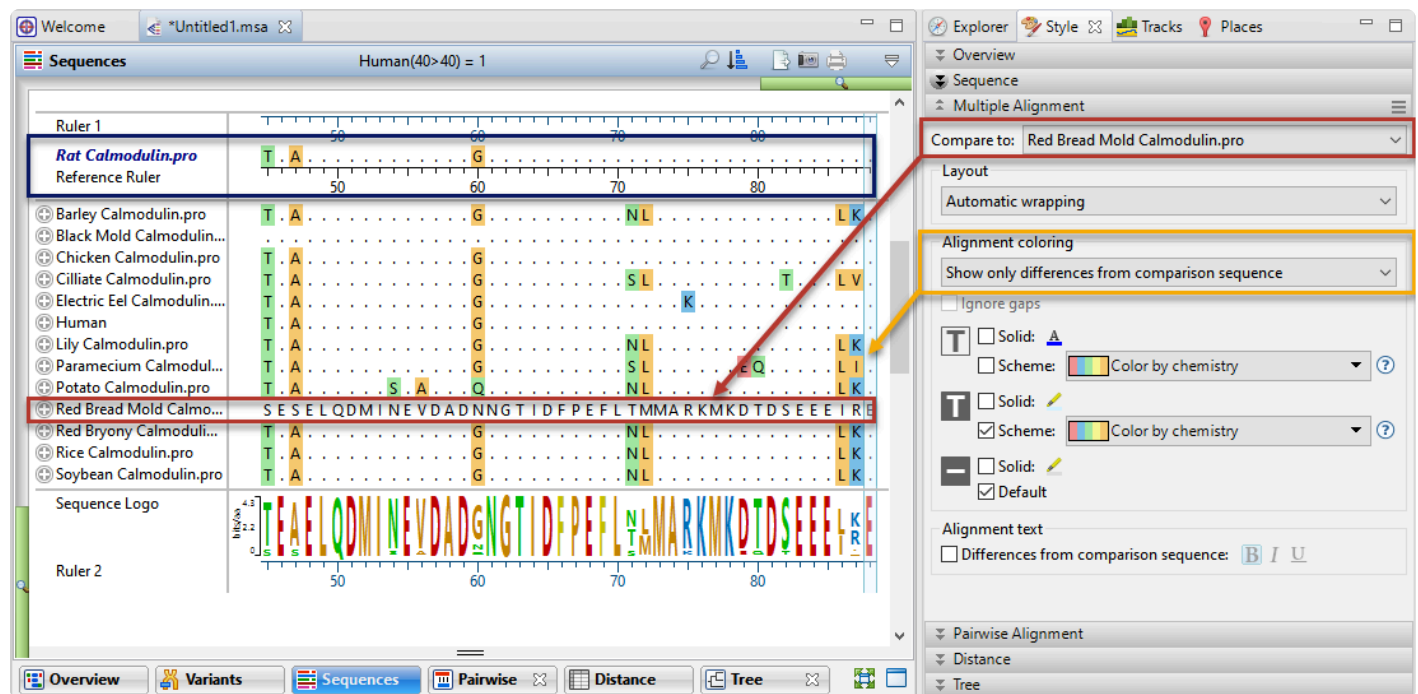
To visually compare sequences in the Overview (Mauve alignments only):

For Mauve alignments, the Overview may show multiple aligned blocks on either forward or reverse strands. The blocks for the Overview comparison sequence are all drawn on the forward strand. Blocks for other sequences are drawn on the reverse strand if they align to the reverse-complement of the corresponding block in the comparison sequence.

You can change the Overview comparison sequence using the procedure below. This procedure is independent of any comparison performed in the Sequences view using the steps in the first section of this topic.

1. Add sequences to the project.

2. Align the sequences using Mauve.

3. Choose which sequence should show all its blocks on the forward strand, using the **Compare to** drop-down menu in the Overview section of the Style panel (**View > Style > Overview**).

Perform a Pairwise Alignment

MegAlign Pro supports local, global, semi-global, and chromosome-based pairwise alignment methods, as well as a specialized proprietary method for aligning cDNAs or other short sequences to a chromosome. Click here for descriptions and a comparison of the different methods.

How to perform a pairwise alignment:

Typical pairwise alignments involving sequences of similar lengths can only be performed when two sequences, and *only* two sequences, have been selected. A common workflow is to first perform a multiple alignment on an entire group of sequences. From the resulting Tree view, two closely related sequences can then be further analyzed by selecting them and performing a pairwise alignment. To learn how to perform this general type of pairwise alignment, see Pairwise alignment for sequences with similar lengths.

As of the Lasergene 17.5 release (July 2023), a new pairwise alignment procedure is available and mandatory for aligning short sequences like cDNAs to a chromosome or other extremely long sequence. For instructions, see Pairwise alignment of a short sequence to a chromosome

Pairwise alignment tutorials:

The following tutorials pertain to "typical" pairwise alignments (see the section above) and use data that can be downloaded from the DNASTAR website:

- Try It! – Follow a multiple alignment with Global pairwise alignments
- Try It! – Align transcripts to genes using Local and Global pairwise alignments
- Try It! – Use a Local pairwise alignment to find a gene within a genome

Pairwise alignment for sequences with similar lengths

The following instructions describe how to align sequences with similar lengths (i.e., within one or two orders of magnitude). DNASTAR has used these pairwise algorithms in this workflow to successfully align protein sequences up to 35,000 bases in length.

> ✳ **Note:** If you need to align cDNAs or other short sequences to a chromosome-length sequence, see Pairwise alignment of a short sequence to a chromosome.

To align two similar-length sequences pairwise:

1. Add sequences to the project.

2. Decide which two sequences you want to align. A typical workflow is to first perform a multiple alignment, then build a phylogenetic tree, then choose two adjacent sequences to compare.

3. From any view, select the two sequences you wish to align. The sequences can vary in length, but both must belong to the same category: DNA/RNA or protein. If one sequence is significantly longer than the other, use drag & drop in the Sequences view or Overview to organize them such that the longer sequence is above the shorter sequence.

4. Choose **Align > Pairwise** or right-click on the selection and choose **Align Pairwise** to launch the Align Pairwise dialog.
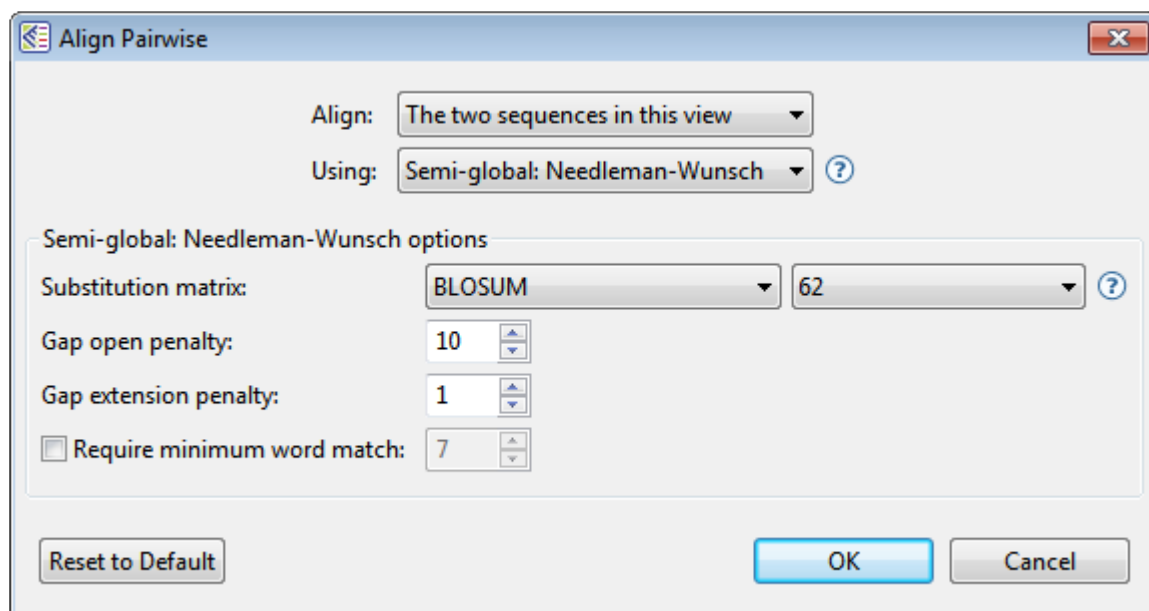
5. Choose the desired settings.

   • **Align** – Use this drop-down menu to specify which sequences to align.

   • **Using** – Use this drop-down menu to choose the desired pairwise alignment method: **Local**, **Global**, or **Semi-Global**.

   • **Substitution matrix** – Use this drop-down menu to choose the substitution matrix. A substitution matrix describes the rate at which a nucleotide or amino acid changes to another nucleotide or amino acid over time. Different matrices are available for nucleotide vs. protein sequences, as shown in table below these instructions. Also see the "Notes" section below that for additional information.

   • **Gap open penalty** – Specify the amount that should be deducted from the alignment score for each gap in the alignment. Gaps of different sizes carry the same penalty. Default is 10.

   • **Gap extension penalty** – Specify the amount that should be deducted from the alignment score after first multiplying it by the length of gaps. Longer gaps have a greater penalty than shorter gaps. Default is 1.

   • **Require minimum word match** – If you want to specify the length of the smallest perfect match of contiguous bases/residues to consider in building an alignment, check the box and enter a value. The default is for the box to be unchecked. If checked, the default value is 7.

6. Press **OK** to begin the alignment.

   During the alignment, MegAlign Pro displays a progress window. In most cases, this will appear and disappear too suddenly to notice it. In the cases of longer alignments, you can interrupt the alignment, if necessary, by clicking its **Cancel** button or view a console window showing the start time and progress of the alignment by clicking its **Show Console** button.

   If an alignment finishes successfully, a Pairwise view opens. If an alignment fails, you will receive a message with recommendations on how to obtain a successful alignment (e.g., by modifying options or choosing a different alignment method).

7. (optional) If the Console is not already open, and you wish to view alignment statistics and other information there, select **View > Console**.

✱ **Note:** Pairwise alignments are not saved when you save or close a MegAlign Pro project.

Substitution matrix descriptions and options:

Available for Sequence Type	Matrix	Description	Secondary Option
Nucleotide	NUC44	DNASTAR's modified version of NCBI's NUC.4.4 algorithm, the modification being that U is treated as a synonym of T. In NUC44, exact matches, and T:U matches score as 5, and mismatches between unambiguous bases [G A T C U] score as -4. Matches between bases and ambiguous symbols [S W R Y K M B V H D N] have intermediate scores. A base versus a 2-way ambiguous category [R Y W S K M] to which it belongs scores as +1, and a mismatch to a 2-way group to which it doesn't belong scores as -4. **Example:** C is in [S R M] but not in [W Y K] . The 3-way groupings are [B V H D] where C is in all but D (which means not C). Therefore, C vs [B V H] scores as -1 while C vs [D] scores as -4.	N/A
Protein	BLOSUM	([Henikoff & Henikoff, 1992](#)). These matrices are ideal for carrying out similarity searches.	Available matrices range from 30-100, and are provided in increments of 5 and 62. Choose larger numbers for less divergent sequences.
Protein	GONNET	Derived from PAM matrices ([Dayhoff et al.](#), 1978) but more sensitive, and based on a much larger data set.	(Unchangeable default of 250)
Protein	IDENTITY	Scores two identical amino acids as 1, and anything else as -10,000.	N/A
Protein	MATCH	Scores two identical amino acids as 1, and anything else as -1.	N/A
Protein	PAM	([Dayhoff et al.](#), 1978). Widely used since the late 1970s.	Available matrices range from 10-500, and are provided in increments of 10. Choose larger numbers for more

			divergent sequences.
Protein	VTML	Derived from PAM matrices ([Dayhoff et al.](), 1978) by [Müller T et al.]() (2002), .	Available matrices range from 10-500, and are provided in increments of 10.

Notes:

- BLOSUM, PAM, GONNET, IDENTITY, and MATCH are part of NCBI's BLAST distribution. For more information, see [NCBI's substitution matrix page]().

- The PAM, GONNET and VTML numbers are based on the presumed millions of years of divergence.

- In BLOSUM, the matrix number is proportional to the presumed degree of similarity. Therefore, BLOSUM100 would be the preferred matrix for near-identical sequences.

- VTLM and GONNET are considered to be updated versions of PAM250.

- In BLOSUM, PAM, and GONNET, match/mismatch scores vary with the series number. Also exact matches vary with the particular amino acid. For example, BLOSUM30 scores W:W as 20 and S:S as 4. BLOSUM100 scores these as 17 and 9, respectively.

Pairwise alignment of a short sequence to a chromosome

With the release of Lasergene 17.5 (July 2023), MegAlign Pro features a unique algorithm—adapted from one used in SeqMan NGen—that supports pairwise alignment of a short sequence (typically a cDNA) to a chromosome-length sequence. This type of alignment is impossible for other algorithms because the exons are miniscule in comparison to the large intron gaps, which range in size from 10,000 to 1 million bp. Using this new aligner, MegAlign Pro can simultaneously perform 25 pairwise alignments.
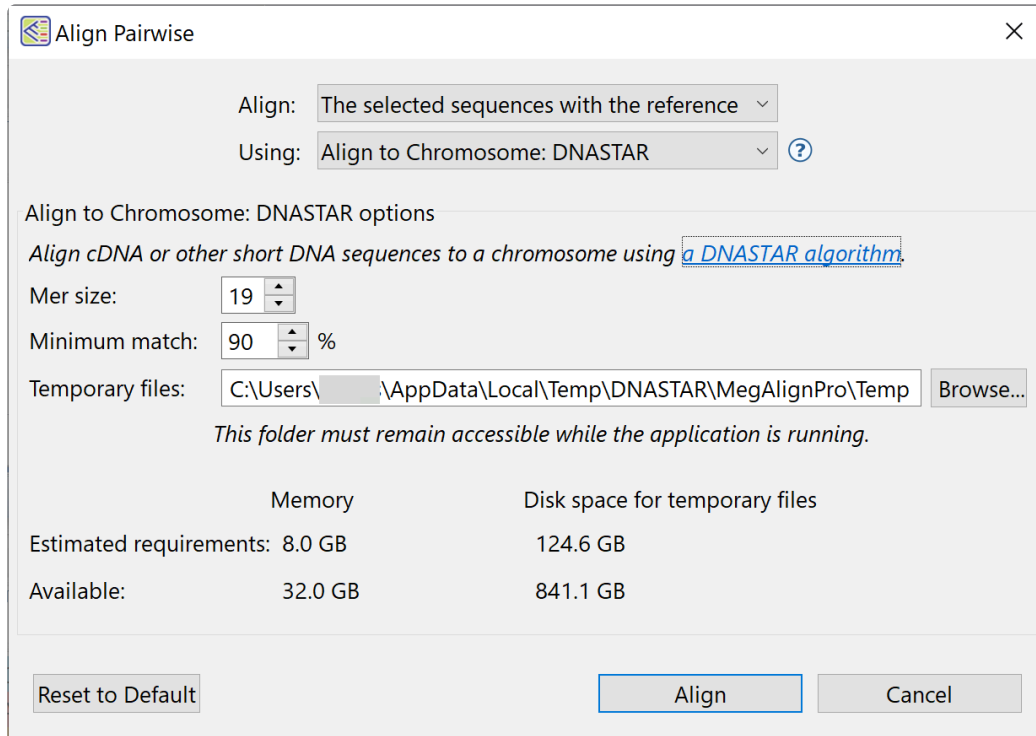
> ✳ **Notes:** If you need to do a more general pairwise alignment, see the topic Pairwise alignment for sequences with similar lengths. If you need to align two chromosome-length sequences together, use SeqMan NGen rather than MegAlign Pro.

To align a short sequence to a chromosome:

1. Prior to starting, make sure that your short sequences consist of extracted CDS features. You can create this type of sequence by using the Extract Features as Sequences template in SeqNinja, an application that is included with each Lasergene package.

2. Start a new project in MegAlign Pro and add one or more CDS sequences and one chromosome sequence to the project.

   In the Overview, the short sequence bars may barely be visible when compared to the long sequence. Upon seeing the difference in sequence lengths, MegAlign Pro automatically and correctly assumes that the long sequence is the reference and that you will be performing one or more pairwise alignments of short sequences to that reference. You do not need to manually specify the reference sequence.

3. Choose **Align > Align Pairwise with Reference** or click the **Align** button (▶). If you have followed the steps above, the Pairwise Options dialog will open and the **Using** menu will say **Align to Chromosome: DNASTAR**.

In addition, you will see that the reference sequence has been marked in bold/italic blue font in the [Sequences view](#).

> ❋ **Note:** If the Pairwise Options dialog does not open or if MegAlign Pro gives an error message or asks if you want to align with Mauve, this indicates that the short sequences are not extracted CDS features. Press **Cancel** (if that option is offered). Then start again from step 1, making sure that all short sequences consist of extracted CDS features.

4. Do not select a different option from the **Using** menu. However, you can adjust other parameters in this dialog if desired.

- **Mer size** – The minimum length of a mer (overlapping region of a fragment read), in bases, required to be considered a match when aligning the sequences. Default is 19.

- **Minimum match** – Specifies the minimum percentage of matches in an overlap required to join two sequences in the same contig. Default is 90%.

- **Temporary files** – The location where temporary files will be stored. To specify or change the location, use the **Browse** button or open your computer's file explorer and drag and drop the directory in the **Temporary files* text box. MegAlign Pro will remember and use the temporary file location for future alignments. Note: Choose a file location on a main drive; do not choose the desktop.

- **Estimated requirements** – Displays the estimated memory and disk space required for the

alignment to run without failure.

- **Available** – Displays the amount of memory and free disk space on your local computer. A yellow warning triangle will appear if you do not have enough available space to meet the estimated requirements.

5. Press **Align**. Unless you selected specific sequences prior to step 4, a pairwise alignment will be made between each short sequence in the project and the long reference sequence. Any sequences that are in the incorrect orientation will be reverse-complemented automatically.

At the conclusion of the alignment process, MegAlign Pro creates a separate Pairwise view for each aligned pair. An empty Pairwise view is also created in case you wish to start a new comparison. To navigate between alignments, click on the **Pairwise view** tab and choose the desired alignment to display.

Each pairwise view shows only the part of the chromosome that the CDS sequence matched. Also, any sequence that was reverse-complemented is designated by a small red left-facing arrow next to the sequence name.

As with all pairwise views in MegAlign Pro, note that the information contained in them is transient. If you close the view, you will need to repeat the steps above in order to recalculate and view the results.

Choosing a pairwise alignment method

MegAlign Pro has four [pairwise alignment](#) algorithms.

Align to Chromosome: DNASTAR is used to perform individual pairwise alignments of 1 to 25 short sequences (typically cDNA sequences) to a chromosome-length reference sequence. This method was released with Lasergene 17.5 in July 2023, and is adapted from a proprietary algorithm used in SeqMan NGen. The new MegAlign Pro algorithm can align sequences in which one sequence is thousands of times longer than the other. The short sequences must consist of extracted CDS features. You can create this type of sequence by using the [Extract Features as Sequences](#) template in SeqNinja, an application that is included with each Lasergene package. To perform a pairwise alignment of this type, see [Pairwise alignment of a short sequence to a chromosome](#).

Local, **Global** and **Semi-Global** are used to align sequences that are similar in length (i.e., within one or two orders of magnitude). To learn how to perform this more general type of pairwise alignment, see [Pairwise alignment for sequences with similar lengths](#).

These three algorithms work in a similar fashion, although they can often produce very different results. All use a method called dynamic programming to find the best scoring alignment between two sequences. Alignment scores are computed by adding up per-base match scores and subtracting a penalty for opening a gap (of any length) and another for the number of positions that have gaps. The match scores are based on a scoring matrix such as NUC42 or BLOSUM62. It's always a good idea to explore the effects of various settings of these three parameters to see if you can get a more desirable outcome. Depending on your two sequences, the three methods can potentially yield widely different results, so it's important to understand how they differ.

- Local Pairwise Alignment – This alignment is a modernized variant of the one described by Smith-Waterman (1981), is designed specifically to find the highest scoring aligned segments of two sequences, even if the full extent of the two is not included in the final alignment. (Note: in MegAlign Pro, the "Show Context" check-box in the Style Panel lets you display any unaligned parts of the sequences flanking the aligned segments). Local alignment reports the highest scoring contigous segment of alignment between two sequences, even if the full extent of one or both of the sequences is not included in the final alignment. Local alignments are ideal for finding a short sequence within a larger sequence. Flanking segments of sequences that are not within the aligned segment can be visualized in MegAlign Pro by checking the **Show context** box in the [Pairwise Alignment section](#) of the of [Style panel](#).

- Global Pairwise Alignment – The alternative to locally aligning is to align globally. To do this MegAlign Pro uses two variants of the Needleman and Wunsch (1970) algorithm. Global aligners don't try to find the best scoring segment, but instead require that the full extent of both sequences be included in their results. There is no requirement or guarantee that the best scoring pair of aligned segments from a local alignment will be aligned in a global alignment. Global alignment includes the full length of both sequences, even if this requires padding one or more of the sequence ends with gaps. MegAlign Pro considers overhangs and underhangs created this way as unaligned context. One example of a

situation where a global alignment is preferred over a local alignment is when there are multiple, but disjoint, segments of aligned sequence. Examples when a global alignment is a good choice: 1) aligning a CDS or mRNA sequence to a gene that contains introns; 2) aligning two sequences that differ because of the presence of large insertions, such as might be caused by transposable elements. In both cases, a local alignment is less likely to reflect the full alignment, especially if the lengths of the unalignable inclusions are long relative to the gap extension penalty.

• Semi-Global Pairwise Alignment – A relatively new approach that is particularly suitable when the two sequences differ greatly in length. When that happens, the longer sequence will have overhangs on either end of the alignment. Since overhangs are represented with gaps, a global aligner will attempt to increase the match score and minimize accumulated gap penalties by aligning parts of the shorter sequence to overhanging sequence region(s). This effect can produce a number of unrealistic, usually small aligned segments spaced by gaps near the ends of the alignment. Semi-global alignment is designed to address this problem by not penalizing gaps in overhangs (aka "end gaps"). Semi-global alignment is similar to global alignment, except the gaps placed at the ends of sequences are not penalized. A semi-global alignment might be more useful than global alignment in situations where long leading/trailing gaps might be suppressed in favor of a result that contains segments of aligned sequences punctuated by gaps.

The differences between these three pairwise approaches really can make an impact in the resulting alignment, but the choice of which to use really depends on your task. When two sequences are nearly identical (check by performing a multiple alignment and consulting the Distance view), all pairwise methods should work equally well. For basic cases, such as aligning two genes or proteins, Local alignment is a good starting point, but when things get more complicated, Global or Semi-Global may be the best choice.

Try It! – Follow a multiple alignment with Global pairwise alignments

This tutorial demonstrates a situation in which a pairwise alignment can help resolve a confusing placement of gaps within a multiple alignment. In this case, a multiple protein sequence alignment suggests that the protein sequence from a specific organism (*Tupaia chinensis*, the Chinese tree shrew) is severely truncated at its C-terminus and ends with a run of 27 residues that seem to be unrelated to the other members of the alignment, including a relative (*Sorex araneus*, the Eurasian shrew). A pairwise global alignment of the two shrew sequences, however reveals that a more likely interpretation is that the *T. chinensis* sequence contains a deletion of 235 residues followed by a terminal stretch of 32 amino acids that is nearly identical to that of the *S. araneus* sequence. Here the pairwise alignment suggests that first-pass multiple sequence alignment is not optimal. Armed with this information, you can try changing the alignment engine and gap penalties to see if a more reasonable result can be achieved. Another technique that might help is to use sub-alignments to refine the overall multiple sequence alignment.

This example also demonstrates the power of using pairwise and multiple alignments together to help interpret specific relationships between sequences that might have become obscured by gaps which were added during the multiple alignment process. With MegAlign Pro, it is simple to generate many pairwise alignments without ever having to disturb the multiple sequence alignment, which has the larger picture. This is far more convenient that starting over with several different documents.

To perform the analysis:

1. If you have not yet downloaded and extracted the tutorial data, click here to download it. Then decompress (unzip) the file archive using the method of your choice.

2. Double-click on the *Angiomotin_vertebrates.clustalo.msa* project file to launch it in MegAlign Pro.

   This project contains a collection of Angiomotin proteins from a diverse set of vertebrate species, already multiply-aligned with Clustal Omega. The figure below shows the alignment with the Sequences view and Overview scrolled to show the two tree shrew sequences. For more detail, you can expand the image below by clicking on it.
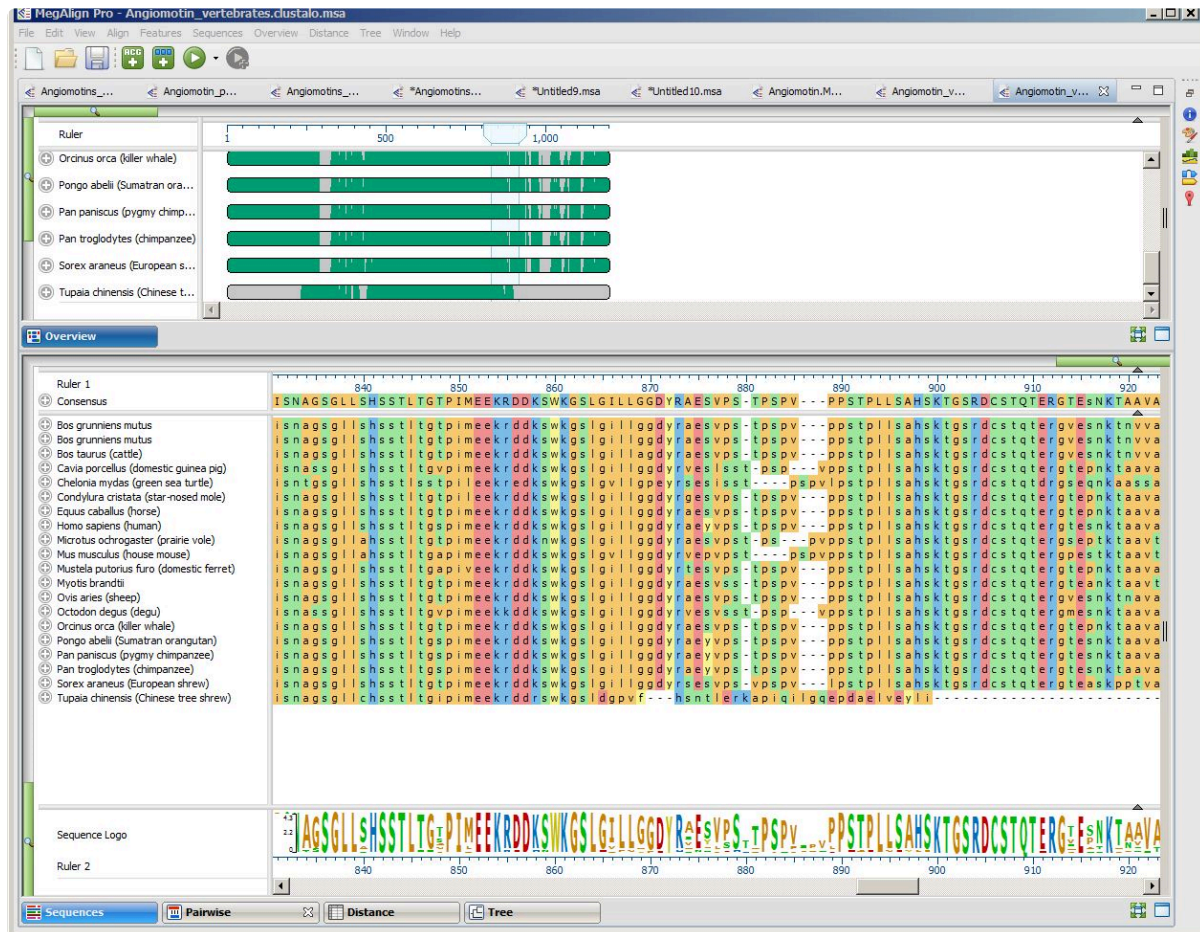
3. Notice that the *Tupaia* sequence (the lowermost item on each list) seems to end abruptly with an unaligned run of amino acids. If you can't see the abrupt end, move the Sequence view's green horizontal zoom slider to the left until the Sequences view appears as in the image above.

4. To investigate this using pairwise alignment, click on the name of the *Tupaia* sequence to select it, and then **Ctrl-click** (Win) or **Cmd-click** (Mac) on the name of the *Sorex* sequence to add it to the selection. You can select these sequences in either the Sequences view or the Overview.

5. Right-click on either name and select **Align Pairwise**.

6. In the Align Pairwise dialog box, select **Global: Needleman-Wunsch** from the **Using** drop-down menu. Since you are specifically interested in the C-terminal ends of the sequences, the **Global** option is ideal.

7. Keep the default settings for all other dialog options and click **OK**. The alignment will run and a Pairwise view will appear.

8. Scroll down the Pairwise view to the end of the alignment and notice that it is now clear that the *Tupaia* sequence contains a large deletion and that the C-terminal ends of the two shrew sequences are nearly identical.

Pairwise alignments can also be used to help interpret multiple alignments with gaps that might obscure the relationships between sequences. While the Pairwise view is still open, try some **Global** alignments between different species pairs. This is easily done by changing the sequence(s) in the drop-down menus at the top of the Pairwise view.

9. Change the sequence in the right drop-down menu to to *Homo sapiens* (human).

10. Scroll through the updated alignment and observe that the two sequences match pretty well, despite some mismatches and gaps. The header shows a similarity calculation (**Similar**) of **91.3**.

11. Change the sequence shown in the left drop-down to *Pan troglodytes* (chimpanzee). The human and chimp sequences are **98.4%** similar, with only an11-residue gap differentiating them. This similarity can be seen with a much greater degree of clarity than could be seen looking at the Overview of the original multiple alignment. (Click on the image below to expand it.)
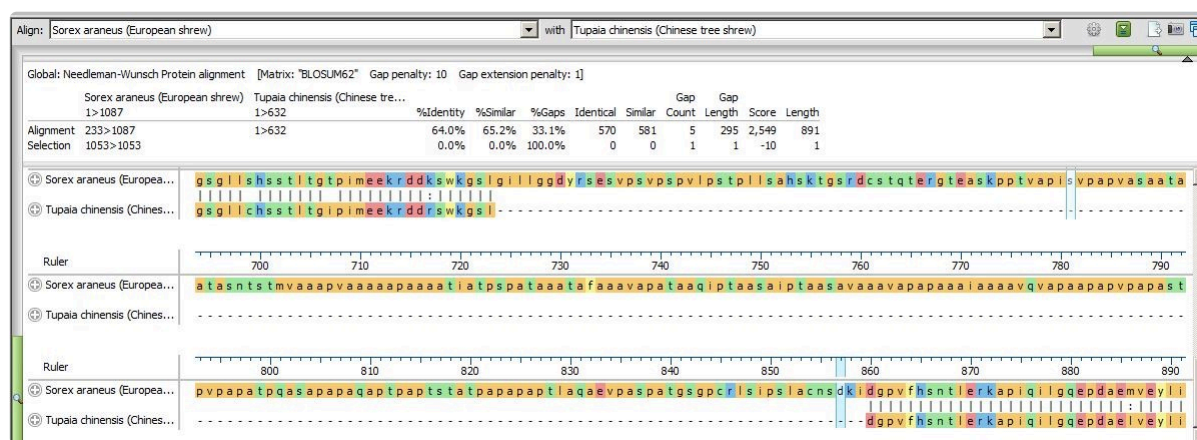
Try It! – Align transcripts to genes using Local and Global pairwise alignments

This three-part tutorial illustrates the utility of pairwise alignments when comparing mRNA transcripts to their cognate genes. This example begins with multiple alignments of the alcohol dehydrogenase (ADH) gene from *Drosophila melanogaster* and four mRNA transcript isoforms. Comparing the overview of the alignment to the gene's annotations reveals a few problems that are easily resolved with pairwise alignments.

As in Try It! – Follow a multiple alignment with Global pairwise alignments, a global alignment is the most suitable approach, because long gaps corresponding to introns may make it hard for a local aligner to join segments from exons. With a local alignment the ends of a transcript may not be included among the aligned segments.

Start with Part A: Compare results from three multiple alignment methods.

Part A: Compare results from three multiple alignment methods

This part of the tutorial demonstrates that if you don't have annotations to rely on for validating multiple alignments, a pairwise alignment is a good alternative.

This tutorial uses a default MUSCLE alignment of the *Drosophila melanogaster* ADH gene and the mRNA transcripts for isoforms F,C,E and H. In this example, some of the annotations of the original GenBank entry, NT_033779.5, were removed for the sake of clarity.

To begin the analysis:

1. If you have not yet downloaded and extracted the tutorial data, click here to download it. Then decompress (unzip) the file archive using the method of your choice.

2. Double-click on the *DmADH.muscle.msa* project file to launch it in MegAlign Pro.

   The Overview for this alignment is shown in the image below. (Click on the image to make it larger.)
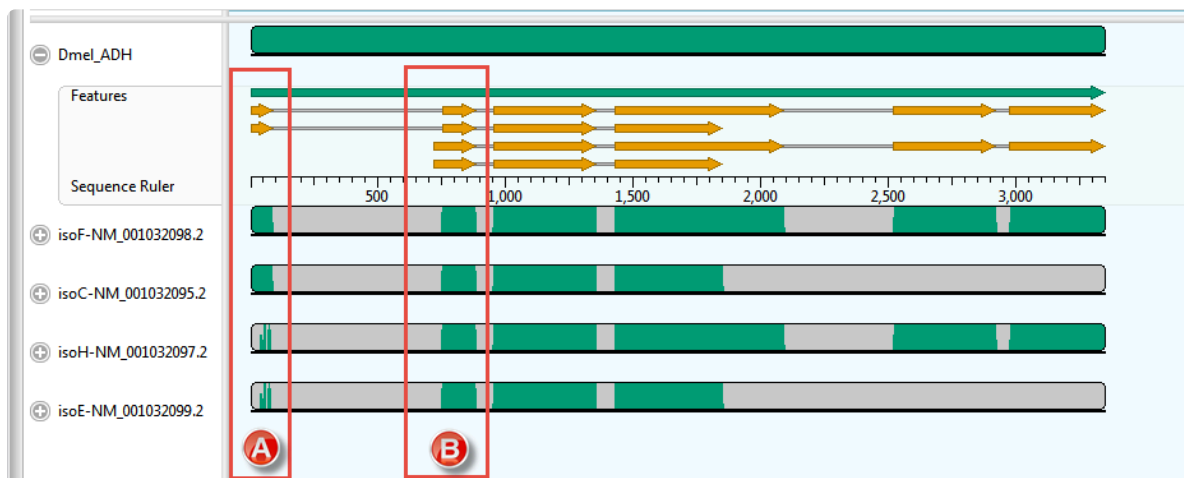
   In this example, the transcripts, represented by the four lower sequences, have already been arranged in the same vertical order as the mRNA-features for the ADH gene, represented by orange arrows in the upper sequence.

3. Hover your mouse over some of the orange mRNA features and verify that they are in the same order as the transcripts represented by the lower four sequences.

   As expected, there is a good correspondence of the aligned regions of the transcripts to the exons of the gene. But a closer examination reveals a few problems, one of which is indicated by the red boxes in the image above. Notice that that isoF and isoC are annotated to code for Exon I (box A), which matches the multiple alignment result. However the 5' end of isoH and isoE are also aligned,

somewhat poorly, to this exon. How can you tell? Note that there are only two yellow arrows in box A, but there are green bars in all four transcript rows for box A, the bottom two of which are much sparser than the others This part of the lower two transcripts — which have no corresponding yellow arrows — should really have aligned with the start of Exon II (box B).

As a reminder, the example above used the MUSCLE alignment method. Before testing whether or not pairwise alignment can resolve this problem, check out how some of the other three multiple alignment engines handle these particular sequences.

4. Multiply realign the sequences using Clustal Omega (**Align > Realign Using Clustal Omega**).

5. Examine the alignments of transcript sequences to the first two exons of the gene's sequence. In this case, there are no green bars at all in box A.

6. Realign again, this time using MAFFT (**Align > Realign Using MAFFT**). For this example data set only MAFFT gives a result that is consistent with the annotations. (Click on the image to make it larger.)
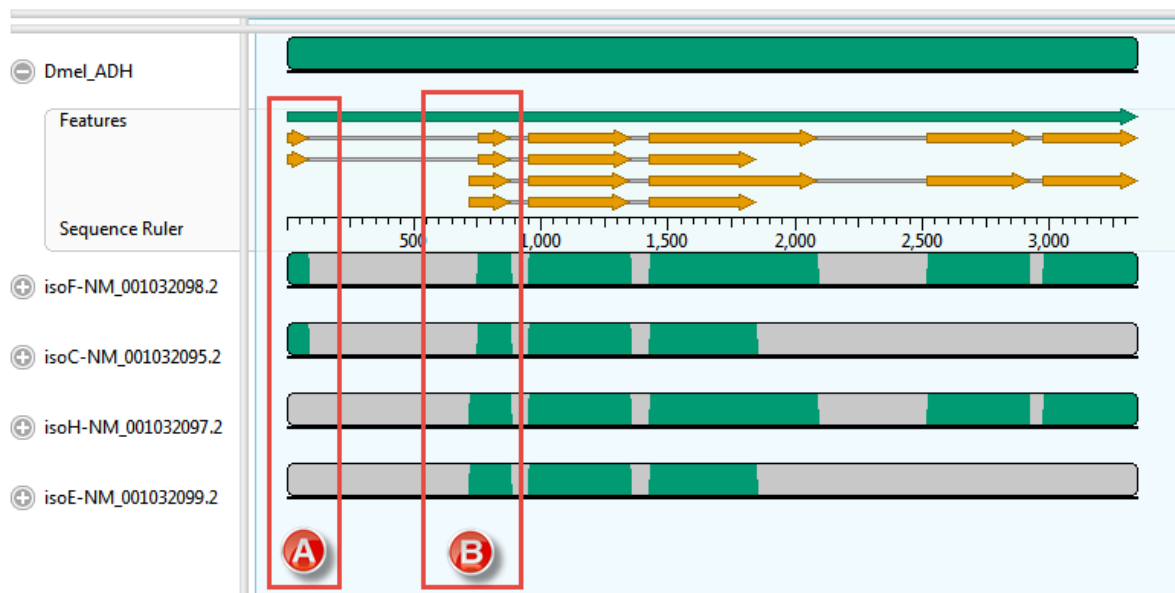
Proceed to Part B: Use a Local pairwise alignment method.

Part B: Use a Local pairwise alignment method

Now that you have finished Part A of the tutorial, this part shows how you can use a local pairwise alignment to resolve the correct mapping of the first intron in isoF and isoC.

1. Click on Pairwise view tab. Look at the left drop-down menu at the top of the view to see that the uppermost sequence (Dmel_ADH) has automatically been selected as the *reference*.

2. In the right drop-down menu, choose "isoF-NM_001022098.2". I The default *Local: Smith Waterman* method is used to create the alignment

3. To make sure all features will be displayed:

   a. Open the **Tracks** side-panel.

   b. Under **Pairwise details**, click once on the word **Features**.

   c. Move the **Height** slider to the right until you can see all the feature rows.

4. In the Pairwise view, click on the plus-sign icon next to the name **Dmel_ADH** to reveal the Features and Sequence Ruler tracks.

   A correct alignment would show the 5' end of the isoF sequence aligned with the two orange arrows that begin at the very start of the Dmel_ADH sequence. Instead, however, the beginnings of both the ADH reference sequence and the mRNA are shown as dimly colored *context*; in other words, as unaligned flanking sequence.
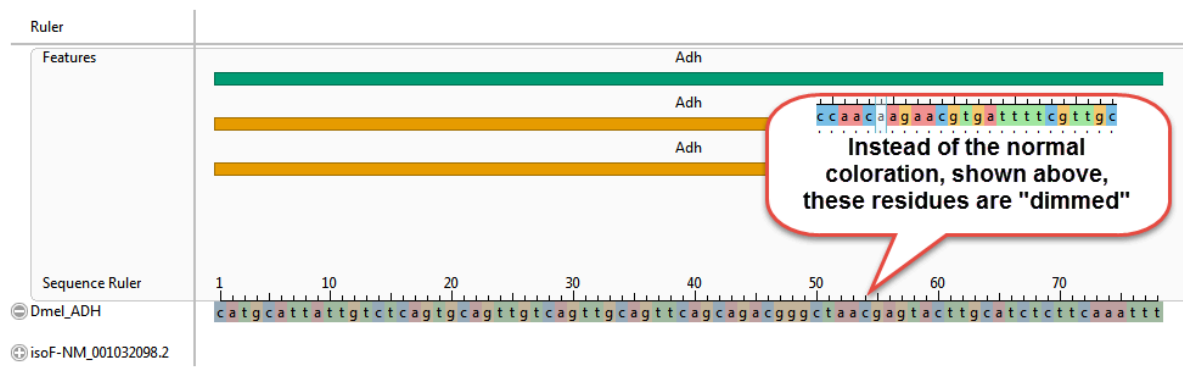
5. Scrolling down through the alignment observe that the 5' end of the transcript sequence has aligned, albeit poorly, with a region beginning within the first intron. Meanwhile, a solid ungapped alignment doesn't start until 5' end of the second exon annotated for either isoF or isoC. (See the boxed area in the image below.)
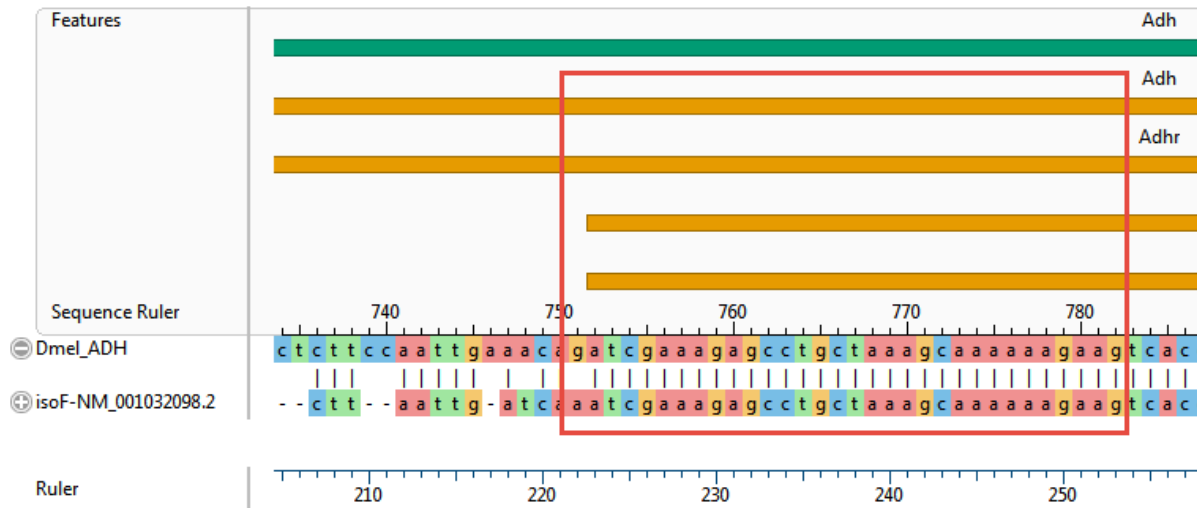
Clearly, a local pairwise alignment was not an improvement over the first two multiple alignments tried in Part A.

Most likely, introducing a sufficient number of gaps to span the intron would reduce the score so much that the aligner can find a higher scoring segment that contains numerous short gapped regions.

Proceed to Part C: Use a Global pairwise alignment method.

Part C: Use a Global pairwise alignment method

Since the Local pairwise alignment in Part B was not an improvement over the multiple alignments in Part A, you will now try a Global pairwise alignment. This type of alignment forces the ends of both sequences to be aligned.

1. Click the gear icon on the Pairwise view toolbar and select **Global: Needleman-Wunsch** from the "Using" drop-down menu. Then press **Align**.

2. After aligning, the display options that you set above remain in effect, but the sequences' detail tracks will collapse. Expand the Dmel_ADH track again by clicking on its plus sign.

3. To see the "big picture," zoom all the way out using the green "zoom slider" above the view. The pairwise alignment now appears similar to the Overview. You can see that the global alignment using isoF matches Dmel_ADH's annotations. (Click on the image to expand it.)
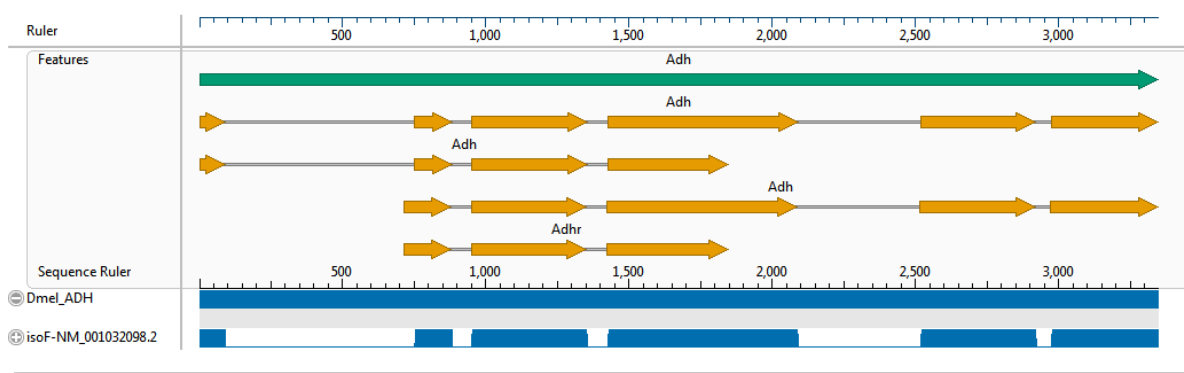
MegAlign Pro allows you to open additional pairwise views so that you can compare alignments with different settings, or even different pairs of sequences. You can use this to compare any of the other three mRNA sequences to Dmel_ADH, while leaving the previous alignment untouched.

4. Clone the current Pairwise view using the "Clone this view" tool situated at the far right side of the toolbar. The two views will start out the same, but they are completely independent.

5. Select isoE-NM001032099.2 as the query sequence, keeping the reference sequence set to Dmel_ADH. The new Global pairwise alignment now gives the expected results.

This is the end of the tutorial.

Try It! – Use Local pairwise alignment to find a gene within a genome

Consider the case of a researcher who is trying to investigate the role of a gene isolated from a yet-unsequenced *Salmonella* strain. This strain has already been demonstrated to be both copper and multi-drug resistant. As demonstrated in the following tutorial, one way to proceed would be to align the sequence of the uncharacterized gene to the genome of a related reference strain. If a reasonable alignment was found, the annotations of the matching segment could then be examined to infer the function of the cloned gene.

> ✳ **Note:** To align the large genomic data set used for this tutorial, we recommend using a machine with a minimum of 8 GB — and preferably 16 GB — of RAM.

To perform the analysis:

1. If you have not yet downloaded and extracted the tutorial data, click here to download it. Then decompress (unzip) the file archive using the method of your choice.

2. Double-click on the *Salmonella_CT18_plus_gene.msa* project file to launch it in MegAlign Pro.

   The document contains just two unaligned sequences: the genomic sequence of reference strain *S. enterica serovar Typhi*, strain CT18, and the unknown sequence.

3. Select both sequences, right-click on the selection, and then select **Align Pairwise** from the context

menu.

4. In the ensuing dialog box, note that the default pairwise alignment method is **Local: Smith-Waterman**. Since this method is the best choice for finding a small segment of similarity within a larger sequence such as chromosome, keep the default settings and click the **OK** button. (Note: If any long gaps were needed for alignment, you could always realign later using lower gap penalties.)

This is a good time to discuss the concept of "reference" (or "target") sequence versus the "query" sequence. In MegAlign Pro, these two sequences are defined through the two drop-down menus at the top of the view. The left menu should be used for the reference, and the right menu for the query sequence. Most of the time, the sequence selected to be the "reference" is somewhat arbitrary. However, when one sequence is much longer than the other — as in this example — the longer sequence should be used as the reference.

After the alignment has completed, the Pairwise view header indicates that the **706 bp** query matches a **702 bp** segment in the CT18 genome with **98.6%** Identity and has **1** gap that is **4** bp long.

Local: Smith-Waterman DNA alignment		[Matrix: "NUC44"	Gap penalty: 10	Gap extension penalty: 1]					
	NC_003198.1 (CT18) 1>4809037	unknown 1>706	%Identity	%Gaps	Identical	Gap Count	Gap Length	Score	Length
Alignment	279279>279980	1>706	98.6%	0.6%	696	1	4	3,443	706

5. Scroll down through the alignment until you locate the gap, which begins at alignment position 452.

Notice that the gap is shown within the reference, indicating that there is a 4 bp insertion in the unknown sequence. Since the insertion length is not a multiple of three, the gap most likely represents a frame-shift in the unknown sequence.

6. In the **Tracks** side panel on the right, locate the **Pairwise Details** section and check the box next to **Translation**.
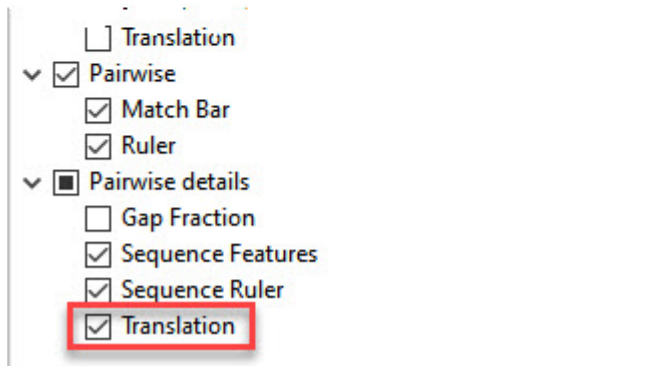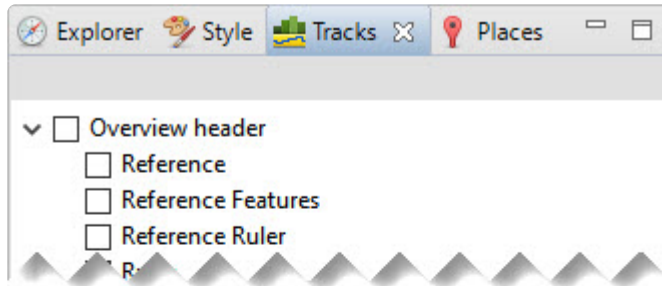
7. Returning your attention to the Pairwise view on the left, open the <u>tracks</u> for the unknown sequence by clicking the plus sign corresponding to that sequence.

    The translation tracks (labeled 1, 2, 3) clearly indicate that the frame-shift has introduced an in-frame stop beginning at base 530 of the query sequence (red box in the image below). In other words, it represents a nonsense mutation. This is good evidence that this is indeed a mutant gene compared to the non-drug resistant reference strain.
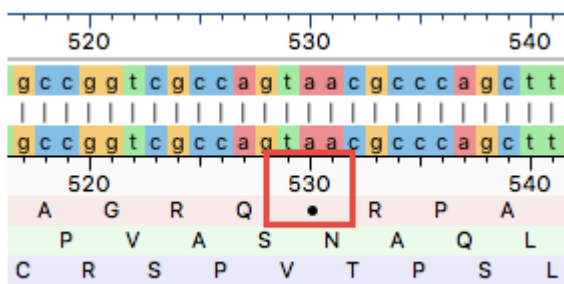
    To find more clues to the identity of the unknown gene, you will next investigate the CT18 reference sequence annotations at this position.

8. Collapse the tracks for the unknown sequence by clicking on any of the minus signs corresponding to that sequence.

9. Expand the detail tracks for the reference (CT18) by clicking any of the corresponding plus signs for that sequence.

Observe that the indel's position overlaps a gene and a CDS that are identified as *STY0266*.

10. Click anywhere on the *STY0266* feature to select it.
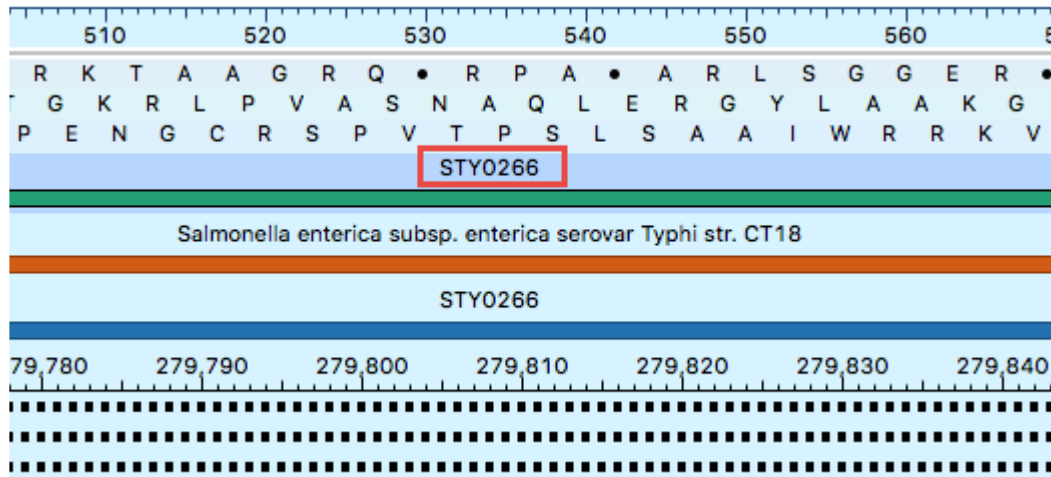
11. Examine the annotations for *STY066* in the Details panel. If the Details panel is not visible, reveal it using **View > Details**.
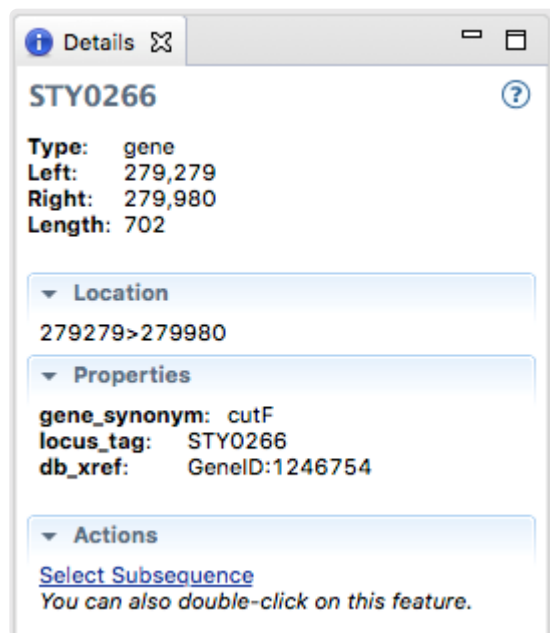
As shown in the Details panel, the gene sequenced from the unknown, copper-resistant strain is a defective version of *cutF* (also called *nlpE*) which encodes a copper homeostasis protein that has been shown (Nishino, et al. 2010) to be associated with elevated multidrug and copper resistance in *E. coli*.

Phylogenetic Trees

MegAlign Pro lets you create and display one or more phylogenetic trees for a multiple alignment. The trees can use different algorithms, or the same algorithm with different settings. To learn more, see the following topics:

- Generate a phylogenetic tree
  - Advice on choosing the best tree for your data (DNASTAR blog post)
- Get to know the Tree view
- Work with multiple trees
- Change the look of the tree (fonts, colors, decimal places shown, etc.)
- Reroot a tree
- Import a Newick-format tree
- Export a tree to a tree viewer

Generate a phylogenetic tree

To generate a new phylogenetic tree:

1. Add sequences and perform a multiple alignment.

2. Use any of the following methods to direct MegAlign Pro to create a tree using the algorithm you specify. For helpful advice on choosing the best tree for your data, see this December 2020 DNASTAR blog post by Brian Walsh, PhD.

   - Click on the **Tree** tab to open the Tree view. Choose one of the two available tree-building algorithms.

   Neighbor Joining: BIONJ   Maximum Likelihood: RAxML

      ◦ **Neighbor joining: BIONJ** uses the BIONJ algorithm (Gascuel, 1997, a variant of the Neighbor-Joining algorithm (Saito and Nei, 1987) that was used in classic MegAlign.

      ◦ **Maximum likelihood: RAxML** was developed by Stamatakis (2014). Comprehensive information about options associated with this algorithm can be found in the RAxML Manual (PDF).

   - In the Trees section of the Explorer panel, press the green plus-sign tool (⊞) to the right of the table. In the **Using** drop-down menu, select the desired algorithm.

   - Use the menu commands **Tree > Compute Phylogeny Using > Neighbor Joining:BIONJ** or **Tree > Compute Phylogeny Using > Maximum Likelihood: RAxML**. Choose **Maximum likelihood (RaxML)** if your data set is divergent, if it contains many taxa and/or very long sequences, or if you need support (bootstrapping) values. To see support values, you need to check bootstrapping in the parameters dialog. If bootstrap value is 50% or below, it's a coin flip, most use 70% cutoff to trust phylogenetic analysis data. Choose **Neighbor joining: BIONJ** if your project includes fewer than 4 sequences, the minimum needed to use **Maximum likelihood (RaxML)**.

3. Customize settings based on the algorithm you chose.

   - If you chose **Neighbor joining: BIONJ**:
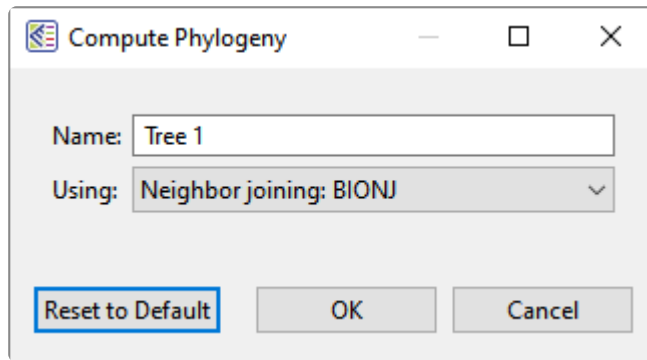
- ◦ Type in a new **Name** for the tree or keep the default name of **Tree <number>**.

- ◦ If you change your mind and want to use the other algorithm, choose it from the **Using** drop-down menu.

- ◦ Press **OK** to generate the tree.

- If you chose **Maximum Likelihood: RAxML**:
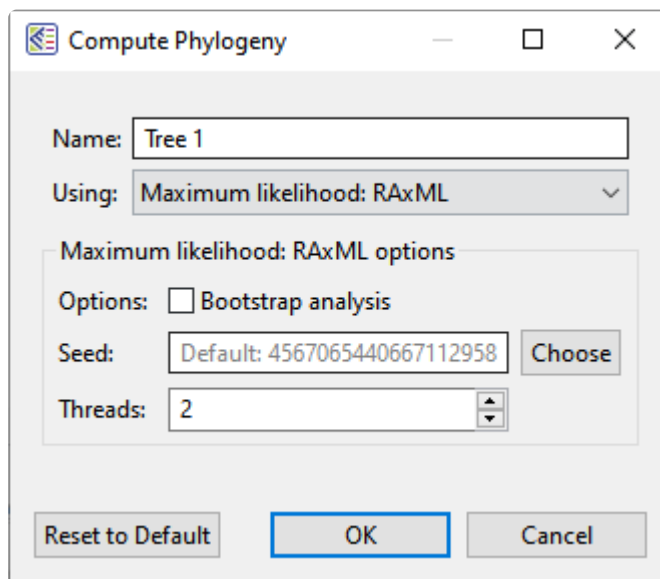
- ◦ Type in a new **Name** for the tree or keep the default name of **Tree <number>**.

- ◦ If you change your mind and want to use the other algorithm, choose it from the **Using** drop-down menu.

- ◦ To add a bootstrapping step to the calculation, place a checkmark next to **Bootstrap analysis**; if you do so, another option appears. Type in the number of **Iterations** or keep the default value of 100.

- ◦ Keep the default **Seed** value, type in a new integer, or select **Choose > Random** to enter a random value in the box.

- Enter the desired number of **Threads**. The default value is 2.

- Press **OK** to generate the tree.

4. (optional) Use the Style panel's Tree section to customize the tree as desired.

5. (optional) Create another phylogenetic tree using a different algorithm or different settings for the same algorithm. See Work with multiple trees.

Try it! – Determine the strain of an experimental viral genome sample

In this tutorial, you will use MegAlign Pro to compare an experimental SARS-CoV-2 sample to references sequences from four SARS-CoV-2 strains in order to see which strain is present in the sample. This type of analysis can be used with any viral genome samples that have been sequenced and assembled into an accurate draft genome by your sequencing core.

1. [Download the tutorial data](#) and extract it to any convenient location (i.e., your desktop).

2. Launch MegAlign Pro. From the Welcome screen, choose **New blank alignment project**.

3. Use **File > Add Sequences** to add the four (known) COVID-19 variant *.fasta* files and the experimental *Spike-var.fasta* sample from the tutorial data folder.

   If a popup asks the file type, specify the sequences are **DNA**.

4. Use **Align > Align Using MAFFT** to perform a multiple alignment. The alignment will finish in 1-2 minutes.

5. To generate a phylogenetic tree, click the **Tree** tab at the bottom of the view. Select **Maximum Likelihood: RAxML**. Keep all defaults and press **OK**. The tree will take 1-2 minutes to generate.

6. Once the tree is displayed, move the green slider at the top of the Tree view all the way to the right to better differentiate the branches. There are four strains represented, and these may appear in any order as you scroll down the tree.

Reference group name	SARS-CoV-2 strain
B.1.1.7	Alpha
B.1.351	Beta
B.1.617.2	Delta
C.37	Lambda

7. Differentiate the strains even more by applying a different background color to sequences from each strain. To do this:

   a. In the Tree view, select all members of a strain by clicking the horizontal root anchoring its clade.

   b. Choose **View > Style > Tree** to open the part of the Style panel that controls the appearance of the tree.

c. In the Style panel, click the **Background** box to add a checkmark, then click the white box to the right to choose a new shade from the color chooser.
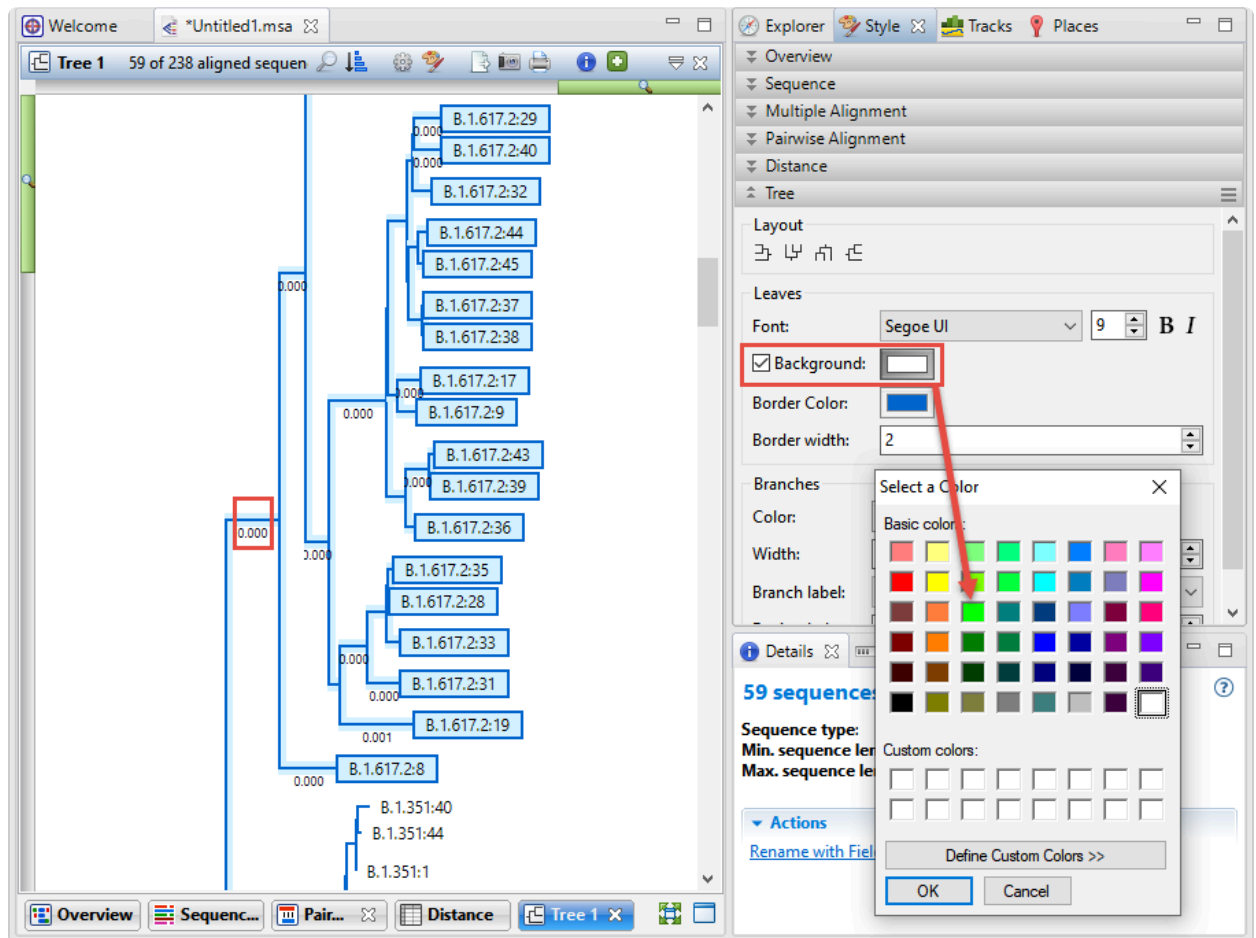
In the image below, the Beta strain was assigned a green background and the Delta strain was assigned a salmon background.

8. In the Tree view, locate *SRR13380669_NC_045512.2*, the experimental draft genome that you added as *Spike-var.fasta* from the tutorial data folder. Note that this strain is a close relative of the "Beta" strain, which was first noted in South Africa in January 2021.
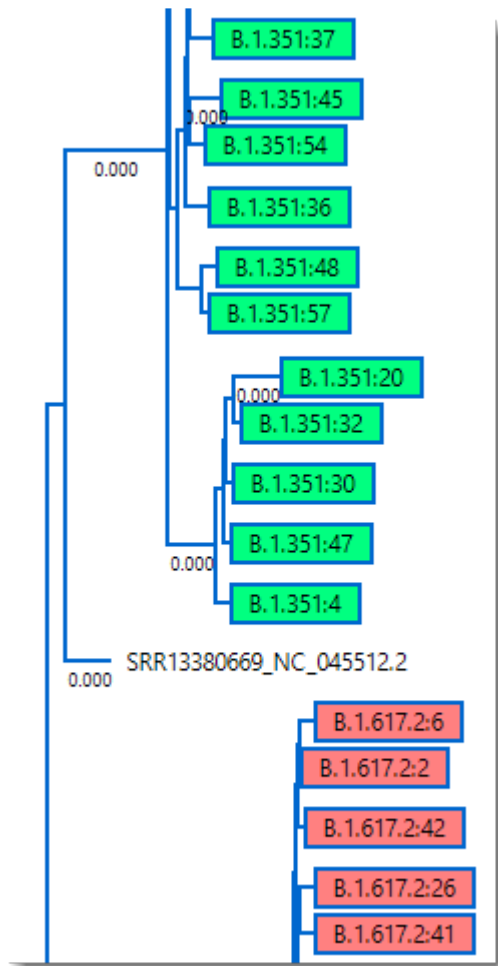
9. To confirm this observation, open the Distance table by clicking the **Distance** tab at the bottom of the window. Select the experimental sample row, **SRR13380669_NC_045512.2** (typically the bottom-most row), then choose **Distance > Order Sequences by Distance from Selection**.

10. The lower-left triangle of the table shows distances between sequences, with white being most related and dark red being least related. Note that all visible distances in the visible portion of this triangle are "0.00." To see differences between the sequences more clearly, locate the **Distance** section of the Style panel (on the right) and change the **Decimal places** from **2** to **4** (or **5**).

    Scrolling down, notice that most of the closely related sequences—shown in white or very light pink—are from the "Beta" strain **B.1.351**. Two random examples within these light areas are shown in the image below.

Work with multiple trees

What are dynamic trees and how do they differ from "snapshots?":

MegAlign Pro allows you to compare multiple phylogenetic trees generated using custom settings. The phylogenetic algorithm you choose dictates which customizations affect the tree and whether a tree can be dynamic, static, or either.

Neighbor Joining: BIONJ trees are "dynamic." Dynamic trees are recomputed automatically when their input changes. Changes made to rooting, zoom level, ordering of nodes and any changes made in the Tree section of the Style panel will affect only the individual BIONJ tree to which changes are applied. However, changes made to the alignment method or its parameters or to the distance parameters will affect the structures of all BIONJ trees equally. To compare two BIONJ trees before and after making changes to the alignment method or distance parameters, use the **Take snapshot** tool (🖼) to preserve a copy of each view showing the tree in its current state. By taking additional snapshots as you change parameters, you can then compare multiple versions of the tree. To learn how to view trees side-by-side, see the final section in this topic.

Maximum likelihood: RAxML trees are created as snapshots from the start. You can generate multiple RAxML trees, each with different alignment methods, RAxML parameters, etc. Each tree will be generated as a unique snapshot that is unaffected by subsequent customizations. Note that distance parameters have no effect on RAxML trees.

How can you determine whether a tree is dynamic or whether it is a snapshot? Snapshots are distinguished from dynamic trees by having the word "Snapshot" in the header, and by featuring the snapshot icon in the leftmost column in the Trees section of the Explorer panel. You can apply custom styles to snapshot trees just like you can with dynamic trees.

Controlling which tree is currently displayed:

There are three methods for choosing which tree is displayed in the Tree view.

- **View > Tree** commands – Use **View > Tree > (Tree name)** to make that tree active.

- **Explorer panel** – The Trees section of the Explorer panel contains a table listing the name of each tree, the method used to create it, and the time the tree was built. Snapshot versions of trees have a snapshot icon in the leftmost column. To make a particular tree active, click its row in the table. In the example below, three trees are available and their names are: **Tree** (the default tree), **Tree 1** and **Tree 2**.
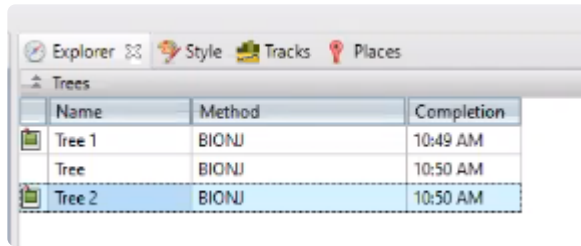
- **Tree view tab** – The Tree view tab shows the number of available trees as follows: **(Number of trees) Tree**. The trees themselves are numbered as specified during their creation (see steps above). In the example below, three trees are available, and their names are: **Tree** (the default tree), **Tree 3**, and **Tree 5**.
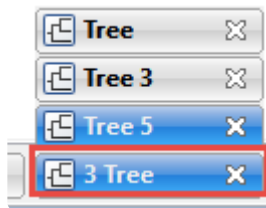
Displaying multiple trees side-by-side:

Right-click on the **Tree** tab and choose **Detach 'n' Tree Views**. You can then drag the views with your mouse to position them as desired.

Note that each detached view has its own **File**, **View** and **Tree** menu. If you wish to use commands such as **Tree > Root On**, you need to use the menu in the detached view rather than the menu in the main MegAlign Pro window.

Reroot a tree

Each tree in the Tree view displays an incoming root edge, which is intended to represent the location of a possible common ancestor. The root edge appears on the left of a tree drawn in the default orientation.
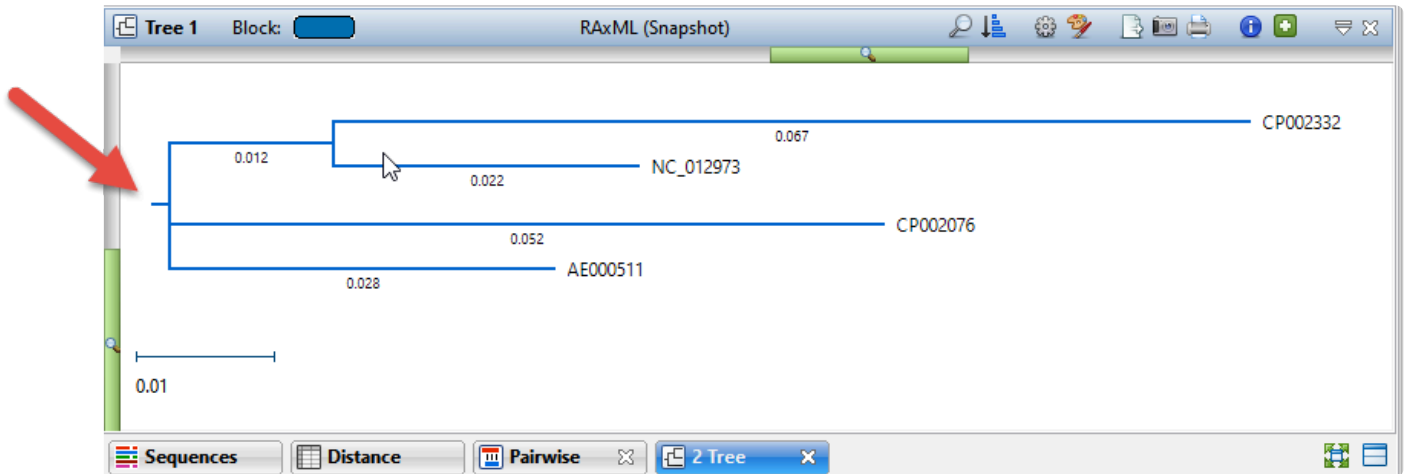
For purposes of visualization or publication, you may wish to root the tree or choose a different location for the root. For example, you may wish to balance the tree by moving the root to a long branch.

To reroot a tree, use the **Tree > Root On** subcommands or press the **Sort sequences** tool (⬇≣) in the Tree view and use the **Root On** commands there.

- **Root On > Selected Branch** places the root at the base of the currently selected branch. After selecting CP002076 in the example tree above, this command yielded the following tree:
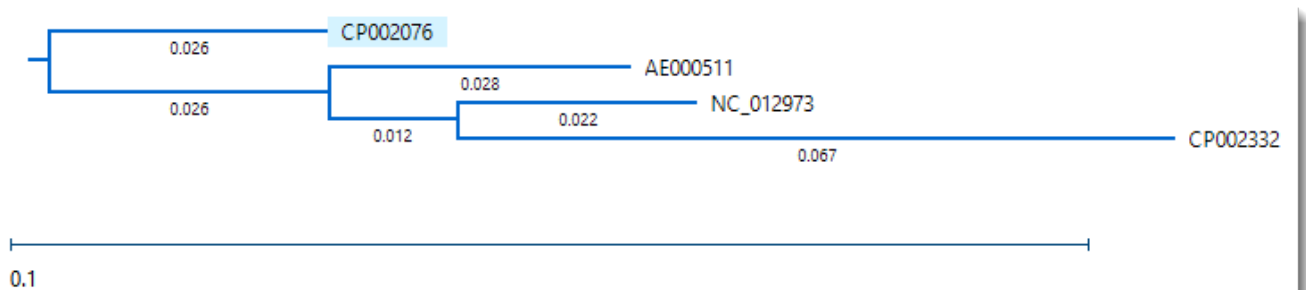
- **Root On > Midpoint Branch** computes the midway point between the two leaf nodes that are farthest apart and places the root there. This can result in a more balanced tree. Using this command with the example tree yielded the following tree:
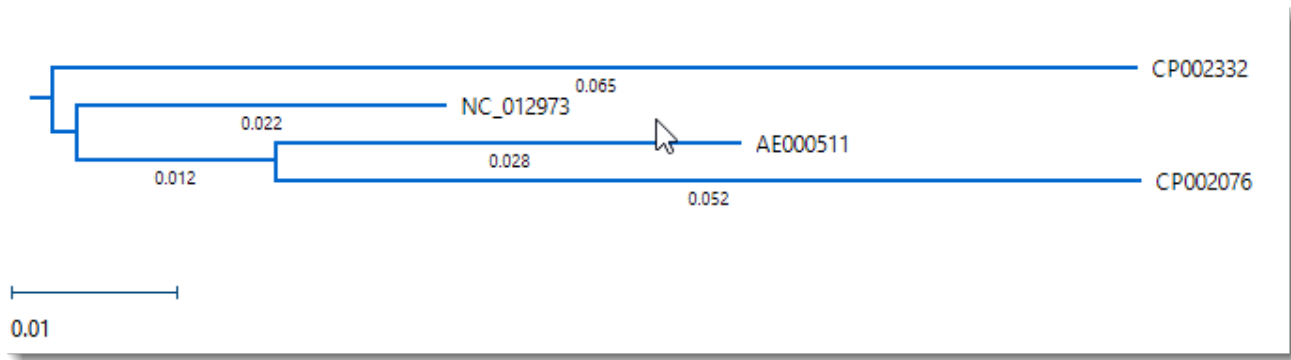
Import a Newick-format tree

MegAlign Pro can open Newick trees with *.newick*, *.txt* or *.tree* extensions.

To import a Newick-format tree file:

1. Make any MegAlign Pro project tab active, or open a new, blank project. Trees cannot be opened from within the [Welcome screen](#)

2. Choose **File > Add Trees** or click to the right of the **Add Sequences** tool () at the top left of the window and choose **Add Trees**.

3. Select one or more Newick tree files and choose **Open**. Each tree is added as a named tree in the **Tree** view tab and the **Explorer** panel, as described above. The tree name is the same as the file name.

4. (optional) If desired, repeat Step 2 to add additional trees.

Trim an Alignment

After performing a multiple or pairwise alignment, you may wish to trim off one or both ends before realigning, saving or exporting.

To trim one or both ends from an alignment:

Perform a multiple sequence alignment then use either of the methods below:

- Use your mouse to grab and drag the gray slider located on either end of the Sequences view. The selected portion will be trimmed off instantly. If you release the mouse button in the wrong location, use **Sequences > Clear Alignment Trim** to restore the trimmed portion, then try again.
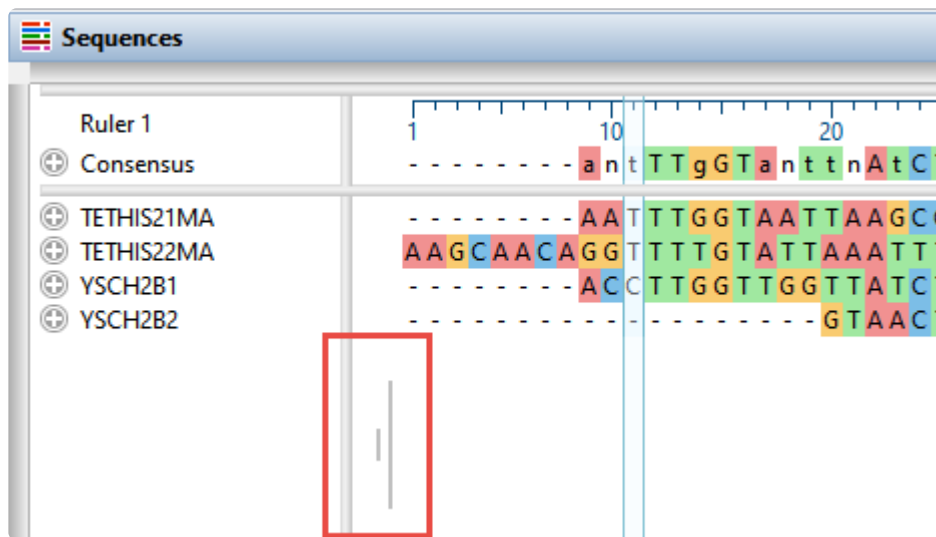
- Select a portion of the consensus sequence in the Overview or Sequences view. Then use the **Align > Trim Alignment to Selection** command.

  - To remove the region from the 5' end of the sequence to the 5' end of the selection, choose **From Start**.

  - To remove the region from the 3' end of the selection to the 3' end of the sequence, choose **From End**.

  - To remove everything except the selection, choose **From Start and End**.

To restore all trimmed regions:

Use **Sequences > Clear Alignment Trim**.

Compute Variants

After performing an assembly, MegAlign Pro can calculate variants between other sequences and the currently-specified reference sequence, or between all sequences and the consensus. For instance, you might download an annotated reference sequence that you found online and use this workflow to compare your own sequences to it.

To compute variants against the consensus sequence:

1. Remove a specified reference sequence (if one exists) by selecting it in a view and choosing **Sequences > Stop Using as Reference**.

2. Perform an alignment using the **Align > (Re)Align Using > (algorithm name)** command for the algorithm you want to use.

3. Click **Compute Variants** in the Variants view or choose **Variants > Compute Variants**. Results are displayed in the Variants view.

To compute variants against a reference sequence:

1. Perform any type of alignment using the **Align > (Re)Align Using > (algorithm name)** command. If you are running a MAFFT alignment, you can optionally specify a reference sequence during setup to guide the alignment. However, this is not necessary for later variant computation.

2. If you have not already done so, specify a reference sequence. If you ran a MAFFT alignment using a reference, note that you can choose a different reference in this step for use in variant analysis.

   By default, the reference sequence is displayed in bold blue italics to differentiate it from the other sequences.
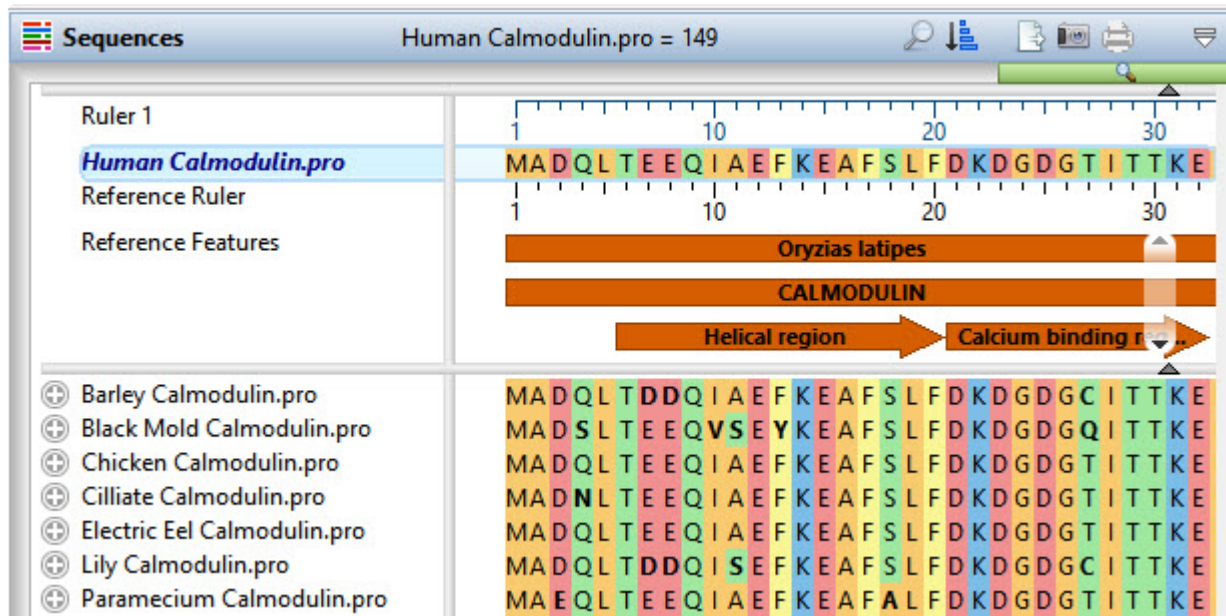
3. Click **Compute Variants** in the Variants view or choose **Variants > Compute Variants**. Results are displayed in the Variants view.

⁑ **Note:** If Mauve was chosen as the alignment method in either procedure above, variants are computed only for the currently-selected block. Each time you select a different block, you may choose to compute its variants, if they haven't already been computed.

Copy, Paste and Delete

In many cases, once you have selected text, data, sequences or subsequences, you can copy the information to the clipboard.

> ✳ **Note:** Copy functions are limited by application memory and operating system buffer size. Therefore, for very long sequences or subsequences, we recommend exporting data to a file rather than copying it.

To copy the selected text as it appears in the view:

Choose **Edit > Copy**, press **Ctrl/Cmd+C**, or use the context (right-click) menu to choose **Copy**.

- If a portion of sequence or one or more features are selected, this command copies the selected subsequence(s) as one long string, including the gaps. The subsequence is copied as one long string, without line breaks. If the selection extends across multiple members of the alignment, names are included as prefixes.

- If a selection is made in the Distance view, this command copies the selection in a tab-delimited format suitable for use with spreadsheet programs like Microsoft Excel®.

- If a selection is made in the Pairwise view, this command also copies the header information.

- In the Details panel, contents of each section must be copied separately.

To copy the selected text or data in FASTA format:

Choose **Edit > Copy As FASTA**, press **Ctrl/Cmd+Shift+C**, or use the context (right-click) menu to choose **Copy As**.

To paste text from the clipboard into a text field:

Choose **Edit > Paste** or use **Ctrl/Cmd+V**. This command is disabled in all but a few circumstances.

To delete selected text from a text box:

Press the **Delete** key or use **Edit > Delete**.

Export

The following topics discuss how to:

- [Export an image of the view](#)

- [Export data to a file](#)

- [Export a tree to a tree viewer](#)

- [Export data from the Text, Table or Pairwise views](#)

If you [renamed sequences manually](#) or [automatically](#), learn how to [retain custom names during export](#).

Export an image of the view

To export an image of a view :

1. Choose **File > Export Image > (View Name)** or click the **Export image** tool (📷) at the upper right of a view that allows exporting.

2. Specify the file name and choose the export format: Adobe PDF (*.pdf*), Microsoft PowerPoint-optimized (*.pptx*), PNG image (*.png*) or JPG image (*.jpg*, *.jpeg*).

> ✱ **Notes:** If you choose Adobe PDF format and a particular font is not supported there, a similar font will be substituted for the original one automatically. If you choose the format that is optimized for use with Microsoft PowerPoint, the MegAlign Pro image is converted into an editable set of objects that can be grouped, ungrouped and moved within the target application. To constrain an object (e.g., a MegAlign Pro track) within Microsoft PowerPoint to move only vertically, rather than laterally, hold down the **Shift** key.

If exporting the Tree view, the export procedure is now complete.

If exporting the Overview, Sequences view or Pairwise view, the following dialog (or an abbreviated version) will open.
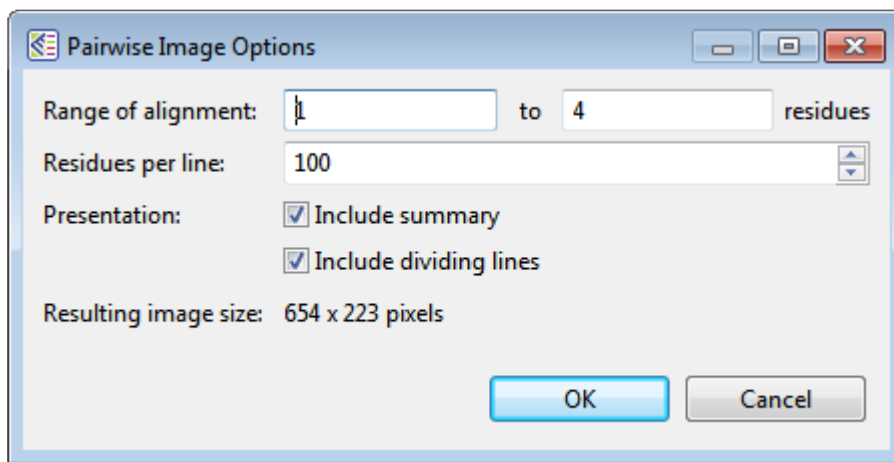
3. Choose export options.

   • Use the **Range of view** or **Range of alignment** boxes to specify the range of residues that should appear in the image. The default is the current viewport range.

   • Type in the number of **Residues per line**. By default, MegAlign Pro places the entire range on a single line.

- Check the **include summary** box if you want to include the summary (above the sequences area) in the image.

- Check the **Include dividing lines** box if you wish to include separators between the header, sequences, and footer.
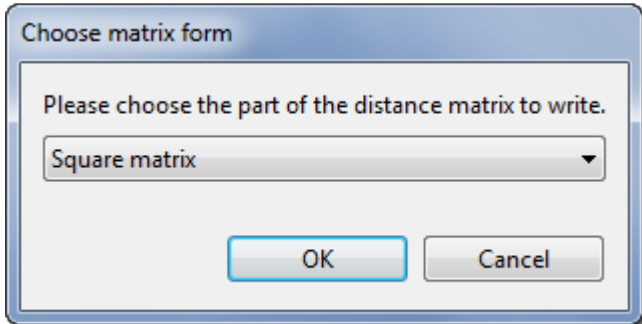
The **Resulting image size** updates automatically based on the options you choose.

4. Click **OK** to export the image using the specified options or **Cancel** to close the dialog without exporting.

The following video shows how to export an image from MegAlign Pro and edit it in PowerPoint.

Export data to a file

Several commands in the **File** menu allow you to export alignments, distance matrices and phylogenetic trees, as well as aligned sequences and the consensus sequence. Unless you wish to export all sequences in the project, you must first make a selection.

Command	What is exported?
Active Block Alignment	Aligned sequences in the active block with gaps included (Mauve alignments only). This option supports exporting to a wide variety of formats, including FASTA, GenBank, MEGA, MegAlign, MSF and Nexus.
Active Block Alignment Report	See Export the alignment report.
Active Block Subsequences	A selected range of the aligned sequences in the active block (Mauve alignments only).
Aligned Subsequences	A selected region of the aligned sequences, including gaps.
Alignment	All aligned sequences, including gaps.
Alignment Report	See Export the alignment report.
All Data	Aligned sequences, distance table and phylogenetic tree information. Supports Nexus formats (*.nex*, *.nxs*) only.
Consensus	The consensus sequence resulting from the current alignment.
Distance Matrix	Matrix information from the Distance view Supports comma- or tab-separated formats (*.csv* or *.tab*), PHYLIP format (*.dist*) or Nexus formats (*.nex*, *.nxs*). After selecting a name and format, click **Save** to launch the Choose Matrix Form dialog: Choose matrix form Please choose the part of the distance matrix to write. Square matrix OK Cancel Use the drop-down menu to choose whether to save the entire ("square") matrix or only the

upper or lower triangle.

Note: If you choose to save in .dist format, the sample names in the saved file will automatically be truncated to 10 characters.

Selected Active Block Aligned Sequence	Aligned sequences in the selected block (Mauve alignments only), including gaps.
Selected Aligned Sequence(s)	Selected aligned sequence(s), including gaps.
Selected Sequence(s)	Selected individual sequence(s), exported as unaligned individual sequences.
Sequences	All sequences in the project, exported as unaligned individual sequences. Supports various GenBank and FASTA formats; MegAlign format (.*meg*); or Nexus formats (.*nex*, .*nxs*). When exporting to most formats, gaps are represented as embedded dashes. When exporting to GenBank, gaps in nucleotide sequences are represented as "n's" and those in amino acid sequences are represented as "x's."
Tree	Phylogenetic tree information from the Tree view. See Export a tree to a tree viewer.
Variants	Variant table information from the Variants view.

Export the alignment report

To export the alignment report for the selected block from a Mauve alignment, use **File > Export Data > Alignment Report**. To export the report for any other alignment type, use **File > Export Data > Alignment Report**. In both cases, the report is saved as a *.txt* file.

- From the **Sequences** drop-down menu, choose **All** to create a report for all sequences or **Top** to create a report only for the uppermost sequence in the Sequences view.

- In the **Include** section, check items you with to include and remove checkmarks from the others.

- In the **Left** and **Right** sections, check boxes showing where, if anywhere, you wish to include the sequence **Name** and/or **Position**. To right-justify any labels chosen to occupy the left margin, check **Right-justify**.

- To wrap sequences (rather than allowing them to go off the page), check **Wrapping** to wrap sequences in the report rather than allowing the sequence to go off the page. If you elect to wrap, enter the number of residues per line in the text box.

- To include a tab or space between residues, check the **Extra space between residues** tab and choose **Tab** or **Space** from the drop-down menu.

- To group residues and include a tab or space between them, check **Residue groups**, make a numberican selection from the drop-down menu, then choose **Tab** or **Space** from the second drop-down menu.

- To include gaps at the end of the sequence, check **End gaps**.

- If your alignment consists of protein sequences, you can show their 3-letter IUPAC amino acid codes in the report by checking **3-letter amino codes**.

After making the desired selections, press **OK** to generate the report. You will then be prompted to save the report.

Export a tree to a tree viewer

To export a phylogenetic tree to a file:

Use **File > Export Data > Tree** to export tree data in Newick (*.newick*) or Nexus (*.nex*, *.nxs*) formats . The saved file can be opened by several free and commercial evolutionary biology software programs.

To set the default application for opening saved trees, see Set preferences.

To export and open the phylogenetic tree in a third-party tree viewer:

The following procedure lets you automatically launch a third-party tree viewer and view the tree within that application.

1. Outside MegAlign Pro, download the desired tree viewer and install it on your computer. One compatible viewer is Archaeopteryx (Win only).

2. In MegAlign Pro, set the default application for opening saved trees using the **File > Send Tree To** menu command.

3. Perform an alignment.

4. Export the tree using **File > Send Tree to (Tree Viewer)**. If you have chosen a compatible tree viewer, this command launches the viewer with the active MegAlign Pro project tree open. If a tree viewer has not yet been specified, you will receive the following message:
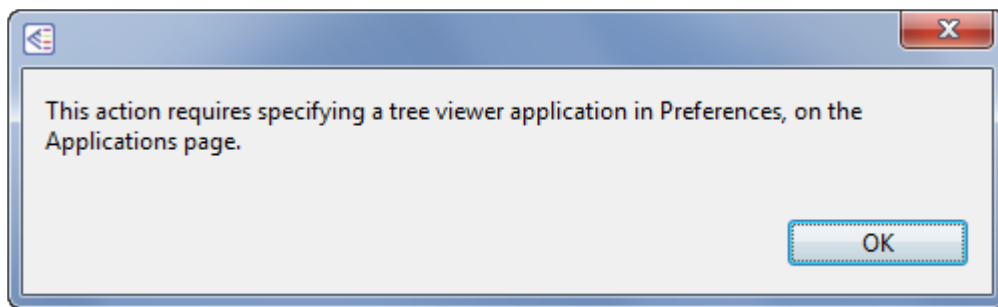
   Click **OK** to launch the Application preferences dialog, where you can indicate the viewer location.

> ✳ **Note:** If you export a tree containing duplicate sequence names, you may receive a warning when you try to open it in a third-party application. To fix this issue, return to MegAlign Pro, rename one of the sequences, and then re-export the tree.

Retain custom names during export

After using **File > Rename** or **File > Rename with Fields**, the custom names are only exported as part of the sequence data in certain cases. Raw data sequences imported into MegAlign are never modified in any way, and always retain their original names.

The following table lists the cases in which custom sequence names are exported:

Export target	Notes
MegAlign	Send an alignment to MegAlign using **File > Send to MegAlign**. Or use **File > Export Data**, save as a MegAlign (*.meg*) file and then open the file in MegAlign. In either case, MegAlign truncates imported labels to a maximum of 39 characters.
Tree viewers	Send an alignment directly to a tree viewer with **File > Send Tree to (Tree Viewer)**. Or save a tree file via **File > Export Data > Tree**, and then open it in a tree viewer. The custom label will be retained.
Sequences and alignments	Custom labels are exported to Nexus, and MegAlign formats. They are not exported to FASTA or GenBank formats.
Distance matrices	Custom labels are exported to distance matrices of all formats (e.g., *.csv*, *.tab*, *.dst*, and *.nex*).

Export data from the Text, Table or Pairwise views

The **Export data** tool (📄), located in the header above the Text, Table or Pairwise views, lets you export data from those views. This tool provides a choice of three options, one of which has four sub-options.

Export Data	>	Matches Without Gaps...
Download Sequences...		Aligned Matches...
Set Access ID...		Table...
		Text...
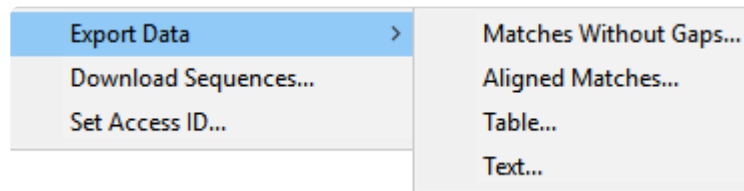
Options include:

- **Export Data >**

  ◦ **Matches without Gaps** – To export an ungapped FASTA (*.fasta*) file consisting of matching regions of displayed matches, but not the query sequence itself.

  ◦ **Aligned Matches** – To export a gapped FASTA (*.fasta*) file with displayed matches aligned against the query.

  ◦ **Table** – To save Table view contents in either Comma- or tab-separated value (*.csv* or *.tab*) formats.

  ◦ **Text** – To save the Text view contents in Text (*.txt*) format.

- **Download Sequences** – To download the selected match sequence in GenBank (*.gbk*) format.

The commands above all launch a Save As dialog. Choose a file name, location and file type (fixed, in some cases) then press **Save**.

- **Set Access ID** – To launch the Access ID dialog, in which you can specify your NCBI Access ID (optional). See Access ID dialog for information.

Print Images

To print an image of any MegAlign Pro view, whether it is currently active or not:

Choose **File > Print** or use **Ctrl/Cmd+P**. The Print Document dialog opens:
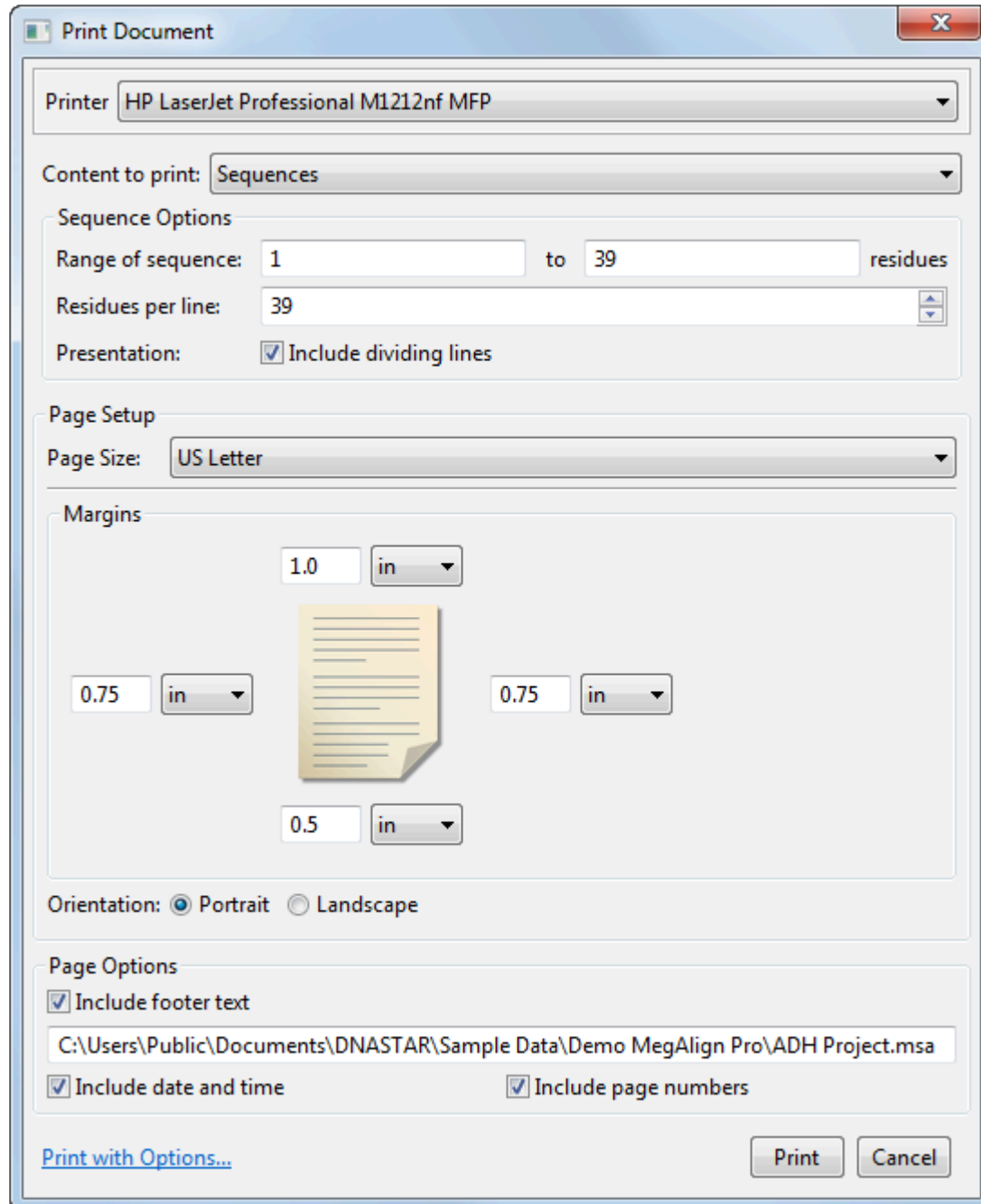
- Use the **Printer** drop-down menu to choose the desired printer from a list of available options.

- Specify the view you wish to print using the **Content to print** menu. Choices may include: **Overview**, **Sequences View**, or **Tree View**. Depending upon the selection, additional options may appear just

below this drop-down menu.

- ◦ **Overview** and **Sequences View** - To print only a sub-range of the alignment, type the left and right residues into the **Range of alignment** boxes (if an alignment has taken place) or **Range of sequence** boxes (if alignment has not taken place).

  ◦ **Sequences** - To specify how many residues to include on one line, type in a number from 1-100 next to **Residues per line**. Check the **Include dividing lines** box if you wish to include separators between the header, sequences, and footer.

  ◦ **Tree** - Check **Fit to page** to automatically size the tree to fit on a single page of the size specified in **Page Size** (see below).

- Choose which **Page Size** to print on. (**Note:** If you wish to change the default print size setting, see Change the default print size.)

- Specify the top, bottom, **Margins** by typing a number in the associated box and choosing the units (**inches**, **cm**, etc.) from the nearby drop-down menu.

- Choose whether to print in **Portrait** or **Landscape** orientation.

- Under **Page Options**, use the checkboxes to specify what to include in the footer: **footer text**, **date and time** and/or **page numbers**. If you check **Include footer text**, the footer will incorporate the text just below the checkbox. By default, this is the path and MegAlign Pro project name (or only the default project name, for unsaved documents), but you may type in any text that you wish.

- Click the **Print with options** link if you want to launch the default Print dialog for your computer.

Choose **Cancel** to exit from the MegAlign Pro Print Document dialog without printing. Click **Print** to print using the selected parameters, or **Print with Options** to open the Print dialog for your computer's operating system. The generic dialog can be used to specify additional options, such as the page range and number of copies to print.

Change the default print size

Unless otherwise specified on your computer, the default printing size for all DNASTAR applications is "U.S. Letter" size (8.5 × 11 inches = 21.6 × 28 cm). You can temporarily change to another size, such as A4, from within the application. However, the next time you print, the default size will again be U.S. Letter size.

To correct this issue, you need to change your computer's default printing size:

On Windows 7 and Windows 8:

1. Open the Control Panel.

2. In the Hardware and Sound category, click on **Printers** or **View devices and printers**.

3. Right-click on your printer and choose **Properties**.

4. In the **General** tab, click on the **Printing Properties** button.

5. In the ensuing **Printing Shortcuts** tab, choose the desired default printing size from the **Printing Sizes** drop-down menu.

6. Click the **Apply** button, then the **OK** button.

On Macintosh:
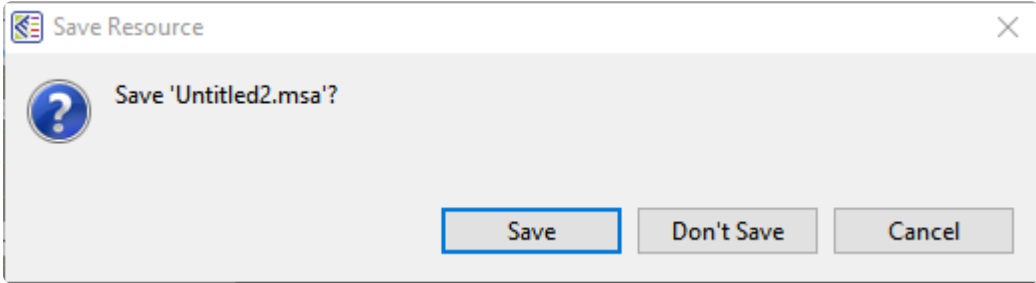
1. Open System Preferences.

2. Click on **Printers and Scanners**.

3. Choose the desired size from the **Default paper size** drop-down menu.

✳ **Note:** Changes to the default printing size will not be applied until after you restart the DNASTAR application.

Save, Close and Exit

The following table describes how to save files, close projects and exit from MegAlign Pro:

Task	How to…
Save all of the open projects simultaneously	Choose **File > Save All** or press **Ctrl/Cmd+Shift+S**. Each project is saved with an *.msa* extension. By default, MegAlign Pro saves to the desktop unless directed to save elsewhere. An auxiliary file folder with the extension *.msa.index* may also be saved.
Save a new project or overwrite a previously saved version of the project	Choose **File > Save**, select the **Save** tool (💾), or use **Ctrl/Cmd+S**. If you want to preserve the earlier version of a project, or store a copy of the project with a different name or location, use **File > Save As** instead. By default, MegAlign Pro saves to the desktop unless directed to save elsewhere. Each project is saved with an *.msa* extension. An auxiliary file folder with the extension *.msa.index* may also be saved.
Close an active project	Choose **File > Close Editor** or **Ctrl/Cmd+W**. If multiple projects are open, first select the tab corresponding to the project you wish to close. To reopen files that you just close, use **File > Reopen Closed File** or **Ctrl/Cmd+Shift+O**.
Exit from MegAlign Pro	On Windows, use **File > Exit.** On Macintosh, use **MegAlign Pro > Quit MegAlign Pro**. The shortcut **Cmd+Q** works on both platforms.
Move or copy a project to another location on your computer	When you save a MegAlign Pro project, two files are saved: the *.msa* project file and and an auxiliary folder with the extension *.msa.index*. If you decide to move or copy the project, we recommend moving/copying both of these files in tandem. If not, you may have to recompute variants when you reopen the saved project file.
Close or quit when there are unsaved documents	If you select **File > Close**, or if you quit MegAlign Pro when just one unsaved project is open, MegAlign Pro will open the Save Resource dialog. Click **Save** to save changes, **Don't Save** to close/quit without saving changes, or **Cancel** to return to MegAlign Pro without closing/quitting. If you select **File > Close**, or if you quit MegAlign Pro when multiple unsaved projects are open, MegAlign Pro will open the Save Documents dialog.
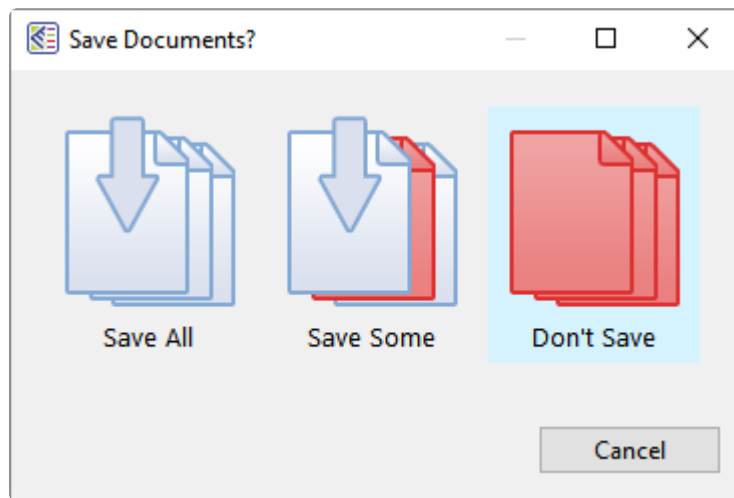
To save all changed projects, choose **Save All**. To choose which projects to save, choose **Save Some**. To exit from MegAlign Pro without saving any projects, choose **Don't Save**. Click **Cancel** to return to MegAlign Pro without saving the checked projects.

Appendix

Turn off usage logging

By default, usage logging is enabled for Lasergene. To opt in/out of usage logging:

1. Launch the DNASTAR Navigator.

2. Use whichever of the following commands is available; the command varies by Lasergene version and operating system.

   - **Window > Preferences**

   - **View > Preferences > Lasergene**

   - **DNASTAR Navigator > Preferences > Lasergene**

   - **DNASTAR Navigator > Settings**

   The following dialog appears:
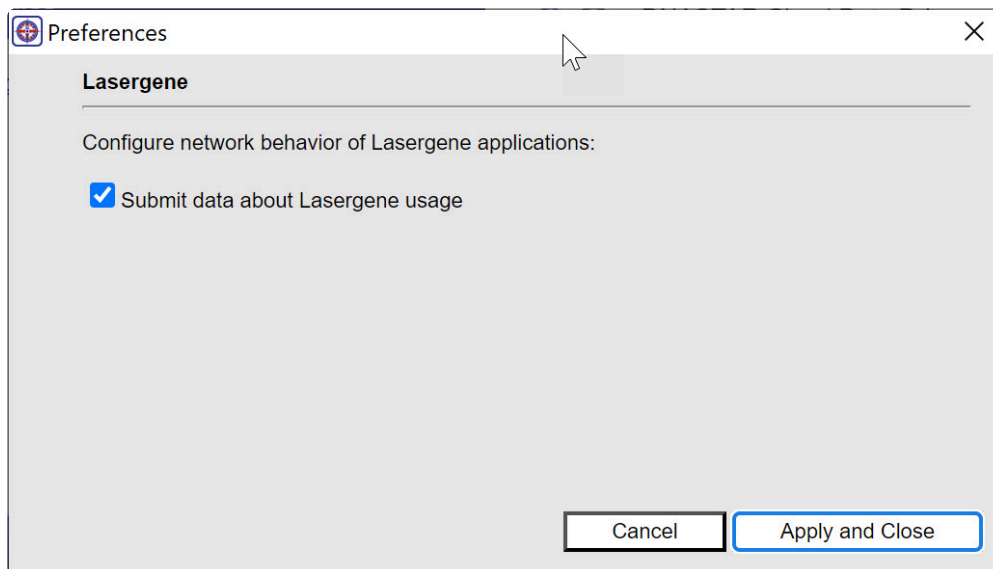
3. Check or uncheck the box, as desired.

4. Click **Apply and Close** or **Apply** and **OK**, depending on which option is available.

Set preferences

To set preferences on Windows, choose **Edit > Preferences**. On Macintosh, choose **MegAlign Pro > Preferences** (macOS 12 and lower) or **MegAlign Pro > Settings** (macOS 13 and higher). Click any of the four items on the left to open the preferences dialog for that item, then follow the steps in the sections below. After setting preferences, click **Apply** and **OK** to save your changes or **Cancel** to ignore any changes. Alternatively, click **Restore Defaults** to return to the default settings for that dialog.

Applications:

The Applications dialog designates an (optional) Newick or Nexus viewer of your choice to use in advanced editing of phylogenetic trees. Note that the download, use, or specification of a tree viewer is optional, and is only necessary if you wish to use MegAlign Pro's **File > Send Tree to [Tree Viewer Name]** command.
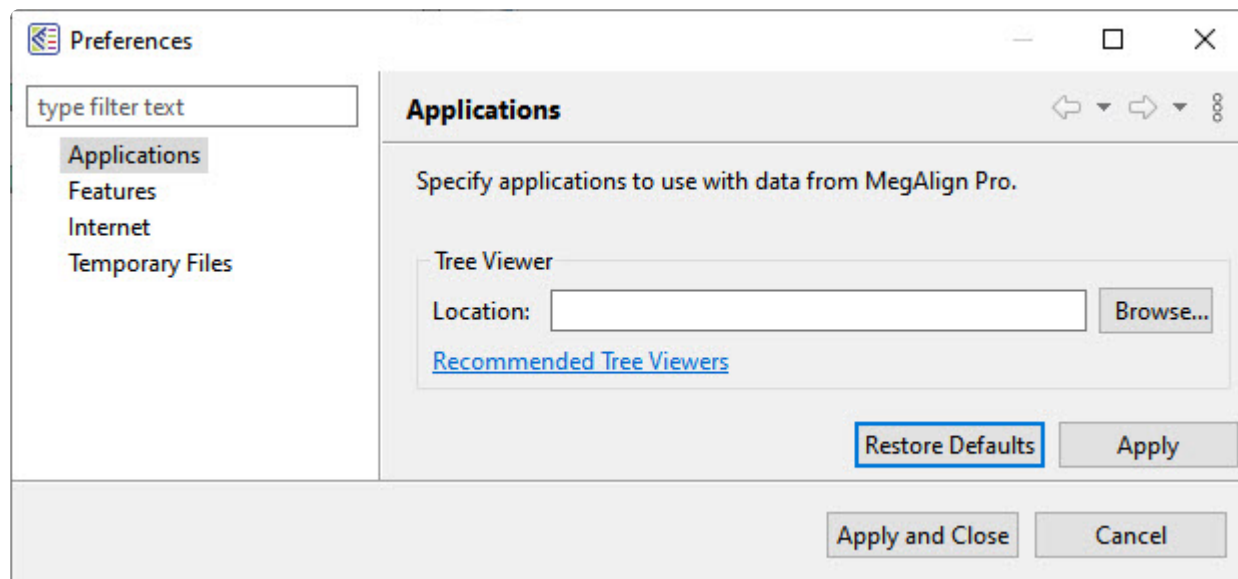
By default, the **Location** box is empty. If you already have tree viewing/editing software on your machine, use the **Browse** button to navigate to the application file, and then click **OK**. Most software that can read a Newick- and/or Nexus-formatted tree file will work here.

If you would like to download free tree viewing/editing software, click the link provided to visit this page of our website. After downloading appropriate software, the **Location** box will usually update automatically. If not, use the **Browse** button as described above.

Features:

The Features dialog lets you specify which features are hidden by default in new documents. By default, this dialog specifies that **exon**, **PDB**, and **STS** features be hidden.
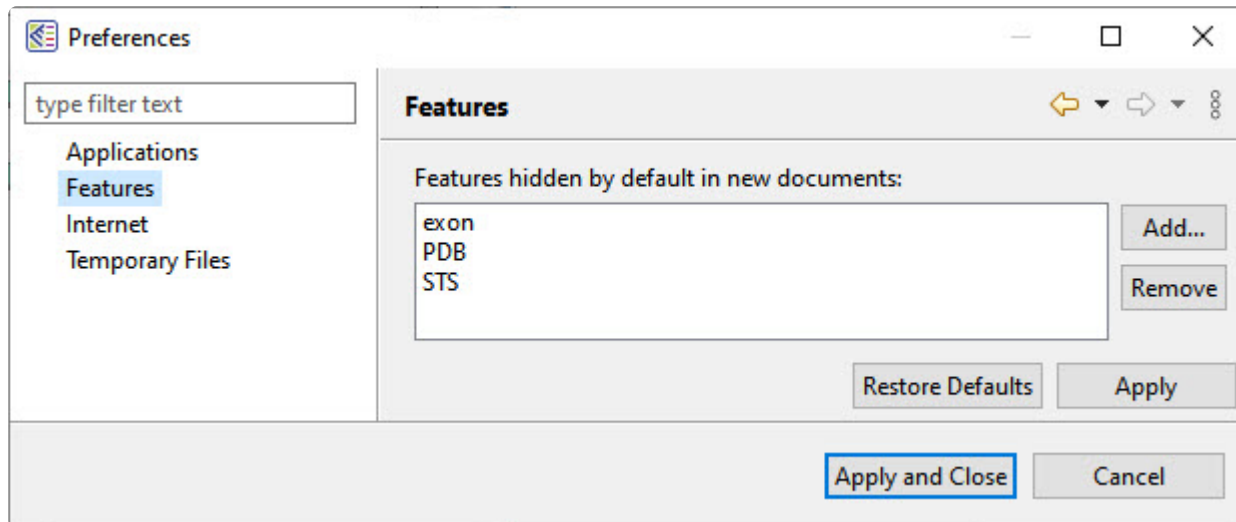
To remove a feature type from the list, select it and choose **Remove**.

To add a feature type to the list, click **Add**. In the ensuing dialog, type the feature type(s) you would like to hide. To specify multiple types, leave a space between each; no punctuation is necessary. Then press **OK**.
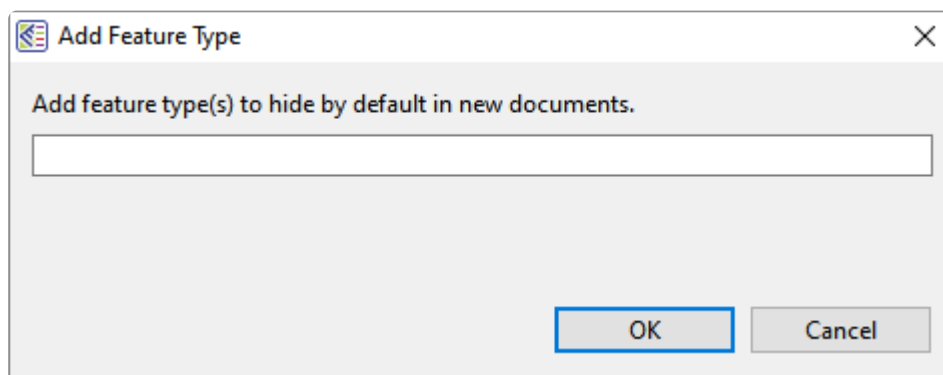
Internet:

The Internet Preferences dialog lets you choose whether you wish to be notified about Lasergene updates.
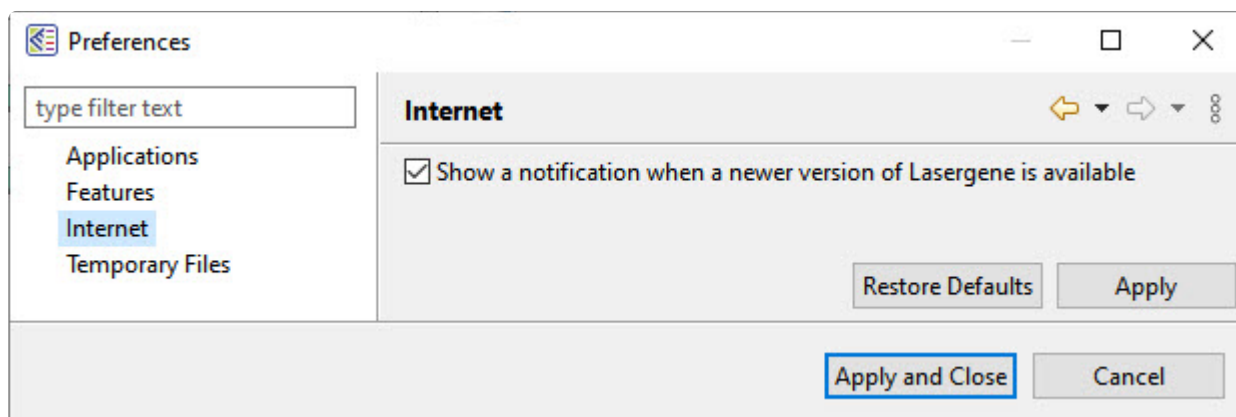
Check the box if you would like MegAlign Pro to display a notification when a newer version of Lasergene is available. Uncheck the box if you do not wish to receive this notification.

Temporary Files:

MegAlign Pro creates and uses temporary files while a project is open. The need for temporary file space grows with the number and length of the sequences and amount of track data in open alignment projects. If you expect the temporary files will be too large to fit in the default location, you may specify a different location for these files using the Temporary Files Preferences dialog.
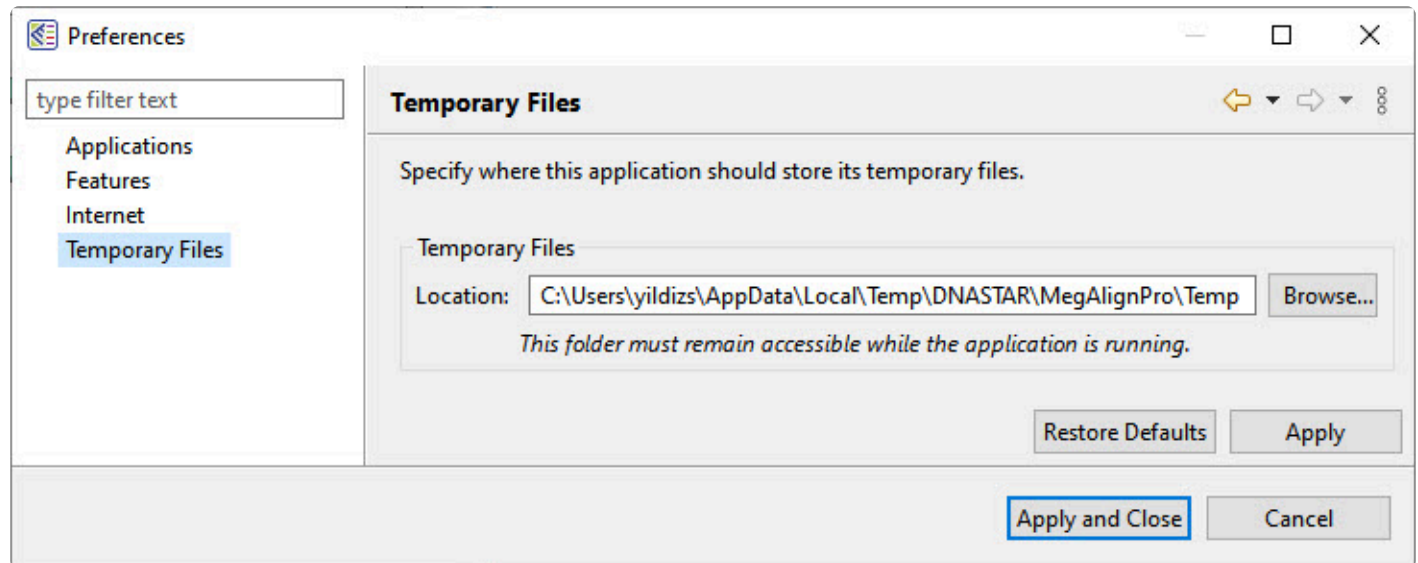
To change the temporary file location, close all MegAlign Pro projects and then return to this dialog. Select the **Browse** button, navigate to the new location and click **OK**. As noted in the dialog, the location you choose must remain accessible while MegAlign Pro is running. The new location will be used for all projects that are subsequently created or opened.

Change the default application for opening files

Windows 7 and Windows 8:

1. Open the **Control Panel**.

2. Go to **Programs > Default Programs > Associate a file type or protocol with a specific program.**

3. Select extension you wish to change in the **Extensions** list and then click the **Change program** button.

4. Click the **Browse** button in the Open With dialog box.

5. Navigate to and select the application you want the files to open in. (Lasergene applications are located by default in **C:\Program Files\DNASTAR\Lasergene 'x'** or **C:\Program Files (x86)\DNASTAR\Lasergene 'x'**, where 'x' represents the version number).

6. Click **Open**.

7. Click **OK** to close the Open With dialog box, and then click **Close**.

Macintosh:

1. Select a file of the type you wish to change in Finder.

2. Go to **File > Get Info**.

3. Click the dropdown arrow in the **Open With** field, and navigate to the application you want the files to open in. (Lasergene applications are located by default in **Applications:DNASTAR:Lasergene 'x'**, where 'x' represents the version number).

4. Click the **Change All** button to make the application you selected the default for all files of this type.

Clean cached files on startup

Prior to the Lasergene 14.0 release, the default MegAlign Pro startup ran a "*-clean*" option. This option automatically removed some cached files on application restart.

The "-clean" option is no longer executed at each startup. However, if you are a Windows user, you may execute an explicit "-clean" in the unusual event that cached data has become corrupted:

1. Close MegAlign Pro if it is currently open.

2. Launch the Command Line (C:\Windows\system32\cmd.exe).

3. Type the path to the MegAlign Pro application, followed by a space and the *–clean* command.

    C:\Program Files (x86)\DNASTAR\DNASTAR\Lasergene 14\MegAlign Pro.exe –clean

Troubleshoot the error "A different version is running"

If you receive the error "A different version of Lasergene is running" when you try to launch a Lasergene application, but another version of Lasergene actually isn't running, there may be a lingering .state file from another version that needs to be deleted.

The solution is to locate and delete files named *STARDM*.state*, …where '*' represents the version number.

These files may exist in the following directories:

Windows 7 and Windows 8:

- *C:\Users\username\AppData\Local\DNASTAR\DataManager*

- *C:\ProgramData\DNASTAR\DataManager*

Macintosh:

- *Hard Drive:Library:Preferences:DNASTAR:DataManager*

- *Hard Drive:Users:username:Library:Preferences:DNASTAR:DataManager*

If you find any of the *.state* files listed above, delete them. Once the files are deleted, the error should no longer appear.

Supported file types

For a list of file formats supported by this DNASTAR product, please see the File Formats page of our website.

IUPAC codes

IUPAC codes for amino acids:

1-Letter Code	3-Letter Code	Amino Acid
A	Ala	Alanine
B	Asx	Aspartic Acid or Asparagine
C	Cys	Cysteine
D	Asp	Aspartic Acid
E	Glu	Glutamic Acid
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
J	Xle	Leucine or Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
O	Pyl	Pyrrolysine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
U	Sec	Selenocysteine
V	Val	Valine
W	Trp	Tryptophan
X	Xaa	Unspecified or unknown
Y	Tyr	Tyrosine

Z	Glx	Glutamic Acid or Glutamine

IUPAC codes for nucleotides:

Symbol	Description	Bases Represented			
A	adenosine/adenine	A			
C	cytidine/cytosine		C		
G	guanosine/guanine			G	
T	thymidine/thymine				T
U	uridine/uracil				U
W	weak	A			T
S	strong		C	G	
M	amino	A	C		
K	keto			G	T
R	purine	A		G	
Y	pyrimidine		C		T
B	not A		C	G	T
D	not C	A		G	T
H	not G	A	C		T
V	not T	A	C	G	
N or -	any base (not a gap)	A	C	G	T

✳ The non-standard nucleotide codes "I" (Inosine) or "X" (xanthine) are not supported.

Installed Lasergene file locations

The following file names use 'x' to represent the version number.

File Category	Application	Path
Application[2]	ArrayStar	Windows: C:\Program Files (x86)\DNASTAR\Lasergene x\ArrayStar
	SeqNinja (command line)	Windows: C:\Program Files (x86)\DNASTAR\Lasergene x\SeqNinjaCL
	All others	Windows: C:\Program Files (x86)\DNASTAR\Lasergene x Macintosh: /Applications/DNASTAR/Lasergene x
Data Manager[2] (*DMx, DMx.exe*)	SeqBuilder Pro, Protean 3D, GeneQuest, MegAlign Pro	Windows: C:\Program Files (x86) \DNASTAR\Lasergene x Macintosh: /Applications/DNASTAR/Lasergene x
Data Manager State File	SeqBuilder Pro, Protean 3D, GeneQuest, MegAlign Pro	Windows: C:\Program Data\DNASTAR\DataManager Macintosh: ~/Library/Application Support/DNASTAR/DataManager, ~/Library/Preferences/DNASTAR/DataManager
License Manager[2]	All	Windows: C:\Program Files (x86)\DNASTAR\License Manager Macintosh: /Applications/DNASTAR/License Manager
Server License File (*lservrc*), Server License Manager [2]*, Server Executables (_lserv, lservnt.exe)*	All	Windows: C:\Program Files (x86)\DNASTAR-LicenseServer\Server

		Macintosh: ~/Library/ DNASTAR-LicenseServer
Standalone & Trial Licenses (*.license), License Server Client License (*.lshost), Key Server Client License (*.keyhost)	All	Windows: C:\Program Data\ DNASTAR\Licenses Macintosh: ~/Library/ Application Support/ DNASTAR/Licenses
Preferences	Protean 3D, Navigator, SeqNinja (DNA*), GenVision Pro	Windows: C:\Users\<user>\DNASTAR Macintosh: ~/Library/ DNASTAR
	ArrayStar	Windows: C:\Users\<User>\AppData\ Roaming\DNASTAR\ArrayStar
	All others	Windows: C:\Users\<user>\AppData\ Local\DNASTAR\ Macintosh: ~/Library/ Preferences

✱ AppData is a hidden folder in Windows. To unhide the folder, go to **Organize > Folder** and **Search Options > View > Show Hidden files and folders**.

Research references

Chou PY and Fasman GD (1978). "Prediction of the secondary structure of proteins from their amino acid sequence." Advances in Enzymology 47: 45-148.

Darling AE, Mau B, and Perna NT (2010). "progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss, and Rearrangement." PLoS One. 5(6):e11147. (See the PLOS ONE entry.)

Dayhoff MO, Schwartz R and Orcutt BC (1978). "A model of evolutionary change in proteins". Atlas of protein sequence and structure (volume 5, supplement 3 ed.). Nat. Biomed. Res. Found. pp. 345–358. ISBN 0-912466-07-3.

Deléage G and Roux B (1987). "An algorithm for protein secondary structure prediction based on class prediction." Protein Engineering 1: 289-294.

Edgar RC (2004). "MUSCLE: multiple sequence alignment with high accuracy and high throughput." Nucleic Acids Res. 32(5):1792-1797. (See the PubMed entry).

Edgar RC (2004). "MUSCLE: a multiple sequence alignment method with reduced time and space complexity." BMC Bioinformatics (5):113 (See the PubMed entry).

Eisenberg D, Weiss RM and Terwilliger TC (1984). "The hydrophobic moment detects periodicity in protein hydrophobicity." Proceedings of the National Academy of Science (USA), 81(1): 140-144.

Garnier J, Gibrat JF, Robson B (1996). "GOR method for predicting protein secondary structure from amino acid sequence." Methods Enzymol. 266:540-53. [Reference pertains to MegAlign Pro's GOR IV algorithm].

Garnier J, Osguthorpe DJ and Robson B (1978). "Analysis of the accuracy and implications of simple method for predicting the secondary structure of globular proteins." Journal of Molecular Biology, 120: 97-120. [Reference pertains to MegAlign Pro's Garnier-Robson method].

Garnier J and Robson B (1990). "The GOR Method for Predicting Secondary Structures in Proteins." In Prediction of Protein Structure and the Principles of Protein Conformation, ed. G.D. Fasman (2nd ed). Plenum Press, New York. [Reference pertains to MegAlign Pro's GOR II algorithm].

Gascuel O (1997). " BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data." Molecular Biology and Evolution 14:685-695. (See the PubMed entry.) [[MegAlign Pro's tree-building algorithm]

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD and Bairoch A (2005). "Protein Identification and Analysis Tools on the ExPASy Server." In The Proteomics Protocols Handbook, ed. John M. Walker, Humana Press, pp. 571-607.

Guruprasad K, Reddy BVB and Pandit MW (1990). "Correlation between stability of a protein and its

dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence." Protein Eng., 4, 155 – 161.

Henikoff S & Henikoff JG (1992). "Amino Acid Substitution Matrices from Protein Blocks". PNAS. 89 (22): 10915–10919. (See the PubMed entry)

Hopp TP and Woods KR (1981). "Prediction of protein antigenic determinants from amino acid sequences." Proceedings of the National Academy of Science, 78: 3824-3828.

Jameson BA and Wolf H (1988). "The antigenic index: a novel algorithm for predicting antigenic determinants." CABIOS, 4: 181-186.

Kabsch W and Sander C (1983). "Dictionary of protein secondary structure: pattern recognition of hydrogen bonded and geometrical features." Biopolymers, 22: 2577-2637.

Katoh M, Kumar M (2002) "MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform." (Nucleic Acids Res. 30:3059-3066).

Kimura M (1980). "A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences." J. Mol. Evol., 16: 111-120.

Kimura M (1983). The Neutral Theory of Molecular Evolution. Cambridge University Press, Cambridge.

Kyte J, Doolittle RF. (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol., 157, 105-132. [Hydrophobicity color scheme]

Lehninger A (2005). Principles of Biochemistry, 4th Edition, Chapter 3, page 78, Table 3-1.

Lehninger A, Nelson D, and Cox M (2005). "Principles of biochemistry." New York: W.H. Freeman.

Müller T, Spang R, Vingron M (2002). "Estimating amino acid substitution models: a comparison of Dayhoff's estimator, the resolvent approach and a maximum likelihood method." Mol Biol Evol. 2002 Jan;19(1):8-13. (See the PubMed entry)

Müller T & Vingron (2000). "Modeling amino acid replacement." J Comput Biol. 2000;7(6):761-76. (See the PubMed entry)

Nakashima H, Nishikawa K and Ooi T (1986). "The folding type of a protein is relevant to the amino acid composition." Journal of Biochemistry (Tokyo), 99: 153-162.

Nishino, Kunihiko & Yamasaki, Seiji & Hayashi-Nishino, Mitsuko & Yamaguchi, Akihito. (2010). "Effect of NlpE Overproduction on Multidrug Resistance in Escherichia coli." Antimicrobial agents and chemotherapy. 54. 2239-43. 10.1128/AAC.01677-09. [View on PubMed]

Parker JMR, Guo D and Hodges RS (1986). "New hydrophilicity scale derived from high-performance liquid

chromatography peptide retention data: Correlation of predicted surface residues with antigenicity and x-ray-derived accessible sites." Biochemistry, 25: 5425-5432.

Saito N and Nei M (1987). "The neighbor-joining method: a new method for reconstructing phylogenetic trees." Mol Biol Evol. 1987 Jul;4(4):406-25.

Schneider TD and Stephens RM (1990). "Sequence logos: a new way to display consensus sequences." Nucleic Acids Research, 18 (20): 6097-6100 doi:10.1093/nar/18.20.6097. (See the PubMed entry) [Sequence Logo track]

Sievers F, Wilm A, Dineen DG, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins D (2011). "Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega." Molecular Systems Biology 7, Article number: 539 (See the article in Nature)

Sonnhammer ELL and Hollich V (2005). "Scoredist: A simple and robust protein sequence distance estimator." BMC Bioinformatics 2005, 6:108. (See on Biomed Central.) [Scoredist model]

Stamatakis A (2014). "RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies". Bioinformatics, 2014. [RAxML tree-building algorithm]

Tamura K and Nei M (1993). "Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees." Mol Biol Evol. 1993 May;10(3):512-26.

Taylor WR (1997). "Residual colours: a proposal for aminochromography." Protein Engineering vol.10 no.7 pp.743 – 746. (See the PDF file.) [Taylor color scheme]

Thompson JD, Higgins DG and Gibson TJ (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Research, 22:4673-4680. [Clustal X color scheme]

Thorne JL, Kishino H, Painter IS (December, 1998). "Estimating the rate of evolution of the rate of molecular evolution". Mol. Biol. Evol. 15 (12): 1647–57. (See the PubMed entry the PubMed entry)

Troshin PV, Procter JB and Barton GJ (2011). "Java Bioinformatics Analysis Web Services for Multiple Sequence Alignment – JABAWS:MSA." Bioinformatics; doi: 10.1093/bioinformatics/btr304.

Waterhouse AM, Procter JB, Martin DMA, Clamp M and Barton GJ (2009). "JalView version 2: a multiple sequence alignment and analysis workbench." Bioinformatics 25:1189 – 1191. (See on PMC.) [Discusses methodology later used to create the Taylor and Zappo color schemes]

Welling GW, Weijer WJ, van der Zee R and Welling-Wester S (1985). "Prediction of sequential antigenic regions in proteins." FEBS Letters, Vol. 188, No. 2: 215-218.

Yang ZR, Thomson R, McMeil P and Esnouf RM (2005). RONN: the bio-basis function neural network

technique applied to the detection of natively disordered regions in proteins. Bioinformatics 21: 3369-3376.