

User Guide to



MegAlign Pro

DNASTAR, Inc. 2020

Table of Contents

Welcome to MegAlign Pro.....	5
MegAlign Pro Tutorials	6
Create or Open a Project.....	7
Navigate between projects	9
MegAlign Pro Interface.....	10
Views	12
Overview	14
Overview prior to performing an alignment	16
Overview after aligning multiple genomes with Mauve	17
Overview after other alignments	19
Overview after aligning and entering additional sequences.....	20
Using the Overview as a sequence navigation tool	21
Sequences view.....	22
Pairwise view	24
Distance view	27
Tree view.....	29
Text view	32
Table view	34
Change the appearance of a view	36
Split and resize panes in a view.....	38
Zoom in or out on a view.....	40
Panels.....	42
Style panel.....	43
Overview section	45
Sequence section	47
Available color schemes.....	48
Multiple Alignment section	52
Pairwise Alignment section	54
Distance section	57
Distance metric options	58
Gap treatment options.....	59
Tree section	60
Tracks panel	62
Layout section	64
Options section	65
Features panel.....	66
Places panel	67
Favorites section	69

Recent section	70
Details panel.....	71
When no document, or a blank document, is open.....	73
When a non-empty document is open.....	74
When one sequence is selected	75
When a portion of a sequence is selected	77
When multiple sequences are selected.....	78
When a portion of multiple sequences is selected.....	79
When one track is selected.....	80
When multiple tracks are selected	81
When one feature is selected	82
When multiple features are selected.....	84
When a portion of a pairwise alignment is selected.....	85
Jobs panel	86
Console panel.....	88
Hide, show, move, or resize items in the MegAlign Pro window.....	90
Sequences	91
Add ungapped sequences to a project	92
Advantages of using SeqNinja files.....	95
Add gapped sequences to a project	97
Add multi-segment sequences to a project.....	98
Try it! – Add and align multi-segment sequences	100
Add sequences from a saved project to an active project	101
Manually specify sequence type.....	102
Change the order of sequences	103
Rename sequences manually	104
Rename sequences automatically using specified data fields.....	105
Remove sequences from a project	107
Tracks	109
Ruler tracks.....	112
Features tracks	114
Consensus track	117
Consensus Match track.....	118
Sequence Logo track	120
GC Content track	122
Gap Fraction track	124
Numeric tracks	126
Match Bar track.....	129
Translation track	131
Features	133
Map features (copy features to another sequence).....	134

Features screen.....	137
Options screen	139
Output screen	141
Output files	142
Mapping features in MegAlign Pro vs. SeqNinja.....	143
Remove features.....	144
Try it! – Map features	145
Part A: Add, rename and align sequences	146
Part B: Map a single feature	147
Part C: Map all features	149
Part D: Map a filtered set of features	150
Part E: Compare output files	152
Part F: Export the annotated target sequence	153
Search Within an Open Project.....	154
Make a Selection	156
Choose the Alignment Type.....	158
Search for Sequences Online	160
Search for sequence matches (BLAST).....	161
Search wizard: Query	162
Search wizard: Search.....	165
Search wizard: Options.....	167
Search wizard: Job	170
Search for text or accession number matches (Entrez)	171
Filter sequence search results	174
Filter text search results	176
Access ID dialog	177
Perform a Multiple Alignment	178
Perform an initial multiple alignment.....	180
Notes regarding multiple alignments	182
Troubleshoot multiple alignments.....	184
Multiple alignment methods and options.....	187
Clustal Omega alignment options	189
Clustal W alignment options.....	192
MAFFT alignment options	195
MUSCLE alignment options	198
Mauve alignment options	202
Modify a multiple alignment.....	204
Subalign sequences.....	205
Merge two existing alignments (“Profile alignment”)	207
Merge unaligned sequences into an existing alignment	209

Align selected unaligned sequences with an existing alignment	210
Unalign aligned sequences	211
Try it! – Perform a Clustal Omega alignment	213
Part A: Add and align sequences	214
Part B: Navigate using the Overview	215
Part C: Change sequence names and rendering in the Sequences view	216
Part D: Change metrics and rendering in the Distance view	220
Part E: Change the appearance of the Tree view	222
Try it! – Perform a MUSCLE alignment with multi-segment sequences	224
Part A: Add and align multi-segment files	225
Part B: Replace a sequence in the Overview and view the phylogenetic tree	229
Try it! – Perform a genomic alignment with Mauve	232
Part A: Add genomic sequences using drag & drop	233
Part B: Perform a Mauve alignment using modified parameters	235
Part C: View a genomic alignment in the Overview	237
Part D: View the phylogenetic tree	240
Perform a Pairwise Alignment	242
Comparison of pairwise alignment methods	245
Substitution matrices	247
Try It! – Follow a multiple alignment with Global pairwise alignments	249
Try It! – Align transcripts to genes using Local and Global pairwise alignments	252
Part A: Compare results from three multiple alignment methods	253
Part B: Use a Local pairwise alignment method	255
Part C: Use a Global pairwise alignment method	257
Try It! – Use Local pairwise alignment to find a gene within a genome	258
Copy, Paste and Delete	262
Export	263
Export an image of the view	264
Export data to a file	266
Export data to MegAlign	269
Export a tree to a tree viewer	270
Retain custom names during export	272
Export data from the Text, Table or Pairwise views	273
Print Images	274
Change the default print size	276
Save, Close and Exit	277
Appendix	279
Turn off usage logging	280
Set preferences	281


Set application preferences	282
Set Internet preferences	283
Set temporary file preferences	284
Change the default application for opening files	285
Clean cached files on startup	286
Troubleshoot the error “A different version is running”	287
Supported file types	288
IUPAC codes	289
Installed Lasergene file locations	291
Research references	293

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.

The following video is a brief walk-through of the MegAlign Pro application. Note that this video does not show all available alignment methods.

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**.

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](#)

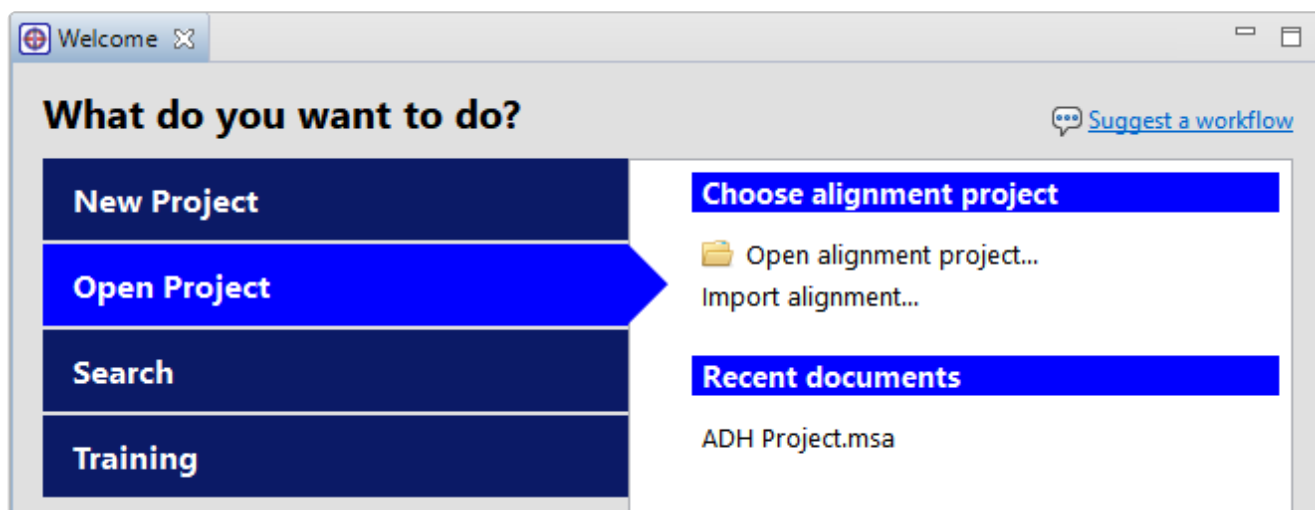
Create or Open a Project


The Welcome “project” opens when you launch MegAlign Pro, and is used for creating or opening projects. As you add projects, the Welcome “project” will remain as the leftmost tab unless you manually move or close it. Note that 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.

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.




- Choose **File > Open** from the menu.
- Click the **Open a file** tool ()
- Press **Ctrl/Cmd+O**.

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 choose. 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.
- 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**.

To create a new project based on a classic MegAlign project file or a gapped FASTA file:

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. The file must be in MegAlign (*.meg*) or gapped FASTA format (*.fasta*, *.fas*, *.fap*, etc.).
3. Click **Open**.

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:

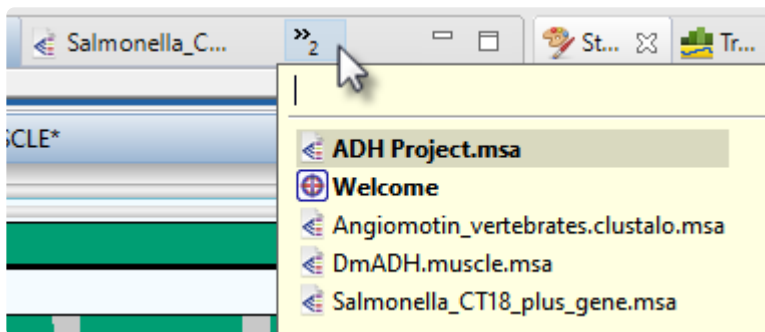
Select the **Window > New Window** command.

To navigate between open projects:

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

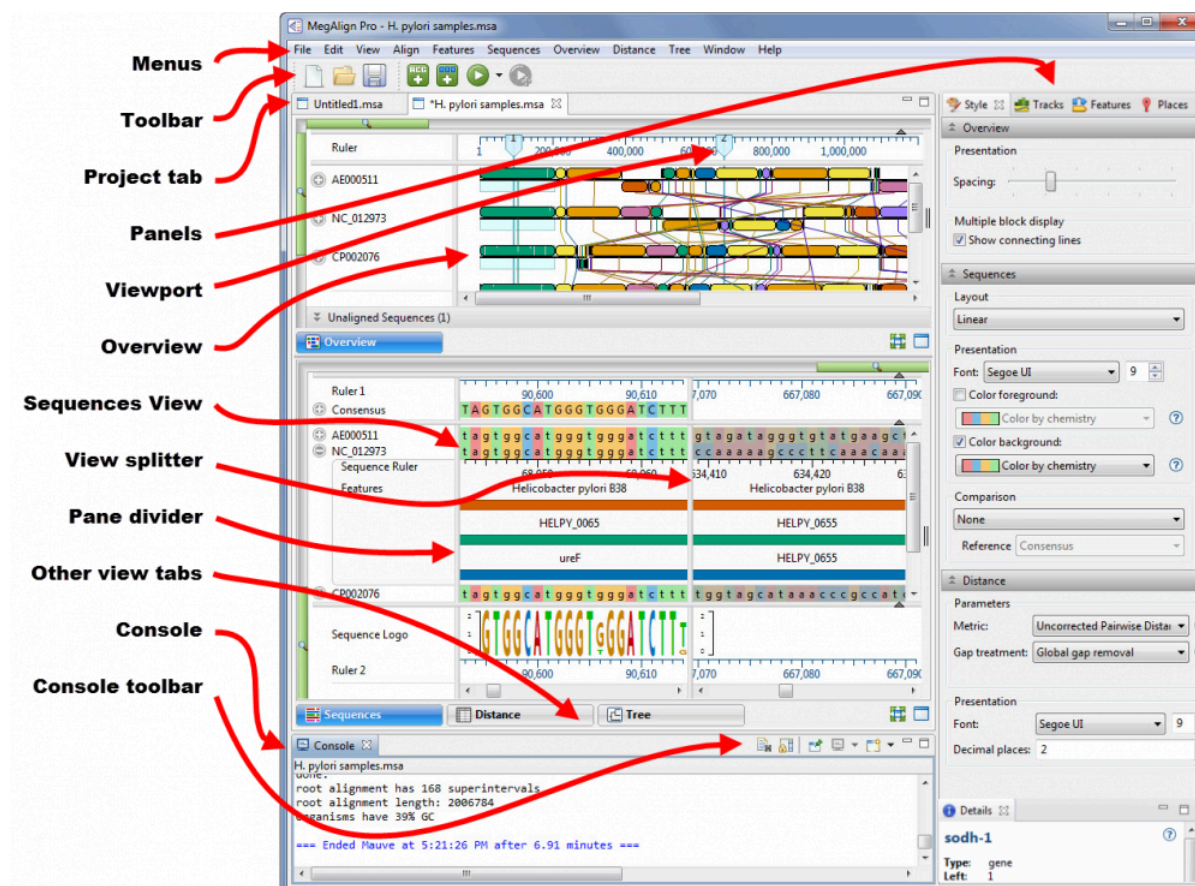
One tab, the Welcome screen, is always present by default unless it has been manually closed. To make another project active, click on its tab.

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 <i>.msa</i> 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.

Views

The majority of the MegAlign Pro window is comprised of one or two “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 change the look of a view:

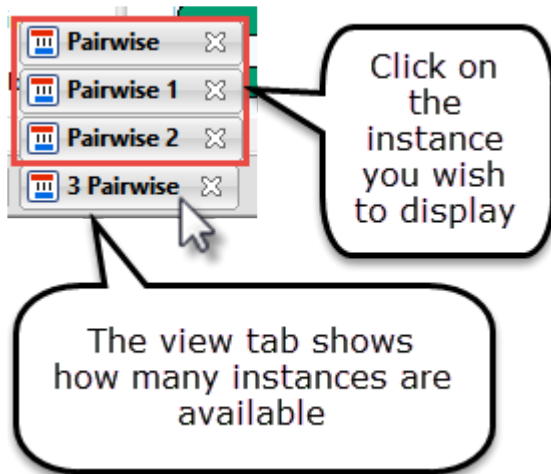
- [Change the appearance of a view](#)
- [Split and resize panes in a view](#)
- [Zoom in or out on a view](#)

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:

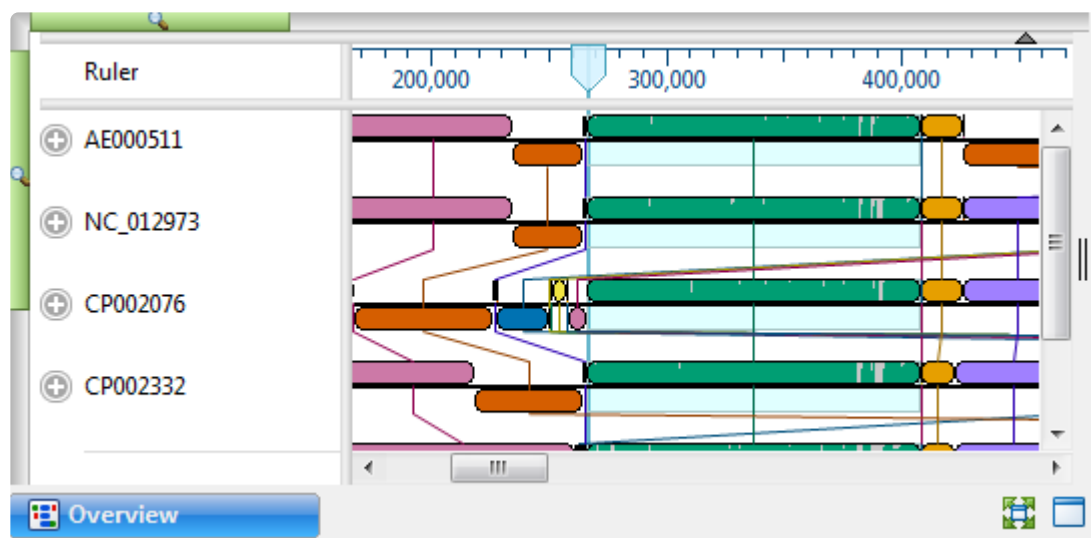
1. Hover over the view tab until you see a popup showing each instance by name.
2. Click on the desired instance.



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.



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](#)

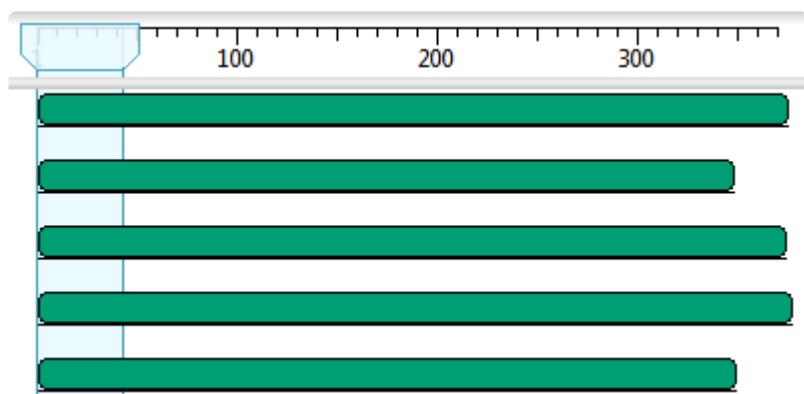
The following table describes ways to control the appearance of the Overview:

Task	How to...
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 .
To 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.
To change settings related to the Overview	use the Overview section of the Style panel .
To 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.
To center the Overview display over the active block	Choose Overview > Show Active Block .
To 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.

Overview prior to performing an alignment

The [Overview](#) has different appearances, depending on the stage and workflow.

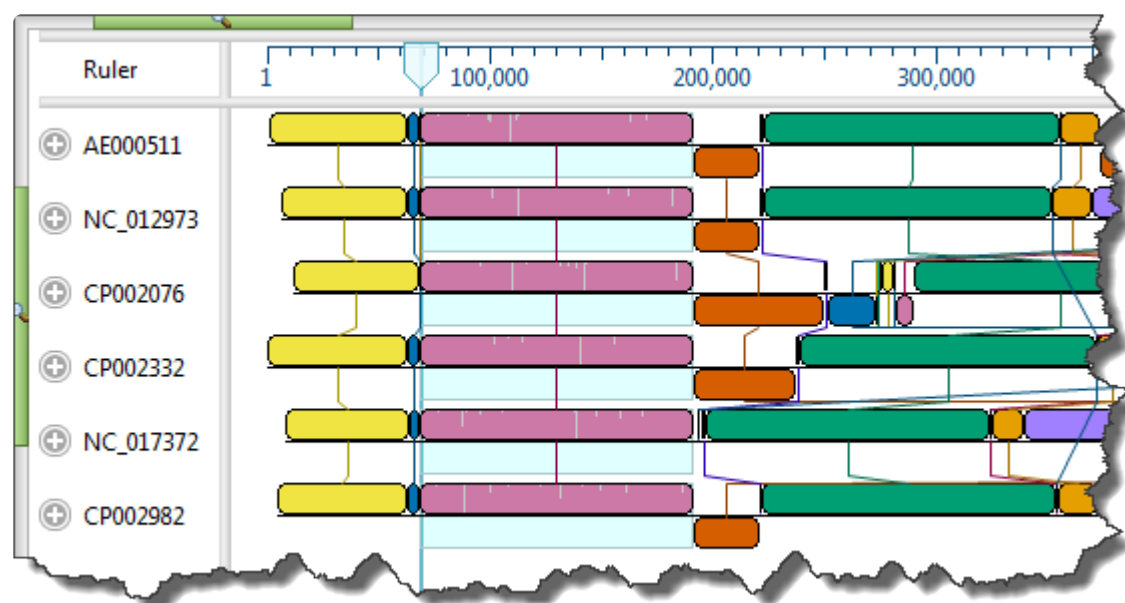
Prior to alignment, each sequence in the Overview appears as a green rectangle (“sequence block”) with no gaps. The length of each block is proportional to the ungapped length of its corresponding sequence.



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.

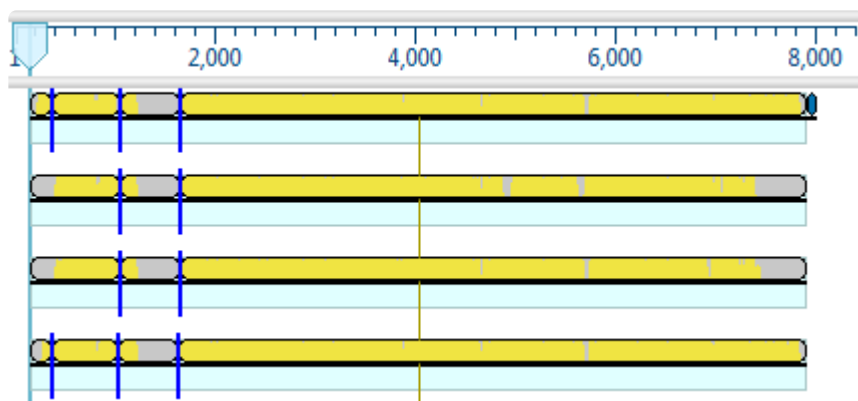


To orient the alignment around an active block, double-click on the sequence block. If necessary, individual sequences flip (i.e., are reverse complemented) to orient properly for the new alignment. Sequence rulers reorient along with the strands, and display the new coordinates. Gaps in the active block 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.

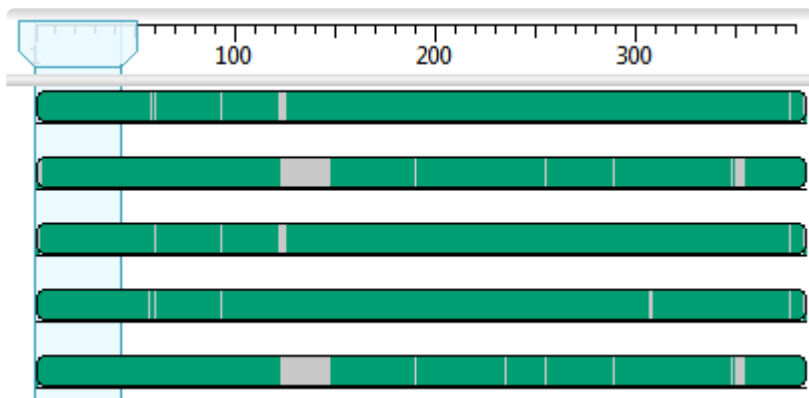


Overview after other alignments

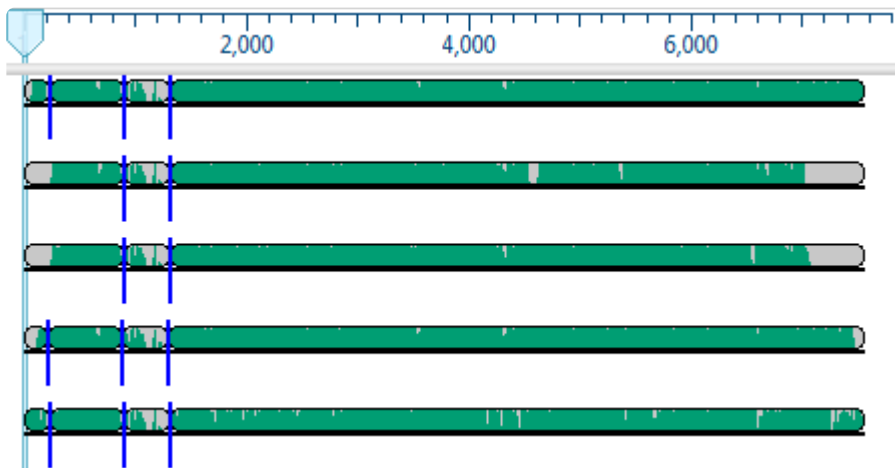
The [Overview](#) has different appearances, depending on the stage and workflow.

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.



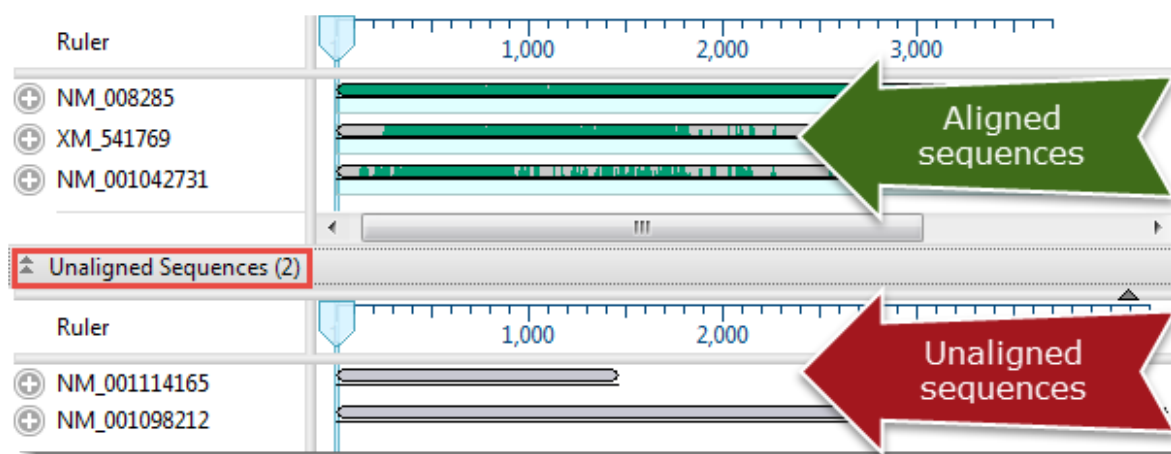
After aligning sequences end-to-end (see Add Multi-Segment Sequences), vertical blue bars are used to show the beginnings and ends of the component sequences (i.e., the segments). These bars are displayed regardless of the type of alignment performed.



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:

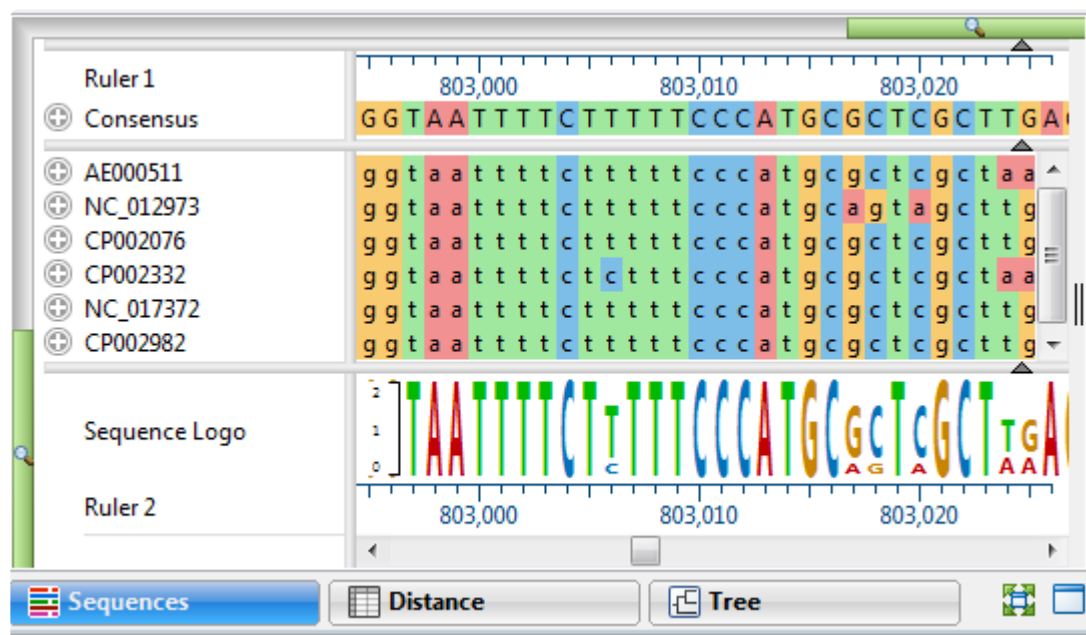
- In any project, move the slider to a desired region by dragging it with the mouse.
- In a single-block project, double-click anywhere on a block to center the slider over that position.
- In a multi-block project, double-click on an un-highlighted block to make it the active block and to move the slider 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

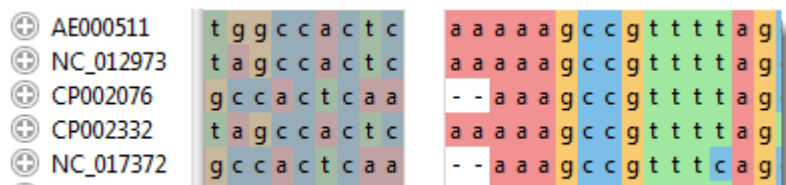
The Sequences View displays added sequences and aligned sequences. Sequence names selected in this view are simultaneously selected all other views in MegAlign Pro. Similarly, sequence names selected elsewhere will be highlighted in the Sequences View.

To access the view, click on the **Sequences** tab.



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](#).

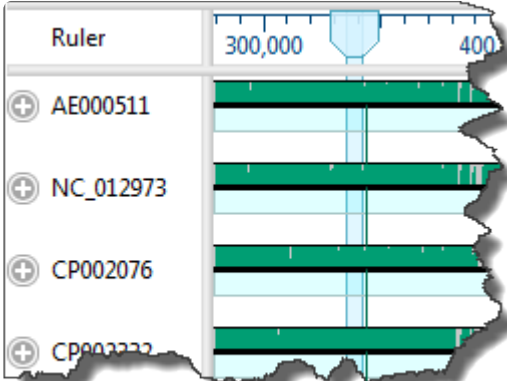

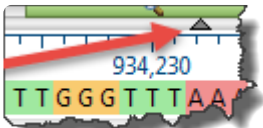
When viewing a multi-block alignment produced by [Mauve](#) and using a minimal horizontal zoom, the currently-selected alignment block will appear the same color as in the [Overview](#). Flanking sequences in adjacent blocks will appear gray. At higher zoom levels, the residues for the active block 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.

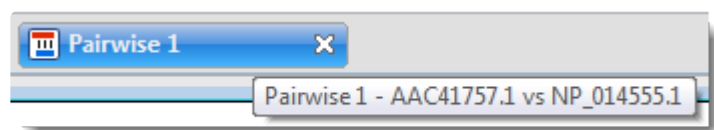
The following table describes ways to control the appearance of the Sequences view:

Task	How to...
To change settings related to the Sequences view	Use the Sequences section of the Style panel . The Sequences > Color Sequence Foreground (Background) commands act as alternatives to checking/unchecking the Color foreground (background) boxes in the that section.
To control the portion of sequence displayed in the view	<p>The Overview's "viewport indicator" highlights the portion of sequence displayed in the displayed in the Sequences View. You can drag the indicator to scroll the Sequences view along the sequence. If you split the Sequences view, there will be more than one viewport indicator in the Overview. If you split the Overview, a viewport indicator may appear in more than one of the sub-panes.</p> 
To show/hide a vertical guideline	<p>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.</p> 
To show/hide the header, sequences area, or footer in the Sequences view	<p>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.</p> 

Pairwise view

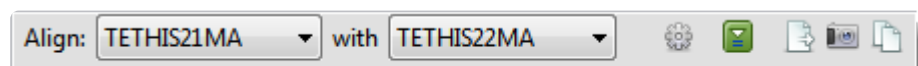
The Pairwise view displays sequences that have been aligned pairwise. Sequence names selected in this view are simultaneously selected all other views in MegAlign Pro. Similarly, sequence names selected elsewhere will be highlighted in the Pairwise view.

The Pairwise view opens automatically after you [perform a pairwise alignment](#). To return to the view at a later time, click on the **Pairwise** tab. If there are multiple Pairwise tabs, hover over each tab to see a tooltip showing which sequences were involved in that alignment.








By default, the Pairwise view consists of three areas: toolbar, summary, and sequences.

Toolbar area:



This area consists of two drop-down menus and five tools.

Tool	Description
Align and with drop-down menus	The left drop-down menu is used to specify the target sequence, which should be the longest sequence of the two. The right drop-down menu is used to specify the query sequence. Specifying a different sequence in either menu will initiate a new pairwise alignment.
Change alignment options in this view 	To replace the current pairwise alignment, use this tool to open the pairwise options dialog. Set options as desired and press Align .
Create new pairwise alignment from selection 	Select a portion of the alignment and use this tool to perform a subalignment. The subalignment will open in a new Pairwise view.
Export data	Shortcut to File > Export Data > Pairwise . The summary and sequences areas are both exported.

	
Export image 	Shortcut to File > Export Image > Pairwise . The summary and sequences areas are both exported.
Clone this view 	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.

Summary area:

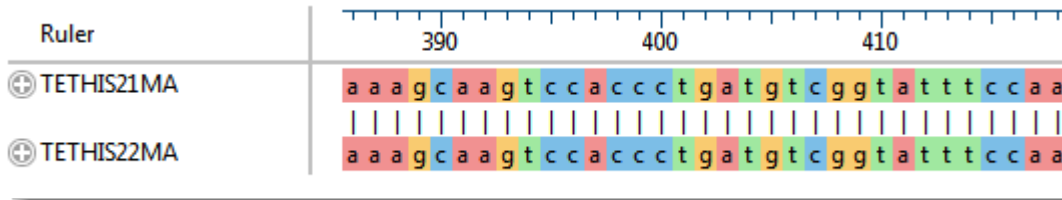
Local: Smith-Waterman DNA alignment [Matrix: "NUC44" Gap penalty: 10 Gap extension penalty: 1]

	TETHIS21MA 1>906	TETHIS22MA 1>905	%Identity	%Gaps	Identical	Gap Count	Gap Length	Score	Length
Alignment	13>862	1>902	69.2%	19.2%	671	46	186	2,307	969
Selection	24>264	8>263	60.4%	25.6%	172	20	73	447	285

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

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 sequence” first, and “query sequence” second.

A variety of optional [data tracks](#) can be displayed in this section. The Pairwise view has a different appearance depending on the [zoom level](#).

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.

Block: Metric: Uncorrected Pairwise Distance Sequences: 6 Global gap removal Residues considered: 11,748

		A	B	C	D	E	F
A	AE000511		0.08	0.10	0.09	0.08	0.10
B	NC_012973	0.08		0.10	0.09	0.07	0.10
C	CP002076	0.10	0.10		0.12	0.09	0.05
D	CP002332	0.09	0.09	0.12		0.10	0.12
E	NC_017372	0.08	0.07	0.09	0.10		0.10
F	CP002982	0.10	0.10	0.05	0.12	0.10	

Sequences

Distance

Tree

Above the matrix is a row containing the following information:

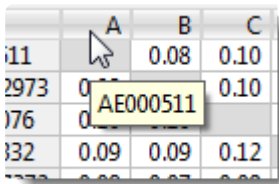
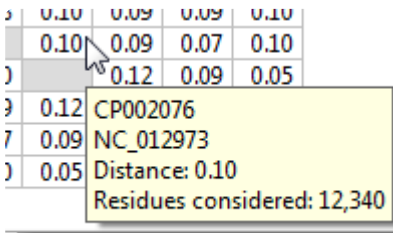
- **Block** – (multi-block [Mauve](#) alignments only) The color of the alignment block currently selected in the Overview.
- **Metric** – The **Metric** (calculation method) specified in the [Distance section](#) of the [Style panel](#).
- **Sequences** – The number of sequences in the alignment or 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 gap. This field shows both the minimum and maximum number of residues considered, if they differ.

Here are some tips for understanding the information shown in the matrix:

- 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.

- The upper and lower triangles of the matrix are reflections of one other and contain the same information.
- 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.
- 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.

The following table describes ways to control the appearance of the Distance view:

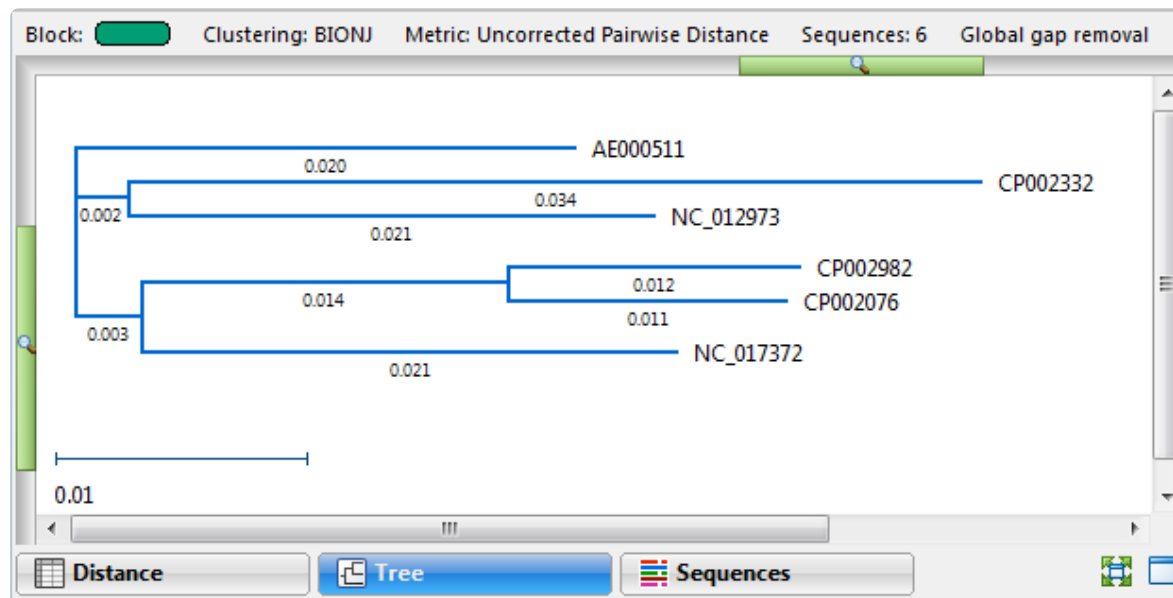
Task	How to...
To change settings related to the Distance view	Use the Distance section of the Style panel .
To display a tool tip with the name of the sequence corresponding to a cell	<p>Hover over any row/column header or any cell on the diagonal of the matrix.</p>  <p>The screenshot shows a portion of a distance matrix. The columns are labeled A, B, and C. The rows are labeled with sequence identifiers: 11, 2973, 176, and 132. A mouse cursor is hovering over the cell for sequence 11 under column A, which displays a tooltip with the sequence name 'AE000511'. The diagonal cells (where sequence names meet) are highlighted in yellow.</p>
To display a tool tip with the two sequence names and the distance between them	Hover over a distance cell in the table that is not on the diagonal. For projects using Pairwise gap removal (see Distance section), the tool tip also contains the number of residues considered for that pair.
To select all cells in the table	Choose Distance > Table > Select All .
To use editing commands with the Distance table	Choose Distance > Table > Copy (Select All, Rename, Rename with Fields) .
To freeze scrolling	<p>Choose Distance > Table > Freeze Scrolling Region. In a large table, this command allows you to compare rows near the top with those further down.</p>  <p>The screenshot shows a distance matrix with a mouse cursor hovering over a cell. A tooltip is displayed, showing the sequence names 'CP002076' and 'NC_012973', the distance 'Distance: 0.10', and the number of residues considered 'Residues considered: 12,340'.</p>

Tree view

The Tree View shows evolutionary relationships estimated from the multiple sequence alignment, and is calculated using the currently selected [Distance metric](#). The tree is built using the BioNJ algorithm ([Gascuel, 1997](#)), a variant of the Neighbor-Joining algorithm ([Saito and Nei, 1987](#)) used in MegAlign. The following video is a brief overview of the Tree view:

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.

The Tree 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 Tree tab (the dark blue part of the image, below).



Branch lengths are displayed, by default, and a distance scale appears at the bottom left of the image. Phylogenetic trees are considered to be reflections of the evolutionary relationships between the aligned sequences. Due to the nature of the tree-building algorithm, note that the sum of branch lengths between any pair of sequences is not expected to equal the pair-wise distance reported in the Distance view. This is because the computed branch lengths are conditioned by the relationships to other sequences represented on the tree.

Note: Due to the difference in clustering algorithms, trees built in MegAlign 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.

Above the tree is a row containing the following information:

- **Block** – (multi-block Mauve alignments only) The color of the alignment block currently selected in the [Overview](#).
- **Clustering** – The clustering method used is BioNJ ([Gascuel, 1997](#)).
- **Metric** – The [distance metric](#) (calculation method) specified in the [Distance section](#) of the [Style panel](#).
- **Sequences** – Number of sequences in the alignment or in the active block.
- **Global/Pairwise gap removal** – The gap treatment specified in the 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.

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.

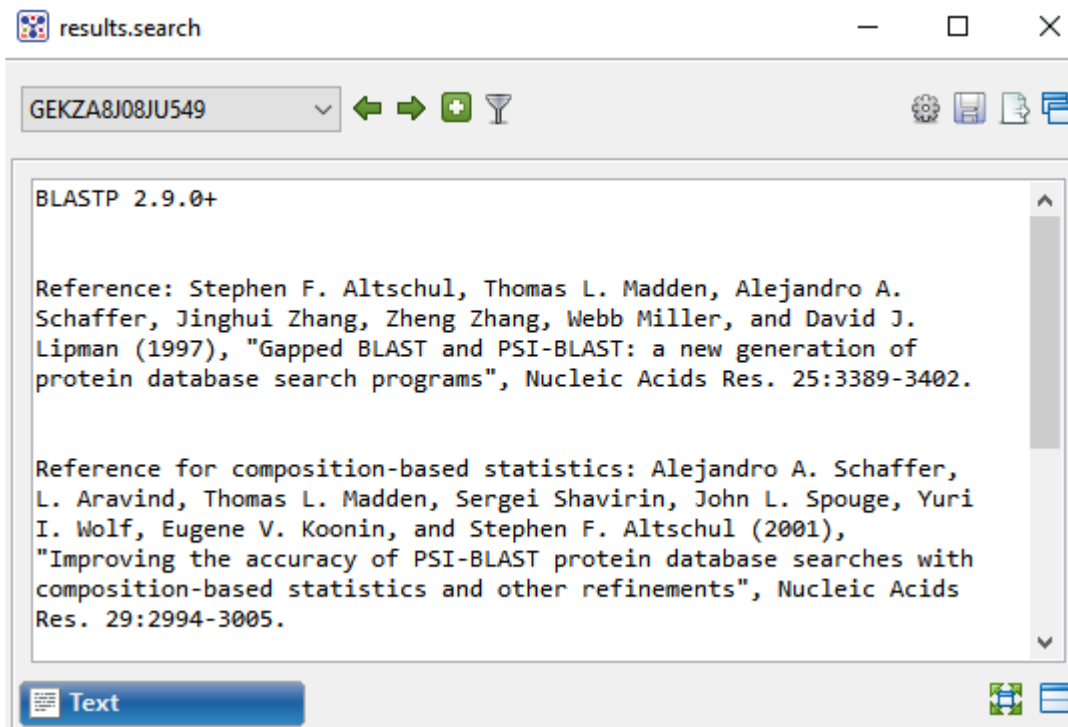
The following table describes ways to control the appearance of the Tree view:

Task	How to...
To change settings related to the Tree view	Use the Tree section of the Style panel .
To redraw the tree according to the default settings	Use Tree > Reset Phylogenetic Tree View or Ctrl/Cmd+Shift+R .

Text view

After the [Jobs panel](#) shows the **Status** of a [BLAST search](#) or [Entrez search](#) is **Done**, pressing the “Done” hyperlink opens the search results simultaneously in the Text, [Table](#), and [Pairwise](#) views. All three views are contained in a large pop up window that is separate from the main window.

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.

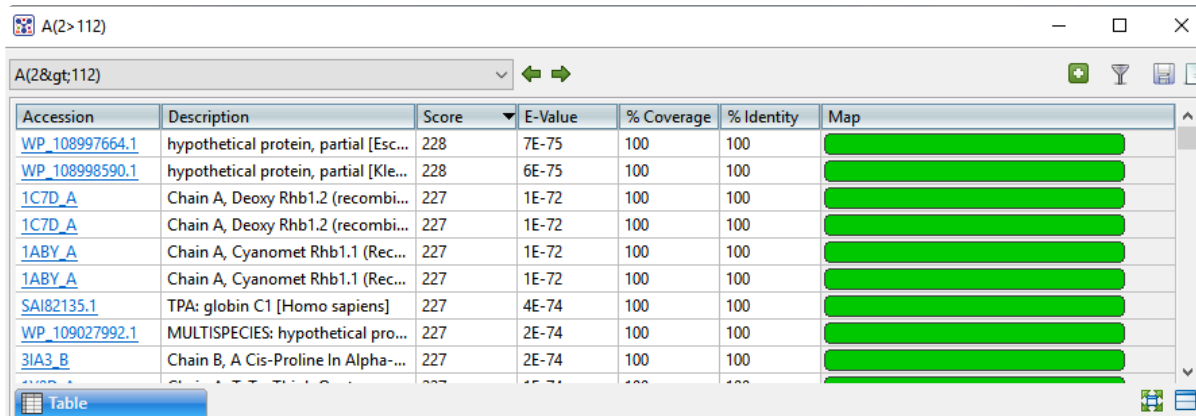
Tasks that pertain to the Text view:

Task	How to...
To make selections in the view	See Make a Selection .
To copy all text	Right-click anywhere in the view and choose Copy .
To select all text	Right-click anywhere in the view and choose Select All .

Table view

After the [Jobs panel](#) shows the **Status** of a [BLAST search](#) or [Entrez search](#) is **Done**, pressing the “Done” hyperlink opens the search results simultaneously in the [Text](#), Table, and [Pairwise](#) views. All three views are contained in a large pop up window that is separate from the main window.

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

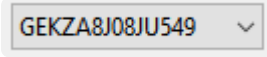










Accession	Description	Score	E-Value	% Coverage	% Identity	Map
WP_108997664.1	hypothetical protein, partial [Esc...	228	7E-75	100	100	
WP_108998590.1	hypothetical protein, partial [Kle...	228	6E-75	100	100	
1C7D_A	Chain A, Deoxy Rhb1.2 (recombi...	227	1E-72	100	100	
1C7D_A	Chain A, Deoxy Rhb1.2 (recombi...	227	1E-72	100	100	
1ABY_A	Chain A, Cyanomet Rhb1.1 (Rec...	227	1E-72	100	100	
1ABY_A	Chain A, Cyanomet Rhb1.1 (Rec...	227	1E-72	100	100	
SAI82135.1	TPA: globin C1 [Homo sapiens]	227	4E-74	100	100	
WP_109027992.1	MULTISPECIES: hypothetical pro...	227	2E-74	100	100	
3IA3_B	Chain B, A Cis-Proline In Alpha...	227	2E-74	100	100	

Columns in the view's table are described below:












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.

Tasks that pertain to the Table view:



Task	How To...
To launch the NCBI page for matching database entry	Click the hyperlink in the table's Accession column.
To copy all information from a search result row	Right-click on the row and choose Copy .
To 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.
To display a different search job	Select it from the drop-down menu in the upper left of the window ().
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.
To display search results for previous query (in a single job with multiple queries)	Use the Previous query tool ().
To display search results for the next query (in a single job with multiple queries)	Use the Next query tool ().
To filter results from a sequence search	Use the Filter results tool (). Filtering for this situation is described here .
To filter results from a text search	Use the Filter results tool (). Filtering for this situation is described here .
To add sequences or matches	Use the Add sequences or subsequences tool (). Sequences are added back into the host application, not the job's document.
To launch the Search Online dialog	Use the Change search strategy tool (). See Search for Sequences Online for details.
To save the search job to disk	Use the Save as tool ().
To export data from the view	Use the Export data tool (), described in the topic Export data from the Text, Table or Pairwise views .

Change the appearance of a view

The following table shows methods for changing the appearance of one or more [views](#):

Task	How to...
To show a view	Use View > Show Document View > (View Name) .
To split the project into two sections that are arranged as a “vertical stack”	Use Sequences > Split, Overview > Split , or click on the Split tool () under the view.
To minimize all but the active view	Use Sequences > Remove Split, Overview > Remove Split , or click on the Remove Split tool () under the view.
To show the active view in full-screen mode	Use View > Full Screen or press Ctrl/Cmd+Shift+F .
To exit full-screen mode and return to normal mode	Press the Esc key or Ctrl/Cmd+Shift+F , or click on the Exit Full Screen tool () at the bottom of the full-screen view.
To detach the view	Use View > Detach Document View > (View Name) or use the context (right-click) option of the same name. Detaching a view turns it into a free-floating window.
To reattach all detached views to their original locations	Use View > Reattach All Views .
To show the active view in full-screen mode	Use the View > Full Screen command or the the Full screen tool () at the bottom right corner of the view.
To exit full screen mode	Press the Esc key or use the Exit full screen tool () at the bottom of the full-screen view.
To configure the number and location of visible views.	<p>Use the View > Change layout command or the Change layout tool (image varies) in the bottom right corner of the view. The following options are available:</p> <ul style="list-style-type: none"> • One – ( One) – To show only the current view. • Two Horizontal ( Two Horizontal) – To show the current view on top and one additional view below it. • Two Vertical ( Two Vertical) – To show the current view on the left, and one additional view to the right. • Three Horizontal ( Three Horizontal) – To show the current view on top and two additional views below it. • Three – Left ( Three - Left) – To show the current view on the left and two additional views stacked one on top of the other on the right. • Three – Right ( Three - Right) – To show the current view

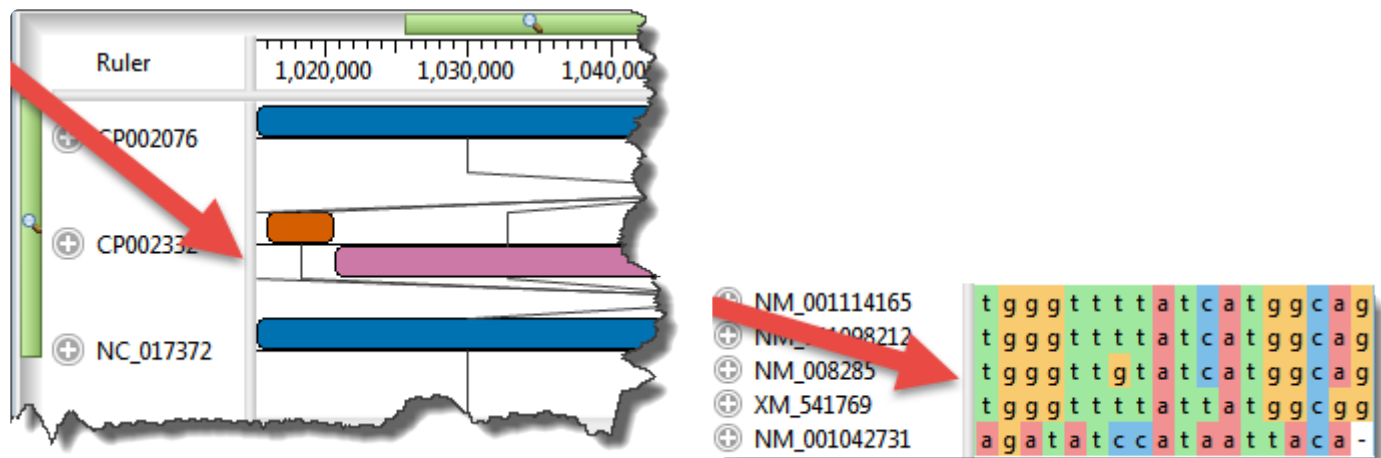
on the right and two additional views stacked one on top of the other on the left.

- **Three – Top** ( **Three - Top**) – To show the current view on the top and two additional views side-by-side on the bottom.
- **Three – Bottom** ( **Three - Bottom**) – To show the current view on the bottom and two additional views side-by-side on the top.

Split and resize panes in a view

To change the relative sizes of panes:


To change the sizes of panes (e.g., the sequence names and the alignment blocks areas in the [Overview](#)), drag the light gray divider bar between them to the left or right. The images below show examples in the [Overview](#) (left) and [Sequences view](#) (right).

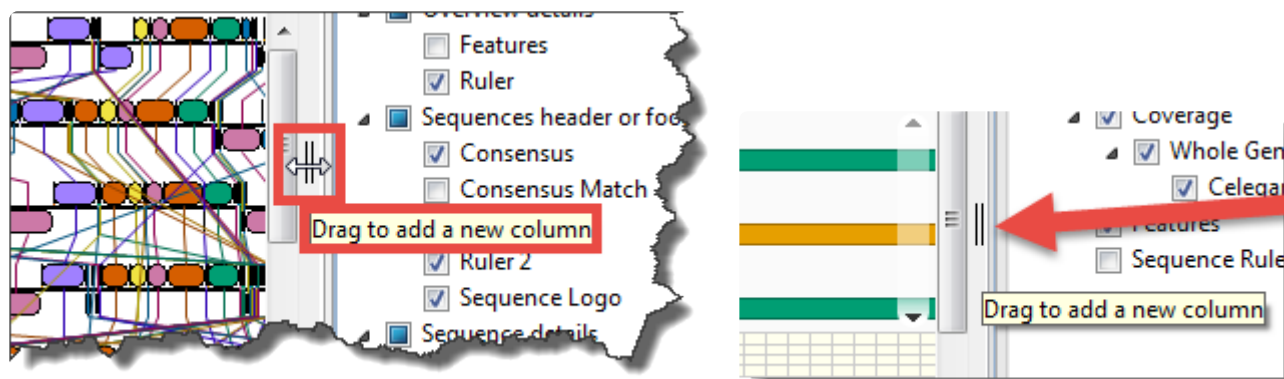


To split the sequence portion of the view into multiple panes:

To view different sections of the sequences simultaneously, you can split the view into any number of panes.


- To split the Overview in half, use **Overview > Split View**. To split the Sequences view in half, use **Sequences > Split View**.

- To split the view manually, drag the view-splitter tool (), found at the right edge of the view, in a leftward direction. The images below show examples in the [Overview](#) (left) and [Sequences view](#) (right).



When the Sequences view is split, the Overview provides a separate “viewport indicator” for each section. Splits are ignored when printing or exporting an image, although the layout is preserved.

To split the sequence portion of the Sequences view horizontally:

To view different sections of the sequence simultaneously, you can split the view into horizontal panes. From the [Style panel](#), use the [Sequences section Layout](#) drop-down menu to select **Automatic wrapping**. The Sequences view will immediately wrap horizontally. You can then add one or more splits by dragging the split tool at the bottom of the view () in an upward direction.

If you were using a **Linear** layout and had already split the sequence portion of the view vertically, then choosing **Automatic wrapping** will change vertical splits to horizontal ones. Also, the scroll bars will move from the bottom of the view to the right.

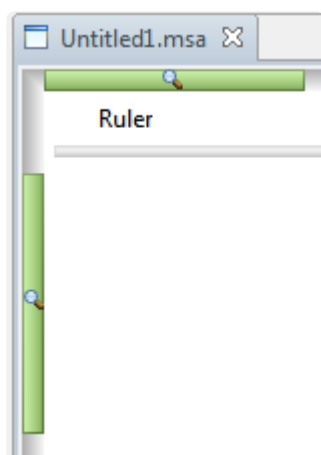
When the Sequences view is split, the Overview provides a separate “viewport indicator” for each section. Splits are ignored when printing or exporting an image, although the layout is preserved.

To remove vertical or horizontal splits:

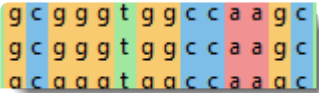




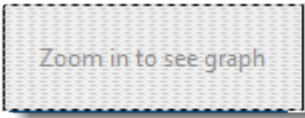


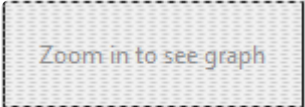
To remove an existing split, drag the divider as far as you can to either side (vertical splits) or downward (horizontal splits) and drop it there. Or, to automatically remove the right-most split, use **Overview > Remove Split**. To split the Sequences view in half, use **Sequences > Remove Split**.

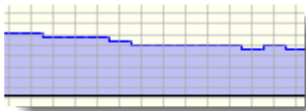

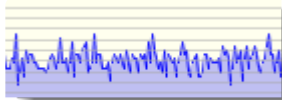
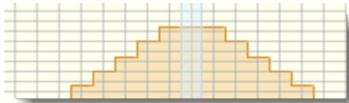
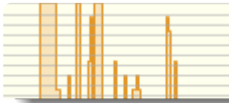
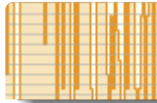
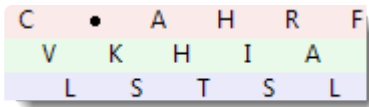


Zoom in or out on a view

Three of MegAlign Pro's views ([Sequences](#), [Tree](#) and [Overview](#)) all 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.



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.	 Some tracks cannot be displayed at low zoom levels, and prompt you to zoom in.
Consensus Match			 Zoom in to see graph

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

Panels


Settings panels allow you to optimize the look, contents and layout of the MegAlign Pro view. To learn what can be done within each panel and their “areas” (expandable sections within each panel), click the links below.

- [Style](#)
- [Tracks](#)
- [Features](#)
- [Places](#)
- [Details](#)
- [Console](#)

Style panel

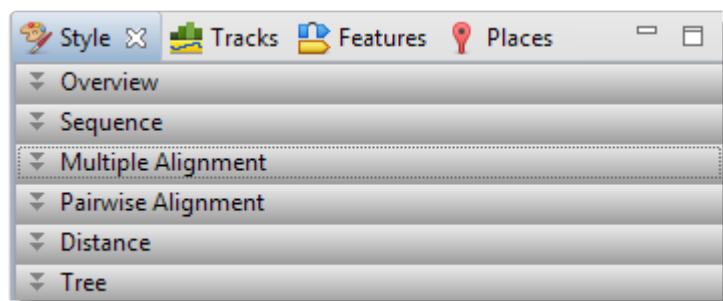
The Style panel has expandable sections with options affecting the [Overview](#), [Sequences](#), [Distance](#) and [Tree](#) views, as well as [multiple](#) and [pairwise](#) alignments.

To open the Style panel:

Click on the **Style** tab (). 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**.

To change settings in the Style panel:



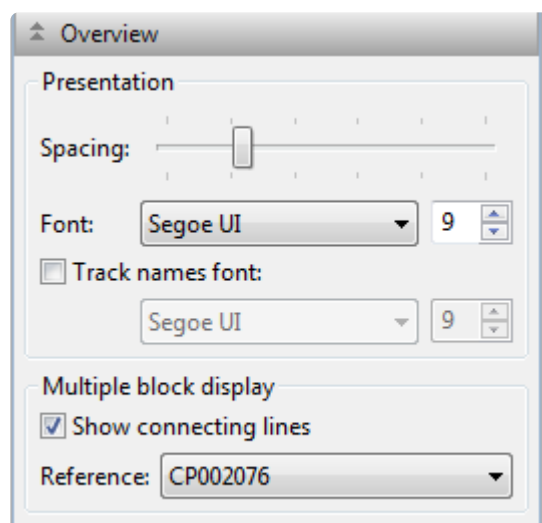
Click the links below to learn about each of the sections in the Style panel:



- [Overview section](#)
- [Sequence section](#)
- [Multiple Alignment section](#)
- [Pairwise Alignment section](#)
- [Distance section](#)

- [Tree section](#)

Overview section

To access options affecting the [Overview](#) display, click on the [Style panel](#) expand bar entitled **Overview**, or choose **View > Style > Overview**.

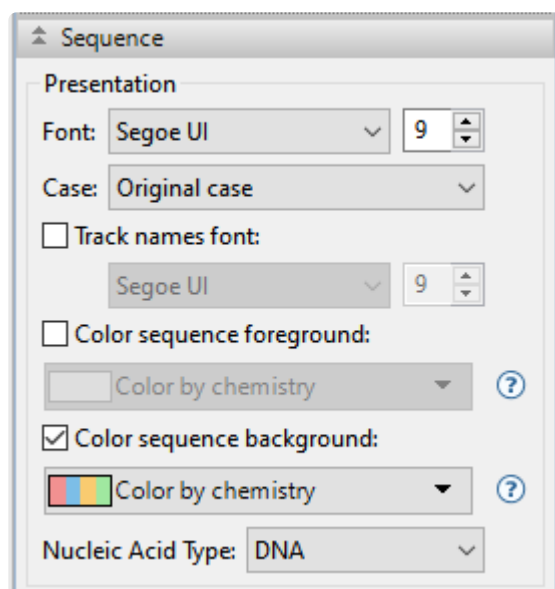




Task	How to...
To increase/decrease the amount of vertical space between sequence blocks	Use the Spacing slider. This command does not affect the sizes of the alignment blocks, but only the amount of white space between them. To make the alignment blocks increase/decrease in size, instead use the Overview's zoom in/out sliders.
To 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.
To 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 track labels.
To toggle between displaying or hiding 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.

To specify the reference sequence	To use one of the aligned sequences as the reference, make a selection from the Reference 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 selected reference will be placed on the forward strand. Blocks for non-reference sequences can appear on either strand.
-----------------------------------	---

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...
To change font options for the ruler numbers in the Sequences 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 select font options for other items in the Sequences 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 change the sequence color scheme in the Sequences view	To change the letter color in the Sequences view, check the Color foreground box. To change the background color (boxes around each letter), check the Color background box. Then make selections from the corresponding drop-down menus. Color schemes are described in Available color schemes . As alternatives to checking/unchecking these boxes, check or uncheck the menu items Sequences > Color Sequence Foreground (Background) .

Available color schemes

The [Sequence section](#) allows you to select color schemes for sequences in the [Sequences view](#). Each available color scheme is described below:

Color by Chemistry - This is the default color scheme used by multiple DNASTAR applications. Amino acids are colored according to their side chain chemistry:

Nucleotide
A
C
G
T/U

Side Chain Chemistry - Amino Acids
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)

Shapely - This color scheme matches the RasMol amino acid and RasMol nucleotide [color schemes](#), which are, in turn, based on Robert Fletterick's "Shapely models."

Nucleotide
A
C
G
T
U

Amino Acids
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)

Taylor - This color scheme is taken from [Taylor](#) and is also used in JalView ([Waterhouse et al., 2009](#)).

Nucleotides
A
C
G
T/U

Amino Acids
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)

Zappo - This scheme colors residues according to their physico-chemical properties, and is also used in [JalView](#) ([Waterhouse et al., 2009](#)):

Nucleotide
A
C
G
T
U

Physico-Chemical Property - Amino Acids
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)

Solid – These scheme colors residues with the color shown. After selecting **Solid**, you can specify the color by clicking on the colored box to the right of the menu.

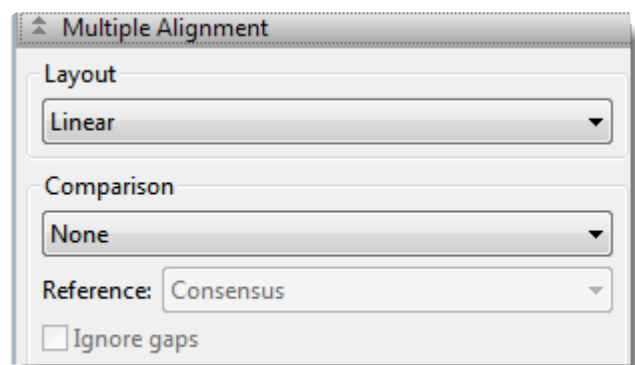


* **Note:** In Mauve alignments, “dimmed” (dull) colored residues represent flanking sequences immediately adjacent to the alignment block currently selected in the Overview.

⊕ AE000511	t g g c c a c t c	a a a a a g c c g t t t t a g
⊕ NC_012973	t a g c c a c t c	a a a a a g c c g t t t t a g
⊕ CP002076	g c c a c t c a a	- - a a a g c c g t t t t a g
⊕ CP002332	t a g c c a c t c	a a a a a g c c g t t t t a g
⊕ NC_017372	g c c a c t c a a	- - a a a g c c g t t t c a g

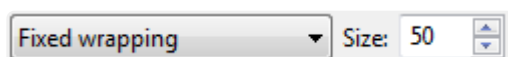
Multiple Alignment section

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**.



- Use the **Layout** drop-down menu to specify the wrapping style for rows in the Sequences view. 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.

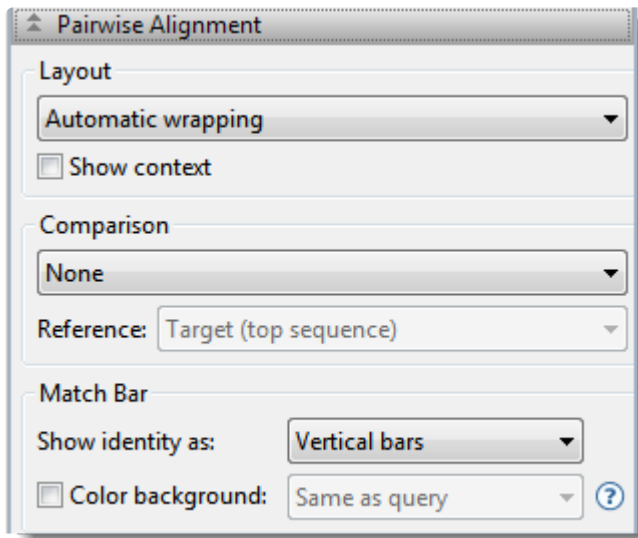
- Use the **Comparison** drop-down menu to specify how to color the sequences and the consensus to facilitate visual comparisons. Sequence comparisons can only be performed with aligned sequences. Choose between:
 - **None** – All residues appear in the color schemes chosen in **Color foreground** and **Color background** (see row above).
 - **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.

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**.



- Use the **Layout** drop-down menu to specify the wrapping style for rows in the Sequences view. 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.

- A local alignment reports the highest scoring contiguous 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 sequences within a larger sequence. Flanking segments of sequences that are not within the aligned segment can be visualized by checking the **Show context** box. The flanking segments appear in a dimmer color scheme than the alignment.

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** -----

- Use the **Comparison** drop-down menu to specify how to color the sequences and the consensus to facilitate visual comparisons. Sequence comparisons can only be performed with aligned sequences. Choose between:
 - **None** – All residues appear in the color schemes chosen in **Color foreground** and **Color background** (see row above).
 - **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 one of the two comparison options described in this bullet point.
- The Match bar section affects display of the optional [Match bar track](#). This track is used to show similarities and differences between the query and target sequences:
 - Use the **Show identity as** drop-down menu to choose how to show the connection between the query and the target. 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	(0,1)
Period	Neutral mismatch	0
Space	Unfavorable mismatch	< 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 [Match Bar track options](#) section. Choices made in either location are instantly reflected in the other location.

Distance section

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

Task	How to...
To change the metric used to calculated distance	Use the Metric drop-down menu. For detailed information, see Distance metric options .
To specify when gaps should be ignored	Use the Gap Treatment drop-down menu. For detailed information, see Gap treatment options .
To change 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.
To change 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 change 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.

Distance metric options

The metric used to calculate distances (which affect the [Distance](#) and [Tree](#) views) is specified in the [Distance section](#) of the [Style panel](#). To open the section, click on the expand bar entitled **Distance**, or choose **View > Style > Distance**.

To change the distance metric, use the *Metric* drop-down menu. Note that metric selection is disabled for projects without sequences.

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: $\%ID = 100 * (1 - \text{distance})$.
- **Kimura** – (protein sequences only) – The [Kimura](#) model (1983), which should not be used for very divergent sequences, uses the formula: $D = -\ln(1 - p - 0.2 p^2)$, 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.

Gap treatment options

Gap treatment options affect the calculation of distances, which, in turn, affect the [Distance](#) and [Tree](#) views. One of two options may be specified in the [Distance section](#) of the [Style panel](#). To open the section, click on the expand bar entitled **Distance**, or choose **View > Style > Distance**.

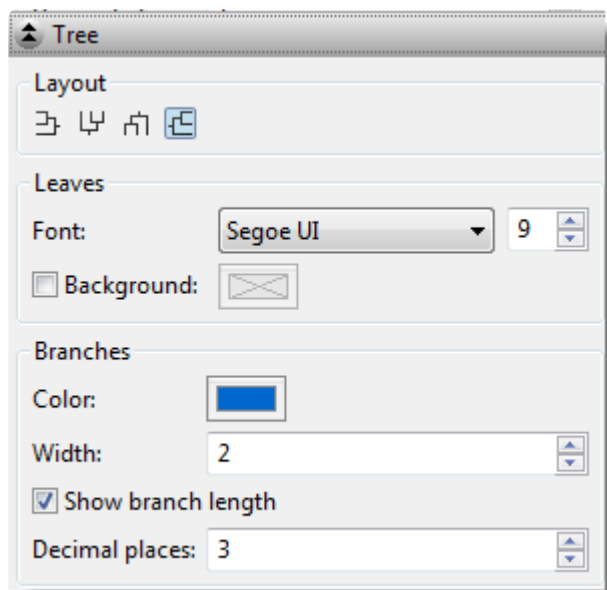
Use the **Gap Treatment** drop-down menu to choose between these options:



- **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.


The banners at the tops of the Distance and Tree views will show a range in the field **Residues considered** if different numbers of residues are considered for different pairs of sequences. This happens frequently when **Pairwise gap removal** is checked. However, if the minimum and maximum values differ greatly, we recommend selecting **Global gap removal** instead.

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...
To change the layout of the Tree view	Click on any of the four orientations to show the phylogenetic tree with that layout (left, up, down, right).
To change the font style used in the “leaves” of the Tree 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 a row of dashes.
To change the font size used in the “leaves” of the Tree view	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 change the background color in the Tree view	Check the box to the left of Background , and then click the box to the right to open the color picker.
To change the branch color in the Tree view	Click the Color box in the Branches section to open the color picker.
To change the branch width in the Tree view	Type a number to the right of Width or use the corresponding up and down arrows () to select a font size. Allowed values = 1 to 20.
To toggle between showing/hiding the	Check or uncheck the box to the left of Show branch length .

branch lengths	
To change the number of decimal places for the branch length	Type a number to the right of Decimal places , or use the corresponding up and down arrows () to specify the number of decimal places for the branch length. Allowed values are 1-7. This field is only enabled when Show branch length is checked.

Tracks panel

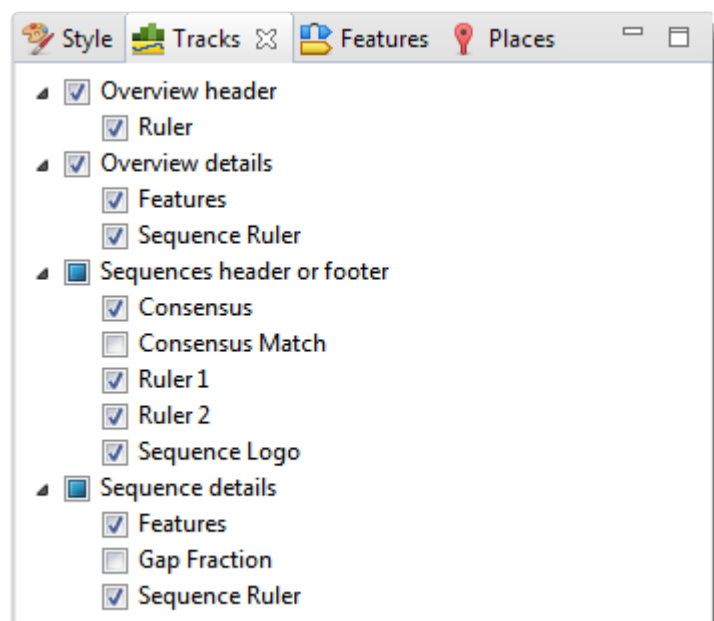
The Tracks panel lets you manage the tracks displayed in the [Overview](#) and the [Sequences view](#).

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:




The uppermost part of the Tracks panel lets you specify which tracks are visible, and configure their options and appearance. Depending whether the project contains DNA/RNA or protein sequences, different subsets of track types will be listed. For instance, in the absence of an alignment, **Consensus**, **Consensus Match**, **Sequence Logo** and **Gap Fraction** are hidden from the list. For those tracks that are available, some will be displayed and other will be hidden, by default. Put check marks next to available tracks that you wish to display, and remove check marks next to 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:

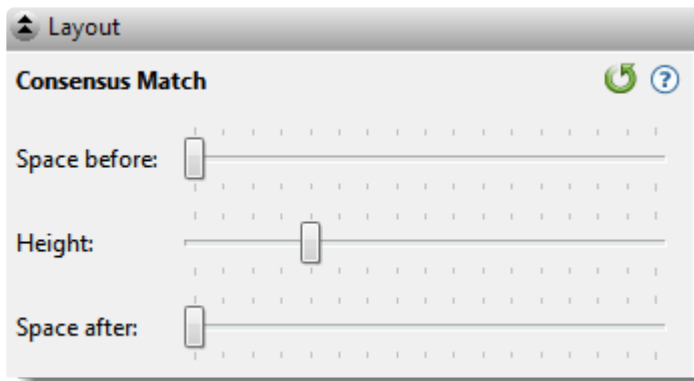
- [Layout](#)
- [Options](#)


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

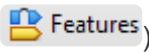
To access settings for the track type currently selected in the [Tracks panel](#), click on the Tracks panel expand bar entitled “Options,” or choose **View > Tracks > Options**.

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

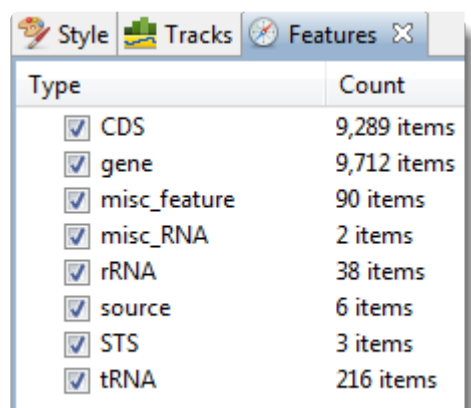
The Features panel lets you specify which feature types should be visible in the [Overview](#) and the [Sequences view](#) once a sequence has been expanded.

To open the Features panel:

Click on the **Features** tab (). 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**.

To change settings in the Features panel:



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

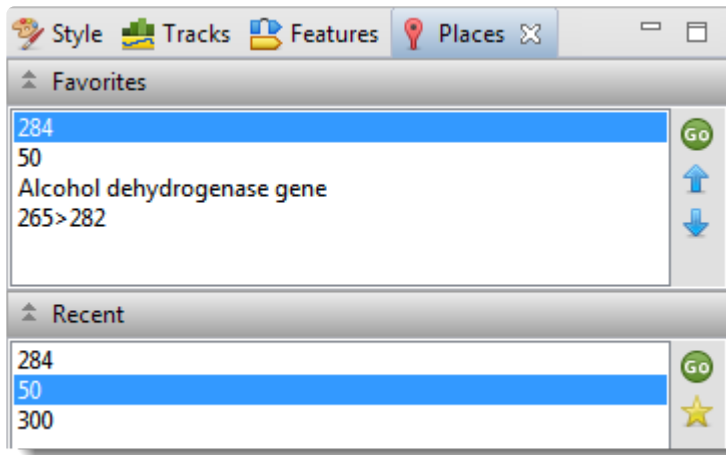
The Places panel lets you quickly navigate to a sequence location or range that you have visited before. The following video shows to bookmark favorite locations using the Places panel:

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:

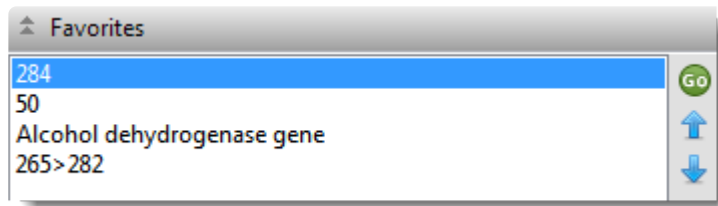
- [Recent section](#)
- [Favorites section](#)

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 choosing **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

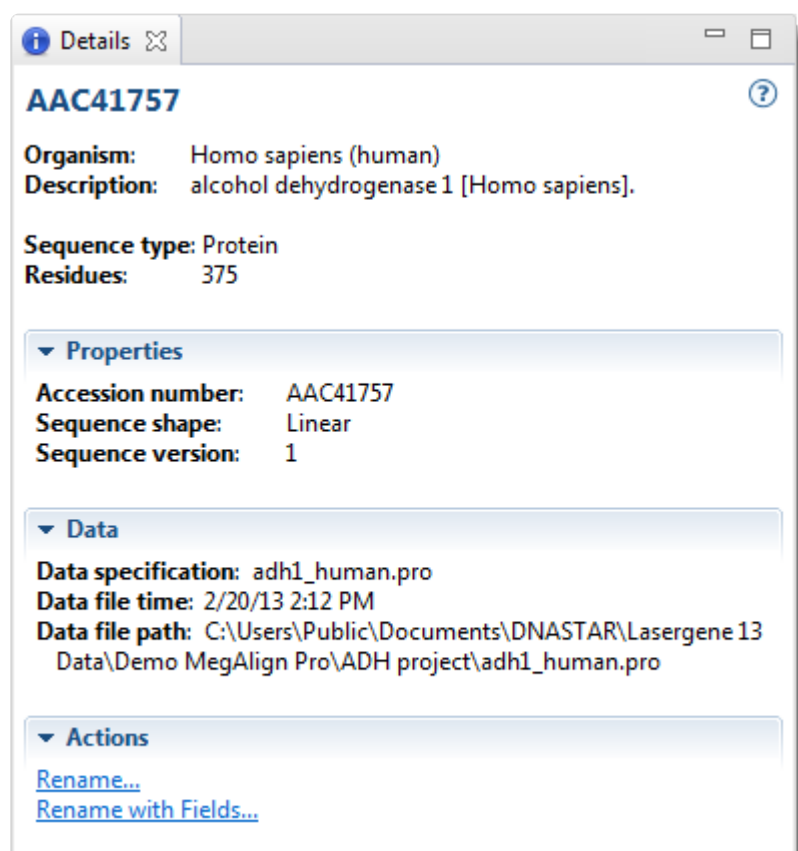
The Details panel shows details of the application, selection or project.

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.

<p>Heading</p> <p>Getting Started in MegAlign Pro</p> <p>Create a new alignment project Open a file Import an alignment</p> <p>Add sequences Add sequences end-to-end</p> <p>Watch a training video</p>	<p>One or more links are available. The first three options are available for blank documents, but unavailable when no document is open:</p> <ul style="list-style-type: none"> • Create or open a project • Import an alignment • Add sequences to the project (or add them end-to-end). • Watch a training video
<p>Recent Documents</p> <p>▼ Recent Documents</p> <p>RP1 project.msa ADH Project.msa Drosophila project.msa ADH project.msa</p>	<p>Open a recent document.</p>

When a non-empty document is open

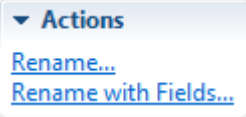
If a document is open and contains sequences (whether they have been aligned or not) and nothing is selected, the [Details panel](#) contains the following information:

<p>Heading</p> <p>ADH Project.msa</p> <p>Sequence type: Protein</p>	<ul style="list-style-type: none"> • The project filename. • Sequence type – DNA, RNA, or Protein.
<p>Alignment</p> <p>▼ Alignment</p> <p>Alignment: Clustal Omega Sequences: 5</p> <p>Sequence type: Protein Min. sequence length: 348 Max. sequence length: 377</p>	<p>This section is only displayed for, and only pertains to, sequences in the project that have already been aligned.</p> <ul style="list-style-type: none"> • 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.
<p>Unaligned Sequences</p> <p>▼ Unaligned Sequences</p> <p>Sequences: 5 Sequence type: Protein Min. sequence length: 348 Max. sequence length: 377</p>	<p>This section is only displayed for, and only pertains to, sequences in the project that are not included in the alignment.</p> <ul style="list-style-type: none"> • Sequences – Number of unaligned sequences. • Sequence type – DNA, RNA, or Protein. • Min. and Max. sequence lengths – Lengths of the shortest and longest unaligned sequences.
<p>Actions</p> <p>▼ Actions</p> <p>Add sequences Add sequences end-to-end Align with options</p>	<ul style="list-style-type: none"> • 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:

<p>Heading</p> <p>AAC41757</p> <p>Organism: Homo sapiens (human) Description: alcohol dehydrogenase 1 [Homo sapiens].</p> <p>Sequence type: Protein Residues: 375</p>	<p>The default name, if present, appears in blue above 'Organism.'</p> <ul style="list-style-type: none"> • 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.
<p>Properties</p> <p>▼ Properties</p> <p>Accession number: AAC41757 Sequence shape: Linear Sequence version: 1</p>	<ul style="list-style-type: none"> • 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).
<p>Data</p> <p>▼ Data</p> <p>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</p>	<ul style="list-style-type: none"> • 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.
<p>Actions</p>	<ul style="list-style-type: none"> • Rename – Shortcut to Edit > Rename (See

 <p>▼ Actions Rename... Rename with Fields...</p>	<p>Rename sequences manually).</p> <ul style="list-style-type: none">• Rename with Fields – Shortcut to Edit > Rename with Fields (See Rename sequences automatically using specified data fields).
--	--

When a portion of a 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:

<p>Heading</p> <div data-bbox="126 646 516 800"> <p>Selected Region</p> <p>Sequence: XM_541769</p> <p>Alignment selection: (1,538>1,555)</p> </div>	<ul style="list-style-type: none"> • 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. <p>When the sequence is not part of an alignment, (i.e., is listed with 'unaligned sequences'), then then Alignment selection is replaced with Sequence selection. In this case, the coordinates are the ungapped sequence coordinates rather than the alignment coordinates.</p>
<p>Actions</p> <div data-bbox="126 1066 342 1192"> <p>▼ Actions</p> <p>Copy</p> <p>Copy as FASTA...</p> </div>	<ul style="list-style-type: none"> • 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:

<p>Heading</p> <div data-bbox="126 468 477 630"> <p>2 sequences</p> <p>Sequence type: Protein</p> <p>Min. sequence length: 348</p> <p>Max. sequence length: 375</p> </div>	<ul style="list-style-type: none"> • 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.
<p>Actions</p> <div data-bbox="126 756 375 840"> <p>▼ Actions</p> <p>Rename with Fields...</p> </div>	<ul style="list-style-type: none"> • 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:

<p>Heading</p> <p>Selected Region</p> <p>Sequences: 4</p> <p>Alignment selection: 2,148>2,173</p>	<ul style="list-style-type: none"> • Sequences – The number of sequences in the alignment. • Alignment selection – The inclusive coordinates of the beginning and end of the selection.
<p>Actions</p> <p>▼ Actions</p> <p>Copy</p> <p>Copy as FASTA...</p>	<ul style="list-style-type: none"> • 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.

<p>Heading</p> <p>Features</p> <p>Summary: The Features track displays the pre-alignment coordinates of each sequence.</p>	<ul style="list-style-type: none"> • Type of track (in blue). • Summary – Basic information about what is displayed in the selected track.
<p>Actions</p> <p>▼ Actions</p> <p>Rename...</p> <p>Rename with Fields...</p> <p>Show track layout</p> <p>Show track options</p>	<p>Only a subset of links may be available, depending on the selected track.</p> <ul style="list-style-type: none"> • 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 [Overview](#) or [Sequences view](#), the [Details panel](#) contains the following information.

<div>Heading</div> <div><div>2 tracks</div><div>Summary: Multiple tracks are selected. They can be moved or hidden together.</div></div>	<ul style="list-style-type: none">• Number of tracks currently selected (in blue).• Summary – Description of options available for the multiple track selection.
<div>Actions</div> <div><div>Actions</div><div>Show tracks panel</div></div>	<ul style="list-style-type: none">• Show tracks panel – Shortcut to View > Tracks > Tracks.

For additional information about using tracks, see the topics [Tracks panel](#) and [Tracks](#).

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:

<p>Heading</p> <div data-bbox="126 514 506 709"> <p>alcohol_DH_class_I_II_IV</p> <p>Type: Region Left: 3 Right: 375 Length: 373</p> </div>	<ul style="list-style-type: none"> • 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.
<p>Location</p> <div data-bbox="126 884 279 968"> <p>▼ Location</p> <p>3>375</p> </div>	<p>Range of the feature in nucleotides or residues.</p>
<p>Properties</p> <div data-bbox="126 1062 453 1268"> <p>▼ Properties</p> <p>db_xref: taxon:10090 organism: Mus musculus chromosome: 3 map: 3 strain: C57BL/6</p> </div>	<p>Properties shown vary by feature and may include the organism, gene name NCBI feature qualifiers, etc.</p>
<p>Notes</p> <div data-bbox="126 1367 685 1451"> <p>▼ Notes</p> <p>note: class I, II, IV alcohol dehydrogenases; cd08299</p> </div>	<ul style="list-style-type: none"> • note – The NCBI feature qualifier of the same name.
<p>Actions</p> <div data-bbox="126 1644 578 1812"> <p>▼ Actions</p> <p>Select Subsequence Select Subsequences Across Alignment <i>You can also double- or triple-click on this feature.</i></p> </div>	<ul style="list-style-type: none"> • 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


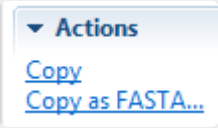
If two or more features have been selected in a track in the [Overview](#) or [Sequences view](#), the [Details panel](#) contains the following information:

<p>Heading</p> <div data-bbox="126 499 548 724"> <p>3 features</p> <p>Type: Region, source, Protein Left: 1 Right: 374 Min. Length: 368 Max. Length: 374</p> </div>	<ul style="list-style-type: none"> • 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.
<p>Properties</p> <div data-bbox="126 886 568 1003"> <p>▼ Properties</p> <p>db_xref: taxon:10090, GeneID:15465, MGI:107619</p> </div>	<p>Properties, if shown in this case, will vary by feature and may include the organism, gene name NCBI feature qualifiers, etc.</p>

For additional information about working with features, see [Features panel](#) and [Features](#).

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:

<p>Heading</p> 	<ul style="list-style-type: none"> • Selected Region – (in blue) Unchangeable default value. • Pairwise alignment selection – The inclusive coordinates of the beginning and end of the selection.
<p>Actions</p> 	<ul style="list-style-type: none"> • Copy – Shortcut to Edit > Copy (See Copy, Paste and Delete). • Copy As FASTA – Shortcut to Edit > Copy As FASTA (See Copy, Paste and Delete).

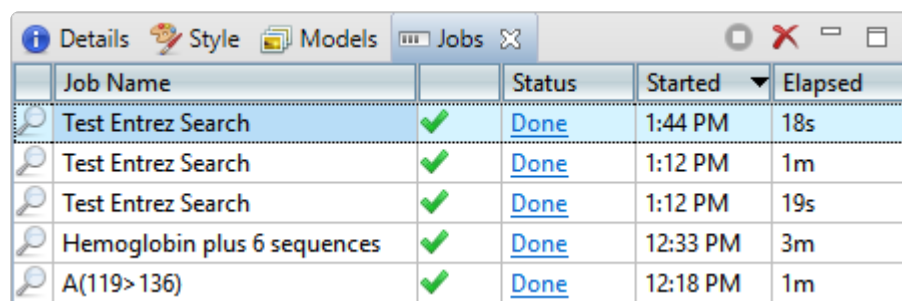
Jobs panel

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 Searches**.





Job Name	Status	Started	Elapsed
Test Entrez Search	Done	1:44 PM	18s
Test Entrez Search	Done	1:12 PM	1m
Test Entrez Search	Done	1:12 PM	19s
Hemoglobin plus 6 sequences	Done	12:33 PM	3m
A(119>136)	Done	12:18 PM	1m
















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	<p>A word and icon indicating the status of the search.</p> <ul style="list-style-type: none"> • 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.

 Details  Style  Models  Jobs 		
Job Name	Status	
 Hotspot for 7TIM (2>248, 2>248)		Done
 GEKZA8J08JS1CF plus 5 sequences		Expiring at 6/6/19 5:33 PM
 Variant for 4HHB (1 position)		Done
 Hemoglobin plus 6 sequences		Expiring at 6/6/19 12:33 PM
 A(119>136)		Expiring at 6/6/19 12:18 PM

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.

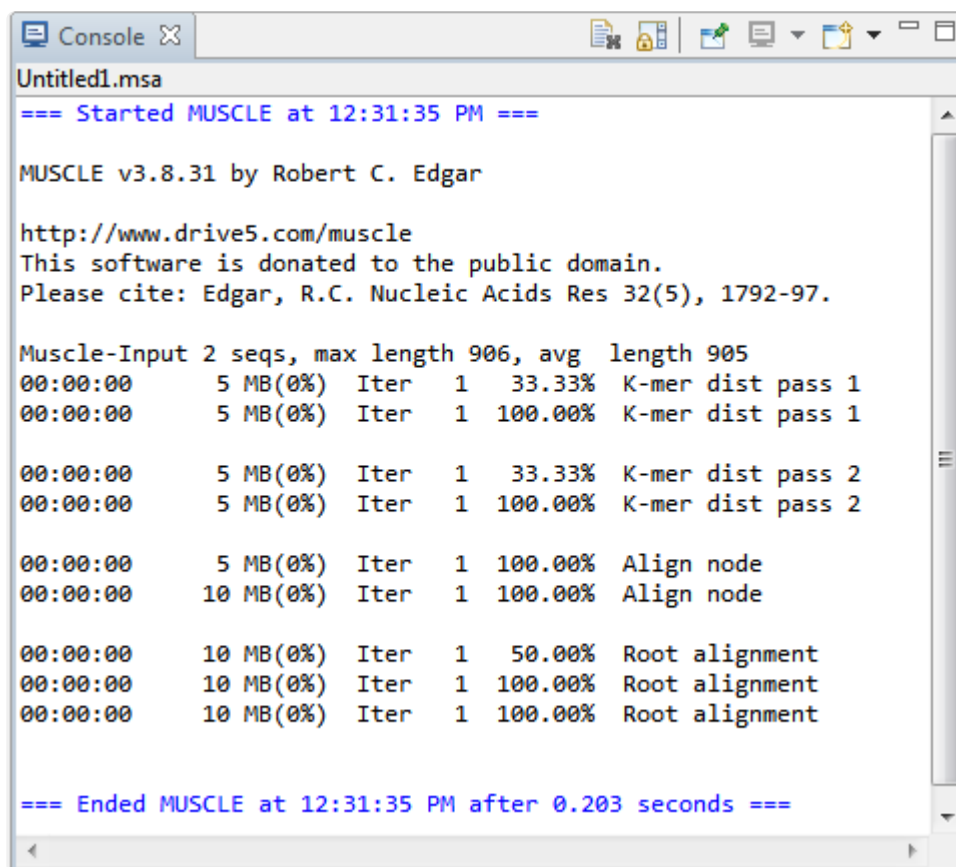
Console panel

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:



```

Console
Untitled1.msa
=== Started MUSCLE at 12:31:35 PM ===

MUSCLE v3.8.31 by Robert C. Edgar

http://www.drive5.com/muscle
This software is donated to the public domain.
Please cite: Edgar, R.C. Nucleic Acids Res 32(5), 1792-97.

Muscle-Input 2 seqs, max length 906, avg length 905
00:00:00    5 MB(0%) Iter  1  33.33% K-mer dist pass 1
00:00:00    5 MB(0%) Iter  1 100.00% K-mer dist pass 1

00:00:00    5 MB(0%) Iter  1  33.33% K-mer dist pass 2
00:00:00    5 MB(0%) Iter  1 100.00% K-mer dist pass 2

00:00:00    5 MB(0%) Iter  1 100.00% Align node
00:00:00   10 MB(0%) Iter  1 100.00% Align node









00:00:00   10 MB(0%) Iter  1  50.00% Root alignment
00:00:00   10 MB(0%) Iter  1 100.00% Root alignment
00:00:00   10 MB(0%) Iter  1 100.00% Root alignment

=== Ended MUSCLE at 12:31:35 PM after 0.203 seconds ===
  
```

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 [Hide, show, move, or resize items in the MegAlign Pro window](#)).

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.

Hide, show, move, or resize items in the MegAlign Pro window

Controls for MegAlign Pro's docking area, panels and expand bars are described in the following table. Not all options in the table are applicable to every item.

Task	"How To"
To show or open a hidden item	Use the associated View menu command, click on a closed expand bar, or click on a plus sign.
To 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.
To increase or decrease the size of an item	Drag the bottom or left sides of the item or use the Maximize icon () or Restore icon (). To view more of an open expand bar, close neighboring bars and/or increase the size of the whole panel.
To move an item	Using the mouse, grab a project, view or panel tab, then drag and drop it in the desired location.
To move a view or panel up/down	Right-click on the view or panel tab and choose Move Up (Down) .
To return to the default MegAlign Pro display	Choose View > Reset Panel Layout .
To return to a view or bring a floating view to the top	Choose View > Show Document View > (View Name) . This command can also be used to regenerate a Pairwise view that was closed.
To scroll vertically	Use the vertical scrollbar or the mouse wheel.
To scroll horizontally	Use the horizontal scrollbar, or hold down the Alt key (Win) or Option key (Mac) and drag the mouse.

Sequences

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

- [Ungapped sequences, including SeqNinja files](#)
- [Gapped sequences](#)
- [Multi-segment sequences](#)
- [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:

- [Manually specify sequence type](#) (DNA, RNA, protein)
- [Reorder sequences](#)
- Rename sequences [manually](#) or [automatically, using specified data fields](#)
- [Remove sequences from a project](#)

Add ungapped sequences to a project

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

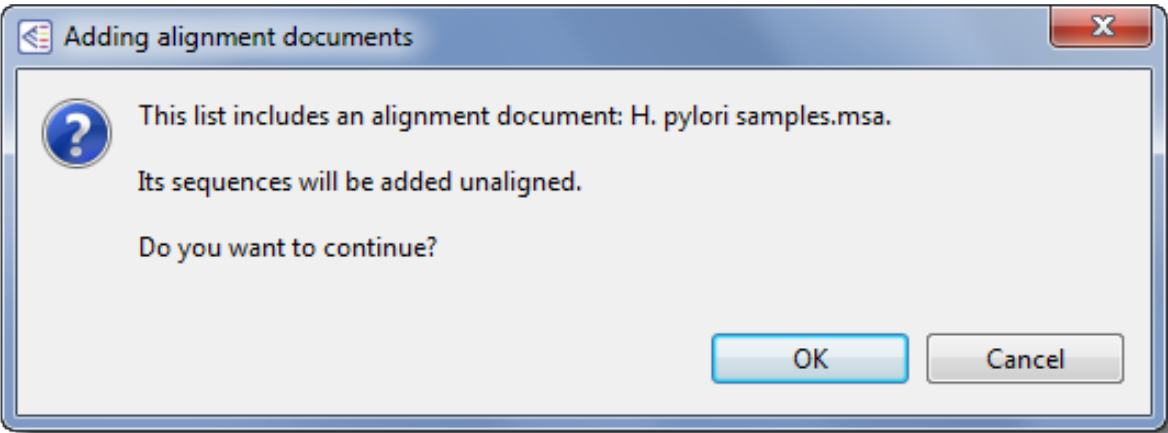
You can add standard DNA, RNA or protein sequences from a wide variety of [file types](#), including SeqNinja .star files. (See [Advantages of using SeqNinja files](#)).

- Choose **File > Add Sequences**, select the **Add Sequences** button () , or press **Ctrl/Cmd+E**. Then choose one or more sequences from the file browser.
- Drag & drop one or more sequence files (but not folders or MegAlign Pro .msa projects) on an open project.

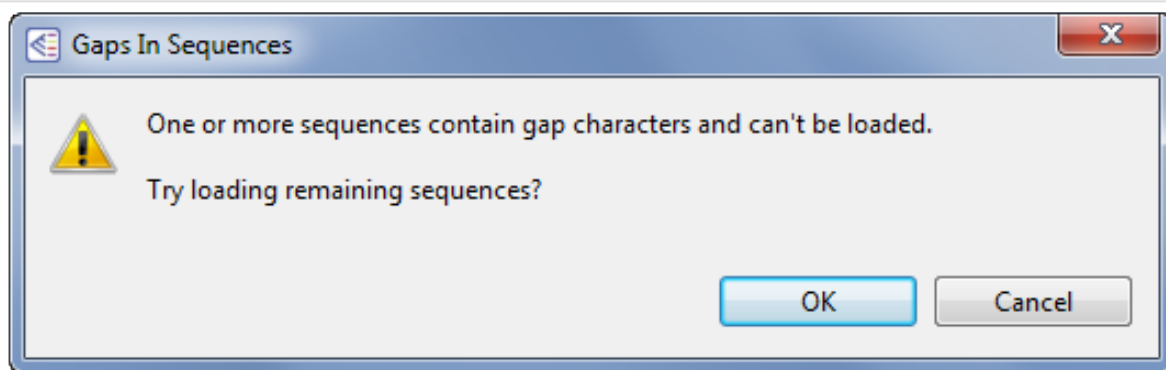
A project can consist of any combination of file types, including single and multi-segment sequences, but all sequences must be either DNA, RNA or protein.

✿ **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.

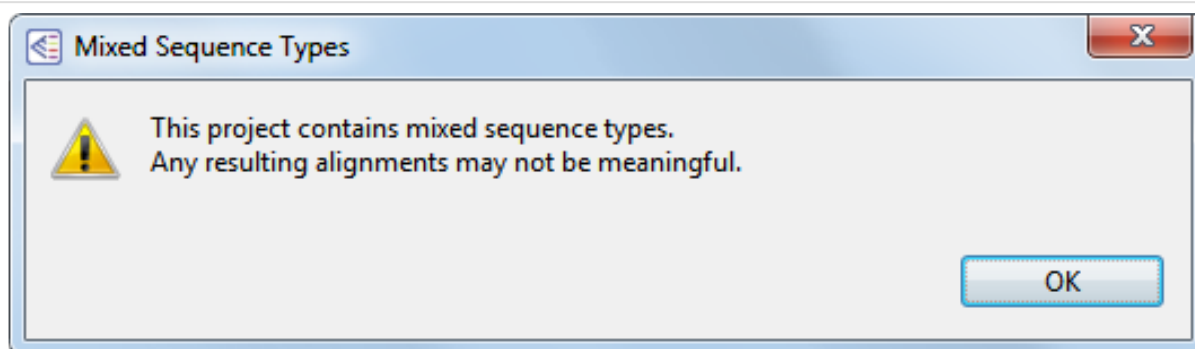
Situation	Message and response
Adding a MegAlign Project (.msa file)	 <p>Click OK to add unaligned sequences from the saved project to the current project. This is one way to remove gaps from sequences so that they can be used in MegAlign Pro.</p>

Adding a sequence that contains gap characters (e.g., dashes or question marks)



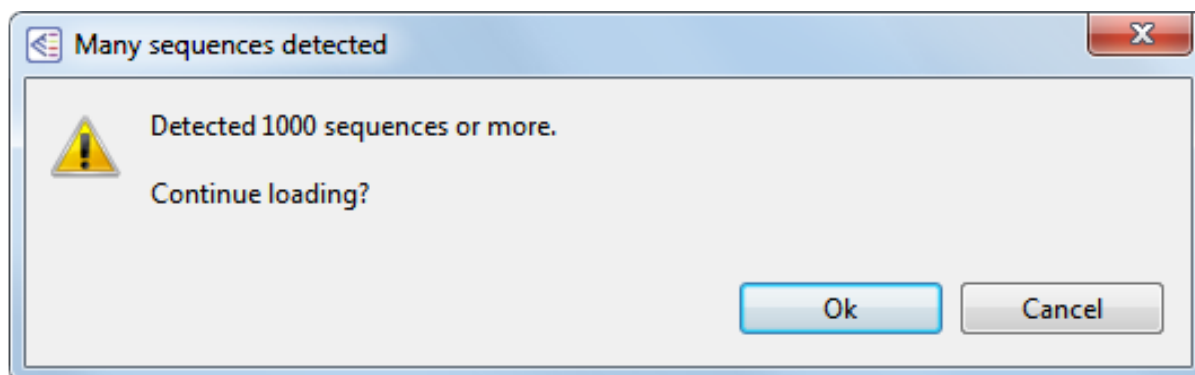
Click **OK** to add only those sequences without gaps to the project. (There is a separate method for [adding gapped sequences](#)).

Attempting to enter mixed sequence types (both protein and nucleotide) to the Sequences View

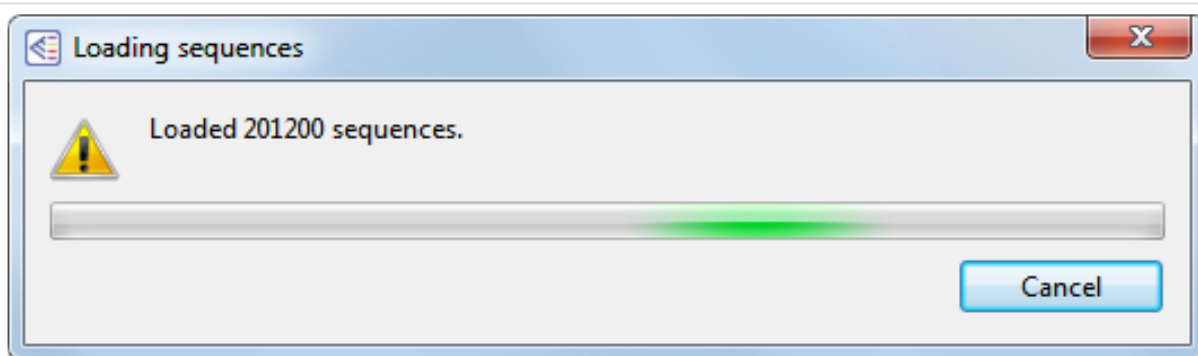


Click **OK** to add all the sequences to the project. Afterwards, manually remove any sequences that are not of the desired category.

Attempting to load more than 1000 sequences

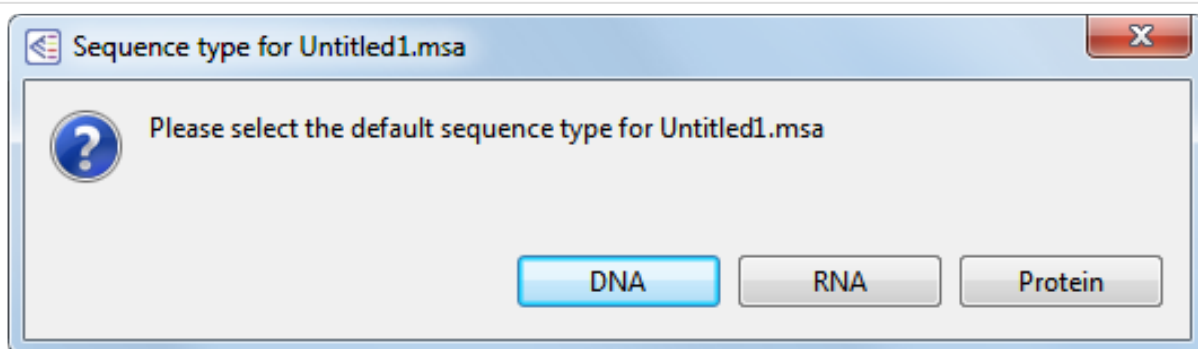


After clicking **OK**, a progress bar is displayed:



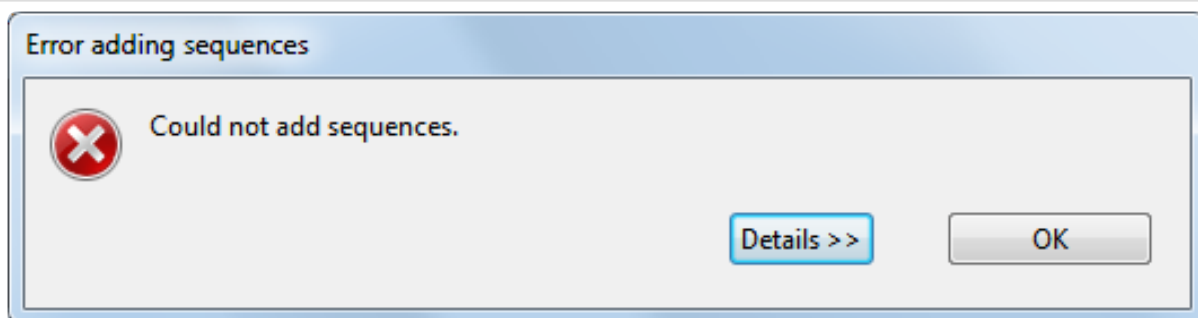
Press **OK** to continue loading all of the sequences. MegAlign Pro will display a progress bar (see lower image, left) with a running count of the number of sequences loaded so far. At any point, you may cancel the entire sample upload by pressing **Cancel**. This terminates the entering step without 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.

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.

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 “Unaligned 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.

- Launch DNASTAR [SeqMan Pro](#), then drag and drop the gapped sequence file(s) onto the SeqMan Pro window.
- Select all sequences in the Unassembled Sequences using **Ctrl+A** (Win) or **Cmd+A** (Mac).
- Use **Contig > Export Sequences > Single File**.
- 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 multi-segment sequences to a project

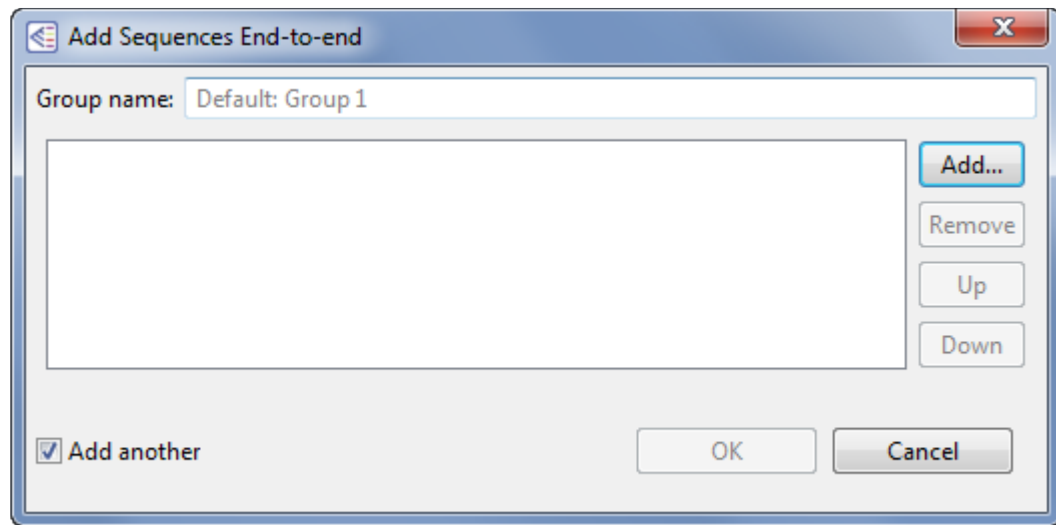
Sequences that consist of more than one chromosome, contig, or fragment are called multi-segment sequences. By entering a group of sequences into MegAlign Pro using a special command, the entire data set will be concatenated end-to-end and treated as a single multi-segment sequence.

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.

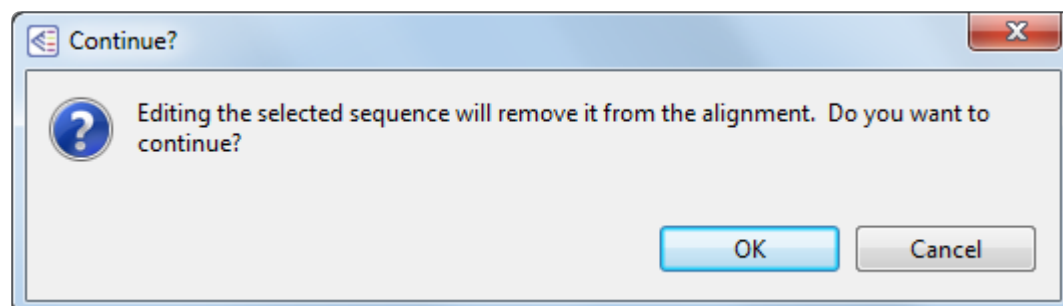
Click **OK** to keep your changes and add the sequence group(s) to the project, or **Cancel** to leave the dialog without adding the sequence group(s) to the project.

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 download and extracted the tutorial data, [click here to download it](#). Then decompress (unzip) the file archive using the method of your choice.

Add sequences from a saved project to an active project

To add sequences from a saved project to an active project:

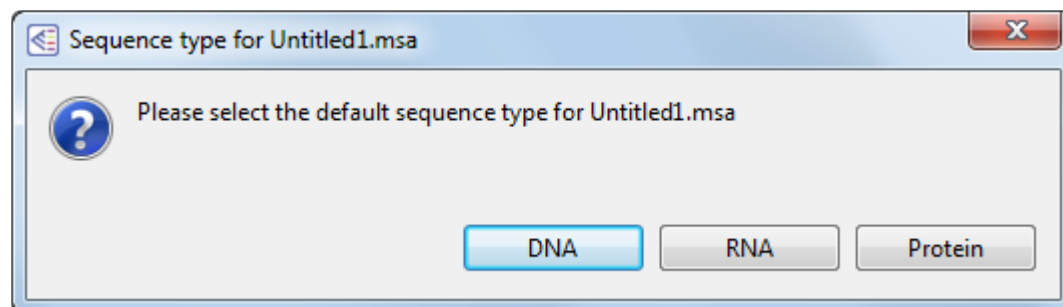
Choose **File > Add Sequences** and select a MegAlign Pro project file (*.msa*). All sequences in the *.msa* file will be added to the “Unaligned Sequences” area of the active project.

Manually specify sequence type

To manually specify a sequence type for the active project:

In most cases, MegAlign Pro can read whether a sequence is DNA, RNA or protein using clues in the sequence itself. In rare cases, you may need to over-ride or manually specify the sequence type.

Choose **Edit > Change Sequence Type**. After you select the desired sequence type by pressing its button, MegAlign Pro will treat all sequences in the project as being of the specified type.



Change the order of sequences

To change the order of sequences in views:

Sequences may be rearranged one at a time or as contiguous groups using drag and drop. To move several sequences as a group, select them with **Ctrl/Cmd+click** or **Shift+click**, then drag them to the desired location and release the mouse button. Any ordering changes made to the [Sequences view](#) are simultaneously made to the [Distance view](#) and to the [Overview](#).

To reorder sequences in the other views in the same order as the Tree view:

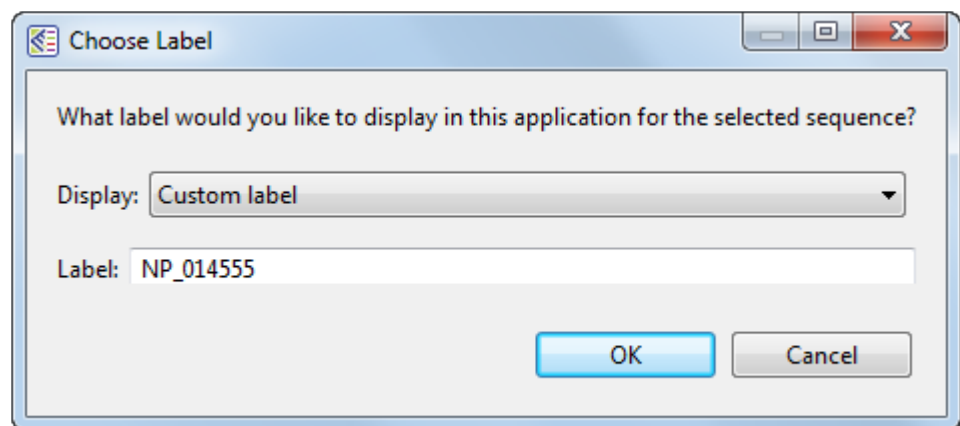
The Tree view sequences are ordered top-to-bottom in the vertical formats and left-to-right in the horizontal ones. This means that the order of sequence names shown in the phylogenetic tree may vary from ordering seen in the Sequences and Distance view and the Overview. To force the other views to reorder their sequences to agree with the [Tree view](#), use **Tree > Order Sequences Like Tree**.

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**. Alternatively, select the sequence name and choose **Edit > Rename**.

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.

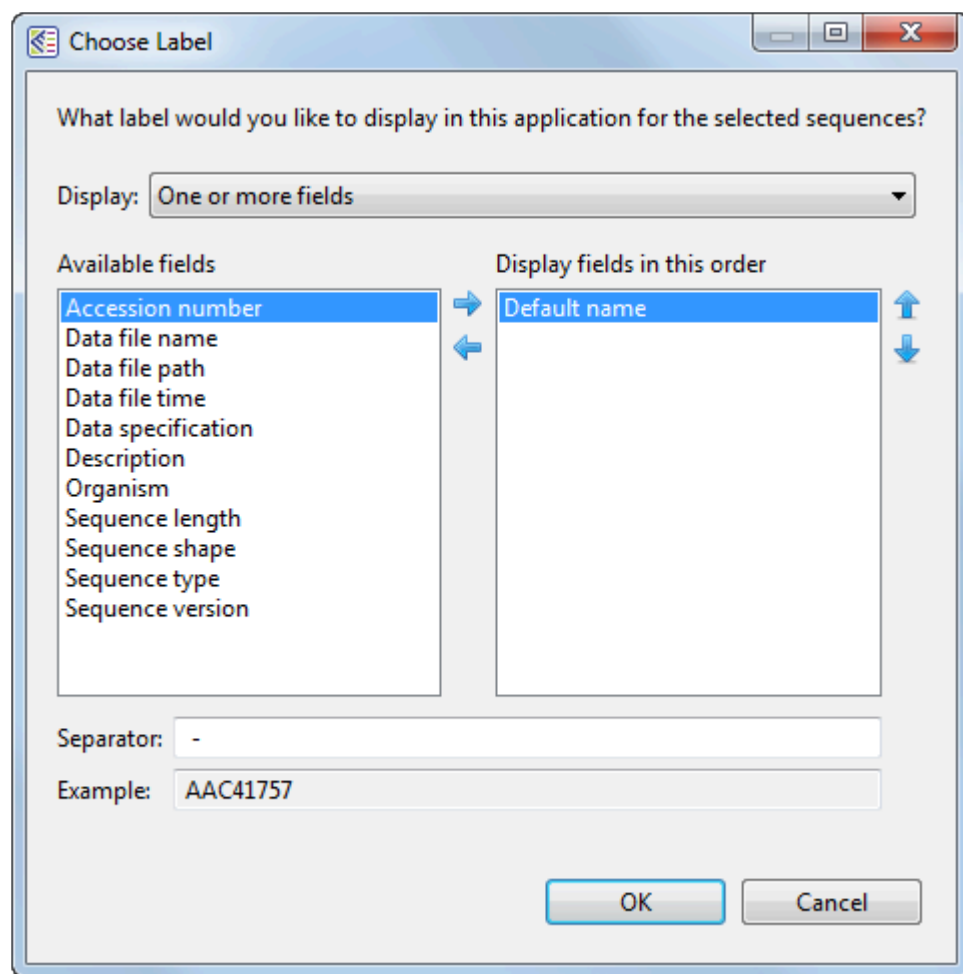
Rename sequences automatically using specified data fields

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

To rename sequences using specified data fields:

1. Select one or more sequence names from any MegAlign Pro view.
2. Right-click on the selection and choose **Rename with Fields**, or instead choose **Edit > Rename 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.

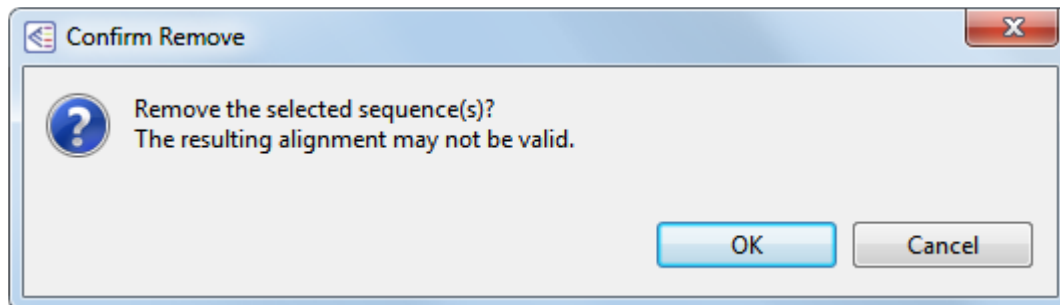
✿ **Note:** When you export data for a renamed sequence, the new name may be exported or not, depending on the situation. See [Retain custom names during export](#) for details.

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.

Tracks




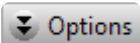
“Tracks” are rows of information that can be displayed or hidden, as desired, in the [Overview](#) and [Sequences view](#). Tracks can pertain to a header, footer, or individual sequence row. The following video provides a brief overview of how tracks are used in MegAlign Pro:

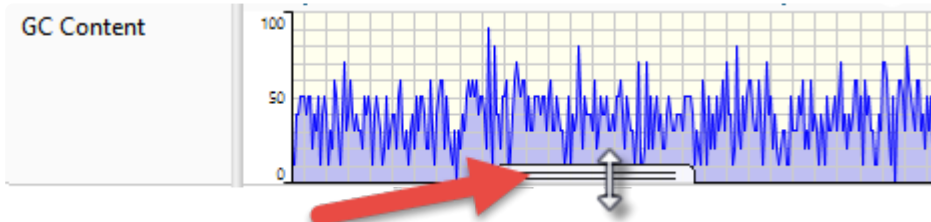
The following tracks can be applied, removed or (in most cases) edited:

- [Ruler](#)
- [Feature](#)
- [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...
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).
To 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.
To expand/ collapse tracks for selected sequences	<p>Do any of the following:</p> <ul style="list-style-type: none"> 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	<p>Ensure that no sequences are selected.</p> <ul style="list-style-type: none"> 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.
To hide tracks	<p>Do one of the following:</p> <ul style="list-style-type: none"> <u>To hide a specific track from all views</u>, uncheck the box next to the name of the track in the Tracks panel. <u>To hide a specific track in one view only</u>, 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. <u>To hide all tracks in one view only</u>, 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. <u>To hide all tracks in all views</u>, do a combination of the above.
To access the options for a particular type of track	<p>Do either of the following:</p> <ul style="list-style-type: none"> 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 (), then do either of the following to specify the track of interest

	by clicking on its name in the Tracks panel .
To 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.
To increase/decrease the vertical height of a visible track	<p>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.</p> 
To see how the zoom level can affect the appearance of tracks	See Zoom in or out on a view .

Ruler tracks

To view consensus rulers:

Blue consensus rulers number and tick any gaps found in the consensus. The visibility of consensus rulers in the [Overview](#), [Sequences view](#) and [Pairwise view](#) is dependent on:

- Which boxes you check in the [Tracks panel](#) (see image below).
- - **Ruler** – Overview header & Pairwise view
 - **Sequence Ruler** – Overview footer
 - **Ruler 1** – Sequences View header
 - **Ruler 2** – Sequences View footer



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

To view gap-aware sequence rulers:

Unlike the blue consensus rulers described above, black “gap-aware” sequence rulers use pre-alignment numbering for the sequences. The visibility of sequence rulers in views is dependent on:

- Which boxes you check in the Tracks panel. Boxes for **Sequence Ruler** appear in three parts of the Tracks panel: **Overview Details**, **Sequence Details**, and **Pairwise Details**.
- Whether or not tracks are visible in the views. To display hidden tracks, click on the plus sign next to a sequence in the view of interest.

✿ **Note:** 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.

To change Ruler options:

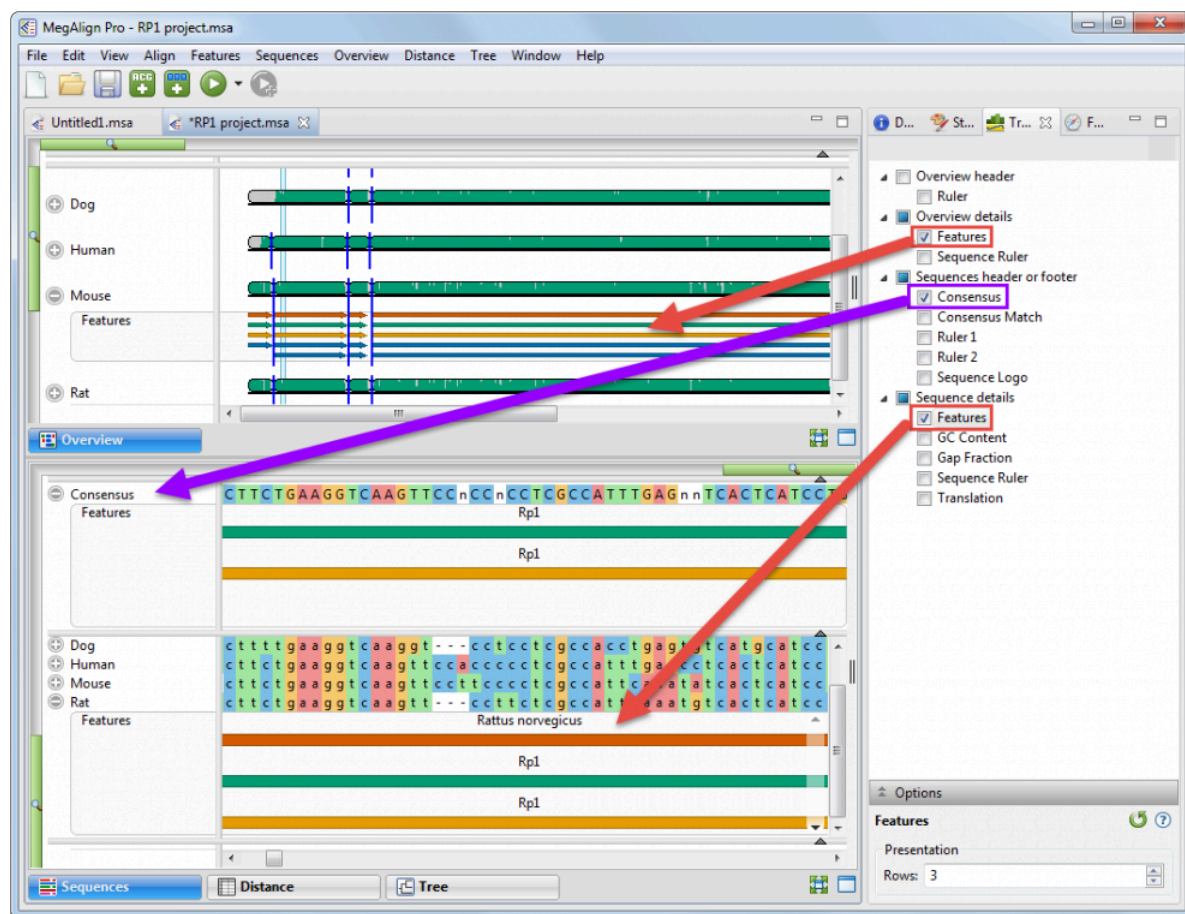
Currently, the Ruler tracks do not have editable options. However, you can add white space above and/or below the ruler using sliders in the [Layout section](#).

Features tracks

To display sequence features:

The visibility of feature (annotation) 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 Overview or Sequences view.
- Which boxes you check in the [Tracks panel](#) (see image below). Features can be displayed in individual views by clicking the **Features** boxes under: Overview details, Sequence details, and Pairwise details. Hover the mouse over a feature to view a tooltip with details about the feature. 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.
- **Features** (under Pairwise details*) – Check this box to display any available sequence features as Pairwise



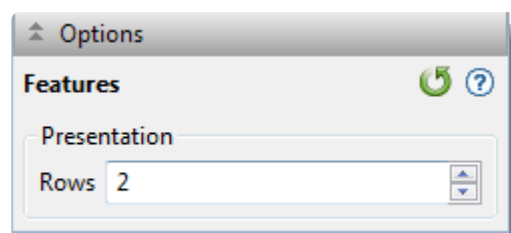
To view a tooltip with details about a feature, hover over it with the mouse. Or select one or more features by clicking on them. You can then view information about them in the [Details panel](#).

The same default color scheme is used for features in both MegAlign Pro and [GenVision Pro](#).

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

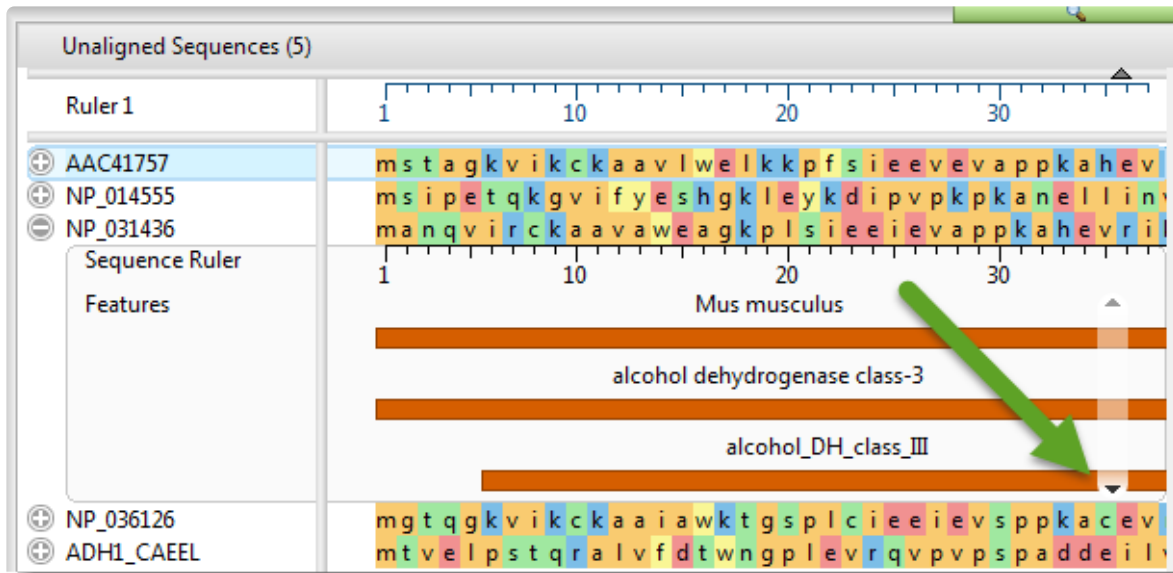
To change Features options:

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](#).

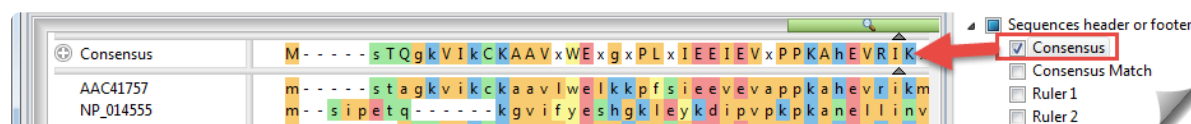
Click  if you wish to return to the default values.

Consensus track

To display the consensus:

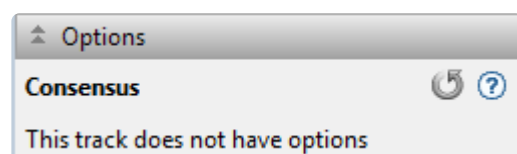
After any type of alignment, a consensus sequence is automatically displayed in the [Sequences view](#) header. To display or hide this consensus sequence, check or uncheck the **Consensus** box in the [Tracks](#) [panel](#). This box is only visible after performing an alignment.

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.



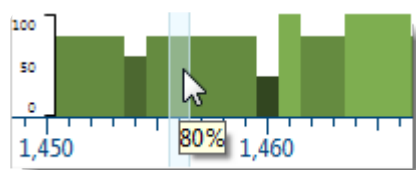
Consensus Match track

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

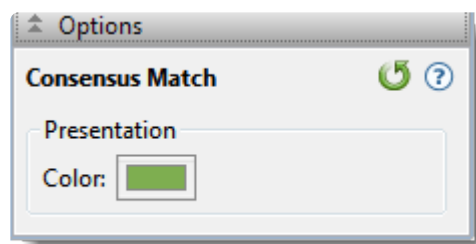
match is the total score for the called character divided by the number of sequences at the position.

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.

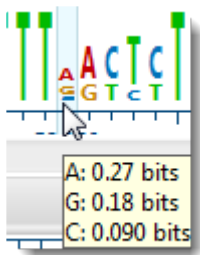
Sequence Logo track

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.

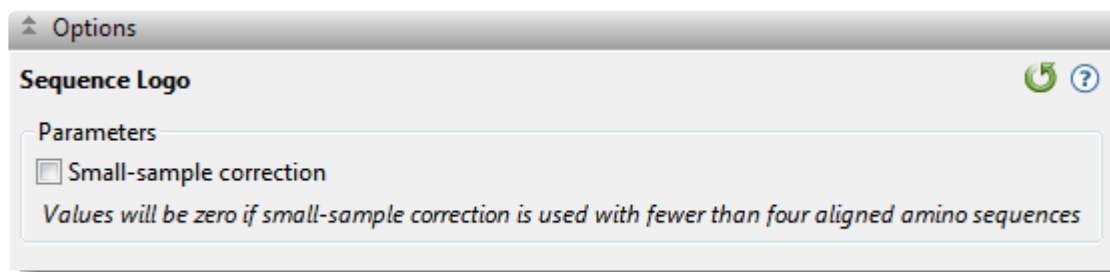
How Sequence Logo is calculated:

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 is proportional to the amount of information available for each character. 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](#).



- 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. We recommend checking **Small sample correction** when there are relatively few sequences without gaps represented at a given position. This correction has no effect when there are sufficient numbers of sequences to provide a good estimate of information content. Values will be zero if **Small sample correction** is used with fewer

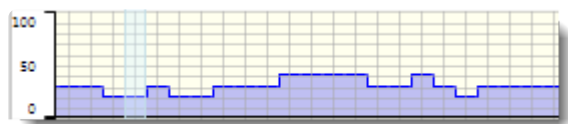
than four aligned amino acid sequences.

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.

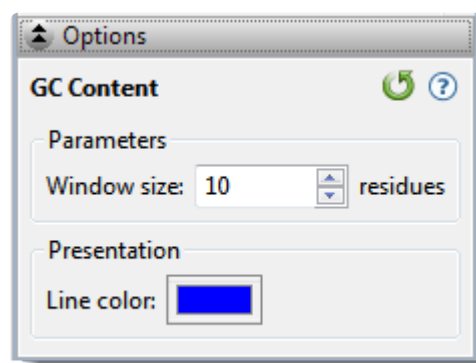


The visibility of GC Content tracks in the [Sequences view](#) is available only for nucleotide (DNA) sequences, and is dependent on:

- 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.

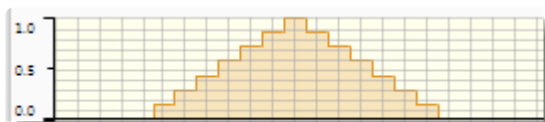
- By default, GC Content is shown in tones of blue. To choose another color, click on the color box to the right of **Line color**.

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.

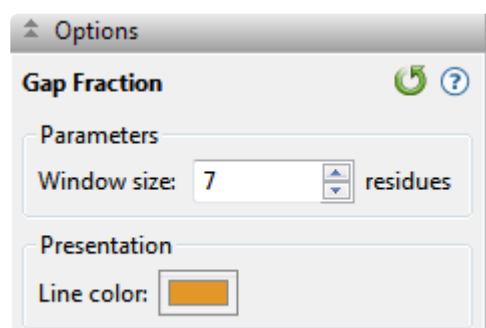


The availability of a Gap Fraction track in the [Sequences view](#) or [Pairwise view](#) is available only for aligned sequences, and is dependent on:

- 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

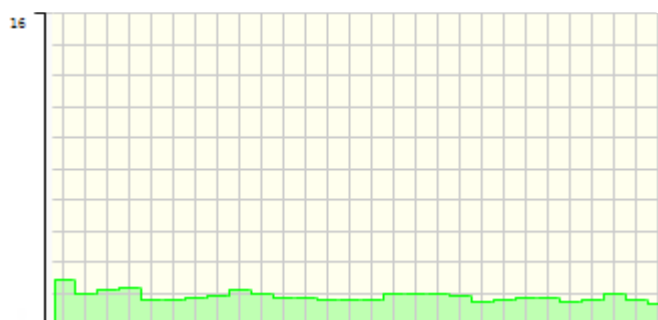
A numeric track displays a plot of values along one or more sequences in the project.

To display numeric tracks:

The visibility of numeric tracks in the [Sequences view](#) is dependent on:

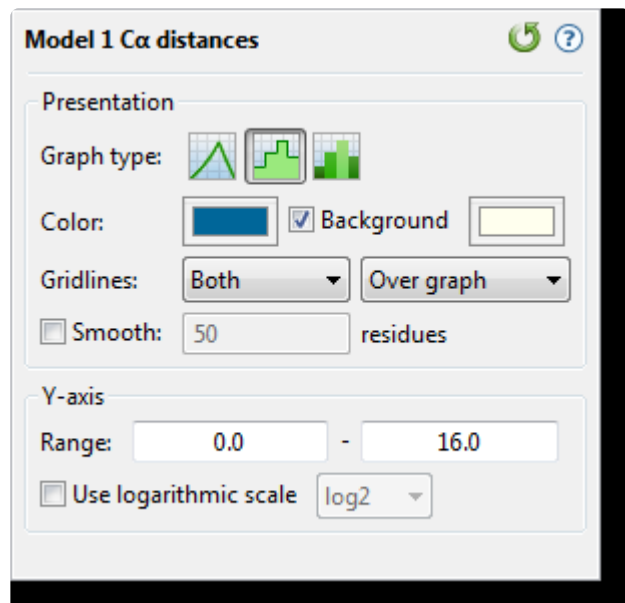
- The origin of the project. For example, the Model 'n' C α Distances track is only available only for the query sequence in a project sent to MegAlign Pro using NovaFold's **Send template alignment to MegAlign Pro** link. It is not available for template sequences in those projects, nor for MegAlign 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' C α Distances track is a graph with a y-axis from 0-16. It shows the predicted pairwise distances between C α 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**.

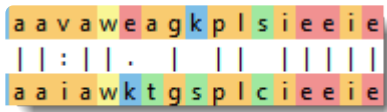
✿ **Note:** The log of 0 is not defined. Therefore, log values for the track are calculated as follows: For $y > 0$, $y' = \log(y)$; otherwise, it is displayed as a **0**.

Click  if you wish to return to the default values.

Match Bar track

To display the Match Bar track:

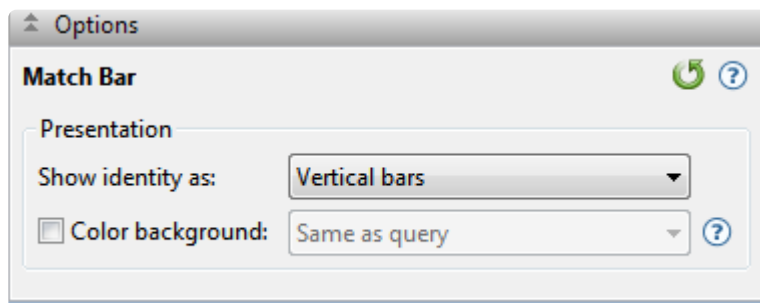
The Match Bar track is used to show similarities and differences between the query and target sequences. This track is available in the [Pairwise view](#) only after [performing a pairwise alignment](#).



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. 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	(0,1)
Period	Neutral mismatch	0
Space	Unfavorable mismatch	< 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.

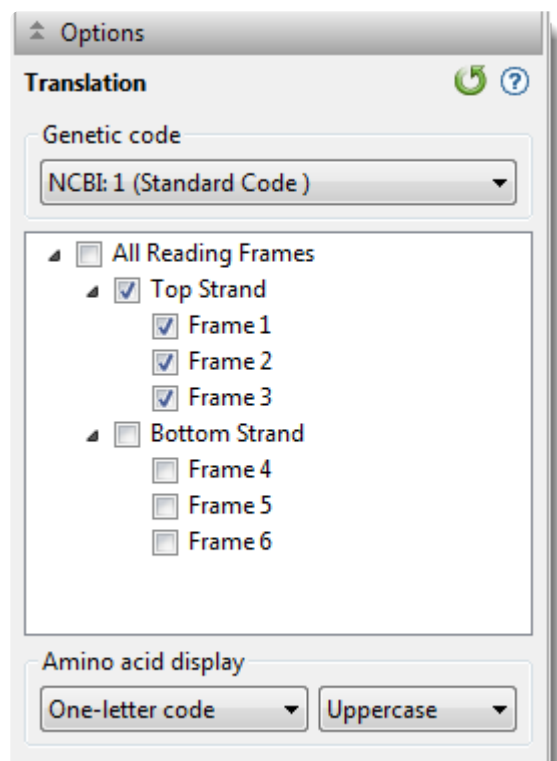
Translation track conventions:

- 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.


1	L	L	L	R	Q	•
2	C	S	C	A	N	E
3	L	A	P	M	S	L
4	A	R	A	G	I	L
5	K	R	W	H	A	E
6	Q	E	Q	A	L	S

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.

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](#)

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:

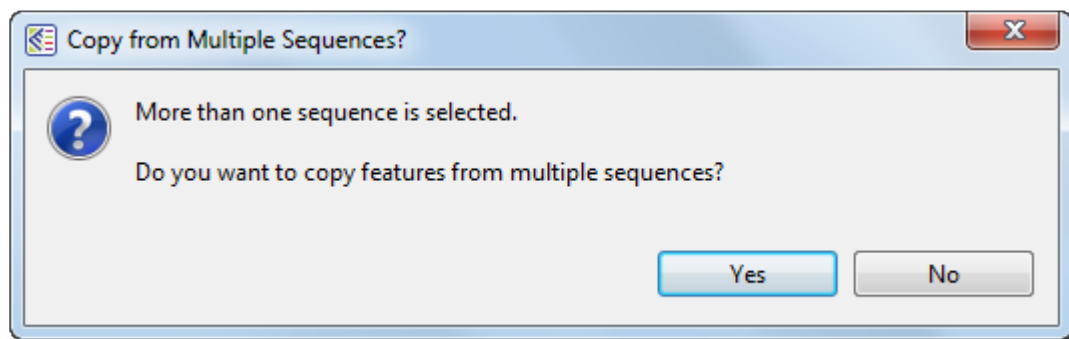
To map features from one or more source sequences to a target sequence:

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
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.

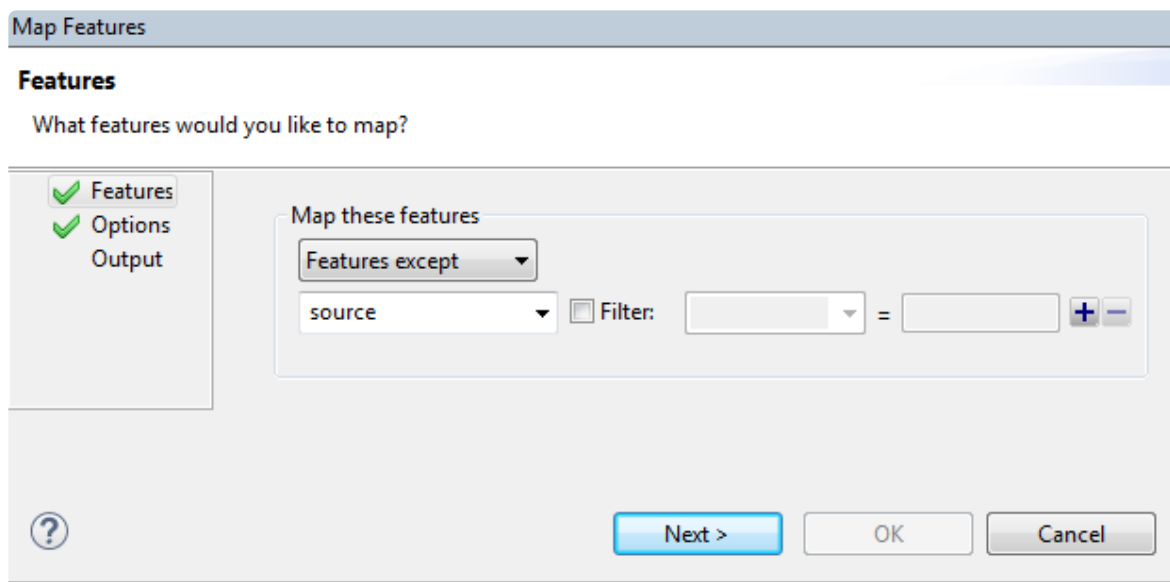
Things to note when mapping features:

- 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

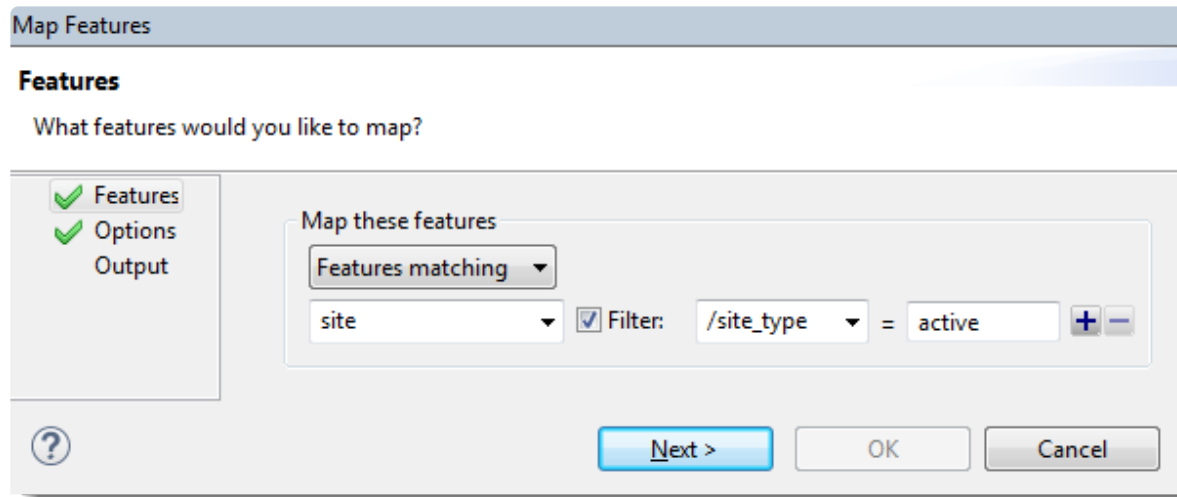
The Features screen is the first wizard screen of the [Map Features dialog](#), and allows you to choose the features you want to include or exclude in the mapping.

✿ **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:



The dialog box is titled "Map Features". It has a section titled "Features" with the question "What features would you like to map?". On the left, there is a list of three items: "Features" (checked), "Options" (checked), and "Output" (unchecked). In the center, there is a section titled "Map these features" containing a dropdown menu set to "Features matching", a dropdown menu set to "site", a checked checkbox labeled "Filter:", a dropdown menu set to "/site_type", an equals sign, a text field containing "active", and plus/minus buttons. At the bottom, there are three buttons: "Next >" (highlighted in blue), "OK", and "Cancel". A help icon (?) is located in the bottom left corner.

Map Features

Features

What features would you like to map?

✓ Features
✓ Options
Output

Map these features

Features matching

site Filter: /site_type = active

Next > OK Cancel

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.

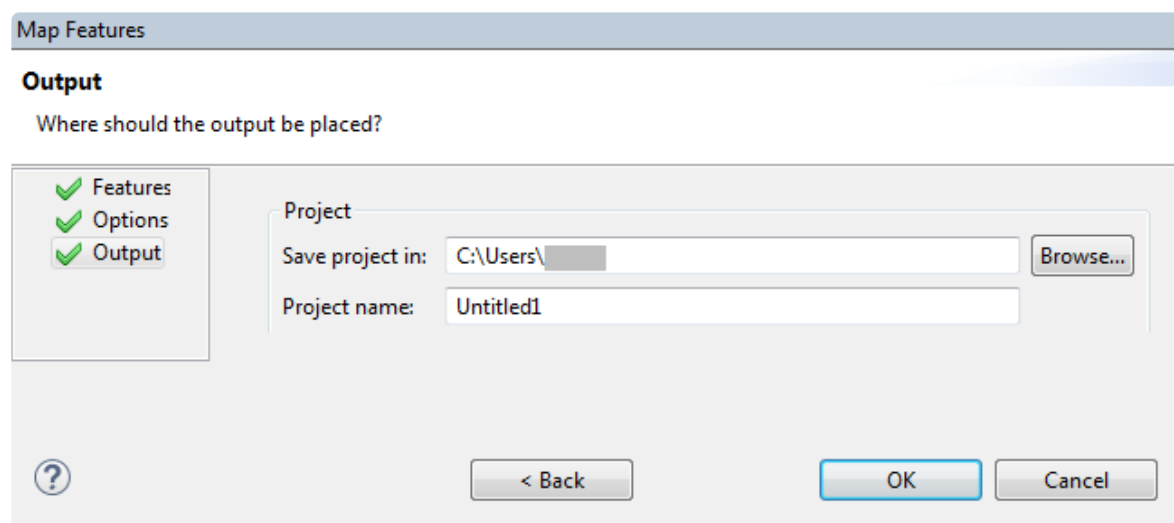
Feature coverage is less sensitive to insertions and substitutions than is sequence identity. The default for both Mapping items is 80%.

- 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
<i>PROJECT_map_features_console.txt</i>	Copy of the text that appeared for this mapping in the Console side panel in the MegAlign Pro user interface.
<i>PROJECT_map_features_report.tsv</i>	See the text and table below this one for more information.
<i>PROJECT_map_features_unmapped.tsv</i>	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 <i>.vcf</i> 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:

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)** from the context menu.

To remove all features from the consensus:

Select the consensus and choose **Features > Remove Features from Active Consensus**; or right-click on the consensus 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](#).

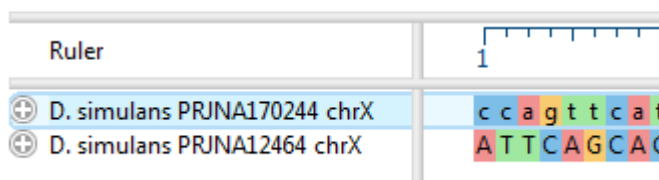
Part A: Add, rename and align sequences


Before mapping features in MegAlign Pro, you first need to [add](#) and [align](#) the sequences.

1. If you have not yet download 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.




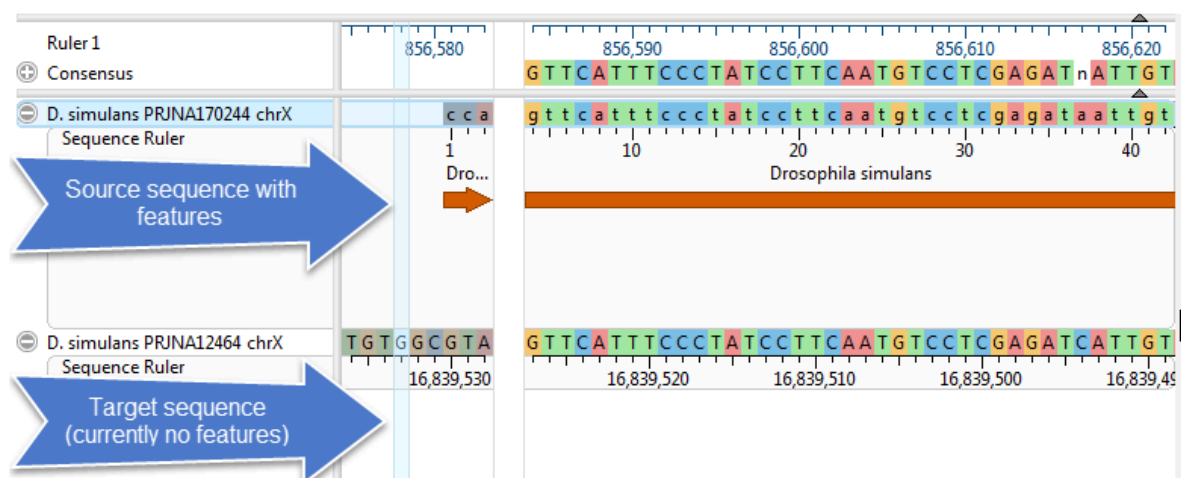
8. To align sequences with [Mauve](#) using the default parameters, choose **Align > Align Using Mauve** or click the **Align** tool arrow () and choose **Align Using Mauve**.
9. Use **File > Save** to save the project with the name *Drosophila project.msa*.

Proceed to [Part B: Map a single feature](#).

Part B: Map a single feature

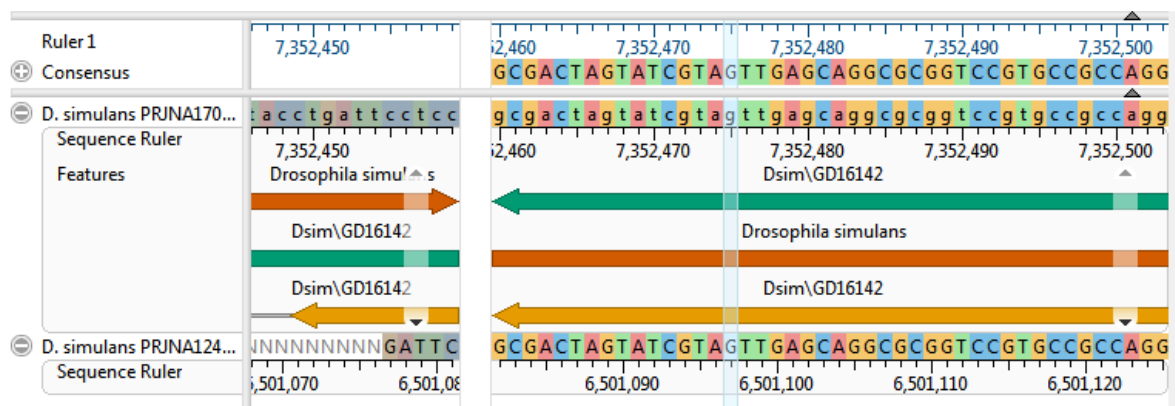
In this part of the tutorial, you will map a single selected feature from the source sequence to the target sequence. In order to choose the appropriate mapping command, you will need to know where the source and target sequences appear in relation to one another in the list on the left of the views. 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.

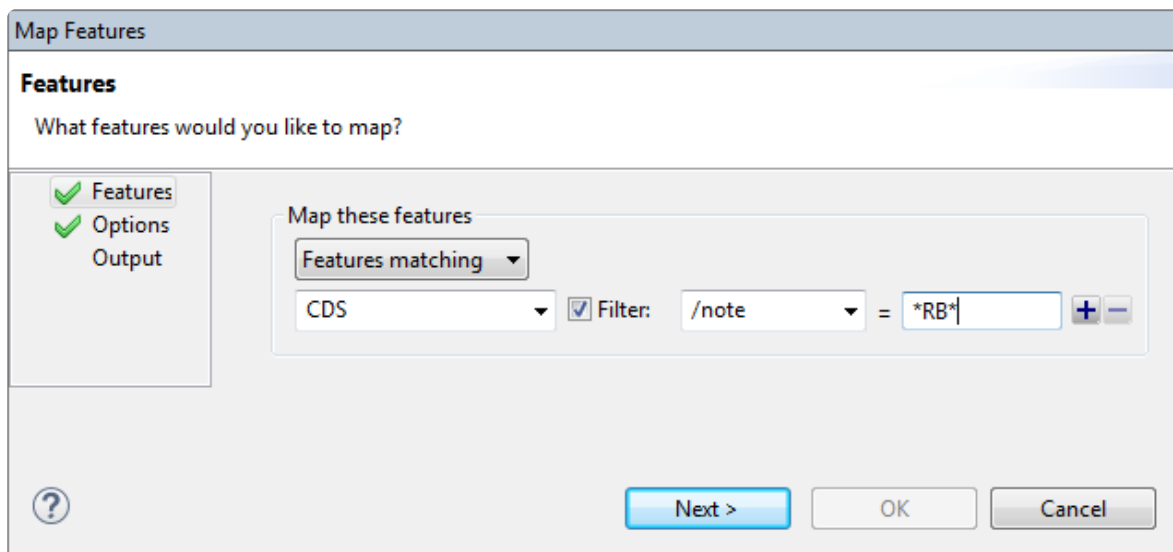
1. Right-click on the name of the source (upper) sequence and select **Features > Map Features to Bottom Sequence**. The Map Features wizard launches.
2. In the [Features screen](#), use the drop-down menu to select **All Features**, then press **Next**.
3. In the [Options screen](#), keep the defaults and press **Next**.
4. In the [Output screen](#), 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 [Part D: Map a filtered set of features](#).

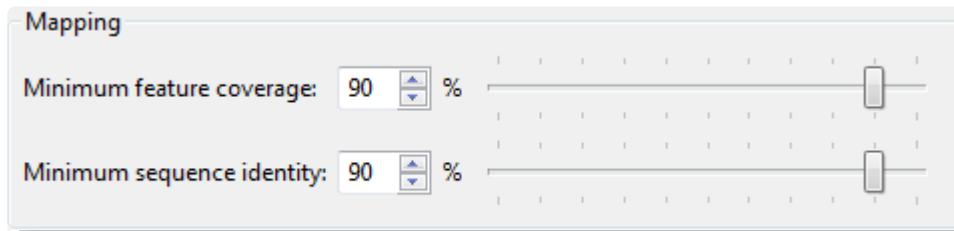
Part D: Map a filtered set of features

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**.



The image shows a 'Mapping' dialog box with two settings. The first setting is 'Minimum feature coverage' set to 90%. The second setting is 'Minimum sequence identity' also set to 90%. Each setting has a numeric input field with up/down arrows and a corresponding slider bar to the right.

Mapping

Minimum feature coverage: 90 %

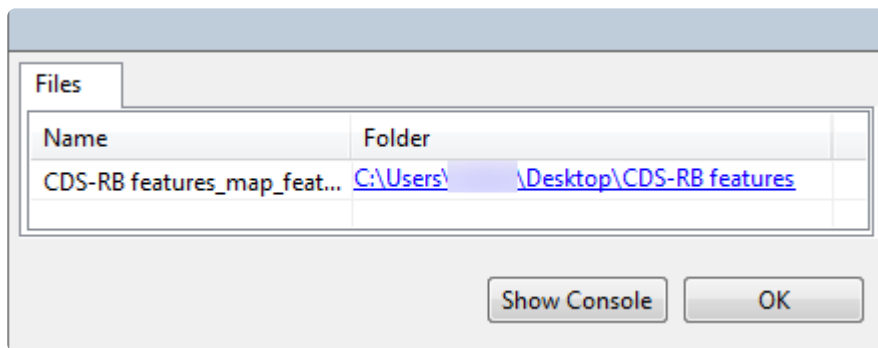
Minimum sequence identity: 90 %

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 [Part D](#) 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.

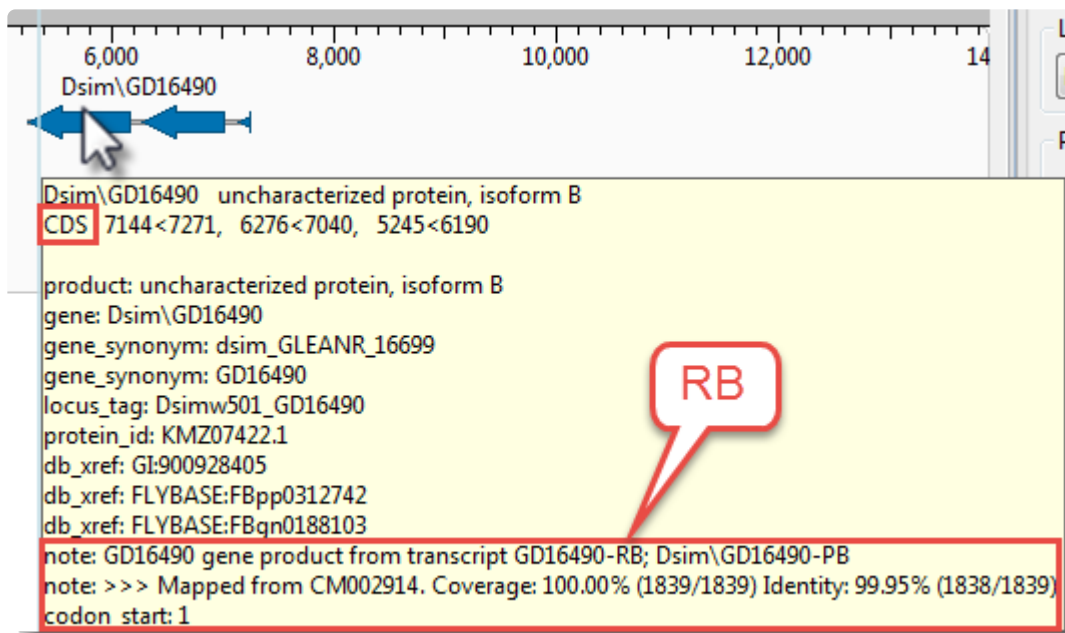
Proceed to [Part F: Export the annotated target sequence](#).

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.

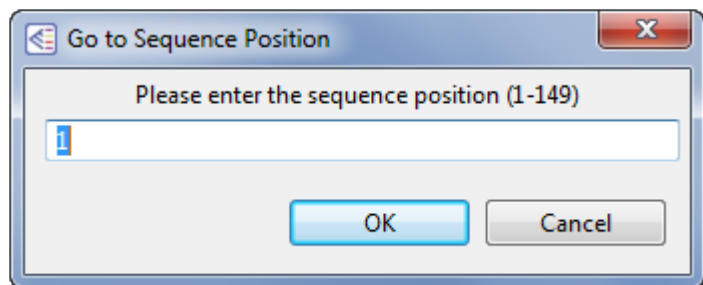
This is the end of the Map Features tutorial.

Search Within an Open Project

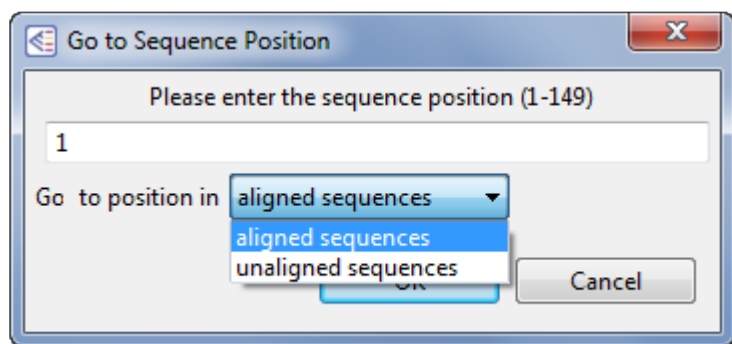
To go to a specified position:

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**.

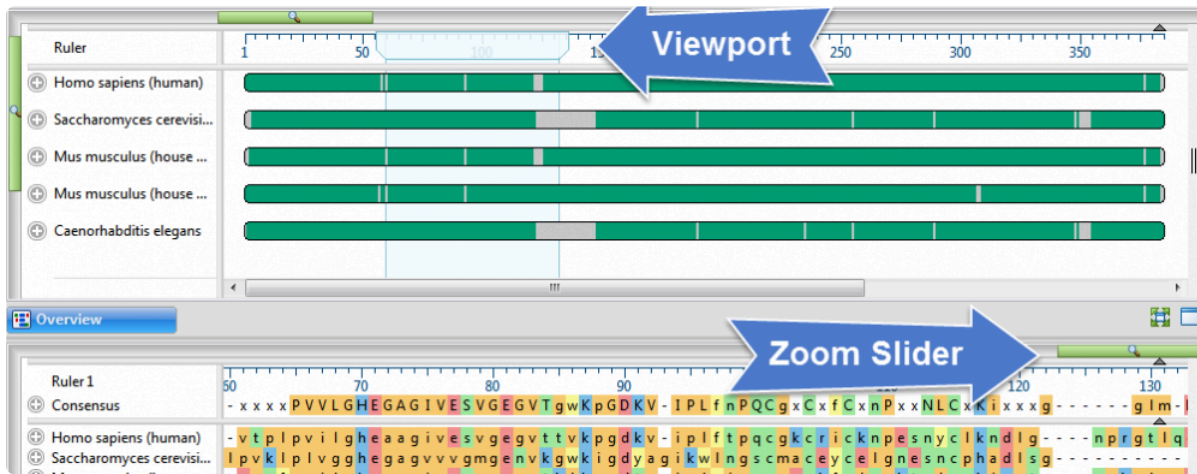


- If the project contains a mixture of aligned and unaligned sequences, type in a number and use the drop-down menu to choose whether to go to the position in the aligned or unaligned sequences. Then click **OK**.



To see a specific range in the Sequences view:

In the [Overview](#), double-click on an 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. Using the [zoom sliders](#) in the Sequences view controls the width of the viewport, and thus, the amount of visible sequence.



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	<p>Click on the name of any sequence, then use Ctrl/Cmd+A or Edit > Select All.</p> <ul style="list-style-type: none"> 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	<p>Either:</p> <ul style="list-style-type: none"> 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.
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	Drag the mouse across the open track while holding down the left mouse

sequence associated with a section of track (e.g. a Gap Fraction peak, etc.)	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	<p>To expand a selection consisting of one or more features to a selection containing just the sequence corresponding to the feature(s), either:</p> <ul style="list-style-type: none"> Choose Edit > Select Subsequence or use the context (right-click) option of the same name. Double-click on the feature under that sequence.
The portion of all aligned sequences associated with a feature	<p>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:</p> <ul style="list-style-type: none"> 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 endpoint has been selected	Use Ctrl/Cmd+Shift and click on the other endpoint.
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.

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.



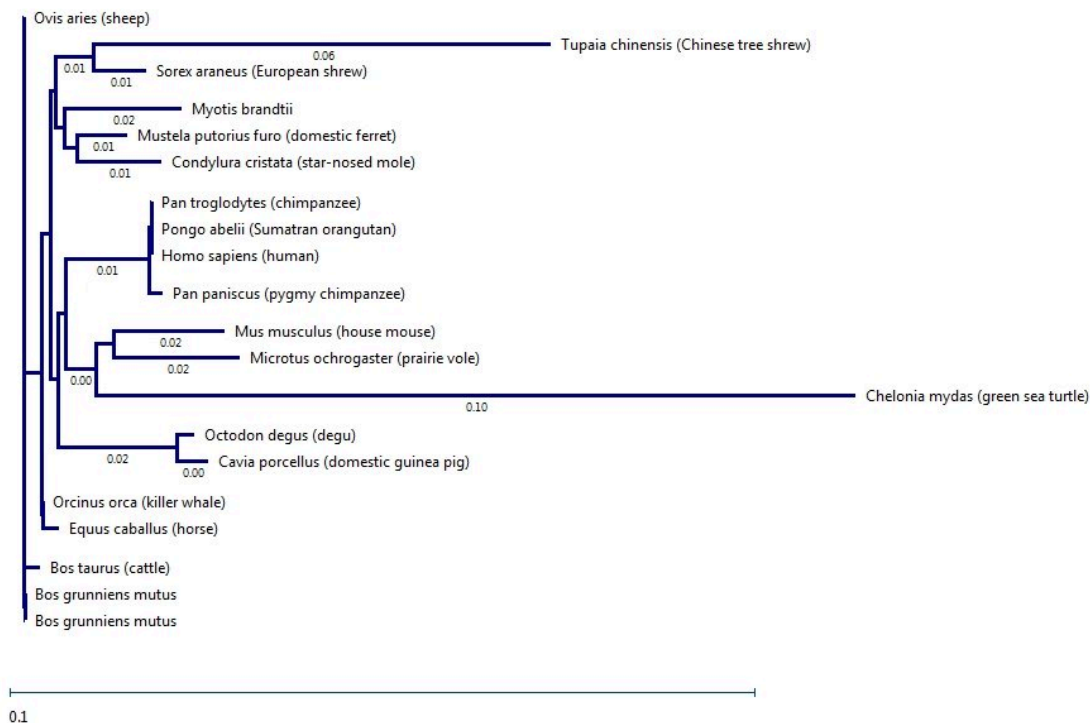
Choose the Alignment Type

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.

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 > Sequence Search	Launches the wizard at the Query screen .
Use the "Change search strategy" tool () from the Text , Table or Pairwise views	
From the Welcome tab, choose Search and then Search for sequences .	
Search > Sequence Search with Selection	Launches the wizard at the Query screen , whose table is pre-populated with the currently-selected portion of sequence.
Press Ctrl/Cmd+B	
Search > Sequence Search with All Sequences	Launches the wizard at the Query screen , whose table is pre-populated with all sequences in the document.

Enter

Enter sequences in FASTA format

```
>Hemoglobin (Alpha Chain)
VLSPADKTNVKAAWGKVGAGHAGEYGAEALERMFLSFPTTK
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

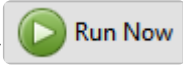
The Search screen is the second of four screens in the [BLAST Search wizard](#). The Search screen is used to choose the BLAST program and database.

- 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.

Search Online

Options

Choose common options that apply to all searches.

☒ Query
☒ Search
☒ **Options**
☒ Job

Selection Matches

Entrez query:

☐ Ignore low complexity in query sequence
☐ Ignore low complexity during database lookup
☐ Ignore lower case letters

? ... Key ✓

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:

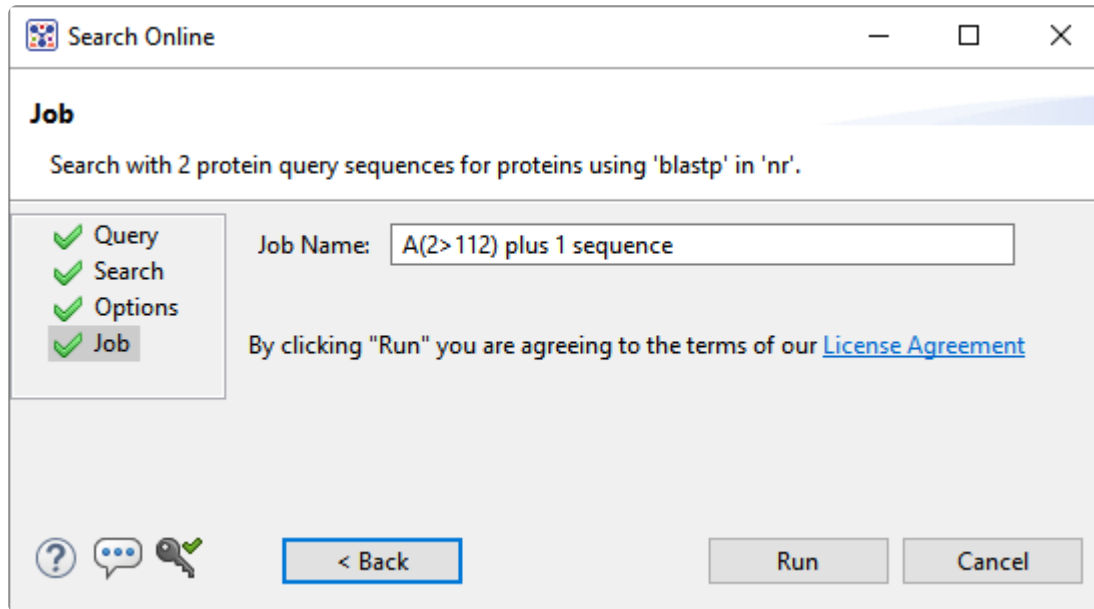
- To continue filtering, press **Next** to continue to the [Job screen](#).
- To start the search immediately, press the **Run Now** button () or **Run** button. The [Jobs panel](#) in the application will open and contain information on the status of the search. During the search, a progress dialog 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)

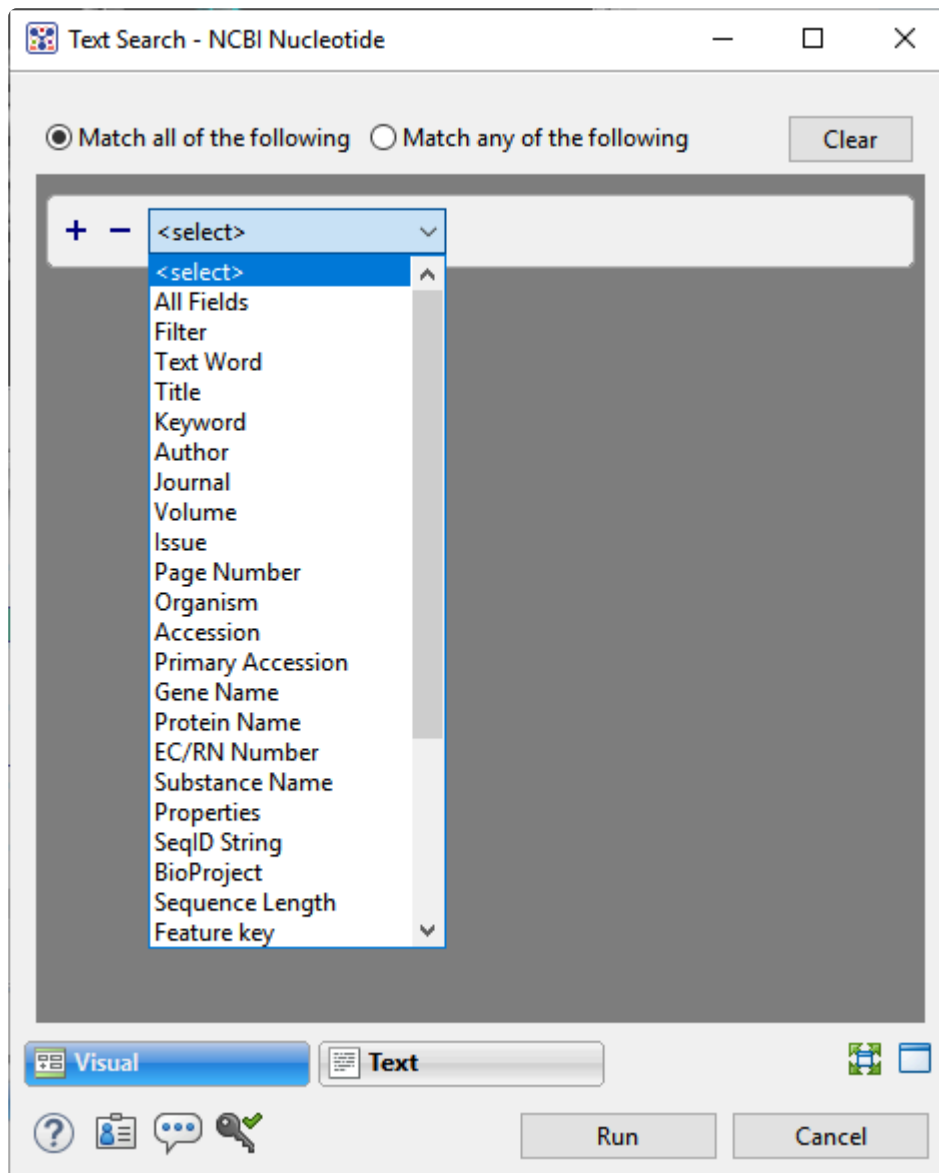
The Entrez Search wizard is used to [search](#) a text query or accession number against NCBI's Entrez databases for nucleotides or proteins.

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 [Text view](#).

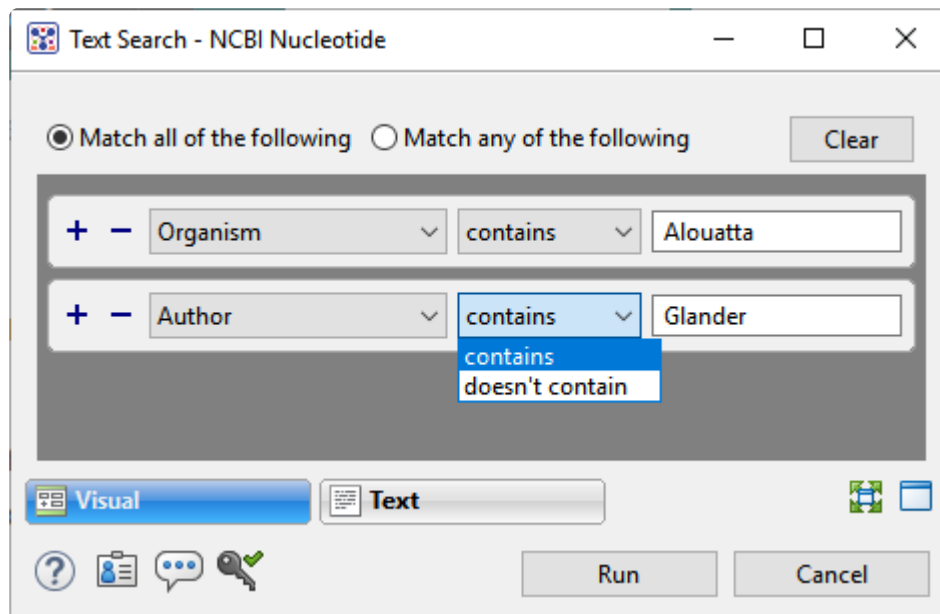
1. Use the **<select>** drop-down menu to choose a criterion. The field headings vary depending upon the online database that you are using. Choosing **All fields** will cause the application to search every word of every listing for the query term in the database.



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 should be found in any results (in the case of **contains**) or should not 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 (-).

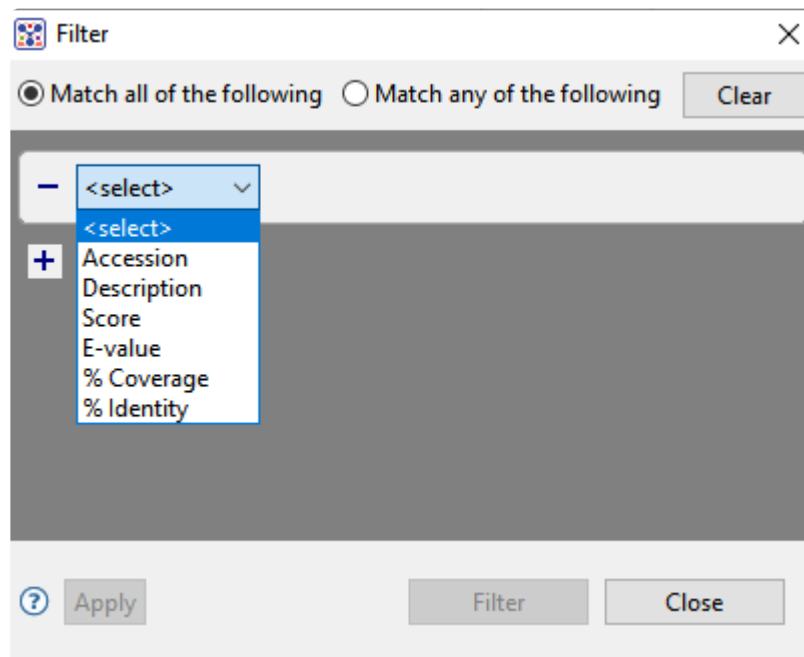


3. Once you have set up all desired criteria, go to the top of the dialog and choose whether to **Match all of the following** or **Match any of the following** criteria.
4. To learn about the optional tools in the bottom left corner of the screen, see the bottom of [this topic](#).
5. Click **Run** 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 sequence search results

When you press the **Filter results** tool () in the header above the [Table](#) or [Pairwise](#) views, there are two options: **Filter** and **Clear Filter**. Clicking on the **Filter** option opens the following dialog:



- Use the **<select>** drop-down menu to choose a filter type, then enter the desired options.

Filter type	Description	Options
Accession	The Accession number.	
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 coincidentally. (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%.	

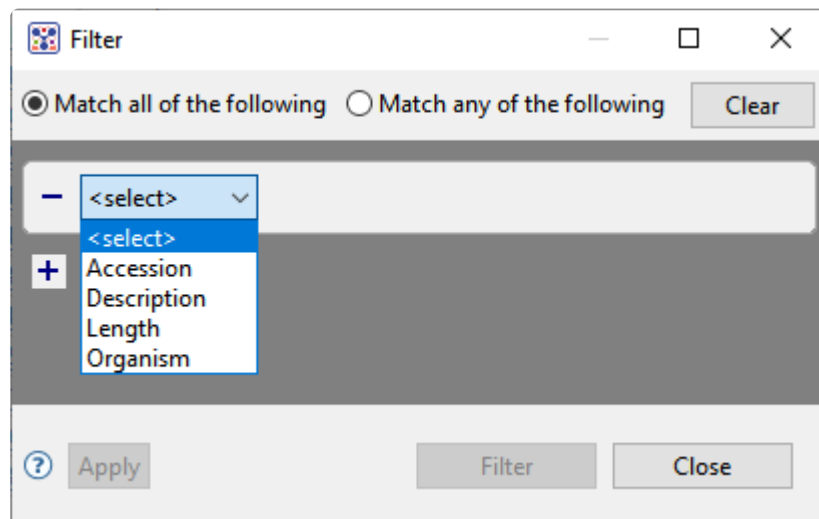
- To add another filter, press the plus icon (+). To delete a filter row, press the minus icon (-).
- Once you have set up all desired queries, go to the top of the dialog and choose whether to **Match all of the following** or **Match any of the following** queries.

Once you are finished setting up filters, do any of the following:

- To apply the filters, press **Apply** or **Filter**.
- To clear all filters and start over, press **Clear**.
- To exit the dialog without filtering, press **Close**.

Filter text search results

When you press the **Filter results** tool (🔍) in the header above the [Table](#) or [Pairwise](#) views, there are two options: **Filter** and **Clear Filter**. Clicking on the **Filter** option opens the following dialog:



- Use the **<select>** drop-down menu to choose a filter type, then enter the desired options.

Filter type	Description	Options
Accession	The Accession number.	
Description	The NCBI “Definition” field text.	
Length	The length of the match sequence in bases.	
Organism	The species name.	

- To add another filter, press the plus icon (+). To delete a filter row, press the minus icon (−).
- Once you have set up all desired queries, go to the top of the dialog and choose whether to **Match all of the following** or **Match any of the following** queries.


Once you are finished setting up filters, do any of the following:

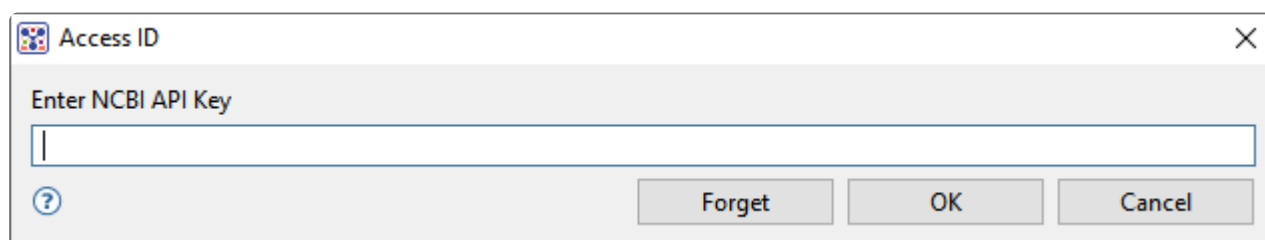
- To apply the filters, press **Apply** or **Filter**.
- To clear all filters and start over, press **Clear**.
- To exit the dialog without filtering, press **Close**.

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.



The image shows a dialog box titled "Access ID" with a close button (X) in the top right corner. Inside the dialog, there is a label "Enter NCBI API Key" above a text input field. Below the input field, there is a help icon (question mark in a circle) on the left and three buttons: "Forget", "OK", and "Cancel" on the right.

- 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**.

Perform a Multiple Alignment

Did you arrive here by selecting the  DNASTAR Navigator workflow **Molecular Biology > Multiple sequence alignment**? If so, you're in the right place!

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.

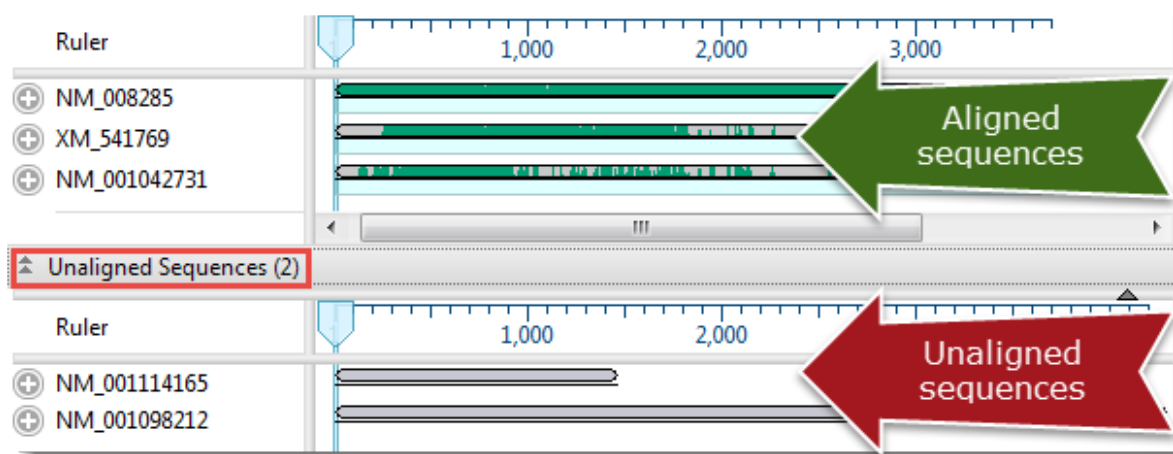
- [Try it! – Perform a Clustal Omega alignment](#)
- [Try it! – Perform a MUSCLE alignment with multi-segment sequences](#)
- [Try it! – Perform a genomic alignment with Mauve](#)

The following video is a quick introduction to performing multiple alignments in MegAlign Pro:

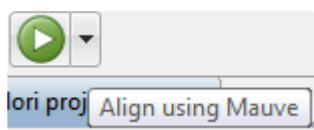
Perform an initial multiple alignment

To perform the initial multiple sequence alignment:

1. [Select](#) two or more sequences to align. 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.
 - 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 [Troubleshoot multiple 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

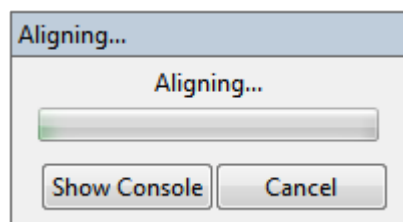
addition, a progress bar will open, but the alignment will not finish. To circumvent this issue, **Cancel** the alignment, drag the sequences into a different order in the Sequences View or Overview, and then try the alignment again.

- 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.

Troubleshoot multiple alignments

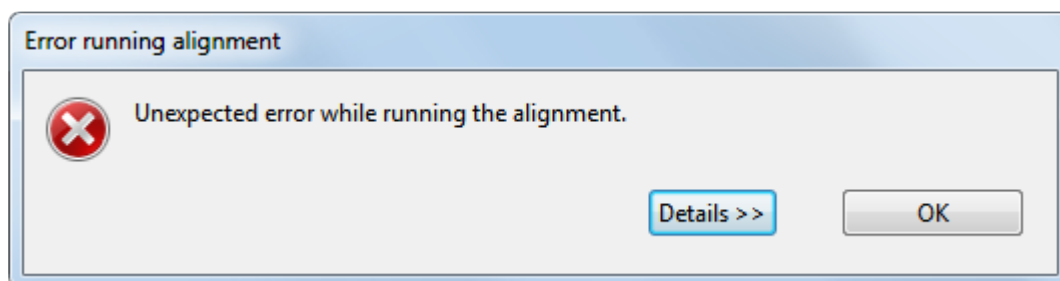
Error message that sequences are the wrong type or are too long:

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](#).

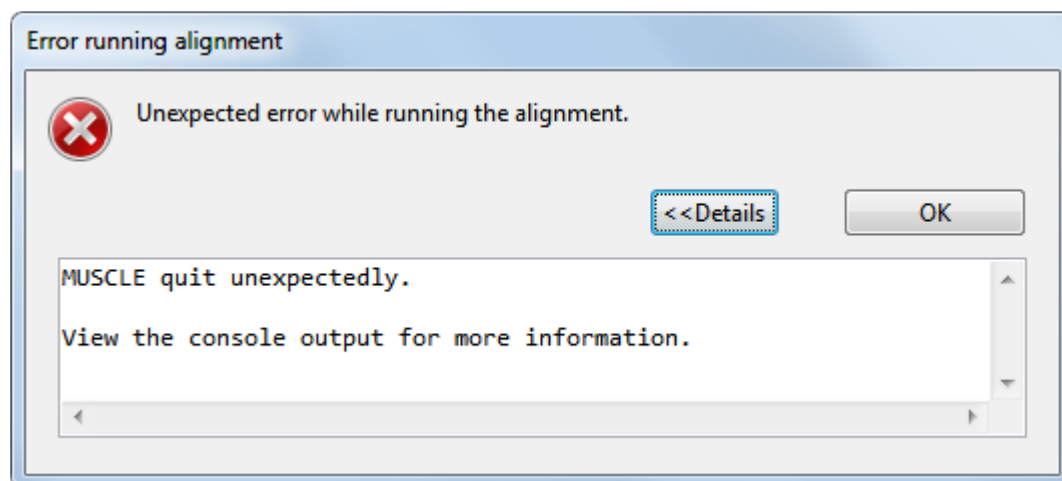


Message that an unexpected error occurred:

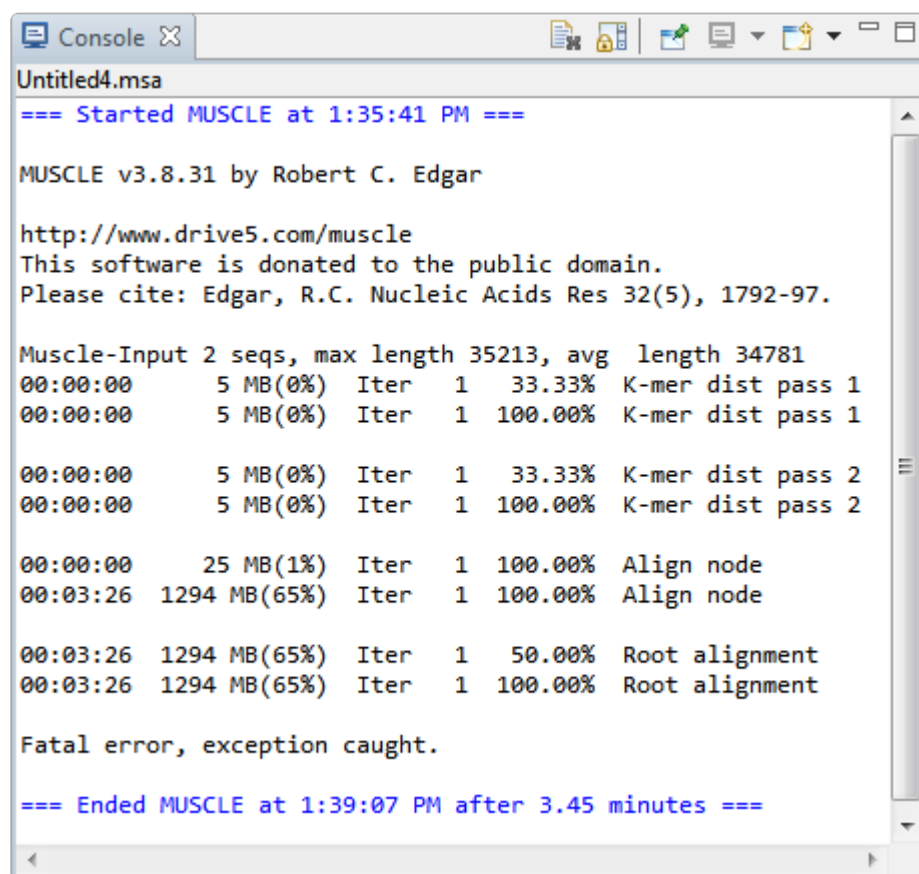
During a multiple sequence alignment, you may receive the following warning:




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.

Multiple alignment methods and options

MegAlign Pro offers both gene-level and genome-level multiple alignment algorithms.

Gene-level alignment of either protein or nucleotide sequences:

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; [Kato 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.
- **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.

Genome-level alignment of nucleotide sequences:

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](#).

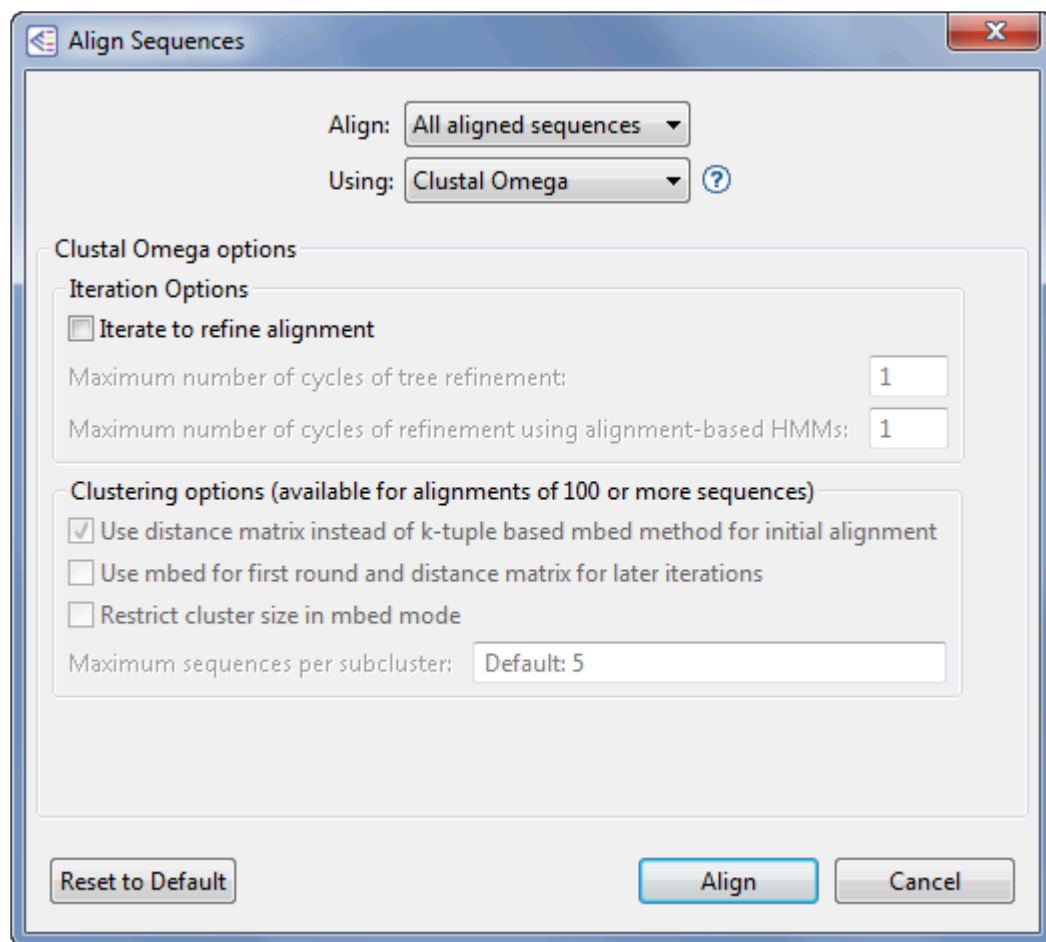
Mauve has high capacity and uses MUSCLE to create multiple alignments for each block that contains more than a single sequence. The Mauve algorithm is currently the only MegAlign Pro alignment method that is:

- 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).

Align Sequences

Align: **All aligned sequences**

Using: **ClustalW** ?

ClustalW options

Scoring matrix: **IUB**

Iteration method: **None** Maximum Iterations: **1**

Multiple alignment parameters

Gap open penalty: Default: 10

Gap extend penalty: Default: 0.20

Delay divergent sequences: Default: 30 %

Transition weight: Default: 0.50

Pairwise alignments for guide tree

☒ Slow, accurate ☐ Fast, approximate

Gap open penalty: Default: 10

Gap extend penalty: Default: 0.1

Reset to Default **Align** **Cancel**

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.

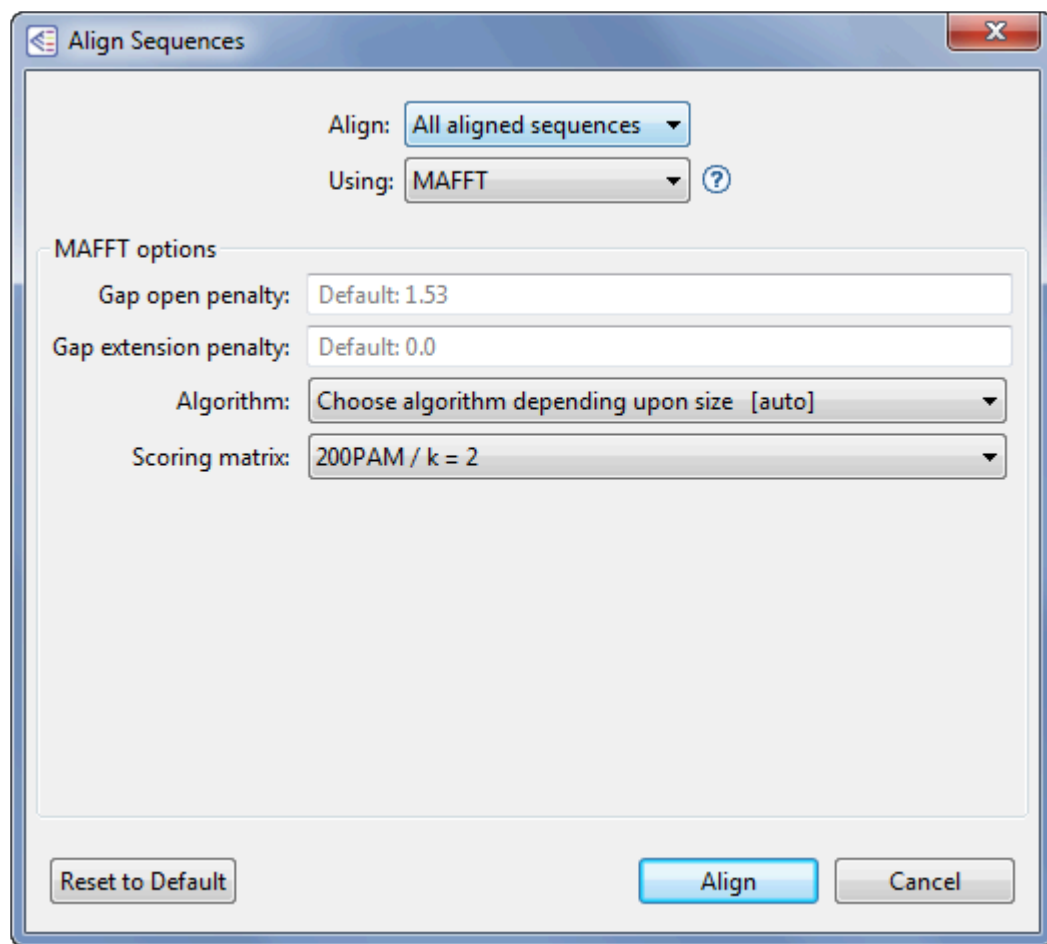
After making your changes:

- 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. 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 **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]	(default)		
Very slow	global homology	< 200	g-ins-i
Very slow	one conserved domain	< 200	l-ins-i
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).

The screenshot shows the 'Align Sequences' dialog box with the following settings:

- Align:** All aligned sequences
- Using:** MUSCLE
- MUSCLE options:**
 - Substitution matrix:** MUSCLE default
 - Gap open penalty:** Default: 2.9
 - Gap extension penalty:** Default: 0.0
 - Center penalty:** Default: 0.52
 - Maximum iterations:** Default: 8
 - ☐ Use Brenner's alignment
 - ☐ Disable anchors optimization
- Initial iterations:**
 - Cluster method:** UPGMB *
 - Distance measure:** Kmer 6-6 *
 - Sequence weight:** ClustalW *
- Later iterations:**
 - Cluster method:** UPGMB *
 - Distance measure:** Kimura % identity *
 - Sequence weight:** ClustalW *

* indicates default value

Buttons: Reset to Default, Align, Cancel

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(N ²) 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	weight1

Later iterations	weight2
------------------	---------

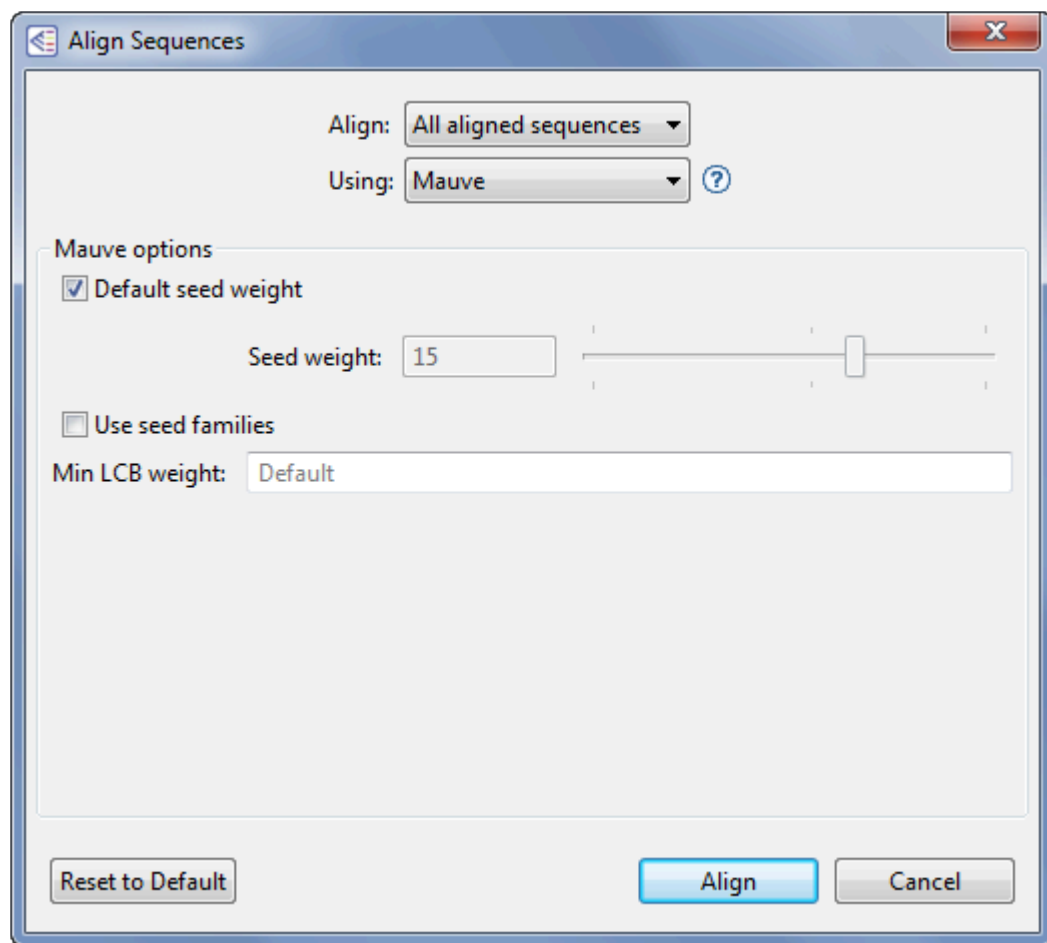
After making your changes:

- 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

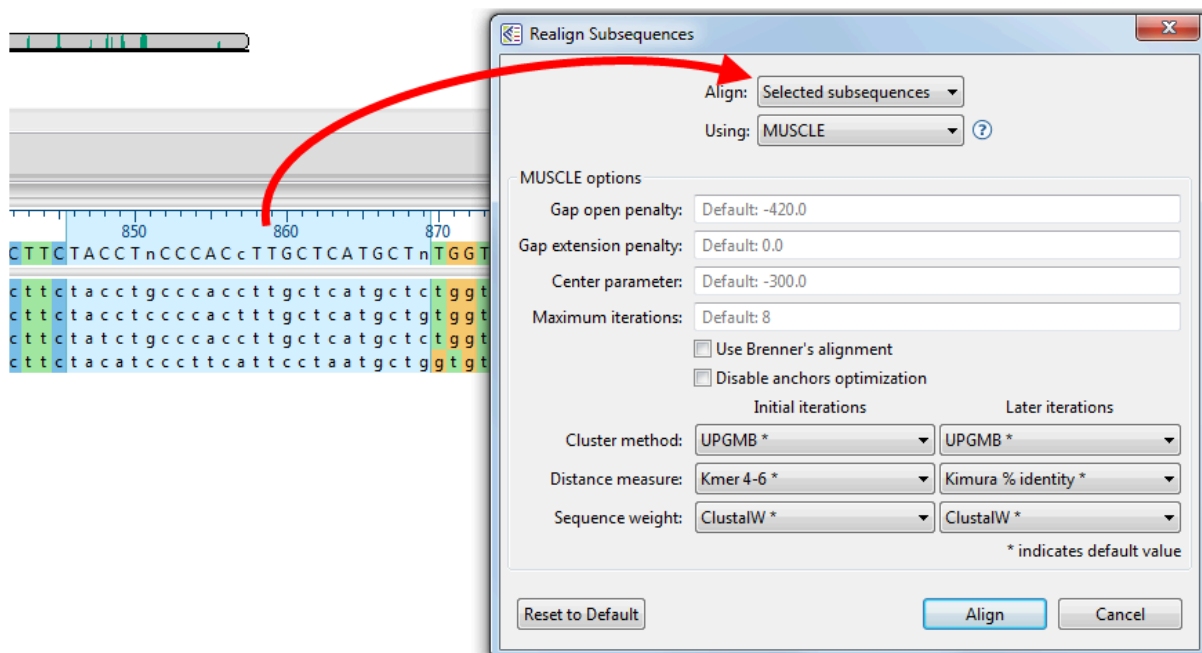
The following video illustrates some of these modification procedures:

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:

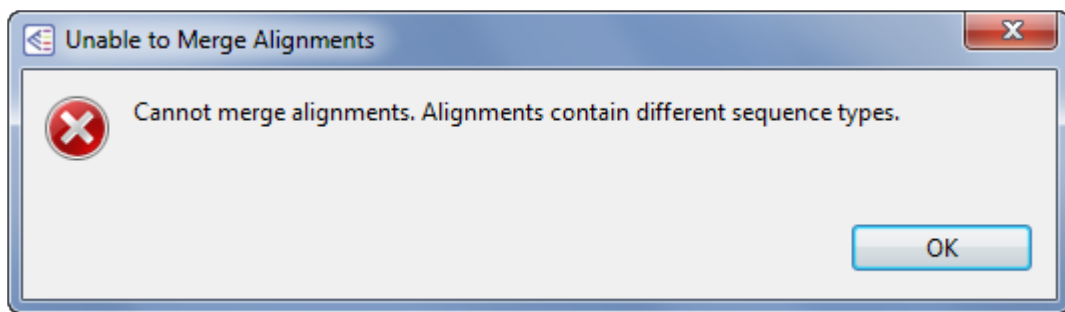
- Subalignments are not supported when the document contains any [end-to-end sequences](#).
- Occasionally, the new subalignment may no longer be the same as the optimal alignment your alignment method would generate from scratch for this particular set of sequence coordinates. You may want to realign 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 discontinuous 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

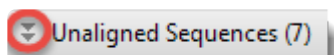
the “Unaligned Sequences” area of the current project.


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.



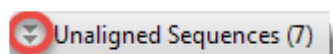
3. (optional) An upcoming step will overwrite the current alignment. Therefore, you may wish to save the original alignment before proceeding further.
4. [Specify which unaligned sequences](#) to add to the current alignment.
5. Select **Align > Merge Unaligned into Existing Alignment** or press the **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.

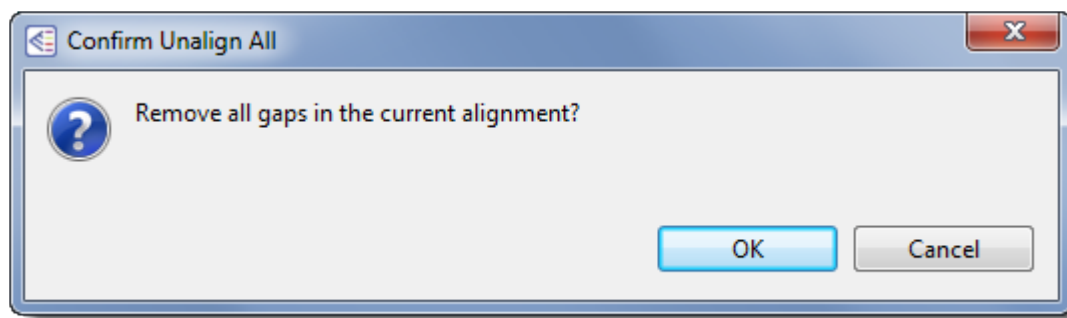


3. (optional) An upcoming step will over-write the current alignment. Therefore, you may wish to save the original alignment before proceeding further.
4. Specify which unaligned sequences to include in the new alignment by clicking on them or by using **Ctrl/Cmd+click** or **Shift+click**.
5. Select **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:

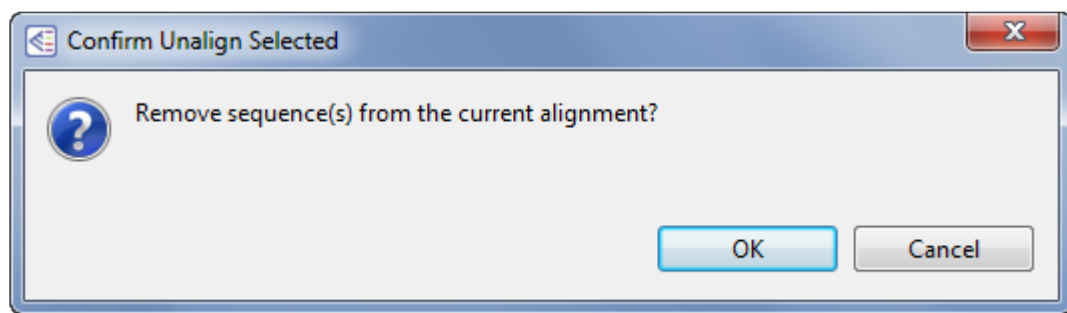
1. Choose **Align > Unalign All**.
2. When the following message appears, click **OK**.



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:

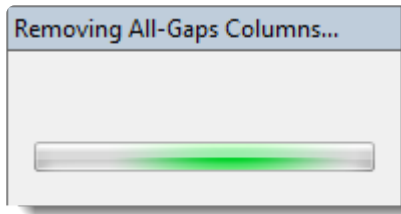
1. [Select one or more of the aligned sequences](#) in the Overview or Sequences view.
2. Choose **Align > Unalign Selected**.
3. When the following message appears, click **OK**.



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.



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.

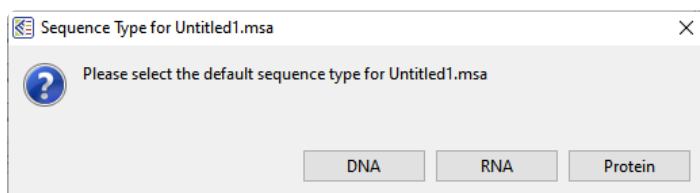
The L1 gene is not the cause of variation in pathogenicity. However, variations of the L1 gene are used to tell various strains apart and group them according to relatedness. The results of this tutorial reveal a deeply branched tree revealing an ancient phylogeny for HPV within the human genome.

Begin with [Part A: Add and align sequences](#).

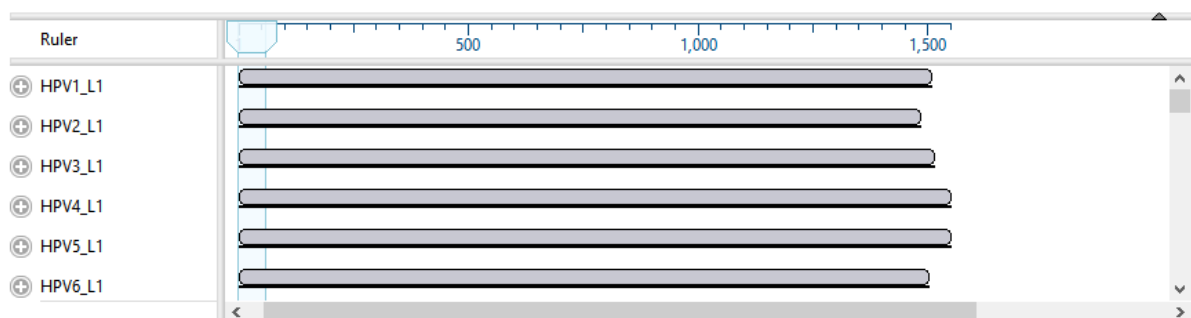
Part A: Add and align sequences


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 download 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.

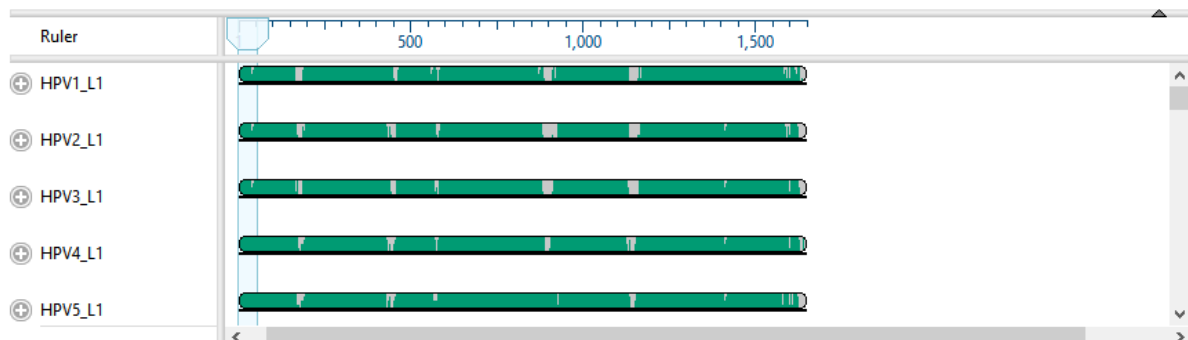


7. To align sequences with Clustal Omega using the default parameters, choose **Align > Align Using Clustal Omega** or click the **Align** tool arrow () and choose **Align Using Clustal Omega**. Sequences now appear mostly in green, with gray representing gaps.

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.

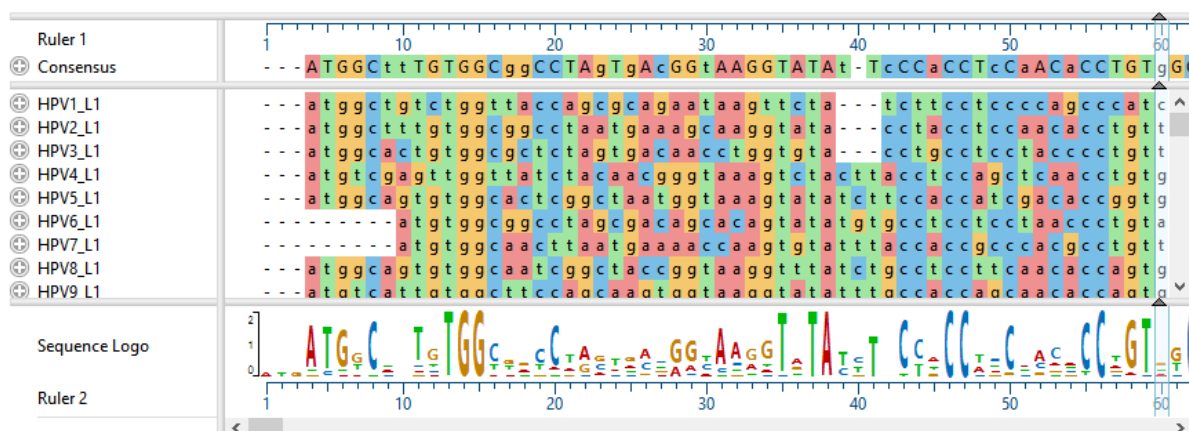
- Using the mouse, slide the viewport to the far right, and then back to the left. At the same time, notice how the Sequence view slides in the same way.
- To navigate to a particular residue in the Sequences view, double-click anywhere on a block in the Overview. The viewport will move so that it is 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

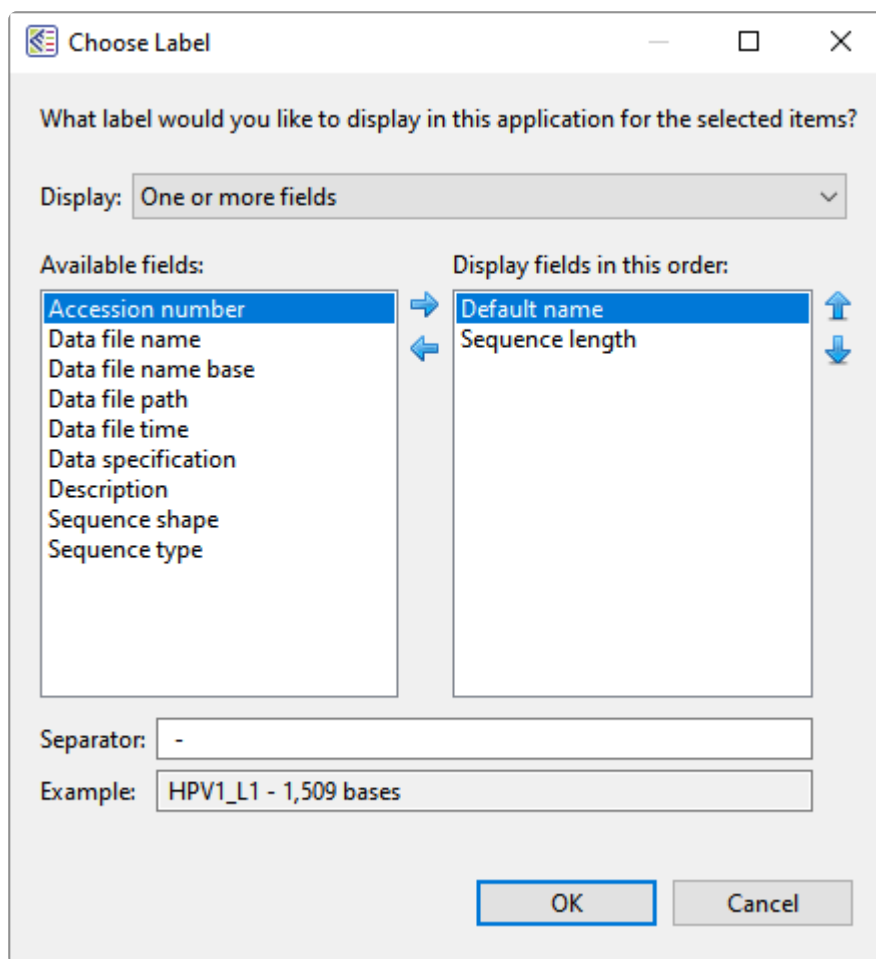
In [Part B](#), you looked at the [Sequences view](#) while learning the function of the "viewport." You will now explore the Sequences view in more depth. The Sequences view displays entered sequences and aligned sequences. Sequences selected in this view are simultaneously selected in all other MegAlign 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 with Fields**, or choose **Edit > Rename 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**.

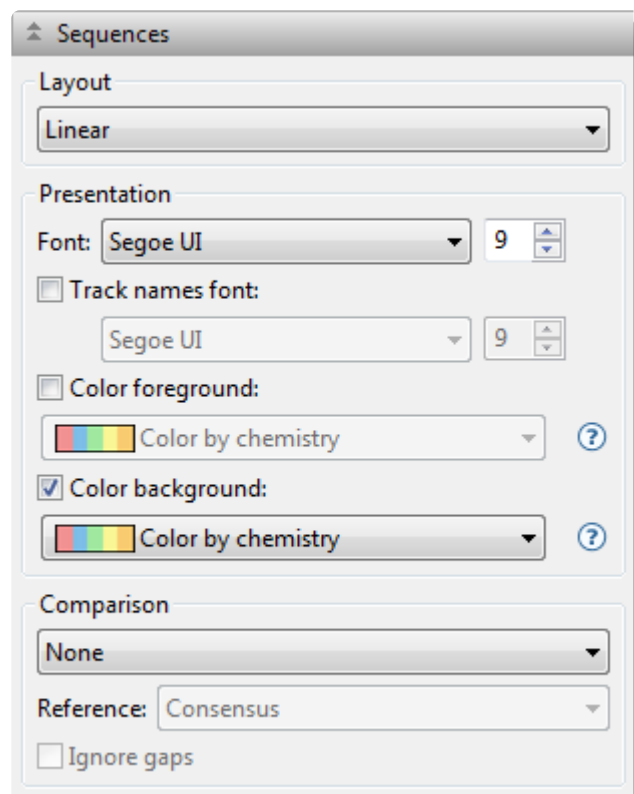
The sequence labels now reflect their lengths in base pairs:

⊕ HPV1_L1 - 1,509 bases	- - - a t g g c t g
⊕ HPV2_L1 - 1,485 bases	- - - a t g g c t t
⊕ HPV3_L1 - 1,515 bases	- - - a t g g c a c
⊕ HPV4_L1 - 1,551 bases	- - - a t g t c g a
⊕ HPV5_L1 - 1,551 bases	- - - a t g g c a g
⊕ HPV6_L1 - 1,503 bases	- - - - - - - a
⊕ HPV7_L1 - 1,518 bases	- - - - - - - a
⊕ HPV8_L1 - 1,545 bases	- - - a t g g c a g
⊕ HPV9_L1 - 1,524 bases	- - - a t a t c a t

- 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.
- Use **File > Save** to save the project as *HPV project.msa*.
- 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

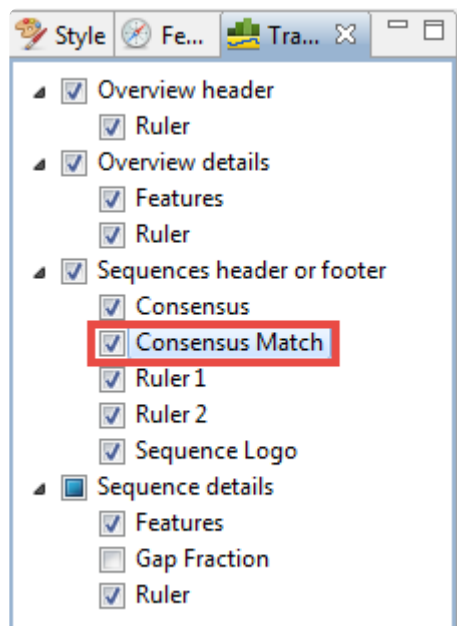
visible.

- The appearance of the sequences can be changed using settings in the Sequences area 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.

- 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 header 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.



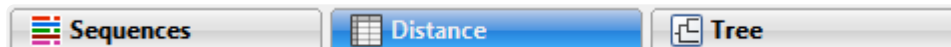
8. Hide the footer by clicking on the small gray arrow near its top right corner (). The Sequence Logo, Ruler and Consensus Match will disappear. Click on the arrow again to redisplay the footer. (Note: For future reference, you can close other panes the same way, including the footer and the main parts 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.

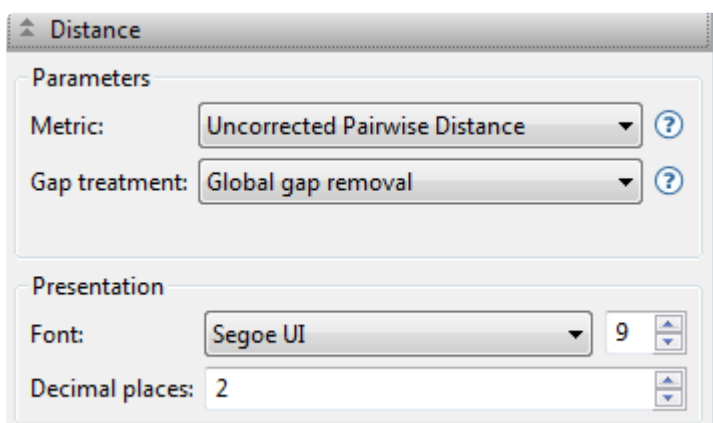
- To display the Distance view, click on the Distance tab at the bottom of the MegAlign Pro window.



The Distance view matrix is calculated, by default, using Uncorrected Pairwise Distance.

Distance		Metric: Uncorrected Pairwise Distance																Sequences: 143				Global gap removal				Residues considered: 1,240			
		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W					
A	HPV1_L1 - 1,509 bases		0.43	0.43	0.40	0.41	0.42	0.42	0.42	0.40	0.42	0.41	0.40	0.42	0.39	0.39	0.39	0.41	0.42	0.39	0.40	0.39	0.40	0.41					
B	HPV2_L1 - 1,485 bases	0.43		0.31	0.43	0.42	0.33	0.35	0.41	0.43	0.32	0.33	0.43	0.35	0.41	0.43	0.33	0.42	0.32	0.42	0.42	0.43	0.44	0.43					
C	HPV3_L1 - 1,515 bases	0.43	0.31		0.42	0.41	0.33	0.34	0.41	0.40	0.17	0.31	0.40	0.32	0.41	0.42	0.32	0.42	0.29	0.41	0.40	0.42	0.43	0.41					
D	HPV4_L1 - 1,551 bases	0.40	0.43	0.42		0.35	0.39	0.39	0.38	0.37	0.41	0.39	0.37	0.39	0.34	0.37	0.38	0.38	0.38	0.34	0.35	0.35	0.39	0.38					
E	HPV5_L1 - 1,551 bases	0.41	0.42	0.41	0.35		0.39	0.39	0.20	0.27	0.41	0.40	0.18	0.40	0.22	0.28	0.39	0.29	0.39	0.22	0.22	0.22	0.30	0.31					
F	HPV6_L1 - 1,503 bases	0.42	0.33	0.33	0.39	0.39		0.31	0.40	0.41	0.33	0.14	0.40	0.22	0.40	0.40	0.29	0.40	0.32	0.40	0.40	0.41	0.41	0.41					
G	HPV7_L1 - 1,518 bases	0.42	0.35	0.34	0.39	0.39	0.31		0.40	0.39	0.34	0.30	0.39	0.30	0.38	0.38	0.28	0.38	0.31	0.39	0.38	0.40	0.40	0.40					
H	HPV8_L1 - 1,545 bases	0.42	0.41	0.41	0.38	0.20	0.40	0.40		0.30	0.39	0.40	0.18	0.40	0.22	0.29	0.38	0.30	0.40	0.22	0.24	0.23	0.32	0.30					
I	HPV9_L1 - 1,524 bases	0.40	0.43	0.40	0.37	0.27	0.41	0.39	0.30		0.41	0.40	0.28	0.39	0.27	0.21	0.39	0.21	0.39	0.28	0.27	0.27	0.26	0.25					
J	HPV10_L1 - 1,512 bases	0.42	0.32	0.17	0.41	0.41	0.33	0.34	0.39	0.41		0.33	0.41	0.32	0.39	0.43	0.31	0.42	0.30	0.41	0.39	0.40	0.42	0.41					
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		0.39	0.22	0.39	0.40	0.28	0.40	0.31	0.40	0.38	0.41	0.40	0.40					
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		0.39	0.22	0.29	0.38	0.30	0.38	0.21	0.22	0.23	0.32	0.31					
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		0.39	0.39	0.28	0.38	0.30	0.40	0.39	0.39	0.39	0.39					
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		0.28	0.36	0.27	0.37	0.16	0.14	0.13	0.29	0.28					
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		0.37	0.19	0.39	0.28	0.28	0.29	0.25	0.25					
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	0.27		0.39	0.30	0.27	0.27	0.39	0.39	0.39					

- Click on the Style tab and experiment with the settings in the Distance parameters section. Observe how these changes affect the look of the Distance view.

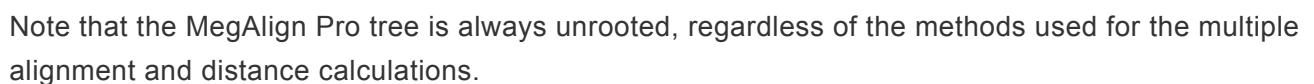
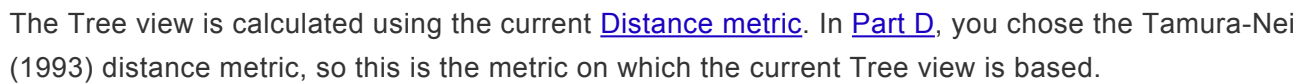


- Change the distance **Metric** to **Tamura-Nei (1993)**. (Note: Different sub-sets of metrics are available for protein sequences and nucleotide sequences).
 - Change the **Font** style and/or size using the middle drop-down menu and text box.

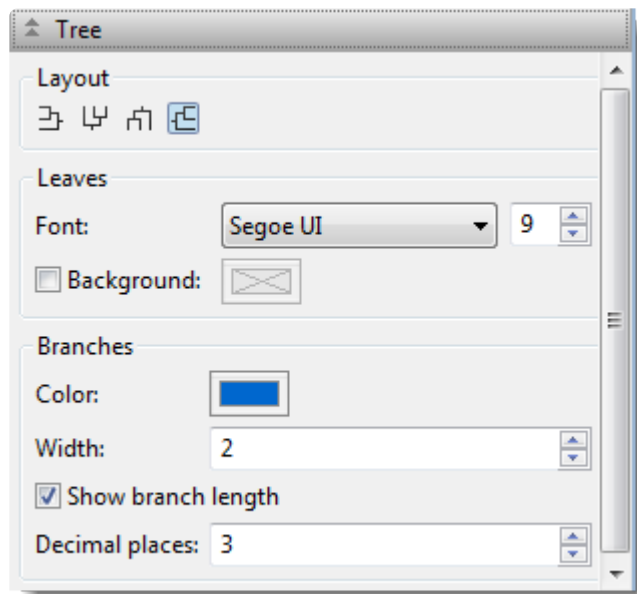
- Change the number of **Decimal places** displayed in the Distance View.
3. Try selecting, copying, and exporting the distance table.
- - *To select* a portion of the table, use the mouse. To select the entire table, use **Edit > Select All** or **Ctrl/Cmd+A**.
 - *To paste the selected information into a spreadsheet application* such as Microsoft Excel[®], use **Edit > Copy** or **Ctrl/Cmd+C**.
 - To export the distance matrix in your choice of available formats_, use **File > Export Data > Distance Matrix**.


Proceed to [Part E: Change the appearance of the Tree view](#).

1. To display the Tree view, click on the **Tree** tab at the bottom of the MegAlign Pro window.



- Page 222 of 294




- - Change the **Layout** by clicking any of the four position icons at the top of the Tree section.
 - Change the font style and/or size used for the **Leaves** using the middle drop-down menu and text box.
 - Add a background color for the text, if desired, by checking the **Background** checkbox, then clicking on the white box on the right to make a selection from the color picker.
 - Change the color of the branches by clicking on the Branches **Color** box and making a selection from the color picker.
 - Show or hide the **branch length** by checking and unchecking the corresponding box. While the box is checked, try adjusting the number of **Decimal places**.
- 3. In the [Distance section](#) on the right, use the distance **Metric** dropdown menu to return to **Uncorrected Pairwise Distance**. Observe how the Tree view changes. For greatest accuracy, we recommend using an evolutionarily adjusted metric such as Tamura-Nei (1993) rather than “Uncorrected Pairwise Distance.”
- 4. 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.

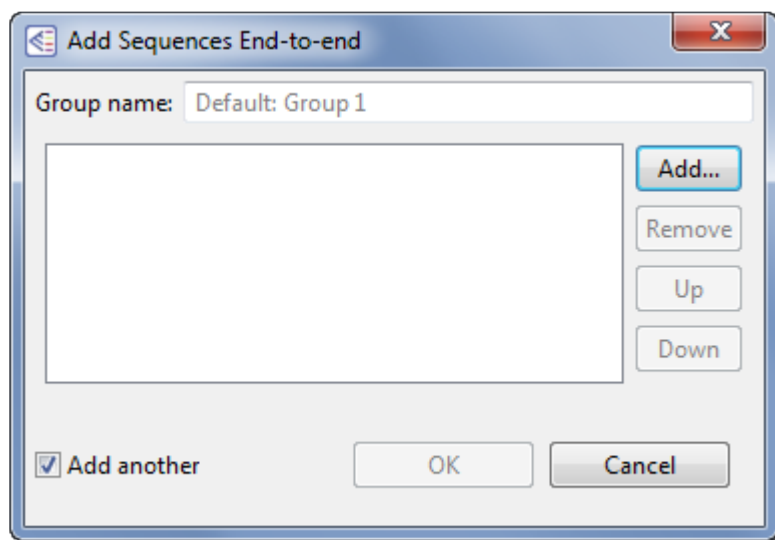
Suppose you want to string together the consensus sequences from all the contigs in an assembly, or chromosomes in a genome, or exons in a gene, etc. If you enter these sequences using **File > Enter**

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.

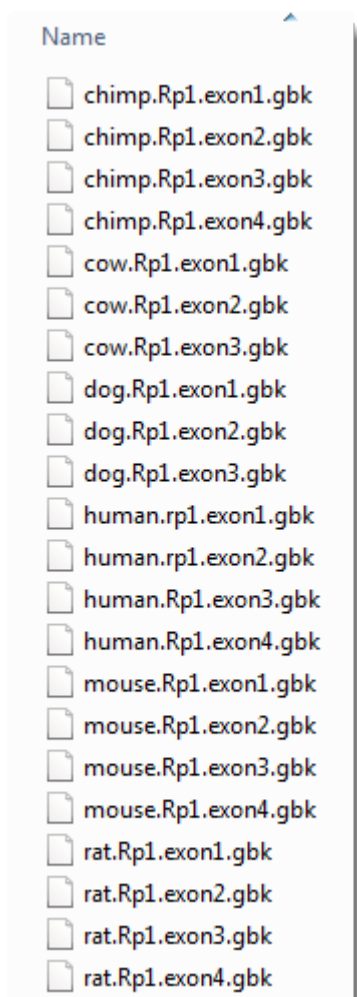
Begin with [Part A: Add and align multi-segment files](#).

Part A: Add and align multi-segment files

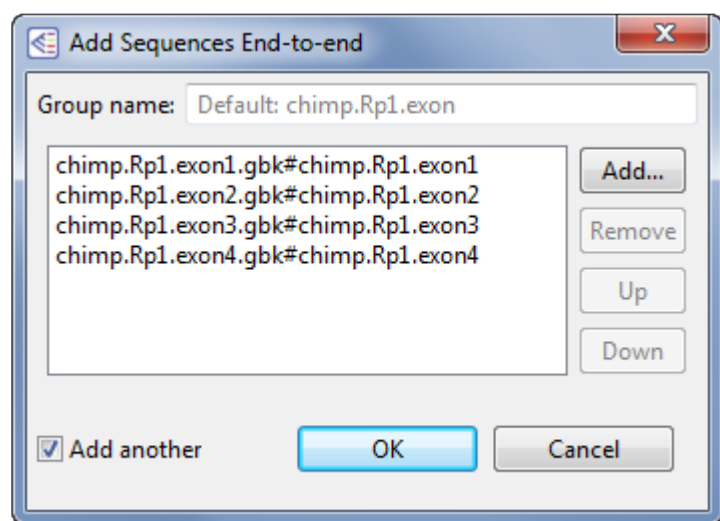
1. If you have not yet download and extracted the tutorial data, [click here to download it](#). Then decompress (unzip) the file archive using the method of your choice.
2. If a MegAlign Pro project is already open, choose **File > New Alignment Project**. Otherwise, launch MegAlign Pro.
3. Choose **File > Add Sequences End-to-end**, or click on the **Add sequences end-to-end in project** tool ().



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.

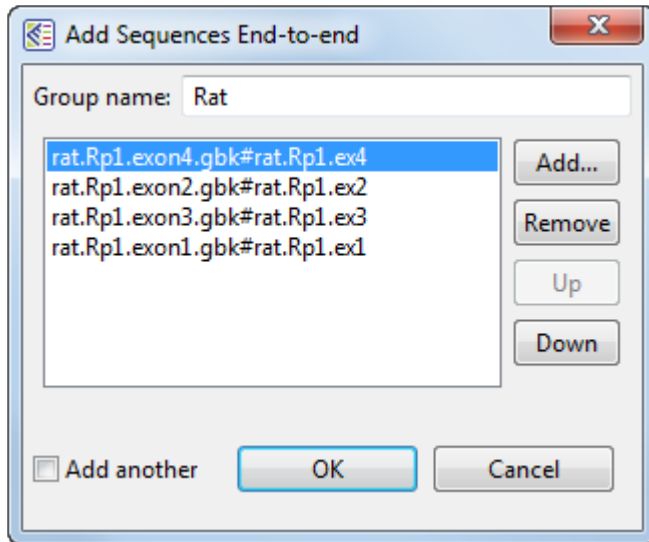


- Use **Shift+click** to select only the four "chimp" segments, and then click **Open**. In the Add Sequences End-to-end dialog, the files will appear in ascending alphabetical order.



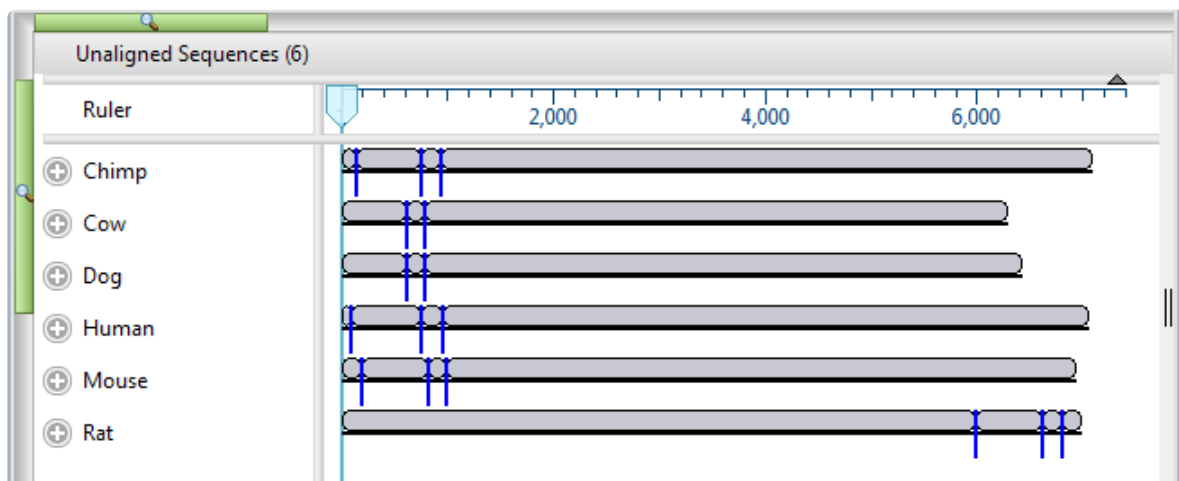
- In the **Group name** text box, replace the default name with **Chimp**, and then click **OK**.

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 [Part B: Replace a sequence in the Overview and view the phylogenetic tree](#).

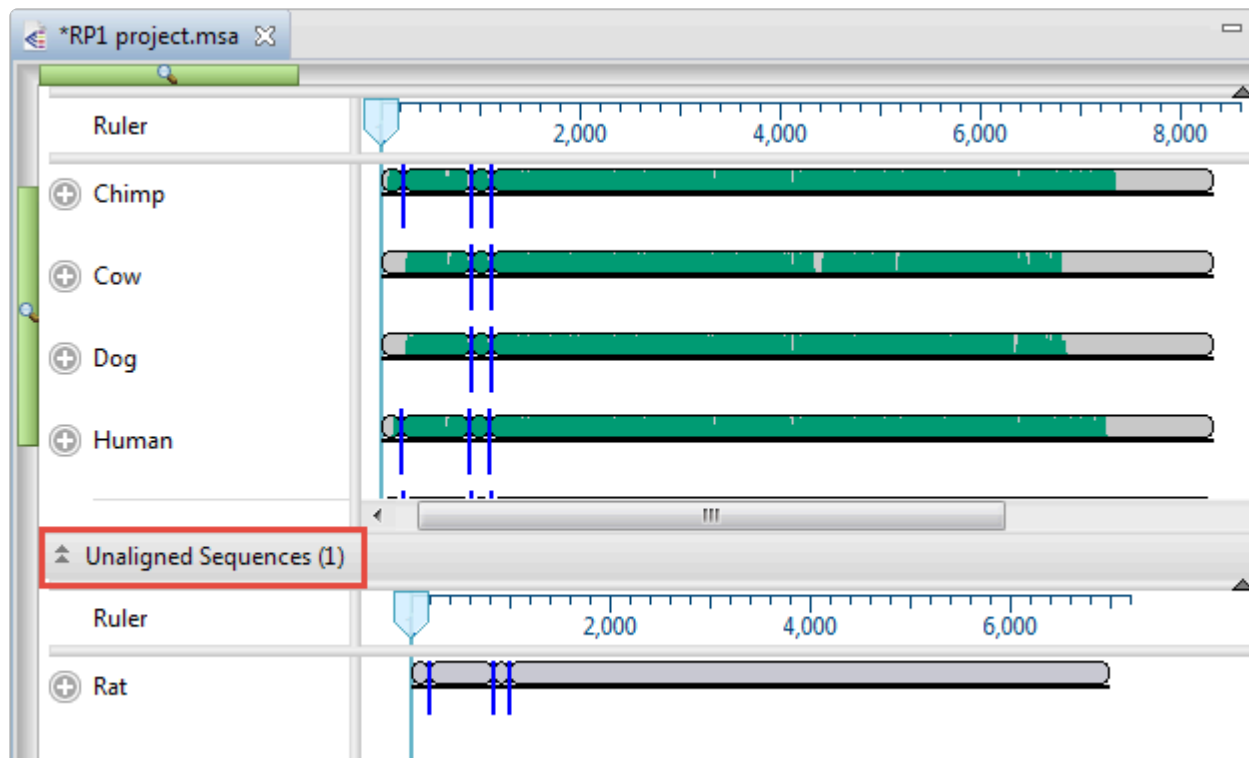
Part B: Replace a sequence in the Overview and view the phylogenetic tree

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. 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**.

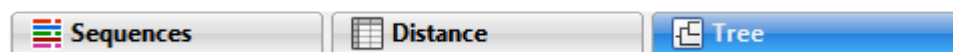
Observe that the Rat sequence has been added to a new area of the view called "Unaligned Sequences." Click on the expand arrows to reveal the sequence, shown in the image below.



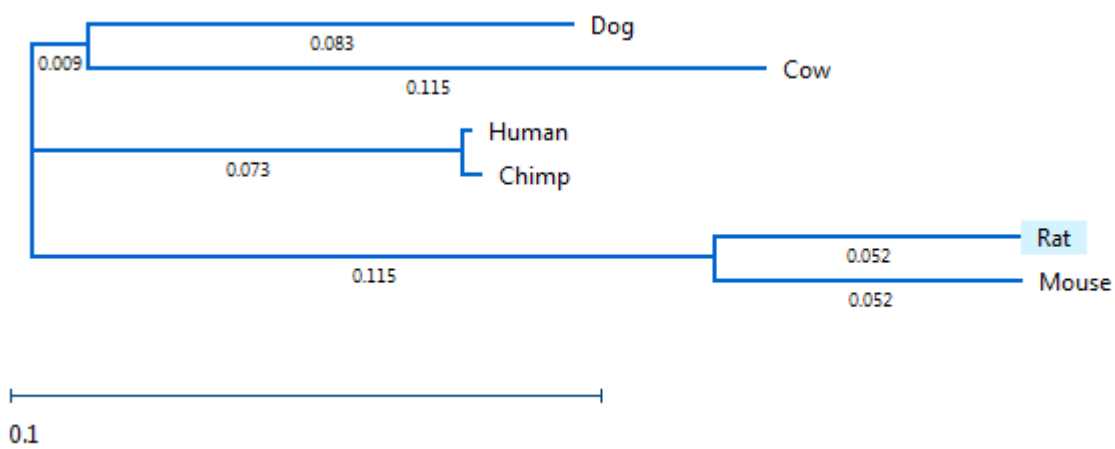
6. To perform the MUSCLE alignment with the "Rat" group in the correct order, select the name "Rat" on the left of the Overview and choose **Align > Realign with Selected Unaligned**. This command tells MegAlign Pro to perform a novel alignment that includes all the sequences in the existing alignment 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.

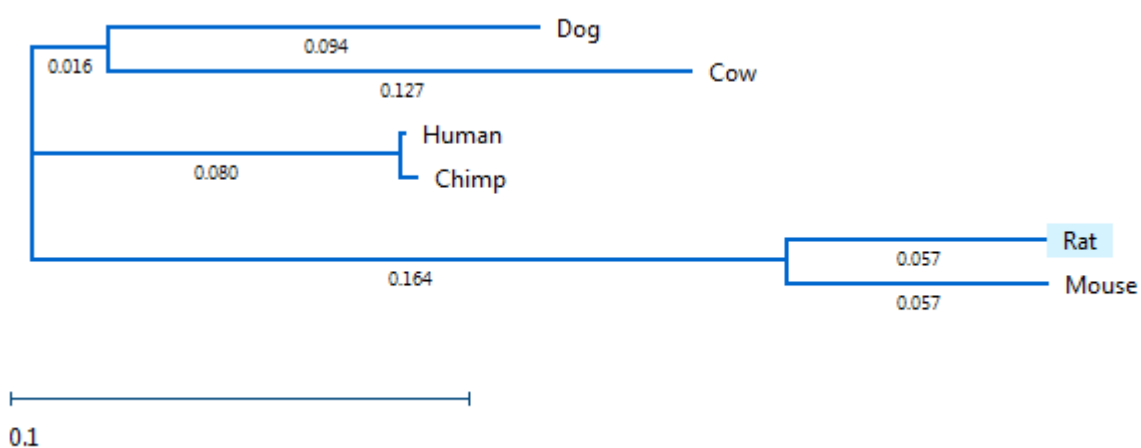
8. Display the [Tree view](#) by clicking on the **Tree** tab at the bottom of the MegAlign Pro window.




The phylogenetic tree graphically illustrates plausible evolutionary relationships for the seven species.



9. Using the [Distance section](#) on the right, change the **Metric** from **Uncorrected Pairwise Distance** to **Tamura-Nei (1993)**. Unlike in Try it! – Perform a Clustal Omega alignment, 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.

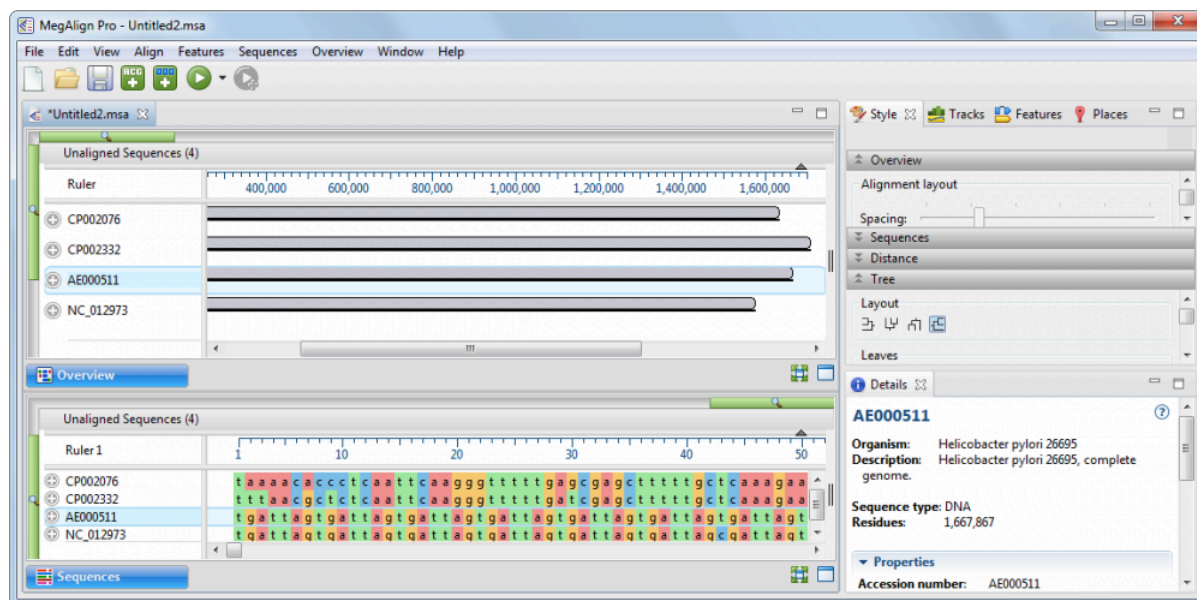
Begin with [Part A: Add genomic sequences using drag & drop](#).

Part A: Add genomic sequences using drag & drop

1. If you have not yet download 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*.

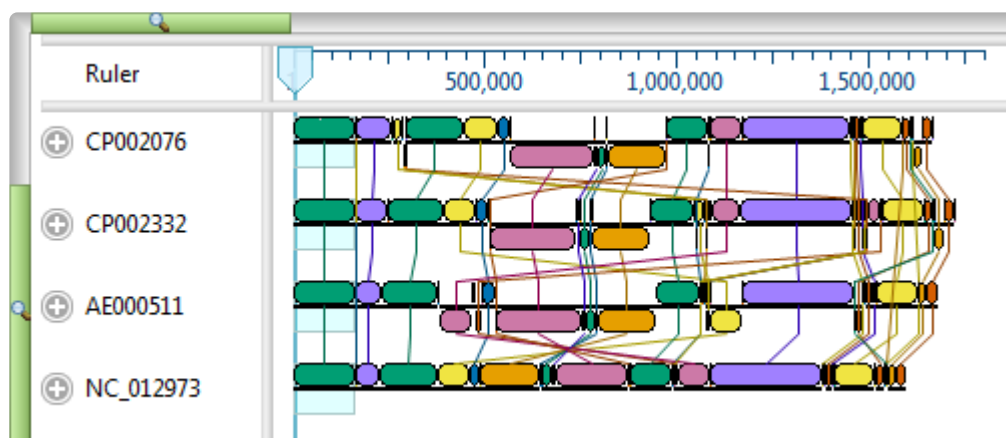
Proceed to [Part B: Perform a Mauve alignment using modified parameters.](#)

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.

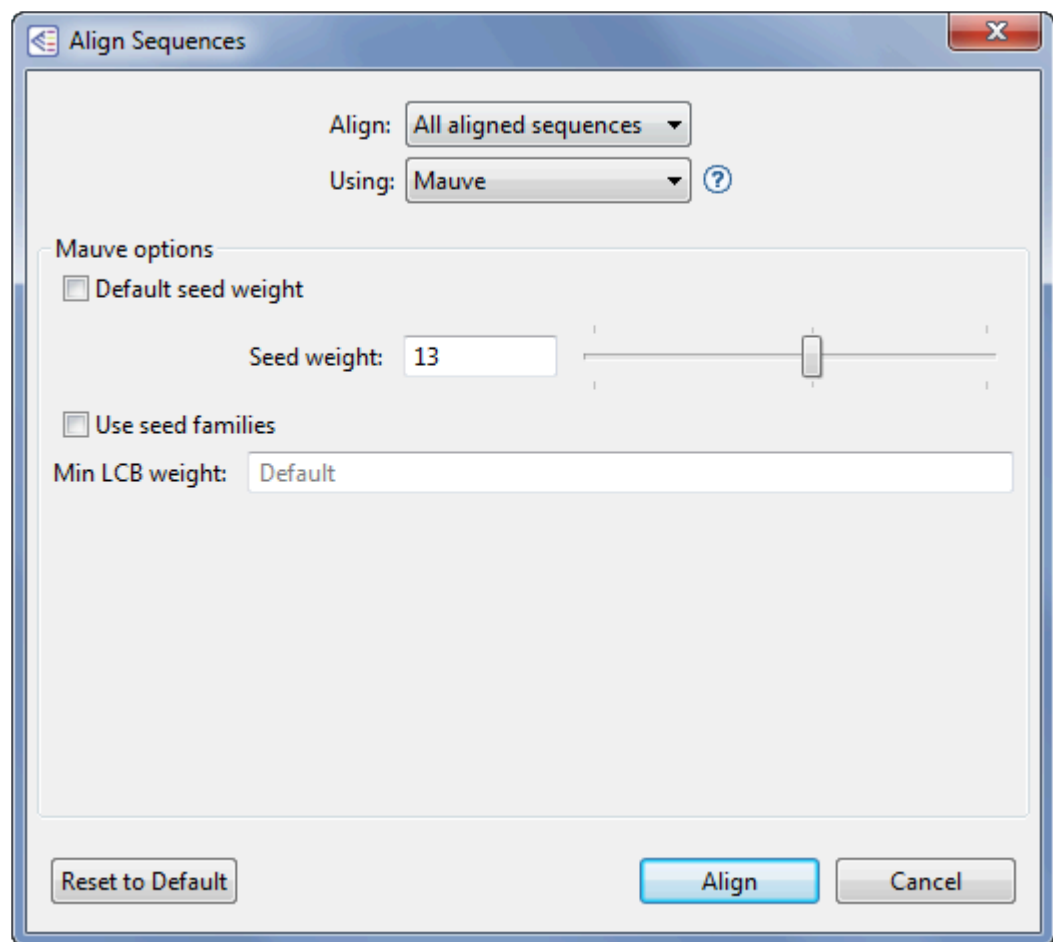


The default settings used in creating this alignment are suitable for most projects, including this one. However, you will make a change to Mauve's **Seed weight** parameter in order to see how this affects the alignment results.

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.

- 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.



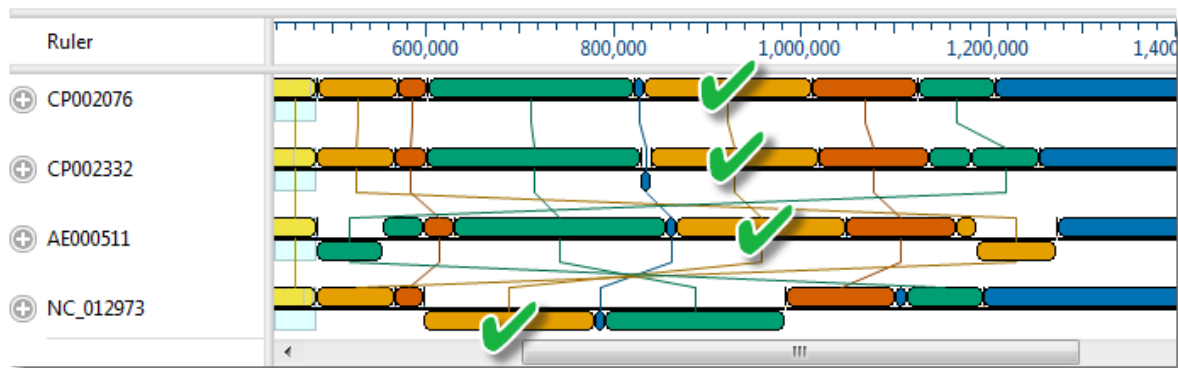
- Choose **Align** to perform the Mauve alignment again using the selected parameters.

Proceed to [Part C: View a genomic alignment in the Overview](#).

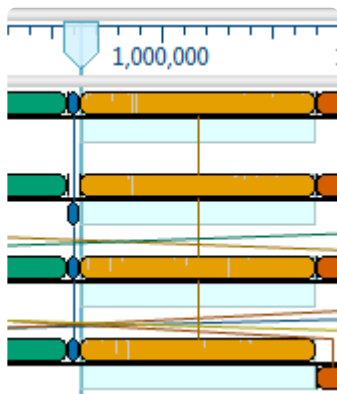
Part C: View a genomic alignment in the Overview

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.

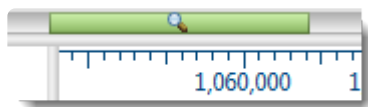


Observe that the active blocks are now in a single, highlighted column, with each block having equal width. The viewport is located on the left-most end of the blocks.



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.

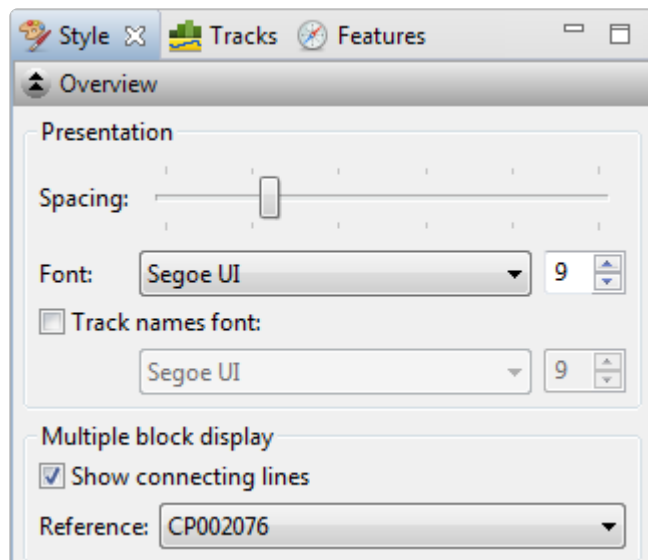
2. In the Overview, slide the viewport to the right edge of the active blocks. You can use the horizontal green slider above the view to expand this area so you can see it more clearly.



- 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.




- 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 **Reference**, change it back to CP002076 before you are finished.

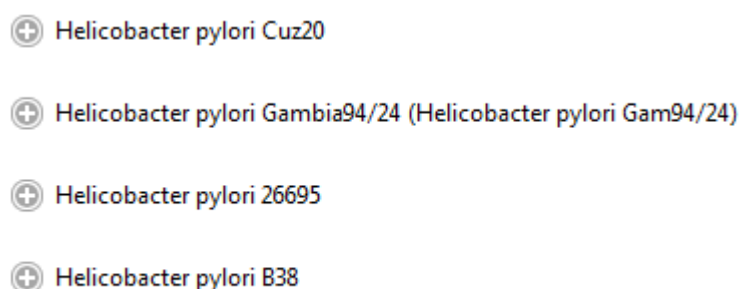
Proceed to [Part D: View the phylogenetic tree.](#)

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.

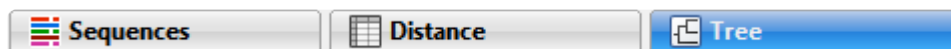
1. The current names of each sequence originated from the accession numbers that were embedded in the input files. Before viewing the phylogenetic tree, you can change these names to show the organism from which each sequence originated:
 - a. Use **Ctrl+A** or **Cmd+A** to select all of the sequence names.
 - b. Right-click on the highlighted group and select **Rename with Fields**, or choose **Edit > Rename 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:

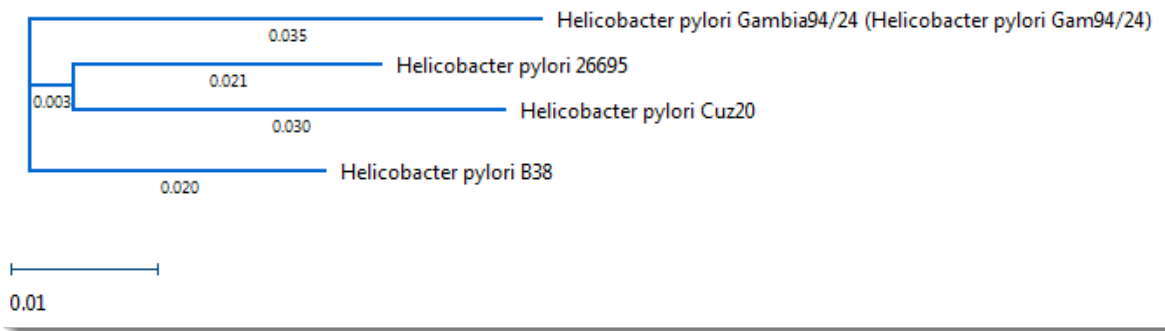


+ Helicobacter pylori Cuz20
 + Helicobacter pylori Gambia94/24 (Helicobacter pylori Gam94/24)
 + Helicobacter pylori 26695
 + Helicobacter pylori B38

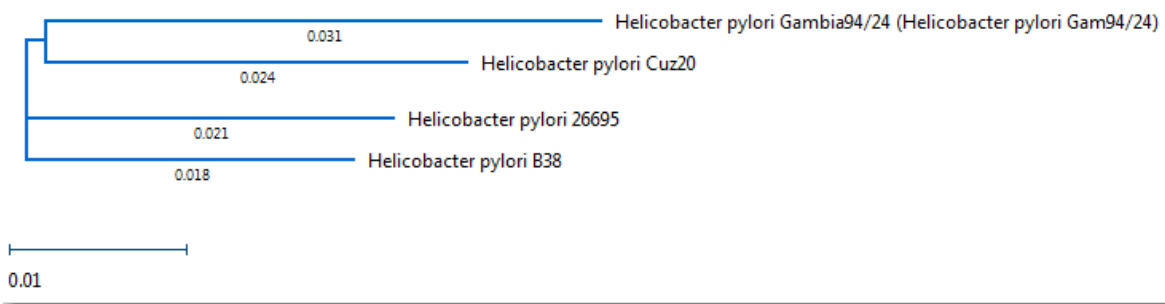
2. To display the Tree View for the active blocks, click on the Tree tab at the bottom of the MegAlign Pro window.



The contents of the [Tree view](#) (and also the [Distance view](#)) are specific to the currently selected block only.



- In the Overview, double-click on different blocks in the Overview and see how the branch lengths, or (sometimes) the tree itself, change.




- To order sequences in the other views so they match the tree, select **Tree > Order sequences like tree**.

- If desire, save the project using **File > Save**, or by clicking on the **Save** tool ().

This is the end of the Mauve alignment tutorial.

Perform a Pairwise Alignment

Did you arrive here by selecting the  DNASTAR Navigator workflow **Molecular Biology > Pairwise sequence alignment**? If so, you're in the right place!

This application supports [local, global and semi-global pairwise alignment methods](#). Pairwise alignments 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.

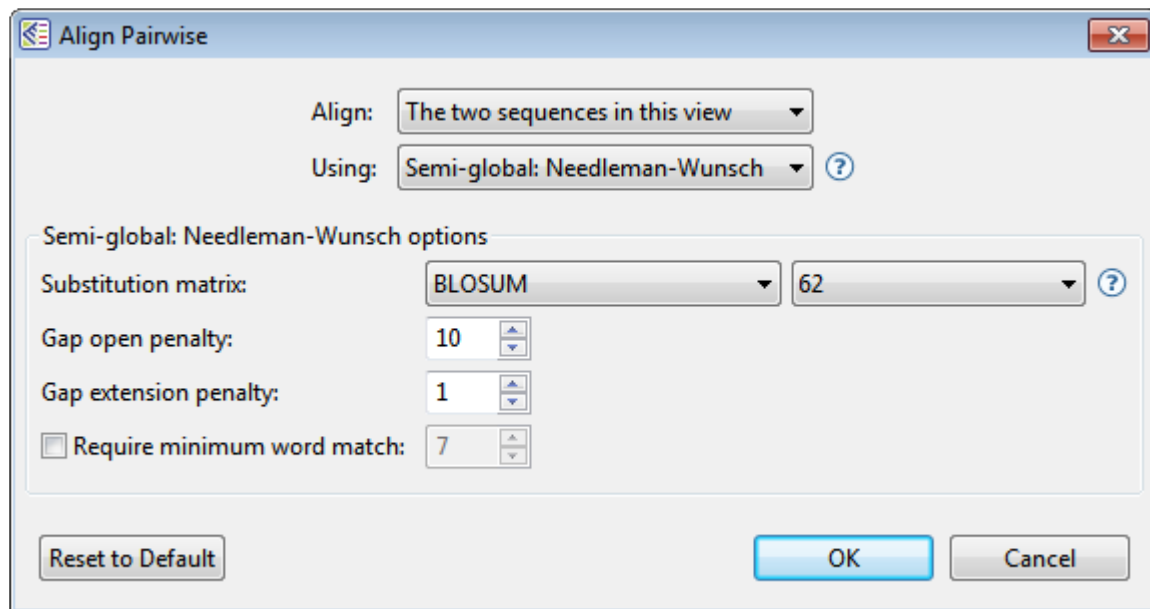
Pairwise alignment tutorials:

The following tutorials use free 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](#)

To perform a pairwise sequence alignment:

1. Decide which two sequences you want to align. The sequences can be any length (DNASTAR has successfully aligned protein sequences up to 35,000 bases 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.
2. [Select](#) the two sequences and choose **Align > Pairwise** or right-click on the selection and choose **Align Pairwise**. Dialog options vary depending on whether protein or nucleotide sequences were selected. An example is shown below.



3. Use the **Align** drop-down menu to specify which sequences to align.
4. In the **Using** drop-down menu, choose the desired pairwise alignment method:
 - Global alignment, represented in MegAlign Pro by the **Global Needleman-Wunsch** algorithm, 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 alignment, represented by the **Semi-global Needleman-Wunsch** algorithm, 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.
 - Local alignment, represented by the **Local Smith-Waterman** algorithm, reports the highest scoring contiguous 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 [Style panel](#).

See [this page](#) from Trinity College-Dublin's Department of Genetics for a good synopsis of Needleman-Wunsch vs. Smith-Waterman algorithms.

Note: When two sequences are nearly identical (check by [performing a multiple alignment](#) and consulting the [Distance view](#)), all pairwise methods should work equally well.

5. Use the drop-down menu(s) to choose the [Substitution matrix](#) and related options, if any
6. Specify the **Gap open penalty**, the amount deducted from the alignment score for each gap in the alignment. Gaps of different sizes carry the same penalty. Default is 10.
7. Specify the **Gap extension penalty**, the value which will 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.
8. If you wish to specify the length of the smallest perfect match of contiguous bases/residues to consider in building an alignment, check the **Require minimum word match** box and enter a value. The default is unchecked; if checked, the default value is 7.
9. Press **OK**.

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).

10. (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.

Comparison of pairwise alignment methods

The three types of alignment—Local, Global and Semi-Global—are actually quite similar, 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:

MegAlign Pro's local alignment algorithm, 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).

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.

Semi-Global Pairwise Alignment:

Semi-global alignment is 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").

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. 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 way to go. We have included three tutorials to demonstrate some of the differences between these methods.

- [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](#)

Substitution matrices

A substitution matrix describes the rate at which a nucleotide or amino acid changes to another nucleotide or amino acid over time. When [performing a pairwise alignment](#), you can specify the desired substitution matrix in the (Pairwise) Alignment Options dialog.

Available matrices for nucleotide sequences:

Matrix	Description
NUC44	<p>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.</p> <p>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.</p>

Available matrices for protein sequences:

Matrix	Description	Secondary option
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.
GONNET	Derived from PAM matrices (Dayhoff et al., 1978) but more sensitive, and based on a much larger data set.	(Unchangeable default of 250)
IDENTITY	Scores two identical amino acids as 1, and anything else as -10,000.	N/A
MATCH	Scores two identical amino acids as 1, and anything else as -1.	N/A
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.
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 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.

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 than starting over with several different documents.

To perform the analysis:

1. If you have not yet download 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_vertibrates.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 an 11-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

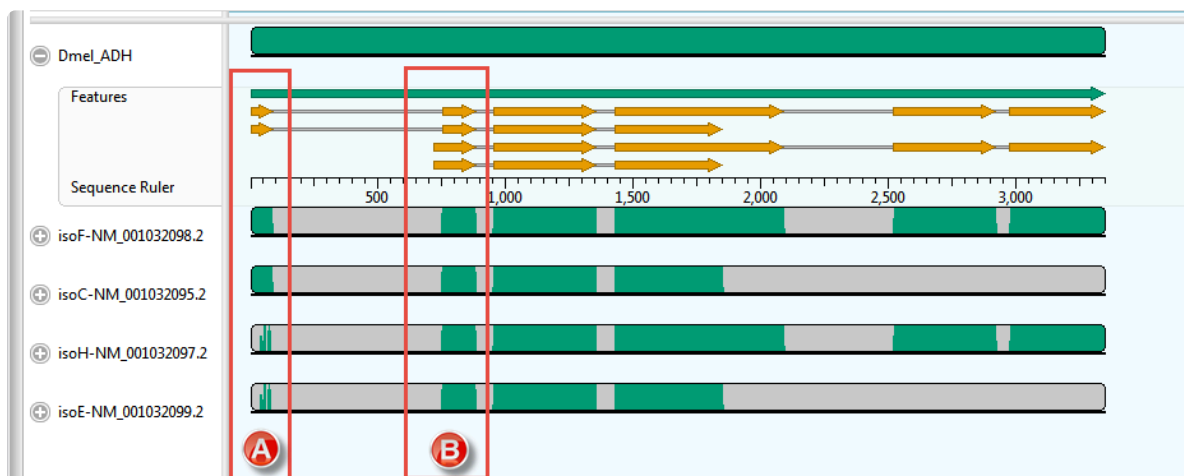
This part of [the tutorial](#) demonstrates that if you don't have annotations to rely on for validating multiple alignments, a pairwise alignment is a good alternative.

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 download 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.

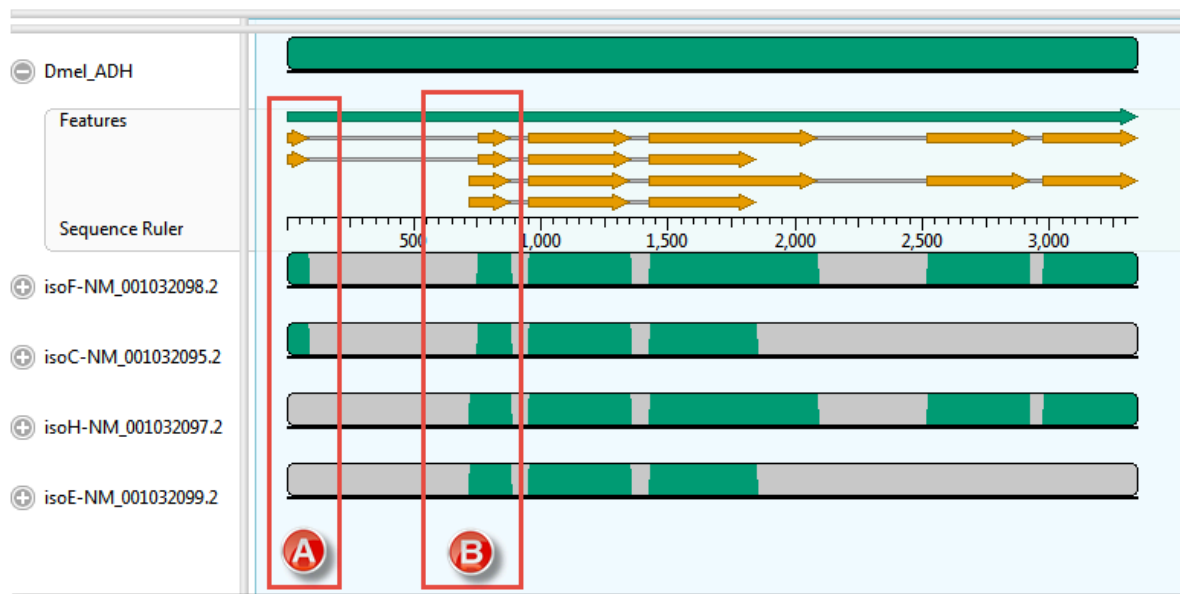
3. Hover your mouse over some of the orange mRNA features and verify that they are in the same order as the transcripts represented by the lower four sequences.

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.)



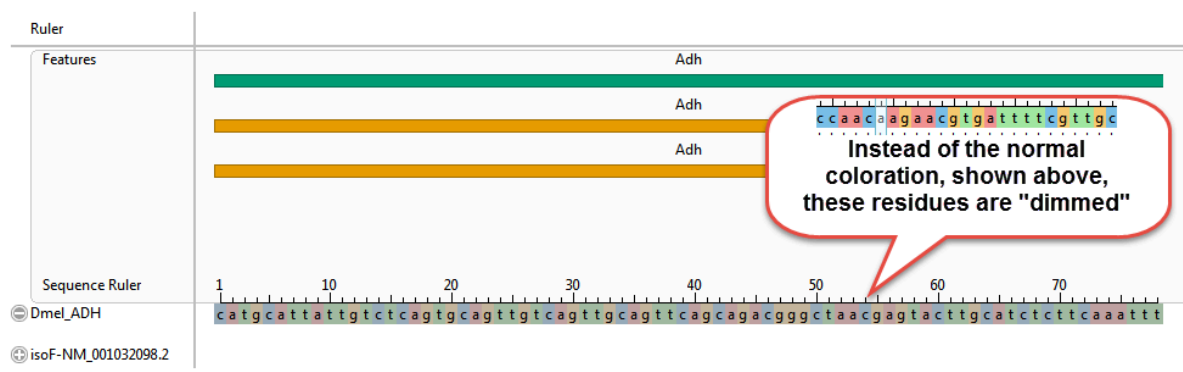
Proceed to [Part B: Use a Local pairwise alignment method](#).

Part B: Use a Local pairwise alignment method

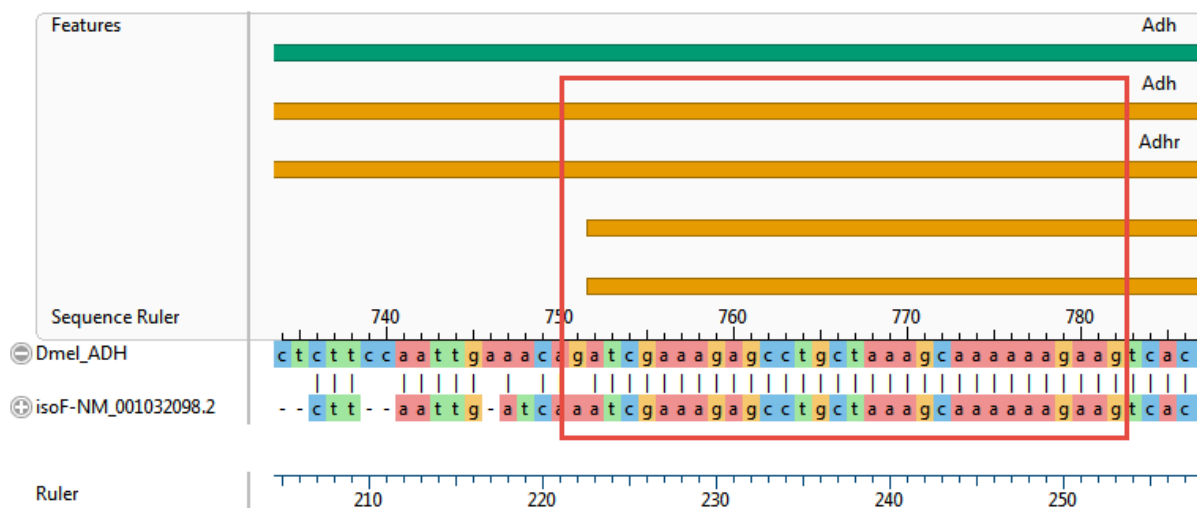
Now that you have finished [Part A](#) of the tutorial, this part shows how you can use a local pairwise alignment to resolve the correct mapping of the first intron in isoF and isoC.

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". 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](#).

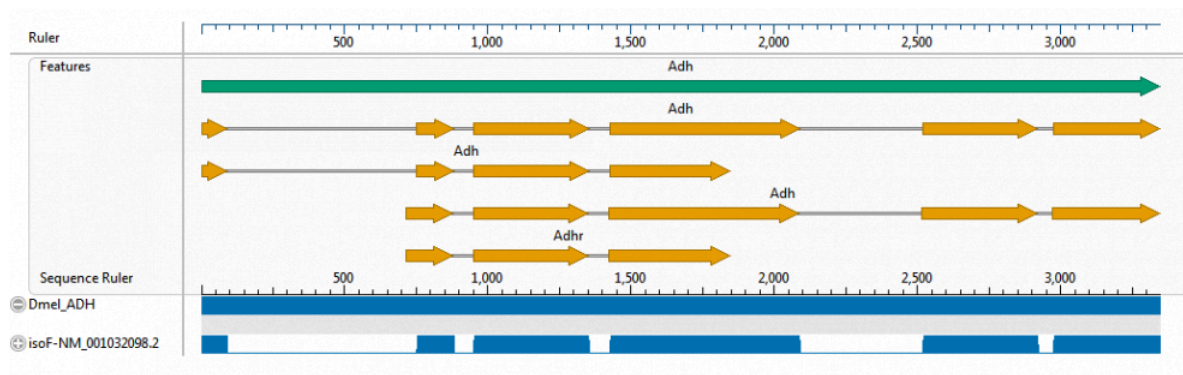
Most likely, introducing a sufficient number of gaps to span the intron would reduce the score so much that the aligner can find a higher scoring segment that contains numerous short gapped regions.

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 download and extracted the tutorial data, [click here to download it](#). Then decompress (unzip) the file archive using the method of your choice.
1. 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.



2. Select both sequences, right-click on the selection, and then select **Align Pairwise** from the context menu.
3. 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.

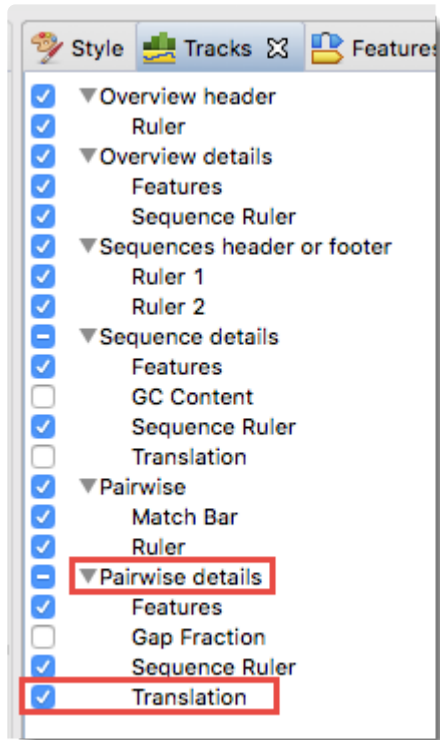
Local: Smith-Waterman DNA alignment [Matrix: "NUC44" Gap penalty: 10 Gap extension penalty: 1]									
	NC_003198.1 (CT18)	unknown				Gap	Gap		
	1>4809037	1>706	%Identity	%Gaps	Identical	Count	Length	Score	Length
Alignment	279279>279980	1>706	98.6%	0.6%	696	1	4	3,443	706

4. Scroll down through the alignment until you locate the gap, which begins at alignment position 452.



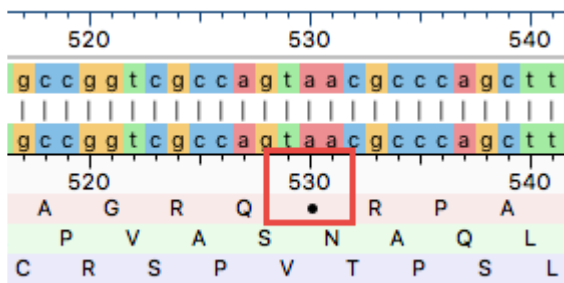
Notice that the gap is shown within the reference, indicating that there is a 4 bp insertion in the unknown sequence. Since the insertion length is not a multiple of three, the gap most likely represents a frame-shift in the unknown sequence.

5. In the **Tracks** side panel on the right, locate the **Pairwise Details** section and check the box next to **Translation**.



6. Returning your attention to the Pairwise view on the left, open the [tracks](#) 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.

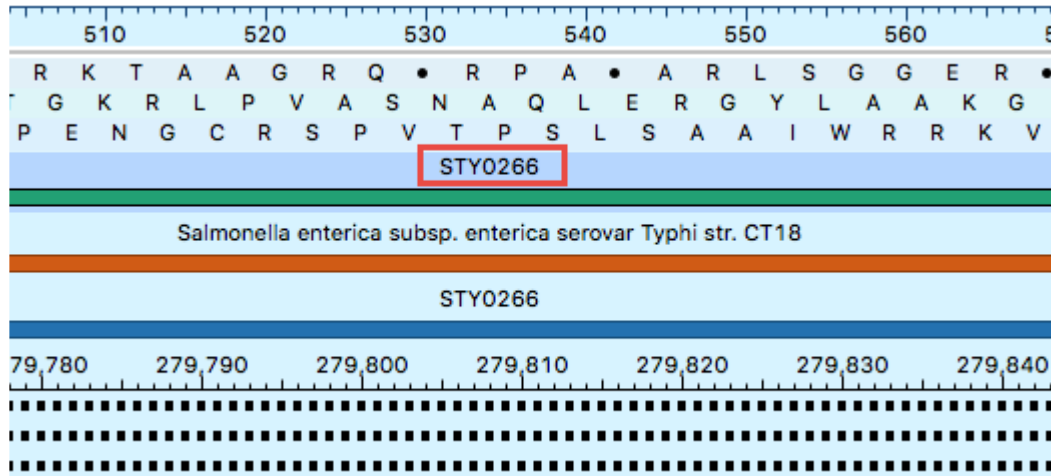


To find more clues to the identity of the unknown gene, you will next investigate the CT18 reference sequence annotations at this position.

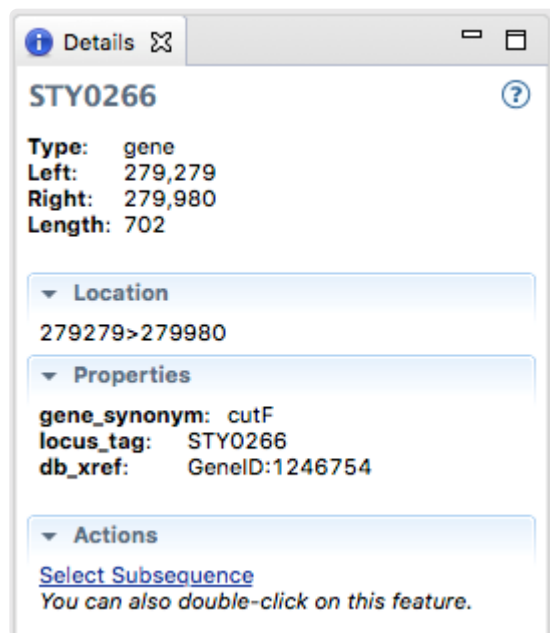
7. Collapse the tracks for the unknown sequence by clicking on any of the minus signs corresponding to that sequence.
8. Expand the detail tracks for the reference (CT18) by clicking any of the corresponding plus signs for that sequence.

Observe that the indel's position overlaps a gene and a CDS that are identified as *STY0266*.

9. Click anywhere on the *STY0266* feature to select it.



10. 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*.

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 data to MegAlign](#)
- [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

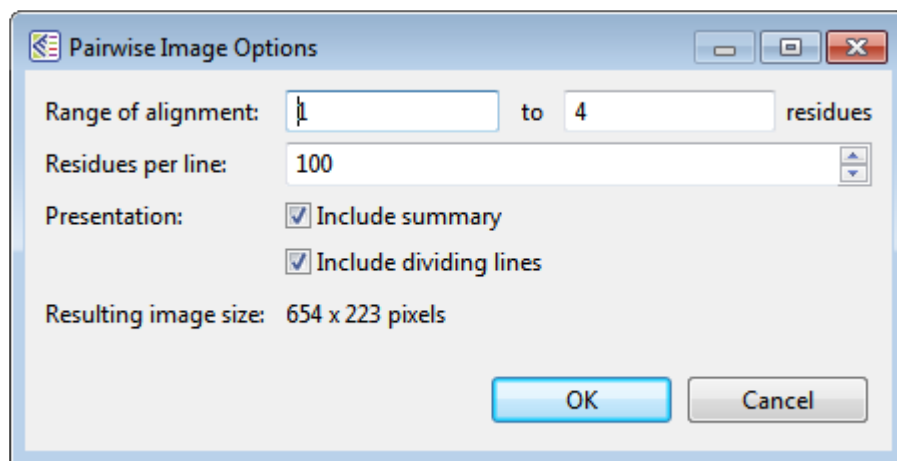
To export an image of a view :

1. Choose **File > Export Image > (View Name)**.
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 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.

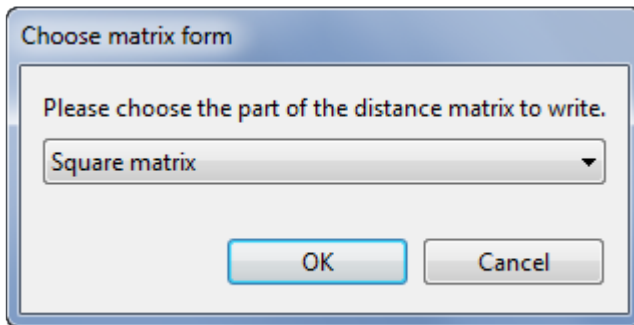
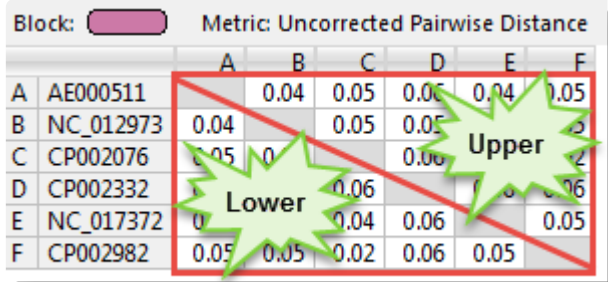
4. Click **OK** to export the image using the specified options or **Cancel** to close the dialog without exporting.

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](#).

Item to export	Command	Notes / Export format(s)
All sequences in the project	File > Export Data > Sequences	Various GenBank and FASTA formats; MegAlign format (<i>.meg</i>); or Nexus formats (<i>.nex</i> , <i>.nxs</i>). 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.”
Selected aligned sequences	File > Export Data > Aligned Sequences	
Selected portion(s) of the aligned sequences	File > Export Data > Aligned Subsequences	
Selected aligned sequences in the active block (requires Mauve alignment)	File > Export Data > Active Block Sequences	
Selected aligned sequences with gaps removed	File > Export Data > Aligned Sequences Without Gaps	Various GenBank and FASTA formats; MegAlign format (<i>.meg</i>); or Nexus formats (<i>.nex</i> , <i>.nxs</i>). Gaps are removed.
Aligned sequences in the active block with gaps removed (requires Mauve alignment)	File > Export Data > Active Block Sequences without Gaps	Various GenBank and FASTA formats; MegAlign format (<i>.meg</i>); or Nexus formats (<i>.nex</i> , <i>.nxs</i>). Gaps are removed. If you wish to make a new MegAlign Pro (<i>.msa</i>) file based on the active sequence block, use this command. Later, you can add the exported sequence to a MegAlign Pro project and use it as one of the sequences in a multiple alignment.
Consensus sequence from the Sequences view	File > Export Data > Aligned Consensus	Various GenBank and FASTA formats; MegAlign format (<i>.meg</i>); or Nexus formats (<i>.nex</i> , <i>.nxs</i>). 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.”
Consensus sequence from the active block of the Overview (requires Mauve alignment)	File > Export Data > Active Block Consensus	
Selected portion(s) of the aligned sequences in the active block (requires Mauve alignment)	File > Export Data > Aligned Active Block Subsequences	
Consensus sequence from the Sequences view with gaps	File > Export Data > Aligned Consensus	

removed	Without Gaps	
Consensus sequence from the active block of the Overview with gaps removed (requires Mauve alignment)	File > Export Data > Active Block Consensus without Gaps	
Phylogenetic tree information from the Tree view	File > Export Data > Tree	<p>Newick (<i>.newick</i>) format or Nexus formats (<i>.nex</i>, <i>.nxs</i>). These file extensions can be opened by many free and commercial evolutionary biology applications. Visit this page of our website to download a viewer. To set the default application for opening saved trees, see Set application preferences.</p> <p>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.</p>
Matrix information from the Distance view	File > Export Data > Distance Matrix	<p>Comma- or tab-separated formats (<i>.csv</i> or <i>.tab</i>), PHYLIP format (<i>.dist</i>) or Nexus formats (<i>.nex</i>, <i>.nxs</i>).</p> <p>After selecting a name and format, click Save to launch the Choose Matrix Form dialog:</p>  <p>Use the drop-down menu to choose whether to save the entire ("square") matrix or only the upper or lower triangle.</p>  <p>Note: If you choose to save in <i>.dist</i> format, the sample names in the saved file will automatically be truncated to 10 characters.</p>
Aligned sequences, distance table and phylogenetic tree	File > Export Data > All Data	Nexus formats (<i>.nex</i> , <i>.nxs</i>) only.

information		
-------------	--	--

Export data to MegAlign

You can export sequences and alignments to classic MegAlign, where you may realign using a different method, perform bootstrapping, or view alignment reports.

Timing	Command	Notes
Before performing an alignment	File > Send Sequences to MegAlign	Creates and opens a new MegAlign project containing the same sequences and sequence order as in the MegAlign Pro project.
After performing an alignment	File > Send Aligned Sequences to MegAlign	Sequences retain their MegAlign Pro alignment.
After performing an alignment and selecting a portion of the consensus	File > Send Aligned Subsequences to MegAlign	Sequences retain their MegAlign Pro alignment.
After performing a Mauve alignment and selecting an alignment block in the Overview	File > Send Active Block to MegAlign	Exports only the sequences in the active block.

Export a tree to a tree viewer

As described in [Export data to a file](#), Phylogenetic trees can be exported in Newick (*.newick*) or Nexus (*.nex*, *.nxs*) formats using **File > Export Data > Tree**. The saved files can be opened by many free and commercial evolutionary biology software programs.

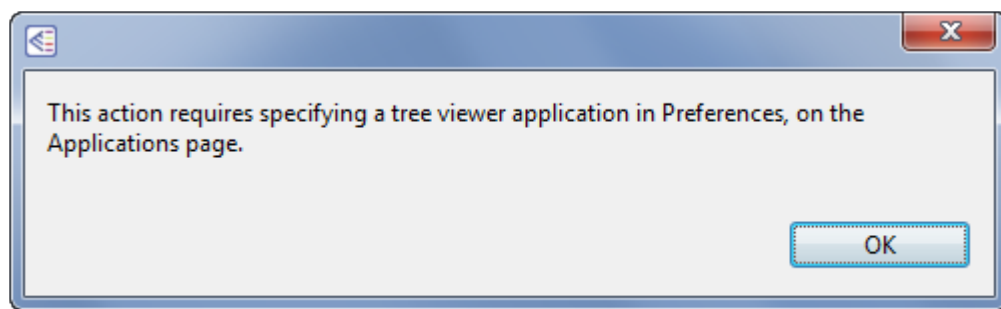
The following procedure goes a step further. It lets you automatically launch a third-party tree viewer and view the tree within that application.

To export and open the phylogenetic tree in a third-party tree viewer:

1. Outside MegAlign Pro, download the desired tree viewer and install it on your computer. Compatible viewers include [Archaeopteryx](#) (Win only) and [FigTree](#).

✿ **Note:** When installing FigTree on a Macintosh computer, you may receive a message that the application is damaged. You may need to change your Security & Privacy settings to allow applications downloaded from “Anywhere” prior to installation.

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

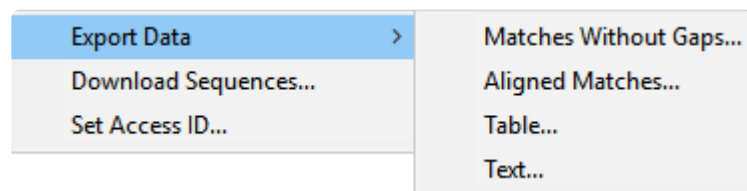
After using **File > Rename** or **File > Rename with Fields**, the custom names are only exported as part of the sequence data in certain cases. Raw data sequences imported into MegAlign are never modified in any way, and always retain their original names.

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.



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.

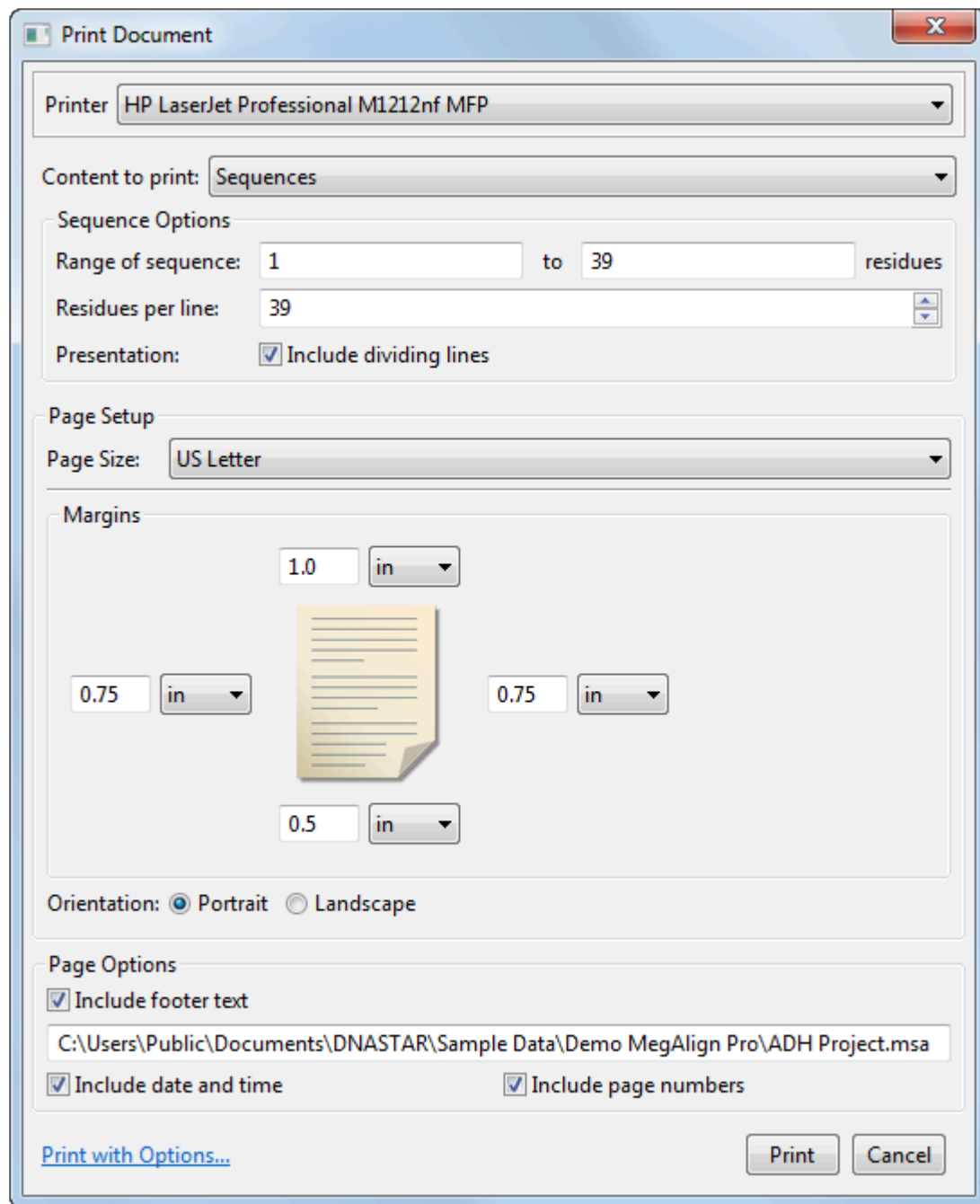
The commands above all launch a Save As dialog. Choose a file name, location and file type (fixed, in some cases) then press **Save**.

- **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

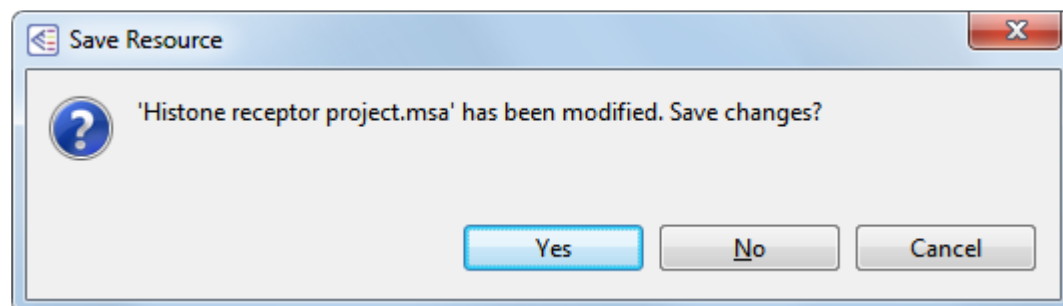
To save, close, or exit:

The following table describes how to save files, close projects and exit from MegAlign Pro:

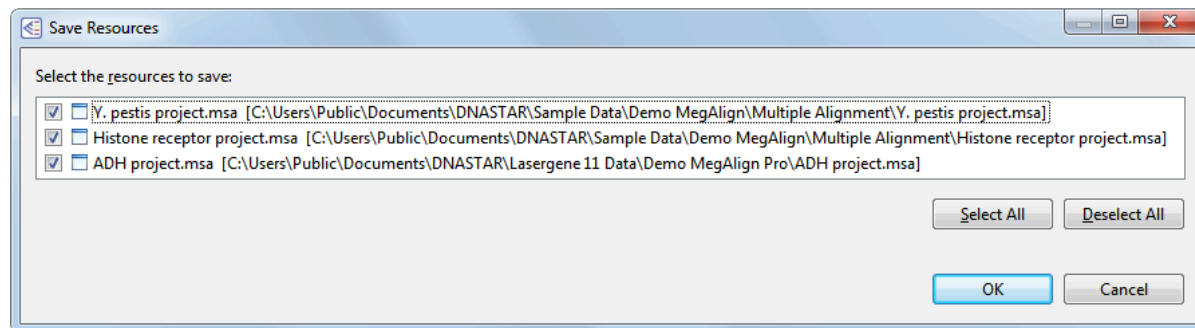
Task	How to...
To save all of the open projects simultaneously	Choose File > Save All or press Ctrl/Cmd+Shift+S .
To save a new project in multiple sequence alignment (.msa) format, or to overwrite a previously saved version of the project	Choose File > Save , select the Save tool () , or use Ctrl/Cmd+S . 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.
To close an active project	Choose File > Close or Ctrl/Cmd+W . If multiple projects are open, first select the tab corresponding to the project you wish to close.
To 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.

Closing or quitting 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 **Yes** to save changes, **No** to close/quit without saving changes, or **Cancel** to return to MegAlign Pro without closing/quitting.



If you quit MegAlign Pro while multiple unsaved projects are open, MegAlign Pro will instead display the Save Resources dialog.



- To save all changed projects, choose **Select All**. To close all changed projects without saving the updates, pick **Deselect All**. To save only some of the projects, add/remove checkmarks individually.
- Click **OK** to save only the checked projects and to quit MegAlign Pro. Click **Cancel** to return to MegAlign Pro without saving the checked projects.

Appendix

Turn off usage logging

By default, [usage logging](#) is enabled in Lasergene version 11 and later. To opt out of usage logging, launch the DNASTAR Navigator and go to **View > Preferences > Lasergene** (Win) or **DNASTAR Navigator > Preferences > Lasergene** (Mac). Uncheck the box and click **Apply** and then **OK**.

Set preferences

You can set preferences in MegAlign Pro for [application](#), [Internet](#), and [temporary files](#).

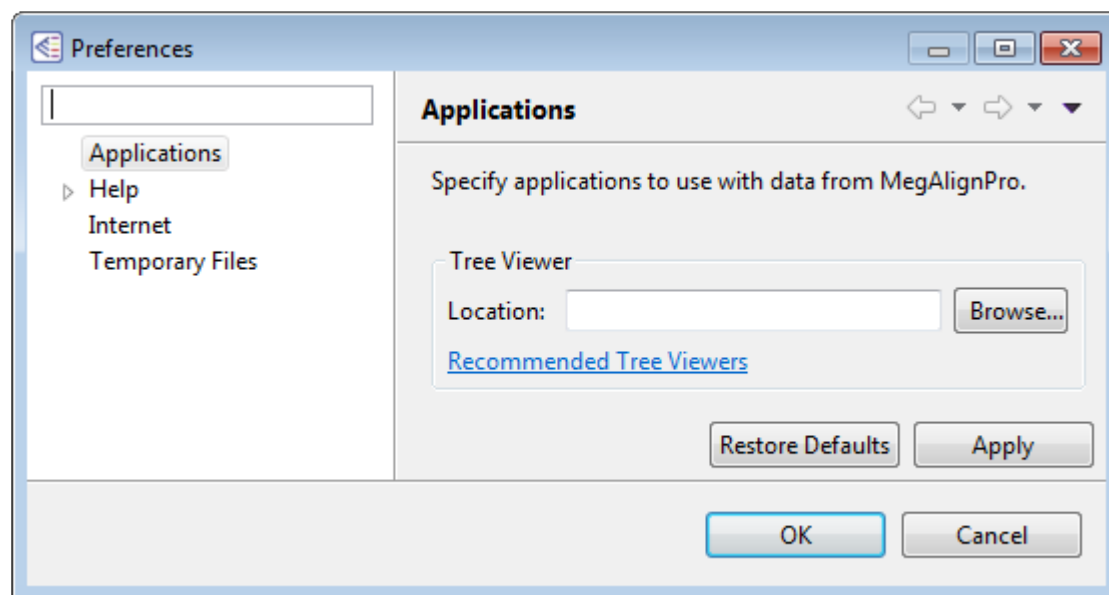
To open the Set Preferences dialog:

On Macintosh, choose **MegAlign Pro > Preferences**. On Windows, choose **Edit > Preferences**.

Set application preferences

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.

To access the dialog on Windows, use **Edit > Preferences**. On Macintosh, use **MegAlign Pro > Preferences** or **Cmd+,**. If the Applications section is not already active, click on "Applications" on the left.



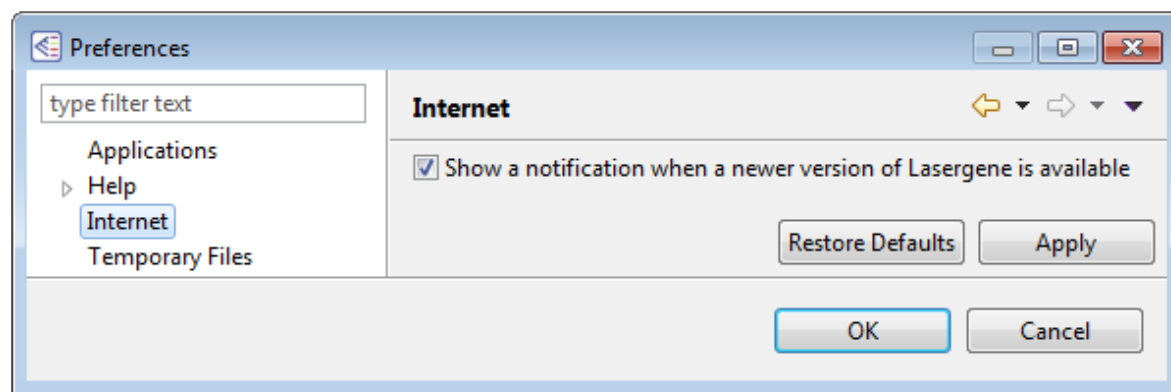
- By default, the **Location** box is empty. If you already have tree viewing/editing software on your machine, use the **Browse** button to navigate to the application file, and then click **OK**. Most software that can read a Newick- and/or Nexus-formatted tree file will work here.
- If you would like 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.

Click **Apply** and **OK** to save your changes or **Cancel** to ignore any changes. Click **Restore Defaults** to return to the default (temporary) storage location.

Set Internet preferences

The Internet Preferences dialog lets you choose whether you wish to be notified about Lasergene updates.

To access the dialog, use **Edit > Preferences** (Win) or **MegAlign Pro > Preferences** (Mac). If the Internet section is not already active, click on the word "Internet" on the left.



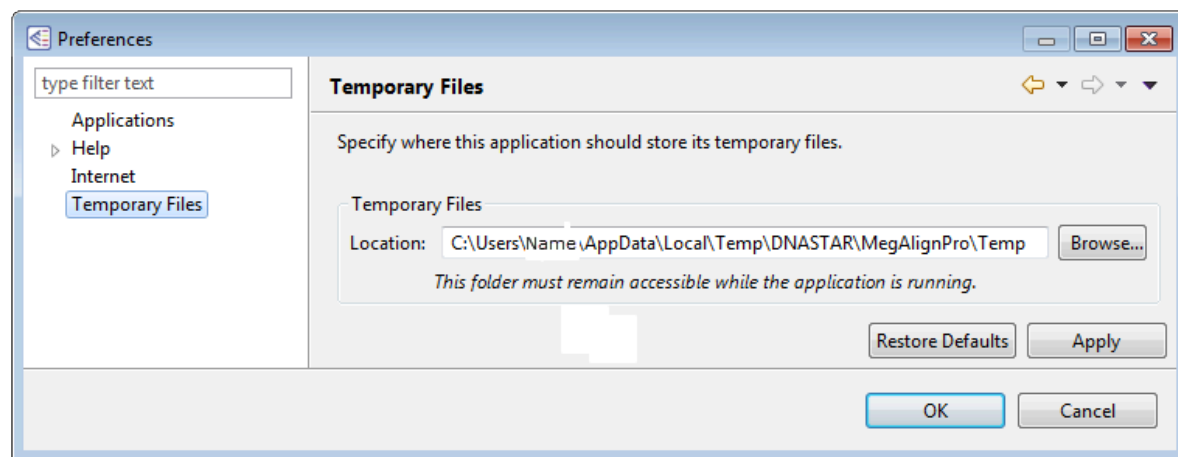
- Check the box if you would like MegAlign Pro to display a notification when a newer version of Lasergene is available. Uncheck the box if you do not wish to receive this notification.

Click **Apply** and **OK** to save your changes or **Cancel** to ignore any changes. Click **Restore Defaults** to return to the default settings for all four drop-down menus.

Set temporary file preferences

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 access the dialog, use **Edit > Preferences** (Win) or **MegAlign Pro > Preferences** (Mac). If the Temporary Files section is not already active, click on the phrase "Temporary Files" on the left.



- 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.

Click **Apply** and **OK** to save your changes or **Cancel** to ignore any changes. Click **Restore Defaults** to return to the default (temporary) storage location.

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 (<i>DMx</i> , <i>DMx.exe</i>)	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 (<i>Iservrc</i>), Server License Manager ² , <i>Server Executables</i> (<i>_Iserv</i> , <i>Iservnt.exe</i>)	All	Windows: C:\Program Files (x86)\DNASTAR- LicenseServer\Server Macintosh: ~/Library/ DNASTAR-LicenseServer
Standalone & Trial Licenses (<i>*.license</i>), License Server Client License (<i>*.lshost</i>), Key Server Client License (<i>*.keyhost</i>)	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

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.

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](#)).

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]

Henikoff S & Henikoff JG (1992). “Amino Acid Substitution Matrices from Protein Blocks”. PNAS. 89 (22): 10915–10919. (See the [PubMed entry](#))

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.

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](#))

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](#)

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]

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]

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](#))

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]