Multiple Sequence Alignment
Building Multiple Sequence Alignment
Lec4: Coloring MSA for Publication
Preparing Your Multiple Alignment for Publication

- You want to show this alignment to your colleagues, you want to include it in publications

- In short, you need a high-impact picture to convince people that your research is going well.

Introducing the Multiple Sequence Alignment Viewer (MSAV)
The MSA Viewer

- A general purpose alignment viewer
- Useful for assessing sequence conservation, sequencing errors or variations
- Flexible options for coloring alignments
- Plugs into NCBI alignment tools (BLAST, COBALT)
- Also accepts custom alignment data
- Easily embedded in your own web pages

\[ <\text{ncbi}>/\text{tools/msaviewer}/ \]
Display Feature: Annotations

(+) to see annotated:
- Domains
- Binding sites
- Motifs
- Other features

Requires NCBI accessioned sequences
Display Feature: Colors

Tools Menu
Colors emphasize amino acid properties conservation column average compared to Master

- BLOSUM45
- BLOSUM62
- BLOSUM80

Conservation
Column Quality score - Protein
Show Differences
Frequency-Based Difference
Hydropathy Scale
Membrane preference
Rasmol Amino Acid Colors
Shapely Amino Acid Colors
Signal sequence
Size
Display Feature: Residue Context

Click amino acid position in sequence
position in alignment
sequence statistics
column statistics

Alignment View: 286 - 341 (56r shown)

Descriptions
Seq Start
NP_000612.1
NP_001005349.1
NP_001001157.1
NP_0099950.1
NP_001269288.1
NP_001126918.1
NP_001075253.1
NP_001028138.1
NP_001005869.1
NP_000882.1
NP_001233657.1
NP_765400.1

Seq End
Organism
Homo sapiens

Alignment Pos: 307
Sequence Pos: 307

Taxonomy: Mesocricetus auratus (golden hamster) [rodents]

Amino acid: Threonine
Matches: 5 of 12 rows (41.7%)
Mismatches: 2 of 12 rows (16.7%)
Gaps: 5 of 12 rows (41.7%)
Unaligned: 0 of 12 rows (0.0%)
Total rows: 12
Sequence is aligned from residues 1 to 471
Amino acids aligned: 471
Welcome to NCBI Multiple Sequence Alignment Viewer 1.5.2

To get started click on the ? at the upper right to review the Help documentation.

Example Links:

Protein alignment, no master
This link provides a look at protein multiple alignment with no master sequence set.

Protein alignment, master set to AC128628
This link provides a look at protein multiple alignment with master sequence set to one of the proteins, AC128628.

MUSCLE alignment
This link provides a look at protein multiple alignment resulting from a run of MUSCLE program.

Alignment from Blast RID
Alignment from Blast RID
Alignment from ID data
Trace assembly
Another alignment with master
And another one
And another one

To see your own alignment, Upload your data

You can also embed Multiple Sequence Alignment Viewer on your own page.

This is an example of static embedding.
This is an example of dynamic embedding.
This is an example of interaction with SV on the same page.
This is an example of MSAV event handling.
This is an example of different coloring options.
This is an example of Alignment Coordinate transformation.
Multiple Sequence Alignment Viewer 1.5.2

Alignment

Protein: 1 - 361 (361 shown)

<table>
<thead>
<tr>
<th>Description</th>
<th>Seq Start</th>
<th>Alignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC28628</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AC27343</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AC27341</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MY2724</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AM24532</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Contact | Copyright | Disclaimer | Privacy | Accessibility
National Center for Biotechnology Information, US National Library of Medicine
8600 Rockville Pike, Rockville, MD USA 20894

Introduction to Bioinformatics Online Course: IBT
Multiple Sequence Alignment | Prof. Ahmed M. Alzohairy
Welcome to NCBI Multiple Sequence Alignment Viewer 1.5.2

To get started click on the ? at the upper right to review the Help documentation.

Example Links:

- Protein alignment, no master
  This link provides a look at protein multiple alignment with no master sequence set.

- Protein alignment, master set to AC128628
  This link provides a look at protein multiple alignment with master sequence set to one of the proteins, AC128628.

- MUSCLE alignment
  This link provides a look at protein multiple alignment resulting from a run of MUSCLE program.

- Alignment from Blast RID
- Alignment from Blast RID
- Alignment from ID data
- Trace assembly
- Another alignment with master
- And another one
- And another one

To see your own alignment, Upload your data

You can also embed Multiple Sequence Alignment Viewer on your own page.

This is an example of static embedding.

This is an example of dynamic embedding.

This is an example of interaction with SV on the same page.

This is an example of MSAV event handling.

This is an example of Different coloring options.

This is an example of Alignment Coordinate transformation.
BLAST returns separate alignments for each query, and these separate alignments can further be ordered into sets offering consistent non-overlapping query and subject coverage. The sequence viewer offers the ability to evaluate the original BLAST hits on the fly and link together alignments that meet a strict definition of non-overlapping query and subject coverage.
Using Boxshade

Boxshade is a utility that allows you to put some life into your alignment. It shades columns according to their level of conservation and produces files that you can easily manipulate for inclusion in reports or articles.

BoxShade Server
Pretty Printing and Shading of Multiple-Alignment files. This server uses version 3.21 of BOX SHADE, written by K. Hofmann and M. Baron. BOX SHADE is in the ...

BoxShade Server - EMBnet node Switzerland
BOX SHADE is a program for pretty-printing multiple alignment output. ... running here at the ISREC in Lausanne/Switzerland using the URL: BOX_form.html

boxshade - BioMed Central
Boxshade is a program for creating visually pleasing images of multiple alignments of protein ... Homepage, http://www.ch.embnet.org/software/BOX_form.html.

How to generate a publication-quality multiple sequence ... [PDF]
alignment using another program called Boxshade. • Download the
Pretty Printing and Shading of Multiple-Alignment files

This server uses version 3.21 of BOXSHADE, written by K. Hofmann and M. Baron. BOXSHADE is in the public domain and available from Source Forge http://sourceforge.net/projects/boxshade/

This server takes a multiple-alignment file in either GCG's MSF-format or Clustal ALN-format. Output can be created in the following formats:

- Postscript/EPS (using shaded background)
- RTF old (using colors)
- RTF new (using shaded background)
- XFIG-files (using shaded background)
- ASCII (showing similarities)
- ASCII (showing differences)
- HPGL (using colors)
- PICT (for later editing on MACs and PCs)

If you have problems using this server (like getting no result), read this and see the FAQ list.

<table>
<thead>
<tr>
<th>Output format</th>
<th>Postscript_portrait ▼</th>
</tr>
</thead>
<tbody>
<tr>
<td>Font Size</td>
<td>10 ▼</td>
</tr>
<tr>
<td>Consensus Line</td>
<td>no consensus line ▼</td>
</tr>
</tbody>
</table>

Fraction of sequences that must agree for shading: 0.5 ▼
When pasting MSF or ClustalW files, please make sure that the pasted text starts with the header line of the alignment and contains no extra blank lines at the bottom.
Generates a file that most word processors (such as Microsoft Word) can read.

If your alignment is long, select a small font size.

Consensus sequence that contains the most common amino acid for each column.
If you request 0.5, for example, it means you want half the residues to be conserved for some shading to occur.

Select the format of the multiple sequence alignment

Input sequence format

CLUSTAL FORMAT for T-COFFEE Version_4.26
[http://www.tcoffee.org], CPU=3.97 sec, SCORE=69, Nseg=9, Len=158
BoxShade

BOXSHADE result for

BOXSHADE has now created the output file that can be downloaded.

The command line used was:
ulimit -t 30; /mnt/local/bin/boxshade -def -numdef -out=wwwtmp/1.BOX.2744.4160.txt -in=wwwtmp/1.BOX.2744.4160.ali -par=pt10.par -type=2 -dev=4 -thr=0.5 >/dev/null2410 423

Here is your output number 1

Important note: If you get an error message when clicking on the link above Please read this page
Conservation doesn’t necessarily mean identity in Boxshade. Similar residues, such as isoleucine and valine, also account for conservation. Two types of shading exist:

- **Black**: Identical amino acids or nucleotides
- **Gray**: Similar amino acids
MView for colorful MSA
Introduction to Bioinformatics Online Course: IBT

Multiple Sequence Alignment | Prof. Ahmed M. Alzohairy
Practical
(to try in your own time)

1. Point your browser to www.ch.embnet.org/software/BOX_form.html.
2. Choose RTF_new from the Output Format drop-down menu.
3. Select the font size you want.
4. Choose Add a Consensus Line with Letters from the Consensus Line drop-down menu.
5. Select the fraction of sequences you would like shaded.
6. Select the format of the multiple sequence alignment you want to use.
7. Paste your multiple sequence alignment into the Sequence window.
8. Click the Run Boxshade button.
9. Click the here is your output link, save to a local file, then open the local file with Word.
Introduction to Bioinformatics Online Course: IBT

Multiple Sequence Alignment | Prof. Ahmed M. Alzohairy
By
Ahmed Mansour Alzohairy
Department of Genetics, Zagazig University, Zagazig, Egypt