Supporting online material for: MSAViewer: interactive JavaScript visualization of multiple sequence alignments

Guy Yachdav, Sebastian Wilzbach, Benedikt Rauscher, Robert Sheridan, Ian Sillitoe, James Procter, Suzanna Lewis, Burkhard Rost &Tatyana Goldberg

Material

Table SOM_1: Comparison of MSAViewer features with popular alignment analysis tools.

Features	***MSAViewer[1]***	***Jalview Desktop[2]***	***Jalview Lite[3]***	***JSAV[4]***	***alignmentviewer.org[5]***	***ProViz[6]***	***Wasabi[7]***
GENERAL INFORMATION							
Programing language	JavaScript	Java	Java	JavaScript	Java	JavaScript	JavaScript (frontend), Python (backend)
Web compatible	yes	no	yes	yes	yes	yes	yes
Extendable API by a developer	yes	no	no	yes	no	no	yes
DATA IMPORT							
Data source	File, URL, drag & drop	File, URL, textbox	Textbox	Array of JavaScript objects	File, drag & drop	File, URL, drag & drop, textbox	File,URL, drag & drop
Alignment formats	FASTA, Clustal	FASTA, Clustal, MSF, PFAM,PIR, BLC, AMSA, JalView, Stockholm, RNAML, PHYLIP	FASTA, Clustal, MSF, PFAM,PIR, BLC, AMSA, JalView, Stockholm, RNAML, PHYLIP	n.a.	FASTA, Stockholm, plain text	FASTA, UniProt term triggers a visualization of a precomputed alignment	FASTA, Clustal, Phylip, NEXUS, HSAML and through Ensembl REST API
Sequence annotation formats	GFFv3 (non-hierarchical), JalView	JalView, GFFv1	JalView, GFFv1	n.a.	Tab separated text file	Automatic retrieval of annotations from various resources	n.a.
Phylogenetic tree format	Newick	Newick	Newick	n.a.	n.a.	n.a.	Newick, extended Newick, NEXUS, HSAML and through Ensembl REST API
DATA EXPORT							
Alignment formats	FASTA	FASTA, MSF, CLUSTAL, BLC, PIR, PFAM, JalView, AMSA, STH, Phylip	FASTA, MSF, CLUSTAL, BLC, PIR, PFAM, JalView, AMSA, STH, Phylip	FASTA	n.a.	FASTA	FASTA, Phylip, NEXUS
Sequence annotation formats	GFF	GFF, JalView	JalView	n.a.	n.a.	n.a.	n.a.

Image	PNG	HTML, EPS, PNG, SVG	n.a.	n.a.	n.a.	PDF	n.a
URL for submission to another web site	yes	Includes web services for bioinformatic data analysis	n.a.	yes	no	yes	yes
CONSENSUS SEQUENCE							
Mode of calculation	Most frequent residue per column	Most frequent residue per column	Most frequent residue per column	Input sequence most similar to other sequences (centroid sequence)	First sequence of the alignment	First sequence of the alignment	n.a.
NAVIGATION THROUGH THE ALIGNMENT							
Overview panel	yes	yes	yes	no	yes	yes	no
Mode of navigation	Scrolling, "jump to a column", panning	Scrolling, keyboard shortcuts, "Jump to sequence position"	Scrolling, keyboard shortcuts, "Jump to sequence position"	Scrolling	Scrolling	Scrolling, keyboard shortcut on a feature, specified range of resides	Scrolling
Motif search (regular expression)	yes	yes	yes	no	no	yes	no
Sequence sorting by ...	Sequence identifiers, sequence names, sequences,% of gaps, % identity to consensus	Sequence names, length, group, pairwise identity, sequence features (scores, counts), associated tree, alignment annotation scores	Sequence names, length, group, pairwise identity, sequence features (scores, counts), associated tree, alignment annotation scores	Sequences and regions of sequences	% identity to consensus.	Manual reordering, restriction to model organisms	Reordering through rearranging or pruning phylogenetic tree branches
COLOURING/ HIGHLIGHTING/ HIDING							
Flexible coloring schemas	yes	yes	yes	yes	no	no	yes
Hide columns by …	Selection, % identity to consensus, % gaps	Selection and/or annotation	Selection and/or annotation	n.a.	n.a.	Gaps in the query sequence	% gaps
Hide sequences by …	Selection, % identity to consensus, % gaps	Selection	Selection	Selection	% identity to consensus, % gaps	Selection	Selection

ADDITIONAL VISUALIZATION ELEMENTS							
Sequence logo	yes	no	no	no	yes	no	no
Nucleotides/ amino acids conservation scores	yes	yes	yes	no	no	yes	no
Occupancy per column (gap weights)	yes	no	no	no	no	no	no
Phylogenetic tree	yes	yes	yes	no	no	no	yes
ADDITIONAL FEATURES							
Recognition of identifiers for linking to external databases	UniProtKB, GenBank	EMBL-EBI	EMBL-EBI, PDB	n.a.	UniProtKB	UniProtKB	Ensembl GeneID, GeneTreeID

[1] MSAViewer is our JavaScript-based tool described in the main text.

[2] Jalview Desktop (Waterhouse, et al., 2009) is an example of a powerful and a highly used desktop application.

[3] Jalview Lite (Waterhouse, et al., 2009) is an example of a lightweight Java applet for use in web applications.

[4] JSAV (Martin, 2014) is an example of a recently developed JavaScript-based tool.

[5] http://www.alignmentviewer.org/ is an example a of server side-rendered graphics.

[6-7] ProViz (Jehl, et al., 2016) and Wasabi (Veidenberg, et al., 2016) are examples of backend server-based tools.

Section SOM_1: MSAViewer implementation

The MSAViewer is a JavaScript application that renders an MSA as a Canvas element. Canvas is a web technology adopted by modern web browsers that renders graphical presentation without relying on browser plugins. Whereas other drawing technologies, like Scalable Vector Graphics (SVG), strain the browser memory by maintaining an object model for each rendered graphical element, Canvas objects are drawn on the fly, and are thus a more efficient approach for visualizing large data sets. By using the Canvas to render only the visible part of the alignment, fewer resources are needed. Basic support for Canvas is now available in all major browsers (http://caniuse.com/#search=canva). To support the integration of the tool with other web applications, the viewer was designed with modularity in mind allowing the separation of user controls from the MSA presentation. Programmatic access to the functionalities of each module further enables integration. These impliemntation decisions imrpoved the performance of the MSA viewer as compared to existing JavaScript multiple sequence alginment viewers. A perofrmnace survey of existing JavaScript-based MSA viewers is given in Supplementary Table_SOM2.

Table SOM_2: Performance profiling of MSAViewer against JSAV.

PFAM family	*PF15698 (150 proteins)*[8]		*PF01437 (3239 proteins)* [9]		*PF00982 (9717 proteins)* [10]	
Method *Task*	*MSAViewer*	*JSAV*	*MSAViewer*	*JSAV*	*MSAViewer*	*JSAV*
Loading[1]	1.22	50.55	1.09	1796.67	1.07	N/A
Scripting[2]	312.66	266.04	86.27	1153.33	361.93	N/A
Rendering[3]	22.50	484.65	1.40	19733.33	1.18	N/A
Painting[4]	10.86	174.97	5.98	366.56	4.22	N/A
Other[5]	40.60	141.09	52.72	828.09	340.93	N/A
Idle[6]	2773.33	2803.33	3030.00	3013.33	3083.33	N/A
Total[7]	3161.17	3920.64	3177.47	26891.32	3792.66	N/A

MSAViewer and JSAV (Martin, 2014) are JavaScript-based applications. While MSAViewer uses Canvas, JSAV uses DOM objects to render and draw the alignment. To eliminate the effect of network speed and data file transfer times, we downloaded all source files and data files, and used them from the same disk drive. Performance timings showed in the table are in milliseconds and were collected from the event categories in the "Timeline" tab of the Google Developer Tools kit. These event categories include:

[1] *Loading:* parse HTML, finish loading, receive data, receive response and send request.
[2] *Scripting:* animation, garbage collection, script evaluation, function call and event handling.
[3] *Rendering:* layout, recalculate style, scroll and invalidate layout (triggered by an event that modified the document's layout).
[4] *Painting:* paint, composite layers and image handling.
[5] *Other:* events not included in the above categories.
[6] *Idle:* time spent on doing nothing.
[7] *Total:* total aggregate time. Note that due to events taking place in parallel, the actual time to complete a presentation is usually shorter than total aggregate time.

Data: We report the timing profiles (averages of 3 runs per test case) for MSAViewer and JSAV using the following alignments extracted from the PFAM database (Finn, et al., 2014), release October 2015:

[8] *PF15698 : Phosphatase:150 sequences (1% of PFAM families are of this size)*
[9] *PF01437: Plexin repeat: 3239 sequences (40% of PFAM families are of this size)*
[10] *PF00982: Glycosyltransferase family 20: 9717 sequences (60% of PFAM families are of this size)*

Note: for the alignment PF00982 (9717 sequences) we could not produce timing profiles for JSAV and concluded that the viewer has exceeded the amount of memory allowed by the browser.

The performance profile was conducted on a system with the following hardware/software specifications:

Model: Apple MacBook Pro 13-inch, Mid 2012
OS: OS X 10.9.5 (13F34)
CPU: 2.5 GHz Intel Core i5
RAM 2X4GB DDR 1600Mhz
Disk APPLE HDD HTS545050A7E362 (5400 RPM)
Software: Google Chrome Version 46.0.2490.80 (64-bit)

References for Supporting online material

Finn, R.D.*, et al.* (2014) Pfam: the protein families database, *Nucleic acids research*, **42**, D222-230.

Jehl, P.*, et al.* (2016) ProViz-a web-based visualization tool to investigate the functional and evolutionary features of protein sequences, *Nucleic acids research*.

Martin, A.C. (2014) Viewing multiple sequence alignments with the JavaScript Sequence Alignment Viewer (JSAV), *F1000Research*, **3**, 249.

Veidenberg, A., Medlar, A. and Loytynoja, A. (2016) Wasabi: An Integrated Platform for Evolutionary Sequence Analysis and Data Visualization, *Molecular biology and evolution*, **33**, 1126-1130.

Waterhouse, A.M.*, et al.* (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench, *Bioinformatics*, **25**, 1189-1191.